International Journal of Engineering and Techniques - Volume 12 Issue 3, May-June - 2026

RESEARCH ARTICLE OPEN ACCESS

DEVELOPMENT OF AN INTELLIGENT SYSTEM FOR DETECTING PHISHING EMAILS IN WEB ENVIRONMENTS

Nawfal Raad Mahmood[1], Alaa Jabbar Qasim Almaliki[2]
1,2(School of Computing, College of Arts and Sciences, Universiti Utara Malaysia, 06010, Sintok, Kedah, Malaysia
Email: nawfal1986rm@gmail.com, alaa.jabbar@uum.edu.my)

Abstract:

Phishing attacks by email continue to be one of the biggest cyber security threats facing people in the internet age. As phishing attacks become more advanced and innovative, there are more and more examples of AI being used to generate realistic-looking phishing emails. This work proposes and tests a framework for phishing email detection by combining hybrid deep learning and explainable AI (XAI). A large-scale, multi-source data set was built, consisting of 106,135 emails from three open corpora: Enron Email Dataset, Phishing Email Dataset, and the Education-Targeted Phishing Email Dataset. Various classification architectures were implemented and tested under the same conditions, including traditional machine learning models (SVM, Logistic Regression, Random Forest), deep learning models (CNN, LSTM, GRU, RNN, BiLSTM), transformer-based models (BERT, RoBERTa), dual hybrid models, and multi-model hybrid models. The GRU-RoBERTa hybrid model obtained the best results, with 97.81% accuracy, 97.34% precision, 98.27% recall and 97.80% F1-Score, better than all standalone models. To improve interpretability and uncover the most important phishing signals, LIME and SHAP explainability techniques were incorporated. A working prototype was implemented on the web and shown to run on standard CPU computers. The results validate the potential of a hybrid architecture comprising transformer-based contextual learning with sequential neural networks for effective and interpretable phishing detection in real-world cybersecurity applications.

Keywords — **BERT, Cybersecurity, Deep Learning, Explainable Artificial Intelligence, RoBERTa.**

I. INTRODUCTION

Phishing email attacks remain a significant security threat in today's web-based environment due to their exploitation of human trust, realistic email appearances, and resistance to traditional security measures. The widespread adoption of online communication platforms, cloud-based services, and web applications has led to increased exposure to phishing attacks for both individuals and organizations, resulting in substantial financial losses, operational disruptions, and reputational damage. [1], [2] In recent years, phishing attacks have evolved from easily identifiable emails characterized by poor grammar, spelling errors, and suspicious links to highly sophisticated campaigns that generate contextually relevant and semantically coherent content. The rapid advancement of Artificial Intelligence (AI), Natural Language Processing (NLP), and Large Language Models (LLMs) has further intensified this challenge by enabling attackers to automatically produce phishing emails that convincingly mimic genuine human communication. [3] Experimental results demonstrate that emails generated by advanced language models such as GPT-4 can successfully bypass traditional spam filters at a high rate, exposing critical vulnerabilities in current detection systems. [4].

Current phishing detection methods primarily involve blacklist filtering, rule-based approaches, and handcrafted feature extraction techniques. While these methods have proven effective against earlier phishing attacks, recent studies indicate that

traditional detection techniques lack adaptability, contextual understanding, and robustness against dynamically generated phishing content. [1], [5] To address these limitations, deep learning and transformer models such as BERT and RoBERTa have demonstrated significant success by leveraging bidirectional attention and contextual semantic learning, substantially outperforming traditional methods. [6] The integration of transformer-based contextual learning with deep learning mechanisms has further enhanced detection capabilities and robustness. [5], [7] Despite these advances, several research gaps remain. Prior studies often rely on datasets with limited diversity and few sources, evaluate models independently without standardised benchmarking, and lack emphasis on interpretability. [1], [8]. The absence of interpretable mechanisms, such as LIME and SHAP, reduces transparency and limits practical applicability in real-world cybersecurity environments.

