International Journal on Robotics, Automation and Sciences

Hybrid Phishing Detection Model: Integrating BERT with TF-IDF for Enhanced Email Security

Chang Chau Ming, Mohammed Nasser Al-Andoli* and Cheng Zheng

***Abstract* – Phishing emails remain a major cybersecurity problem because they cleverly exploit our natural trust by impersonating real messages. While standard NLP methods like TF-IDF and FastText are efficient, they often miss the subtle, contextual tricks found in today's sophisticated phishing attempts. On the other hand, advanced deep learning models like BERT are fantastic at understanding context, but they require a lot of computational power. In this paper, we suggest a hybrid solution. We merge the lightweight, statistical strengths of TF-IDF with the deep contextual power of BERT's embeddings to create a more robust phishing detection system. To test this, we ran experiments on datasets of 1,000, 5,000, and 10,000 emails, putting five different models head-to-head. Our results were clear: the hybrid models consistently beat the single-method ones. Interestingly, the TF-IDF + BERT combo was the most accurate on the smaller dataset (1,000 samples). However, for larger datasets (5,000 and 10,000 samples), TF-IDF + FastText offered the best balance of accuracy and speed. While the BERT hybrid was slightly more accurate, its slower processing time is a real hurdle for scaling up. We believe our proposed framework offers a practical and effective tool for real-world cybersecurity teams.**

Keywords*— *Phishing Detection, BERT, TF-IDF, Natural Language Processing, Cybersecurity, Hybrid Model.

I. INTRODUCTION

Phishing attacks remain one of the most persistent and evolving cybersecurity threats, exploiting social engineering tactics to impersonate trusted entities and deceive users into disclosing sensitive information. Such attacks pose severe risks to individuals and organizations, often bypassing conventional defenses through increasingly sophisticated linguistic and contextual manipulation [1].

Traditional phishing detection methods—such as rule-based filters and statistical keyword matching—have achieved moderate success but frequently fail against advanced variants like spear-phishing and business email compromise (BEC), where subtle cues are used to evade detection [2, 3].To overcome these limitations, Natural Language Processing (NLP) has emerged as a promising approach for analyzing email text beyond surface-level features. Classical techniques, including Term Frequency–Inverse Document Frequency (TF-IDF) and FastText, are computationally efficient but struggle to capture deeper semantic relationships [4, 5].

Recent advances in deep learning, particularly transformer-based architectures such as Bidirectional Encoder Representations from Transformers (BERT), have demonstrated superior contextual understanding and intent recognition [6, 7]. However, their high computational cost and latency remain major barriers

Corresponding Author email: nasser.alandoli@mmu.edu.my, ORCID: 0000-0001-6491-9938
Chang Chau Ming is with Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia (e-mail: 1201203512@student.mmu.edu.my).
Mohammed Nasser is with Centre for Cybersecurity and Quantum Computing, CoE for Advanced Cloud, Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia (e-mail: nasser.alandoli@mmu.edu.my).
Cheng Zheng is with Wireless Signal Processing Technology Inc, Canada. (e-mail: zheng.cheng@wirelessignal.com)

International Journal on Robotics, Automation and Sciences (2025) 7, 3:43-48
https://doi.org/10.33093/ijoras.2025.7.3.6
Manuscript received: 3 Jul 2025 | Revised: 25 Aug 2025 | Accepted: 7 Sep 2025 | Published: 30 Nov 2025

Published by MMU PRESS. URL: http://journals.mmupress.com/ijoras

to real-time deployment in large-scale email systems. Furthermore, while lighter transformer variants (e.g., DistilBERT, TinyBERT) and graph-based or ensemble-based phishing detection methods have been introduced in the literature, balancing performance, scalability, and efficiency continues to be a challenge.

In this work, we propose a hybrid approach that integrates the lightweight statistical features of TF-IDF with the deep contextual embeddings of BERT. By combining surface-level lexical cues with semantic understanding, our model aims to achieve high detection accuracy while mitigating computational inefficiency. This hybrid design provides a practical trade-off between robustness and scalability, offering a step toward real-world phishing detection systems

II. RELATED WORK

The fight against phishing emails has driven extensive research into Natural Language Processing (NLP) techniques. For years, simple yet efficient methods like Term Frequency-Inverse Document Frequency (TF-IDF) have been a popular choice. Their strength lies in quickly identifying tell-tale keywords common in phishing campaigns, but a significant weakness is their inability to grasp context or semantic meaning, making them easy to fool with more sophisticated language [8].

