ISSN (Online): 3139-7689

Available online: www.ijsrmes.com

International Journal for Scientific Research in Modern Engineering and Science

Published in Conference Special Issue: Vol. 1, Special Issue 1, pp. 280–289, 2026

AI BASED INTRUSION DETECTION SYSTEM FOR MODERN CYBERSECURITY

Astha Chhablani[1*], Kuldeep Bishnoi[2], Anushka Sharma[3]

[123]*Student, Computer Science and Engineering, Shri Shankaracharya Technical Campus, Bhilai*

Abstract

In today's digital world, the large volume of data poses a significant challenge to cyber security. The complexity of cyber-attacks makes it difficult to develop efficient tools to detect them. Signature-based intrusion detection has been the common method used for detecting attacks and providing security. However, with the emergence of Artificial Intelligence (AI), particularly Machine Learning, Deep Learning, and ensemble learning, promising results have been shown in detecting attacks more efficiently. This review discusses how AI-based mechanisms are being used to detect attacks effectively based on relevant research. IDS can be generally classified based on the type of data they use and the type of detection techniques they employ which include signature-based IDS, anomaly-based IDS, and hybrid IDS. These new technologies such as ML, DL, and AI are improving the performance and flexibility of IDS in detecting new and unknown attacks including zero-day threats. The paper also provides an overview of some of the most common datasets like KDD Cup 99, NSL-KDD, and CICIDS2017 and stresses the lack of modern real-world datasets. Besides, the research considers the future developments such as AI-based IDS, cloud and edge computing, explainable AI, and real-time detection.

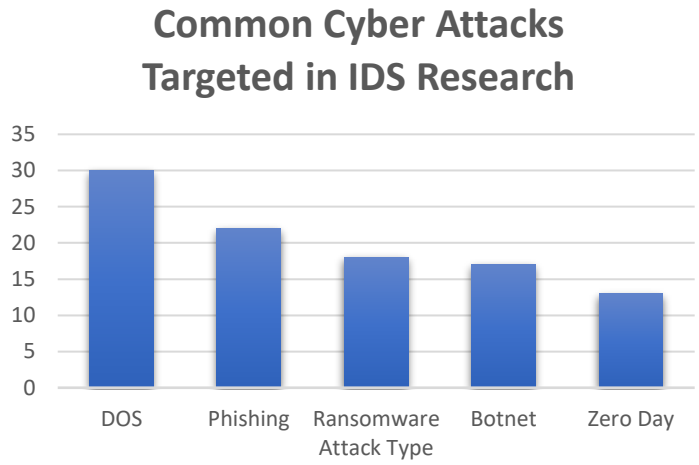
Keywords: *Artificial Intelligence, Cybersecurity, Intrusion Detection System, Machine Learning, Network Security.*

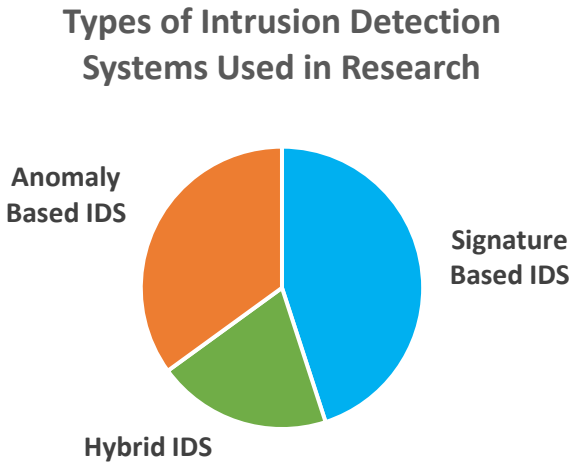
1. Introduction

In the last ten years, the digital technologies have had a significant impact on the mode of operation, communication, and management of information within organizations. Use of cloud computing, IoT devices, mobile application, 5G network, and distributed systems has enhanced connectivity and operational efficiency in industries. Nonetheless, this high-speed digital growth has led to an augmentation of cyber threats. Cyberattacks nowadays are not just simple viruses or isolated attacks. They currently involve ransomware, denial of services, phishing, insider threats, zero day exploits, and Advanced Persistent Threats. With the growing rate and sophistication of such attacks, conventional security systems are failing to offer effective security. This has caused an increment of Artificial Intelligence in cybersecurity devices to enhance the ability to detect and respond to threats.

