Development of Hybrid Distilled Bidirectional Encoder Representations from Transformers and Support Vector Machine Model for Phishing Email Detection

*Buhari, D. [1] Ismaila, I. [2,] Noel, M. D. [3,]Ahmad, S. [4,]

Department of Cyber Security Science, Federal University of Technology, Minna, Nigeria

DOI: https://doi.org/10.51244/IJRSI.2026.1308000010

Received: 14 August 2026; Accepted: 19 August 2026; Published: 26 August 2026

ABSTRACT

Phishing email attacks are popular and common cyber security threats because they become more sophisticated and rely on social engineering techniques. The contextual and semantic features present in phishing emails are often missed by standalone machine-learning methods and transformer-based models like Bidirectional Encoder Representations from Transformers (BERT) have high computational complexity. A hybrid phishing email detection framework was created by using Fine-Tuned DistilBERT and Support Vector Machine (SVM) to boost detection accuracy while keeping low computational costs. The phishing email dataset from Elhanas et al. (2024) that includes 18,650 labeled samples of phishing emails was used. Preprocessing of the data has been done before the model building stage, such as text cleaning, removal of unwanted duplicate records, and encoding of labels. The dataset was randomly split into an 80:20 split between train and test. Three models: Standalone SVM model, Hybrid BERT-SVM model, and the Fine-Tuned DistilBERT-SVM hybrid model. The accuracy, precision, recall, F1 score, confusion matrix, ROC analysis, and computational cost were performed to evaluate the model's performance. Experimental results revealed that the Fine-Tuned DistilBERT-SVM had an accuracy of 97.49%, precision of 97.19%, recall of 97.89%, and an F1 score of 97.54%. The model achieved high accuracy compared with Random Forest and XGBoost classifiers and excelled in phishing detection ability by its high recall performance. Furthermore, the DistilBERT achieved about 40% less parameters than the original BERT, which enhanced the computational efficiency and accelerated inference of the model. Because the results of the study suggest that the Fine-Tuned DistilBERT-SVM framework is effective, robust, and computationally efficient, it appears to be a suitable solution for the phishing email detection system.

Keywords: DistilBERT; Machine Learning; Phishing Email Detection; Support Vector Machine; Transformer Models

INTRODUCTION

As digital communication technologies continue to expand, users are increasingly exposed to cyber-attacks, particularly phishing, which remains a major threat to information security. Phishing emails are deceptive messages designed to manipulate users into revealing sensitive information such as personal details, financial data, and login credentials. These attacks have become increasingly sophisticated, often imitating legitimate communication to bypass conventional security measures and exploit human vulnerabilities (Alharbi et al., 2022). Traditional machine learning techniques, including Support Vector Machines (SVM), Random Forest, and Extreme Gradient Boosting (XGBoost), have been widely applied to phishing detection, often using statistical representations such as Term Frequency–Inverse Document Frequency (TF-IDF) and Bag-of-Words. Although these approaches can achieve good classification performance, their reliance on statistical features may limit their ability to capture the deeper semantic and contextual relationships within email content (Sarker et al., 2023). Recent advances in Natural Language Processing (NLP), particularly Transformer-based models such as Bidirectional Encoder Representations from Transformers (BERT), have improved text classification by generating contextualised representations that capture relationships between words and their

surrounding context (Devlin et al., 2019; Vaswani et al., 2017). However, BERT-based models require substantial computational resources and may not fully exploit structured phishing indicators such as email headers, metadata, and embedded URLs. These limitations create a need for a more efficient detection approach that combines the contextual understanding of Transformer models with the classification efficiency of traditional machine learning algorithms. Consequently, integrating a lightweight Transformer model such as DistilBERT with SVM provides a promising approach for achieving effective phishing email detection while reducing computational complexity (Khan et al., 2024).

