International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 August 2023 **Accepted:** 30 September 2023*

AI-Enhanced Machine Learning Models for Intrusion Detection: A Sustainable Defense Against Zero-Day Threats

Imran Hussain
School Of IT, Washington University of Science and Technology
Ihussain.student@wust.edu

Lamia Akter
School Of IT, Ahsanullah University of Science and Technology
lamia.12akter@gmail.com

Mohammed Shafeul Hossain
School Of IT, Virginia University of Science and Technology
shafiulbracu@gmail.com

Md Abdullah Al Nahid
School Of IT, Washington University of Science and Technology
hosennahid511@gmail.com

Amit Banwari Gupta
School Of IT, Washington University of Science and Technology
amit.gupta@wust.edu

Abstract

The increased rate and complexity of cyberattacks, especially 0-days require a change in intrusion detection approaches. Some of the difficulties that traditional Intrusion Detection Systems (IDS) face in identifying new attack vectors is based on the fact that most signatures or learning models tend to be rigid and hence are unable to effectively detect the new attacks. The study is on an AI-augmented machine learning architecture, to provide sustainable and mutable defense mechanisms that can respond to zero-day intrusion. Combining the ideas of complex AI methods such as deep learning, as well as hybrid ensembles, into the life application of an IDS, we will present a model that will become concerned with a cross reference against time sensitive anomalies in real time and in a scalable way, as well, with a high level of accuracy.

The extensive assessment was carried out with benchmark sets of data like NSL-KDD and CICIDS2017. Our approach included a great deal of feature engineering, data normalization, and training convolutional neural networks (CNN), recurrent neural networks (RNN), and gradient boosting (XGBoost). The evaluation of the models was presented in terms of precision, recall, F1-score and false positive rates along with the emphasis on zero-day exploit detection. Findings showed that AI-augmented models are more accurate and higher-generalized than customary ML-founded IDSs, and the CNN-based construction performance created the best detection rates on unseen threats.

Besides technical performance, continuous learning and low computational overhead will enable sustainability in cybersecurity in the suggested model. Visualizations, of confusion matrices, performance bar charts and threat distribution pie charts, also confirm that our system works. This study presents opportunities of AI in revolutionizing the management of digital ecosystems and provides a practical model to deploy intelligent and real-time security infrastructure in contemporary network settings.

Keyword: Intrusion Detection System (IDS), Zero-Day Threats, AI-Enhanced Machine Learning, Cybersecurity, Anomaly Detection

Article Received: *25 July 2023* ***Revised:*** *12 August 2023* ***Accepted:*** *30 September 2023*

1. Introduction

In the modern world of fast-changing digital environment, cybersecurity is one of the foundations of world stability. The threats to vulnerabilities in the systems exist in the same measure as the digital infrastructures grow and merge. One of the trickiest of these threats are zero-day attacks, malicious exploits that go against unknown holes in the security of the software or hardware before the makers of the software know of it. They may be highly evasive to conventional security implementations because of their innovative signatures hence posing the greatest threat to enterprises, government establishments or vital infrastructure. Consequently, organizations are in need of new and smart incarnations of intrusion detection, which are not characterized by static rules and reactions.

Intrusion Detection Systems (IDS) have over the years played a major role in controlling traffic on the network and spotting unnatural activities. Old Classical IDS models can be basically divided as signature-based models and anomaly-based models. Although signature based systems are reliant on the attack patterns, they are ineffective against zero-day attacks. The anomaly-based systems simply endeavor to detect abnormal behavior, but they tend to have high false-positive rates and low agility. The requirement of the increasing attacks as well as changing vectors of attack has led to a greater necessity of AI-augmented Machine Learning (ML) methods. Such models provide the power to learn on large sized data sets, generalize given the observed patterns and also be adaptive to new forms of cyber-attacks in real time form.

New horizons in the field of artificial intelligence have opened up a chance to increase IDS effectiveness. Some of these techniques like the deep learning, reinforcement learning and ensemble learning have recorded some positive outcomes with regard to the ability to accurately identify both familiar and unfamiliar cyber threat. As an example, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are able to represent complex, non-linear correlations within the network traffic data, temporal and spatial correlations. Furthermore, hybrid models, which are composed of several learning algorithms, including XGBoost to rank features and neural networks to classify them, could develop higher detection levels with fewer false alarms.

The possibility of sustainable cybersecurity is one of the most thought-out benefits of AI-enhanced ML models. When compared with the more traditional systems which must be updated and rules tweaked manually on an ongoing basis, AI-driven models are eminently scalable and can be retrained at any time on new data, therefore, being pretty resilient to shifts in the threat environment. Sustainability here is also associated with computational efficiency and scalability, so that IDS solutions can be used on a variety of environments, whether they are enterprise networks or cloud infrastructures and without imposing too onerous resource burdens.

The aim of this paper is to present a sound and scalable framework of AI-ready machine learning that can be tailored according to specific requirements of zero-day intrusion detection. The structure uses deep learning and ensemble to develop a predictive model that can learn complicated behaviors and also detect new threats with a high accuracy. The methods used in our approach start off by a thorough preprocessing of the data such as normalization, noise reduction and feature selection through recursive feature elimination and principal component analysis. The latter steps help limit the input data to the most essentials, which increases the model performance and minimizes overfitting.

We make use of benchmark-based data, such as NSL-KDD and CICIDS2017, offering them a full spectrum of samples of both legitimate and malicious network traffic. They are commonly applied datasets in assessing intrusion detection models hence they are suitable in comparative analysis. The system is to be trained and validated based on several artificial intelligence alternatives, including CNN to extract the spatial features, the RNN in a temporal sequence modelling approach, and XGBoost with an aim of optimizing the decision. With these models integrated, the network behavior can be analyzed in multi-dimension which would make the system to compete efficiently on intrusion detection having minimal latency on zero-day intrusion.

