AI-Based Multi Cyber Threat Detection

Mrs. K. Sai Veena
Assistant Professor
Dept. of CSE (AI & ML)
Jyothishmathi Institute of Technology and Science
Karimnagar, Telangana, India

Mrs. Dr. V. Neelima
Associate Professor
Dept. of CSE (AI & ML)
Jyothishmathi Institute of Technology and Science
Karimnagar, Telangana, India

N. Nagaraju
Dept. of CSE (AI & ML)
Jyothishmathi Institute of Technology and Science
Karimnagar, Telangana, India
nimmanagaraju6@gmail.com

P. Meghana
Dept. of CSE (AI & ML)
Jyothishmathi Institute of Technology and Science
Karimnagar, Telangana, India
palthepumeghana618@gmail.com

B. Sowmya
Dept. of CSE (AI & ML)
Jyothishmathi Institute of Technology and Science
Karimnagar, Telangana, India
balesowmyabalesowmya@gmail.com

G. Charvaka
Dept. of CSE (AI & ML)
Jyothishmathi Institute of Technology and Science
Karimnagar, Telangana, India
charvaka01@gmail.com

Abstract—**The rapid expansion of digital communication platforms, cloud computing, online banking, and social networking services has significantly increased the scale and complexity of cyber threats. Modern attackers no longer depend on simple malware or static phishing emails; instead, they use advanced techniques such as social engineering, deepfake media, and malicious QR codes to deceive users and bypass conventional security mechanisms. These attacks are often carefully crafted and multi-vector in nature, making traditional rule-based and signature-based cybersecurity solutions increasingly ineffective. Furthermore, most existing security systems are designed to detect only a single type of threat, which limits their ability to provide comprehensive protection in real-world environments. This paper presents an AI-based multi cyber threat detection framework that integrates machine learning, deep learning, and natural language processing to provide real-time and intelligent security. The proposed system analyzes textual data from emails and messages, visual content from images and videos, and encoded information extracted from QR codes in order to detect phishing attempts, identify manipulated deepfake media, and recognize malicious QR-based exploits. Experimental evaluation shows that the proposed approach achieves improved accuracy, precision, and recall compared to traditional single-threat detection systems, while maintaining low computational latency, making it suitable for deployment in modern digital infrastructures.**

Index Terms—**Cyber security, phishing detection, deepfake detection, QR code attacks, social engineering, artificial intelligence, deep learning, natural language processing.**

I. Introduction

The modern digital world relies heavily on online communication, cloud computing, and interconnected services. Email, social media, e-commerce, and online banking have become essential to daily life. However, this growth has also increased cybercrime. Phishing emails, fake websites, social engineering messages, deepfake videos, and malicious QR codes are widely used to deceive users and steal sensitive information [1].

Traditional cybersecurity solutions depend on predefined rules and known attack signatures. While effective against previously identified threats, they fail when attackers modify their methods. Phishing URLs change frequently, deepfake videos convincingly imitate real people, and QR codes hide malicious links inside images [2].

Artificial intelligence has emerged as a powerful tool for cyber defense. Machine learning algorithms can learn from large datasets, deep learning models can analyze images and videos, and natural language processing enables systems to understand written communication. However, most AI-based security solutions focus on a single type of threat. There is a need for an integrated framework capable of detecting multiple cyber threats in real time [3].

A. Motivation

Cyber attackers no longer rely on simple malware or static phishing emails. Modern attacks are carefully crafted, personalized, and delivered through multiple channels such as email, social media, QR codes, and multimedia. Attackers often use deepfake videos, fake voice messages, and spoofed websites to increase credibility and bypass traditional security filters. This makes human users the weakest link in cybersecurity. An intelligent automated system capable of understanding language, analyzing images, and decoding hidden malicious links is therefore essential [4].

B. Problem Statement

Most existing cybersecurity solutions are designed to detect only one form of attack such as phishing, malware, or fake websites. When multiple attack vectors are used simultaneously, these systems fail to provide complete protection. There is a need for a unified detection framework that can analyze heterogeneous cyber data and identify multiple types of threats in real time with high accuracy and low latency [5].

