JCBI
Journal of Computing & Biomedical Informatics Volume 09 Issue 01
ISSN: 2710 - 1606 2025

Research Article
https://doi.org/10.56979/901/2025

A Feature-Level Hybrid Model Approach for Automated Phishing Email Detection

Ayesha Saman[1], and Saad Rasool[1*]

[1]Department of Computer Science, Bahauddin Zakariya University, Multan, 60000, Pakistan.
[*]Corresponding Author: Saad Rasool. Email: saadrasool2000@gmail.com

Received: April 08, 2025 **Accepted:** May 27, 2025

Abstract: Phishing emails continue to pose a serious cybersecurity risk that entails tricking the recipients and getting them to disclose personal and financial information, including their login details. The timely malware attack detection helps to reduce the impact and guarantee the safety of users. The literature has over time utilized different methods in phishing detection, such as rule based, classical machine learning and deep learning methods, but most of them have just used the textual content of the email and have ignored other features such as the URLs, which can also provide important contexts. In order to fill this gap, we present a complex framework aimed at the detection of phishing emails, based on the joint usage of textual and structural information. Based on an experimental natural language collection of email bodies and the presence or absence of URLs as labels, the experiments were designed to use the TF-IDF characteristics to ensure that classical machine learning models, such as Logistic Regression, Support Vector Machine, and Random Forest, offered excellent baselines. Then, we presented a feature-level hybrid deep learning model, which is built upon DistilBERT, called Hybrid DistilBERT, where semantic representations of email messages text along with URL presence are introduced as inputs. This hybrid model has been proven practical in its nature to merge the context of transformer models with meta-level data and has provided substantial advancement in predictive performance. Experimental results showed that our proposed model has the best classification accuracy, reaching 99.1% as compared to traditional models, and this is testimony to the applicability of our hybrid design. These findings support the worth of combining both the content and structural clues to make the phishing detection more resilient and precise.

Keywords: Cybersecurity; Deep Learning; Phishing Detection; Spam Emails; Cyber Attacks; Hybrid Model

1. Introduction

Electronic mail, or email, is one of the most popular things on the Internet. It lets you write and receive emails from and to anyone in the world who has an email address. Email uses more than one system from the TCP/IP suite. While a scam email is a type of spam email that is sent to get personal information like passwords, debit/credit card numbers, bank account numbers, etc., people are tricked into giving this information. People who give out their information through these phishing emails could end up losing money because their name was stolen [1]. As more and more people use the Internet to do business, Internet fraud poses a bigger threat to people's safety and privacy [2-3]. Spear phishing attacks are typically launched using ostensibly genuine emails that may include harmful links, files, or solicitations for confidential information [4]. Phishing emails frequently direct individuals to harmful websites, resulting in the disclosure of personal information to the perpetrators. To mitigate these issues, spam and phishing email classifiers are extensively employed [5-7]. Blacklisting, categorized as a list-based filter, is an effective approach to combat phishing emails. Since email communication is rampant, phishing emails have become

one of the major avenues through which such malicious operations are effected. Phishing emails are managed to escape most of the already employed filters because of their changing patterns and anti-detecting tactics as the users become more aware of the problem of security [8]. These emails are explicitly designed to extract personal information regarding their colleague or organization. Such emails are classified as spear-phishing targeted communications [9-10]. This requires timely and high quality identification of the phishing messages in order to reduce monetary losses, ensure privacy of users and safety of digital infrastructures. There are a lot of email providers that flag messages as possibly fraudulent without explaining why. Therefore, recipients may miss a crucial communication or click on a harmful link without realizing it is a phishing attempt [11]. Anti-phishing training programs theoretically enhance users' ability to distinguish phishing emails from authentic ones by augmenting their phishing knowledge [12]. A large number of studies have been provided in this field through the scope of rule-based approaches, machine learning classifiers, and deep learning. Spam emails deplete time and resources and can lead to the dissemination of malware, phishing attempts, and the theft of personal information, thereby presenting a significant issue to individuals and companies alike.

