AI-Based Predictive Cyber Threat Intelligence System Using Natural Language Processing (NLP)

KUSHAL S M
Dept. of CSE (AI & ML)
ATME College of Engineering
Mysore, India
kushal2003sm@gmail.com

PONNANNA K V
Dept. of CSE (AI & ML)
ATME College of Engineering
Mysore, India
karthamadaponnanna@gmail.com

SANJAN B M
Dept. of CSE (AI & ML)
ATME College of Engineering
Mysore, India
sanjanacharaya1234@gmail.com

VISHNU S
Dept. of CSE (AI & ML)
ATME College of Engineering
Mysore, India
vishnuvishu031@gmail.com

Mrs. Khateeja Ambareen
Assistant Professor
Dept. of CSE (AI & ML)
ATME College of Engineering
Mysore, India
khateejaambareen.ci@atme.edu.in

***Abstract*—In today's rapidly evolving digital landscape, the frequency and sophistication of cyberattacks are increasing at an alarming rate. From large corporations to small businesses, no organization is immune to the risks posed by cyber threats. Traditional cybersecurity defences are struggling to keep pace with these ever-evolving attacks, making it essential to find smarter, more efficient ways to stay protected. This project proposes the development of an Automated Cyber Threat Intelligence (ACTI) system powered by artificial intelligence (AI). The system is designed to detect, analyse, and respond to a wide range of cyber threats, such as malware, phishing, and ransomware attacks, in real time. By leveraging cutting-edge AI techniques such as machine learning and natural language processing (NLP), the ACTI system will analyse data from a variety of sources, including Open-Source Intelligence (OSINT) feeds, to provide actionable insights and proactive defences. Our goal is to create a system that not only automates threat detection but also enhances security teams' ability to respond quickly and effectively, protecting against threats before they can do harm.**

***Index Terms*—Cyber Threat Intelligence (CTI), Artificial Intelligence (AI), Natural Language Processing (NLP), Predictive Analytics, Machine Learning, Anomaly Detection, Artificial Immune System (AIS)**

I. Introduction

In the hyper-connected digital era, cyber threats have become increasingly frequent, dynamic, and difficult to detect. Organizations now face the inevitability of an attack, shifting the paradigm from *if* to *when*. Traditional cybersecurity approaches, often manual and rule-based, are struggling to keep pace with rapidly evolving attack vectors such as phishing, malware, ransomware, and DDoS attacks. These conventional systems are rigid, reactive, and require significant human effort, leading to delayed responses and critical security gaps.

Machine learning (ML) and early Natural Language Processing (NLP) methods have been applied to improve detection, but they often produce high false positives, depend on large labeled datasets, and fail to grasp the contextual nuances of unstructured threat data, especially from informal sources like dark web forums.

To address this growing gap, this paper proposes an AI-powered Predictive Cyber Threat Intelligence (CTI) System. The system intelligently integrates advanced NLP and methodologies inspired by the Artificial Immune System (AIS). It is designed to autonomously collect and analyze threat data from diverse sources, including OSINT feeds, security blogs, and dark web forums.

The primary objectives of this work are:

1) To develop an NLP-based system for the automated extraction of threat intelligence from unstructured cybersecurity text.
2) To predict the severity and category of emerging cyber threats using AI-based classification and anomaly detection techniques.
3) To evaluate the system's performance using comprehensive metrics.

By fusing NLP-based intelligence extraction with AIS-inspired behavioral anomaly detection, the proposed system can identify both known and unknown threats in real-time. It provides real-time alerts, severity predictions, and preventive recommendations, empowering security teams with proactive and adaptive defense capabilities.

The remainder of this paper is organized as follows: Section II reviews related work in AI and CTI. Section III details the system architecture and methodology. Section IV discusses the implementation and experimental setup. Section V presents and analyzes the results. Finally, Section VI concludes the paper.

II. Related Work

The application of AI to CTI is a rapidly growing field. Alturkistani and Chuprat (2024) systematically reviewed 41 studies, categorizing CTI advancements into conventional, AI-based, and LLM-based techniques. They noted that while

AI/ML improves predictive accuracy, challenges in explainability, real-time data, and integration remain.

Trifonov et al. (2018) highlighted the shift from traditional cybercrime to cyber warfare, emphasizing that no single AI method is sufficient. They discussed a successful Multi-Agent System for Tactical CTI, but noted that Operational CTI (predicting adversary behavior) is still in its early stages.

Ismail (2024) provided a critical evaluation of how AI and NLP are transforming cybersecurity. The study emphasized AI's role in detecting complex anomalous patterns and NLP's effectiveness in analyzing textual data for threats like phishing. Together, they enhance automated incident assessment and reduce false positives.

