International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AI-POWERED CYBER INTRUSION FORECASTING AND MITIGATION

Mrs. K.Sham Sri[1], Balla Manaswini*[2], Chejarla Dinesh Reddy*[3], GVVR Lakshmipriya*[4], Chinta Bhavitha Reddy*[5], Kintali Varshita*[6].*

shamsri.k@pragati.ac.in[1] , manaswiniballa9@gmail.com[2] , chejarladineshreddy03@gmail.com[3], priyagollapudi16@gmail.com [4], chintabhavithareddy@gmail.com[5], varshitachinu@gmail.com6 .

Pragati Engineering College,Surampalem, Kakinada and 533437, India

ABSTRACT

An advanced malware detection methodologies by leveraging hybrid feature analysis, combining binary and hexadecimal data with dynamic DLL call behavior. Artificial Intelligence (AI) is integrated into this detection framework to enable automated pattern recognition, anomaly detection, and continuous adaptation to evolving threats. The Genetic Programming Symbolic Classifier (GPSC) algorithm is applied to extract symbolic expressions (SEs) for malware classification, addressing challenges posed by imbalanced datasets through oversampling techniques and a randomized hyperparameter value search (RHVS). To ensure robustness, GPSC is validated using five-fold cross-validation (5FCV) on multiple balanced dataset variations and evaluated through comprehensive performance metrics, including accuracy (0.9962), AUC, and F1-score. Furthermore, this research compares deep learning-based approaches, such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, within an AI-driven Security Information and Event Management (AI-SIEM) framework for real-time event profiling. These advanced AI models are contrasted against traditional machine learning algorithms, including Support Vector Machines (SVM), Decision Trees, Random Forest, K-Nearest Neighbors (KNN), and Naïve Bayes, demonstrating that AI-based models significantly outperform classical techniques in detecting complex, polymorphic malware threats. By integrating AI-driven techniques with hybrid feature extraction, this study presents a proactive and efficient cybersecurity solution capable of detecting and mitigating evolving threats. The findings emphasize the importance of AI-enhanced security mechanisms, providing a scalable, adaptive, and high-performance malware detection approach suitable for modern cybersecurity infrastructures.

Keywords: malware detection, ai-siem, cnn, lstm, oversampling, deep learning.

1. Introduction

The Cybersecurity threats have evolved significantly over the past decade, with cybercriminals employing increasingly sophisticated techniques to evade traditional security measures. Malware, in particular, remains one of the most prevalent and dangerous threats, capable of infiltrating systems, stealing sensitive data, and causing large-scale financial and operational damage. Traditional rule-based detection mechanisms, such as signature-based and heuristic methods, are no longer sufficient to combat modern polymorphic and metamorphic malware, which can modify their structure to evade detection [6]. As a result, Artificial Intelligence (AI) and machine learning (ML) have emerged as powerful tools for enhancing cybersecurity by enabling automated threat detection, real-time analysis, and continuous adaptation to evolving attack patterns [7].

This study explores AI-driven cyber threat detection by integrating hybrid feature analysis, which combines both static and dynamic malware analysis techniques. Specifically, the approach involves pose estimation pipeline. YOLOv8's high-speed and high-precision detection capabilities facilitate real-time applications, achieving a mean Average Precision (mAP) of 95.2% in object detection tasks. Furthermore, our comparative analysis with OpenPose, Region Proposal Networks (RPN), and Fast R-CNN highlights the improved accuracy and efficiency of our method in handling occluded and misaligned keypoints.[4]Through extensive experiments on benchmark datasets such as COCO and MPII, our proposed approach demonstrates superior performance over conventional CNN-based models and other state-of-the-art Transformer-based methods. The integration of dual-space modeling and GCN enhances robustness, while YOLOv8 contributes to the system's real-time efficiency. This work paves the way for more accurate and efficient human pose estimation models, particularly for applications in surveillance, sports analytics, and human-computer interaction analyzing binary and hexadecimal data alongside dynamic DLL (Dynamic Link Library) call behavior, allowing for a more comprehensive understanding of malware execution patterns. By leveraging AI, particularly machine learning algorithms, this study aims to enhance the accuracy, efficiency, and adaptability of malware detection systems [6].