A. Problem Statement

Despite their high success rate in evading traditional security protocols and becoming more sophisticated, phishing emails continue to be a significant cybersecurity risk. Current detection methods are mostly based on machine learning algorithms or handcrafted features that have limited understanding of context, poor generalization, and lack of adaptability to AI-generated phishing content. Moreover, the explainability of the model is not explored in most previous studies, which results in low user trust and limited deployment in security-sensitive applications. [8].

B. Research Objectives

The goals of this study are:

- to implement and benchmark traditional machine learning models on a unified large-scale multi-source phishing email dataset;
- to implement and compare the performance of standalone deep learning models (CNN, LSTM, GRU, BiLSTM, RNN) with transformer-based models (BERT, RoBERTa) under unified experimental settings;
- to design and evaluate dual hybrid models that combine transformer-based contextual learning with deep learning mechanisms;
- to evaluate multi-model hybrid models to further enhance phishing detection capability;
- to select the best-performing model for the phishing detection task based on accuracy, precision, recall and F1-score;
- to integrate LIME and SHAP explainability techniques into the best-performing model to make decisions more transparent in modern web contexts.

II. RELATED WORK

Great strides have been made in phishing email detection research in the past ten years, developing from statistical machine learning techniques to deep learning approaches, transformer-based techniques, and intelligent hybrid models. Table I gives an overview of the literature.

TABLE I
Summary of Related Phishing Email Detection Studies

Reference	Method Model	Key Findings	Limitation
Rathee & Mann (2022)	ML Comparative (SVM, RF, LR, NB)	Acceptable performance with low complexity.	Limited contextual understanding; no deep learning.
Atawneh & Aljehani (2023)	BERT+LSTM hybrid	99.61% accuracy; outperformed standalone models.	Limited XAI and multi-model hybrid evaluation.
Altwaijry et al. (2024)	Deep Learning Models	Strong detection performance across architectures.	Limited to DL; no transformer comparison.
Li et al. (2024)	BERT Transformer Model	Improved semantic detection; fewer false positives.	Large dataset and GPU dependency; no hybrid.
Jamal et al. (2024)	DistilBERT, RoBERTa (IPSDM)	Up to 99% accuracy with reduced overfitting.	English only; no hybrid or XAI.
Gupta et al. (2024)	BERT+CNN Hybrid	Improved enterprise phishing detection accuracy.	Large dataset; enterprise-specific scope.
Alhuzali et al. (2025)	ML vs DL vs Transformers (10 datasets)	BERT and RoBERTa outperformed traditional ML.	Experimental inconsistency; no XAI.
Hosseinzadeh et al. (2025)	BERT+CNN+GRU+MGO	96.8% accuracy; improved robustness.	High overhead; limited explainability.
Lim et al. (2025)	LogReg+LIME+SHAP+DeepSeek v3 LLM	98.4% accuracy; 94.2% explanation accuracy.	LLM API dependency; added latency.
Uddin et al.	RoBERTa+LI	98.45% accuracy;	Single-

(2026)	ME+LITA	improved interpretability.	architecture focus.

A. Traditional Machine Learning Approaches

Phishing detection has been extensively studied using traditional machine learning techniques, which are relatively simple and demonstrate good computational efficiency. [2] Handcrafted textual and structural features have been employed to achieve acceptable performance with algorithms such as SVM, RF, LR and Naïve Bayes[1]. However, these models heavily depend on manual feature engineering and struggle to capture complex contextual relationships present in emails. Furthermore, traditional methods have exhibited poor generalization across different phishing datasets, which vary in writing style and attack techniques. [1].

B. Deep Learning Architectures

Deep learning has been used to automatically extract semantic features from raw email content in order to classify phishing emails. CNN, LSTM, GRU, and BiLSTM networks constitute the foundation of contemporary methods. By identifying contextual dependencies and local characteristics in sequential data, these models can automatically learn features from text. Recent methods in particular have attained excellent accuracy, such as 99.61%. [9] using BERT-LSTM, and strong detection performance. [10] through comparative evaluation of multiple architectures. Nevertheless, deep learning methods need a lot of training data and are vulnerable to overfitting when the dataset is small or highly unbalanced.