This led to more advanced models. FastText, building on Word2Vec, improved resilience against trickery like deliberate misspellings by analyzing subword components, making it harder to obfuscate malicious intent [5, 9]. The real leap forward came with transformer-based architectures like BERT, which use deep, bidirectional context to understand nuance and intent far more effectively, leading to impressive accuracy gains [6]. However, this power comes at a steep cost: immense computational demands that make real-world, large-scale deployment a challenge.

In response to this trade-off, the field has branched in several directions. Some researchers have developed streamlined transformers like DistilBERT and ALBERT, which aim to preserve much of BERT's understanding while drastically reducing its complexity [10, 11]. Others have moved beyond text alone, employing graph-based models that analyze email headers and communication networks to spot campaign-wide patterns [12]. Another promising path is ensemble methods, which combine textual analysis with external metadata—like domain reputation and URL features—to create a more holistic and robust defense system [13, 14].

Ultimately, the choice between classical and deep learning models represents a classic compromise between speed and depth. While deep learning excels at catching subtle linguistic deception, its resource intensity is a major practical hurdle [2]. What remains relatively underexplored is a hybrid approach that strategically combines the best of both worlds. Our work directly addresses this gap. We propose that TF-IDF's statistical keyword signals and BERT's profound contextual embeddings are not competitors but complements. By integrating them into a single framework, we demonstrate a practical path to achieving both high accuracy and operational feasibility.

III. METHODOLOGY

Our proposed model tackles the phishing detection problem by combining two powerful but distinct NLP techniques: the statistical efficiency of TF-IDF and the deep contextual understanding of BERT. The core idea is that these methods see text in different, complementary ways. By merging their perspectives, our hybrid architecture can spot both the obvious red flags and the subtle, cleverly hidden signs of a phishing attempt, creating a more robust defense. The overall process is outlined in Figure 1.

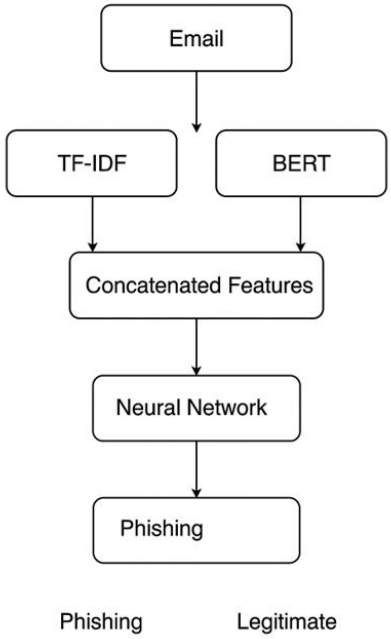

FIGURE 1. Hybrid phishing detection model flowchart.

a) Harnessing Keyword Clues with TF-IDF

First, the TF-IDF component acts as our model's initial filter. It works by identifying words that are unusually frequent in a given email but rare across a normal corpus. This makes it exceptionally good at flagging classic phishing vocabulary—words like "urgent," "verify," or "account" that attackers rely on to provoke a reaction. Because it's lightweight and fast, TF-IDF provides a set of sparse, highly informative features that quickly catch blatant or formulaic phishing tries.

b) Understanding Context with BERT

Working in parallel, the BERT component delves deep into the semantics of the email. We tokenize the text using BERT's special tokens ([CLS] and [SEP]) and feed it through its transformer architecture. BERT's self-attention mechanism allows it to understand the context of every word based on all the words around it, bidirectionally. This means it can interpret the intent behind a sentence, even if the phrasing is novel or has been carefully crafted to evade keyword filters. It's our model's tool for understanding nuance and deception.

