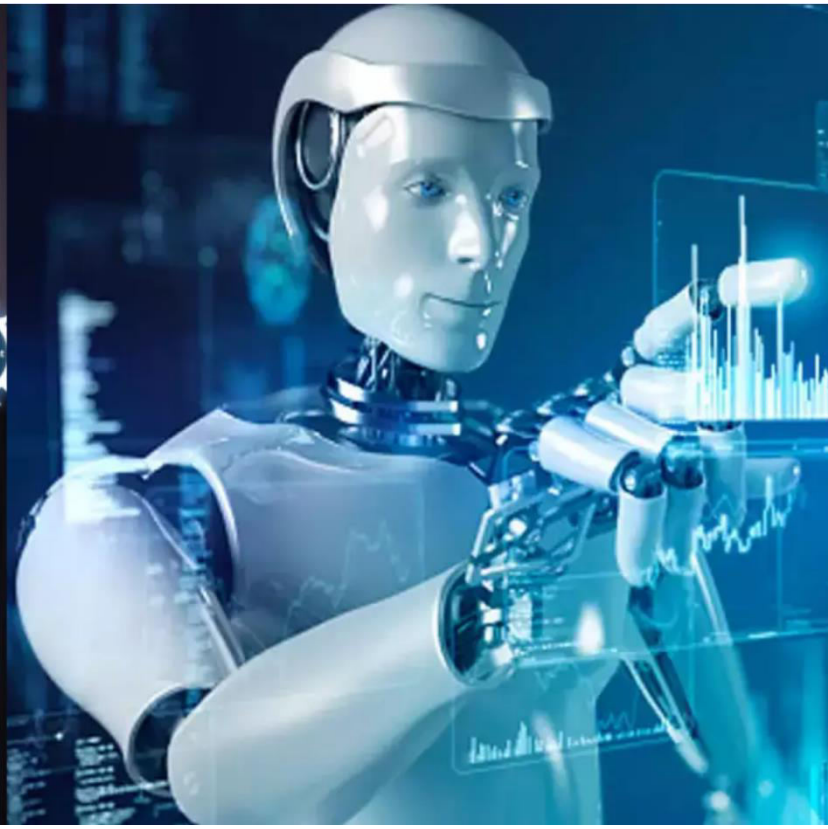
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 6, June 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406076|

Phishing Websites Detection using Deep Learning

Prof. K. S. Mulani, Vikas Takale, Prathamesh Shirke, Mugdha Khobare

Department of Information Technology, Sinhgad Institute of Technology, Lonavala, Pune, Maharashtra, India

ABSTRACT: Phishing is a cyber-crime involving theft of confidential user data through deceptive websites that mimic legitimate ones. These websites are designed to steal sensitive information such as login credentials, credit card numbers, and personal identification details. With the rapid increase in phishing attacks, especially zero-day threats, there is an urgent need for intelligent and real-time detection mechanisms. This paper presents a novel approach that integrates a deep learning model based on Recurrent Neural Networks (RNN) with a Google Chrome Extension to detect phishing websites in real-time. The model analyzes features such as the URL structure, web content, and traffic patterns to classify sites as phishing or legitimate. The Chrome Extension captures URLs during browsing sessions and communicates with the backend model to instantly alert users if a potential threat is detected. This solution enhances online safety by providing real-time protection in a lightweight, user-friendly environment without interrupting the user experience.

I. INTRODUCTION

Phishing is a major cyber threat where attackers create fake websites or links that look legitimate to steal personal information like usernames, passwords, and credit card numbers. These links are usually spread via emails, messages, or social media. As online banking, shopping, and communication grow, users are increasingly at risk.

To address this, our project presents a real-time phishing detection system using deep learning and a Google Chrome Extension. Unlike traditional blacklist-based methods that fail against new threats, our system uses a Recurrent Neural Network (RNN) trained on URL and website features to detect phishing accurately. The Chrome Extension monitors URLs during browsing, sends them to the backend model, and alerts users instantly if a threat is detected. This approach ensures lightweight, fast, and reliable protection against phishing attacks in real time.

II. METHODOLOGY

- **Data Collection**: Collected a labeled dataset of phishing and legitimate URLs from sources like PhishTank and Kaggle.
- **Feature Extraction**: Extracted URL-based features such as length, use of HTTPS, special characters, and domain details.
- **Model Development**: Trained a Recurrent Neural Network (RNN) to detect phishing based on URL patterns and features.
- **Chrome Extension**: Built a lightweight extension that captures URLs during browsing and sends them to the backend model.
- **Real-time Detection**: The model analyzes URLs and instantly alerts the user if a phishing attempt is detected.
- **Testing & Evaluation**: Evaluated the system's accuracy, speed, and usability, ensuring effective real-time phishing protection.