C. Transformer-Based Models

Self-attention mechanisms that model entire text sequences simultaneously have significantly enhanced transformer-based architectures in the phishing email detection task. [6]. BERT integrates bidirectional contextual learning, enabling it to represent semantic relationships between both preceding and succeeding contexts. RoBERTa further optimizes classification accuracy by employing improved pretraining methods. Fine-tuned DistilBERT models have demonstrated 99% accuracy without overfitting, as reported. [11] Despite these advantages, transformer architectures require substantial computational resources and involve complex fine-tuning procedures.

D. Hybrid Intelligent Architectures

Multiple learning mechanisms have been combined to develop hybrid intelligent architectures that achieve superior performance in the phishing detection task[2]. In enterprise phishing detection. [7] proposed a BERT-CNN hybrid model that integrates transformer-based contextual embeddings with convolutional feature extraction. [9] introduced a hybrid BERT-LSTM model achieving an impressive accuracy of 99.61%, outperforming the previously mentioned models. Hosseinzadeh et al. (2025)[5] demonstrated that multi-model hybrids incorporating BERT, CNN, GRU, and Moth-Flame Optimization performed well, achieving an accuracy of 96.8% and exhibiting enhanced robustness.

E. Explainable Artificial Intelligence for Phishing Detection

As phishing detection models become increasingly complex, there has been growing interest in explainable AI (XAI) techniques, which aim to enhance transparency and interpretability[3]. LIME generates local explanations by simplifying models around specific prediction instances, enabling analysts to identify influential textual features within individual email examples. LIME was integrated with a transformer-based phishing detection model, achieving 98.45% accuracy while improving interpretability and explanation quality. [8], [12] introduced the EXPLICATE framework, which integrates LIME, SHAP, and the DeepSeek v3 large language model, attaining 98.4% phishing detection accuracy and 94.2% explanation accuracy, utilizing both a GUI application and a Chrome extension.

F. Research Gaps

Despite significant advancements, several research gaps persist:

- The absence of unified comparative frameworks that simultaneously evaluate machine learning, deep learning, transformer, and hybrid architectures under consistent experimental conditions.
- Limited utilization of large-scale multi-source phishing datasets.

- Insufficient investigation of multi-model hybrid architectures combining three or more learning mechanisms.
- Limited XAI integration within hybrid phishing detection frameworks.
- Insufficient attention to AI-generated phishing content detection.

This study addresses all five gaps through a systematic unified comparative investigation.

III. PROPOSED METHOD

A. Research Design

In this research, a quantitative experimental research design is used for systematic investigation and comparison of several intelligent classification architectures for phishing email detection. The experimental procedure is as follows: constructing datasets, pre-processing the datasets, implementing models, comparing models, creating hybrid models, and analysing the explainability of the model. The research design and experimental process used in this study are shown in Figure 1.