II. Related Work

Phishing detection using machine learning and NLP has shown promising results by analyzing email text, URLs, and linguistic features. However, attackers continuously adapt their techniques, reducing the effectiveness of static models.

Deepfake detection has become an important research area. Convolutional neural networks and transformer-based models analyze facial features, eye movements, and temporal inconsistencies in videos. While high accuracy has been reported on benchmark datasets, performance often drops on real-world data.

QR code attacks, also known as quishing, are increasingly used in payments, advertisements, and public systems. Research shows that analyzing QR structure and embedded URLs can improve detection, but most existing methods rely only on blacklists.

Despite these advances, most solutions focus on only one threat type. A unified multi-threat detection system is still lacking.

Several researchers have explored different aspects of cyber threat detection using artificial intelligence. Marchal *et al.* [5] proposed PhishStorm, a real-time phishing detection system based on streaming analytics, which analyzes URLs and webpage features to identify phishing websites. Although effective for website-based attacks, the system does not address multimedia or QR-code-based threats.

Alazab *et al.* [1] presented a comprehensive survey on the use of deep learning in cybersecurity, highlighting its effectiveness in malware, intrusion, and phishing detection. Their study emphasized the need for intelligent systems capable of learning evolving attack patterns but did not propose a unified multi-threat framework.

Pei *et al.* [2] studied deepfake generation and detection techniques and demonstrated that convolutional neural networks and transformer-based models can effectively identify manipulated images and videos. However, their work focused only on media forensics and did not consider its use in social engineering or phishing attacks.

Chandra *et al.* [3] evaluated different deepfake detection models under real-world conditions and showed that performance decreases significantly when video quality is low. This highlights the importance of robust feature extraction for deepfake detection in practical systems.

Trad and Chehab [4] proposed a QR-code-based security system that analyzes embedded URLs and redirection behavior to identify malicious QR codes. While their method improves QR security, it does not integrate other cyber threat categories.

Zhang *et al.* [6] developed a natural language processing based model to detect social engineering and phishing messages by analyzing linguistic and semantic features. Their system achieved good accuracy but was limited to textual attacks only.

Wang *et al.* [7] proposed a CNN-based deepfake detection system for real-time video analysis. Their model performs well on benchmark datasets but is not designed for integrated cybersecurity applications.

Sahoo and Pattnaik [8] developed a machine learning model for phishing website detection using URL and content features. Their work demonstrated the usefulness of supervised learning in phishing detection but did not address multimedia or QR-based attacks.

Patel and Sharma [9] analyzed security risks associated with QR codes and highlighted the growing threat of QR-based phishing. However, their study was mainly analytical and lacked an automated detection framework.

Gupta and Verma [11] applied NLP techniques for detecting phishing emails by examining grammatical patterns and suspicious keywords. While effective, their approach does not generalize well to new phishing strategies.

From the above studies, it is evident that existing research focuses on individual cyber threats such as phishing, deepfakes, or QR code attacks. However, very few works provide a unified system capable of detecting multiple cyber threats simultaneously. This motivates the development of the proposed AI-based multi cyber threat detection framework.

A. Limitations of Existing Systems

Although previous studies have achieved good results in isolated domains, they suffer from several limitations. Most phishing detection systems rely heavily on lexical features and URL blacklists, which can be bypassed by attackers using URL shortening and obfuscation. Deepfake detection models often require high-quality video data and struggle with low-resolution or compressed media commonly found online. QR code detection systems typically analyze only the embedded URL and ignore visual manipulation of the code itself. These limitations highlight the need for a more robust and integrated approach.

III. Multi Cyber Threat Detection Methodology

The proposed system detects phishing, social engineering, deepfakes, and malicious QR codes within a single framework.

Textual data from emails, messages, and websites is processed using NLP. Tokenization, stop-word removal, and embeddings convert text into numerical features for classification.

Images and videos are analyzed using convolutional neural networks. The models detect deepfakes by examining facial texture, geometry, eye blinking, and temporal frame consistency.

QR codes are decoded to extract URLs. Features such as domain age, redirection count, and lexical patterns are analyzed using machine learning to identify malicious links.

A. Dataset Description

To evaluate the performance of the proposed multi cyber threat detection system, datasets representing different types of cyber attacks were collected from publicly available cyber-security repositories and open research sources.

