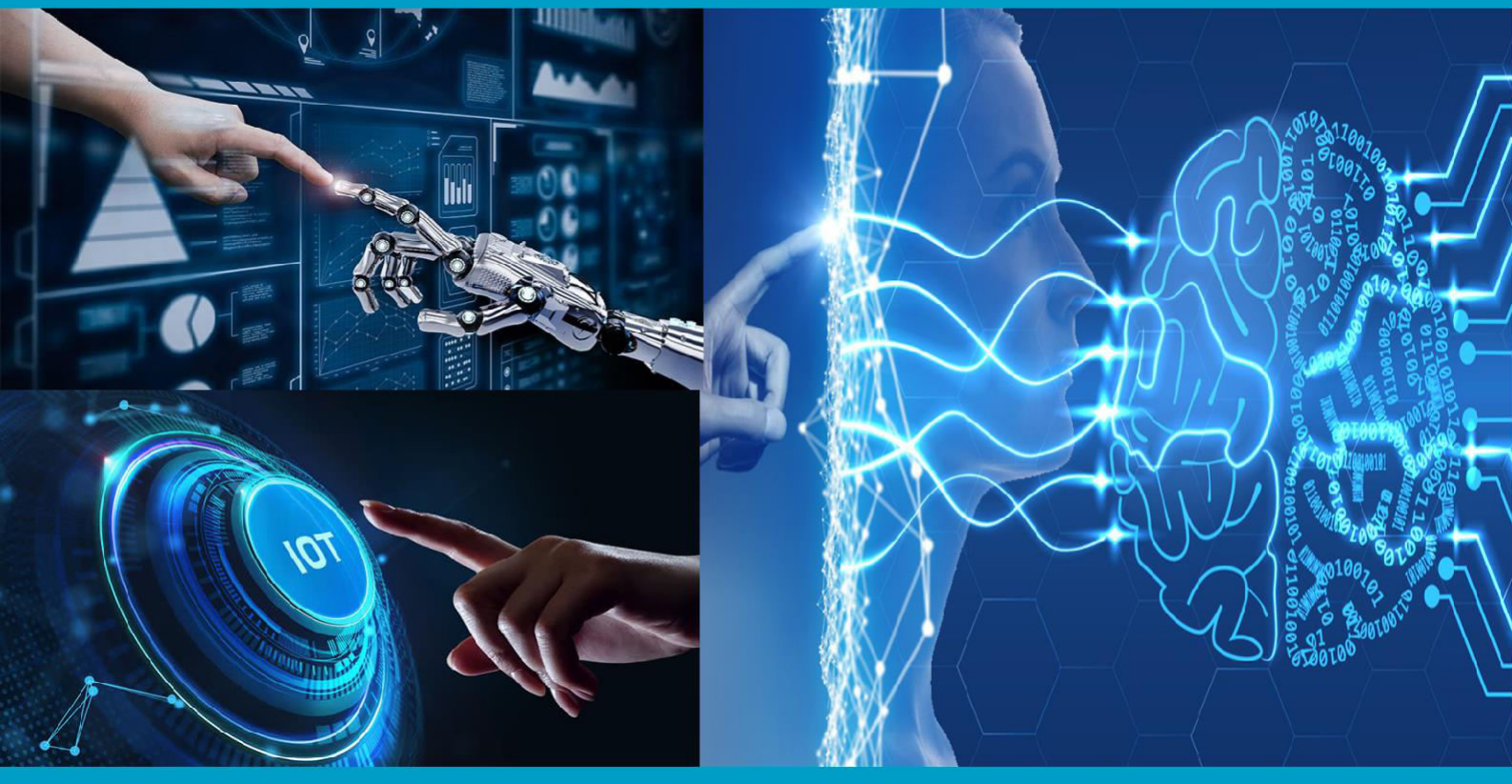
IJMRSETM

e-ISSN: 2395 - 7639



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING, TECHNOLOGY AND MANAGEMENT

Volume 12, Issue 5, May 2025

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.214

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

| ISSN: 2395-7639 | www.ijmrsetm.com | Impact Factor: 8.214 | A Monthly Double-Blind Peer Reviewed Journal |

| Volume 12, Issue 5, May 2025 |

Phishing Link Detector and Vulnerability Scanner

Prof.Deeksha K R [1], Ms.Aishwarya[2], Ms.Sunitha[2], Mr.Vikram[2], Mr.Vikram Shetty[2]

Professor, Dept. of ISE, YIT Moodbidri, Mangalore, Karnataka, India[1]

B.E Student, Dept. of ISE, YIT Moodbidri, Mangalore, Karnataka, India[2]