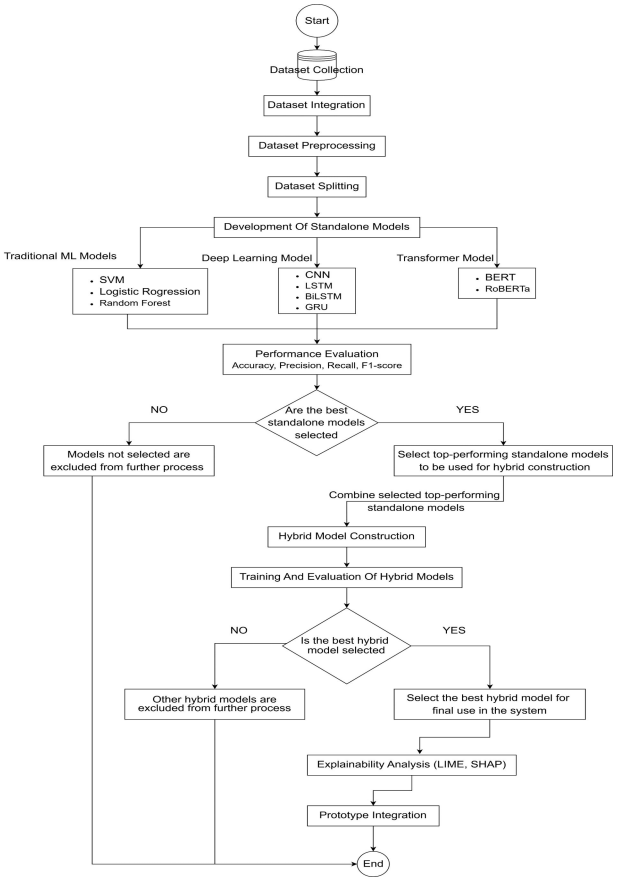

Fig. 1 Research design of the proposed intelligent phishing email detection framework

B. Dataset Collection and Construction

A large-scale multi-source phishing email corpus was built by systematically integrating three publicly available email corpora. Table II shows the main datasets used, and Figure 2 shows the construction process of the datasets.

TABLE II
Summary of Datasets Used in This Study

Dataset	Source	Raw Size	After Preproc.	Purpose
Enron Email Dataset	Kaggle (wcukierski)	517,401	240,039	Legitimate emails
Phishing Email Dataset	Kaggle (naserabdullahalam)	52,542	44,954	Phishing emails (5 sub-corpora)
Education-Targeted Phishing	Kaggle (tanvirahmed0981)	16,932	16,234	Domain-specific phishing

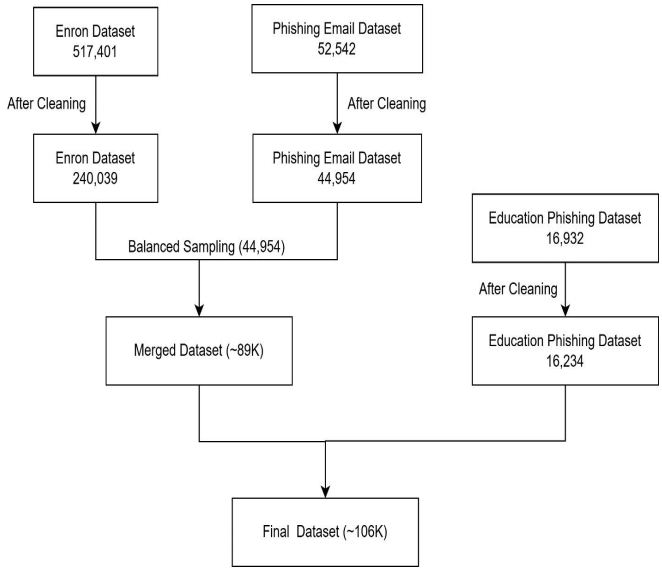

Fig. 2 Dataset construction workflow

The final consolidated corpus after integration, deduplication and preprocessing consisted of 106,135 email records — with 53,686 (50.6%) legitimate and 52,449 (49.4%) phishing — and the class distribution was almost balanced after the process. The dataset was stratified and split into 80% as the training set (84,908 samples), 10% as the validation set (10,613 samples) and 10% as the test set (10,614 samples).

C. Data Preprocessing Strategy

A unified preprocessing pipeline was applied across all collected datasets. The preprocessing process included:

- duplicate removal and missing value handling;

- text cleaning, removal of HTML tags, special characters, URLs, and irrelevant metadata;
- normalization, lowercase conversion, whitespace standardisation, and lemmatization;
- tokenization suitable for different intelligent architectures and dataset balancing procedures.

