IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14427

Threat Intelligence System for Cyber Attacks

Kunal Naikade[1], Ritesh Patni[2], Prathmesh Jaiswal[3], Smita Chunamari[4]

Computer Department, University of Mumbai, Navi Mumbai, India[1-4]

Abstract: Cybersecurity threats have become increasingly sophisticated, with phishing attacks remaining one of the most prevalent and damaging forms of cybercrime. This project focuses on developing a threat intelligence system designed to detect and mitigate phishing links in real time. By utilizing advanced machine learning algorithms and natural language processing techniques, the system analyzes URLs, email content, and website characteristics to identify malicious patterns indicative of phishing attempts. The model is trained on large datasets of legitimate and fraudulent links to maximize detection accuracy and reduce false positives. Additionally, the system integrates threat intelligence feeds to enhance adaptability against evolving attack strategies. The ultimate goal of this project is to provide a proactive cybersecurity solution that identifies phishing threats before they compromise user data or organizational networks. If effectively implemented, the system can strengthen online security, prevent financial losses, and support the broader effort toward safer digital ecosystems.

I. INTRODUCTION

The "Threat Intelligent System for Cyber Attacks" project focuses on developing an intelligent framework capable of detecting and preventing phishing links and other malicious online activities. In today's digital landscape, cyber threats have become increasingly complex, targeting individuals, businesses, and even government infrastructures. Among these, phishing remains one of the most common and dangerous tactics used by attackers to steal sensitive information, credentials, and financial data. This project aims to design a system that leverages threat intelligence principles and machine learning algorithms to identify and classify phishing links effectively. By analyzing patterns in URLs, email content, and website behavior, the system can differentiate between legitimate and fraudulent sources. The model will be trained using large datasets containing both benign and malicious links to ensure high detection accuracy and adaptability to new attack vectors. Furthermore, the project integrates real-time threat updates and feedback mechanisms to enhance the system's awareness of emerging cyber threats. The ultimate goal is to create an efficient and automated solution that strengthens cybersecurity defenses, reduces human error in threat identification, and mitigates the risk of data breaches. Through this approach, the proposed system contributes to building a more secure and resilient digital environment

Techniques, and Procedures (TTPs) and Indicators of Compromise (IoCs) used by specific threat actors. This project aims to bridge the gap between information and action, empowering security teams to anticipate attacks before they penetrate the perimeter. Through the integration of machine learning and automated workflows, the system filters out noise to focus on high-priority risks tailored to an organization's unique vulnerabilities. Ultimately, this platform shifts the defensive posture from guesswork to evidence-based decision-making, significantly reducing the impact and frequency of successful breaches.

II. MOTIVATION

The motivation behind this project is to develop an intelligent threat detection system capable of identifying phishing links and preventing malicious activities before they cause harm. By incorporating machine learning, pattern analysis, and real-time threat intelligence, the system aims to provide stronger, faster, and more adaptive protection against cyber attacks. This approach not only enhances cybersecurity but also fosters greater user trust and data safety in digital environments.

With the rapid growth of digital communication and online transactions, cyber threats have evolved into a major global concern. Among these, phishing attacks pose a significant challenge due to their ability to deceive users into revealing sensitive information such as passwords, financial details, and personal data. Traditional security systems often fail to detect these threats promptly because attackers continuously change their tactics to bypass existing defenses.

III. LITERATURE SURVEY

The evolution of **Cyber Threat Intelligence (CTI)** reflects a shift from manual oversight to automated, predictive defense frameworks necessitated by increasingly sophisticated digital adversaries. Early defensive efforts focused on

DOI: 10.17148/IJIREEICE.2026.14427

traditional statistical methods and signature-based detection, which, while effective against known malware, struggled with the scale and speed of modern polymorphic threats.