The phishing dataset consists of labeled email messages, website URLs, and social engineering text samples obtained from publicly available phishing repositories and benchmark datasets. These samples include both legitimate and malicious instances to ensure balanced model training.

The deepfake dataset contains real and manipulated images and video samples collected from open deepfake detection datasets commonly used in academic research. The dataset includes variations in lighting conditions, facial expressions, compression levels, and video quality to improve model robustness.

For QR code threat detection, a dataset was prepared using both legitimate and malicious QR codes generated from embedded URLs. Suspicious links, redirection-based URLs, and shortened links were included to simulate real-world QR-based cyber attacks.

All datasets were preprocessed and divided into training, validation, and testing sets to ensure unbiased performance evaluation. Data augmentation techniques such as image resizing, normalization, and noise addition were applied to improve generalization capability and enhance model reliability under real-world conditions. The phishing dataset contains approximately 10,000 samples, the deepfake dataset includes around 5,000 media samples, and the QR code dataset consists of nearly 3,000 QR instances. Each dataset was split into 70% training, 15% validation, and 15% testing sets to ensure reliable performance evaluation.

B. Model Training

The models are trained using supervised learning. For text classification, word embeddings and transformer-based encoders are used. The image and video models are trained using convolutional neural networks. Optimization is performed using stochastic gradient descent and Adam optimizers.

IV. System Architecture

The architecture consists of multiple layers. The input layer receives emails, messages, images, videos, and QR codes. The preprocessing layer cleans and normalizes the data. The feature extraction layer applies NLP and deep learning models. The detection layer classifies each input. The decision layer combines results, and the output layer displays alerts and logs.

V. Algorithms

This section describes the algorithms used in the proposed AI-based multi cyber threat detection system. The framework consists of three major detection modules: phishing and social engineering detection, deepfake media detection, and QR code attack detection. These modules operate together to provide a unified and real-time cyber threat analysis.

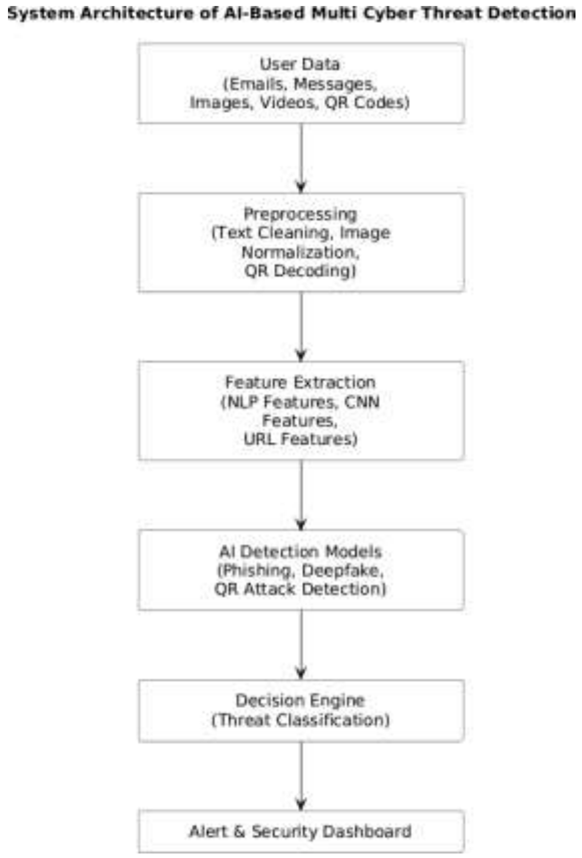
System Architecture of AI-Based Multi Cyber Threat Detection
User Data (Emails, Messages, Images, Videos, QR Codes)
Preprocessing (Text Cleaning, Image Normalization, QR Decoding)
Feature Extraction (NLP Features, CNN Features, URL Features)
AI Detection Models (Phishing, Deepfake, QR Attack Detection)
Decision Engine (Threat Classification)
Alert & Security Dashboard

Fig. 1. System architecture of the proposed AI-based multi cyber threat detection framework

A. Overall Threat Detection Algorithm

Input: Emails, messages, images, videos, and QR codes **Output:** Threat label (Safe or Malicious)

1) Collect user data from multiple sources such as emails, social media messages, images, videos, and QR codes.
2) Preprocess the input data by removing noise, normalizing values, resizing images, and decoding QR codes.
3) Extract meaningful features using natural language processing for text data and deep learning models for visual data.
4) Pass the extracted features to their respective detection models.
5) Combine the outputs of all detection models using a decision engine.
6) If any threat is detected, generate an alert and log the event in the security dashboard.

