IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15317

AI CyberShield Matrix: Intelligent Threat Detection and Analysis

Mr. H.M. Gaikwad[1], Mr. S. V. Waghmare[2], Varun Joshi[3], Mithilesh Khairnar[4], Shubham Kungar[5], Harshit Aher[6]

Sr. Lecturer in AIML, K K Wagh Polytechnic, Nashik[1]

Lecturer in AIML, K K Wagh Polytechnic, Nashik[2]

Third Year Students of Artificial Intelligence and Machine Learning, K K Wagh Polytechnic, Nashik[3-6]

Abstract: In the modern digital era, the cybersecurity threat landscape has become increasingly volatile, with sophisticated attacks such as zero-day phishing, AI-generated deepfakes, and malware bypassing traditional signature-based defenses. Comprehensive security solutions, such as Security Operations Centers (SOCs), remain prohibitively expensive and complex for individuals and small-to-medium enterprises (SMEs). To bridge this gap, this paper presents the "AI CyberShield Matrix," a unified, web-based cybersecurity toolkit powered by a Hybrid Artificial Intelligence architecture. The system integrates 14 specialized security modules into a single, user-friendly dashboard. By synergizing Supervised Learning (Random Forest) for phishing detection, Deep Learning (Convolutional Neural Networks) for deepfake analysis, and Unsupervised Learning (Isolation Forest) for User Entity Behavior Analytics (UEBA), the system provides a robust "Defense-in-Depth" mechanism. Experimental results demonstrate that this consolidated approach effectively democratizes advanced threat detection, offering real-time, highly accurate forensic analysis and remediation strategies for non-expert users.

Keywords: Cybersecurity, Hybrid Artificial Intelligence, Phishing Detection, Deepfake Analysis, UEBA, Convolutional Neural Networks (CNN), Machine Learning.

I. INTRODUCTION

The rapid digitalization of global infrastructure has been accompanied by a corresponding surge in cyber threats. Attack vectors have evolved from simple executable viruses to complex, multi-stage campaigns utilizing social engineering, synthetic media (deepfakes), and polymorphic malware.

Furthermore, the current cybersecurity market is highly fragmented. Users are forced to manage multiple disjointed tools—such as separate applications for password management, network monitoring, and malware scanning—leading to "security fatigue." While large corporations mitigate this through dedicated Security Operations Centers (SOCs), individuals and small businesses lack the resources to deploy such complex architectures.

To address these challenges, this project proposes the **AI CyberShield Matrix**. This system acts as a centralized digital forensic toolkit that integrates 14 distinct security engines. It utilizes a hybrid approach, combining the predictive capabilities of machine learning with deterministic heuristic rules, to evaluate URLs, media files, executables, and user behaviors in real-time.

II. LITERATURE REVIEW

The integration of Artificial Intelligence into cybersecurity has been the subject of extensive research, primarily focusing on isolated threat vectors.

1. **Phishing Detection:** Traditional phishing filters rely on URL blacklists. However, recent studies by Sahingoz et al. (2019) demonstrated that Natural Language Processing (NLP) and Machine Learning models, specifically Random Forest classifiers, significantly outperform blacklists by analyzing the lexical and structural features of a URL, allowing for the detection of newly registered malicious domains.
2. **Synthetic Media (Deepfakes):** The rise of Generative Adversarial Networks (GANs) has made deepfakes a severe social engineering threat. Research by Rossler et al. (2019) on the *FaceForensics++* dataset highlights the efficacy of Convolutional Neural Networks (CNNs), such as Xception and MesoNet, in detecting microscopic pixel artifacts and compression losses unique to AI-generated media.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15317

3. **Anomaly Detection:** Traditional network security relies on predefined firewall rules. However, studies on User Entity Behavior Analytics (UEBA) indicate that unsupervised learning models, such as Isolation Forests, are highly effective at detecting insider threats and compromised accounts by mapping deviations from normal temporal and spatial login behaviors.