- **In the paper titled "Detecting Phishing Websites Using Machine Learning Techniques" [1],** the authors Mohammad Basit and Muhammad Raza Naqvi propose a machine learning–based phishing detection framework that analyzes both URL-based and webpage content features to identify phishing websites. The study employs algorithms such as Random Forest, Decision Tree, and Support Vector Machine (SVM) to classify malicious**.**

- **In the paper titled "PhishTank: An Automated Real-Time Phishing Detection Framework" [2],** the authors
S. Narayanan and R. Shukla introduce a real-time phishing link detection system that utilizes ensemble learning techniques to analyze suspicious URLs. The proposed system is lightweight and designed for quick implementation in real-time environments, making it suitable for dynamic web security applications. Despite its advantages, the framework sometimes produces a higher false positive rate when legitimate websites exhibit link structures similar to phishing pages. This study was published in the Journal of Network and Computer Applications by Elsevier in the year 2021 with contributions from researchers at Indian Institute of Technology, achieving detection accuracy above 95%.

- **In the paper titled "Intelligent Phishing Email Detection Using NLP and Deep Learning" [3],** the authors present a deep learning framework that integrates Natural Language Processing (NLP) techniques to analyze the semantic meaning and linguistic patterns within email messages to detect hidden phishing intent. The model focuses on identifying subtle textual cues that traditional detection systems might overlook, significantly improving precision for phishing email detection. However, the approach is computationally expensive and requires a large dataset for training, which can limit real-time deployment. This work was published by Springer in the book series Lecture Notes in Networks and Systems in the year 2022, achieving an accuracy of up to 96%.

In the paper titled "Hybrid Machine Learning Approach for Phishing URL Detection" [4], the authors L. Al-Saffar and R. Hussain propose a hybrid model that combines Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to improve phishing URL detection performance. The hybrid architecture leverages the pattern recognition capabilities of CNNs and the sequential analysis strengths of RNNs, leading to improved detection speed and reliability. Nevertheless, the model can be sensitive to imbalanced datasets and may incorrectly classify highly targeted or newly emerging phishing campaigns. This research was published in the Computers & Security journal by Elsevier in the year 2023 and achieved a detection rate above 97%, outperforming several traditional machine learning methods.

- **In the paper titled "Advanced Threat Intelligence for Phishing Detection Using Ensemble Learning" [5**], the authors T. Zhang and F. Chen present a phishing detection framework that integrates real-time threat intelligence with ensemble learning models to dynamically update the detection system. The approach enables the model to adapt quickly to evolving phishing techniques, improving its ability to detect newly emerging attacks. However, the architecture requires a complex infrastructure and large training datasets tomaintain consistent performance. This study was published in IEEE Transactions on Information Forensics and Security in the year 2024 and achieved an average detection accuracy of approximately 98%, supporting more effective early prevention of phishing links.

IV. PROBLEM STATEMENT

Cyber attacks, especially phishing, remain one of the most common and effective methods used by attackers to compromise online security. Despite extensive cybersecurity measures, detecting phishing links remains difficult because attackers frequently modify URLs, content, and website features to appear legitimate.

V. PROPOSED SYSTEM

- Data Collection and Preprocessing Collect a comprehensive dataset containing legitimate and phishing URLs from reliable sources such as PhishTank, UCI Machine Learning Repository, and open-source cybersecurity databases. Preprocess the data by cleaning, tokenizing, and extracting key features like URL length, domain age, SSL certificate validity, and redirection patterns. This step ensures consistency, removes duplicates, and enhances data quality for model training.
- Implement advanced feature engineering techniques to identify indicators of phishing activity. These include lexical features (presence of special characters, domain structure), host-based features (IP address usage, DNS information), and content-based features (HTML tags, embedded links). The extracted features serve as the input to

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14427

the learning model for accurate classification.

- Utilize machine learning frameworks such as Scikit-learn, TensorFlow, or PyTorch to develop classification models. Algorithms such as Random Forest, Gradient Boosting, and Neural Networks will be trained to distinguish between phishing and legitimate URLs. The model training process will employ cross-validation and hyperparameter tuning to optimize.
- Evaluate the developed model using metrics such as accuracy, precision, recall, and F1-score. The testing phase will involve unseen phishing datasets to assess robustness and generalizability. Comparative analysis will be conducted with existing detection frameworks to validate improvements in performance and reliability.