Emeka et al. (2023) focused on *predictive* CTI, which marks a significant shift from reactive to proactive cybersecurity. They detailed how AI (ML, NLP, Deep Learning) enables the forecasting of threats by identifying attacker behavior patterns from aggregated data, leading to early detection and faster response.

Rajaram et al. (2022) explored AI-driven threat detection leveraging big data. They noted that traditional tools are overwhelmed, making AI essential for advanced analysis of malware, intrusions, and insider threats. Patel et al. (2024) further emphasized AI's role in predictive intelligence and automated response, highlighting the reduction in human error and the ability to provide 24/7 monitoring.

While these studies confirm the promise of AI and NLP, they also highlight persistent gaps in adaptability, scalability, and contextual insight, which our proposed hybrid NLP-AIS system aims to address.

III. System Architecture and Methodology

The proposed system employs a multi-stage pipeline, as shown in Fig. 1, to ingest, process, analyze, and act upon cyber threat data.

A. Data Collection

The system aggregates data from diverse, real-time sources:

- **Twitter API:** Integrates with the v2 API to collect social media intelligence using targeted query parameters (e.g., "phishing OR ransomware") and handles rate limiting.
- **Dark Web Intelligence:** A module (currently mocked, pending production) designed to scrape underground forums and ransomware group communications.
- **MITRE ATT&CK Framework:** Integrates with the MITRE database for structured threat intelligence on tactics and techniques.

Data is collected in parallel, preprocessed, and stored in a PostgreSQL database and JSONL files for persistence and analysis.

B. Data Preprocessing

Raw data undergoes a rigorous 6-stage preprocessing pipeline, illustrated in Fig. 2.

Key stages include:

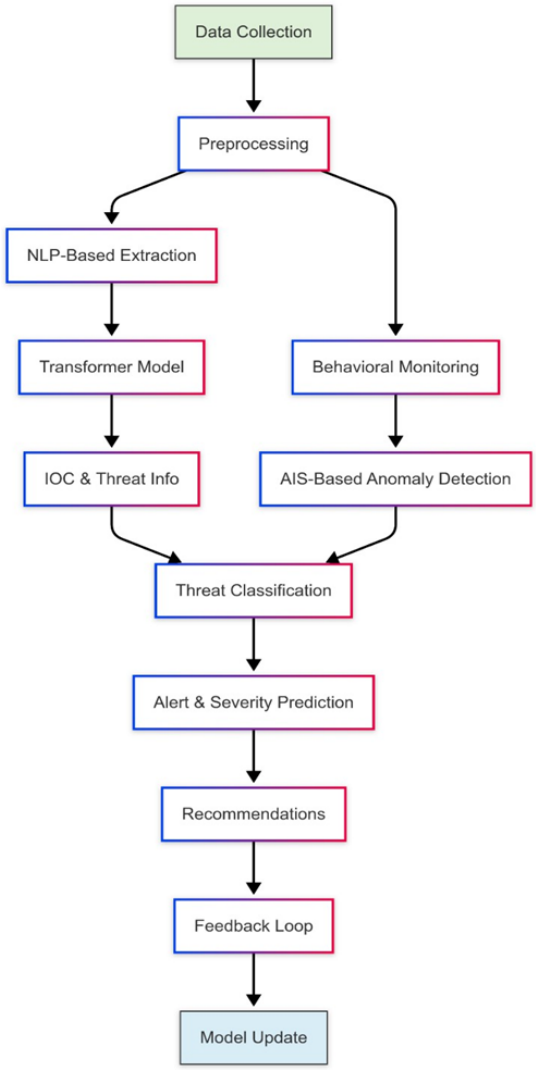

Fig. 1. Flow Chart of The Methodology

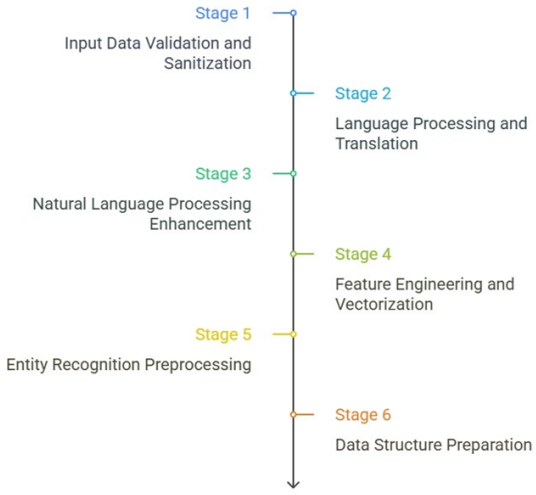

Fig. 2. Data Preprocessing Pipeline Stages

1) **Validation and Sanitization:** Text is extracted, null values are filtered, and content is cleaned of URLs and special characters.
2) **Language Processing:** Non-English text is detected using `langdetect` and automatically translated to English via the Google Translate API for consistent processing.
3) **NLP Enhancement:** Text is processed with `spaCy` for tokenization, lemmatization, and the removal of stop words and punctuation.

