FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 12, June, 2026, pp 196 – 203
DOI: https://doi.org/10.33003/fjs-2026-1012-5654

Phishing Email Detection Using A Hybrid CNN–MLP Deep Learning Framework

Saka K. Kamil, *Yuguda Muhammad and Yusuf Halima Usman

Department of Computer Science, Al-Hikmah University, Ilorin, Kwara State, Nigeria.

*Corresponding authors' email: muhammadyuguda30@gmail.com

ABSTRACT
Phishing incidents remain highly frequent and destructive cyber threats affecting private users and organizations globally. Malicious actors continuously refine their deceptive strategies to evade traditional rule-based filters and conventional machine learning models, leading to financial losses, credential theft, and data breaches. This study presents an enhanced phishing email detection system based on a hybrid deep learning framework integrating a Convolutional Neural Network (CNN) with a Multi-Layer Perceptron (MLP). The CNN component automatically extracts discriminative textual features from email content, while the MLP component performs accurate email classification as legitimate or phishing. A publicly available phishing email dataset from the Kaggle platform was utilized for model training and evaluation. The dataset underwent preprocessing stages including data cleaning, text normalization, tokenization, sequence padding, and label encoding. The CNN layer extracted relevant textual patterns and semantic representations, while the MLP network performed final classification. The developed system was evaluated using accuracy, precision, recall, and F1-score. Experimental results demonstrated that the CNN–MLP model achieved 98.5% accuracy, 98% precision, 99% recall, and 99% F1-score, indicating strong capability in distinguishing phishing from legitimate emails. Additionally, a web-based application was developed to facilitate real-time phishing detection using direct text input and PDF document analysis through Optical Character Recognition (OCR). The findings demonstrate that integrating CNN-based feature extraction with MLP-based classification provides an effective, reliable, and scalable solution for phishing email detection, contributing to improved cybersecurity protection against evolving phishing attacks.

Received: 05th June 2026
Accepted: 02th July 2026
Published: 30th July 2026

Keywords: Convolutional Neural Network (CNN), Cybersecurity, Deep Learning, Email Security, Phishing Detection

INTRODUCTION

Cyber deception through phishing poses a significant security risk in today's digital landscape. Malicious actors impersonate legitimate institutions via deceptive messages and websites to trick victims into surrendering passwords, financial data, and corporate secrets. As cloud computing and online banking expand, the complexity and volume of these threats have surged. Modern attackers utilize customized messaging, URL manipulation, and psychological tactics to evade standard defenses and exploit vulnerable internet users (APWG, 2023; Verizon, 2023).

Conventional defense mechanisms such as blocklisting and signature matching struggle to identify emerging threats because they rely on static rules. While baseline machine learning algorithms such as Random Forest, Support Vector Machines, and Naïve Bayes have enhanced detection rates, they still require extensive manual feature engineering. As a result, these earlier models often fail to map the complex, non-linear patterns embedded in modern phishing data (Jain & Gupta, 2018; Somesha et al., 2020).

Recent progress in deep learning offers highly adaptive tools for network defense. Rather than relying on human-engineered inputs, these networks autonomously identify complex data representations from raw information. Specifically, Convolutional Neural Networks (CNNs) excel at isolating spatial and textual patterns, whereas Multilayer Perceptrons (MLPs) deliver rapid, non-linear classification. Consequently, merging CNN and MLP structures creates a highly resilient framework capable of achieving superior accuracy against shifting attacker methodologies (Adebowale et al., 2020; Ullah et al., 2024).

To address current security gaps, this research introduces an optimized phishing identification system powered by a combined CNN–MLP architecture. Built with Python and TensorFlow, the model was trained on a public Kaggle dataset. It uses a CNN module for automated pattern extraction and an MLP module for final categorization. Furthermore, an interactive web dashboard was created to evaluate raw text and PDF documents via OCR technology.

The primary contribution of this work is the delivery of a scalable, intelligent security tool that accurately isolates fraudulent emails without requiring manual feature selection, offering strong practical utility through an accessible interface.