A. System Design and Architecture

The Threat Intelligent System for Cyber Attacks is designed to automatically detect phishing links and malicious URLs using machine learning and threat analysis techniques. The system analyzes URL and content-based features, classifies them as legitimate or malicious, and provides real-time detection reports. Below is the layout process of the proposed system:

- Data collection: Collect phishing and legitimate URLs from reliable cybersecurity datasets such as PhishTank, Kaggle, and open-source threat repositories. These datasets serve as labeled inputs for model training and evaluation.
- Preprocessing: Clean and normalize the collected data by removing duplicates, correcting missing values, and standard.
- Model selection: Choose appropriate machine learning algorithms such as Random Forest, Gradient Boosting, or Neural Networks.
- Training: Train the model using labeled data with known classifications. Employ cross-validation and hyperparameter tuning to improve accuracy and prevent overfitting. The training dataset will include varied phishing patterns to enhance adaptability.
- Evaluation: Evaluate model performance using metrics such as accuracy, precision, recall, and F1-score.
- Postprocessing: Apply threshold optimization and bias correction to minimize false positives and false negatives.
- Deployment: Deploy the final model as a user-interactive web tool or API-based service capable of real-time URL analysis. The system continuously monitors incoming data, flags potential phishing attempts, and generates alert reports for users and administrators.

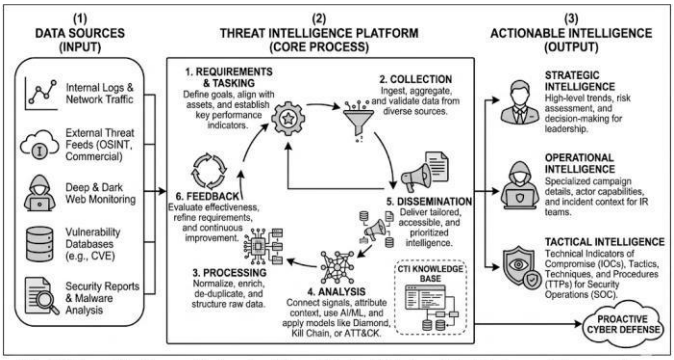

Fig. 1. Threat Intelligence System for Cyber Attacks: High-Level Architecture and Process Flow.

Fig. 1. Architecture design

B. System Architecture Overview

The proposed system for **cyber-attack threat intelligence** utilizes machine learning and data analytics. The architecture focuses on threat data collection, preprocessing, model training, and continuous monitoring for improved cyber defense. Fig. ?? illustrates the complete workflow.

1) *Data Acquisition and Conversion:* Raw cyber threat data is collected from multiple sources such as network logs, security alerts, phishing URLs, and threat intelligence feeds. These data sources include firewall logs, intrusion detection system (IDS) alerts, and publicly available phishing datasets.

2) *Preprocessing Pipeline:* The preprocessing stage involves several sequential data preparation steps:
   - Data cleaning to remove duplicate, incomplete, or irrelevant records.
   - Feature extraction from URLs, email headers, and network traffic patterns.
   - Encoding of categorical data and normalization of numerical features.
   - Removal of noise and outliers to improve the reliability of the dataset.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14427

3) *Data Augmentation and Cross-Validation:* To improve the model's generalization capability and prevent overfitting, the dataset is divided into training and testing sets using an **80/20 split**.

4) *Modeling and Feature Extraction:*Several machine learning algorithms are applied to detect cyber threats and phishing attacks. These include Random Forest, Decision Tree, Support Vector Machine (SVM), and ensemble learning techniques. Feature extraction focuses on identifying suspicious URL structures, abnormal network behavior, and malicious patterns within system logs.

5) *Evaluation and Metrics:* Model performance is eval-Model performance is evaluated using standard classification metrics such as accuracy, precision, recall, F1-score, and confusion matrix. Additional evaluation methods such as ROC curves and detection rate comparisons are used to analyze the efficiency of different algorithms. These metrics help determine the effectiveness of the threat detection system.