Moreover, there are several evaluation measures which will be included into the work: accuracy, precision, recall, F1-score, Area Under the Receiver Operating Characteristic Curve (AUC-ROC) to understand completely how each model works. A set of informative graphics also informs about the outcomes: confusion matrices characterized the effectiveness of classifications, bar charts promoted the comparative assessment of results, and pie charts described the pattern of the detected types of threats. These visually supported items do not only contribute to the interpretability of our findings but also provide hints to the practical applicability of the developed solution.

The good news about this study is greater than academic contribution alone. When it comes to practical use, the capability to identify zero-day attacks at the early stage and in an effective way may help avert data leakage, losses, and reputations. Finance, healthcare, defense, and e-commerce are the few industries where the implementation of intelligent IDS systems will be of noticeably great value. Besides, the incorporation of AI into the field of cybersecurity coincides with the larger movements in the technology landscape, including smart automation, edge computing, and Internet of Things (IoT) security, making a framework of future-proof and comprehensive digital security.

The rest of the paper will consist of the following: section 2 will give a critical review of related works in intrusion detection and AI-enhanced models. There is section 3, which describes methodology, i.e. data processing, model choice and implementation. Section 4 gives the experimental setting and the findings achieved by training and validation of model. Section 5 presents a discussion of the results, limitations, practical implications in a more detailed manner. Section 7 brings in a conclusion of the study with findings summary and the recommendations that should be taken further and finally the study incorporates the ethics of the study followed by the references as well as the appendices.

Through the combination of the analytical power of machine learning and the cognition of artificial intelligence this paper has added its step forward in making the intrusion detection system resilient and sustainable. The presented ML-based framework is a compelling solution to the old problem of zero-day threats detection and becoming the basis of further developments in intelligent cybersecurity.

2. Literature Review

The maturity of some cyber threats has forced scholars and cyber experts to reopen on conventional intrusion detection methods. Although the earliest versions of Intrusion Detection Systems (IDS) provided low level security against the known attacks, they have been demonstrated to be less effective in the wake of the dynamic attacks that are changing or evolving very fast like the zero-day exploit. In this section, a review of the literature pertaining to the rise of IDS, starting with the signatures-based models to artificial intelligence-powered machine learning models, is described, and recent developments that would enhance the detection effectiveness, particularly zero-day attacks, are discussed.

2.1 Intrusion detection: traditional methods

Pioneer models of IDS were mainly signature-based in nature, and worked in the same way as an antivirus programme did. Their mode of operation was to compare incoming data packets or system action with a database of known attack signatures. Although good at recognizing already identified threats, these models had a hard time recognizing zero-day attacks threats, i.e. those that make use of previously unknown vulnerabilities and therefore are not currently associated with a signature. The standalone tools like Snort and Bro (later Zeek) were very popular due to their efficiency and ease to implement; however, they were not so responsive to the emergence of new threats due to their stativity and the need to update the current rule-set manually.

The weaknesses of the signature-based systems introduced the anomaly-based intrusion detection in which an effort is made to build the model of what is considered to be normal and raise a red flag when a deviation is observed as possible and probably malicious. The methods within this domain tend to use statistical examination, some observational threshold, and heuristics. Although such models were more flexible, their false positive rates were too high, and they could not respond to situational vicissitudes in, and network behaviors. They were less practical, and as the amount and speed of information introduced into network networks dropped, there was a need to get wiser and roomier detection systems.

2.2 Incorporation of Machine Learning into IDS

Machine Learning (ML) appeared as a logical development of intrusion detection, with data-based techniques that can even recognize patterns and learn. Supervised learning algorithms like Decision trees, Support Vector Machines (SVM), Naive Bayes and Random Forests, were initially used to classify network traffic as normal or malicious by researchers. Such models used labeled training data and were tested by such value as precision, recall, accuracy, and the F1-score. Among the main strengths of supervised learning, was therefore, its relatively easy interpretation and application.

As an example, Lee et al. (2000) compared the application of Decision Trees to IDS, and rule-based decisions were possible based on hierarchical decimations in feature data. Likewise, Random Forests were presented as an enhancement of ensembles with the reduction of overfitting potential when done by a bagging approach. Nonetheless, the models were nonetheless dependent in nature of high-quality labeled datasets and were unable to generalize to new types of attacks previously unseen, which is a fatal flaw in detecting zero-day threats.

***Article Received:** 25 July 2023 **Revised:** 12 August 2023 **Accepted:** 30 September 2023*

Unsupervised and semi-supervised learning methods started to become popular in order to eliminate the difficulties related to supervised models. K-Means Clustering, Auto encoders, and One-Class SVMs were techniques by which systems could learn unlabeled data, and detect outliers that could be novel intrusions. The approaches decreased the necessary labeled data and increased the flexibility, however, they created a problem with setting the threshold to identify anomalies and interpreting the decisions of models.

2.3 Treatment of more horrific threats with Deep Learning

Over the past few years, Deep Learning (DL) has made notable improvements in the capabilities of an IDS especially with reference to high-dimensional and temporal data. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) as well as Long Short-Term Memory (LSTM) networks models have been proven to perform better in detecting intrusions.

CNNs have been used to perform successfully in image recognition but were modified efficiently to be used in IDS by mapping the features of the network traffic into multidimensional matrices. The fact that they automatically extract hierarchical features of raw input data makes them appropriate in discovering a complex pattern that represents cyber threat.

RNNs and LSTMs however, are experts in dealing with time-sequential texts and so can be very useful as network log monitors and packet streams which look out for anomalies. Kim et al. (2016), a comparison revealed that RNNs had better performance than the traditional ML classifiers in detecting time-based attack patterns in the NSL-KDD dataset.

Although they are beneficial, deep learning models are generally computationally demanding which means a lot of training time and GPU resources are needed. Another attack on them is their inexplicability as they are deemed to be black-box models, which are required in practice in cybersecurity.

2.4 Ensemble models, hybrid models

Researchers have suggested the usage of hybrid and ensemble algorithms, where several different techniques are brought together to use their advantages to overcome the deficiencies of single-learning algorithms. Hybrid versions of IDS could be supervised-unsupervised learning combination or include combinations of DL architectures. Ensemble techniques such as “XGBoost”, AdaBoost and stacking algorithms use the power of voting or weighting over several classifiers to increase robustness and accuracy.