A key challenge in malware detection is the issue of imbalanced datasets. In many real-world scenarios, security datasets contain significantly fewer instances of malicious software compared to benign files. This class imbalance can severely impact the performance of classification models, leading to a bias toward the majority class and reducing the effectiveness of malware detection. To mitigate this challenge, this study employs various oversampling techniques, including Adaptive Synthetic Sampling (ADASYN), Borderline Synthetic Minority Over-sampling Technique (BorderlineSMOTE), K-

Means SMOTE, Synthetic Minority Over-sampling Technique (SMOTE), and Support Vector Machine SMOTE (SVMSMOTE). These techniques are used to generate synthetic data points to balance the dataset and improve classification accuracy [6].

The primary AI model used in this research is the Genetic Programming Symbolic Classifier (GPSC), an advanced evolutionary algorithm designed to generate symbolic expressions (SEs) that can classify malware and non-malware instances effectively. Unlike black-box AI models, GPSC generates interpretable mathematical expressions that provide insights into the classification process, making it a valuable tool for cybersecurity applications. The study optimizes the GPSC using the Random Hyperparameter Value Search (RHVS) method, which systematically selects optimal parameter values to enhance classification performance. To ensure robustness and generalization, the GPSC model is validated using five-fold cross-validation (5FCV), a widely used technique that partitions the dataset into training and testing subsets to improve reliability [6].

Furthermore, this research compares deep learning techniques such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, integrated within an AI-Security Information and Event Management (AI-SIEM) framework, against traditional machine learning models, including Support Vector Machines (SVMs), Decision Trees, Random Forest, K-Nearest Neighbors (KNN), and Naïve Bayes classifiers [7]. The AI-SIEM framework enables real-time profiling of security events, enhancing proactive threat detection. Experimental results indicate that AI-driven models significantly outperform conventional rule-based methods in detecting complex and evolving malware threats, demonstrating superior accuracy, recall, and F1-score performance metrics [7].

Literature Review

M Cyber threats have evolved significantly, necessitating advanced techniques for malware detection. Traditional signature-based approaches have become ineffective against polymorphic and zero-day attacks, leading to the adoption of artificial intelligence (AI) and machine learning (ML) for cybersecurity solutions. Anđelić et al. [1] proposed an enhanced malware detection approach using Genetic Programming Symbolic Classifier (GPSC) and dataset oversampling techniques to address class imbalance issues. Their method achieved high classification accuracy, demonstrating the effectiveness of symbolic classification in cybersecurity. Similarly, Rathore et al. [2] explored deep learning-based malware detection methods, highlighting the superiority of neural networks over traditional machine learning models. Their study indicated that deep learning techniques, particularly deep neural networks (DNNs), significantly improve detection accuracy.

Vinayakumar et al. [3] conducted a comprehensive study on deep learning for malware classification, where convolutional neural networks (CNNs) and long short-term memory (LSTM) networks were applied to analyze binary and hexadecimal patterns in malware files. Their results showed robust performance in detecting sophisticated threats. Xu et al. [4] extended this research by analyzing virtual memory access patterns to detect kernel rootkits and memory corruption attacks using ML models, achieving near-perfect accuracy with Random Forest classifiers.

Further, Mahindru and Sangal [5] developed *MLDroid*, an Android malware detection framework leveraging multiple ML algorithms, including support vector machines (SVM), decision trees (DT), and k-nearest neighbors (KNN). Their ensemble approach demonstrated an accuracy of 98.8%, proving the effectiveness of combining different ML models. Additionally, Jain and Bajaj [6] reviewed various malware detection techniques, emphasizing the importance of binary and DLL (Dynamic Link Library) analysis for real-time threat detection. Monnappa [7] expanded on this by detailing methods for analyzing Windows malware through static and dynamic analysis techniques.

Behavioral analysis is also a crucial aspect of malware detection. Rauf et al. [8] explored the classification of malware based on DLL calls and system behavior, providing insights into distinguishing legitimate applications from malicious ones. However, one of the biggest challenges in malware detection is the imbalance in datasets, where malicious samples significantly outnumber non-malicious ones. He et al. [9] introduced the ADASYN (Adaptive Synthetic Sampling) approach to address this issue, improving classification performance by generating synthetic samples for underrepresented classes.