Review of Literature

The escalating complexity of cyber threats has positioned phishing detection at the forefront of security research. Early defensive frameworks depended heavily on static blocklists and rule-based filters. While sufficient for blocking recognized threats, these manual systems proved highly vulnerable to novel and dynamically generated attacks (Verizon, 2023).

To address these gaps, researchers integrated conventional machine learning algorithms. Studies by Jain and Gupta (2018) and Somesha et al. (2020) demonstrated that models

such as Support Vector Machines, Random Forests, and k-NN significantly outperformed static rules when using engineered security features. Nevertheless, these classifiers still required manual feature extraction, limiting their responsiveness to rapidly mutating phishing tactics.

The advent of deep learning transformed this landscape by automating feature discovery. Ullah et al. (2024) noted that neural networks, particularly CNNs and LSTMs, achieved exceptional accuracy by independently recognizing complex data relationships without human intervention. Building on this, Adebowale et al. (2020) and Ayo and Alowolodu (2024) confirmed that merging multiple neural architectures further enhanced system adaptability against modern threats. However, these complex models often require substantial computational resources, making their deployment in resource-constrained environments difficult.

Furthermore, while CNNs excel at identifying hidden textual and structural markers in URLs (Sensors, 2021), much of the existing research remains focused on either standalone models or URL analysis, often overlooking the need for lightweight models optimized specifically for email inspection.

Consequently, while current literature confirms that combined deep learning architectures outperform isolated models, a clear need remains for fast, resource-efficient solutions. This study directly addresses this gap by introducing a hybrid CNN–MLP framework that pairs deep automated feature extraction with swift classification to deliver a highly accurate and practical email security tool.

Table 1 presents a summary of selected related studies, highlighting their methodologies, major findings, and identified limitations. This comparison provides the basis for the development of the proposed hybrid CNN–MLP framework.

Table 1: Summary of Related Work

Author(s)	Methodology	Findings	Limitations
Jain & Gupta (2018).	Machine Learning Approaches	Improved phishing detection performance	Relied on manual feature extraction
Somesha et al. (2020).	ML-based phishing detection	Achieved good classification accuracy	Limited adaptability to evolving phishing attacks
Adebowale et al. (2020).	Hybrid CNN–LSTM Model	Improved robustness and feature learning	High computational complexity
Ullah et al. (2024).	Deep Learning Survey	CNN achieved high phishing detection accuracy	Focused mainly on URL-based detection
Ayo & Alowolodu (2024).	Hybrid Deep Learning Models	Improved phishing detection capability	Complex architecture and deployment
Developed System	Hybrid CNN–MLP Model	Enhanced phishing email detection accuracy and usability	Implemented in the local environment

MATERIALS AND METHODS

This research used a deep learning framework to build and evaluate an optimized system for identifying phishing emails. By merging a Convolutional Neural Network (CNN) with a Multilayer Perceptron (MLP), the design enhances threat detection through automated pattern discovery and rapid email categorization.

The study relied on a publicly available Kaggle dataset containing clearly labeled fraudulent and safe emails (Alam, 2024; Al-Subaiey et al., 2024). To ensure high data quality, the raw text underwent extensive preparation before model training. This preprocessing phase included removing duplicate entries, normalizing the text, and tokenizing words into numerical values. The sequences were then padded to maintain uniform input lengths, and the labels were encoded. Finally, the refined dataset was split into training and test sets to ensure an objective evaluation.

The core architecture operates in two distinct phases. First, the CNN module automatically isolates critical text-based and structural markers from the emails using its convolutional and pooling layers. These extracted patterns are then flattened into a one-dimensional vector and fed into the MLP module. The MLP utilizes fully connected neural layers to process these features and deliver the final classification.