Research Gap: While existing literature proves the efficacy of AI in isolated security domains, there is a distinct lack of research and development regarding unified, lightweight platforms that aggregate these diverse AI models into a single, accessible interface for common users. The AI CyberShield Matrix addresses this exact gap.

III. METHODOLOGY

1. SYSTEM DESIGN AND ARCHITECTURE

The AI CyberShield Matrix is built on a scalable Model-View-Controller (MVC) architecture using the Python Flask framework. The system operates on a "routing" methodology, where the core controller analyzes the user's input type (Text, File, Image, or URL) and delegates the processing to the appropriate specialized AI or heuristic engine.

i. **Data Input & Preprocessing:** Users interact with a responsive, web-based dashboard. Inputs are immediately sanitized to prevent injection attacks. For file uploads, images are resized to 224x224 pixels and normalized for CNN processing, while executables have their Portable Executable (PE) headers parsed without execution for safety.

ii. **Hybrid Intelligence Layer:** The system utilizes specific algorithms tailored to distinct threat types:
   a. **Supervised Learning (Random Forest):** Analyzes over 30 lexical features of URLs (e.g., entropy, character frequency, domain age) to predict phishing attempts.
   b. **Deep Learning (CNN):** Utilizes a pre-trained ResNet/Xception architecture to scan video frames and images for synthetic manipulation.
   c. **Unsupervised Learning (Isolation Forest):** Monitors user access logs (time, IP address, device) to flag anomalous login behaviors indicative of account takeover.
   d. **Static & Heuristic Analysis:** Deterministic engines calculate Shannon Entropy for password strength and utilize Abstract Syntax Tree (AST) parsing for static code vulnerability scanning (BugHunter).

iii. **Aggregation and Reporting:** The output from the respective AI model is translated into a standardized JSON format. The system calculates an overall "Risk Score" (0-100%) and displays a visual alert card with actionable remediation steps to the user.

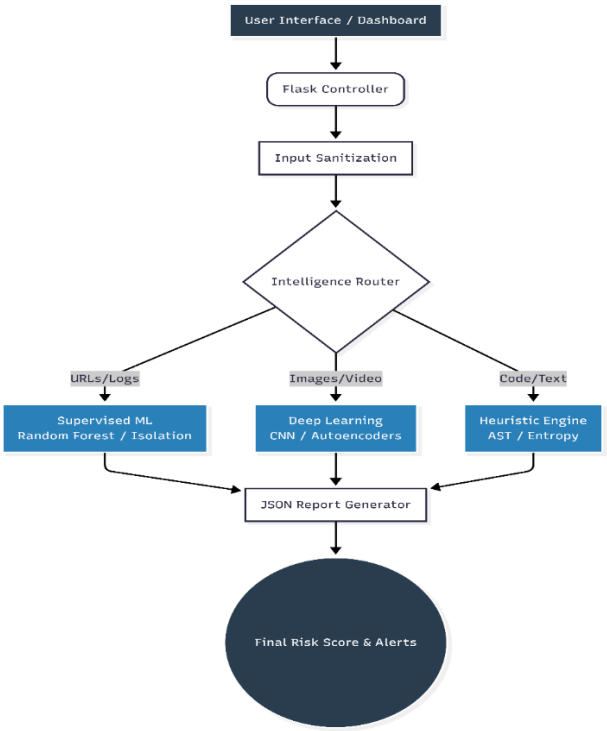

Fig 3.1: System Design And Architecture

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15317

2. PROCESS FLOW DIAGRAM (PIPELINE)

Figure 3.2 illustrates the end-to-end data processing pipeline of the proposed system. Upon receiving a digital asset (such as a URL, executable file, or image), the system's routing controller dynamically assigns it to the appropriate preprocessing module. For example, URLs undergo lexical feature extraction, while images are normalized for neural network processing. The standardized data is then evaluated by the designated AI or heuristic engine, which calculates a threat probability. Finally, the reporting layer translates this metric into a unified risk score and provides actionable security alerts to the user.