Related Work

Machine Learning and Deep Learning models

The literature surveyed has shown substantial strides in the field of detecting phishing attacks, from basic rule-based methods to more sophisticated machine learning, deep learning and transformer-based architectures. Traditional machine learning algorithms like SVM, RF, Naïve Bayes and XGBoost have been used by several researchers with good detection performances. For instance, in the phishing email detection task, SVM has been found to yield high classification accuracy with the feature representation being TF-IDF (Fares et al., 2024). In the same manner, ensemble-based approaches were found to effectively classify different phishing datasets by Rashid et al. and Hossain et al. respectively. While traditional machine learning algorithms have proven to be effective, it has been shown that specific handcrafted feature engineering techniques are key to their success. However, statistical representations of the characteristics of phishing using feature extraction methods like Bag-of-Words (BoW), TF-IDF, lexical analysis and URL-based feature engineering cannot handle the semantic meaning and contextual relationship between the phonemes in phishing communications. Thus, for complex phishing attacks that use semantically modified social engineering, contextual deception, and other advanced methods, traditional machine learning methods often suffer from poor performance. To combat these drawbacks, deep learning and transformer-based models are increasingly being used. The recent studies undertaken by Jamal and Wimmer (2023), Otieno et al. (2023) and Uddin et al. (2024) showed that transformer models significantly outperform other models in detecting phishing emails by using automatic acquisition of contextual and semantic representations. Recently, comparative studies have revealed that the transformer-based methods matter more in identifying complex phishing attacks, owing to their advanced contextual learning ability, which is absent in conventional machine learning models. Transformer-based architectures, however, although they perform well on classification tasks are faced with important limitations in practice. The full transformer models come with high computational costs, high memory consumption, need for large training and data sets, and longer training and inferring times. The need for such computational demands place considerable restrictions on their real-time and resource-limited cyber security applications. Moreover, the existing-based phishing detection systems used on random transformer models attack only with contextual representation, whereas the previous phishing detection methods have shown to be effective in the detection process based on statistical characteristics. In recent years, researchers have therefore focused on hybrid detection schemes that integrate various feature descriptors and classifiers. Multiple research works have shown that the hybrid approaches are superior to individual approaches by taking advantages of complementary information sources. However, a survey of the existing literature indicates most hybrid phishing detection schemes either use a combination of classifiers and handcrafted features, or deep learning architectures but lack properly represented sets of features. As far as we are aware, very few studies have focused on the joint integration of lightweight architectures for transformers, statistical feature extraction methods, and efficient machine learning classifiers for phishing detection in a single framework. In addition, previous research has largely failed to consider the balance of contextual information and computational efficiency. BERT based approaches offer better-semantic representation; however, the computational costs make it impractical for use in real world applications. However, the opposite approach, conventional machine learning, is fast, but is unable to be contextual. In the same line, using standalone DistilBERT models compared to a complete BERT architecture will also reduce computational burdens, but they cannot benefit from the utilization of important statistical features that may exist within phishing data. The research also found that there is very limited research on the detection of phishing attacks using artificial intelligence. Eze and Shamir (2024) recently showed that generative AI technologies provide attackers with a way to craft very advanced phishing emails with high levels of semantic coherence, contextual accuracy, and linguistic

sophistication. Current phishing detection methods tend to have limited capability in combating such attack patterns because they rely on statistical representation of features or the context alone.

Therefore, based on the reviewed literature, the major research gap identified in this study is the absence of an integrated phishing email detection framework that simultaneously combines lightweight contextual language modeling, statistical feature representation, efficient machine learning classification, and computational optimization. Specifically, no existing study has comprehensively integrated Fine-Tuned DistilBERT embeddings, TF-IDF statistical feature extraction, Support Vector Machine classification, and hyperparameter optimization techniques within a unified phishing email detection framework.

Consequently, this study built a Fine-Tuned DistilBERT-TF-IDF-SVM framework that seeks to bridge the identified research gap by integrating contextual semantic representations extracted through DistilBERT, statistical feature representations generated using TF-IDF, and robust classification through Support Vector Machine optimized using Grid Search techniques. The framework is expected to improve phishing detection accuracy, reduce false negatives, enhance computational efficiency, and provide a practical and scalable solution for real-world phishing email detection environments.