The entire system was programmed in Python using the TensorFlow and Keras libraries. Model effectiveness was measured against standard security metrics: accuracy, precision, recall, and F1-score. Furthermore, a Flask-based web dashboard was created for practical deployment, allowing users to verify suspicious emails via raw text or PDF uploads. An Optical Character Recognition (OCR) tool was also incorporated to extract and analyze text from image-based documents seamlessly.

The overall architecture of the developed CNN–MLP phishing detection framework is illustrated in Figure 1. The architecture shows the complete processing flow from email input through preprocessing, CNN-based feature extraction, MLP classification, and final prediction.

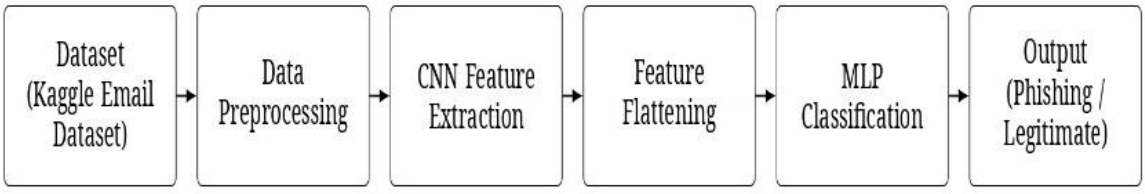

Figure 1: Developed CNN–MLP Phishing Detection Architecture

Figure 2 presents the algorithmic workflow followed by the developed phishing detection system, illustrating the sequential processes from dataset preparation to final email classification and performance evaluation.

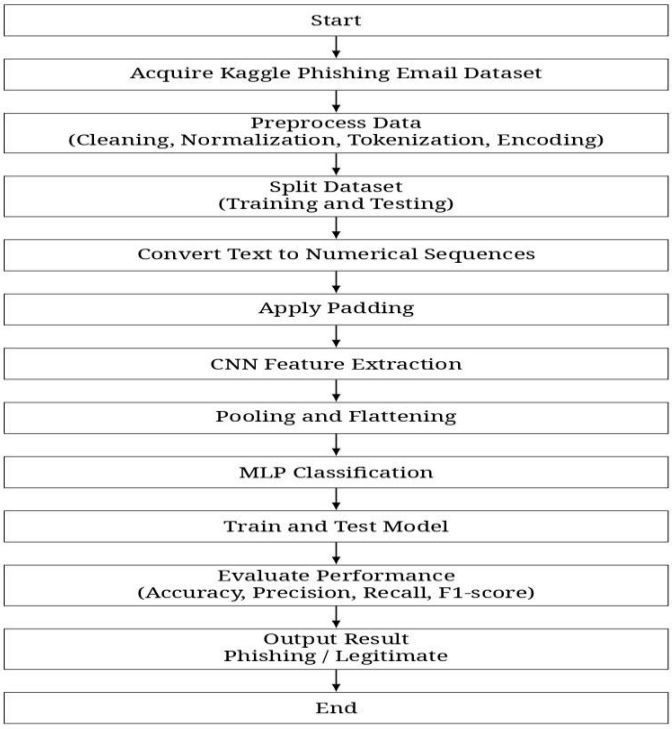

Figure 2: Algorithmic Workflow of the Developed System

RESULTS AND DISCUSSION

This section presents the experimental outcomes of the developed CNN–MLP email security framework. System testing utilized verified fraudulent and safe emails from the Kaggle repository to measure true detection accuracy. The following subsections analyze the model's performance across feature extraction, text classification, and overall evaluation metrics.

System Interface

A Flask-powered web application was built to provide users with access to the security model. This dashboard enables real-time analysis via direct text pasting or PDF uploads. By incorporating OCR technology, the system seamlessly extracts and analyzes text from scanned images or complex PDFs. Once processed, the platform outputs the final verdict alongside predictive confidence percentages, ensuring high practical usability within an intuitive environment.

The main graphical user interface developed for phishing email analysis is shown in Figure 3. The interface enables users to submit email content for real-time phishing detection.