6) *Prediction and Output:* The trained system analyzes incoming network data and classifies it into categories such as phishing attack, malware activity, suspicious traffic, or normal behavior. The system generates alerts for detected threats and provides detailed information about the type and severity of the attack.

7) *Continuous Learning:*To ensure long-term effectiveness, the system supports continuous learning by periodically updating the dataset with newly identified cyber threats. The models are retrained using updated threat intelligence data, enabling the system to adapt to evolving cyber attack techniques and maintain high detection accuracy.

8) This architecture integrates efficient data collection, machine learning–based threat detection, and continuous intelligence updates to provide an effective **Threat Intelligence System for Cyber Attacks** capable of detecting and preventing emerging cybersecurity threats.

9) *Data Flow Diagram*

Fig. 2. Data Flow Diagram Level 0

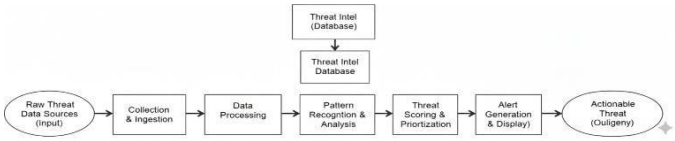

Fig. 3. Data Flow Diagram Level 1

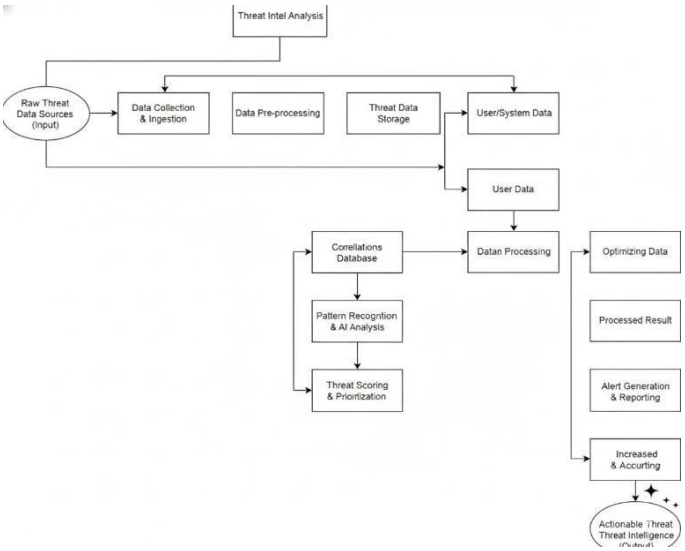

Fig. 4. Data Flow Diagram Level 2

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14427

VI. HARDWARE AND SOFTWARE REQUIREMENTS

A. *Software Requirements*

- Programming Language: Python (for data analysis, cyber threat detection and model development).
- Libraries and Frameworks:

NumPy, Pandas (for data manipulation) Scikit-learn (for machine learning models)

TensorFlow or PyTorch (for deep learning models) Scapy or PyShark (for network packet analysis and monitoring)

Matplotlib, Seaborn (for data visualization and threat analysis dashboards)

BeautifulSoup or Requests (for web data extraction and phishing URL analysis)

- Development Environment : Jupyter Notebook or any Python IDE (e.g., PyCharm, VS Code)
- Data Storage and Management : Local storage for smaller cybersecurity datasets (e.g., AWS S3, Google Cloud Storage) for storing large-scale threat intelligence data and logs.
- Version Control : Git and GitHub or GitLab for version control and collaboration.

B. *Hardware Requirements*

- Computing Resources : A high-performance laptop or desktop with at least 8GB of RAM and a multi-core CPU for running cybersecurity analysis and machine learning models.GPU(e.g,NVIDIA) for faster processing and training of advanced threat detection models.
- Storage: At least 100GB of available storage space for storing cybersecurity datasets, network logs,phising. Cloud Computing Resources (optional but recommended for large datasets or complex models) : Access to cloud-based GPU/TPU instances from platforms like AWS,Google Cloud, or Azure.