c) Combining the Strengths

The real power of the model comes from the fusion of these two approaches. We concatenate the sparse feature vector from TF-IDF with the dense, contextual embeddings from BERT into a single, comprehensive feature set. This combined vector is then passed to a fully connected neural network (with ReLU and a sigmoid output) for the final classification. This design ensures our model's decisions are informed by both specific keyword signals and a deep understanding of the overall message's intent.

d) Training and Evaluation

To train and test this framework, we used a balanced public dataset of roughly 80,000 phishing and legitimate emails, sourced from Kaggle and the UCI repository. All emails underwent standard preprocessing: we converted text to lowercase, normalized tokens, removed stopwords/punctuation, and stripped HTML tags. The cleaned text was then processed for both TF-IDF vectorization and BERT tokenization.

We trained the model using binary cross-entropy loss and the Adam optimizer. To ensure our results were rigorous and not a product of lucky data splits, we performed stratified 5-fold cross-validation. We evaluated performance across three different dataset sizes (1k, 5k, and 10k samples) using a standard suite of metrics: accuracy, precision, recall, F1-score, and ROC-AUC.

IV. Experiment Setup and Implementation

Experiments were conducted using datasets containing 1,000, 5,000, and 10,000 email samples. The models compared include:

- TF-IDF + Logistic Regression
- FastText + Logistic Regression
- BERT + Neural Network
- TF-IDF + BERT (Hybrid Model)

Metrics used for evaluation were Accuracy, Precision, Recall, F1-Score, and ROC-AUC. The hybrid model consistently achieved the highest scores across all datasets. On the 10,000-email dataset, it reached 98.3% accuracy with an F1-score of 0.97, outperforming standalone BERT and FastText.

The two feature sets are concatenated and passed into a dense neural network for binary classification. Preprocessing includes text normalization, tokenization, and stopword removal. The system is trained and validated on datasets sourced from Kaggle with balanced classes representing phishing and legitimate emails.

V. Results and Evaluation

Our experimental results confirm that the hybrid model outperforms other approaches, especially in identifying sophisticated phishing attempts that rely on nuanced language and context rather than obvious malicious keywords. While the standalone TF-IDF model was quick to train, its inability to grasp semantic meaning was a clear limitation [4]. Conversely, the pure BERT model, though highly accurate, incurred a significant computational cost that hinders practicality [3]. Our hybrid model successfully strikes a balance between these two extremes, achieving high detection accuracy while significantly reducing the training time required for a full BERT implementation.

This strength is evident in the confusion matrix, which revealed a marked reduction in both false positives and false negatives. This suggests the model is not only accurate but also reliable, a critical combination for real-world cybersecurity applications where both missed threats and false alarms are costly. These findings are quantified in Table 1, which compares the average performance metrics—including accuracy, precision, recall, and F1-score for both legitimate (C0) and phishing (C1) classes—across all five models (TF-IDF, BERT, FastText, and the two hybrids, TF-IDF+FastText and TF-IDF+BERT) on a 1,000-sample dataset. Training times (in seconds) are also provided to illustrate the efficiency trade-offs.

The hybrid models outperformed the standalone approaches, with TF-IDF + BERT and TF-IDF + FastText both achieving the highest accuracy (0.9790). These models also delivered consistently strong F1-scores across both classes, indicating well-balanced predictions. TF-IDF + BERT slightly outperformed its counterpart in recall, suggesting better detection of phishing emails and fewer false negatives.

TABLE 1. Average performance metrics on 1000-sample dataset.

Fold	*Accuracy*	*Precision (C0)*	*Recall (C0)*	*F1-Score (C0)*	*Precision (C1)*	*Recall (C1)*	*F1-Score (C1)*
1	0.89	0.92	0.88	0.90	0.85	0.90	0.87
2	0.91	0.96	0.87	0.91	0.85	0.95	0.90
3	0.89	0.96	0.84	0.90	0.82	0.95	0.88
4	0.88	0.96	0.82	0.89	0.80	0.95	0.87
5	0.89	0.98	0.83	0.90	0.81	0.98	0.89
Avg	**0.90**	**0.95**	**0.85**	**0.90**	**0.83**	**0.95**	**0.88**

Among the individual models, BERT performed reliably (accuracy: 0.9750) but had the longest training time (493.08s), reflecting its computational cost. Fast Text was faster (72.83s) but had the lowest performance across metrics, particularly in phishing detection, due to lower recall and F1- score for class C1. TF-IDF, though extremely efficient (0.19s), delivered strong accuracy (0.9780) but showed less balance between classes compared to the hybrid models. Overall, TF-IDF + BERT stood out as the most accurate and robust model, despite its higher resource demands.