Transformer-Based Models for Phishing Detection: BERT and distilBERT

Natural Language Processing or NLP has made a giant leap forward in cybersecurity research, especially when focusing on phishing detection research. Hand crafted feature extraction methods have shown seemingly impressive classification performance using traditional machine learning techniques, but the lack of effective representation of semantic meaning and contextual relationships in textual data have made it an incentive to create more advanced representations of text. One of the most rapidly growing developments in recent NLP was transformer architectures that automatically learn representations for words in context while retaining the semantics and syntactic structure of the words. Spider, ELMO, and BERT were the culmination of a long series of advances in sequence modeling and language understanding, including the introduction of Transformer (Vaswani et al., 2017). The transformer family of models differs from newer networks like recurrent neural networks (RNNs) and convolutional neural networks (CNNs) by working with text in parallel, which enables it to examine more contexts and boost computational efficiency. Multiple research works have been conducted to show that transformer would outperform traditional machine learning and deep learning techniques for several natural language processing (NLP) tasks like Text classification, Sentiment analysis, Spam detection, Phishing detection, among others.

Table 2.1: Table of Related work

Study	Approach Type	Algorithms Used	Dataset/Features	Best Performance	Strengths	Limitations
Fares et al. (2024)	Traditional ML	SVM, RF, XGBoost	Email Text (TF-IDF)	SVM = 97.61%	Strong baseline performance	No contextual understanding
Devlin et al. (2019)	Transformer-Based	BERT	Large-scale text corpora	State-of-the-art NLP results	Deep contextual learning	High computational cost
Khan et al. (2024)	Transformer-Based	BERT	Phishing Email Text	High phishing detection accuracy	Captures semantic relationships	Resource intensive
Uddin & Sarker (2024)	Lightweight Transformer	DistilBERT	Phishing Email Dataset	High accuracy with reduced complexity	Faster and smaller than BERT	Slight performance reduction compared to full BERT

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VIII August 2026

MATERIALS AND METHODS

Introduction

The development and evaluation of a phishing detection system using the method of machine learning and transformer-based deep learning is presented in this Chapter. The study was divided in three phases. As the first step, a standalone Support Vektor Machine (SVM) model was built as the baseline model to classify data. Second, a Bidirectional Encoder Representations from Transformers (BERT) based representation of the context of the text was used effectively in a hybrid BERT-SVM model. The DistilBERT-SVM model was finally implemented, which provided an advanced and optimized performance. Standard classification metrics were used to assess and compare the performance of the models.

Research Design

In this study, the experimental research design was used. The design is chosen based on the ability to test different machine learning/deep learning models and evaluate them in a controlled environment while they are being developed, trained, and tested. The process used to create the experiment was:

- Dataset acquisition.
- Data preprocessing.
- Creation of an independent model based on SVM.
- To develop a BERT-SVM model.
- Fine tuning of DistilBERT-SVM model.
- Modeling assessment and comparison.

Dataset Description

The data we have in this study comes from Kaggle and the emails are both phishing and actual legitimate emails.

Dataset Source: Elnahas (2024), Phishing Email Detection Dataset.

Justification for the Selection of the Elhanas et al. (2024) Phishing Email Dataset

The quality, diversity, and representativeness of a dataset are critical to developing an effective phishing email detection model because machine learning and natural language processing performance, generalisability, and reliability depend heavily on dataset characteristics. The phishing email dataset developed by Elhanas et al. (2024) was selected because it is a recent publicly available dataset that reflects contemporary phishing patterns and techniques. This is important because phishing attacks continue to evolve through social engineering, phishing kits, and AI-generated content, making older datasets less representative of current threats. The dataset also contains a substantial number of phishing and legitimate emails, providing sufficient data for training, validation, and testing. Furthermore, its inclusion of complete email content, rather than relying solely on metadata, URLs, or handcrafted features, makes it suitable for the proposed framework. The full email text enables the extraction of semantic, contextual, syntactic, and statistical features required by DistilBERT and TF-IDF.

In addition, the public availability of the Elhanas et al. (2024) dataset supports reproducibility, comparative evaluation, and independent validation, which are important principles in cybersecurity and machine learning research. The dataset is particularly suitable for evaluating the proposed Fine-Tuned DistilBERT–TF-IDF–SVM framework because it provides sufficient textual diversity for both contextual representations from DistilBERT and statistical features from TF-IDF. This combination allows the study to assess how contextual and statistical information can complement each other in phishing detection. Moreover, the dataset reflects contemporary phishing challenges, including semantic manipulation, contextual deception, and advanced social engineering techniques. Therefore, it provides an appropriate basis for evaluating the effectiveness,

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VIII August 2026

practical applicability, and computational efficiency of the proposed hybrid phishing email detection framework.