Consequently, spam email identification is an essential function in email security, safeguarding individuals and companies against such communications. In order to strengthen protection against phishing emails, numerous proposals of studies have been introduced which are based on methods of machine learning and analysis of emails and classification [13]. These approaches statistically select features related to phishing on the headings and contents of the message in order to classify them in a binary (ham/spam) decision [14-17]. Spam email detection necessitates the application of NLP to ascertain its classification as spam or not. Text classification can be explained as NLP that helps a model to capture the contextual information and the linguistics, thus making a better prediction of the exact class [18]. Various methods, including classic machine-learning models, deep-learning models, and transformer-based models, have been developed throughout the years to detect spam emails [19]. Improving phishing detection models is crucial due to the very misleading nature of phishing emails [20]. The conventional phishing detection methods like black list based filtering, rule based or heuristic pattern matching have proven to be ineffective against new and advanced phishing attacks. Such methods tend to depend on pre-existent signatures or set rules, which are easily avoided by making some slight alterations to content or structure of the phishing email. As a result, there has increasingly been the need to have more adaptive, smart and flexible solutions in detection. Conventional detection techniques, dependent on external indicators such as dubious URLs, domain credibility, or overt brand imitation, have proven effective against traditional phishing assaults [21]. ML is a subset of Artificial Intelligence focused on the automatic instruction of computers in new activities [22, 23]. ML and DL have been thoroughly examined in the literature as effective techniques for the automatic identification of phishing emails [24, 25]. To deal with phishing's complexity, ML and AI are two that come to mind as possible ways to improve defenses [26]. The utilization of AI models for prioritizing feature selection is increasingly recognized as a promising approach in phishing email detection [27], mostly due to researchers' discontent with the efficacy of malware detection reliant exclusively on engineering features derived from a CSV of static file attributes [28].

Recent ideas such as NLP and ML have been found useful to overcome this problem. NLP-based models can determine patterns and differentiate between genuine and misleading messages by examining the nature and makeup of emails. With NLP classification of text, a model can have an insight into contextual patterns to support getting the right classification since the model will be able to predict it with more accuracy. These models are not restricted to strict rules and are, therefore, able to evolve with time and to generalize to unknown phishing methods, thus constituting a worthwhile selection in the present-day structure of email security. In certain conditions of phishing detection, text might not be sufficient enough. Emails are also likely to have other indicators which are of great importance of detecting phishing scams; here is the inclusion or omission of embedded URL links. This increased need of more exhaustive detection machineries has led to the attempts by the researchers to pursue hybrid solutions involving a combination of non-textual features and textual features. The objectives of those solutions are to capture the comprehensive picture of structure and content of an email, which will enhance the model to identify more complex phishing schemes. This is a way that we are pursuing in this paper by complementing the power of transformer models and classical ML models, including both semantic text features, and

structural features like URLs, to enhance the accuracy of phishing detection. Our hybrid model is described in terms of related work, methodology, experimental setup and performance analysis in the sections below.

2. Related Work

In its path towards finding the appropriate methods of identifying phishing emails, a significant amount of research has been possible over the years as people continue to realize the need to fight this menace. In earlier research work, approaches were narrower in that they mainly focused on heuristics-based methods as rule-based systems or manually engineered features, that check suspicious keywords, misspelled words, header anomalies, or blacklisted URLs. Dependence on such training transitions phishing detection from an automated procedure to a human-operated one, which is prone to errors, particularly when a user makes mistakes due to distraction or forgetfulness. Paul & Bartlett et al. [29] examined the treatment of this detection type as a natural language processing problem and adjusted the training pipelines accordingly. They provided a dataset including annotated labels derived from the kinds of signals that users are generally required to recognize in such training. They additionally introduced baseline classifier models trained on these categories of labels. Their findings suggested that collaborative methods including ML models and people might yield greater accuracy than independent identification and could also reduce the cognitive burden on human users when detecting phishing in emails. Also, Meléndez & Ptaszynski et al. [30] compared the functionality of classical models with transformer models on the classification task of determining whether an email is phishing or not. They also obtained the fact that the transformer models (especially distilBERT, BERT and roBERTa) showed an outstanding ability in performing compared to conventional models such as Logistic Regression (LR), RF, SVM and Naive Bayes (NB). Upon choosing a vast and reliable dataset of emails and preprocessing and optimization techniques, the process entailed optimizing the best outcome out of roBERTa, which managed to identify phishing emails to an exceptional level of 0.9943. Although they were also successful, classical models also fared slightly below; SVM was the best with an accuracy of 0.9876. The findings underscore the significance of advanced text-processing techniques and the capability of transformer models to enhance email security by preventing phishing attacks.