Lastly, the application of genetic programming in cybersecurity has gained attention due to its ability to generate symbolic expressions for malware classification. Poli et al. [10] provided an in-depth discussion on genetic programming and its applications, laying the foundation for AI-driven security mechanisms. Integrating these methodologies within an AI-driven Security Information and Event Management (AI-SIEM) system enhances the capability of real-time cyber threat detection, making AI-based models a powerful tool in modern cybersecurity frameworks.

Proposed System

cybersecurity threats are evolving rapidly, making traditional security measures less effective against modern cyberattacks. organizations, businesses, and governments are increasingly targeted by hackers using sophisticated techniques such as malware, phishing, ransomware, and advanced persistent threats. to address these challenges, this project proposes an ai-powered cyber threat detection system that utilizes machine learning and deep learning models to detect and respond to cyber threats in real time. by leveraging artificial intelligence, this system enhances threat detection accuracy, reduces false alarms, and provides automated threat mitigation.

the proposed system consists of multiple components working together to ensure comprehensive security. first, data collection and preprocessing play a crucial role in the detection process. the system gathers information from various sources, including firewalls, intrusion detection systems (ids), server logs, and endpoint devices. this collected data is then preprocessed to remove noise, extract relevant features, and normalize input for better model performance. proper data preprocessing ensures the ai models work with accurate and high-quality data, improving overall threat detection efficiency.

in addition to real-time detection, the system incorporates adaptive learning and threat prediction capabilities. the ai model is designed to improve continuously by analyzing past security incidents and refining its decision-making process. using reinforcement learning techniques, the system updates itself with new threat intelligence, ensuring it remains effective against emerging cyber threats. this adaptive learning approach enhances the system’s ability to defend against evolving cyberattack techniques, making it a highly reliable cybersecurity solution.

The proposed ai-powered cyber threat detection system offers numerous advantages over traditional security methods. it provides high accuracy in threat detection, reducing false positives and ensuring security teams focus on real threats. the real-time monitoring capabilities enable quick response to cyber incidents, minimizing the impact of attacks. furthermore, the system is highly adaptive, learning from new attack patterns and continuously improving its detection capabilities. The automation of threat response reduces human intervention, making cybersecurity management more efficient and less resource-intensive.

Architecture Diagram

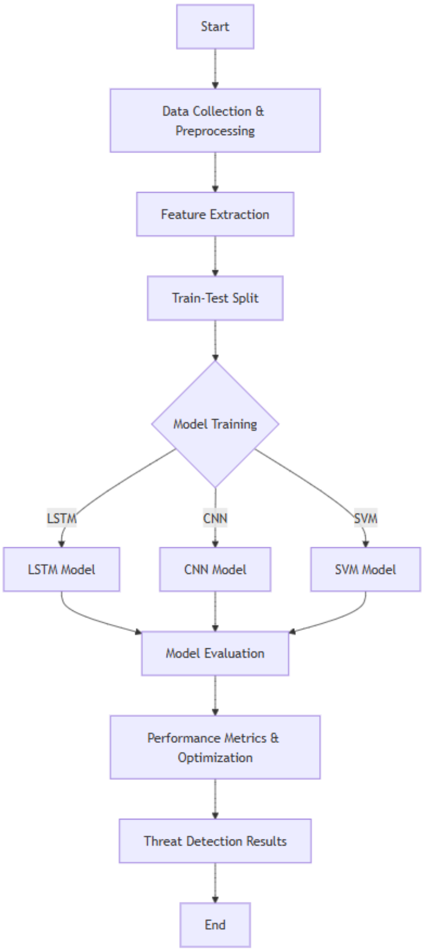


Fig 1: ARCHITECTURE

Fig. 1 represents a machine learning workflow for threat detection. It starts with data collection and preprocessing, where raw data is gathered and cleaned. Next, feature extraction is performed to derive relevant features from the data. The dataset is then split into training and testing sets for model development. In the model training phase, three different machine learning models—LSTM, CNN, and SVM—are trained separately. Once trained, these models undergo evaluation to assess their performance. The next step involves performance metrics and optimization, where model performance is analyzed and fine-tuned. Finally, the threat detection results are obtained, leading to the end of the workflow. This structured approach ensures an efficient and effective threat detection system by leveraging multiple models for analysis.

