United International Journal of Engineering and Sciences
(UIJES – A Peer-Reviewed Journal); ISSN:2582-5887 | Impact Factor:8.075(SJIF)
Volume 5 | Special Issue 1 | 2025 Edition
National Level Conference on "Advanced Trends in Engineering Science & Technology" – Organized by RKCE

CYBER THREAT DETECTION BASED ON ARTIFICIAL NEURAL NETWORKS USING EVENT PROFILES

S Ranga Swamy Reddy[1], M Ramesh[2], CH Dileep[3], P Rajesh[4]

[1,2,3,4]*UG Scholars, Department of Computer Science and Engineering | R K College of Engineering*

Vijayawada | India

sakamreddy2004@gmail.com, rameshmadduluri369@gmail.com, dileepchelluboina143@gmail.com,

pr876605@gmail.com

***Abstract*-One of the major challenges in cybersecurity is the provision of an automated and effective cyber-threats detection technique. In this paper, we present an AI technique for cyber-threats detection, based on artificial neural networks. The proposed technique converts multitude of collected security events to individual event profiles and use a deep learning-based detection method for enhanced cyber-threat detection. For this work, we developed an AI-SIEM system based on a combination of event profiling for data preprocessing and different artificial neural network methods, including FCNN, CNN, and LSTM. The system focuses on discriminating between true positive and false positive alerts, thus helping security analysts to rapidly respond to cyber threats. All experiments in this study are performed by authors using two benchmark datasets (NSLKDD and CICIDS2017) and two datasets collected in the real world. To evaluate the performance comparison with existing methods, we conducted experiments using the five conventional machine-learning methods (SVM, k-NN, RF, NB, and DT). Consequently, the experimental results of this study ensure that our proposed methods are capable of being employed as learning-based models for network intrusion-detection, and show that although it is employed in the real world, the performance outperforms the conventional machine-learning methods.**

Keywords: Cyber-Threats, Machine-Learning Methods, KNN model, CNN model

I.INTRODUCTION

With the increase in-depth integration of the Internet and social life, the Internet is changing how people learn and work, but it also exposes us to increase consequential security threats. Where to know that and how to identify various network attacks, mainly not previously seen attacks. It is an important issue to be solved immediately. There are two primary systems for detecting cyber-threats and network intrusions. It is still hard to know and detect intrusions against intelligent network attacks owing to their high false alerts and the huge amount of security data in this present work our objective is to achieve an automated and effective cyber- threats detection framework using Deep Learning (DL)[1]. It learns normal and threat patterns from collected data, considering the frequency of their occurrence. By minimizing the false positive alerts, it assists the security analysts to rapidly respond to cyber threats. The target is to increase the detection performance by the elimination of improper, noisy

and irrelevant features. Classification algorithms are used to predict the class of an object [2]. In a dataset, the large number of data which may be irrelevant features or redundant features have an important effect in decreasing the accuracy of classification methods. These can increase the complexity of classifiers [3]. The experimental results on different size of dataset demonstrate the effective performance of the proposed data-model. In this I proposed PSO algorithm for better results. It ranks all the attributes and select the features then SVM is employed to classify data. The aim of this is to improve classification accuracy by using PSO algorithm. In this I use NDLKKD CUP dataset it was a developed as a variant of the famous KDD-99 dataset, used for anomaly detection. It has a training set with 125,973 instances and testing set with 22,543 samples, and includes 41 features and 5 classes. These features of this dataset can be divided into three data types i.e., nominal, binary, and numeric[4].

II . LITERATURE SURVEY

In this the main tasks of this is selection of features and selection of classification methods. For this they use PSO as feature selection and SVM as fitness function of PSO. By using this they help to optimize the feature selection process, since it increases classification accuracy but keeps computational resources minimum. Therefore, a good feature selection method based on number of features analysed that sample classification is needed in order to speed up the processing rate, predictive accuracy, and to avoid incomprehensibility. Says that the Genetic algorithm isa powerful tool for reducing the time for finding near optimal subsets of feature from large sets in this the author said two versions for problem of feature selection First version of problem leads to unconstrained combinational optimization in which the error rate is the search criteria. Second version of problem leads to constrained combinational optimization in which the error rate serves as a constraint and the number of features[5].

discussed that from the last decade, anomaly detection has attracted the attention of many researchers to overcome the weakness of signature-based IDSs in detecting novel attacks, and KDDCUP'99is the mostly widely used data set for the evaluation of these systems. It reduces false positive rate based on DAPRA 98 and then later it is updated as kDD 99 dataset. The basic two issues in in this are highly affects the performance of evaluated systems and second results very poor evaluation of anomaly detection approaches. For this issue NSL-KDD new dataset is proposed. It consists of complete selected records of the KDD dataset and it cannot suffer from any shortcomings. In this the author says that the reason for this aspect is distribution of different types of attacks imbalanced and ANN is unstable as it often coverage to local minimum. To solve above two problems the authors proposed FC-ANN. To enhance the detection precision for low frequent attacks and detection stability. The general procedure of FC-ANN is as follows: first fuzzy clustering technique is used to generate different training subsets. Subsequently, based on different training subsets, different ANN models are trained to formulate different base models. Finally, a meta-learner, fuzzy aggregation module, is employed to aggregate these results. Experimental results on the KDD CUP 1999 dataset show that our proposed new approach, FC-ANN outperforms detection precision and detection stability [6-7].