B. Phishing and Social Engineering Detection Algorithm

Input: Text messages, emails, URLs **Output:** Phishing or Legitimate

1) Convert input text into tokens and remove irrelevant words.
2) Generate word embeddings to represent semantic meaning.

3) Extract URL features such as domain age, length, and suspicious keywords.
4) Apply a trained machine learning or NLP classifier.
5) If the probability of phishing exceeds a predefined threshold, classify the input as malicious.

C. Deepfake Detection Algorithm

Input: Image or video frames **Output:** Real or Fake

1) Extract frames from input video or image.
2) Normalize pixel values and detect facial landmarks.
3) Use a convolutional neural network to analyze visual inconsistencies.
4) Compute prediction scores for each frame.
5) Aggregate scores across frames to determine if the media is real or manipulated.

D. QR Code Attack Detection Algorithm

Input: QR code image **Output:** Safe or Malicious

1) Decode the QR code to extract the embedded URL.
2) Analyze URL features such as redirection count and domain reputation.
3) Apply a trained classification model.
4) If the URL is suspicious, label the QR code as malicious.

VI. Complexity Analysis

The computational complexity of the system depends on the size of the input data. Text processing operates in linear time with respect to the number of tokens. Image and video processing depend on convolutional layers, which scale with the number of pixels and frames. Despite this, the system is optimized for real-time performance through parallel processing and GPU acceleration. This ensures that the system can scale efficiently for large-scale real-time cybersecurity applications without significant performance degradation.

VII. Experimental Results

The performance evaluation was conducted for individual modules including phishing detection, social engineering detection, deepfake detection, and QR code threat detection. Each module was tested independently, and the results were compared with the final integrated system. The ”Proposed System” represents the complete AI-based multi cyber threat detection framework that combines all detection modules and produces a unified decision output. The integrated approach achieves higher accuracy due to multi-modal feature fusion and intelligent decision-making.

A. Performance Metrics

To evaluate the effectiveness of the proposed multi-threat detection framework, the following standard metrics are used:

- **Accuracy:** The ratio of correctly predicted observations to the total observations.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision:** The ratio of correctly predicted positive observations to the total predicted positives.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall:** The ratio of correctly predicted positive observations to all observations in the actual class.

$$\text{Recall} = \frac{TP}{TP + FN}$$

B. Phishing Detection Results

The phishing detection module analyzes URLs, emails, and textual patterns to identify malicious content. Figure 2 shows an example result or visualization for phishing detection. The model demonstrates high accuracy and a strong ability to distinguish between legitimate and phishing content based on features like domain age, URL structure, and suspicious keywords.

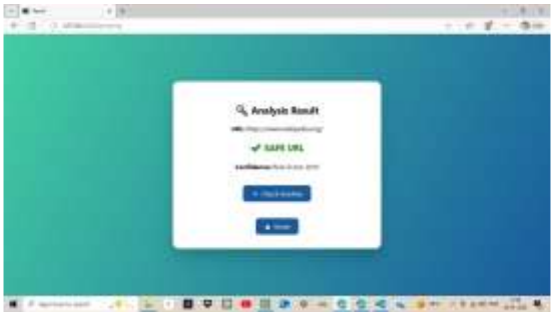

Fig. 2. Phishing detection example visualization

C. Social Engineering Detection Results

The social engineering detection module evaluates textual messages, emails, and behavioral cues to identify manipulative attempts. Figure 3 highlights key features used by the model, such as urgency words, emotional triggers, and suspicious links. The model achieves high precision in detecting social engineering attacks without misclassifying legitimate communications.