C. NLP-Based Threat Intelligence Extraction

This module extracts actionable intelligence from the pre-processed text.

1) Transformer Integration: The system loads transformer-based models like BERT and domain-adapted variants (e.g., ThreatBERT) to understand cybersecurity-specific terminology and narratives.

2) IoC and NER Pipeline: A Named Entity Recognition (NER) pipeline identifies key Indicators of Compromise (IoCs) such as malware families, domains, IPv4 addresses, file hashes (MD5/SHA-1/SHA-256), and CVE identifiers. A secondary regex-based extractor provides a high-speed fallback.

3) TF-IDF Vectorization: To identify threat topics and trends, Term Frequency–Inverse Document Frequency (TF-IDF) is used to transform unstructured text into a weighted feature matrix. The TF-IDF score for a term t in a document d from a corpus D is calculated as:

$$tfidf(t, d, D) = tf(t, d) \cdot idf(t, D) \tag{1}$$

where $tf(t, d)$ is the frequency of term t in document d, and $idf(t, D)$ is the inverse document frequency, calculated as:

$$idf(t, D) = \log \frac{|D|}{|\{d \in D : t \in d\}|} \tag{2}$$

This matrix feeds a Latent Dirichlet Allocation (LDA) model to produce per-document topic distributions.

4) Relation Extraction: A model infers relationships (e.g., *uses*, *targets*, *exploits*) between extracted entities, helping to build a contextual understanding of the threat.

D. Threat Classification and Severity Prediction

1) Feature Fusion: Features from the NLP pipeline (entities, relations, topics) are fused with behavioral features from the AIS-inspired anomaly detection module (monitoring user/network activity). These are concatenated into a single, unified threat vector.

2) Stacked Ensemble Learner: The unified vector is fed into a stacked ensemble classifier for robust prediction. Base models (Random Forest, XGBoost, MLP, RBF-SVM) produce class probabilities. These probabilities are then fed into a Logistic Regression meta-learner, which makes the final, calibrated prediction.

For a K-class problem (e.g., malware, phishing, spam), the meta-learner uses the softmax function to output the final probability P for class c, given the base-model probability vector x' and weight parameters θ:

$$P(y = c|x'; \theta) = \frac{e^{\theta_c^T x'}}{\sum_{j=1}^{K} e^{\theta_j^T x'}} \tag{3}$$

The system classifies threats into categories such as malware, phishing, scam, spam, and legitimate.

3) Severity Prediction Model: A custom, CVSS-like model quantifies risk on a 0-100 scale. The score is calculated based on the base threat class, additive bonuses for risk factors (e.g., +18 for system access, +15 for financial data), model confidence, and multipliers for high-risk combinations. This score is then mapped to four levels: Low (0-25), Medium (26-50), High (51-75), and Critical (76-100).

E. Alert and Recommendation System

Based on the predicted class and severity, the system generates real-time, priority-based alerts. A recommendation engine provides threat-specific actions categorized by urgency: Immediate (0-30 mins, e.g., block IPs), Short-term (1-24 hrs, e.g., deploy patch), and Long-term (1-30 days, e.g., architecture review).

IV. Implementation and Experimental Setup

A. System Requirements

The system was developed and trained using the following hardware and software specifications.

1) Hardware Requirements:

- **Processor:** Intel i7/i9 or AMD Ryzen 7/9
- **RAM:** 16GB (minimum), 32GB+ (recommended)
- **Storage:** 1TB SSD
- **GPU:** NVIDIA RTX 3080/3090 (for deep learning)

2) Software Requirements:

- **OS:** Windows / Linux / macOS
- **Language:** Python 3.x
- **Frameworks:** TensorFlow, PyTorch, Flask
- **Libraries:** NumPy, Pandas, Scikit-learn, Hugging Face Transformers, NLTK, spaCy
- **Database:** PostgreSQL

B. Frontend Dashboard

A modern, responsive web interface was built using the Flask web framework. The dashboard provides an interactive form for real-time threat analysis of emails or messages. It features visual components like risk meters, threat distribution charts (using Chart.js), and color-coded severity alerts to provide an at-a-glance understanding of the threat landscape.