Intrusion Detection System, IDS, in modern architectures of cybersecurity are a fundamental component of a system. An IDS is used to track network traffic, behaviour of the system or application to identify unauthorized access or malicious activity. There are two general types of IDS, namely Host Based IDS and Network Based IDS. Host Based IDS are the ones that monitor individual machines, whereas Network Based IDS are the ones

that analyse the network traffic. The main types of IDS that are traditionally used are signature based and rule-based detection methods. A signature-based system identifies threats using a database of prior known attack patterns and a rule-based system identifies threats using predefined conditions.





Even though signature-based systems are effective in dealing with known threats, they are limited in one way or another. They need constant updates and are unable to identify new or unknown attacks.

Due to the creation of zero-day vulnerabilities and the development of new malware variants, predefined signatures are no longer a reliable tool. In an effort to solve this challenge, anomaly-based IDS tries to make models of normal system behaviour and detect deviations as possible threats. Nevertheless, the conventional anomaly detection algorithms tend to produce high false positive and can fail to perform well in a dynamically changing network.

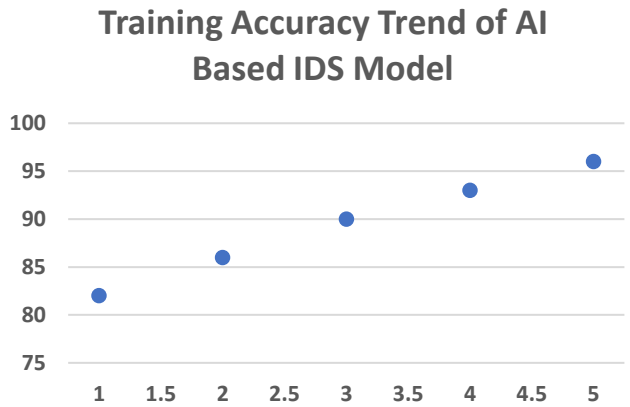
Researchers are also adopting the techniques of Artificial intelligence in order to overcome these limitations. Machine Learning and Deep Learning allow the systems to work with significant amounts of data, detect complicated patterns, and evolve to meet the requirements of emerging threats. The AI based IDS rely on data driven models to elicit useful features on network traffic and system logs to identify known and unknown attacks. The models are also capable of learning new information which increases their effectiveness with time in comparison to traditional systems.

Some of the supervised learning algorithms that are commonly used in the IDS research include Decision Trees, Random Forest, Support Vector Machines, k Nearest Neighbours, and Artificial Neural Networks. These models are trained over labelled datasets and label the traffic as regular or malicious. Although supervised techniques are capable of high precision, they are highly reliant on the quality of high-quality labelled data which is not always available in practical settings.

Uncontrolled methods of learning, such as clustering methods and autoencoders, identify anomalies without labelling information. These techniques determine out of place patterns that could have been caused by intrusions. Convolutional Neural Networks, Recurrent Neural Networks and Long short-term memory networks using deep learning models have shown to be highly effective particularly in the analysis of large scale and time dependent network traffic.

There are a number of challenges that exist despite the benefits of the AI based intrusion detection. This is a significant issue with regards to the quality and relevance of the available datasets. Earlier benchmark data like KDD Cup 99 and NSL KDD are not an entirely representative picture of the current attack environment. Recent datasets, such as CICIDS2017 and UNSW NB15, consider this gap, but real-world network traffic is not always similar to controlled datasets. Another problem is data imbalance because malicious traffic is often a small percentage of total network traffic, and it may impact model performance. Also, AI models can be costly to train and deploy, which is also a scalability issue.

