ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

An Optimized Hybrid Machine Learning and Deep Learning Framework for Phishing Detection

Prof. Usha K[1], Gowri Kannakatti[2], Chithra R[2], Boodalu Priya[2], Bhumika R[2], Gangamma[2]

Assistant Professor, Dept. of CSE, Jain Institute of Technology, Davangere, Karnataka, India[1]

UG Students, Dept. of CSE, Jain Institute of Technology, Davangere, Karnataka, India[2]

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150500064

Received: 01 April 2026; Accepted: 06 April 2026; Published: 01 June 2026

ABSTRACT

Phishing remains one of the most persistent cybersecurity threats, targeting users through fraudulent emails, websites, and evolving digital platforms. Although machine learning (ML) and deep learning (DL) techniques have improved detection rates, existing models still face limitations such as poor adaptability to new attack patterns, reliance on manual feature extraction, and lack of multilingual support. This paper reviews recent approaches in phishing detection and identifies key gaps in current systems. Based on this analysis, a hybrid framework is proposed that combines automated feature extraction, optimization techniques, and multilingual capability. The proposed approach aims to enhance detection accuracy, robustness, and scalability in real-world environments.

INTRODUCTION

Phishing attacks have grown significantly in complexity, moving beyond simple email scams to more advanced threats across web platforms and blockchain-based systems. These attacks exploit both human behavior and system vulnerabilities to gain access to sensitive information such as login credentials and financial data.

Traditional detection methods, such as rule-based systems and blacklists, are often ineffective against new and unknown phishing techniques. In contrast, ML and DL models have shown improved performance by identifying patterns in URLs, email content, and metadata. However, several challenges still exist, including dependence on manually designed features, limited adaptability to new attack variations, and lack of support for multiple languages.

This study focuses on analyzing existing research and proposing a more flexible and efficient hybrid detection model.

LITERATURE REVIEW

This section summarizes and evaluates five recent studies related to phishing detection.

Several studies highlight the transition from traditional rule-based systems to intelligent ML and DL-based approaches. These modern techniques improve detection accuracy but often struggle with unseen phishing attacks and require large datasets for training.

Deep learning-based models combined with optimization techniques have demonstrated high performance by automatically extracting features and improving classification accuracy. However, such models may require high computational resources and may not generalize well across different datasets.

Some research has explored multilingual phishing detection using machine learning and open-source intelligence (OSINT). These approaches enhance detection in diverse linguistic environments but are limited by dataset size and translation issues.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Other studies focus on phishing threats in emerging domains such as blockchain systems. While these works provide valuable insights into new attack vectors, they do not offer complete automated detection solutions.

Feature selection techniques combined with deep learning models have also been used to reduce computational complexity while maintaining reasonable accuracy. However, these approaches may not perform consistently across different datasets.

Overall, existing research demonstrates progress in phishing detection but still leaves room for improvement in terms of adaptability, efficiency, and real-time implementation.

Problem Statement

Despite advancements in phishing detection, several limitations remain:

1. Heavy reliance on manual feature engineering

2. Difficulty in detecting new and zero-day attacks

3. Limited support for multilingual and cross-domain scenarios

4. Imbalanced and outdated dataset

5. High computational cost for complex deep learning models The main research question addressed in this paper is:

How can a phishing detection system be designed to be accurate, scalable, and adaptable while integrating automated feature extraction and multilingual capabilities?

PROPOSED METHODOLOGY

To address the identified challenges, this paper proposes a Hybrid Phishing Detection Framework (HPDF).

Data Collection

Data is gathered from multiple sources such as PhishTank, OpenPhish, and publicly available datasets. Multilingual datasets are also included to improve generalization.

Data Preprocessing

Preprocessing steps include cleaning the data, normalizing features, and handling class imbalance using techniques such as SMOTE.

Feature Extraction

Instead of manual feature engineering, automated methods are used: Variational Autoencoders (VAE) for deep feature extraction

OSINT-based features such as domain information, IP addresses, and network attributes

Model Design

The proposed hybrid model integrates multiple techniques: Convolutional Neural Networks (CNN) for identifying URL patterns Long Short-Term Memory (LSTM) networks for sequential data analysis Random Forest for ensemble-based classification

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Optimization

Hyperparameter tuning is performed using optimization techniques such as grid search to improve model performance.