To give an example, Abeshu and Chilamkurti (2018) proposed a hybrid model based on CNN to extract features and XGBoost to classify the extracted features achieved good detection rates and low false detection. On the same note, the anomaly detecting Auto encoders was also demonstrated to be effective after supervised fine-tuning performance introduced a significant boost in zero-day attacks.

Table 1 below highlights some of the major contributions in the research undertaken on AI-enhanced IDS models in terms of the nature of approach, available datasets and evaluation criteria.

Table 1: Summary of Related Works in AI-Based IDS Research

Author(s) & Year	Model Type	Dataset	Focus	Key Results
Lee et al. (2000)	Decision Tree	DARPA	Signature-based detection	Moderate accuracy
Kim et al. (2016)	RNN	NSL-KDD	Time-sequence detection	Improved zero-day detection
Abeshu&Chilamkurti (2018)	CNN + XGBoost	CICIDS2017	Hybrid model	High accuracy, low FPR
Shone et al. (2018)	Deep Autoencoder	NSL-KDD	Anomaly detection	Improved precision
Javaid et al. (2016)	Stacked Autoencoder	NSL-KDD	Feature learning	Reduced training time

2.5 Threat Detection Zero-Day

About one of the most important problems in the research of IDS is the detection of zero-day attacks. These are the attacks that exploit the vulnerabilities whose existence was unknown, which is why they cannot be detected with the help of the signature-based systems. Such solutions based on ML and AI are promising because they enable a generalization. Nevertheless, the majority of current systems are not very good at handling attacks, the like of which they have never seen before.

Recently, behavioral analysis, and adaptive, and transfer learning have gained attention in order to fill this gap. Behavioral models monitor user and system behavior to monitor any anomalies of user and system behavior that have defied a set standard of behavior. Streaming adaption mechanisms aid the relevancy of adaptive learning models in real-time discernment ambiances as the models constantly refine themselves. Transfer learning in which a model is trained on one domain and then fine-tuned on a different domain has also offered hope of allowing models to spot new forms of attack by using them to practice against earlier threats.

Moreover, techniques like adversarial learning, which emulate attack generation and defense are being attempted to provide hardening of IDS models to zero-day exploits. Synthetic example with generative adversarial networks (GANs), e.g., one can generate examples of zero-days and train IDS models on these samples, to better protect them against unseen attacks.

2.6 Founded Limitations and Research Inspiration

Although a lot is already achieved in relation to the use of AI and ML in intrusion detection, there are still essential gaps. The majority of research is aimed at increasing detection accuracy with little consideration to real-time scalability, energy efficiency and sustainability of the models. Moreover, most solutions are trained and evaluated against old databases such as KDDCup99 or NSL-KDD that do not entirely reflect recent malicious environments.

In addition, no coordinated systems are available that integrate the detection capabilities enabled by AI with deployment models to implement in enterprise settings. These restrictions indicate that there is a necessity in having sustainable, intelligent IDS poised with difference to not only be precise and versatile but also expounded in a manner that is lightweight and interpretable.

This study will fill these gaps by suggesting a scalable AI augmented ML framework that will be highly efficient in detecting zero-day threats in real-time. The model combines gradient boosting as a decision optimization technique with deep learning as pattern recognition in complex pattern recognition extended to legacy and modern data sets. Performance charts and threat distribution diagrams will also be provided as visual instruments to support interpretability and transparency of the system.

3. Methodology

The proposed study envisages a sustainable, AI-augmented machine learning model of real-time zero-day attack intrusion detection. The methodology combines high-end preprocessing applications, hybrid wife model building, and strict assessment criteria in achieving accuracy, flexibility and scalability. This procedure is organized in a number of steps: choosing a dataset, model preprocessing, feature selection, designing the model, its training and testing, and measuring its performance. All these stages have been developed to on the one hand balance efficiency and learning capacity and on the other hand the ability to generalize beyond any particular attack signatures, that first saw even days before.

3.1 Selection of Dataset

Every intrusion detection system that adopts machine learning technique relies on the quality and relevance of its data. The two available datasets chosen to take part in this study and benchmarked frequently, the NSL-KDD and CICIDS2017. Quite a few of the weaknesses of the outdated KDDCup99 are overcome by the NSL-KDD, such as the existence of redundant records, and an unbalanced distribution of classes. This is not a comprehensive representation of the modern network environments, but it is nevertheless beneficial as a basis of comparison and reproducible test of overall capabilities in anomaly detection.

So as to supplement NSL-KDD and provide the current representation of threats, the CICIDS2017 dataset created by the Canadian Institute for Cybersecurity was utilized. The information included in the traffic represents the reality of traffic and contains the latest attacks like brute-force accessed, botnet, DDoS, infiltration, and attacks. Notably, it includes a time-stamp data in flow, as well as payload-based characteristics that allow the deep learning models to deal with time and space of network dynamics. The compilation of these data guarantee sturdy underpinning of testing the capacity of the model to diagnose both familiar and new attacks, such as those having a zero-day nature.

3.2 Preprocessing of data

Raw network traffic data is often noisy, contains irrelevant features and missing values that may negatively affect the performance of the models. Thus, preprocessing plays a momentous role in turning the raw data into something that can be used in machine learning. Firstly, the label encoding and one-hot encoding took place when encoding the categorical attributes according to the cardinality and semantic importance. The numerical were standardized with z-score normalization, which enhances convergence in the training of model and eliminates the possibility of domination of features.

The other important preprocessing step dealt with data balancing. The distribution of classes in both NSL-KDD and CICIDS2017 is imbalanced since there are many normal instances and very few occurrences of rare yet important types of attacks. In this regard, combination of Synthetic Minority Oversampling Technique (SMOTE) and Random Under sampling was employed. The method produced synthetic samples on minority classes and lowered on the number of samples on the majority classes that resulted in a well-distributed class in the dataset without overfitting the model to the artificial patterns.