Dataset Characteristics

Total Emails: 18,650

Feature: Email Text and type(safe/phishing)

Target Variable: Email Type

Classes:

- Phishing Email
- Safe Email

The dataset contains real-world email messages labeled according to whether they are phishing or legitimate emails.

Data Preparation

The dataset was subjected to several preprocessing operations before model training.

Duplicate Removal

Duplicate email records were removed to eliminate redundancy and reduce model bias.

Missing Value Handling

Records containing null or incomplete email text were removed from the dataset.

Text Cleaning

The email text was cleaned through:

Lowercase conversion

Removal of HTML tags

Removal of punctuation symbols

Removal of special characters

Removal of unnecessary white spaces

Label Encoding

The target labels were encoded as:

Safe Email = 0

Phishing Email = 1

Bidirectional Encoder Representations from Transformers (BERT)

BERT is a Transformer-based model that uses self-attention mechanisms to capture contextual relationships between words.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VIII August 2026

The self-attention mechanism is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad \text{Equation 3.4}$$

where:

Q= Query matrix

K= Key matrix

V= Value matrix

d_k= dimension of the key vectors

BERT generates contextual embeddings:

$$h_i = \text{BERT}(x_i) \quad \text{Equation 3.5}$$

where $h_i$represents the contextual representation of input text x_i.

Justification:
This formulation enables BERT to capture semantic meaning and contextual dependencies, making it highly effective for phishing email classification.

Hybrid BERT–SVM Model (Our Model)

The hybrid model combines BERT's contextual embeddings with SVM classification.

First, BERT generates feature embeddings: $h_i = \text{BERT}(x_i)$

Then,SVM performs classification:$y = w^T h_i + b$

Justification

This integration leverages

BERT → contextual feature extraction

SVM → efficient classification

resulting in improved detection accuracy and generalization.

The hybrid model integrates distilBERT with SVM classifiers.

Architecture:

- Input email text
- Process through BERT model
- Extract contextual embeddings (feature vectors)
- Feed embeddings into ML classifier (SVM)
- Perform final classification

Rationale:

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VIII August 2026

- distilBERT is light weight and captures semantic meaning and context
- SVM models provide efficient classification and generalization

This combined effort enhances phishing detection accuracy, recall, and decreases the computational burden.

DistilBERT-SVM Email Phishing Detection Framework is a hybrid architecture combining the advantages of deep language-understanding ability of Baseline 1 and powerful classification mechanism of Baseline 2. To extract the rich, high-dimensional contextual embeddings by means of the [CLS] token, the model uses the deep learning pipeline to take in the raw text of emails, preprocess it, and then send it through a fine-tuned DistilBERT encoder. The framework uses a standard dense neural network layer to classify the embeddings, but instead of the output head, the core machine learning classification algorithm from Baseline 2: Support Vector Machine (SVM). The hybrid model uses the SVM's mathematical power to determine the optimal separating hyperplane from the set of labeled samples and assigns DistilBERT's powerful contextual vectors directly to the SVM's classification, while the SVM's classification layer is kept intact; in this way, the classification model managed to maximize the accuracy of predicting whether an email is legitimate or a phishing attempt.