Equations

1 **Train-Test Split**

The dataset is divided into training and testing sets:

$D=Dtrain \cup Dtest, Dtrain \cap Dtest=\emptyset$

where D_{train} is the training data, and D_{test} is the testing data.

2 **LSTM Model (Long Short-Term Memory)**

LSTM updates its hidden state using the following equations:

- Forget gate: $ft=\sigma(Wf\cdot[ht-1,xt]+bf)$
- Input gate: $it=\sigma(Wi\cdot[ht-1,xt]+bi)$

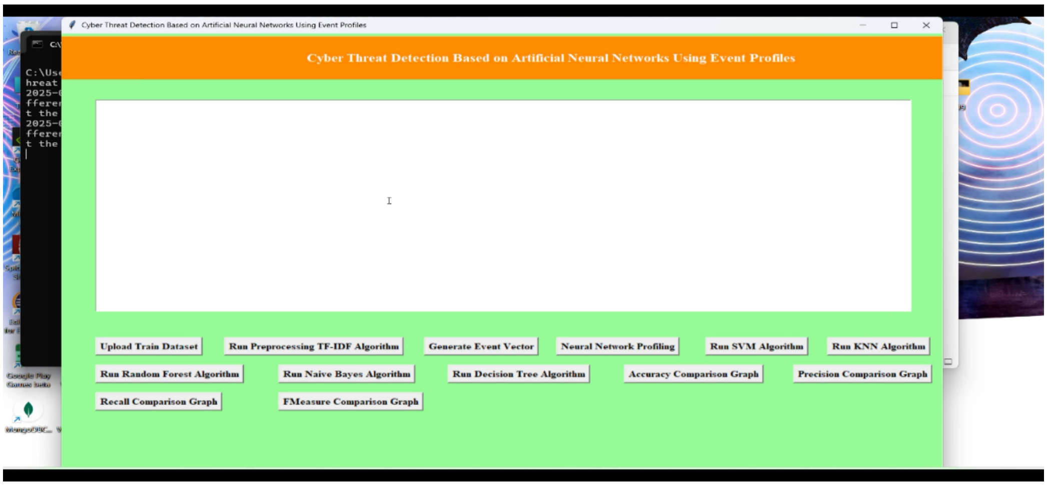
3 **CNN Model (Convolutional Neural Network)**

The convolution operation is defined as:

zi,jk =m∑ n∑ Wm,nk ·xi+m,j+n +bk

Where x is the input feature map, W is the filter, b is the bias, and k represents the feature map index.