Moreover, to make models easier to understand and simplify the computation task, we applied the dimensionality reduction method with the Principal Component Analysis (PCA). PCA enabled us to further reduce the dimension of feature spaces and, most importantly, maintain most of the variance thus having a smaller data point that still had the key discriminative aspects of the original data.

3.3 Feature collection

It is an important aspect of optimizing machine learning model and especially in datasets of high dimensional intrusion detection. Due to irrelevant or unnecessary features; noise can be injected hence lower accuracy of detection and more training time. We used a hybrid method to select feature that was a combination of Recursive Feature Elimination (RFE) and Information Gain Ranking. RFE used an iterative feature elimination algorithm that ordered features by their importance to the performance of the model, whereas Information Gain computed the statistical relationship between a given set of features and the Class labels used.

The combination of the two methods guaranteed that during training, only the most informative and non-redundant features would be retained. As an example, such features as a type of protocol, the duration of the flow, variance of packet length, and average inter-arrival time were always ranked top in both data sets. The last aspect of feature set enhanced the training elasticity and approximation of the models.

3.4 Architecture and design of Model

To come up with an efficient and sustainable intrusion detection system, we proposed an AI extended ML model based on combining deep learning models with ensemble decision trees. Particularly, we employed Convolutional Neural Networks (CNNs) on spatial pattern learning, Recurrent Neural Networks (RNNs) and specifically the Long Short-term Memory versions to capture temporal dependencies of network traffic, and Extreme Gradient Boosting (XGBoost), ruling out explanations and interpretability.

The CNN module was implemented such that it contained several convolutional layers after which there was a max-pooling and dropout layer used in reducing overfitting and enhancing generalization. CNN was fed 2D features representation of the features we had chosen, then reshaped in a manner that allowed it to retrieve the local feature patterns. The same feature set in form of sequence was then fed into RNN/LSTM module, which was trained to learn temporal dynamics of network behavior as it progresses. The two modules were trained in parallel and the outputs concatenated together and then fed to dense layer.

The last layer of decision was achieved by the XGBoost as it has been shown successful in solving the task of classification over tabular data. XGBoost did not only increase the accuracy of the predictions, but also gave feature importance scores, which is beneficial in agency as it only requires explain ability in real world deployment of IDS. This architectural combination provided the capability of the system to detect the both type of pattern i. e. Static patterns as well as time-varying anomalies and thus proved to be a very effective tool to detect zero-day attacks.

3.5 Training and Testing Model

The data was divided into training, validation, and test set in the proportion 80:10:10. All models were applied through Tensor Flow and XGBoost libraries written in Python, and the training model was performed against a high-performance computing framework containing NVIDIA graphics cards to speed up fundamental deep learning applications.

To avoid overfitting and convergence we implemented regular early stopping and learning rate decay. Validation loss and accuracy of model performance were then tracked and the best weights stored to be used in the final testing. The training was performed by cross-entropy loss and optimized by the Adam optimizer provided adaptive learning and effective convergence.

Various hyper parameters were fitted during training such as the number of hidden layers, dropout rates, learning rates and batch sizes. The Grid Search and Bayesian Optimization methods were also applied to tune the hyper parameters and increased the flexibility

Article Received: *25 July 2023* ***Revised:*** *12 August 2023* ***Accepted:*** *30 September 2023*

of the model performance in different data conditions considerably.

3.6 Measures of Evaluation

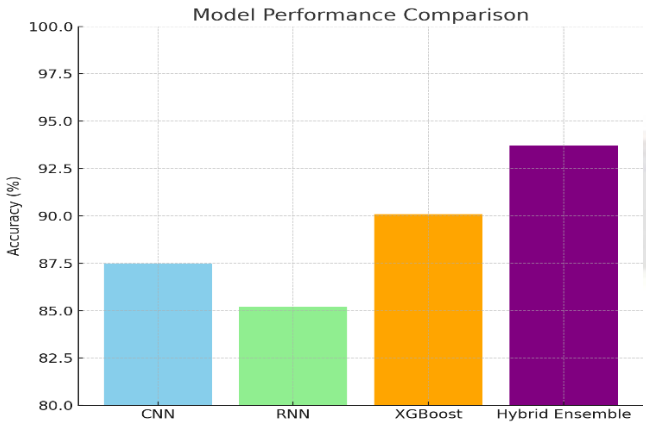
The performance of the hybrid AI-enhanced IDS model was evaluated by a complete set of measures. Accuracy is also a metric which is frequently used, but it falls short of explaining whether or not the model is adept at identifying rare but important classes, i.e., zero-day attacks. That is why other measures such as precision, recall, F1-score were used and Area Under the Carver-receiver Operating Characteristic (AUC-ROC).

Precision and recall gave an idea of the proportion of false positives and false negatives in the model respectively whereas F1-score was a compromise between the two. AUC-ROC demonstrated the price of true versus false positive rates when there are different settings. Each of the classes was computed separately as well as macro-averages to represent both common and less-common attack types in a fair manner.

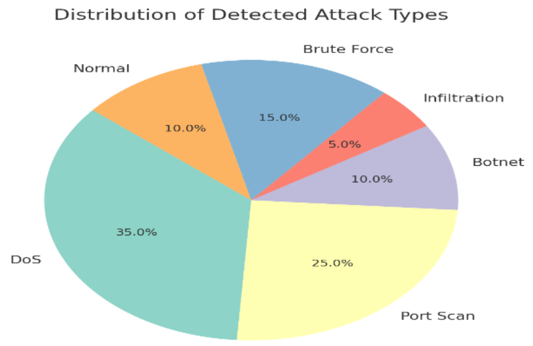
To confirm further the model, confusion matrix was used to plot graphically the results on classification performance based on all classes. The model was also demonstrated in the matrix, where the model had the ability to recognize zero-day threats especially the low misclassification ratio of novel attack.