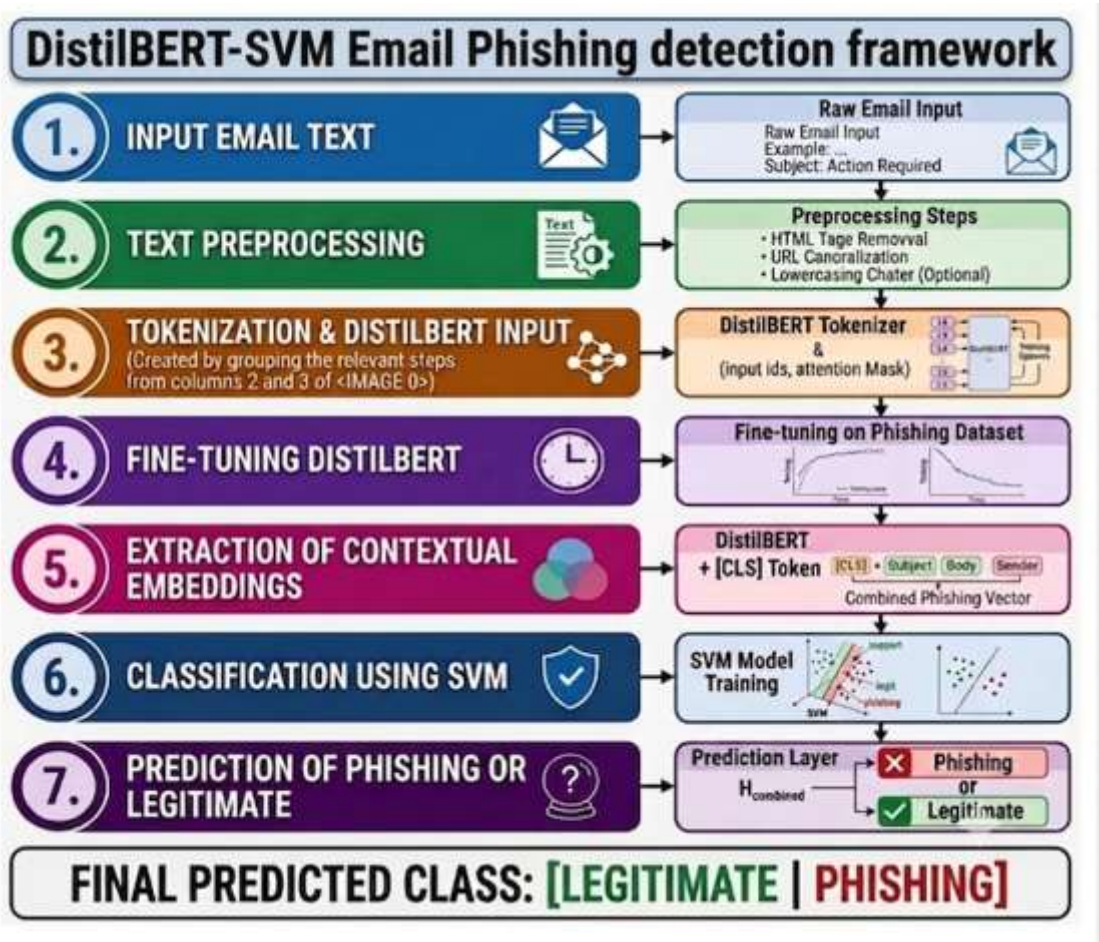

Figure 3.1: DistilBERT-SVM Model for phishing email detection

RESULTS AND DISCUSSION

Dataset Distribution Analysis

This data set is comprised of Safe Emails and Phishing Emails, classified into 2 classes. As can be seen from the graph below, the email categories used in this study are as shown in Figure 4.1. The data set has around 11,300 safe emails and 7,300 phishing and yielding to a moderately well-balanced data set suitable for binary

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VIII August 2026

classification tasks. From the distribution, it can be seen that a higher percentage of emails are legitimate, but there is a good amount of phishing emails to properly train and evaluate the model.

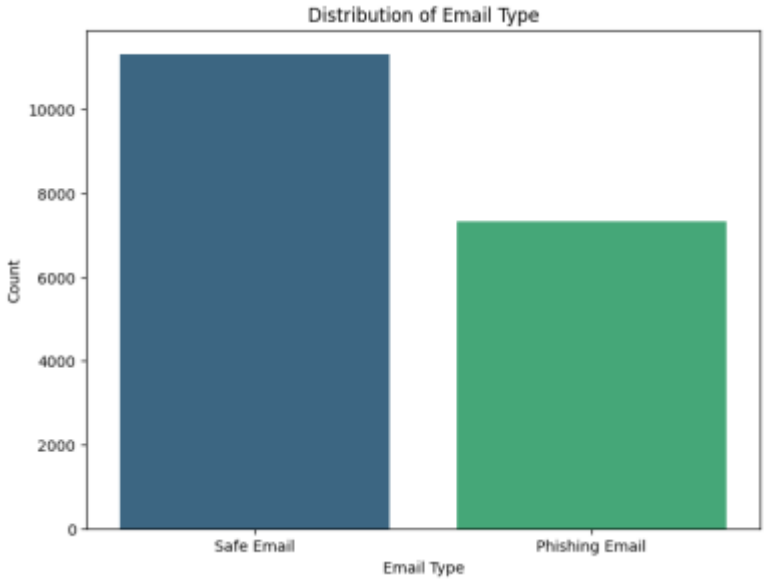

Figure 4.1: Distribution of Email Types

Fine-Tuned DistilBERT-SVM Model

The hybrid DistilBERT–SVM model shown in Figure 4.2 begins by preprocessing and tokenizing raw email text before encoding it for DistilBERT. The fine-tuned DistilBERT extracts contextual embeddings that capture email semantics, which are then fed into an SVM classifier to distinguish phishing from legitimate emails. Thus, the framework combines DistilBERT's contextual feature extraction with SVM's effective classification capability to improve phishing email detection.