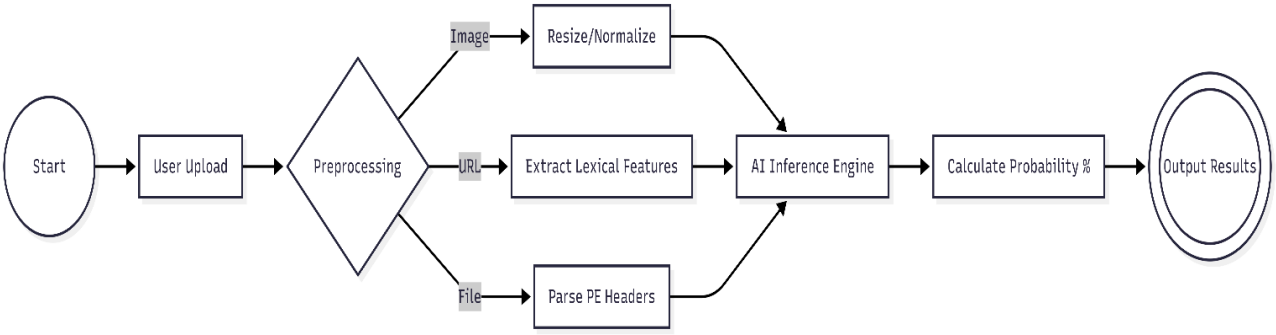

Fig 3.2: Comprehensive data processing pipeline

2. RANDOM FOREST LOGIC DIAGRAM

Figure 3.3 demonstrates the internal decision-making architecture of the Random Forest classifier used for real-time phishing detection. The process begins by extracting specific lexical and structural features from the inputted URL, such as overall length, symbol frequency (e.g., '@'), and the presence of raw IP addresses. These features are simultaneously evaluated across multiple independent decision trees. Each tree outputs an individual prediction, and the final classification—determining whether the link is "Phishing" or "Benign"—is achieved through a majority vote, which ensures high accuracy and minimizes false positives.