Digitalization has become a part today by which every walk of our life has been computerized, and it has made our life more flexible. On one side, we are happy to take the privilege of digitalisation. On the other side, security of our information in the internet is the most concerning element. A variety of security mechanisms, namely cryptography, algorithms which provide access to protected information, and authentication including biometric and stenography, provide security to our information on the Internet. In spite of the above mechanisms, recently, artificial intelligence (AI) has also contributed towards strengthening information security by providing machine learning and deep learning-based security mechanisms. The artificial intelligence (AI) contribution to cybersecurity is important as it

serves as a reaction and a response to hackers' malicious actions. The purpose of this paper is to provide a brief about recent papers that are contributing the information security by using machine learning and deep learning techniques [8].

III. METHODOLOGY

This section describes the dataset used, preprocessing steps, model implementation, prediction intervals, and evaluation metrics for assessing the performance of explanation detects threats most accurately and with fewer false alarms [9-10]

1. Start

This is the entry point of the system.

It indicates the beginning of the process to detect cyber threats using AI.

2. Upload Dataset

The user uploads a dataset containing network traffic records.

Example datasets: NSL-KDD, CICIDS2017, etc.

These datasets include features like duration, protocol type, source IP, destination IP, etc.

Purpose: To provide input data that contains both normal and attack behaviours.

3. Data Preprocessing

This step prepares the raw dataset for analysis.

- **Clean the data:**
  - Remove missing (null or NaN) values.
  - Eliminate duplicate and irrelevant entries.
  - Data transformation:
  - Normalize or scale numerical values.
- **Split the data:**
  - Usually into 80% training and 20% testing datasets.
  - Why important? Ensures that the data fed into the models is clean, consistent, and suitable for learning.

4. Feature Extraction

- Extract important features from the raw data.
- Use TF-IDF (Term Frequency-Inverse Document Frequency):
- Converts security events or logs into numerical vectors.
- Highlights important terms (features) that occur frequently in malicious patterns.
- This helps the model focus on the most informative features.

5. Feature Selection using PSO (Particle Swarm Optimization)

- PSO is a bio-inspired optimization technique.
- It selects the best features from the dataset by:
- Ranking all features based on their importance.
- Eliminating noisy or redundant features.
- After feature selection, the data is more refined for model input.

- Improves model accuracy and reduces computation time.

6. Apply Algorithms

Apply different Machine Learning and Deep Learning models:

- Traditional ML Algorithms:
- SVM (Support Vector Machine)
- K-NN (K-Nearest Neighbours)
- Naive Bayes
- Decision Tree
- Deep Learning Models:
- FCNN (Fully Connected Neural Network)
- CNN (Convolutional Neural Network)
- LSTM (Long Short-Term Memory)

Each model is trained using the training dataset and tested using the test dataset.

7. Model Evaluation

Evaluate the performance of each algorithm.

- Accuracy
- Precision
- Recall
- F1-score
- False Positive Rate (FPR)

The goal is to find which model detects threats most accurately and with fewer false alarms.

8. Graphical Comparison

Plot graphs comparing the accuracy and efficiency of each algorithm.

Visualization helps understand which model is better for real-world implementation.

9. Results & Conclusion

Final outcome is the best-performing model.

It can now be used to detect cyber threats in real-time.

The system can also be improved continuously with new data.

10. End

Marks the completion of the detection process.

Modules description:

- Upload Data set
- Pre processing
- Apply algorithm
- Compare the graph

Upload Data set:

- collecting the data from the www.kaggle.com, which cyber threads data into the XSL or CSV, and we upload it into the application/system
- **Pre-processing**
- pre-processing the data, which was uploaded, i.e., removing the null values and Nan and dividing the data for training and testing.
- **Apply the algorithm**
- apply algorithms (neural network, SVM, KNN, navye, etc) and the individual accuracies.
- **Compare graph**
- We plot a graph among the accuracies of the algorithms.

IV. RESULTS

This figure displays the front-end interface of the AI-SIEM (Artificial Intelligence Security Information and Event Management) system developed in this study. Figure 1: Interface of the Cyber Threats Detection Using AI. It provides a user-friendly view of the system's threat detection process, alert classification, and system status monitoring. The interface was designed for real-time interaction and to assist cybersecurity analysts in detecting and acting upon potential threats efficiently. This figure (not shown here) would typically illustrate metrics such as precision, recall, F1-score, and accuracy, used for evaluating the performance of different models. Figure 2: Demonstration of Performance Metrics ensures a balanced view of how the models perform, especially in differentiating between true and false positives—an essential factor in intrusion detection systems.

The bar chart compares precision values across various machine learning and deep learning models. Figure 3: Comparison of Machine Learning Models This performance comparison is critical to the research aim of evaluating the effectiveness of AI-based cyber-threat detection compared to traditional models.