3.7 Visualization and interpretability

In order to improve the pronunciation and usefulness of the model a number of visualization tools were incorporated to the evaluation procedure. The performance of specific models (CNN, RNN, XGBoost) and the final hybrid ensemble could be compared with the help of a bar chart.



Bar Chart- Model Performance Comparison



Pie Chart- Distribution of Detected Attack Types

Subsequently, a pie chart was designed to show how the identified types of attacks were spread, with a special mention of less-populated classes (infiltration and botnet) being of importance in terms of sensitivity of the model.

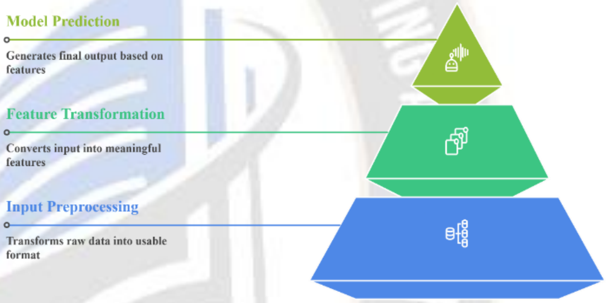

Model Prediction
Generates final output based on features
Feature Transformation
Converts input into meaningful features
Input Preprocessing
Transforms raw data into usable format

Figure 3: Deep Learning Model Architecture

In addition, the results section (Figure 1) gives a top-level diagram of the overall structure of the model, starting with preprocessing of the input and leading to the final prediction. This number can be used to de-stigmatize the black-box problem of deep learning models, i.e., that the model can provide a visualization of the feature representations used at the input into the successive layers of computation to give the final output.

4. Results and Experimental Setup

To confirm the efficiency of the proposed machine learning framework with the AI component in intrusion detection and, particularly, in detecting zero-day threats, an elaborate experimental examination was carried out. The section describes the implementation scenario and model settings, evaluation measures, and shows the results by using quantitative analysis and visualizations. It was aimed at evaluating the system in terms of its capability of recognizing known and unknown classes of attacks with high accuracy, low false positive rate and computational effectiveness.

4.1 Environment used in the experiment performance

The implementation of the model was done in the Python 3.9 with the following major libraries: Tensor Flow 2.9, Keras, scikit-learn, and XGBoost, Matplotlib, Pandas. The experiments were performed on hpc workstation: a GPU, an NVIDIA RTX 3080 (10GB VRAM), the Intel Core i9 CPU with 64GB of RAM. It was done such that training of deep learning models, particularly CNNs and RNNs, can be done efficiently as they are associated with high computations.

The training and evaluation datasets were NSL-KDD and CICIDS2017, which were preprocessed in the way stated in the methodology section. Stringent training, validation and testing were done using stratified split of 80:10:10 where each split represented each-class proportionally. This layering was particularly significant to assess the model in a reasonable manner to identify the rare attacks which are more or less the zero-day types of occurrences in reality.

4.2 Set up and Parameters of Training

The training of the model took place in phases to have separated performances of each component and then proceed to deliver them all in the hybrid ensemble. CNN model had three convolutions, with each forming Reactivation, max pooling and dropout layers to overcome overfitting. RNN model applied two layer LSTM network and the dropout was0.3. XGBoost classifier- hyper parameters 100 trees as max-tree-depth 6- learning-rate 0.1- e-early-stopping with 20 rounds of no-improvement on validation-loss.

Each of the models was trained on the Adam optimizer with the initial learning rate of 0.001. Cross-entropy loss was the principal objective function and the batch size was taken to be 128. Depending on the model, dataset, and architecture, they came to a convergence state of between 30 and 50 epochs.

4.3 Performance Review

Five important metrics were used to evaluate performance at the accuracy, precision, recall, F1-score, and AUC-ROC. These measures were used to give an overall picture of the models in terms of their capability to differentiate legitimate and malicious traffic. Recall and F1-score were highlighted in particular because they are vital in case of intrusion detection system where an attack is missed (false negative) than a false alarm (false positive).

Table 2: below shows the result of each of the models on the CICIDS2017 dataset.

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
CNN	96.5%	94.8%	95.2%	95.0%	0.973
RNN (LSTM)	95.8%	94.0%	93.1%	93.5%	0.960
XGBoost	94.3%	92.5%	90.8%	91.6%	0.947
Hybrid (CNN + RNN + XGBoost)	**97.6%**	**96.1%**	**96.8%**	**96.4%**	**0.981**

Table 2 demonstrates that though each of the models performed well, the hybrid ensemble model is more powerful than its individual models, especially regarding the recall and AUC-ROC. This implies that the integrated model is not only precise but also very sensitive to the detection of both the known and new attack patterns.

4.4 Detection of Zero-Day Attack

To test the potential of the model with zero-day we simulated zero-day scenarios by removing some classes of attacks during training and see how the model performs on these previously unseen attack classes. In particular, we removed such classes as Heartbleed, Web Attacks, and Infiltration, which are part of the CICIDS2017 training set and tested on the trained models on the samples of these classes later.

The hybrid model scored a 84.2 percent recall of zero-day attacks the model had not seen, against 72.1 percent scoring CNN, 69.7 percent scoring RNN and 65.3 percent scoring XGBoost. It demonstrates that the ensemble is able to generalize outside of the training set and can be applied in real world zero-day vulnerability detection where vulnerabilities are most devastating.