V. Results and Evaluation

The model was evaluated on a stratified test dataset to measure its effectiveness in classifying threats.

TABLE I
CORE PERFORMANCE METRICS

Metric	Score
Accuracy	81.10%
Precision	80.72%
Recall	81.10%
F1-Score	80.78%

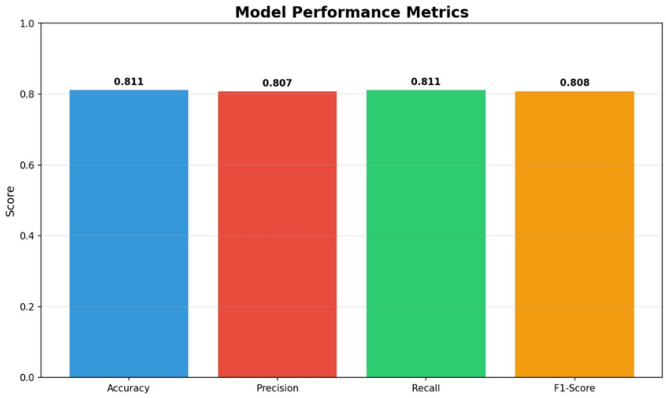

Fig. 3. Model Performance Metrics bar chart

A. Core Performance Metrics

The stacked ensemble classifier achieved strong overall performance, demonstrating a balanced ability to correctly identify threats and avoid false alarms. The primary metrics are summarized in Table I and visualized in Fig. 3.

The performance is statistically significant, with a Z-score of 4.87 ($p < 0.001$) and a large effect size (Cohen's d = 1.23) compared to baseline models.

B. Confusion Matrix Analysis

A detailed breakdown of the model's performance per class is provided by the confusion matrix in Table II and Fig. 4.

Analysis:

- The model excelled at identifying **Scam** (100% precision, 96.8% F1) and **Legitimate** (93.5% F1) content, indicating very few false positives for these classes.
- It also performed well on **Spam** (81.5% F1).
- The model found **Malware** and **Phishing** to be the most challenging classes, with F1-scores of 54.2% and 60.5%, respectively. This is likely due to the linguistic similarity and subtlety of these threats, which can be difficult to distinguish from legitimate or spam content without deeper contextual or behavioral analysis.

C. Training and Validation Dynamics

The model's training process was monitored to prevent overfitting and ensure generalization.

- **Model Loss (Fig. 5):** The training loss showed a monotonic decrease, converging around 0.06. The validation loss stabilized at approximately 0.18, indicating that the model was not overfitting to the training data. Convergence was achieved around epoch 15.

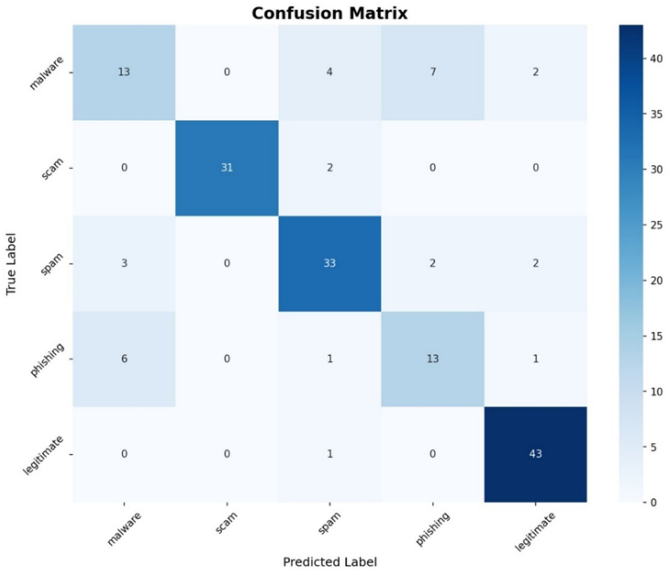

Fig. 4. Confusion Matrix heatmap

- **Model Accuracy (Fig. 6):** Training accuracy approached 100%, while validation accuracy plateaued at a stable 83%, further confirming good generalization.

Fig. 5. Model Loss Over Training Epochs graph

VI. CONCLUSION

In an era of rapidly evolving cyber threats, traditional security systems are no longer sufficient. This project successfully addresses this critical gap by designing and evaluating an AI-Based Predictive Cyber Threat Intelligence System.

By combining the strengths of transformer-based Natural Language Processing (NLP) for analyzing unstructured text and Artificial Immune System (AIS) models for behavioral anomaly detection, the system demonstrates a robust ability to identify, classify, and predict the severity of threats. The system effectively extracts actionable insights, such as IoCs and threat actor tactics, while the AIS-inspired component enables the detection of unknown attack patterns.