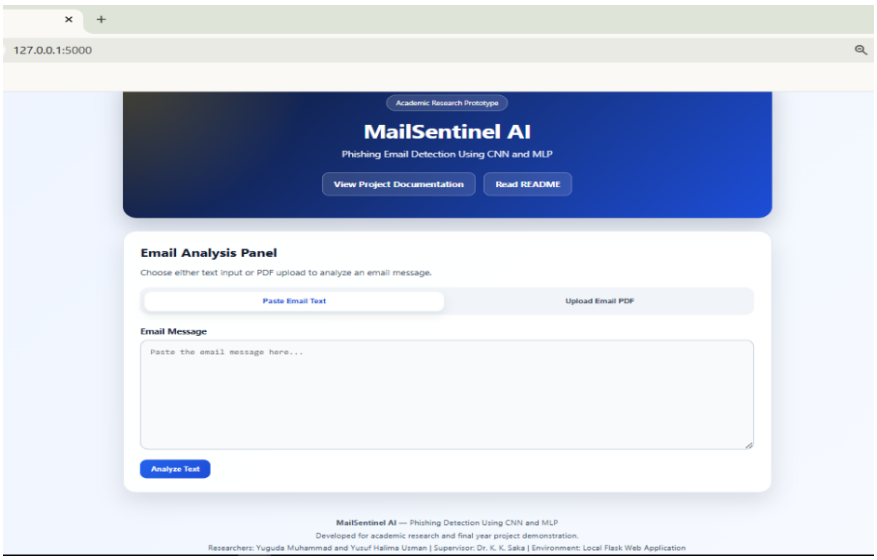

Figure 3: Main System Interface

Figure 4 illustrates the PDF upload interface integrated with Optical Character Recognition (OCR), enabling the system to extract and analyze text from uploaded PDF documents.

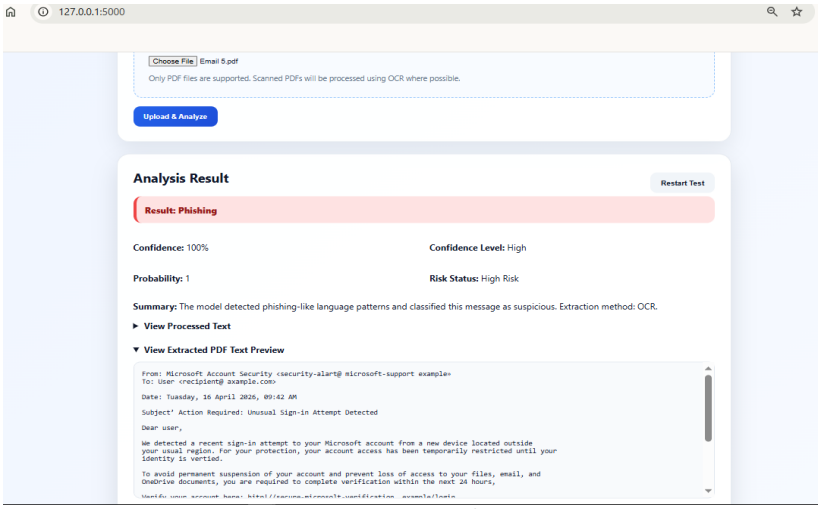

Figure 4: PDF Upload and OCR Extraction Interface

CNN Feature Extraction Result

The Convolutional Neural Network (CNN) component of the developed system automatically extracted phishing-related textual patterns from email content. During training, the CNN successfully learned structural and semantic representations associated with phishing and legitimate emails.

The feature extraction performance of the CNN component during training and testing is summarized in Table 2. The reported processing times and feature representation demonstrate the efficiency of the CNN in automatically extracting meaningful phishing-related features

Table 2: CNN Feature Extraction Processing Result

Parameter	Result
Training Feature Extraction Time	38.6 seconds
Testing Feature Extraction Time	9.4 seconds
Average Extraction Time per Sample	0.12 seconds
Feature Representation Type	Numerical Text Sequence
CNN Output Feature Dimension	128

The results indicate that the CNN component effectively extracted relevant phishing-related features while maintaining stable convergence during training. The low validation loss and high validation accuracy demonstrate the robustness of the feature extraction process.