VII. SYSTEM DEVELOPMENT AND OUTPUT

A. *Algorithms*

- **Random Forest:**

Random Forest [7] is an ensemble machine learning algorithm used for classification and prediction tasks. It works by constructing multiple decision trees during training and combining their outputs to produce a more accurate and stable prediction. Each tree in the forest analyzes different subsets of the dataset and features, which helps reduce overfitting and improves generalization. Random Forest is widely used in cybersecurity for detecting phishing websites, malicious URLs, and abnormal network activities because of its high accuracy and ability to handle large and complex datasets.

- **SVM:**

Support Vector Machine (SVM) [8] is a supervised machine learning algorithm used for classification and regression tasks. The algorithm works by finding the optimal hyperplane that separates data points of different classes in the feature space. In cybersecurity applications, SVM is often used to classify network traffic, phishing links, and malicious activities by identifying patterns that distinguish normal behavior from suspicious activity. For datasets that are not linearly separable, kernel functions are used to map the data into a higher-dimensional space where classification becomes easier. [9]

- **Decision Tree:**

Decision Tree [10] is a tree-structured classification algorithm where the dataset is split into smaller subsets based on feature values. Each internal node represents a decision rule based on a feature, each branch represents an outcome of that rule, and each leaf node represents a final classification result. Decision Trees are easy to interpret and useful for identifying patterns in cybersecurity datasets such as phishing URLs, abnormal login attempts, and suspicious network behavior. However, a single decision tree may sometimes overfit the training data, which is why it is often combined with ensemble methods. [11]

- **XgBoost and AdaBoost:**

XgBoost: XgBoost The search range may include values such as (100, 300, 500, 1000) for the number of estimators,(2, 10, 1) for maximum tree depth, and (0.01, 0.03, 0.05,0.1, 0.3, 0.5) for the learning rate., maximum depth of the tree and learning rate respectively.

AdaBoost: is another ensemble learning algorithm that combines multiple weak classifiers to create a strong classifier. It works by assigning higher importance to incorrectly classified instances and adjusting the weights of training samples during each iteration. In cyber attack detection systems, AdaBoost helps improve the detection of phishing links and malicious patterns by focusing on difficult-to-classify threats. Parameter tuning techniques such as grid-search used to determine optimal learning rates and estimator counts for better perform [12]