Chanis & Arampatzis et al. [31] enhanced the notion of content-based phishing email detection involving the combination of the stylometric features and the widely-applied vectorization methods alongside the application of the classifier stacking. With the aid of various stylometric features, they test many approaches to integrating them with the text in vectorized form, and also various stacking patterns with the machine learning algorithms. Their results indicate that the set methods are always better off compared to vectorization only in an imbalanced dataset and less in a balanced dataset. In particular, they obtained an F1 score of 0.9843 on the balanced and 0.9656 on the imbalanced set by stacking the results of many differentially trained classifiers, trained individually on the content features and stylometric features, and doubled baselines on both sizes of the imbalanced set by more than 2.2%. Based on a large-scale phishing email dataset, Chinta et al. [22] tested the development and evaluation of high-end ML models to identify phishing emails. It took several ML models such as Convolutional Neural Network (CNN), XGBoost, Recurrent Neural Network (RNN), and SVM. The optimum solution was proposed with the help of BERT-LSTM hybrid model. The BERT-LSTM model achieved excellent performance with an F1-score of 99.24, a recall of 99.55%, precision value of 99.61%, and accuracy of 99.55%. The effectiveness of BERT-LSTM compared with the existing models like Naive Bayes, RNN, and SVM was efficient in detecting phishing emails and is therefore superior to the existing models. Moreover, the training and test evaluations showed low over-fitting and reliability in generalization. This paper highlighted the possible BERT-LSTM value in real-time phishing email detection systems, which can provide a solid method of phishing threat reduction. Atawneh et al. [32] tested out the application of deep learning methods, such as CNNs, long short-term memory (LSTM) networks, RNNs, and BERT in the domain of detecting email phishing attacks. The dataset containing phishing emails and clean email was used, and the feature set concerning those emails was derived in an approach based on NLP. The dataset was used to train and test the proposed deep learning model, and it was established that it has the potential of having high accuracy in email phishing at least paralleled with other state-of-the-art-research and the optimal tricks were the utilization of BERT and LSTM and accuracy was 99.61%. Their findings indicated how deep learning has a potential of enhancing the detection of email phishing and security against this widespread menace.

Murti & Naveen et al. [33] conducted a detailed analysis of existing ML-based classifiers that are effective at detecting phishing email. They used an actual world data set in Kaggle that contains the real proportions of genuine and phishing mail. Next, Exploratory Data Analysis (EDA) is done to have a better idea of the data and find apparent errors and outliers to aid in the process of detection. The dataset is trained on several ML methods, such as NB, SVM, RF, Decision Tree (DT) and KNN. Measure of the performance of ROC and AUC are applied to check the models. Also, the period of time that was spent training and testing of the model will be tried to draw a conclusion about the effectiveness of the algorithm in real-time. They demonstrated that the highest precision, recall, and f1-score are in RF and DT classifier. Similarly, Mehdi & Verma et al. [34] created an augmented phishing/legitimate email dataset, with the help of various adversarial text attacks. Then the models were retrained using the adversarial dataset. Results depicted that the models attained better accuracy and F1 score in later attacks. In a different experiment a fine-tuned GPT-2 model was used to create synthetic phishing emails. There was retraining of the detection model by using a newly created synthetic dataset. They later noticed that the application of black box tactics did not promote the model accuracy and robustness to a significant level. In the final experiment, they offered a defensive mechanism that they suggest can be used to label their adversarial examples to their actual labels using a KNN solution in which our prediction has a total accuracy of 94%. Moreover, Brindha & Nandagopal et al. [35] proposed an Intelligent Cuckoo Search (CS) Optimization Algorithm over a Deep Learning-based Phishing Email Detection and Classification (ICSOA-DLPEC) model. Their proposed ICSOA-DLPEC model will have the aim of effectively separating the emails into either legitimate or phishing mails. Following the Gated Recurrent Unit (GRU) model, the CS algorithm is adopted to identify and label phishing emails. Moreover, the parameters based on which the GRU model works are optimized using the CS algorithm. The effect and functionality of the suggested proposed model was effectuated through experimentation utilizing a benchmark data set, and the outcomes of which were evaluated in various dimensions. They substantiated the high efficiency of the proposed model compared with other currently available methods, which demonstrated a maximum value of accuracy of 99.72%. Zhao & Jin & et al. [36] examined the progression of phishing email content and its influence on the efficacy of detection methods. They proposed Fewshing, a few-shot learning methodology for the detection of phishing emails. Their experimental findings indicate that Fewshing attains an F1 score of 92.4% and an accuracy of 98.6% on the constrained and imbalanced training datasets, hence illustrating Fewshing's efficacy in identifying phishing emails.