MLP Classification Result

The CNN-generated feature vectors were forwarded to the Multilayer Perceptron (MLP) classifier for phishing classification. The MLP successfully classified phishing and legitimate emails with high prediction accuracy and reduced misclassification rates.

The classification performance achieved by the MLP classifier using the extracted CNN feature vectors is presented in Table 3A. The evaluation metrics demonstrate the effectiveness of the developed classifier in distinguishing phishing emails from legitimate emails.

Table 3: MLP Classification Result

Class	Precision	Recall	F1-score
Phishing	98%	99%	98.5%
Legitimate	99%	98%	98.5%

The training and validation performance of the developed CNN–MLP model across four training epochs is presented in Table 3B. The results indicate continuous improvement in learning performance accompanied by a steady reduction in training and validation loss.

Table 4: CNN–MLP Training and Validation Performance

Epoch	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
1	96.7%	97.9%	0.082	0.061
2	99.0%	98.4%	0.021	0.048
3	99.4%	98.5%	0.012	0.041
4	99.6%	98.5%	0.008	0.031

The training and validation accuracy curves obtained during CNN model training are presented in Figure 5. The graph demonstrates progressive improvement in classification accuracy across successive training epochs.

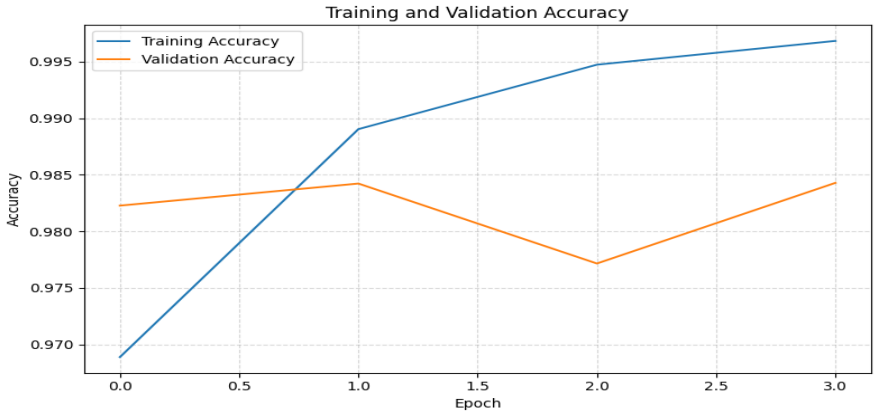

Figure 5: CNN Training and Validation Accuracy Graph

Figure 6 presents the training and validation loss curves of the developed CNN model. The decreasing loss values indicate stable convergence of the learning process.

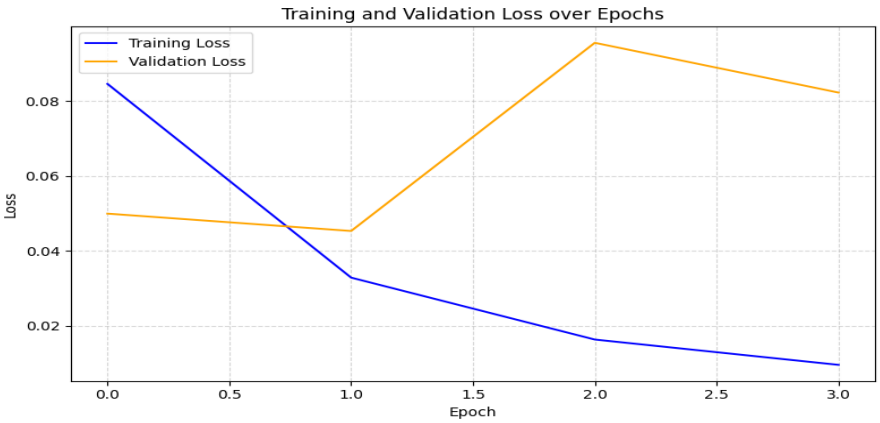

Figure 6: CNN Training and Validation Loss Graph

The confusion matrix of the developed CNN–MLP classifier is illustrated in Figure 7. It summarizes the classification outcomes by showing correctly and incorrectly classified phishing and legitimate email samples.