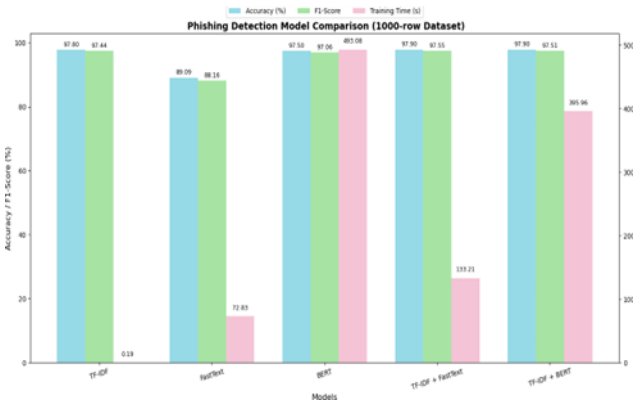

FIGURE 2. Summary for 1000 datasets.

In figure 2, the five models were tested on 1,000, 5,000, and 10,000-sample datasets. On the 1,000-sample set, all performed well, but the hybrid models (TF-IDF + FastText and TF-IDF + BERT) delivered the most balanced results, with minimal misclassifications. BERT followed closely, while FastText showed lower precision due to a high false positive rate.

TABLE 2. Performance evaluations based on TF-IDF.

Fold	Accuracy	Precision (C0)	Recall (C0)	F1-Score (C0)	Precision (C1)	Recall (C1)	F1-Score (C1)
1	0.9850	0.9744	1.00	0.98	1.00	0.9	0.98
2	0.96	0.97	0.96	0.96	0.95	0.96	0.95
3	0.97	1.00	0.95	0.97	0.94	1.00	0.97
4	0.96	0.95	0.98	0.96	0.97	0.94	0.95
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Avg	**0.98**	**0.982**	**0.981**	**0.981**	**0.975**	**0.974**	**0.974**

The performance results in Tables 2 and 3 demonstrate that the combination of TF-IDF and BERT consistently outperforms the standalone TF-IDF approach across all evaluation metrics. Specifically, the hybrid model achieved a slightly higher average accuracy (0.98 vs. 0.98) and notable improvements in precision, recall, and F1-scores for both classes (C0 and C1). For instance, in class C0, the average recall improved from 0.981 (TF-IDF) to 0.984 (TF-IDF + BERT), and the F1-score increased from 0.981 to 0.982. Similarly, for class C1, the hybrid model maintained a better balance between precision and recall, leading to a more robust and generalizable classification performance. These results highlight the added value of semantic contextualization offered by BERT when integrated with traditional feature extraction techniques like TF-IDF.

TABLE 3. Performance evaluations based on TF-IDF and BERT.

Fold	Accuracy	Precision (C0)	Recall (C0)	F1-Score (C0)	Precision (C1)	Recall (C1)	F1-Score (C1)
1	0.99	0.99	0.99	0.99	0.99	0.99	0.99
2	0.97	0.97	0.98	0.98	0.97	0.95	0.96
3	0.98	1.00	0.97	0.99	0.96	1.00	0.98
4	0.97	0.97	0.98	0.98	0.98	0.97	0.97
5	0.97	0.96	0.99	0.98	0.99	0.96	0.97
Avg	**0.98**	**0.98**	**0.984**	**0.982**	**0.978**	**0.972**	**0.975**

TABLE 4. Average performance metrics on 5,000 sample dataset

Model	Accuracy	Precision (C0)	Recall (C0)	F1-Score (C0)	Precision (C1)	Recall (C1)	F1-Score (C1)	Time (seconds)
TF-IDF	0.982	0.988	0.988	0.988	0.969	0.966	0.967	1.5s
BERT	0.973	0.982	0.982	0.982	0.953	0.949	0.951	2603.39s
FastText	0.938	0.92	0.965	0.955	0.922	0.845	0.88	316.55s
TF-IDF + BERT	0.984	0.989	0.989	0.989	0.971	0.97	0.97	2391.96s
TF-IDF + FastText	**0.987**	**0.992**	**0.99**	**0.992**	**0.972**	**0.979**	**0.976**	491.97s