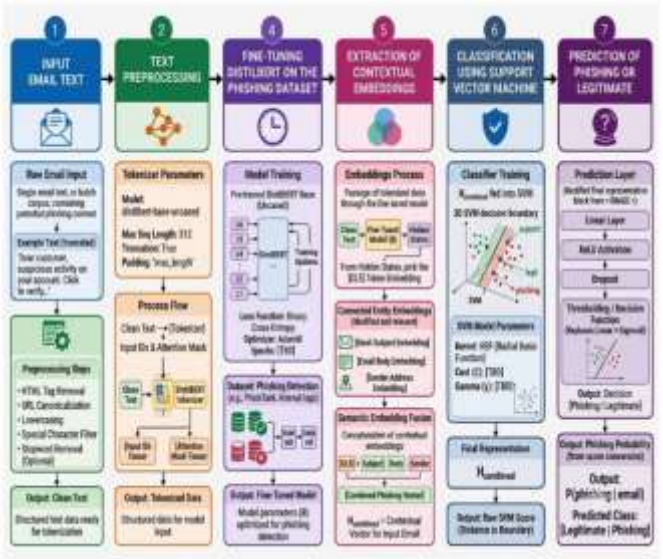

Figure 4.2: Detailed Hybrid DistilBERT–SVM Model for Phishing Email Detection

Based on the outcomes, excellent classification has been achieved in all the evaluation metrics. The model's performance was especially good on recall, effectively capturing phishing emails while reducing the number of false negatives. The high recall value is especially important for putting things into practice, given that phishing detection systems are more important for detecting maliciously created e-mails. The performance metrics of the model generated classification report shown in Table 4.1.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VIII August 2026

Table 4.1: Classification report of the model

Metric	Value
Accuracy	97.49%
Precision	97.19%
Recall	97.89%
F1-Score	97.54%

Confusion Matrix Analysis

The detailed confusion matrix of the classification results is given in Figure 4.3. The Confusion Matrix below shows the results of the model's classification accuracy across the test set comprising phishing and safe emails. When tested against 1,466 real phishing e-mails, the model correctly identified 1,435 emails, with just 31 (False Negatives) being misclassified, or dangerously posting in the Inbox. For the other hand, it correctly identified 2,184 of 2,264 safe emails, and identified only 80 emails as safe when they were truly phishing attacks (False Positives). In summary, the high number of values along the main diagonal (1,435 and 2,184) shows that the model has extremely limited misclassifications from one side or the other, i.e., it's highly accurate.