For traditional machine learning models, TF-IDF vectorization was applied. For deep learning architectures, sequence tokenization with a maximum length of 200 tokens and embedding dimension of 128 was used. For transformer-based architectures, pretrained tokenizers with a maximum sequence length of 128 tokens were employed.

D. Proposed Intelligent Classification Models

Traditional Machine Learning Models: Three baseline classifiers were evaluated using TF-IDF vectorized representations: Logistic Regression (LR), Support Vector Machine (SVM) with a linear kernel, and Random Forest (RF) as an ensemble-based approach. All models were implemented using Scikit-learn with a fixed random seed of 42.

Standalone Deep Learning Models: Five deep learning architectures were implemented using TensorFlow and Keras: CNN for local textual feature extraction, LSTM and GRU for sequential contextual dependency learning, BiLSTM for bidirectional contextual analysis, and RNN as a baseline recurrent architecture. All models used tokenized sequences with a maximum length of 200 tokens, an embedding dimension of 128, trained over 5 epochs with a batch size of 32 using the Adam optimiser.

Transformer-Based Models: BERT and RoBERTa were fine-tuned using pretrained contextual embeddings with a maximum sequence length of 128 tokens, trained over 2 epochs with a batch size of 8 using the AdamW optimiser with a learning rate of 2e-5.

Hybrid Intelligent Architectures: Based on individual model performance, five dual hybrid architectures were constructed: CNN+GRU, CNN+BERT, CNN+RoBERTa, GRU+BERT, and GRU+RoBERTa. Two triple hybrid architectures were subsequently investigated: CNN+GRU+BERT and CNN+GRU+RoBERTa. Transformer-based hybrid models were trained using AdamW (learning rate = 2e-5) for 3 epochs with a batch size of 16 and a maximum sequence length of 128 tokens.

Explainability Analysis: LIME and SHAP techniques were applied to the best-performing model to improve interpretability. LIME was configured to generate local explanations using 500 neighbourhood samples and 15 features per explanation. SHAP analysis was conducted across 20 test emails using a text masker derived from the RoBERTa tokenizer to identify the most influential textual features contributing to phishing classification decisions.

E. Experimental Setup

All experiments were conducted in Google Colab using Python with TensorFlow, PyTorch, Hugging Face Transformers, and Scikit-learn libraries. Traditional ML and preliminary DL experiments used CPU-based execution, while transformer-based and hybrid architectures utilized Tesla T4 GPU acceleration. Table III summarises the training configurations applied across all implemented architectures.

TABLE III
Training Configuration Summary Across All Implemented Architectures

Architecture Group	Models	Optimiser	Epochs	Batch	Max Seq.
Traditional ML	LR, SVM, RF	Solver-based	N/A	N/A	TF-IDF
Deep Learning	CNN, RNN, LSTM, GRU, BiLSTM	Adam	5	32	200
Deep Hybrid	CNN+GRU	Adam	5	32	200
Transformer-Based	BERT, RoBERTa	AdamW (2e-5)	2	8	128
Transformer Hybrid (Dual)	CNN+BERT, CNN+RoBERTa, GRU+BERT, GRU+RoBERTa	AdamW (2e-5)	3	16	128
Transformer Hybrid (Triple)	CNN+GRU+BERT, CNN+GRU+RoBERTa	AdamW (2e-5)	3	16	128

IV. RESULT AND ANALYSIS

A. Performance Evaluation Metrics

The performance of all implemented models was evaluated using four standardised metrics: Accuracy (overall correctly classified samples), Precision (proportion of correctly predicted phishing emails among all predicted phishing),

Recall (proportion of actual phishing emails correctly identified), and F1-Score (harmonic mean of Precision and Recall). Confusion matrix analysis was additionally employed to provide detailed visualisation of classification behaviour across all architectures.

B. Traditional Machine Learning Model Results

The performance evaluation results of the traditional machine learning models are presented in Table IV. Among the three classifiers, SVM demonstrated the strongest overall performance, attributed to its effectiveness in handling high-dimensional TF-IDF feature spaces. Despite acceptable performance, all three models demonstrated limitations in capturing deep contextual relationships and complex phishing communication structures.