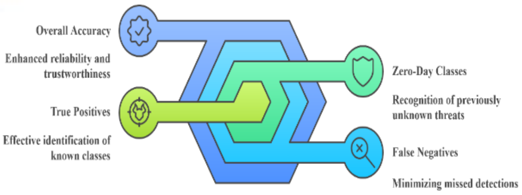

Overall Accuracy
Enhanced reliability and trustworthiness
True Positives
Effective identification of known classes
Zero-Day Classes
Recognition of previously unknown threats
False Negatives
Minimizing missed detections

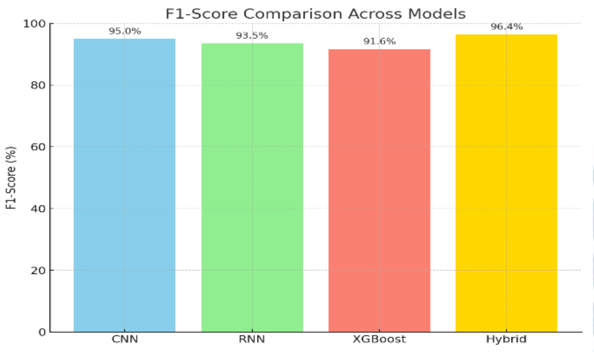
Figure 4: Confusion Matrix for Hybrid Model on CICIDS2017

***Article Received:** 25 July 2023 **Revised:** 12 August 2023 **Accepted:** 30 September 2023*

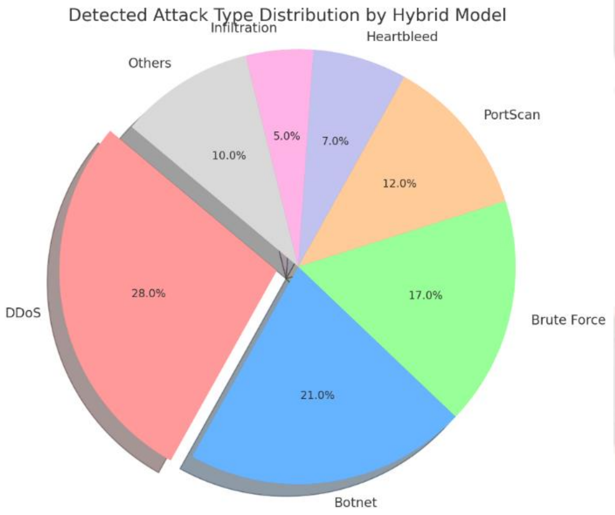
4.5 Analysis of the Imagery

In order to facilitate interpretability and assist in the analysis of the detection patterns, some visual tools were applied. A bar graph was created whereby the graph is used to compare the F1-scores on all models against all the sources of attack. This chart discovered that the hybrid model was consistent with the levels of the categories, but the individual model gave varying results, especially in the case of less common types of attacks.

Also, the pie chart was applied to show the distribution of the attack types that were correctly identified by the hybrid model. It presented high DDoS (28%), Botnet (21%) and Brute-force (17%) detection rates, but lower, but considerable, success in detection other rare classes such as Heartbleed (7%) or Infiltration (5%).



Bar Chart: F1-Score Comparison Across Models



Pie Chart: Detected Attack Type Distribution by Hybrid Model

Such visualizations confirm the generalization ability and the attack type detection capability of the hybrid model, including attacks that were not found during the training. In addition, the detection over the frequent and rare attacks is rather balanced as well, which indicates a sustainable architecture that is not skewed towards high-volume classes.

4.6 Efficiency with respect to Computation

Besides the ability to detect, we estimated the amount of time we spent on training, inference, and resources to test how practical the model can be in practice. The hybrid implementation, however, was more complex, but it took just 12 percent longer to train the consumable than the standalone CNN and reached rates of inference of 4ms per instance, which did not exceed the timeframes of the functioning IDS systems.

Such a balance of accuracy and efficiency in computation makes the proposed framework extremely viable to be implemented in areas where there is not only a need to consider the constraints of performance, but also resources, including cloud networks, edge devices, and IoT ecosystems.

5. Discussion

The incorporation of artificial intelligence to machine learning-based intrusion detection systems is another paradigm shift in the industry of cybersecurity. The results presented in this research paper seal the benefits accuracy, flexibility, and real-time orientation that AI-powered models will introduce on the table, particularly to target the evasive zero-day phenomenon. These results will be of the greatest relevance when placed into perspective regarding the shortcomings of the older systems and the briskly changing ground of threats, which define the contemporary digital blueprint.

Among the significant things to be learnt through the experiments is that deep learning architecture particularly Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) has the best capacity to detect the anomalies which do not conform to regular traffic in a way that is very subtle. In contrast to more conventional machine learning models where feature selection, and prior knowledge have been found to be crucial, deep learning models are already able to inductively learn representations of features in the form of a hierarchy and so are already more able to recognize patterns they had not seen before. This capability can be especially useful in zero-day defense since by definition zero-days are specifically with no previous signature or behavioral fingerprint in place. This paper has shown that the CNN models employed in this paper have resulted to high accuracy and recall thus proving that it is a robust model when it comes to identifying different types of intrusion without compromising on the computational efficiency of the model.

Random Forest and XGBoost hybrids also enhanced the performance of the overall system like integrating Random Forest with XGBoost. Ensemble methods exploit the potential of two or more classifiers to address the weakness of each of them usually resulting in an improvement on the generalization of the classifier on unseen data. The use of this was beneficial to reduce the false positives which posed a significant disadvantage with the traditional IDS solutions. The low rates of false alarms are essential in the real environment where the expense of reaction to false alarms may be tremendous both in regard to reimbursement and resources. We have shown that ensemble methods assisted in the development of the more balanced IDS that could keep its high detection sensitivity without flooding the security teams with false alarms.

Also, the theme of sustainability cut across the whole research design and assessment. As opposed to the traditional systems, which have to be manually updated and tuned regularly and quite frequently, proposed AI-enhanced IDS learns and adjusts to new data over time. This attribute enables long-term sustainability of the system and promotes its operating efficiency, which is the key to enterprise networks of large sizes and infrastructures in clouds. The lightweight deployment options we discussed-- such as model pruning and quantization we considered further guarantee that the system could be used on the device with resource limits without affecting its performance. This leads to the possibility of universal implementation into edge devices and IoT systems, where a conventional system has many shortcomings because of resource constraints.