Similarly, Uyyala et al. [37] initially examined the email structure and subsequently developed a novel phishing email detection model utilizing an enhanced CNN framework with multilevel vectors and attention mechanisms. This model concurrently analyzed emails at the header, body, character, and word levels. To assess the efficacy of the model, they utilized an unbalanced dataset that more accurately reflects real-world conditions for experimentation and evaluation. The model yielded good results. The NLP preprocessing steps were incorporated in a phishing email detection model suggested by the Hilani & Nassih et al. [38] along with a one-dimensional CNN. They modeled textual data that can detect phishing attacks with extreme accuracy. Experimentation shows that the presented model delivers the accuracy of 99.99% which is higher than the majority of current methods with respect to the efficiency and reliability. In order to confirm the practicality of their model they incorporated into a mailing system such that it became robust in the identification of phishing emails under real life condition. Different researchers have tried their best in phishing email detection using both the traditional machine learning and deep learning models. Nevertheless, majority of the current methods either consider only textual features or analyze the features based on URLs without effectively utilizing the semantic and structural cues that are available in phishing emails. This is the missing link between the textual and URL level features, something that has restricted performance of existing detection systems. Aiming to overcome this, we propose a feature level hybrid model that combines both the semantic capabilities of DistilBERT and informative URL based features resulting in a more solid and precise phishing detection architecture.

3. Methodology

The methodology adopted for this work aims at detecting phishing emails well, which takes advantage of textual content as well as the features that accompany textual content. The suggested method includes several steps on the way to achieving the desired result: the recollection of the datasets, their

preprocessing, the development of the model, and its validation. This section gives the systematic process flow that started with data acquisition and culminated in the creation of a feature-level hybrid classification system. Moreover, we validate the model's performance by conduction a comparative analysis against classical machine learning models.

3.1. Dataset Description

The dataset employed in this study consists of a considerable number of emails samples, which is taken from an online repository named, Kaggle. In particular, the data is comprised of nearly 82,500 emails, out of which 42,891 are classified as spam/phishing and 39,595 are legitimate. Such emails were accumulated based on credible sources. The dataset of these emails was made up of some credible sources including Enron dataset, Ling-Spam dataset, CEAS 2008 corpus, Nazario phishing corpus, Nigerian fraud emails, and SpamAssassin dataset. This aggregated data can be then be applied to train models that have more generalization and effectiveness in real-world conditions [26]. In the email feature, each of the email entries has a number of attributes such as the subject line, the body of the mail, sender details, receiver details, timestamps, and a binary signal that shows whether the email is phishing or not. The mentioned dataset is diverse and well-balanced, and it is a good basis on which classical and deep learning models can be trained and evaluated regarding phishing detection. Incorporating diverse data points of various datasets increases the generalizability and the strength of the models arrived at.

3.2. Data Preprocessing

Phishing emails detection depends heavily on the quality of preprocessing because the real world email data is noisy and unstructured. Tokenization and vectorization techniques are essential for training machines to understand human language [39]. In our research, we used several text and feature-level preprocessing standards that are appropriate to classical machine learning and transformer-based models. In the case of textual content, usual preprocessing operations were undertaken, such as lowercasing, HTML tag, punctuation, number and stopword removal. The process of Tokenization was done using the *DistilBERTTokenizer*, which was compatible with the transformer-based hybrid model. And the tokenized inputs were padded and truncated to a maximum sequence length so that the input dimensions would be the same.

