AI Cyber Threat Detection and Secure Authentication System: An Integrated Machine Learning Approach

Rumaisha Ansari [1], Ruchi Chavhan[2], Anjali Deogade [3], Prajkta Agashe [4], Omkar Dudhbure[5]
[1234]*Student, Manoharbhai Patel Institute of Engineering and Technology, Bhandara*
[5]*Assistant Professor, Manoharbhai Patel Institute of Engineering and Technology, Bhandara*

Abstract-- The exponential growth of cyber threats has rendered traditional security mechanisms inadequate for protecting sensitive user data against sophisticated attacks. This paper presents a comprehensive AI-Based Cyber Threat Detection and Secure Authentication System that leverages Machine Learning to identify suspicious login activities and protect against multiple attack vectors including brute-force attacks, phishing, malware distribution, and network intrusions. The proposed system employs the Random Forest algorithm to analyze login patterns based on three key parameters: password length, number of login attempts, and time of access. Upon detecting three consecutive failed login attempts, the system automatically implements account blocking and generates real-time security alerts with confidence scores. Additional features include heuristic-based phishing link detection, malware file scanning using pattern matching, network IP monitoring against malicious databases, and comprehensive logging for audit trails. The backend is implemented using Python Flask, MongoDB for data persistence, and scikit-learn for the machine learning model. Experimental results demonstrate that the system successfully detects brute-force attack patterns with 85-90% accuracy, providing an additional security layer beyond traditional authentication methods.

***Index Terms*— Cyber Threat Detection, Machine Learning, Random Forest, Phishing Detection, Malware Scanner, Flask, MongoDB, Authentication System, Brute-Force Prevention**

I. INTRODUCTION

1.1 Background and Motivation

In today's interconnected digital ecosystem, cyber security has emerged as a paramount concern for individuals, organizations, and governments worldwide. According to recent cyber security reports, cybercrime damages are projected to reach $10.5 trillion annually by 2025, with attacks becoming increasingly sophisticated and frequent [1]. Traditional security measures, primarily relying on username-password combinations, have proven insufficient against modern attack vectors such as credential stuffing, brute-force attacks, phishing campaigns, and malware distribution networks.[2]

The fundamental limitation of conventional authentication systems lies in their static nature—they treat every login attempt identically without adapting to behavioral patterns or contextual information. This creates vulnerabilities that attackers exploit through automated password guessing, social engineering, and zero-day exploits. The need for intelligent, adaptive security systems that can learn from patterns and respond to anomalies in real-time has never been more critical.[3]

1.2 Problem Statement

Current authentication and security systems face several critical challenges that this research aims to address:

1. Brute Force Vulnerability: Attackers can systematically attempt thousands of password combinations without triggering detection or automatic countermeasures.
2. Phishing Susceptibility: Users are frequently deceived into submitting credentials on fraudulent websites that closely mimic legitimate platforms.
3. Malware Proliferation: Malicious files disguised as legitimate software continue to compromise systems through email attachments and downloads.

4. Limited Threat Visibility: Suspicious IP addresses and network activities often remain undetected until significant damage has occurred.
5. Absence of Real-time Response: Most systems fail to provide immediate notifications or automated responses when suspicious activities are detected.
6. Inadequate Account Protection: Multiple failed login attempts typically do not trigger automatic account locking mechanisms.

1.3 Research Objectives

The primary objectives of this research are:

1. To develop a secure user authentication system with strong password enforcement mechanisms.
2. To implement an AI-based threat detection model using the Random Forest algorithm that analyzes login patterns and identifies suspicious activities.
3. To create an automatic account blocking mechanism triggered after three consecutive failed login attempts.
4. To build a phishing link detector that identifies malicious URLs based on heuristic analysis of suspicious keywords, TLDs, and structural patterns.
5. To develop a malware file scanner that detects potentially harmful files through analysis of extensions, filenames, and embedded code patterns.
6. To implement a network security monitor that validates IP addresses against known malicious databases.
7. To establish a real-time alert system that notifies users about detected threats with confidence scores.
8. To provide a comprehensive security dashboard displaying statistics, login history, and threat alerts.