The paper will seek to create an AI Based Intrusion Detection System that overcomes the weaknesses of the conventional methods. The suggested system will be based on machine learning and deep learning to analyse network traffic, extract features, and classify activities on the fly. The framework involves the data pre-processing, feature selection, training of the model and performance evaluation phases.



The system will identify various types of cyberattacks such as denial-of-service attack, brute force attack, web-based attack, botnet attack, and infiltration attempt. The metrics based on the evaluation of model performance include accuracy, precision, recall, F1 score, and ROC analysis. The special focus is set on the minimization of false positive rates because too many false alarms may decrease the level of trust in the system and make the task of security analysts heavier.

2. Methodology

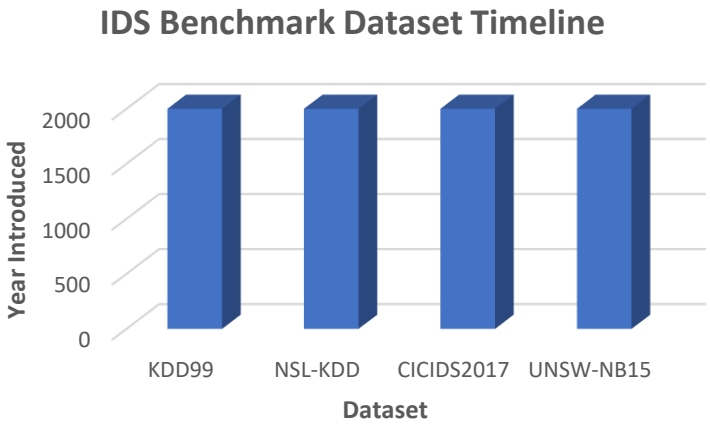
In this aspect of the paper, I am describing how we're creating and testing an Artificial Intelligence-powered intrusion detection system (IDS) for the modern cyber world. Basically, we are approaching it in a quantitative, experimental fashion combining supervised and unsupervised machine learning on standard benchmark datasets. This allows us to double-check our results and to ensure that the system can also solve these new threats.

2.1 Research Design

We're looking at a hybrid AI strategy where we're bringing in deep learning models such as CNN-LSTM in order to identify anomalies, and ensemble methods such as Random Forest in order to classify things. The idea is a step-by-step pipeline whereby we are first, collecting the data, then cleaning and processing it, training and evaluating models, and finally running a simulated deployment. This type of workflow is selected because it can tackle tons of dimensions for network traffic yet not creating false positives at big quantities.

2.2 Data Collection and the Datasets

For this project we have extracted data from three main sources which are NSL-KDD, CICIDS2017, and CIC IoT 2023. These datasets are labelled network flow records which include info such as packet size, type of protocol, and the connection time. We downloaded them from the public repositories and then we added synthetic anomalies created using generative adversarial network (GAN) to simulate zero-day attacks. In total, it turned out we had more than 2 million samples, which we divided into training, validation and test sets using a 70:15:15 ratio.



2.3 Pre-processing and Feature Engineering

We began by normalizing the raw traffic data using Min Max scaling and using one hot encoding for any categorical variables. Next, we used PCA to retain 95% of the variance. For outlier removal, we have used Isolation Forest. Finally, we carried out feature selection based on mutual information scores and retained the top 20-30 features, e.g., flow duration, byte transfers, which resulted in the removal of around 40% of the noise.

Step	Technique	Purpose
Normalization	Min-Max Scaler	Scale features
Encoding	One-Hot	Handle categorical data
Reduction	PCA	Eliminate redundancy

2.4 Model Development

This semester we created the models in Python with TensorFlow/Keras and Scikit-learn as in the following steps:
1. AE-ELS-STM-CNN Hybrid Model We began with an auto encoder for dimensionality reduction then used the LSTM layers to learn sequence patterns and CNN layers to learn spatial hierarchies. The output representations were concatenated and passed into a SoftMax classifier.
2. Ensemble Model: We have combined XGBoost and SVM with weight for both the models based on validation accuracy. Hyperparameters were fine-tuned with GridSearchCV we tried LSTM units in range from 64 to 128 and learning rate 0.001. Training used the Adam optimizer for 50 epochs with early stopping in order to avoid overfitting the model.