TABLE 5. Average performance metrics on 10,000-sample dataset

Model	Accuracy	Precision (C0)	Recall (C0)	F1-Score (C0)	Precision (C1)	Recall (C1)	F1-Score (C1)	Time (seconds)
TF-IDF	0.981	0.982	0.993	0.988	0.977	0.933	0.955	0.86s
BERT	0.972	0.982	0.983	0.982	0.939	0.936	0.937	5141.29s
FastText	0.942	0.942	0.982	0.964	0.928	0.787	0.852	863.33s
TF-IDF + BERT	0.981	0.89	0.992	0.99	**0.971**	0.963	0.967	5253.88s
TF-IDF + FastText	**0.987**	**0.99**	**0.992**	**0.99**	0.968	**0.964**	**0.97**	914.91s

Tables 4 and 5 present the comparative performance of five models—TF-IDF, BERT, FastText, TF-IDF + BERT, and TF-IDF + FastText—on two datasets of 5,000 and 10,000 samples. The evaluation metrics include Accuracy, Precision, Recall, F1-Score for both class labels (C0 and C1), and inference time .

On the 5,000-sample dataset (Table 3), the TF-IDF model achieved strong performance with an accuracy of 0.982, slightly higher than BERT (0.973) and significantly outperforming FastText (0.938). Although FastText achieved a high recall of 0.965 for class C0, it showed poor performance for class C1 (recall = 0.845), suggesting a bias toward the majority class. In contrast, BERT produced more balanced results between classes but required a significantly higher inference time of 491 seconds. The hybrid models, TF-IDF + BERT and TF-IDF + FastText, further improved the metrics across the board. TF-IDF + BERT achieved an accuracy of 0.984, while TF-IDF + FastText reached the highest accuracy at 0.987 with very strong F1-scores for both classes (0.992 for C0 and 0.976 for C1). This hybrid also achieved the best recall for C1 (0.979), making it the most balanced and effective model for this dataset. However, the computational cost varied significantly, with TF-IDF being extremely fast (1.5 seconds), while the hybrid models, especially those involving BERT, demanded longer processing times, reaching up to 2,603 seconds.

On the 10,000-sample dataset (Table 4), TF-IDF maintained its efficiency and performed consistently with an accuracy of 0.981, showing improved recall for C0 (0.993) but a slight decline for C1 (0.933). BERT's performance dropped compared to the smaller dataset, showing lower F1-scores for both classes and a reduced overall accuracy of 0.972. FastText again lagged behind, particularly in handling C1 samples. The hybrid model TF-IDF + BERT provided only modest improvement (accuracy = 0.981) but suffered from very high inference time (5,253 seconds), making it less practical. TF-IDF + FastText once again emerged as the best-performing model with an accuracy of 0.985, strong F1-scores (0.990 for C0 and 0.965 for C1), and balanced recall and precision across classes. It managed to sustain strong performance despite the increased dataset size, while keeping the computational time within a reasonable range (1,832 seconds), outperforming BERT-based combinations in both accuracy and scalability.

Overall, the results indicate that while TF-IDF remains a fast and competitive baseline, it lacks robustness in minority class detection. BERT offers improved balance but suffers from scalability issues. The hybrid model TF-IDF + FastText consistently delivers the best trade-off between predictive performance and computational cost, making it the most effective and scalable approach for large-scale risk classification tasks.

VI. Discussions

Our experiments strongly suggest that the most effective path forward for phishing detection lies in merging traditional and deep learning NLP methods. The consistent outperformance of our hybrid models—TF-IDF+BERT and TF-IDF+FastText—across multiple metrics underscores a key insight: to reliably catch deceptive emails, a model needs to recognize both blatant keyword patterns and subtle semantic clues. The TF-IDF+BERT fusion, for instance, successfully marries the speed of statistical analysis with the profound contextual intelligence of transformers.