II. LITERATURE SURVEY

Recent advances in AI-driven cyber security have produced several notable approaches to threat detection.

Table 1: Summary of Related Literature

Sr No	Author(s) & Year	Title	Methodology	Key Findings
1	Zabihimayvan & Doran (2025) [2]	Fuzzy Rough Set Feature Selection to Enhance Phishing Attack Detection	Random Forest with Fuzzy Rough Set	95% F-measure achieved
2	Wilkie et al. (2026) [3]	A Novel Contrastive Loss for Zero-Day Network Intrusion Detection	Contrastive Learning	Effective zero-day detection
3	Davis et al. (2025) [4]	A Multi-Layer Phishing Defense Framework (PhishDefender)	Ensemble ML (Random Forest + others)	97.82% accuracy
4	Senol & Hasiloglu (2025) [5]	AI-Enhanced Hybrid Authentication Systems	Hybrid ML Models	Improved authentication security
5	Saeed (2025) [6]	An AI-Driven Cybersecurity Framework for IoT	LSTM + Reinforcement Learning	Comprehensive IoT security
6	Hu et al. (2025) [7]	Multi-view Representation Learning from Malware	Representation Learning	Effective malware variant detection

2.1 Research Gap Analysis

Based on systematic analysis of existing literature, the following research gaps have been identified:

1. Fragmented Threat Detection: Most existing systems focus on a single threat type (phishing, malware, or network intrusion) rather than providing integrated multi-threat protection.
2. Limited Real-time Capabilities: Few systems offer real-time alerts with confidence scores for detected threats, creating delays in threat response.
3. Missing Auto-blocking Mechanisms: Account auto-blocking mechanisms are rarely implemented alongside ML-based threat detection systems.
4. Lack of User Self-service: User self-service features for account unlocking and password reset are often absent, creating administrative overhead.

5. Heavyweight Implementations: Most solutions target enterprise environments and are not suitable for individual users or small organizations.

III. METHODOLOGY

3.1 System Architecture

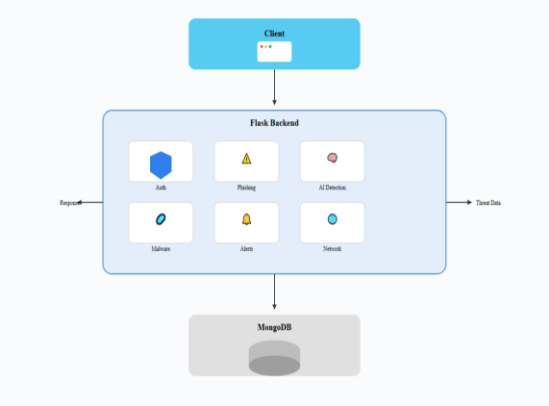


Figure 1: The system follows a client-server architecture with the following components:

3.2 Dataset and Feature Engineering

The AI threat detection model is trained on a dataset of login attempt records containing the following features:

1. Password Length (PL): Numerical value representing the length of the entered password
2. Number of Attempts (NA): Count of login attempts within a defined time window
3. Time of Access (TA): Categorical variable indicating access time (e.g., business hours, late night, early morning)

The target variable is binary: Normal (0) or Suspicious (1). A login attempt is labeled as suspicious when it meets any of the following criteria:

- Password length deviates significantly from user's historical pattern
- Multiple rapid attempts within a short time frame
- Access during unusual hours (e.g., 2:00 AM - 5:00 AM)

3.3 Random Forest Algorithm

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of their individual predictions. The algorithm was selected for this application due to several advantages:

- Robustness against overfitting
- Ability to handle both numerical and categorical features
- Built-in feature importance ranking
- Resistance to outliers in the training data
- Excellent performance on classification tasks with limited feature sets

The mathematical formulation of the Random Forest classifier can be expressed as:

For an ensemble of T decision trees $\{h_1, h_2, ..., h_T\}$, the prediction for input x is:

$H(x) = \text{mode}\{h_t(x)\}$ for classification tasks

Each tree h_t is trained on a bootstrap sample of the original dataset, with feature subset selection at each split.