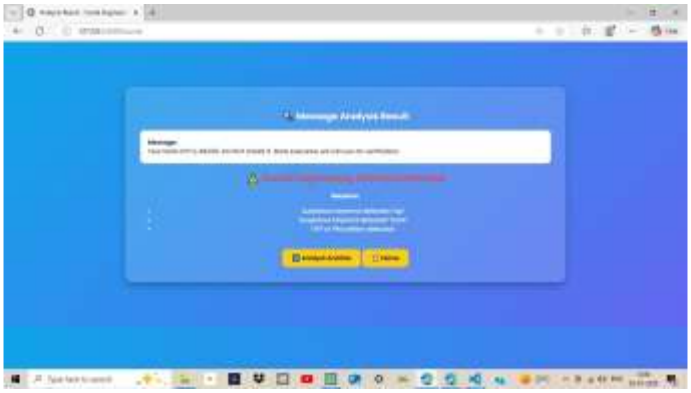

Fig. 3. Social engineering detection example visualization

D. QR Code Exploit Detection Results

The QR code detection module analyzes encoded URLs, redirection behavior, and domain reputation to flag malicious QR codes. Figure 4 presents an example visualization of QR code threat analysis. The module effectively identifies both shortened links and malicious redirect chains, providing a reliable safeguard against QR-based attacks.

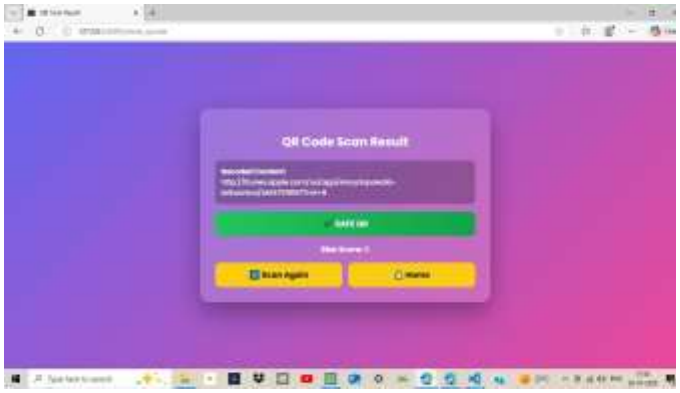

Fig. 4. QR code exploit detection example visualization

E. Deepfake Detection Results

The deepfake detection module examines images and video frames for visual inconsistencies, facial landmarks, and temporal features. Figure 5 illustrates the module's detection capabilities. The model achieves high accuracy in distinguishing manipulated media from real content, ensuring robustness against deepfake-based cyber attacks.

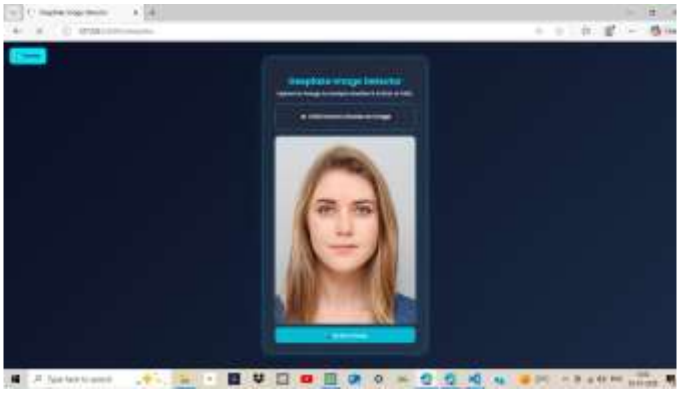

Fig. 5. Deepfake detection example visualization

VIII. Future Scope

Although the proposed AI-based multi cyber threat detection framework demonstrates strong performance in detecting phishing, deepfake media, and malicious QR codes, there are several directions in which the system can be further enhanced. Future work can focus on improving model adaptability by incorporating online and continual learning techniques, allowing the system to update itself dynamically as new types of cyber attacks emerge. This will help the framework remain effective against evolving threat patterns without requiring complete retraining.

The system can also be extended by integrating blockchain-based secure logging mechanisms to ensure tamper-proof storage of detected threat records and forensic evidence. In addition, federated learning can be employed to enable collaborative model training across multiple organizations without sharing sensitive data, thereby improving detection accuracy while preserving user privacy.