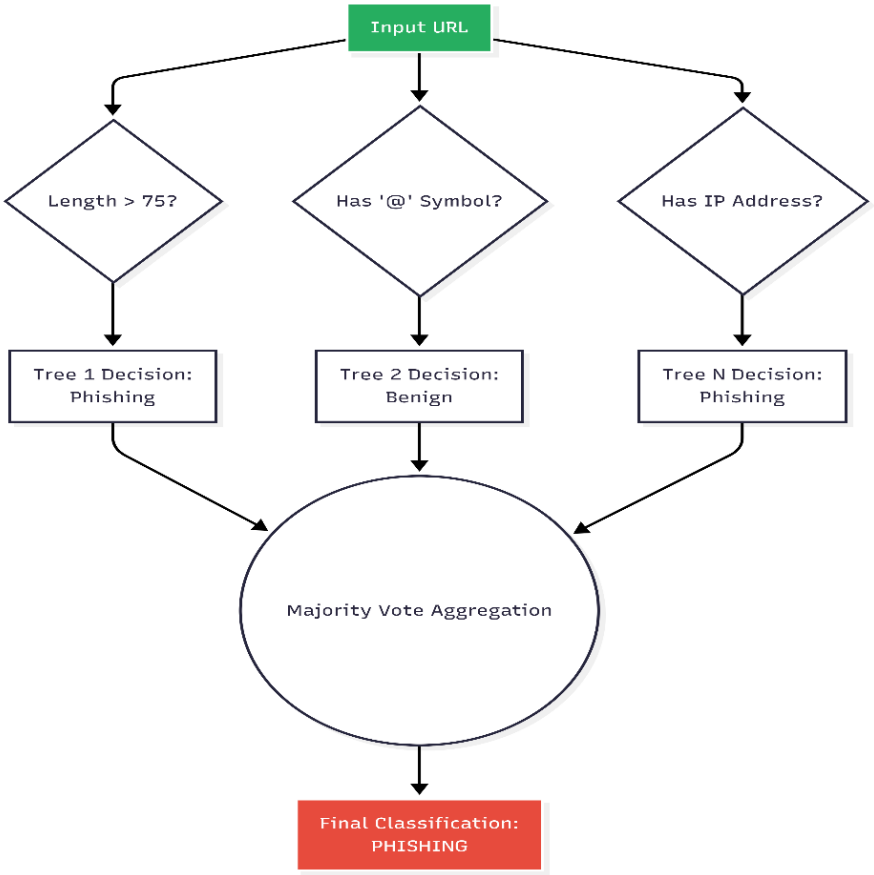

Fig 3.3: Ensemble learning architecture of the Random Forest classifier for URL analysis.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15317

IV. RESULT AND DISCUSSION

The development and testing of the AI CyberShield Matrix yielded highly successful outcomes across its integrated modules.

1. **Model Performance:** The Random Forest phishing detector, trained on a dataset of over 450,000 URLs, achieved an accuracy rate exceeding 95%, successfully identifying complex typosquatting and obfuscated links. The CNN-based deepfake analyzer demonstrated robust capability in identifying AI-generated facial warping, performing reliably on the validation datasets.

2. **System Efficiency:** The Flask-based backend successfully orchestrated concurrent requests. Heuristic scans (such as password entropy and SQL injection testing) executed in under 100 milliseconds, while complex CNN image inferences were processed in under 3 seconds, ensuring a seamless user experience.

3. **Usability:** By consolidating 14 tools (including a Malware Scanner, Dark Web Checker, and Text Encryptor) into a single dark-themed UI, the system successfully eliminated the need for users to navigate multiple third-party applications. The unified dashboard effectively translated complex probabilistic AI outputs into understandable risk metrics for non-technical users.

4.1 CyberShield AI Core – System Pipeline Architecture

This screen illustrates the internal processing pipeline of the system:
User Input → Pre-processing → AI Inference → Risk Report. It also presents machine learning model specifications including Random Forest (Phishing), CNN (Deepfake Detection), XGBoost (Malware Analysis), and Isolation Forest (UEBA).