B.E Student, Dept. of ISE, YIT Moodbidri, Mangalore, Karnataka, India[2]

B.E Student, Dept. of ISE, YIT Moodbidri, Mangalore, Karnataka, India[2]

B.E Student, Dept. of ISE, YIT Moodbidri, Mangalore, Karnataka, India[2]

ABSTRACT: Phishing attacks and software vulnerabilities remain persistent and evolving threats to digital systems. Conventional security approaches, such as rule-based or signature detection, often fall short in identifying novel or adaptive attacks. To address these limitations, machine learning techniques—particularly Random Forest and XGBoost—have emerged as powerful tools for detecting phishing activities and uncovering system vulnerabilities. This study reviews existing literature on these models in cybersecurity contexts, highlighting their implementation in phishing link detection and vulnerability scanning. The analysis identifies methodological trends, evaluates performance outcomes, and outlines current research limitations, laying the foundation for our proposed real-time integrated solution.

KEYWORDS: Phishing Detection, Vulnerability Scanning, Random Forest, XGBoost, Machine Learning, Cybersecurity

I. INTRODUCTION

The increasing frequency and sophistication of cyber threats, such as phishing schemes and software vulnerabilities, pose major challenges to the security of online systems. Phishing typically involves deceiving users into disclosing confidential data by impersonating trusted sources, while vulnerabilities refer to flaws in software that can be exploited by attackers. Traditional defense mechanisms, which rely heavily on predefined signatures or heuristic rules, often struggle to detect new or adaptive threats. In contrast, machine learning models are capable of identifying hidden patterns in data, allowing for more adaptive threat detection. This paper explores the application of machine learning—specifically the Random Forest and XGBoost algorithms—for enhancing phishing detection and automating vulnerability identification.

II. LITERATURE REVIEW

Recent studies have demonstrated the growing effectiveness of machine learning techniques in addressing cybersecurity threats, particularly phishing detection and vulnerability scanning.

Abdelhamid et al. [1] introduced a hybrid classification model integrating K-Nearest Neighbors (KNN) and decision trees to detect phishing attempts. The model exhibited commendable accuracy but encountered scalability limitations in large-scale applications.

Verma and Das [2] conducted a comparative study of several machine learning algorithms for URL-based phishing detection. Their findings indicated that the Random Forest algorithm achieved over 95% accuracy when lexical URL features were used, demonstrating its strength in phishing classification.

Sahingoz et al. [3] applied natural language processing (NLP) techniques alongside various ML classifiers for phishing URL detection. Their study highlighted that ensemble methods like Random Forest and XGBoost consistently outperformed other classifiers in terms of accuracy and robustness.

Xiang et al. [4] developed an improved phishing URL detection model using the XGBoost algorithm. The model achieved an accuracy of 98.2%, showcasing its capability to handle structured datasets efficiently and effectively prioritize feature importance.

Rathod and Rao [5] proposed a real-time phishing detection framework powered by Random Forest. Their system utilized feature reduction techniques to optimize performance and reduce computational overhead in real-time applications.

In the context of vulnerability detection, Aljawarneh et al. [6] introduced a hybrid intrusion detection system that combined Random Forest with deep learning. Their approach successfully enhanced anomaly detection and reduced false positives, making it suitable for vulnerability identification in networked environments.

Collectively, these studies underscore the strong performance of Random Forest and XGBoost in phishing detection and their potential in automated vulnerability scanning. However, challenges such as data imbalance, real-time processing, and deployment in live systems still present opportunities for further research and development.