Model Precision Summary (as shown in Figure 3):

- **CNN (Convolutional Neural Network)**: Highest precision (~90%), indicating superior performance in correctly identifying threats with minimal false positives.
- **SVM & RF (Random Forest)**: Competitive precision (~75%), good but not as robust as CNN.
- **k-NN**: Moderate precision (~68%).
- **LSTM (Long Short-Term Memory)**: Lower precision (~22%), likely due to sequence data challenges or training data limitations.
- **NB (Naive Bayes)**: Low precision (~35%), impacted by its assumption of feature independence.
- **DT (Decision Tree)**: Very low precision (~5%), possibly due to overfitting or poor generalization.

Fig 1 Interface of the Cyber Threats detection using AI

Fig 2 Demonisation of Performance matrices

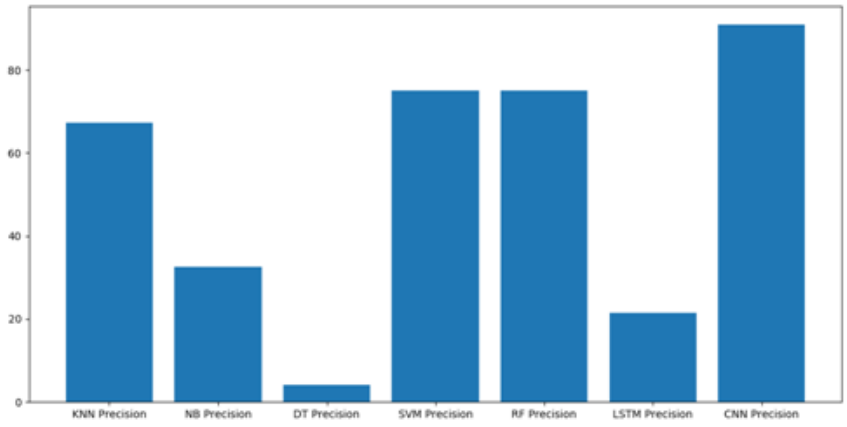

Figure 3 Comparison of Machine Learning models

V. CONCLUSION AND FUTURE SCOPE

In this paper, we have proposed the AI-SIEM system using event profiles and artificial neural networks. The novelty of our work lies in condensing very large-scale data into event profiles and using the deep learning-based detection methods for enhanced cyber-threat detection ability. The AI-SIEM system enables security analysts to deal with significant security alerts promptly and efficiently by comparing long-term security data. By reducing false positive alerts, it can also help the security analysts to rapidly respond to cyber threats dispersed across a large number of security events. For the evaluation of performance, we performed a performance comparison using two benchmark datasets (NSLKDD, CICIDS2017) and two datasets collected in the real world. First, based on the comparison experiment with other methods, using widely known benchmark datasets, we showed that our mechanisms can be applied as one of the learning-based models for network intrusion detection. Second, through the evaluation using two real datasets, we presented promising results that our technology also outperformed conventional machine learning methods in terms of accurate classifications.

VI. REFERENCES

[1] Tu Chung-Jui, Li-Yeh Chuang. Feature selection using PSO-SVM 2007 IAENG International Journal.

[2] W. SIEDLECKI, J. SKLANSKY A NOTE ON GENETIC ALGORITHMS FOR LARGE-SCALEFEATURESELECTION".

[3] M. Tavallaee, E. Bagheri, W. Lu, A.A. Ghorbani, A detailed analysis of the KDD CUP 99 data set in: Computational Intelligence for Security and Defense Applications, 2009. CISDA 2009.IEEE Symposiumon,2009.

[4] G Wang, J Hao, L Huang A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering Vol 37, Issue 9th September 2010 , Pages6225-6232.

[5] Thiyagarajan Paramasivan a Review on Cyber Security Mechanisms Using Machine and Deep Learning Algorithms in book: Advances in Information Security, Privacy, and Ethics(pp.23-41).

[6] Jeng-Fung Chen, Q Hung Do, OrcID and Ho-Nien Hsieh Training Artificial Neural Networks by a Hybrid PSO-CS Algorithm 2015, vol 8, pg292-308.

[7] Mujahid H. Khalifa; Marwa Ammar; Wael Ouarda "Particle swarm optimization for deep learning of convolution neural network" 2017 Sudan Conference on Computer Science and Information Technology (SCCSIT).

[8] Mohammed Harun Babu R, Vinaya Kumar R, Soman KP A short review on Applications of Deep learning for Cyber security 2018 CorpusID:56390787.

[9] T. Kim, S. C. Suh, H. Kim, J. Kim, and J. Kim, An encoding technique for CNN-based network anomaly detection, in Proc. IEEE Int. Conf. Big Data (IEEE Big Data), Seattle, WA, USA, Dec. 2019, pp.29602965.

[10] Shahrzad Zargari; Dave Voorhis , Feature Selection in the Corrected KDD-dataset 2012 Third International Conference on Emerging Intelligent Data and Web Technologies.