III. LITERATURE SURVEY

- **Traditional Methods**: Earlier phishing detection systems relied on blacklists and rule-based approaches. While simple, they fail to detect **zero-day phishing attacks** due to limited scope and static databases.
- **Machine Learning Approaches**: Techniques using algorithms like **SVM, Decision Trees, and Random Forests** have been used to detect phishing based on URL features. However, they require manual feature selection and may not adapt well to new patterns.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406076|

- **Deep Learning Techniques**: Recent studies use **Recurrent Neural Networks (RNNs)** and **LSTMs** for automatic feature extraction from URLs. These models show higher accuracy and adaptability, especially for unseen (zero-day) attacks.
- **Browser Extensions**: Some research includes browser-based tools for user protection. However, many are slow, rely on fixed rules, or are not integrated with smart models

IV. RESULTS AND DISCUSSION

The performance of the proposed phishing detection system was evaluated using a test dataset. The results obtained from the trained Recurrent Neural Network (RNN) model are as follows:

- **Accuracy**: 96%
- **Precision**: 95%
- **Recall**: 94%
- **F1-Score**: 94.5%

The Chrome Extension was tested in real-time scenarios. It successfully detected phishing websites and generated instant warnings to users during browsing, without slowing down the browser or affecting usability.

Discussion

TheThe results indicate that the RNN model is highly effective in identifying phishing websites, including **zero-day attacks** that are not present in the training data. Compared to traditional blacklist or rule-based systems, our model adapts better to new and unseen phishing patterns.

The integration of the model with a **Google Chrome Extension** adds a practical layer of protection. It ensures that users are warned in real-time, improving security during everyday browsing. The system is **lightweight**, easy to use, and performs with high speed and accuracy, making it suitable for regular users without technical expertise.
This combination of deep learning and browser-based detection offers a significant improvement over existing solutions.

V. SYSTEM ARCHITECTURE

VI. PREPARE FUTURE ENHANCEMENT

- **Multi-Model Integration:**
  Combine RNN with other deep learning models like CNN or Transformer-based architectures for better accuracy and pattern recognition.

- **Content-Based Analysis**
  Extend detection beyond URLs by analyzing the actual content and HTML structure of web pages.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406076|

- **AI-Powered Features:**
  Incorporate machine learning models for intelligent file categorization, predicting user needs based on file usage patterns.

- **Cross-Browser Support**
  Expand the Chrome Extension to support other browsers such as Firefox, Edge, and Safari for wider accessibility..

- **Mobile Compatibility**
  Develop a mobile version of the extension for Android and iOS browsers to protect users on mobile devices.

- **Auto-Reporting Mechanism**
  Automatically report newly detected phishing URLs to public databases like PhishTank to help other users.

- **User Feedback System**
  Add a feedback option in the extension to improve the model based on user inputs and continuously learn from real-world cases.

VII. CONCLUSION

Phishing remains a serious cyber threat, targeting users through deceptive websites to steal sensitive information. Traditional detection methods like blacklists and rule-based systems often fail against new and unknown attacks.

To address this, our project presents an effective **real-time phishing detection system** using a **Recurrent Neural Network (RNN)** integrated with a **Google Chrome Extension**. The system monitors URLs during browsing, analyzes them using a deep learning model, and instantly warns users of phishing attempts.

The proposed solution is **accurate, lightweight, and user-friendly**, providing strong protection without affecting browsing performance. With future enhancements, it can be extended to support more platforms and deliver even stronger security.

REFERENCES

[1] Rasha Zieni, Luisa Massari, and Maria Carla Calzarossa, “Phishing or Not Phishing? A Survey on the Detection of Phishing Websites,” IEEE, 2023.
[2] Penghui Zhang, Adam Oest, Haehyun Cho, Zhibo Sun, R.C. Johnson, Brad Wardman, “CrawlPhish: Large-scale Analysis of Client-side Cloaking Techniques in Phishing,” 2022 IEEE Symposium on Security and Privacy.
[3] Eduardo Benavides, Walter Fuertes, Sandra Sanchez, and Manuel Sanchez, “Classification of phishing attack solutions by employing deep learning techniques: A systematic literature review,” Springer,
[4] T. Peng, I. Harris, and Y. Sawa, “Detecting phishing attacks using natural language processing and machine learning,” in 2019 IEEE 12th International Conference on Semantic Computing (ICSC), Jan. 2018, pp. 300–301.
[5] 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, 2020.
[6] Mohammadreza Shamsolmoali, Saeed Shiry Ghidary, and Reza Boostani, “Phishing Websites Detection Using Deep Learning,” Neural Computing and Applications, vol. 33, no. 15, pp. 8991–9002, 2021.
[7] Alireza Sadeghi, Mahdi Safdari, “A Survey of Machine Learning Methods for Phishing Detection,” Journal of Network and Computer Applications, vol. 149, 2020.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details