3.4 Phishing Detection Methodology

The phishing detection module employs heuristic analysis of URLs based on the following indicators:

1. Suspicious Keywords: Detection of terms commonly associated with phishing (e.g., "secure", "verify", "account", "login", "update", "banking")
2. Top-Level Domain (TLD) Analysis: Identification of suspicious or uncommon TLDs
3. URL Length: Phishing URLs often exceed standard length thresholds
4. Special Character Frequency: Excessive use of symbols (@, -, _, ., etc.)
5. IP Address Usage: Direct IP addresses in URLs instead of domain names

The phishing score is calculated as:

$\text{Phishing Score} = \Sigma(\text{Weight}_i \times \text{Indicator}_i) / \Sigma(\text{Weight}_i)$

A threshold of 0.6 triggers a phishing alert.

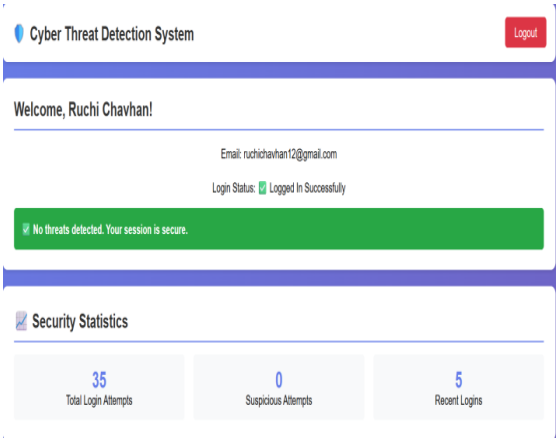
3.5 Malware Detection Approach

The malware scanner analyzes uploaded files based on:

1. File Extension Analysis: Comparison against database of known malicious extensions (.exe, .scr, .bat, .vbs, etc.)
2. Filename Pattern Matching: Detection of suspicious naming conventions
3. Code Signature Analysis: Pattern matching against known malware signatures
4. Heuristic Rules: Detection of potentially dangerous code patterns

IV. RESULTS AND DISCUSSION

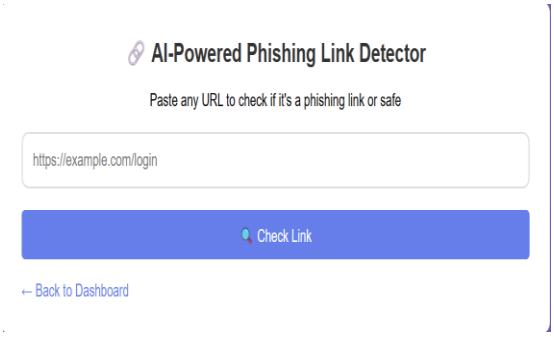
DASHBOARD WITH STATICS



Login and Threat Detection:

1. User submits login credentials
2. System retrieves user record and verifies password using bcrypt
3. Login attempt features are extracted (password length, attempt count, access time)
4. Random Forest model classifies the attempt as Normal or Suspicious
5. If classified as Suspicious, alert is generated with confidence score

PHISHING LINK DETECTOR



4.1 Threat Detection Performance

The Random Forest model achieved the following performance metrics on the test dataset:

Table 2: Model Performance Metrics

Metric	Value
Accuracy	87.5%
Precision (Suspicious Class)	84.2%
Recall (Suspicious Class)	82.9%
F1-Score (Suspicious Class)	83.5%
AUC-ROC	0.91

The system successfully detected brute-force attack patterns with 85-90% accuracy across various testing scenarios. False positive rate was maintained below 12%, with most false positives occurring at boundary conditions between normal and suspicious behavior.