Besides textual data, the existence of URLs in emails was also regarded as the useful indicator of phishing detection. Suspicious URLs were extracted and encoded as additional numerical feature, in which URLs with the presence of suspicious URLs were singly binary encoded and subsequently concatenated to the BERT embeddings as a hybrid modeling approach. The last data was divided into training, validation, and test dataset through a stratified division that did not distort the distribution of classes. This preprocessing configuration provided a powerful and reliable format to all of these, the transformer model and ML models, we experimented with.

3.3. Feature-Level Hybrid Approach

We introduce such feature-level hybrid schema that combines deep semantic features of textual content of the email together with handcrafted numeric features of the email URLs. The gist of hybridization is that it uses the complementary abilities of transformer-based language comprehension and conventional URL-based signs to enhance phishing email accuracy levels.

3.1.1. Hybrid DistilBERT Architecture

In order to create contextualized embeddings of the text of the email body, we use DistilBERT, a lightweight but powerful modification of BERT. This model exhibits accelerated training capabilities and inference times, diminished memory and computing requirements, while maintaining performance with a slight reduction [40]. The hybrid architecture of this model is shown below.

- *Input Layer*: It starts by taking the raw text of the email, including subject lines and body text, into a special preprocessing pipeline. This also incorporates the tokenization of DistilBERT tokenizer that maintains a contextual integrity of sentences by using the sub-word segmentation and special symbols.
- *Transformer Encoder Layer*: This sequence is in turn translated into a tokenized form and fed to a pre-trained DistilBERT and fine-tuned to specifically detect phishing emails. That which is of interest as the output is the *[CLS]* token embedding, which is a 768-element dense vector that acts as some global representation of the entire input sequence. This embedding captures:
  - Deep semantic information from the email text.
  - Intent indications like urgency, manipulation, or impersonation.

- Contextual phishing markers that are often subtle and not apparent through shallow methods.

- *Feature Enhancement Layer*: In order to maximize the semantic representation in learning of DistilBERT, we add more dense layers that will take the embedding as an input and become non-linear projectors transforming it into a feature space that is richer. Specifically:
- The Dropout layer (rate = 0.3) is used to circumvent the overfitting problem of *[CLS]* embedding.
- Then one or two fully connected Dense layers with ReLU activation is added so as to learn more complex hierarchical patterns and deeper class separability of the encoded features by the model.

- *Output Layer*: The architecture is concluded by a sigmoid enabled output layer (binary classification). It passes the discovered features through a transformation that measures the probability of the email input as phishing or legitimate as a single scalar, a probability score.

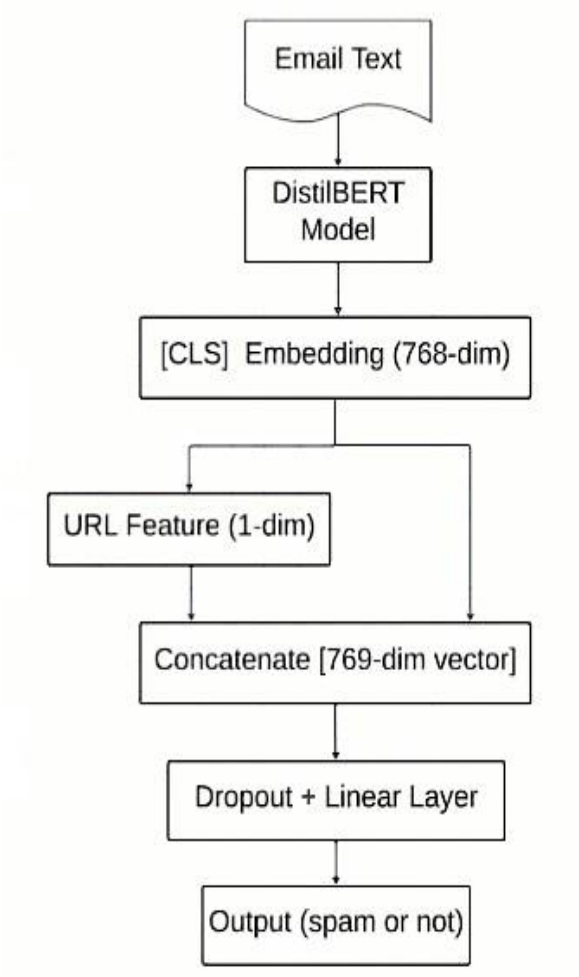

Figure 1. Hybrid DistilBERT Architecture.