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14427

B. Output/Result

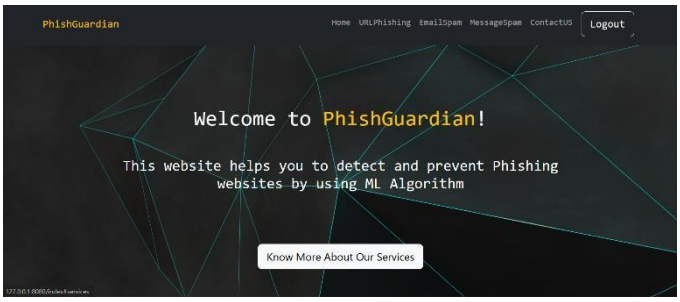

Fig. 5. Interface

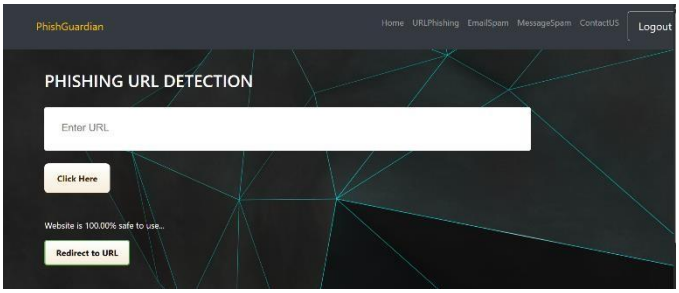

Fig. 6. URL Phising Image

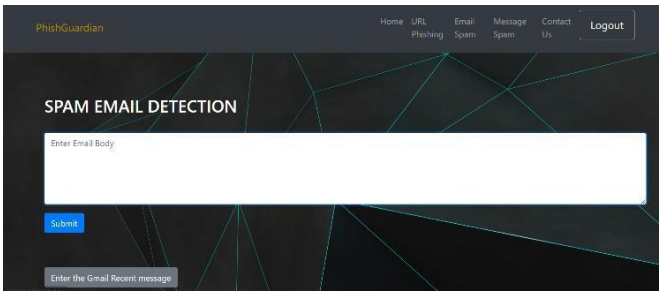

Fig. 7. SPAM Email Image

VIII. FUTURE WORK

Multi-Source Threat Intelligence Integration explore the integration of multiple cybersecurity data source such as network traffic logs, system event logs, phising databases, malware respositories, and real-time threat intelligence feeds to obtain a more comprehensive understanding.

Explainable AI (XAI) Integrate explainable AI techniques to provide insights into the key features and indicators responsible for predicting cyber threats. This will help security analysts understand why a particular activity is classified as malicious, thereby increasing the transparency, interpretability and trustworthiness of threat intelligence system.

IX. CONCULSION

In this project, we have laid the groundwork for a comprehensive **Threat Intelligence System for Cyber Attacks** utilizing cybersecurity datasets such as phishing URLs, network traffic logs, and threat intelligence feeds. Through careful data collection and preprocessing, we have ensured that our dataset is diverse and of high quality, which is essential for training robust machine learning models. Our systematic approach to model development and evaluation has prioritized accuracy and generalizability, which are crucial for the practical implementation of the system in real-world cybersecurity environments.

The proposed system integrates advanced machine learning techniques, leveraging algorithms such as Random Forest, Support Vector Machine (SVM), and Decision Tree to analyze cyber threat data.

DOI: 10.17148/IJIREEICE.2026.14427

The As we move into the next phase of our project in **Semester VIII**, we will focus on the execution of our proposed system. This will involve implementing the trained models, conducting further evaluations, and refining the system based on real-world cyber threat data. Continuous improvement and integration of updated threat intelligence will be essential as we work to enhance the system's accuracy, scalability, and adaptability to evolving cyber attack techniques.

Overall, this project aims to contribute to the advancement of cybersecurity and machine learning by developing an effective threat intelligence framework capable of detecting and analyzing cyber attacks. It also lays the foundation for future research in intelligent cybersecurity systems that can proactively identify threats and protect digital infrastructure.

REFERENCES

[1]. Mohammad Basit, Muhammad Raza Naqvi, Faisal Qamar, and Saqib Anwar. A machine learning approach for phishing website detection. IEEE Access, 8:170–185, 2020.

[2]. S. Narayanan, R. Shukla, and P. Rai. PhishTank: An automated real-time phishing detection framework. Journal of Network and Computer Applications, 182:102994, Elsevier, 2021.

[3]. J. Binu and S. Anitha. Intelligent phishing email detection using natural language processing and deep learning. In Lecture Notes in Networks and Systems. Springer, 2022.

[4]. L. Al-Saffar, R. Hussain, and M. Al-Baity. Hybrid machine learning-based phishing URL detection framework. Computers & Security, 127:103120, Elsevier, 2023.

[5]. T. Zhang, F. Chen, and L. Zhao. Advanced threat intelligence for phishing detection using ensemble learning. IEEE Transactions on Information Forensics and Security, 19:298–312, 2024.

[6]. A. Bhattacharya and S. K. Jha. Phishing URL detection using XGBoost and deep learning models. Journal of Cybersecurity Technology, 8(2):99–115, 2023.

[7]. R. Kumar and P. Gopal. CNN-based phishing detection system for secure internet communication. International Journal of Information Security Science, 13(3):55–66, 2024.

[8]. M. Tiwari, A. Roy, and K. Yadav. Detection of malicious web links using CatBoost and behavioral analysis. Computers and Electrical Engineering, 114:109031, Elsevier, 2024.

[9]. A. Dasgupta, S. Jain, and P. Patel. A comparative study on machine learning algorithms for phishing detection. Journal of Information Security and Applications, 75:103584, 2025.