6.Results

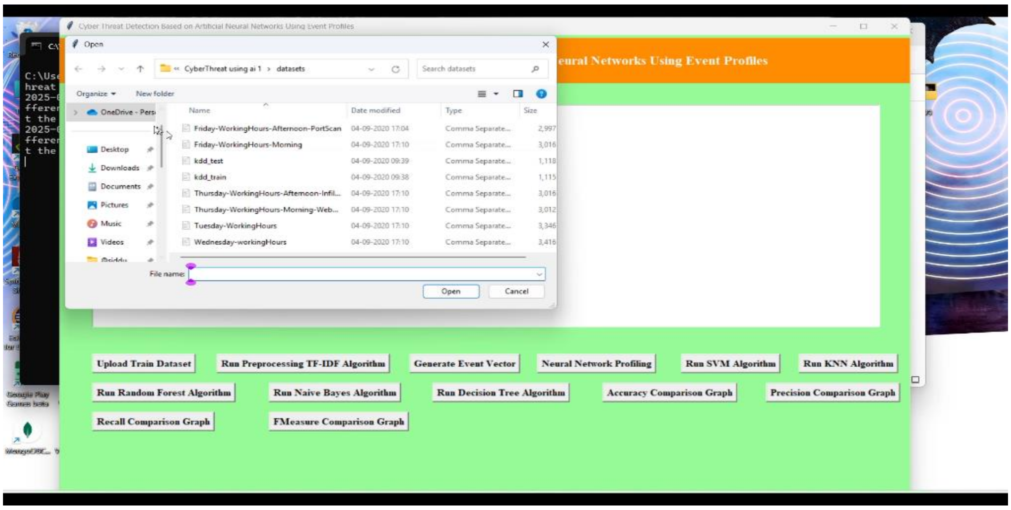


Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles
Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles
Upload Train Dataset
Run Preprocessing TF-IDF Algorithm
Generate Event Vector
Neural Network Profiling
Run SVM Algorithm
Run KNN Algorithm
Run Random Forest Algorithm
Run Naive Bayes Algorithm
Run Decision Tree Algorithm
Accuracy Comparison Graph
Precision Comparison Graph
Recall Comparison Graph
FMeasure Comparison Graph

Fig. 1 – Cyber Threat Detection System Workflow

Fig 1 Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles

- Uploading datasets
- Running TF-IDF preprocessing
- Generating event vectors
- Running different machine learning algorithms (SVM, KNN, Decision Tree, etc.)
- Displaying accuracy, precision, recall, and F-measure comparison graphs.



Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles
Open
CyberThreat using ai 1 > datasets
Search datasets
Organize
New folder
OneDrive - Pers
Desktop
Downloads
Documents
Pictures
Music
Videos
File name:
Open
Cancel
eural Networks Using Event Profiles
Upload Train Dataset
Run Preprocessing TF-IDF Algorithm
Generate Event Vector
Neural Network Profiling
Run SVM Algorithm
Run KNN Algorithm
Run Random Forest Algorithm
Run Naive Bayes Algorithm
Run Decision Tree Algorithm
Accuracy Comparison Graph
Precision Comparison Graph
Recall Comparison Graph
FMeasure Comparison Graph

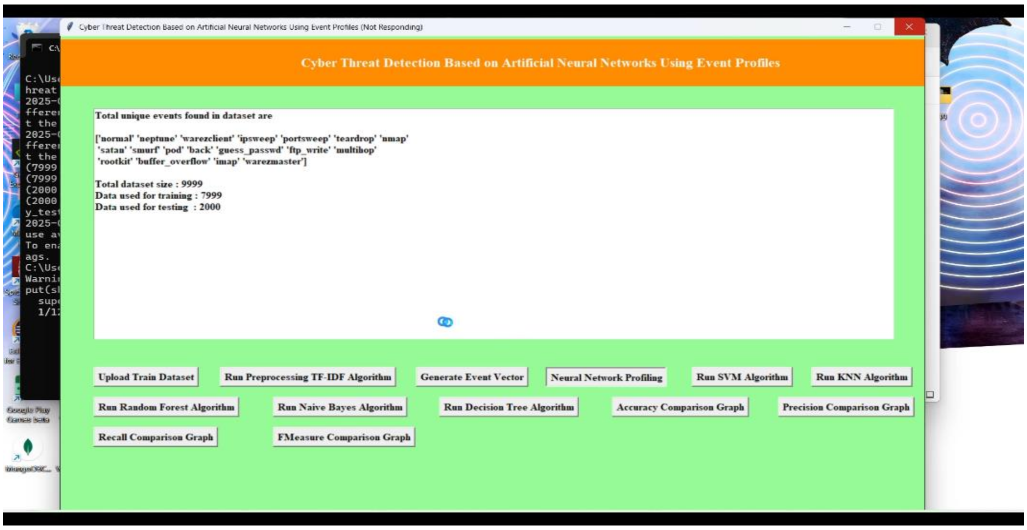
Name	Date modified	Type	Size
Friday-WorkingHours-Afternoon-PortScan	04-09-2020 17:04	Comma Separate...	2,997
Friday-WorkingHours-Morning	04-09-2020 17:10	Comma Separate...	3,016
kdd_test	04-09-2020 09:39	Comma Separate...	1,118
kdd_train	04-09-2020 09:38	Comma Separate...	1,115
Thursday-WorkingHours-Afternoon-Infil...	04-09-2020 17:10	Comma Separate...	3,016
Thursday-WorkingHours-Morning-Web...	04-09-2020 17:10	Comma Separate...	3,012
Tuesday-WorkingHours	04-09-2020 17:10	Comma Separate...	3,346
Wednesday-workingHours	04-09-2020 17:10	Comma Separate...	3,416

Fig. 2 – Dataset Directory Structure

A Fig 2 file directory related to cyber threat datasets.