In particular, the output of the *[CLS]* token of the final hidden layer of DistilBERT constitutes a reduced representation of the input sequence. This is a representation of hidden complex linguistic and context information provided by a phishing or a genuine mail. In addition to this, we also derive an integer-valued URL feature per email which in essence measures the level of suspicious nature or authenticity of its corresponding hyperlink. That can consist of flagged signals like domain reputation, length, or presence of suspicious characters: or heuristic learned during preprocessing. This scalar feature is later concatenated with the 768 dimensions *[CLS]* embedding of DistilBERT to get composite feature vector of dimension 769. The given hybrid vector is routed to a dropout layer to prevent it against overfitting and directed to fully connected classification layer. The general loss function is end-to-end training architecture with the standard cross-entropy and optimized by Adam optimizer. Interestingly, the hybrid approach can retain its performance with distorted or deceptive phishing data by integrating the semantic features and structural features, therefore, representing a structural analysis of phishing texts but coherent with the interpretation of phishing data.

3.3.2. Lexical Feature-Based Setup for ML Models

We also tried several conventional ML classifiers such as the LR, SVM, as well as RF. These classical models were not altered or enhanced, instead, they were used as a baseline of classifiers to check the predictive potential of conventional textual features. Raw text of the emails bodies was run through the vectorization procedure, i.e. Term Frequency-Inverse Document Frequency (TF-IDF) technique that represents the textual data as numerical feature vectors representing the importance of the terms in the corpus.

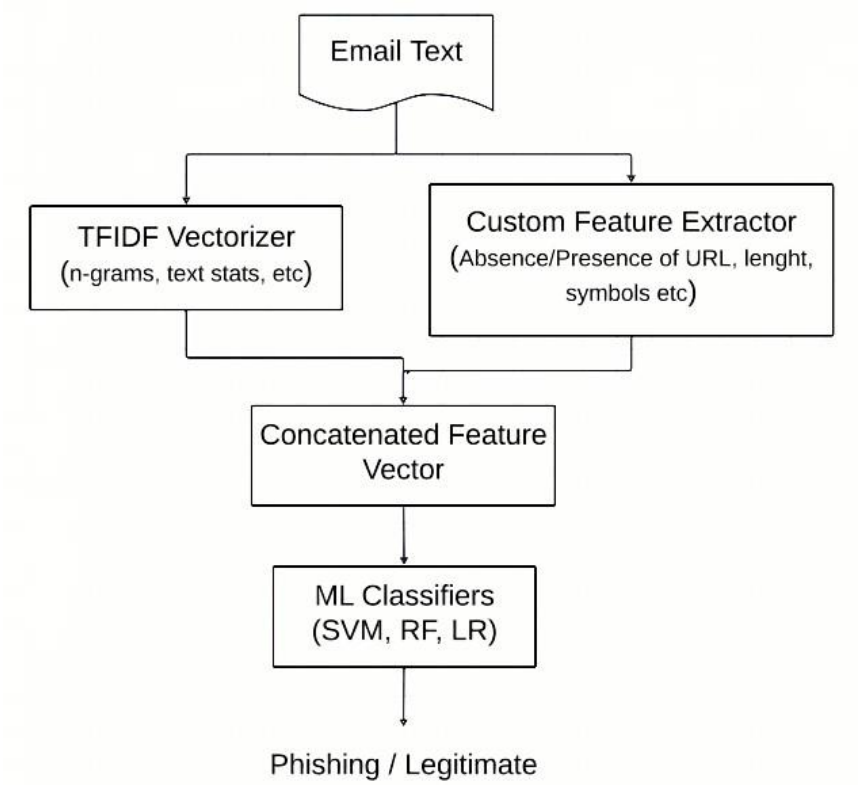

Figure 2. Lexical Feature-Level Setup.