TABLE IV
Performance Evaluation Results of Traditional Machine Learning Models

Model	Accuracy	Precision	Recall	F1-Score
Support Vector Machine (SVM)	94.91%	92.97%	97.04%	94.96%
Logistic Regression (LR)	94.45%	92.33%	96.81%	94.52%
Random Forest (RF)	93.78%	92.73%	94.85%	93.78%

C. Standalone Deep Learning Model Results

The performance of standalone deep learning architectures is presented in Table V. CNN demonstrated the strongest overall deep learning performance through effective local textual pattern extraction. GRU demonstrated computationally efficient sequential contextual learning. LSTM and BiLSTM also showed strong contextual learning capabilities. The conventional RNN architecture demonstrated significantly lower classification performance due to vanishing gradient limitations and weaker long-term dependency learning.

TABLE V
Performance Evaluation Results of Standalone Deep Learning Models

Model	Accuracy	Precision	Recall	F1-Score
CNN	97.22%	96.02%	98.46%	97.22%
LSTM	96.80%	96.71%	96.82%	96.76%
BiLSTM	96.69%	96.10%	97.25%	96.67%
GRU	97.09%	96.55%	97.60%	97.07%
RNN	50.60%	100.00%	0.04%	0.08%

D. Transformer-Based Model Results

The performance of transformer-based architectures is presented in Table VI. Both BERT and RoBERTa achieved superior phishing email classification capability compared to traditional machine learning and standalone deep learning models. Both models achieved an exceptional recall of 99.03%, indicating strong capability in correctly identifying actual phishing emails with minimal false negatives.

TABLE VI
Performance Evaluation Results of Transformer-Based Models

Model	Accuracy	Precision	Recall	F1-Score
BERT	97.75%	96.51%	99.03%	97.75%
RoBERTa	97.64%	96.29%	99.03%	97.64%

E. Hybrid Intelligent Model Results

The performance of all hybrid intelligent architectures is presented in Table VII. The GRU+RoBERTa hybrid architecture achieved the strongest overall performance across all evaluation metrics, attributed to the effective integration of RoBERTa's optimised contextual semantic learning with GRU's efficient sequential dependency analysis. Increasing architectural complexity through triple hybrid integration did not consistently surpass the optimised dual hybrid, indicating that efficient complementary integration is more important than architectural depth alone.

TABLE VII
Performance Evaluation Results of All Hybrid Intelligent Architectures

Model	Accuracy	Precision	Recall	F1-Score
CNN + GRU	97.34%	96.94%	97.71%	97.32%
GRU + BERT	97.71%	96.84%	98.59%	97.70%
GRU + RoBERTa (Best)	97.81%	97.34%	98.27%	97.80%
CNN + BERT	97.76%	96.96%	98.55%	97.75%
CNN + RoBERTa	97.69%	96.68%	98.72%	97.69%
CNN + GRU + BERT	97.17%	96.91%	97.39%	97.15%
CNN + GRU + RoBERTa	97.29%	97.16%	97.35%	97.26%

F. Comparative Overall Analysis

The comparative experimental evaluation revealed a clear performance progression across architecture categories. Traditional machine learning models achieved acceptable baseline performance (SVM: 94.91%), but remained limited in contextual understanding. Standalone deep learning architectures demonstrated improved performance (CNN: 97.22%), while transformer-based models further improved classification (BERT: 97.75%). Hybrid architectures consistently outperformed standalone counterparts, with GRU+RoBERTa achieving the strongest overall performance (97.81% accuracy). Figure 5 illustrates

the confusion matrix analysis of the GRU+RoBERTa hybrid architecture.