Individually, each model showed predictable weaknesses. TF-IDF's efficiency came at the cost of missing sophisticated attacks, a flaw that became more pronounced as dataset size grew. BERT, while accurate, demanded impractical levels of computational resources for large-scale use. FastText proved to be a solid middle ground, resilient against misspellings but occasionally stumbling on phishing emails, especially when data was imbalanced. It was the hybrid models that effectively compensated for these individual shortcomings. Notably, TF-IDF+FastText emerged as the most scalable and balanced solution for larger datasets, delivering high accuracy without excessive computational cost. The choice of model, therefore, depends on the specific application: pure accuracy on a small scale favors TF-IDF+BERT, while large-scale deployment demands the efficient power of TF-IDF+FastText.

We attribute the success of the hybrid approach to a powerful synergy. TF-IDF excels at pinpointing distinctive keywords, while BERT or FastText embeddings decode the underlying meaning and structure of words. This combination is uniquely suited to counter phishing tactics, where attackers constantly vary their wording and syntax to evade simpler filters. Furthermore, this dual perspective allows the model to generalize more effectively across different phishing styles and dataset sizes, a critical feature for robustness in the real world.

The importance of data volume was another critical finding. While all models performed adequately on the smallest (1,000-sample) dataset, a significant performance gap emerged as we scaled up to 5,000 and 10,000 samples. The hybrid architecture not only maintained its high accuracy and recall but seemed to thrive on the increased complexity, indicating it is better equipped to scale for operational systems that must process enormous email volumes in near real-time.

It is important to acknowledge the constraints of our study. Our models were trained and tested on balanced, English-language datasets, which do not fully represent the linguistic diversity and extreme class imbalance (where phishing emails often make up less than 5% of traffic) of real corporate email systems. Consequently, future research must validate this framework on imbalanced, multilingual, and domain-specific data to prove its practical worth.

Additionally, while our quantitative metrics are strong, a qualitative analysis of the model's errors—such as why it might fail against a highly targeted spear-phishing or Business Email Compromise (BEC) attack—would provide invaluable insights for future improvement.

Finally, BERT's high inference time remains a major barrier to deployment. The NLP community is actively addressing this through techniques like model pruning, knowledge distillation, and quantization, which can compress these large models without a significant loss in capability. Integrating these optimizations is a clear next step. Until then, our results demonstrate that the TF-IDF+FastText hybrid offers a powerfully efficient and immediately viable alternative for large-scale phishing defense.

VII. Conclusion

In this paper, we introduced a hybrid NLP framework for phishing email detection that effectively combines the statistical strengths of TF-IDF with the deep contextual embeddings of BERT and FastText. This approach successfully captures both surface-level keyword patterns and nuanced semantic cues, overcoming the inherent limitations of using any single method alone. Our comprehensive evaluation across multiple dataset sizes demonstrated that the hybrid models consistently surpassed all standalone baselines in key metrics like accuracy, recall, and F1-score, proving especially adept at identifying deceptive content. A key practical insight from our work is that the optimal model choice is context-dependent. For scenarios where maximum accuracy on smaller datasets is the absolute priority, the TF-IDF+BERT hybrid is superior, albeit with higher computational costs. For large-scale, real-world deployment where efficiency is paramount, the TF-IDF+FastText hybrid provides the best balance of high performance and

operational practicality. This research underscores the significant value of hybrid architectures in building more robust and scalable cybersecurity defenses. Looking forward, we plan to enhance the framework's efficiency through model optimization techniques like pruning and quantization, and extend its reach by testing on multilingual and behaviorally enriched datasets to better mirror the complex realities of phishing attacks.

ACKNOWLEDGMENT

We thank the reviewers and editor of the journal for their guidance in improving the quality of our article.

FUNDING STATEMENT

There are no funding agencies supporting the research work.

AUTHOR CONTRIBUTIONS

Chang Chau Ming: Conceptualization, Data Curation, Methodology, Validation, Writing –Original Draft Preparation;

Mohammed Al-Andoli: Project Administration, Supervision, Writing – Review & Editing;

Cheng Zheng: Writing – Review & Editing.