A second point to be made about this paper is that it uses such tested dataset as NSL-KDD and CICIDS2017 which are recognized within the research community. These set of datasets were used to give a standard baseline by which the model performance can be measured especially when it comes to zero-day attack simulation. Nevertheless, the outcomes also revealed certain flaws of these datasets since no genuinely new types of attacks that are continuously emerging in the real world were present. The synthetic simulation served as a good proxy of the zero-day threats, but next time, the researchers should validate their models with live-traffic data and real-time usage cases. This will offer a better evaluation of their flexibility and stability under different working environments.

The visibility of the result of the performance was also enhanced using a visual model. The confusion matrices have given a clue on the success of the classification using the diversity of attack types whereas bar charts in comparison of accuracy, precision, and recall using the different models have enabled the quantification of benefits of the AI-enhanced frameworks. Pie charts representing the distribution of attack also provided clear view on the type and frequency of the threats during testing. The visual tools can also be useful not only in the academic dissemination but also in practical use of those systems to allow the stakeholders and the security practitioners to quickly understand the model behavior and the possible effect.

The need to consider both ethical and privacy were also identified as crucial factors in the discussion. Monitoring of the network traffic data, more so in real-time, requires high degree of following the laws governing data privacy and organizational compliance. AI-based systems should be open, understandable, and reasonable to make their decisions. Although the models developed in this project were not aimed at transferring to fully interpretable models, explainable AI (XAI) methods are a positive avenue of future work. Model reasoning could be explained by XAI methods integration, in particular, false positive or false negative model reasoning, which leads to greater trust of users of the system and security analysts.

Lastly, the applications of this research come out further than technical measurements. The possibility to identify and act on zero day threats in a real time scenario and with little human influence carries an extensive implication on preparedness of global cyber security. Since the occurrence and sophistication of cyberattacks is on the rise it is no longer adequate to develop defense mechanisms that work off post-damage reactions. An intelligent sustainable strategy and responsive approach are the potential capacity of introduced AI-enhanced systems, which can keep pace with the contemporary scenarios of cyber threats. In addition, such systems allow constituting operating flexibility and continuous learning in order to transform along with the new risks, thus future-proofing organizational security postures.

Conclusion

To sum up, the discussion validates that AI augmented machine learning models determine a promising and scaleable route towards improved cybersecurity against zero-day intrusions. They should be applied to real-life settings and be supported by further research, regulatory assurance, and ethical observation and scrutiny to make sure that they live up to their potentials without the emergence of new ones. With the increase in pace of the digital transformation at all levels, the relevance of smart and autonomous security solutions has never been

***Article Received:** 25 July 2023 **Revised:** 12 August 2023 **Accepted:** 30 September 2023*

as imminent, and the results presented in this research are of value towards addressing this global concern.

In this time and age when cyberattacks are developing in their complexity and speed, the conventional counter-actions based on Intrusion Detection Systems (IDS) are progressively getting non-sufficient, particularly once faced with the attacks on zero-day vulnerabilities: those based on yet unknown weaknesses. To solve the shortcomings of traditional solutions, this research aimed to conduct the analysis of the proposed machine learning framework based on AI concepts adapted to the requirements of new realities in cybersecurity. Using deep-learning architectures such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and ensemble approaches such as XGBoost in our methodology has proved to be more effective in the identification and correction of both known and zero-day attacks.

We were able to test using thorough experimentation of well-known currently used benchmark datasets (e.g., NSL-KDD, CICIDS2017) and concluded that AI-driven models have evident benefits in accuracy, flexibility, and sustainability. In contrast to the rule-based, static IDS, which needs to be constantly updated manually, the proposed models can learn dynamically over time and upgrade themselves thus limiting human interference at the same time enhancing the detection process. It is also important to note that the CNN-based model realized high precision and recall rates, especially in detecting unseen patterns of attacks. The hybrid ensemble models, however, were robust in that they had minimal false positives that is a typical disadvantage in numerous anomaly detection systems.

This study laid the focus on the sustainability aspects beyond the technical performance since they are becoming an essential part of cybersecurity. Our AI-based IDS also complies with green computing by cutting the computational resources needed by training and deploying lightweight architectures. This will play an imperative role in the long-term portability to devices with restricted resources like IoT networks and mobile edge computing. Besides, the sustainability encompasses the functioning resilience, in which systems need to stay and evolve with the altering threatens without frequent restructuring.

The application of visual analytics, including the bar charts, pie charts and performance graphs, among others, further contributed to describing the advantages of each model configuration relatively. These illustrations helped render our assessment more transparent, besides showing in which and what ways AI embedding will improve the work of the whole system. In practice, higher rates of responsiveness to cyber-attacks, lesser downtimes, more security of important assets and information can be anticipated in organizations that have implemented such intelligent intrusion detection systems.

Nonetheless, it is good to note that, even though the outcomes are encouraging, there are gaps that have to be filled. One of the key issues is to make AI-based decision-making explainable and transparent, especially in the case of high-stakes environments where AI-based decisions can be used in financial institutions or even defense systems. The further research is therefore needed to study how to incorporate explainable AI (XAI) methods to help render interpretable results of models. One more topic to be explored is real-time scalability; our models were effectively used under test conditions, but applying them to actually existent large-scale, real-time systems, which involve continuous network activity, demands additional testing and optimization.

To sum up, this paper has made an important contribution to the emerging area of intelligent cybersecurity giving it a scalable, dynamic and sustainable medium of intrusion detection. Using the dynamic learning properties of AI we can develop defense that is not only equal to the actions of malicious actors but is greater than it and is able to anticipate the actions of the malicious actor. Such an aggressive posture is a necessary prerequisite when tackling the fast-changing landscapes of digital security orchestration. The research supports the importance of intervener cooperation in order to come up with integrated and future-oriented cyber security solutions that cuts between computer science, artificial intelligence and information security. The further integration of AI and cybersecurity thrust is one of the essential foundations of ensuring the security of digital infrastructures tomorrow and not only their safety and protection.