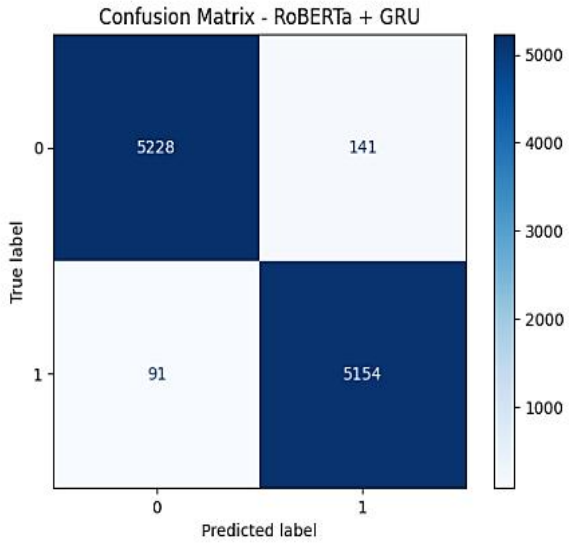

Fig. 5 Confusion matrix of the GRU+RoBERTa hybrid architecture (image by the researcher).

V. CONCLUSION

This study proposed and evaluated an intelligent phishing email detection framework for modern web-based environments using traditional machine learning, deep learning, transformer-based architectures, hybrid intelligent models, and Explainable Artificial Intelligence techniques. A large-scale multi-source phishing email dataset comprising 106,135 email records was constructed through the integration of the Enron Email Dataset, the Phishing Email Dataset, and the Education-Targeted Phishing Email Dataset, thereby improving dataset diversity, contextual variability, and experimental consistency.

The experimental findings demonstrated a clear performance progression across the investigated architecture categories. Traditional machine learning models achieved acceptable baseline performance, with SVM reaching 94.91% accuracy. Standalone deep learning architectures improved performance, with CNN achieving 97.22% accuracy. Transformer-based models further advanced classification through bidirectional semantic representation, with BERT achieving 97.75% accuracy. Hybrid architectures consistently outperformed standalone models, with the GRU+RoBERTa hybrid achieving the strongest overall performance with 97.81% accuracy, 97.34% precision, 98.27% recall, and 97.80% F1-score. Increasing architectural complexity through triple hybrid integration did not consistently guarantee proportional performance improvement beyond optimised dual hybrid architectures.

The integration of LIME and SHAP explainability techniques significantly improved model interpretability, transparency, and trustworthiness. Key phishing-related textual indicators including urgency expressions, credential verification requests, and suspicious action words were identified as the most influential features in phishing classification decisions. The functional web-based prototype confirmed practical deployment capability on standard computing resources without GPU requirements, with response times of approximately 1–2 seconds per email.

A. Practical Implications

Organizations can deploy the GRU-RoBERTa hybrid architecture as an effective phishing email detection system that balances high accuracy with computational efficiency suitable for standard computing environments. Security analysts can utilise the XAI-enhanced system to understand phishing classification decisions through interpretable LIME and SHAP explanations, enabling informed cybersecurity decisions without requiring deep technical knowledge of the underlying model. Researchers can adopt the unified comparative evaluation framework as a reference for systematic evaluation of future intelligent phishing detection architectures.

B. Future Work

Future research may extend the proposed framework toward multimodal phishing detection environments integrating textual, structural, URL-based, and visual phishing analysis mechanisms. Multilingual phishing detection frameworks utilising multilingual transformer architectures such as mBERT and XLM-RoBERTa may improve generalisation across diverse linguistic environments. Lightweight transformer architectures and optimised hybrid deployment strategies may enable real-time enterprise cybersecurity applications with limited computational resources. Furthermore, adaptive continual learning mechanisms may improve model

adaptability against dynamically evolving phishing strategies and AI-generated phishing communication patterns.

REFERENCES

[1] A. Alhuzali, A. Alloqmani, M. Aljabri, and F. Alharbi, “In-Depth Analysis of Phishing Email Detection: Evaluating the Performance of Machine Learning and Deep Learning Models Across Multiple Datasets,” *Appl. Sci.*, vol. 15, no. 6, pp. 1–30, 2025, doi: 10.3390/app15063396.