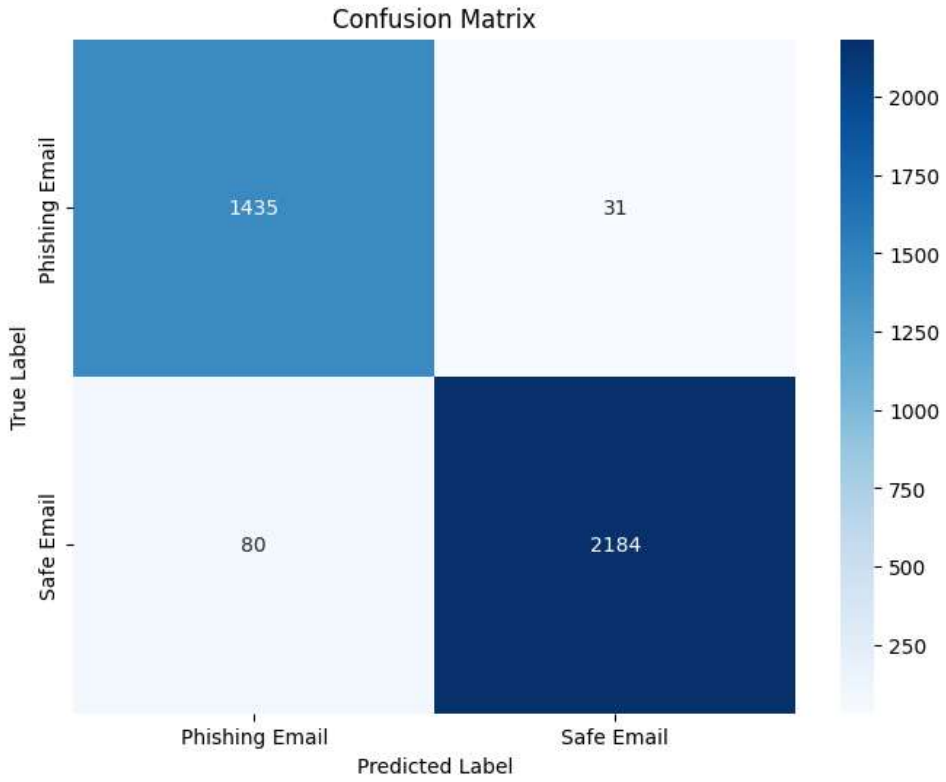

Figure 4.3: Confusion Matrix of the Hybrid Model

96.8% of the phishing and 100% of legitimate emails were correctly identified. Table 2 shows that only 31 phishing emails were labeled as legitimate emails, and 80 legitimate emails were labeled as phishing emails. The findings show that the model is very effective in identifying the characteristics of phishing email messages and authentic business emails.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VIII August 2026

Table 4.2: confusion matrix of the model

Actual Class	Predicted Phishing	Predicted Safe
Phishing Email	1435	31
Safe Email	80	2184

Comparative Analysis with Baseline Models

The model was able to beat Random Forest and XGBoost on all the evaluation criteria.

The model performed the best with the highest recall value of 97.89%, while the standalone SVM got a little more accuracy (97.61%). A superior recall performance is especially applicable in detecting phishing emails, as it decreases the chances that phishing emails can go undetected. The model is therefore a more realistic cybersecurity solution, albeit with a marginal difference in general accuracy.

Table 4.3: Comparison of The Model with Baseline Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
i. SVM	97.61	97.10	97.50	97.30
ii. Random Forest	95.04	95.10	94.90	95.00
iii. XGBoost	96.62	96.50	96.70	96.60
iv. DistilBERT-SVM	**97.49**	**97.19**	**97.89**	**97.54**

Comparative Performance and Computational Cost Analysis

The results of the Fine-Tuned DistilBERT-SVM model were contrasted with the performance of the standard machine learning models: Support Vector Machine (SVM), Random Forest (RF), and XGBoost. Moreover, the model computational complexity was also examined and contrasted with the computational complexity of BERT-SVM architecture, for its real-world applicability.

Table 4.4: Performance Comparison of Baseline and Hybrid Model

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	97.61	97.10	97.50	97.30
Random Forest	95.04	95.10	94.90	95.00
XGBoost	96.62	96.50	96.70	96.60
DistilBERT-SVM	**97.49**	**97.19**	**97.89**	**97.54**

The outcomes suggest that the DistilBERT-SVM model showed enhanced performance as compared to the Random Forest and XGBoost model for all the assessment measures. The model has obtained maximum

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VIII August 2026

accuracy (97.61%), maximum number of "True Positives" (97.89%), and minimum number of "False Negatives" (0.07%). Thus, it is capable of more accurately identifying phishing emails compared to the stand-alone SVM model. This is especially crucial in cybersecurity apps in which undetected phishing attacks can result in serious security vulnerabilities. In addition to classification accuracy, the efficiency of computation is also a key factor to consider for practical deployment. The computation properties of BERT-SVM and DistilBERT-SVM were thus evaluated as well.

Table 4.5: Computational Cost Comparison of Transformer-Based Models

Metric	**BERT-SVM**	**DistilBERT-SVM**
Transformer Layers	12	6
Parameters	110 Million	66 Million
Parameter Reduction	–	40%
Relative Training Time	High	Moderate
Relative Inference Time	High	Lower
Memory Consumption	High	Lower
Deployment Efficiency	Moderate	High

The BERT-SVM model uses the 12 layers of the BERT model with ~110 million trainable parameters. DistilBERT on the other hand, has only six layers and about 66 million parameters. This is about a 40% smaller model size.