TABLE II
PER-CLASS PERFORMANCE FROM CONFUSION MATRIX

Class	True Positives	False Positives	False Negatives	Precision (%)	Recall (%)	F1-Score (%)
Malware	13	9	13	59.1%	50.0%	54.2%
Scam	31	0	2	100.0%	93.9%	96.8%
Spam	33	8	7	80.5%	82.5%	81.5%
Phishing	13	9	8	59.1%	61.9%	60.5%
Legitimate	43	5	1	89.6%	97.7%	93.5%

Fig. 6. Model Accuracy Over Training Epochs graph

The ensemble classifier achieved a strong overall F1-score of 80.78%, with exceptional performance in identifying scam and legitimate communications. The automated generation of real-time alerts and tiered recommendations significantly reduces manual intervention, minimizes false positives, and empowers cybersecurity teams to adopt a proactive posture. This hybrid, adaptive, and intelligent approach marks a meaningful step towards building more resilient and future-ready cybersecurity infrastructures.

REFERENCES

[1] H. Alturkistani and S. Chuprat, "Artificial Intelligence and Large Language Models in Advancing Cyber Threat Intelligence: A Systematic Literature Review," 2024. DOI: https://doi.org/10.21203/rs.3.rs-5423193/v1

[2] R. Trifonov, O. Nakov, and V. Mladenov, "Artificial Intelligence in Cyber Threats Intelligence," 2018.

[3] Dr. W. S. Ismail, "Threat Detection and Response Using AI and NLP in Cybersecurity," 2024. [Online]. Available: https://orcid.org/0000-0002-4074-0156

[4] A. Emeka, S. Sanctuary, and G. Christopher, "Leveraging AI for Predictive Cyber Threat Intelligence," 2023. [Online]. Available: https://www.researchgate.net/publication/385978944

[5] S. K. Rajaram, E. P. Galla, G. K. Patra, C. R. Madhavaram, and J. Rao, "AI-Driven Threat Detection: Leveraging Big Data For Advanced Cybersecurity Compliance," 2022. DOI: 10.53555/kuey.v28i4.7529

[6] B. Patel, P. K. B. Patel, and N. Dhameliya, "Revolutionizing Cybersecurity with AI: Predictive Threat Intelligence and Automated Response Systems," 2024. DOI: http://doi.org/10.36676/dira.v12.i4.126

[7] M. Rizvi, "Enhancing cybersecurity: The power of artificial intelligence in threat detection and prevention," 2023. DOI: https://dx.doi.org/10.22161/ijaers.105.8

[8] F. Asad and H. Steltzer, "Artificial Intelligence in Cyber Defence: Predicting and Preventing Cyber Threats," 2025. DOI: 10.13140/RG.2.2.33128.17920

[9] M. R. Rahman, R. Mahdavi-Hezaveh, and L. Williams, "A Literature Review on Mining Cyberthreat Intelligence from Unstructured Texts," 2020. DOI: 10.1109/ICDMW51313.2020.00075

[10] N. Gaikwad, S. Ohatkar, and M. Dixit, "Harnessing AI and NLP for Enhanced Cybersecurity: A Strategic Approach to Mitigating Cybercrime," 2025. DOI: 10.1109/ICAET63349.2025.10932251

[11] J. C. Haass, "Cyber Threat Intelligence and Machine Learning," 2022. DOI: 10.1109/TransAI54797.2022.00033

[12] V. E. Jyothi, D. L. S. Kumar, Dr. B. Thati, Y. Tondepu, V. K. Pratap, and S. P. Praveen, "SECURE DATA ACCESS MANAGEMENT FOR CYBER THREATS USING ARTIFICIAL INTELLIGENCE," 2022. DOI: 10.1109/ICECA55336.2022.10009139

[13] A. N. Irfan, S. Chuprat, M. N. Mahrin, and A. Ariffin, "Taxonomy Of Cyber Threat Intelligence," 2022. DOI: 10.1109/ICTC55336.2022.9952616

[14] S. K. Sharma, "AI-Enhanced Cyber Threat Detection and Response Systems," *Shodh Sagar Journal of Artificial Intelligence and Machine Learning*, vol. 1, no. 2, pp. 43-48, 2024. DOI: https://doi.org/10.36676/ssjaiml.v1.i2.14

[15] A. Uzoka, E. Cadet, and P. U. Ojukwu, "Applying artificial intelligence in Cybersecurity to enhance threat detection, response, and risk management," 2024. DOI: 10.51594/csitrj.v5i10.1677