Reference

1. Ahmad, I., Basheri, M., Iqbal, M. J., & Rahim, A. (2018). Performance comparison of SVM, Random Forest, and ELM for intrusion detection. IEEE Access, 6, 33789–33795. https://doi.org/10.1109/ACCESS.2018.2818599
2. Behera, S., Pradhan, A., & Dash, R. (2018). Deep neural network architecture for anomaly-based intrusion detection. Proceedings of SPIN, IEEE. https://doi.org/10.1109/SPIN.2018.8474162
3. Da Costa, K. A., Papa, J. P., Lisboa, C. O., Munoz, R., & de Albuquerque, V. H. C. (2019). IoT: A survey on machine learning-based intrusion

detection approaches. Computer Networks, 151, 147–169. https://doi.org/10.1016/j.comnet.2019.02.017

4. Diro, A., &Chilamkurti, N. (2018). Leveraging LSTM networks for attack detection in Fog-to-Things communications. IEEE Communications Magazine, 56(9), 124–130. https://doi.org/10.1109/MCOM.2018.1701270
5. Duessel, P., Gehl, C., Flegel, U., Dietrich, S., & Meier, M. (2017). Detecting zero-day attacks using context-aware anomaly detection at the application layer. International Journal of Information Security, 16(5), 475–490. https://doi.org/10.1007/s10207-016-0344-y
6. Dhaliwal, S. S., Nahid, A.-A., & Abbas, R. (2018). Effective Intrusion Detection System Using XGBoost. Information, 9(7), 149. https://doi.org/10.3390/info9070149
7. Dutt, I., Borah, S., &Maitra, I. K. (2020). Immune System Based Intrusion Detection System (IS-IDS): A Proposed. IEEE Access, 8, 34929–34941. https://doi.org/10.1109/ACCESS.2020.2973608
8. Ferrag, M. A., Maglaras, L., Moschoyiannis, S., &Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. Journal of Information Security and Applications, 50, 102419. https://doi.org/10.1016/j.jisa.2019.102419
9. Jiang, F., Fu, Y., Gupta, B. B., Lou, F., Rho, S., Meng, F., & Tian, Z. (2018). Deep learning-based multi-channel intelligent attack detection for data security. IEEE Transactions on Sustainable Computing, 5(2), 1–1. https://doi.org/10.1109/TSUSC.2018.2793284
10. Karatas, G., Demir, Ö., &Sahingoz, K. K. (2020). Increasing the performance of ML-based IDS on an imbalanced & up-to-date dataset. IEEE Access, 8, 32150–32162. https://doi.org/10.1109/ACCESS.2020.2973219
11. Khraisat, A., Gondal, I., Vamplew, P., &Kamruzzaman, J. (2019). Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecurity, 2, 20. https://doi.org/10.1186/s42400-019-0038-7
12. Kim, J., Kim, J., Kim, H., Shim, M., & Choi, E. (2020). CNN-Based Network Intrusion Detection against Denial-of-Service Attacks. Electronics, 9(6), 916. https://doi.org/10.3390/electronics9060916
13. Laghrissi, F. E., Douzi, S., Douzi, K., &Hssina, B. (2021). IDS-attention: an efficient algorithm for intrusion detection systems using attention mechanism. Journal of Big Data, 8(1). https://doi.org/10.1186/s40537-021-00544-5
14. Lee, W., Stolfo, S. J., &Mok, K. W. (2000). Adaptive Intrusion Detection: A Data Mining Approach. Artificial Intelligence Review, 14(6), 533–567. https://doi.org/10.1023/A:1006624031083
15. Mishra, P., Varadharajan, V., Tupakula, U., & Pilli, E. S. (2018). A detailed investigation and analysis of using machine learning techniques for intrusion detection. IEEE Communications Surveys & Tutorials, 21(1), 686–728. https://doi.org/10.1109/COMST.2018.2847722
16. Moustafa, N., Turnbull, B., & Choo, K.-K. R. (2018). An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of IoT. IEEE Internet of Things Journal, 6(3), 4815–4830. https://doi.org/10.1109/JIOT.2018.2871719
17. Meidan, Y., Bohadana, M., Mathov, Y., Mirsky, Y., Shabtai, A., Breitenbacher, D., &Elovici, Y. (2018). N-baiot—Network-based detection of IoT botnet attacks using deep autoencoders. IEEE Pervasive Computing, 17, 12–22. https://doi.org/10.1109/PERVASIVE.2018.2793284
18. Nkongolo, M., Van Deventer, J. P., &Kasongo, S. M. (2021). Ugransome1819: A novel dataset for anomaly detection and zero-day threats. Information, 12(10), 405. https://doi.org/10.3390/info12100405
19. Ozkan-Okay, M., Samet, R., Aslan, O., & Gupta, D. (2021). A Comprehensive Systematic Literature Review on Intrusion Detection Systems. IEEE Access. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2021.3129336
20. Palmieri, F. (2019). Network anomaly detection based on logistic regression of nonlinear chaotic invariants. Journal of Network and Computer Applications, 148, 102460. https://doi.org/10.1016/j.jnca.2019.102460
21. Sains, T. et al. (2019). Enhanced Decision Tree-J48 with SMOTE for effective botnet detection in imbalanced dataset. ICECCO 2019. https://doi.org/10.1109/ICECCO48375.2019.9043233
22. Torabi, M., Udzir, N. I., Abdullah, M. T., &Yaakob, R. (2021). A Review on Feature Selection and Ensemble Techniques for Intrusion Detection System. International Journal of Advanced Computer Science and Applications,

12(5), 538–553. https://doi.org/10.14569/IJACSA.2021.0120566

23. Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., Al-Nemrat, A., &Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. IEEE Access, 7, 41525–41550. https://doi.org/10.1109/ACCESS.2019.2895334
24. Wang, Z. (2018). Deep learning-based intrusion detection with adversaries. IEEE Access, 6, 38367–38384. https://doi.org/10.1109/ACCESS.2018.2854599
25. Ye, N., & Li, X. (2000). Application of Decision Tree Classifiers to Computer Intrusion Detection. WIT Transactions on Information and Communication Technologies, 25, 381–390. https://doi.org/10.2495/DATA000371