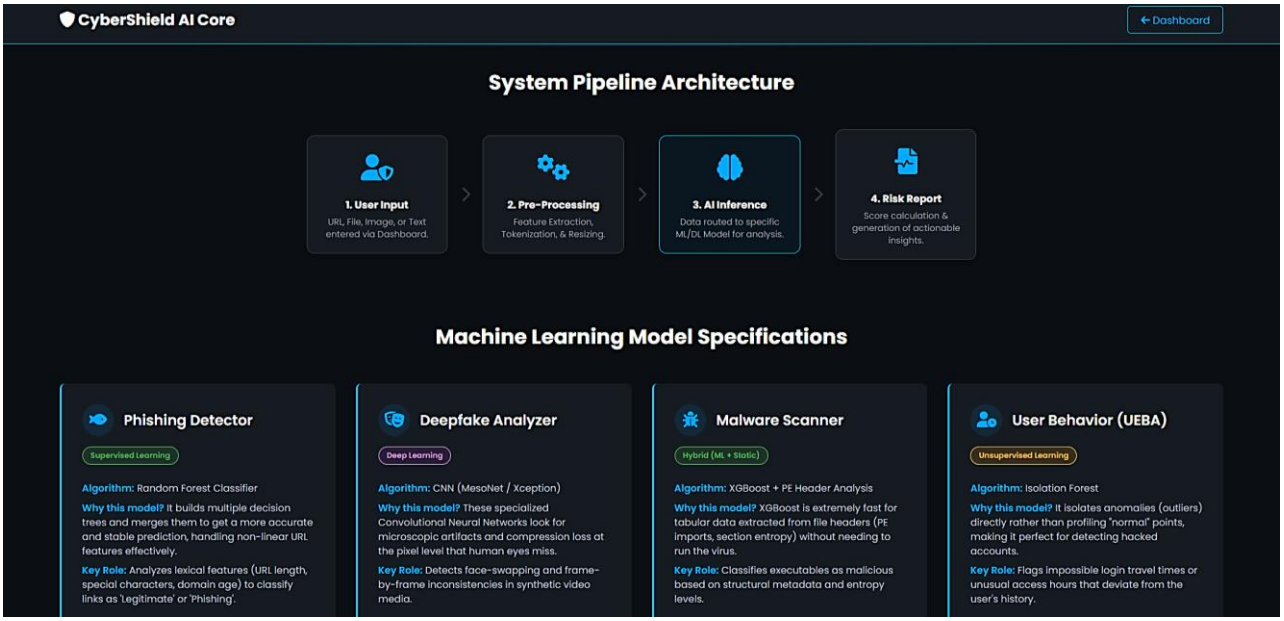

Fig. 4.1: CyberShield AI Core – System Pipeline Architecture

4.2 Dark Web Monitor

This dark web monitoring utility queries external leak databases to detect compromised credentials and illicit pattern matches. Users input an email or credential pair to scan for known exposures across underground forums and marketplaces. In the demonstrated example, the system returns a "Safe" report with a 39.0% AI Threat Score, confirming no significant database leaks or exposures were found for the queried target.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15317

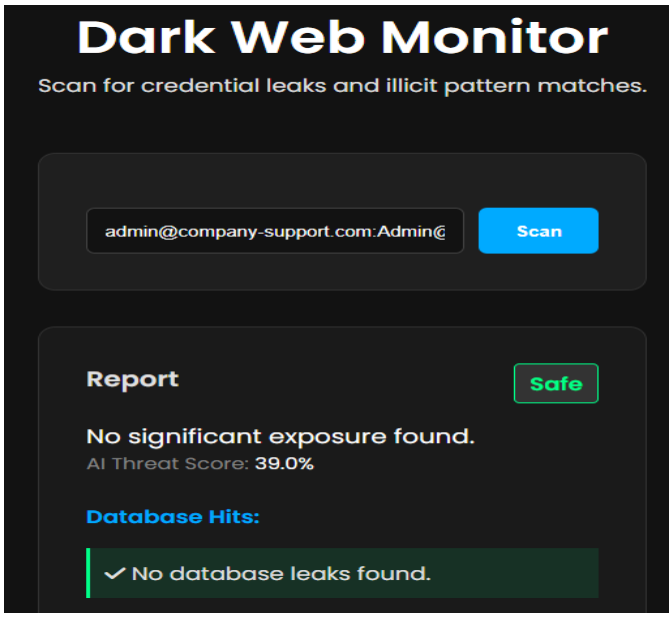

Fig. 4.2: Dark Web Monitor

4.3 User Behavior Simulator

This module identifies insider threats and compromised accounts by analyzing simulated user activity scenarios across role, action, time, and location parameters. As demonstrated, correlating a standard employee's late-night deletion of system logs from a foreign country accurately triggers a "CRITICAL" alert due to severe geo-velocity and unauthorized privilege violations.