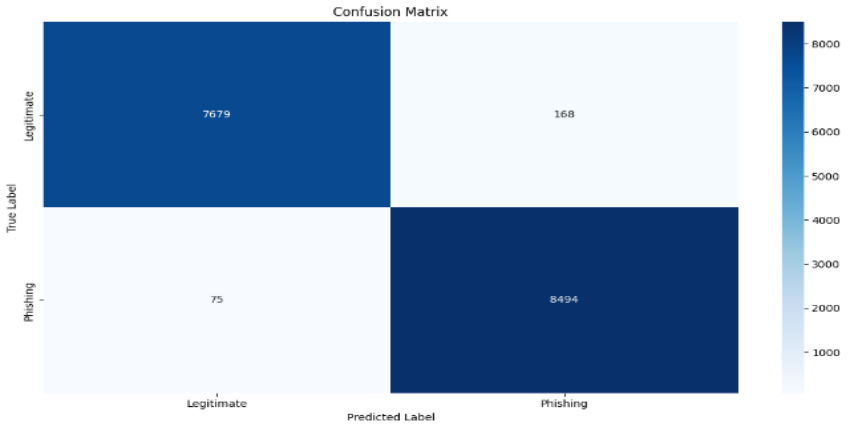

Figure 7: Confusion Matrix of CNN–MLP Classification

Figure 8 presents a sample prediction generated by the developed system for a legitimate email message. The interface displays the predicted class together with the corresponding confidence level and analysis summary.

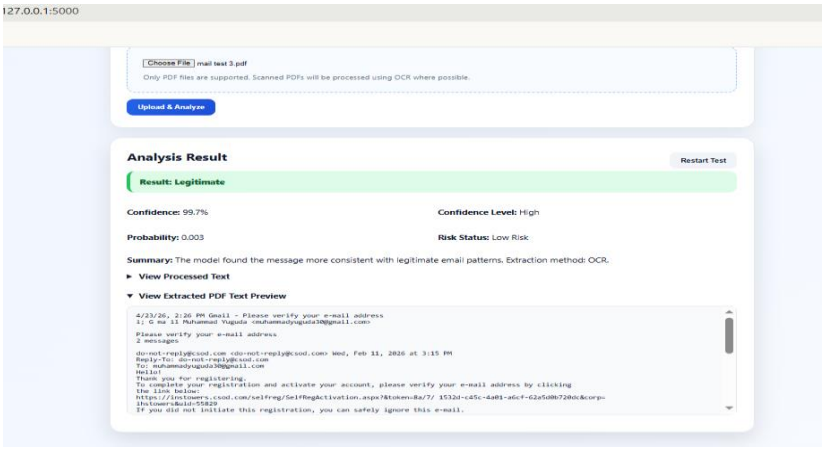

Figure 8: Legitimate Email Detection Result

Figure 9 illustrates the detection result of a phishing email using the developed system. The prediction output confirms the capability of the developed model to identify malicious email content accurately.