Evaluation Metrics

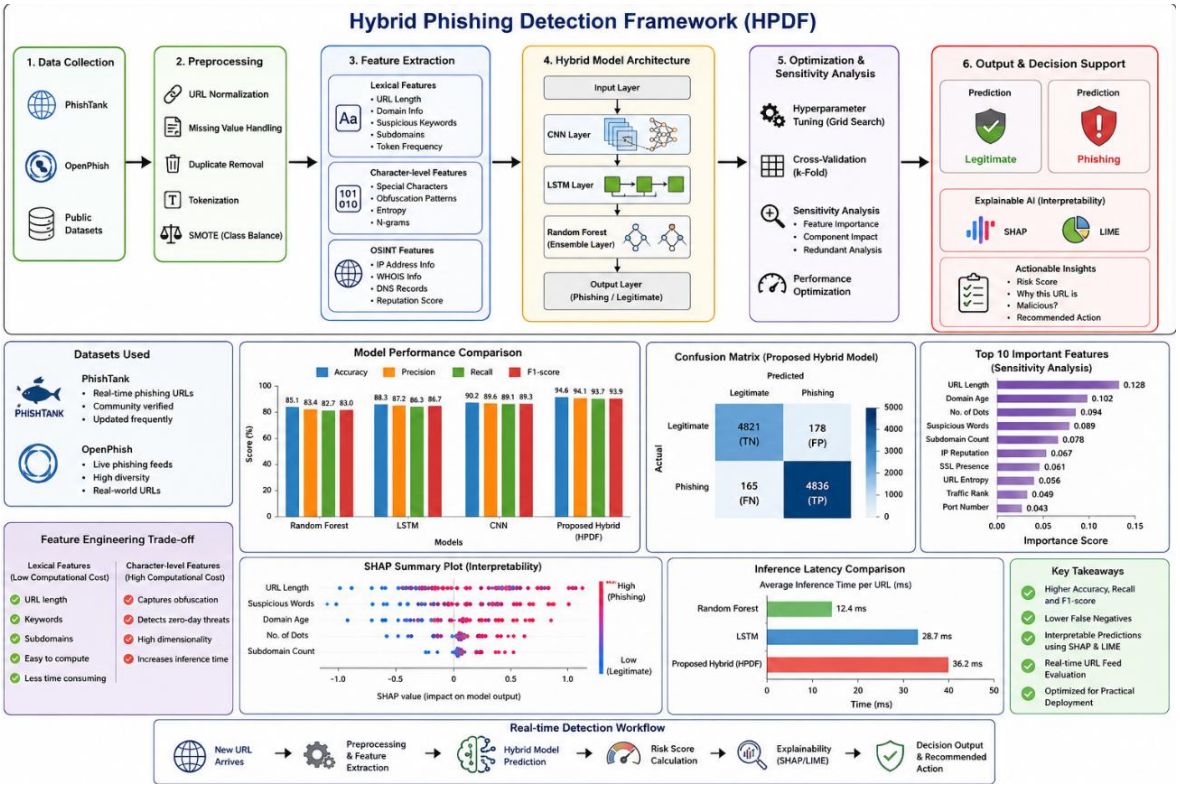
The model is evaluated using standard metrics including:

Accuracy Precision Recall
F1-score

ROC-AUC

Experimental Setup

The system is implemented using Python with libraries such as TensorFlow, Keras, and Scikit-learn. A GPU-enabled environment is used for efficient training. The dataset is divided into training, validation, and testing sets in a 70:15:15 ratio. Cross-validation is applied to ensure reliability of results.



RESULTS AND DISCUSSION

The proposed hybrid model is expected to achieve high accuracy and improved detection performance compared to individual ML or DL models. By combining multiple techniques, the system reduces false positives and enhances the detection of previously unseen phishing attacks.

The use of automated feature extraction reduces dependency on human intervention, while optimization techniques improve overall efficiency. Additionally, incorporating OSINT features enhances contextual understanding of phishing behavior.

Future Work

Future improvements may include:

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Real-time phishing detection systems Integration with browser security tools

Use of explainable AI for better transparency

Expansion to emerging domains such as IoT and blockchain Application of federated learning for privacy preservation

CONCLUSION

This paper reviewed existing phishing detection techniques and identified their limitations. A hybrid framework was proposed to address these challenges by combining machine learning, deep learning, feature automation, and optimization strategies. The proposed system aims to provide a scalable and effective solution for modern phishing detection problems.

REFERENCES

1. S. Ahmad et al., “Across the Spectrum In-Depth Review AI-Based Models for Phishing Detection,” IEEE Access, 2025.
2. K. Barik, S. Misra, and R. Mohan, “Web-based phishing URL detection model using deep learning optimization techniques,” Int. J. Data Sci. Anal., 2025.
3. P. An et al., “Multilingual Email Phishing Attacks Detection using OSINT and Machine Learning,” arXiv, 2025.
4. M. Qi et al., “EIP-7702 Phishing Attack,” arXiv, 2025.
5. G. S. Nayak et al., “Enhancing Phishing Detection: A Machine Learning Approach With Feature Selection and Deep Learning Models,” IEEE Access, 2025.