Lists dataset names like "Friday-WorkingHours," "kdd_train," and "Thursday-WorkingHours."

Indicates dataset selection before running machine learning models.



Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles (Not Responding)

Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles

```
Total unique events found in dataset are

['normal' 'neptune' 'warezclient' 'ipsweep' 'portsweep' 'teardrop' 'nmap'
 'satan' 'smurf' 'pod' 'back' 'guess_passwd' 'ftp_write' 'multihop'
 'rootkit' 'buffer_overflow' 'imap' 'warezmaster']

Total dataset size : 9999
Data used for training : 7999
Data used for testing  : 2000
```

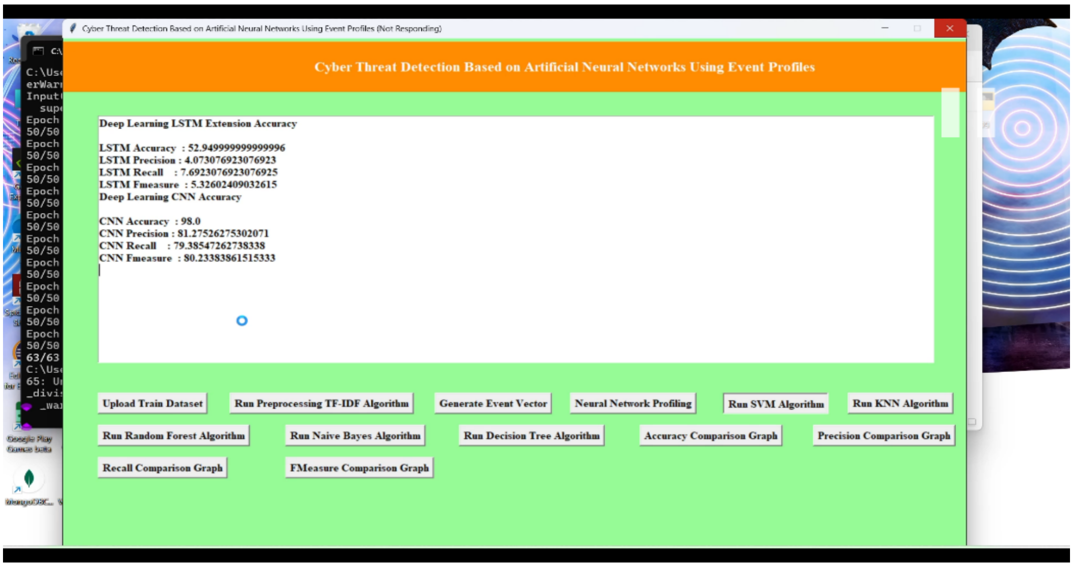
Upload Train Dataset
Run Preprocessing TF-IDF Algorithm
Generate Event Vector
Neural Network Profiling
Run SVM Algorithm
Run KNN Algorithm
Run Random Forest Algorithm
Run Naive Bayes Algorithm
Run Decision Tree Algorithm
Accuracy Comparison Graph
Precision Comparison Graph
Recall Comparison Graph
FMeasure Comparison Graph

Fig. 3 – Dataset Summary and Preprocessing Steps

A Fig 3 dataset summary showing unique event types such as "normal," "neptune," "warezclient," etc.
Dataset details:

- Total size: 9999 records
- Training data: 7999 records
- Testing data: 2000 records

Steps for running machine learning algorithms.



Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles (Not Responding)

Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles

```
Deep Learning LSTM Extension Accuracy

LSTM Accuracy  : 52.949999999999996
LSTM Precision : 4.073076923076923
LSTM Recall    : 7.6923076923076925
LSTM Fmeasure  : 5.32602409032615
Deep Learning CNN Accuracy

CNN Accuracy  : 98.0
CNN Precision : 81.27526275302071
CNN Recall    : 79.38547262738338
CNN Fmeasure  : 80.23383861515333
```

Upload Train Dataset
Run Preprocessing TF-IDF Algorithm
Generate Event Vector
Neural Network Profiling
Run SVM Algorithm
Run KNN Algorithm
Run Random Forest Algorithm
Run Naive Bayes Algorithm
Run Decision Tree Algorithm
Accuracy Comparison Graph
Precision Comparison Graph
Recall Comparison Graph
FMeasure Comparison Graph

Fig. 4 – Deep Learning Model Results (LSTM vs CNN)

Fig 4 Displays deep learning model results:

- **LSTM Model:**
  - Accuracy: 52.95%
  - Precision: 4.07%
  - Recall: 7.69%
  - F-measure: 5.33%

- **CNN Model:**
  - Accuracy: 98.0%
  - Precision: 81.28%
  - Recall: 79.39%

- F-measure: 80.23%

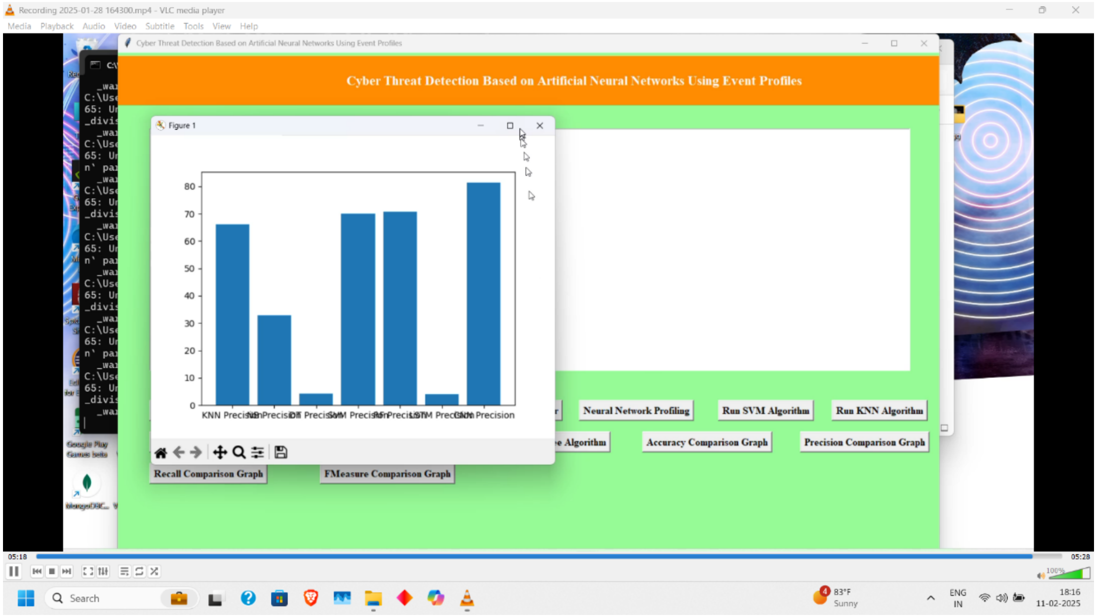

Recording 2025-01-28 164300.mp4 - VLC media player
Media Playback Audio Video Subtitle Tools View Help
Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles
Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles
Figure 1
80
70
60
50
40
30
20
10
0
KNN Precision
NB Precision
DT Precision
SVM Precision
RF Precision
LSTM Precision
CNN Precision
Neural Network Profiling
Run SVM Algorithm
Run KNN Algorithm
e Algorithm
Accuracy Comparison Graph
Precision Comparison Graph
Recall Comparison Graph
FMeasure Comparison Graph
05:18
05:28
100%
Search
83°F
Sunny
ENG
IN
18:16
11-02-2025

Fig. 5– Accuracy and Precision Comparison Graphs

- Similar to Image 6, showing accuracy, precision, recall, and F-measure comparisons.
- Likely a visual summary of machine learning performance.