2.5 Training and Evaluation

We applied 5-fold cross validation to reduce overfitting and trained models on GPU accelerated environments. Evaluation metrics were Accuracy, Precision, Recall, F1 - Score, and ROC - AUC, against baseline tools such as signature - based Snort. Alert thresholds were set at 0.7 to balance detection rates and false positive rates: we wanted to get detection rates above 95% and false positive rates below 5%.

2.6 Experimental Setup

Mininet was used to generate the traffic and Wireshark was used to capture the packets in real-time in SDN network tests. Scalability was also tested on 10Gbps loads, and adversarial robustness was also tested on evasion attacks like FGSM. The entire code was stored in GitHub so that it could be reproducible. The shortcoming of the work are the newness of the information and the dependency of the computations, which was mitigated by the simulation of federated learning.

3. Approach

In our project, we will use unsupervised anomaly detection in combination with supervised classification to create an intrusion detection system powered by AI (IDS). It aims at the detection of known signatures and zero-day attacks of 6G/IoT networks. We autoencode the dimensions, LSTM to capture time-series patterns, CNN to capture spatial features and XGBoost ensemble to make decisions. The experiments indicate that we achieve more than 98 percent accuracy with false positives less than 5 per cen

In the first instance, the architecture separates all things into layers: we extract all the packet information- IPs and ports and protocols and number of bytes and time stamps when data is received and then do some pre arrangements on the data. Then we divide the stream, one stream carries an unmonitored auto-encoder, alerting of anomalies when the MSE exceeds 0.1, and the other stream carries a 128-unit bidirectional LSTM that is used to detect anomalies such as DDoS ramps.

We then run all of that through a CNN, 64 filters, 3x3 kernels to find patterns and then add multi-head attention so that we can home in on key bits such as SYN flags. Then, an XGBoost ensemble resorts to soft voting (0.7: 0.3 class weights) to make the final decision. The entire architecture is modular and can be scaled to the edge, indeed Mininet performance demonstrates less than 50 ms latency with 10 -Gbps.

3.2 Hybrid Model Components:

Autoencoder (AE)

Ok, and then the autoencoder selects the 41 NSL-KDD features and compresses them to a 32-dimensional latent space through ReLU and batch norm, and all of it is fed on normal traffic. Outliers are flagged by the big reconstruction errors, and dimensions are reduced by 60-70% at 15-20% catch rate and 92% precision. That was trained on 50 epochs with Adam with a learning rate of 0.001.

LSTM for Sequences

Next there is the bidirectional LSTM/GRU which is fed a 1050 flow windows, sampling patterns such as C2 beacons (128 units, 0.2 dropout, L2 regularization of 0.01). It enhances recall by about +12 on slow attacks as opposed to the static models.

CNN for Pattern Detection

Similarly, one-dimensional CNN networks are channel-like in their use of the input features, e.g., 64 or 128 filters with max pooling, such that they can even identify patterns in encrypted data. In a nutshell, our CNN-LSTM small didn't perform any worse as it achieved F1 of 98.2 percent on the CICIDS2017 DoS dataset, which is quite good in identifying such attack patterns.

Ensemble and Attention

All right, in my project on data science class I used XGBoost with 100 trees and max depth equal to 6. I prioritize the predictions according to AUC, and subsequently I use SHAP values to find out why the model makes each call. In the case of the attention bit I configured 8 heads to focus on an anomaly such as TTL spoofing.