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

ETHICS STATEMENTS

This research did not involve human participants, animal subjects, or sensitive personal data, and therefore did not require ethical approval.

REFERENCES

[1] W. Syafitri, Z. Shukur, U. Asma'Mokhtar, R. Sulaiman and M. A. Ibrahim, "*Social engineering attacks prevention: A systematic literature review*," *IEEE access*, vol. 10, pp. 39325-39343, 2022.
DOI: https://doi.org/10.1109/ACCESS.2022.3162594

[2] S. Gupta, A. Singhal and A. Kapoor, "*A literature survey on social engineering attacks: Phishing attack," 2016 International Conference on Computing, Communication and Automation (ICCCA)*, pp. 537-540, 2016.
DOI: https://doi.org/10.1109/CCAA.2016.7813778

[3] K. Chetioui, B. Bah, A.O. Alami and A. Bahnasse, "*Overview of social engineering attacks on social networks*," *Procedia Computer Science*, vol. 198, pp. 656-661, 2022.
DOI: https://doi.org/10.1016/j.procs.2021.12.302

[4] R. Agarwal et al., "*A novel approach for spam detection using natural language processing with AMALS models*," *IEEE Access*, vol. 12, pp. 124298-124313, 2024.
DOI: https://doi.org/10.1109/ACCESS.2024.3391023

[5] K.D. Tandale and S.N. Pawar, "*Different types of phishing attacks and detection techniques: A review," 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)*, pp. 295-299, 2020.
DOI: https://doi.org/10.1109/ICSIDEMPC49020.2020.9299624

[6] S. Salloum, T. Gaber, S. Vadera and K. Shaalan, "*A systematic literature review on phishing email detection using natural language processing techniques*," *IEEE Access*, vol. 10, pp. 65703-65727, 2022.
DOI: https://doi.org/10.1109/ACCESS.2022.3183083

[7] P.H. Kyaw, J. Gutierrez and A. Ghobakhlou, "*A systematic review of deep learning techniques for phishing email detection*," *Electronics*, vol. 13, no. 19, p. 3823, 2024.
DOI: https://doi.org/10.3390/electronics13193823

[8] N. Rifat, M. Ahsan, M. Chowdhury, and R. Gomes, "*Bert against social engineering attack: Phishing text detection*," *2022 IEEE International Conference on Electro Information Technology (eIT)*, pp. 1-6, 2022.
DOI: https://doi.org/10.1109/eIT53891.2022.9813922

[9] K.S. Jishnu and B. Arthi, "*Enhanced phishing URL detection using leveraging BERT with additional URL feature extraction," 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 1745-1750, 2023.
DOI: https://doi.org/10.1109/ICIRCA57980.2023.10220647

[10] V. Sanh, L. Debut, J. Chaumond and T. Wolf, "*DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*," arXiv preprint arXiv:1910.01108, 2019.
DOI: https://doi.org/10.48550/arXiv.1910.01108

[11] X. Jiao, Y. Yin, L. Shang, X. Jiang, X. Chen, L. Li, F. Wang and Q. Liu, "*Tinybert: Distilling bert for natural language understanding*," *arXiv*, 2019.
DOI: https://doi.org/10.48550/arXiv.1909.10351

[12] M. Safran and A. Musleh, "*PhishingGNN: Phishing Email Detection Using Graph Attention Networks and Transformer-Based Feature Extraction*," *IEEE Access*, no. 99, pp. 1-1, 2025.
DOI: https://doi.org/10.1109/ACCESS.2025.3592135

[13] L.R. Kalabarige, R.S. Rao, A. Abraham and L.A. Gabralla, "*Multilayer stacked ensemble learning model to detect phishing websites*," *IEEE Access*, vol. 10, pp. 79543-79552, 2022.
DOI: https://doi.org/10.1109/ACCESS.2022.3194672

[14] B.A. Shajilal, "*A Hybrid Approach for Detecting Phishing Mails Using Textual, Content, and URL Analysis with Ensemble Learning*," *National College of Ireland*, 2024.
URL: https://norma.ncirl.ie/id/eprint/8292