[2] D. Rathee and S. Mann, “Detection of E-Mail Phishing Attacks – using Machine Learning and Deep Learning,” *Int. J. Comput. Appl.*, vol. 183, no. 47, pp. 1–7, 2022, doi: 10.5120/ijca2022921868.

[3] C. S. Eze and L. Shamir, “Analysis and Prevention of AI-Based Phishing Email Attacks,” *Electron.*, vol. 13, no. 10, 2024, doi: 10.3390/electronics13101839.

[4] C. Opara, P. Modesti, and L. Golightly, “Evaluating spam filters and Stylometric Detection of AI-generated phishing emails,” *Expert Syst. Appl.*, vol. 276, no. February, p. 127044, 2025, doi: 10.1016/j.eswa.2025.127044.

[5] M. Hosseinzadeh *et al.*, “Improving phishing email detection performance through deep learning with adaptive optimization,” *Sci. Rep.*, vol. 15, no. 1, pp. 1–16, 2025, doi: 10.1038/s41598-025-20668-5.

[6] H. Li, J. Yang, Y. Li, and K. Li, “Email phishing attack detection based on BERT transformer model,” vol. 13395, no. Oece, p. 114, 2024, doi: 10.1117/12.3049161.

[7] B. B. Gupta *et al.*, “Advanced BERT and CNN-Based Computational Model for Phishing Detection in Enterprise Systems,” *C. - Comput. Model. Eng. Sci.*, vol. 141, no. 3, pp. 2165–2183, 2024, doi: 10.32604/cmes.2024.056473.

[8] M. A. Uddin, M. Mahiuddin, and I. H. Sarker, “An Explainable Transformer-based Model for Phishing Email Detection: A Large Language Model Approach,” *Comput. Networks*, vol. 277, no. December 2025, p. 112061, 2025, doi: 10.1016/j.comnet.2026.112061.

[9] S. Atawneh and H. Aljehani, “Phishing Email Detection Model Using Deep Learning,” *Electron.*, vol. 12, no. 20, 2023, doi: 10.3390/electronics12204261.

[10] N. Altwaijry, I. Al-Turaiki, R. Alotaibi, and F. Alakeel, “Advancing Phishing Email Detection: A Comparative Study of Deep Learning Models,” *Sensors*, vol. 24, no. 7, pp. 1–19, 2024, doi: 10.3390/s24072077.

[11] S. Jamal, H. Wimmer, and I. H. Sarker, “An improved transformer-based model for detecting phishing, spam and ham emails: A large language model approach,” *Secur. Priv.*, vol. 7, no. 5, 2024, doi: 10.1002/spy2.402.

[12] B. Lim, R. Huerta, A. Sotelo, A. Quintela, and P. Kumar, “EXPLICATE: Enhancing Phishing Detection through Explainable AI and LLM-Powered Interpretability,” 2025, [Online]. Available: http://arxiv.org/abs/2503.20796

[13] R. Din, A. H. Shakir, S. H. Ali, A. J. Qasim Almaliki, and S. Utama, "Exploring Steganographic Techniques for Enhanced Data Protection in Digital Files," International Journal of Advanced Research in Computational Thinking and Data Science, vol. 1, no. 1, pp. 1-9, 04/19 2024, doi: 10.37934/ctds.1.1.19a.

[14] A. J. Almaliki, O. Ghazali, and R. Din, "Advanced Steganography Methods in Modern Cybersecurity," International Journal of Engineering and Techniques (IJET), vol. 12, no. 3, pp. 172-178, 2026/5.

[15] A. J. Qasim Almaliki et al., "Application of the Canny Filter in Digital Steganography," Journal of Advanced Research in Computing and Applications, vol. 35, no. 1, pp. 21-30, 05/17 2024, doi: 10.37934/arca.35.1.2130.