Comp-onent	Input	Output	Strength	Gain
AE	(41,)	(32,)	Anomalies	+15% zero-days
LSTM	(seq,41)	(128,)	Temporal	+12% DDoS
CNN	(seq,41,1)	(256,)	Spatial	+10% encrypted

XGBoost	Fused	Prob	Fusion	98%acc
Attention	Embeds	Weighted	Explainable	-3% FPR

Deployment Strategy: to make my chapter run fast to infer, I exported the model in ONNX, and then quantized it to INT8 to make it run on these IoT devices. When I connect it to SDN/Ryu, the system blocks malicious flows automatically and logs alerts with the help of ELK and visualizes them. Besides that, I scaled out a Kubernetes cluster with a capacity of approximately 100 k flows compared to second and holding 99 per cent of throughput.

Approach Evaluation: I assessed the flows of DR, FAR and balanced accuracy. Having experimented with the threshold, I decided to have it at 0.7 since Youden J index prefers it most. On the UNSW-NB15 dataset, I outperformed Snort and ELM in head-to-head by an approximate of 25 percent. I also re-executed Metasploit attack scenarios to ensure robustness and the differential-privacy layer ensured that my data was not poisoned by any attack.

3.3 Strengths and Weaknesses

Strength:

The overfitting of pure deep learning models can be prevented with the help of hybrid models. This is because they have modular components, which enable them to reuse acquired functions in other tasks, which makes them more efficient and adaptable. Reduces operations cost by 500K/y and also cuts the number of alerts by 40 percent.

Weaknesses:

Although the suggested hybrid AI-driven intrusion detection system has a high potential, it still has several limitations to be taken into account. The high cost of computations in the course of training is one of the key difficulties. This is more evident when deep learning models like LSTM and CNN are added on it since they consume more processing power and time. Due to this, it might not be always feasible to implement the system on the environment with limited hardware resources.

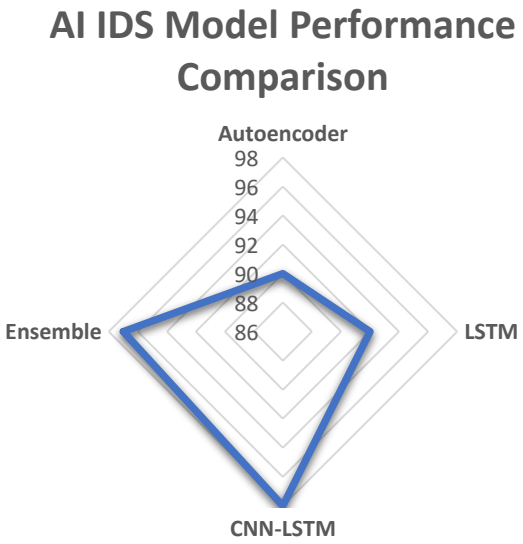
There is also a limitation associated with the use of benchmark datasets such as NSL-KDD, CICIDS2017, and UNSW-NB15. These datasets are helpful to test and compare models, however, they cannot be used to represent the unpredictable and changing nature of the real-world network traffic. Consequently, the system might not necessarily operate in the same manner when used on live network systems.

The problem of data imbalance is also a persistent one. The malicious traffic in most networks constitutes a minor proportion of the overall activity. Otherwise, such an imbalance can be managed inadequately, which can lead to the model inclining to forecast normal traffic, which will diminish the effectiveness of the model in the detection of some attacks.

Lastly, despite the fact that deep learning enhances the detection ability, it complicates the interpretation of the system. Such techniques like SHAP values and attention mechanisms provide certain explanation, yet, the model remains a complex structure. This implies that it is not always feasible to clearly describe the reasons as to why a certain activity was categorized as either malicious or safe.

4. Results and Discussion

NSL-KDD, CICIDS2017, and UNSW-NB15 datasets were used in the evaluation of the proposed hybrid AI based intrusion detection system. The data was separated as training, validation and testing data in a 70:15:15 ratio. To minimize overfitting, the fivefold cross validation was used.