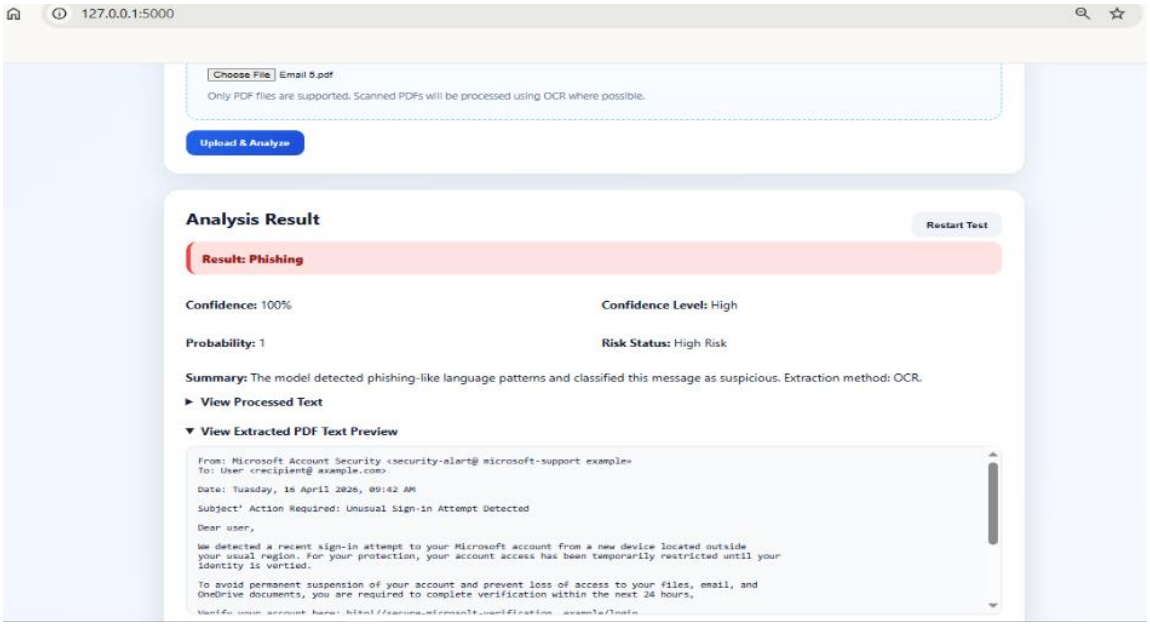

Figure 9: Phishing Email Detection Result

The confusion matrix indicates that the Developed CNN–MLP model achieved low false-positive and false-negative rates. The classification performance confirms that the MLP classifier effectively leverages the extracted CNN features to distinguish phishing emails from legitimate emails.

Performance Evaluation

The overall performance evaluation of the developed CNN–MLP model is summarized in Table 4 using standard evaluation metrics, including accuracy, precision, recall, and F1-score.

Table 4: Performance Evaluation Metrics of the Developed CNN–MLP Model

Metric	Result
Accuracy	98.4%
Precision	98.0%
Recall	99.0%
F1-score	98.5%

The overall performance comparison of the developed CNN–MLP model based on the evaluation metrics is illustrated in Figure 10. The chart summarizes the model's accuracy, precision, recall, and F1-score.

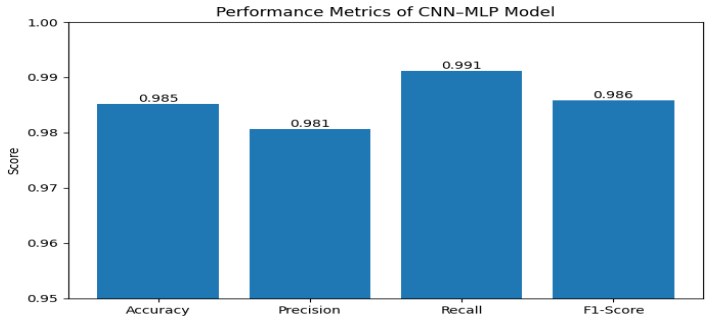

Figure 10: Performance Evaluation Comparison Chart

The hybrid framework delivered outstanding results across every category. Notably, the superior recall score highlights the system's reliability in detecting actual threats, while the strong precision score confirms a low false alarm rate.

Comparison with Existing Studies

To validate its effectiveness, the developed system architecture was benchmarked against prominent deep learning models from recent literature, focusing strictly on detection accuracy and structural complexity.

Table 5 compares the performance of the developed CNN–MLP model with selected existing phishing detection studies reported in the literature.