III. RESULT

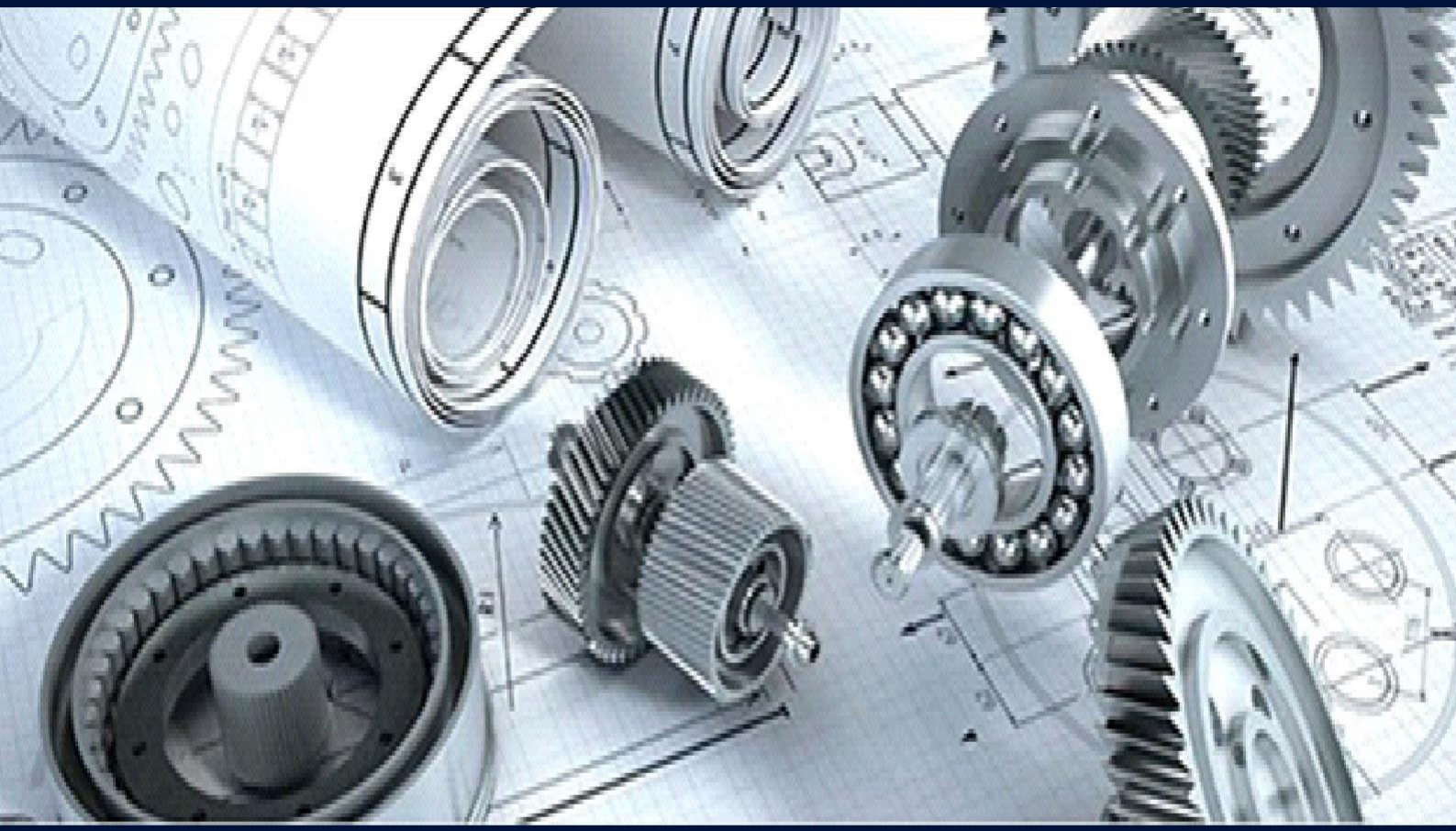
From the reviewed literature, Random Forest and XGBoost consistently deliver high accuracy in phishing detection, often exceeding 95%. XGBoost, in particular, excels in handling structured data and imbalanced datasets. However, integration with real-time vulnerability scanning is limited in current research. Most systems are tested in controlled environments without deployment in actual systems. The survey confirms the potential of combining these models to create a robust, real-time detection system.

IV. CONCLUSION

Machine learning provides powerful tools for enhancing cybersecurity through automated phishing detection and vulnerability scanning. Random Forest and XGBoost are standout algorithms due to their accuracy, speed, and adaptability. The literature reveals a trend toward hybrid and ensemble models but also highlights gaps such as deployment challenges and outdated datasets. Our project aims to build on these findings by integrating these techniques into a real-time system capable of both phishing detection and vulnerability scanning.

REFERENCES

[1] Abdelhamid, N., Ayesh, A., & Thabtah, F. (2014). Phishing detection based on hybrid classifiers. Expert Systems with Applications.
[2] Verma, R., & Das, A. (2017). URL-based phishing detection using machine learning. In IEEE ICCSP.
[3] Sahingoz, O. K., et al. (2019). Machine learning based phishing detection from URLs. Expert Systems with Applications.
[4] Xiang, G., et al. (2020). An improved URL phishing detection model using XGBoost. Computers & Security.
[5] Rathod, V., & Rao, A. (2021). Real-time phishing detection system using Random Forest. In Springer ICCIDS.
[6] Aljawarneh, S., et al. (2018). Anomaly-based intrusion detection system using hybrid techniques. Journal of King Saud University - Computer and Information Sciences.



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING, TECHNOLOGY AND MANAGEMENT

+91 99405 72462 +91 63819 07438 ijmrsetm@gmail.com