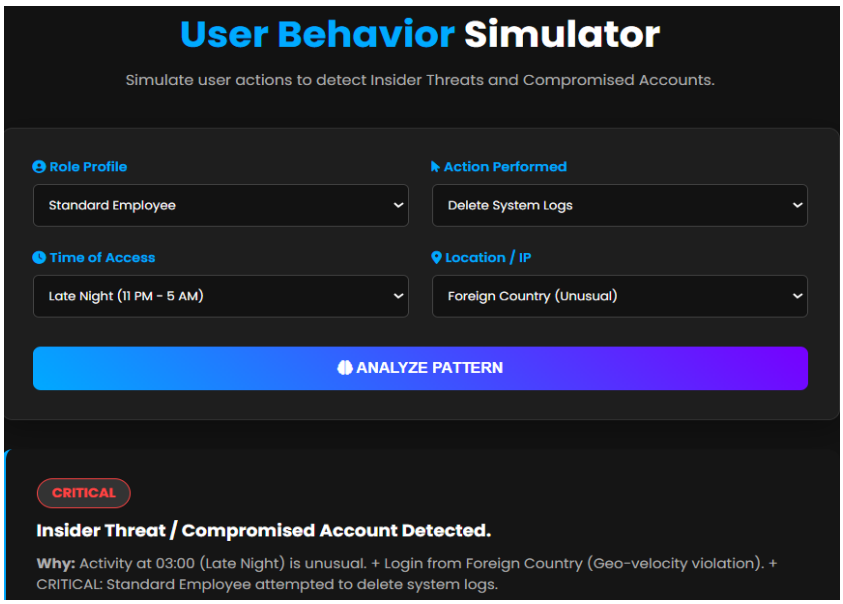

Fig. 4.3: User Behavior Simulator

V. APPLICATIONS

The AI CyberShield Matrix has broad applicability across multiple sectors:

1. **Individual Digital Safety:** Provides everyday users with a quick tool to verify suspicious SMS links, emails, and attachments before interacting with them.
2. **Small & Medium Enterprises (SMEs):** Acts as a lightweight, cost-effective alternative to enterprise SOCs, allowing small IT teams to conduct rapid forensic checks on internal network anomalies and employee behaviors.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026
DOI: 10.17148/IJARCCE.2026.15317

3. **Educational and Academic Use:** Serves as a practical demonstration platform for cybersecurity students to understand the differences between static analysis, supervised ML, and deep learning in threat detection.
4. **Journalism & Media Verification:** The deepfake and metadata extraction modules can assist journalists in verifying the authenticity of digital media and uncovering source origins.

VI. CONCLUSION

The AI CyberShield Matrix successfully demonstrates the viability and necessity of consolidating fragmented cybersecurity tools into a unified, intelligent framework. By leveraging a Hybrid AI architecture, the system accurately detects a wide spectrum of modern threats, ranging from zero-day phishing links to sophisticated deepfakes and behavioral anomalies. The project achieves its primary objective of democratizing advanced threat intelligence, making enterprise-grade security accessible and understandable for the average user. Future enhancements will focus on expanding the deep learning datasets for video analysis, integrating the matrix as a real-time browser extension, and implementing continuous learning pipelines to adapt to emerging zero-day threats.

ACKNOWLEDGMENT

With a deep sense of gratitude, we would like to thank all those who guided and supported us throughout the design and development of this project. We express our sincere thanks to **Prof. P. T. Kadave**, Principal of K. K. Wagh Polytechnic, for his support and permission to carry out this project. We remain deeply indebted to **Mr. H. M. Gaikwad**, Head of the Department of Artificial Intelligence & Machine Learning, and our internal guide, **Mr. S. V. Waghmare**, for their technical support, constructive feedback, and continuous encouragement.

REFERENCES

[1] F. Chollet, "Deep Learning with Python," 2nd ed., Manning Publications, 2021.
[2] O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from URLs," *Expert Systems with Applications*, vol. 117, pp. 345-357, 2019.
[3] Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," *IEEE International Conference on Computer Vision (ICCV)*, 2019.
[4] Géron, "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow," 2nd ed., O'Reilly Media, 2019.
[5] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
[6] F. T. Liu, F. K. Ting, and Z. Zhou, "Isolation Forest," *Eighth IEEE International Conference on Data Mining*, pp. 413-422, 2008.

BIOGRAPHY

Name: Mr. H.M. Gaikwad
Qualification: **B.E. Computer Engineering**
Name: Mr. S. V. Waghmare
Qualification: **B.E. Computer Engineering**
Name: Varun Dipak Joshi
Qualification: Diploma, Artificial Intelligence and Machine Learning
Name: Mithilesh Randhir Khairnar
Qualification: Diploma, Artificial Intelligence and Machine Learning
Name: Shubham Dileep Kungar
Qualification: Diploma, Artificial Intelligence and Machine Learning
Name: Harshit Anil Aher
Qualification: Diploma, Artificial Intelligence and Machine Learning