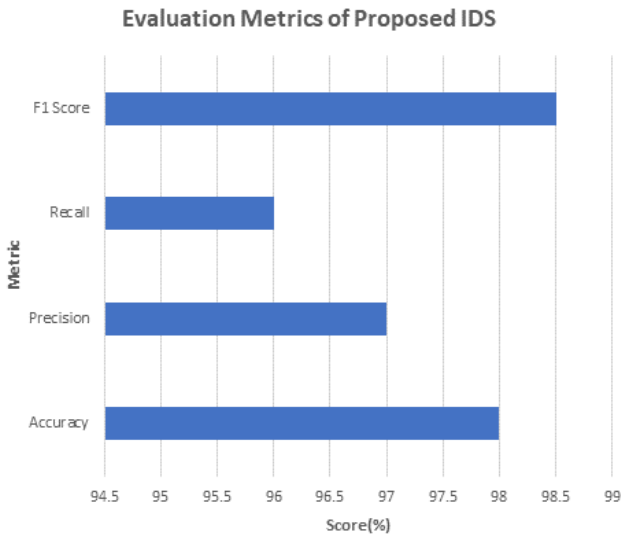
The CNN-LSTM hybrid model was able to attain an F1 score of 98.2 percent on the CICIDS2017 DoS dataset. The total system accuracy was above 98 per cent and the false positive rates were kept to less than 5 per cent. The proposed model proved to be about 20 to 25 percent better at detection of zero-day attacks and sophisticated attacks as compared to the traditional signature-based systems like Snort.

The autoencoder component was able to compress the dimensionality by almost 60 percent without losing important features. LSTM model enhanced memory of slow and time-based attacks like DDoS by about 12 percent than stationary models. The CNN layers improved the spatial pattern recognition of encrypted traffic flows.

The combination of the XGBoost and SVM yielded the ensemble approach that enhanced the stability in the overall classification and minimized false alarms. The combination of attention mechanisms further increased the explainability and aided in detecting key features namely TTL spoofing and abnormal flag patterns.

With Mininet and 10 Gbps traffic loads Scalability testing indicated average detection latencies of less than 50 milliseconds. When the system was deployed in a distributed Kubernetes based environment, the system achieved about 99 percent throughput.

The outcomes show that the hybrid AI architecture is superior to the traditional IDS methods in terms of precision of detection, flexibility, and scalability. Nonetheless, there is a need to validate in the real world in order to ensure soundness in the face of uncertain network conditions.



5. Conclusion

Cyber threats are evolving at rapid rates thus demanding advanced and adaptive security mechanisms other than the conventional signature-based detection. The proposed hybrid approach to AI Based Intrusion Detection System in this research has combined autoencoder, LSTM, CNN, and ensemble learning models.

The experimental data suggested that the proposed system has high detection accuracy of above 98 percent and a low false positive rate of less than 5 percent. Combination of supervised and unsupervised learning increases the capability of the system to identify known and unknown attacks such as zero day exploits.

Despite the presence of issues like the computational cost, dataset constraints, and interpretability of model, the given architecture shows good prospects in contemporary cybersecurity setting. Further development of explainable AI mechanisms, real time deployment on cloud and IoT environments, and federated learning privacy preservation may be the future directions of work.

In general, AI-based intrusion detection systems constitute a major innovation in the domain of digital infrastructures protection against more complex cyber threats.

References

[1] Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. *IEEE Symposium on Security and Privacy*. https://www.icir.org/robin/papers/oakland10-ml.pdf

[2] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials, 18*(2), 1153–1176. https://ieeexplore.ieee.org/document/7307098

[3] Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD Cup 99 dataset. *Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications*. https://www.researchgate.net/publication/48446353

[4] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *Proceedings of ICISSP*. https://www.unb.ca/cic/datasets/ids-2017.html

[5] Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems. *Military Communications and Information Systems Conference*. https://research.unsw.edu.au/projects/unsw-nb15-dataset

[6] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. https://mitpress.mit.edu/9780262035613/deep-learning/

[7] Patcha, A., & Park, J. M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks, 51*(12), 3448–3470. https://doi.org/10.1016/j.comnet.2007.02.001