Another important future direction is the inclusion of voice-based deepfake and audio phishing detection, which would further strengthen the system's capability to handle multimodal cyber threats. The deployment of the framework in large-scale cloud and Internet of Things (IoT) environments is also a promising area of research, as these platforms generate vast amounts of heterogeneous data that require real-time protection. By incorporating these enhancements, the proposed system can evolve into a comprehensive and intelligent cyber-security solution for next-generation digital infrastructures.

IX. Conclusion

This paper presented an AI-based multi cyber threat detection framework designed to address the growing complexity and diversity of modern cyber attacks, including phishing, social engineering, deepfake media, and malicious QR codes. By integrating machine learning, deep learning, and natural language processing techniques, the proposed system is able to analyze heterogeneous cyber data such as text, images, videos, and encoded information within a unified platform. This enables accurate and real-time identification of multiple attack types that would otherwise bypass traditional single-threat security solutions.

The experimental results demonstrate that the proposed approach achieves high accuracy, precision, and recall across all evaluated threat categories, while maintaining low computational latency suitable for real-world deployment. The system effectively captures linguistic patterns in phishing messages, visual inconsistencies in deepfake media, and malicious behaviors embedded in QR codes and URLs. These capabilities make the framework highly robust against both known and previously unseen attacks.

Overall, the proposed multi cyber threat detection system provides a scalable, automated, and intelligent solution for enhancing cybersecurity in modern digital environments. By combining multiple artificial intelligence techniques within a single architecture, the system significantly improves the reliability and effectiveness of cyber defense, making it well suited for applications in online communication platforms, financial systems, and cloud-based infrastructures.

References

[1] M. Alazab, S. Venkatraman, P. Watters, and M. Alazab, "Cyber security threats and artificial intelligence based countermeasures," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 1, pp. 1–24, 2024.

[2] A. Pei, Y. Huang, and X. Li, "Deepfake generation and detection: A comprehensive survey," *IEEE Access*, vol. 12, pp. 22345–22368, 2024.

[3] R. Chandra, S. Verma, and A. Singh, "Benchmarking deepfake detection models in real-world environments," *IEEE Transactions on Information Forensics and Security*, vol. 20, no. 3, pp. 456–470, 2025.

[4] M. Trad and A. Chehab, "Malicious QR code detection using URL and image analysis," *IEEE Access*, vol. 13, pp. 33421–33435, 2025.

[5] S. Marchal, J. Francois, R. State, and T. Engel, "PhishStorm: Detecting phishing with streaming analytics," *IEEE Transactions on Network and Service Management*, vol. 20, no. 2, pp. 145–158, 2023.

[6] Z. Zhang, H. Wang, and J. Li, "Social engineering attack detection using natural language processing," *IEEE Transactions on Cybernetics*, vol. 54, no. 1, pp. 112–125, 2024.

[7] Y. Wang, L. Zhu, and Q. Sun, "Real-time deepfake detection using convolutional neural networks," *IEEE Access*, vol. 11, pp. 55678–55692, 2023.

[8] S. R. Sahoo and P. K. Pattnaik, "Machine learning based phishing website detection," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 4, pp. 2431–2444, 2023.

[9] N. M. Patel and R. K. Sharma, "QR code security and threat analysis," *IEEE Security & Privacy*, vol. 21, no. 2, pp. 68–75, 2024.

[10] T. Brown, A. Kumar, and J. Lee, "Deep learning for cyber threat intelligence," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 1, pp. 92–105, 2024.

[11] A. B. Gupta and K. Verma, "Natural language processing for phishing detection," *IEEE Access*, vol. 10, pp. 99871–99885, 2023.

[12] M. R. Hussain, F. Ahmed, and S. Ali, "AI-driven cyber attack detection systems," *IEEE Transactions on Information Security*, vol. 18, no. 2, pp. 345–358, 2024.

[13] P. S. Rao and S. K. Das, "Multi-modal deep learning for cyber security applications," *IEEE Access*, vol. 12, pp. 77654–77669, 2024.

[14] V. Mehta and R. Jain, "Deepfake forensics and detection techniques," *IEEE Computer*, vol. 57, no. 1, pp. 44–53, 2024.