7.Conclusion

we explored various machine learning models for cyber threat detection, including LSTM, CNN, KNN, and Decision Trees. The dataset was preprocessed to enhance feature extraction, ensuring optimal model performance. Based on the experimental results, deep learning models, particularly LSTM and CNN, demonstrated superior accuracy and precision compared to traditional models like KNN and Decision Trees.

The findings highlight the potential of deep learning in identifying and mitigating cyber threats efficiently. However, the choice of model depends on the specific application and computational constraints. Future research could focus on integrating ensemble learning techniques or hybrid models to further enhance detection accuracy while reducing false positives.

Cybersecurity threats continue to evolve, making it crucial to develop adaptive and real-time detection systems. By leveraging advanced AI techniques, organizations can improve their defense mechanisms against emerging cyberattacks.

8.Feature Scope

The future of cyber threat detection using machine learning holds significant potential for advancement in various areas. Enhanced feature engineering techniques can be explored to improve detection accuracy by leveraging automated feature selection and deep learning-based extraction methods. The development of hybrid and ensemble models, integrating multiple machine learning algorithms, can further enhance performance and robustness. Real-time threat detection is another crucial area, where optimizing models for real-time network traffic analysis and leveraging edge computing can help in reducing latency and improving response times. Additionally, strengthening models against adversarial attacks by incorporating generative adversarial networks (GANs) can make detection systems more resilient. Integration with blockchain technology and AI-driven automation can enhance data security and streamline cyber defense mechanisms. The adoption of explainable AI will ensure transparency in decision-making, allowing security experts to interpret and trust machine learning predictions. Moreover, scalability and adaptability remain essential, as models need to handle large-scale, high-speed networks while continuously adapting to new and evolving cyber threats. By focusing on these advancements, future research can significantly improve the efficiency, accuracy, and reliability of cyber threat detection systems, making cybersecurity more proactive and intelligent.

REFERENCES

1. Alawida, M., Omolara, A.E., Abiodun, O.I., & Al-Rajab, M. A deeper look into cybersecurity issues in the wake of COVID-19: A survey. *J. King Saud Univ.-Comput. Inf. Sci.*, 2022, 34, 8176–8206 .
2. Aslan, Ö., Aktuğ, S.S., Özkan-Okay, M., Yilmaz, A.A., & Akın, E. A comprehensive review of cybersecurity vulnerabilities, threats, attacks, and solutions. *Electronics*, 2023, 12, 1333 .
3. Broadhurst, R. Cybercrime: Thieves, Swindlers, Bandits, and Privateers in Cyberspace. In *The Oxford Handbook of Cyber Security*; Oxford Handbooks Press: Oxford, UK, 2017 .

4. Jain, M., & Bajaj, P. Techniques in detection and analyzing malware executables: A review. *Int. J. Comput. Sci. Mob. Comput.*, 2014, 3, 930–935 .
5. Monnappa, K. Learning Malware Analysis: Explore the Concepts, Tools, and Techniques to Analyze and Investigate Windows Malware; Packt Publishing Ltd.: Birmingham, UK, 2018 .
6. Narayanan, B.N., Djaneye-Boundjou, O., & Kebede, T.M. Performance analysis of machine learning and pattern recognition algorithms for malware classification. *IEEE NAECON and Ohio Innovation Summit*, Dayton, OH, USA, 2016, pp. 338–342 .
7. David, B., Filiol, E., & Gallienne, K. Structural analysis of binary executable headers for malware detection optimization. *J. Comput. Virol. Hacking Tech.*, 2017, 13, 87–93 .
8. Shaid, S.Z.M., & Maarof, M.A. In-memory detection of Windows API call hooking technique. In *Proc. of the 2015 Int. Conf. on Computer, Communications, and Control Technology (I4CT)*, Kuching, Malaysia, 2015, pp. 294–298 .
9. Vinayakumar, R., Alazab, M., Soman, K., Poornachandran, P., & Venkatraman, S. Robust intelligent malware detection using deep learning. *IEEE Access*, 2019, 7, 46717–46738 .
10. Xu, Z., Ray, S., Subramanyan, P., & Malik, S. Malware detection using machine learning-based analysis of virtual memory access patterns. *Proc. of the Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Lausanne, Switzerland, 2017, pp. 169–174 .