A Review on AI in Cybersecurity Threat Detection and Response Using CNN

Meenu[1], Gurkirat Singh[2], Ranju Marwaha[3],Sarita Borkar[4]

[1]Assistant Professor,[2]Assistant Professor,[3]Assistant Professor, [4]Assistant Professor Department of Computer Application, Sri Sukhmani Institute of Hospitality and Management, Dera Bassi, Punjab, India

ABSTRACT

Cybersecurity has become one of the most critical global concerns as cyberattacks grow in scale, diversity, and sophistication. Traditional security systems based on signatures and manual rule definitions are unable to detect newly emerging and zero-day threats. Recent advances in Artificial Intelligence (AI) and Deep Learning (DL) have shown promising results in automating threat detection. Convolutional Neural Networks (CNN), known for their superior feature extraction capability, are widely used for malware classification, intrusion detection, phishing analysis, and anomaly detection. CNNs can automatically learn spatial–temporal patterns from network traffic, system logs, and binary malware images without manual feature engineering. This review presents a detailed analysis of CNN-based methods for cybersecurity threat detection and response, highlights major research findings, identifies limitations, and discusses future directions.

INTRODUCTION

As cyber threats continue to evolve in complexity and scale, traditional cyber security measures often struggle to keep up with new attack techniques. This has led to the increasing adoption of Artificial Intelligence (AI) to enhance threat detection and response capabilities. Among the various AI techniques, Convolutional Neural Networks (CNNs) have proven to be particularly effective in analyzing complex data patterns and detecting anomalies in various cyber security domains.

Originally designed for image recognition, CNNs are now being applied to a range of cyber security challenges, including network traffic analysis, intrusion detection, malware detection, and phishing prevention. By automatically learning patterns from large datasets, CNNs can identify subtle anomalies that may indicate a cyberattack, such as DDoS attacks, botnets, or data exfiltration.

CNNs excel at processing high-dimensional data, making them ideal for identifying malicious activities in large-scale, complex environments. Their ability to detect emerging threats in real-time and adapt to new attack methods offers a powerful tool for cyber security professionals. Despite challenges like data labelling and interpretability, the use of CNNs in cyber security represents a significant step forward in proactive threat detection and mitigation.

LITERATURE REVIEW

Researchers worldwide have proposed CNN-based AI systems to identify malicious activities across networks, IoT devices, cloud platforms, and endpoint environments.

Shone et al. (2018) developed a deep CNN-based Intrusion Detection System using UNSW-NB15, showing higher accuracy than SVM and Random Forest.

Saxe & Berlin (2019) introduced one of the earliest CNN-based malware detectors by converting binary files into grayscale images, achieving strong generalization on unseen malware families.

Vinayakumar et al. (2021) designed a multi-layer CNN for large-scale network threat detection using CIC-IDS2017, significantly reducing false positives.

Kim & Park (2020) proposed a hybrid CNN-LSTM architecture for Android malware detection, combining spatial and sequential features.

Alom et al. (2022) applied CNNs for IoT attack classification, emphasizing lightweight architectures suitable for edge devices.

Zhang et al. (2021) used CNN and autoencoders for anomaly detection in cloud logs, improving zero-day attack detection.

Liu et al. (2023) explored transfer-learning-based CNNs for phishing URL detection, demonstrating improved performance with small datasets.

Mohan et al. (2024) proposed a CNN-autoencoder for DDoS detection using CIC-DDoS2019, providing robustness against noisy traffic.

Rahman et al. (2023) highlighted CNN strengths in identifying obfuscated malware samples by extracting pixel-level spatial patterns.

Haque et al. (2022) designed a CNN model for encrypted traffic classification, enabling threat detection without decrypting packets.

Sun et al. (2021) compared CNN, LSTM, and GRU for network intrusion detection and concluded that CNN delivers the best trade-off between accuracy and speed.

Overall, literature demonstrates that CNNs outperform traditional ML in identifying complex and evolving cyber threats.

PROPOSED METHODOLOGY

The proposed CNN-based cybersecurity threat detection methodology involves several sequential stages. First, raw cybersecurity data—such as network traffic captures, malware binaries, system logs, and IoT communication streams—are collected from datasets like CIC-IDS2017, UNSW-NB15, and custom organizational logs. The collected data undergo pre-processing, which includes noise removal, normalization, feature scaling, encoding (e.g., converting traffic to 2D matrices or malware binaries to images), and labeling. The processed data are then fed into a CNN architecture comprising convolutional layers for spatial pattern extraction, pooling layers to reduce dimensionality, and fully connected layers for classification. The model learns malicious patterns such as abnormal packet distributions, code structure anomalies, or unusual patterns in log sequences. Finally, the classification output determines whether a given input is benign or malicious, and automated response mechanisms such as blocking IPs, isolating devices, or generating real-time alerts are triggered.

Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a deep learning architecture designed to automatically extract hierarchical features from structured data. While originally developed for image processing, CNNs efficiently analyze cybersecurity data by transforming traffic flows, binary files, or log events into visual or matrix formats. A CNN consists of convolutional layers that detect local patterns, ReLU activation to introduce non-linearity, pooling layers to reduce complexity, and fully connected layers for final classification. The convolution operation enables CNNs to capture attack signatures, abnormal behavior, and hidden relationships within traffic patterns or malware code. This makes CNNs highly suitable for cybersecurity applications where feature engineering is complex and threats are dynamic.

DATA COLLECTION

- Network traffic (PCAP files)
- System logs (authentication logs, event logs)
- Malware binaries converted to grayscale images
- IoT/sensor network communication data

- Cloud service logs (API calls, VM activity)

PRE-PROCESSING

- Noise removal
- Normalization / feature scaling
- Label encoding
- Converting raw traffic to 2D arrays
- Splitting into training and testing datasets

FEATURE EXTRACTION

Initial convolutional layers extract low-level cybersecurity features like:

- Traffic anomalies
- Opcode distributions
- Binary file textures
- URL patterns
- Log event correlations

CNN MODEL LAYER

- Convolution Layer
- Pooling Layer
- Dropout Layer
- Fully Connected Layer
- Softmax Output Layer

PREDICTION

The CNN generates the final classification output:

- Normal Traffic
- Malware
- DDoS Attack
- Botnet Activity
- Phishing
- Anomalous Behaviour

Normal Traffic :- Normal traffic refers to the typical or expected flow of data across a network or system. This traffic represents regular user or application behaviour and is usually predictable, based on what is considered standard operations.

Malware: Malware is any software intentionally designed to **harm, exploit, or infiltrate** computers, networks, or devices. It can steal data, disrupt operations, spy on users, or give attackers unauthorized access.

DDoS Attack: A DDoS (Distributed Denial of Service) attack is a type of cyber attack where multiple systems (often compromised computers or devices) are used to flood a target system, service, or network with massive amounts of traffic, overwhelming its capacity and making it **unavailable** to legitimate users.

Botnet Activity: Botnet activity refers to the actions and operations carried out by a botnet, which is a network of compromised computers or devices(zombies) controlled by a malicious actor (botmaster). These infected devices are used to perform a variety of malicious activities, often without the knowledge of the device owners.

Phishing: Phishing is a type of cyber attack in which attackers impersonate legitimate entities (such as banks, social media platforms, or well-known companies) to trick individuals into revealing sensitive information, such as **passwords**, **credit card details**, or **personal data**. It's one of the most common forms of cybercrime and can lead to identity theft, financial loss, and data breaches.

Anomalous Behaviour: Anomalous behaviour in cyber security refers to actions, patterns, or events that deviate from the established normal behaviour of a system, network, or user. Identifying anomalous behaviour is crucial for detecting security threats, such as intrusions, data breaches, or malicious activities like malware infections, insider threats, and external attacks.

Anomalies can be detected in network traffic, user activity, file access patterns, or system operations. When something behaves unexpectedly, it might indicate that something malicious is happening, and it's worth investigating further.

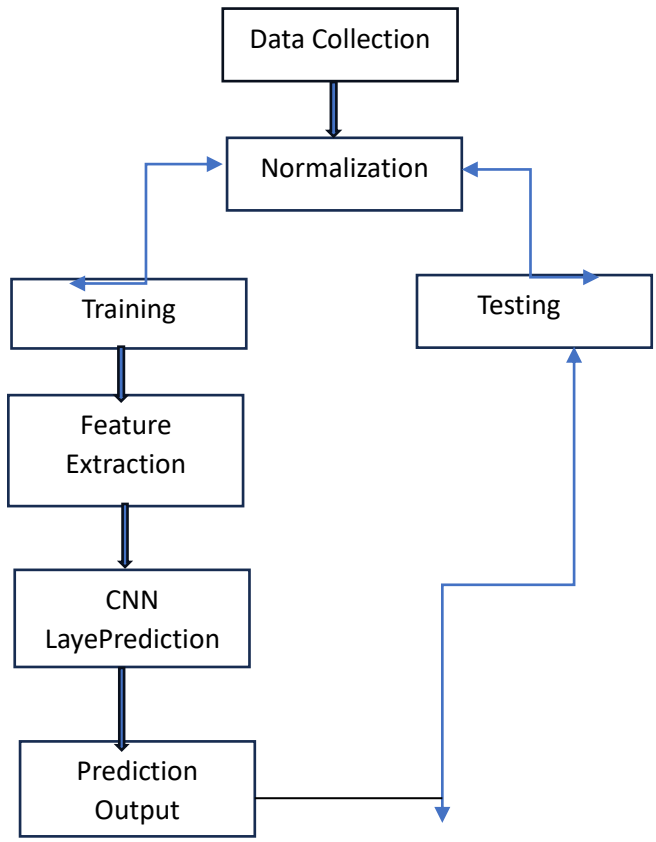

Figure 1.1 Framework of Proposed Model

EXPECTED RESULT

The expected outcome of applying CNN to cybersecurity threat detection is a significant improvement in identifying malicious activities with high accuracy and reduced false positives. CNNs are expected to detect both known and unknown attack patterns, outperforming traditional ML and signature-based systems. The model is anticipated to successfully classify DDoS attacks, malware variants, botnets, and anomalies in real time. Moreover, CNNs provide robustness against noisy traffic, obfuscation techniques, and encrypted data, making them suitable for real-world cybersecurity environments. Automated response mechanisms help reduce detection latency and enhance network resilience.