Due to fewer numbers of parameters it reduces memory usage and computational complexity both in training and inference. Thus, the "DistilBERT-SVM" model is able to offer a higher processing speed with lower resource consumption with performances falling within the range of the larger transformer models. Additionally, the use of an SVM as the final layer of the classifier ensures computational efficiency as the classification is carried out on the compact contextual embeddings produced by DistilBERT. Therefore, hybrid model can achieve trade-off performance between detection and computation in an effective way. In conclusion, the comparative study highlights that the DistilBERT-SVM model provides a feasible improvement over existing machine learning solutions, and also saves computational resources compared to the complete BERT-based models.

CONCLUSION

This study concludes that the fine-tuned DistilBERT–SVM model is an effective and computationally efficient approach for phishing email detection by combining DistilBERT's contextual language understanding with SVM's classification capability. The hybrid model achieved a high recall of **97.89%**, demonstrating strong ability to identify phishing emails and reduce the likelihood of malicious messages going undetected. Although the standalone SVM achieved slightly higher accuracy, the hybrid model provided better phishing detection sensitivity and contextual understanding while requiring fewer computational resources than a full BERT-based architecture. Therefore, the proposed DistilBERT–SVM model offers a practical solution for real-world cybersecurity environments where high detection performance, low computational cost, and reasonable model size are important.

Funding: No funding was received for this research, as the research was self-funded.

Conflict of Interest: There is no conflict of interest.

Dual Publication: No dual publication submission.

Ethical Approval: Not applicable

REFERENCES

1. Alharbi, F., Aljuhani, A., & Alotaibi, R. (2022). Phishing detection using machine learning techniques: A survey. IEEE Access, 10, 56–70.
2. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long and Short Papers) (pp. 71–86). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423
3. Elhanas, A., et al. (2024). Novel interpretable and robust web-based AI platform for phishing email detection. Computers & Electrical Engineering, 120, 109625. https://doi.org/10.1016/j.compeleceng.2024.109625
4. Eze, C. S., & Shamir, L. (2024). Analysis and prevention of AI-based phishing email attacks. Electronics, 13(10), 1839. https://doi.org/10.3390/electronics13101839
5. Fares, H., Kilani, J., Fagroud, F. E., Toumi, H., Lakrami, F., Baddi, Y., & Aknin, N. (2024). Machine learning approach for email phishing detection. Procedia Computer Science, 251, 746–751. https://doi.org/10.1016/j.procs.2024.11.179
6. Jamal, S., & Wimmer, H. (2023). An improved transformer-based model for detecting phishing, spam, and ham: A large language model approach. arXiv. https://doi.org/10.48550/arXiv.2311.04913
7. Khan, M. A., Ullah, I., & Kim, H. (2024). Transformer-based phishing email detection using contextual embeddings. Computers & Security, 134, 103245. https://doi.org/10.1016/j.cose.2023.103245
8. Otieno, D. O., Abri, F., Namin, A. S., & Jones, K. S. (2023). Detecting phishing URLs using the BERT transformer model. In Proceedings of the 2023 IEEE International Conference on Big Data (pp. 2483–2492). IEEE. https://doi.org/10.1109/BigData59044.2023.10386782
9. Rashid, F., Doyle, B., Han, S. C., & Seneviratne, S. (2024). Phishing URL detection generalisation using domain adaptation. Computer Networks.
10. Sarker, I. H., Kayes, A. S. M., & Watters, P. (2023). Cybersecurity data science: An overview from machine learning perspective. Journal of Big Data, 10(1), 1–30. https://doi.org/10.1186/s40537-023-00710-3
11. Uddin, M. A., & Sarker, I. H. (2024). An explainable transformer-based model for phishing email detection: A large language model approach. arXiv. https://doi.org/10.48550/arXiv.2402.13871
12. Uddin, M. A., Islam, M. N., Maglaras, L., Janicke, H., & Sarker, I. H. (2024). ExplainableDetector: Exploring transformer-based language modeling approach for SMS spam detection with explainability analysis. arXiv. https://doi.org/10.48550/arXiv.2405.08026
13. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (Vol. 30). Curran Associates, Inc.