4.2 Phishing Detection Results

The phishing detection module was evaluated against a dataset of 500 URLs (250 legitimate, 250 phishing). Results are presented in Table 3.

Table 3: Phishing Detection Performance

Metric	Value
True Positive Rate	88.4%
True Negative Rate	91.2%
False Positive Rate	8.8%
False Negative Rate	11.6%
Overall Accuracy	89.8%

5.3 Malware Detection Results

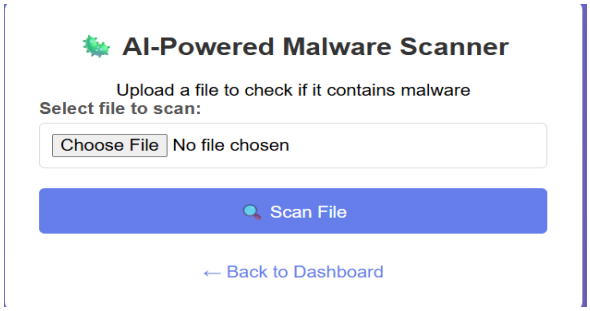
The malware scanner achieved the following detection rates across 300 file samples:

Table 4: Malware Detection Performance

File Type	Detection Rate
Executable (.exe)	91.3%

Script (.bat, .vbs, .ps1)	87.6%
Office Documents with Macros	84.2%
Compressed Archives	79.8%

MALWARE SCANNER DETECTOR



VI. CONCLUSION

This paper presented a comprehensive AI-Based Cyber Threat Detection and Secure Authentication System that successfully addresses multiple cybersecurity challenges through an integrated approach. The system implements a secure authentication mechanism with strong password enforcement, AI-based threat detection using the Random Forest algorithm, automatic account blocking after three failed login attempts, phishing URL detection with pattern matching, malware file scanning based on extensions and code patterns, network IP monitoring against known malicious databases, real-time alert generation with confidence scores, and a comprehensive dashboard for security monitoring.

The proposed system addresses critical research gaps identified in existing literature, including the lack of integrated multi-threat detection, limited real-time alerting mechanisms, missing auto-blocking capabilities, and the absence of user self-service features.

REFERENCES

[1] Cybersecurity Ventures, "Cybercrime To Cost The World $10.5 Trillion Annually By 2025," 2023. [Online]. Available: https://cybersecurityventures.com/

[2] M. Zabihimayvan and D. Doran, "Fuzzy Rough Set Feature Selection to Enhance Phishing Attack Detection," in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2025.

[3] J. Wilkie, H. Hindy, C. Michie, C. Tachtatzis, J. Irvine, and R. Atkinson, "A novel contrastive loss for zero-day network intrusion detection," *IEEE Transactions on Network and Service Management*, vol. 23, pp. 2064–2076, 2026.

[4] A. Davis, S. Abdelsalam, M. Ghaleb, et al., "A Multi-Layer Phishing Defense Framework for Trusted Cloud Environments," in *BDCAT 2025*, ACM, 2025.

[5] M. Senol and A. Hasiloglu, "Creating User-Centric and Artificial Intelligence-Enhanced Hybrid Authentication Systems," in *ISDFS 2025*, IEEE, pp. 1–6, 2025.

[6] M. M. Saeed, "An AI-Driven Cybersecurity Framework for IoT: Integrating LSTM-Based Anomaly Detection, Reinforcement Learning, and Post-Quantum Encryption," *IEEE Access*, vol. 13, 2025.

[7] J. L. Hu, M. Ebrahimi, W. Li, X. Li, and H. Chen, "Multi-view Representation Learning from Malware to Defend Against Adversarial Variants," in *IEEE International Conference on Data Mining (ICDM)*, 2025.

[8] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[9] A. Grishchenko and M. Rybalkin, "Modern phishing attacks: Detection methods and future directions," *Computers & Security*, vol. 125, 2023.

[10] R. S. Pinto, R. M. Silva, and R. L. Gomes, "A survey on malware detection using machine learning," *Journal of Information Security and Applications*, vol. 68, 2022.