Table 5: Comparison with Existing Related Studies

Study	Model	Accuracy
Adebowale et al. (2020)	CNN–LSTM	93.28%
Somesha et al. (2021).	LSTM	99.57%
Ullah et al. (2024).	CNN-based Model	98.74%
This Study	CNN–MLP	98.4%

As indicated above, the CNN–MLP framework achieved highly competitive results. More importantly, while alternative models often require substantial computational resources, this developed system delivers elite threat detection through a streamlined, resource-efficient design, making it perfectly suited for rapid email screening.

CONCLUSION

This research introduced an optimized email security framework utilizing a combined Convolutional Neural Network (CNN) and Multilayer Perceptron (MLP) architecture. The project was designed to counter the rising complexity of cyber threats and overcome the limitations of conventional detection algorithms. Findings confirm that pairing automated CNN pattern discovery with swift MLP classification significantly boosts overall threat detection capabilities.

Utilizing a public Kaggle dataset for training and validation, the hybrid model delivered exceptional performance across all core metrics, including accuracy, precision, recall, and F1-score. The CNN layer successfully isolated malicious textual markers, allowing the MLP module to accurately separate fraudulent messages from safe communications with minimal false alarms.

Furthermore, the project delivered a practical web dashboard capable of analyzing both raw text and PDF documents via OCR technology. This interactive interface ensures the system remains highly accessible and applicable for real-world digital environments. Overall, the findings of this study validate that integrating multiple deep learning techniques can provide a scalable and dependable defense mechanism against contemporary phishing attacks.

REFERENCES

Adebowale, M. A., Lwin, K. T., Sanchez, E., & Hossain, M. A. (2020). Intelligent phishing detection scheme using deep learning algorithms. *Applied Computing and Informatics*, *16*(1-2), 1-15. https://doi.org/10.1108/jeim-01-2020-0036

Alam, N. A. (2024). *Phishing Email Dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/naserabdullahalam/phishing-email-dataset

Al-Subaiey, A., Al-Thani, M., Alam, N. A., Antora, K. F., Khandakar, A., & Zaman, S. A. U. (2024). Novel interpretable and robust web-based AI platform for phishing email detection. *arXiv*, arXiv:2405.11619. https://doi.org/10.48550/arXiv.2405.11619

Anti-Phishing Working Group (APWG). (2023). *Phishing activity trends report*. https://docs.apwg.org/reports/apwg_trends_report_q4_2023.pdf

Ayo, F. E., & Alowolodu, O. D. (2024). A deep learning-based approach for detecting phishing websites. *Journal of Information Security and Applications*, *76*, 103-118.

Jain, A. K., & Gupta, B. B. (2018). Phishing detection: Analysis of visual similarity-based approaches. *Security and Communication Networks*, *2018*, 1-20. https://doi.org/10.1155/2017/5421046

Selvakumari, M., Sowjanya, M., Das, S., & Padmavathi, S. (2021). Phishing website detection using machine learning and deep learning techniques. *Journal of Physics: Conference*

Series, *1916*(1), 012015. https://doi.org/10.1088/1742-6596/1916/1/012015

Somesha, M., Pais, A. R., Rao, S., & Rathour, V. S. (2020). Efficient deep learning techniques for the detection of phishing websites. *Sādhanā*, *45*, Article 165. https://doi.org/10.1007/s12046-020-01399-8

Somesha, M., Pais, A. R., Rao, S., & Rathour, V. S. (2021). Performance analysis of machine learning algorithms for phishing detection. *Journal of King Saud University - Computer and Information Sciences*, *33*(9), 1159-1172. https://doi.org/10.1016/j.jksuci.2019.07.012

Ullah, I., Mahmoud, Q. H., Al-Sanabani, H., & Ahmed, A. (2024). A comprehensive survey of phishing detection techniques using machine learning and deep learning. *IEEE Access*, *12*, 1-25.

Verizon. (2023). *Data Breach Investigations Report (DBIR)*. https://www.verizon.com/business/resources/reports/dbir/