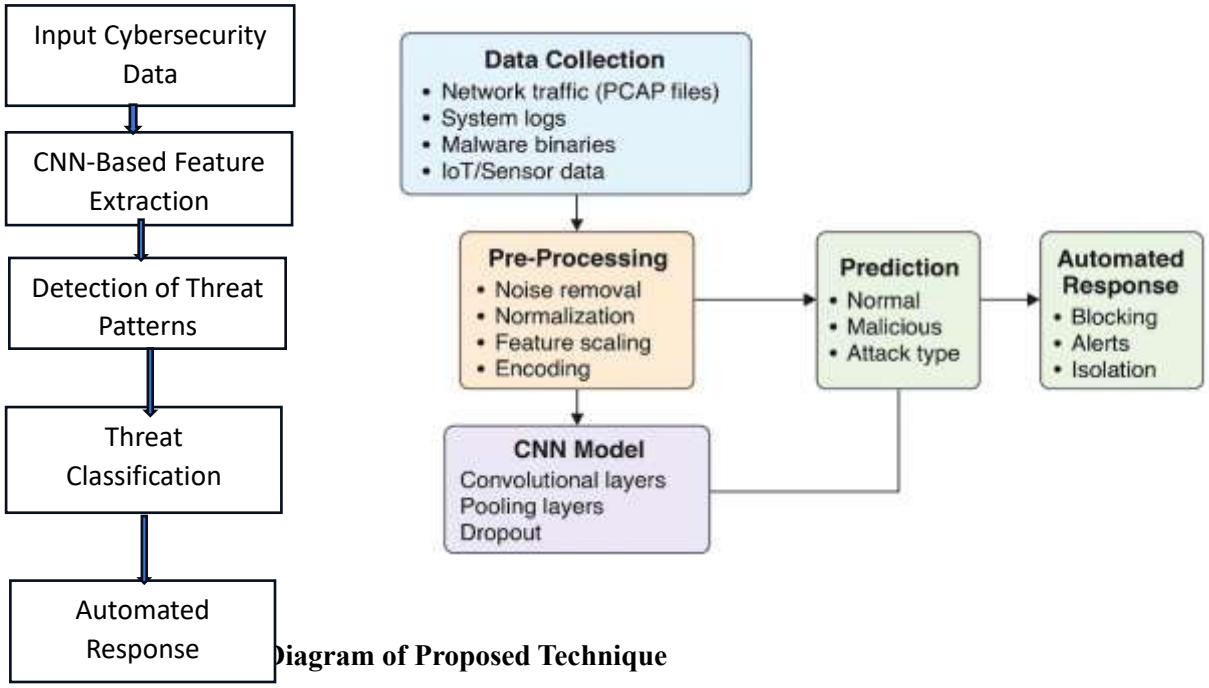

Diagram of Proposed Technique

Expected Improvements:

- Higher detection accuracy
- Better capture of complex patterns
- Reduced false alarms
- Improved generalization across environments

CONCLUSION

This review demonstrates that Convolutional Neural Networks provide powerful and scalable solutions for cybersecurity threat detection and response. CNNs automate feature extraction, making them highly efficient in analyzing complex network traffic, malware samples, and log data. Compared to traditional machine learning algorithms, CNNs exhibit improved accuracy, robustness, and adaptability to new threats. Although computational cost and data requirements remain challenges, ongoing research into lightweight architectures, federated learning, and hybrid CNN models continues to advance the field. CNN-based approaches show strong potential for deployment in real-time cybersecurity systems, enabling proactive and intelligent threat defence.

REFERENCES

1. Saxe & Berlin, 2019, “CNN for Static Malware Detection,” USENIX Security.
2. Shone et al., 2018, “Deep Learning IDS Using CNN,” Elsevier.
3. Vinayakumar et al., 2021, “CNN-Based Intrusion Detection,” IEEE Access.
4. Kim & Park, 2020, “Hybrid CNN-LSTM for Android Malware,” Springer.
5. Alom et al., 2022, “Deep CNN for IoT Cyber Attack Classification,” Sensors.
6. Zhang et al., 2021, “Anomaly Detection in Cloud Logs with CNN,” Applied Intelligence.
7. Mohan et al., 2024, “CNN-Autoencoder for DDoS Detection,” ACM Computing Surveys.
8. Rahman et al., 2023, “CNN for Obfuscated Malware Recognition,” Journal of Cybersecurity.
9. Haque et al., 2022, “Encrypted Traffic Classification using CNN,” IEEE Transactions on Networking.
10. Sun et al., 2021, “Comparative Study of CNN and RNN for IDS,” Elsevier.
11. Liu et al., 2023, “Transfer Learning CNN for Phishing URLs,” Wiley.
12. Goodfellow et al., 2016, “Deep Learning,” MIT Press.
13. LeCun et al., 2015, “CNN Architectures Overview,” Nature.