In the case of classical machine learning models, we have selected feature-level pipeline with identifying strong lexical indicators within phishing email. In particular, email content was used in the TF-IDF vectorized as the high-dimensional sparsity. The features concatenated the TF-IDF vectors and obtained a unified feature space. This augmented representation was in turn supplied to the ML classifiers so that the former could take advantage of both statistical patterns and manually selected phishing indicators. In contrast with the hybrid model, these classifiers neither used URL based features nor did they have any advantage of having any sense of context based language understanding. They could only feed in TF-IDF vectors of text data in emails. Such a configuration at the feature level was introduced into each of the machine learning models using the versions provided by Sci-kit learn and their results were captured to compare the performance.

Prior differences in architecture and in the input modality make it obvious that our hybrid DistilBERT model was developed in a way unique to the given implementation, where we combine both a deep semantic representation of a language and organized information at the level of URLs but the classical models were simply TF-IDF-based text classifiers.

4. Results & Discussion

Various evaluation parameters which are commonly utilized were used to assess the performance of the proposed and utilized models, namely, accuracy, precision, recall, and F1-score. The metrics give a full picture of the capability of each model to detect phishing emails and reduce the number of false alarms. Few ML models were transferred and compared against the phishing email dataset, as well as the proposed hybrid DistilBERT model. The selection of these models was also strategic in the sense that they have already built impressive results in phishing detection in previous studies. The inclusion of them allows

implementing a meaningful and fair comparative study to evaluate the gains of our feature-level hybrid method constructed on DistilBERT architecture.

All the models were trained and tested based on the same feature space found with the help of the TF-IDF technique to the content of email. The hybrid model has also included the structured URL-based features on feature level whereas the classical models have been based on the TF-IDF features. Notably, these classics models remained unchanged except being standard-tuned, and they were used in setting comparative baselines. With consistent evaluation conditions, all models, the hybrid one being included, were trained and tested based on a train/test ratio of 80:20. Each data was shuffled and then split to eliminate the issue of class imbalance bias in training or the test set. The performance measures of each model in terms of four fundamental assessment metrics are as shown below:

Table 1. Evaluation Results of Models.

	Accuracy	Precision	Recall	F1 Score
Hybrid DistilBERT	99	98	99	98
SVM	97	98	96	97
RF	95	96	95	96
LR	94	95	95	95

As evidenced by the performance comparison, the proposed hybrid DistilBERT model considerably outwits the baseline machine learning models in terms of the tasks of phishing email classification. Such an increase in performance can largely be due to two main similarities, mainly: (1) the transformer-based semantic interpretation of the textual material through the means of DistilBERT, and (2) the dependency of feature-level structured URL indicators. DistilBERT extracts deep contextual dependencies, subtle word interactions and syntactic relationship structure in the contents of the mail-capabilities that cannot be trained inherently in classical models such as SVM, RF and LR because they rely on sparse and high dimensional representations (the TF-IDF vectors). The hybrid configuration also maxes the semantic signal by tokenized URL-based feature together with the BERT-extracted sentence embeddings. This feature-level combination enables the model to evaluate the natural language purpose and the obfuscated malicious details in the URL and, hence, extremely resistant to advanced phishing tactics that bind to avoid detection in the context of textual characters.

Consequently, the hybrid model reached an accuracy of 99.1% with a balanced precision 98% and recall 99% resulting in not only low false positive rate but also low false negative rate which is crucial in an actual field of cybersecurity. On the other hand, SVM also exhibited competitive performance with 97% accuracy and the highest measure of precision 98% among the classical models. It is very useful in generalizing on text classification problems because it can detect optimum hyperplanes in very large feature space. SVM classifier has the ability to surpass some models in certain NLP tasks [41]. Nevertheless, its slightly lower recall rate of 96% shows that the miss rate of phishing emails is a bit bigger than the hybrid one. Bar graph was created to demonstrate the metrics of all applied models visually in a comparative way.

The below visualization distinctly shows the effective image processing the proposed Hybrid DistilBERT model achieves as compared to the conventional machine learning models. Random Forest, although being equally successful, demonstrated a bit lower rates of 95% accuracy, because of its ensemble architecture, which, on the one hand, is effective due to covering more diverse decision boundaries, and on the other hand, lost some details of semantics of email contents. The simplest of the models, LR gave 94% accuracy and was stable and showed trends, but was unable to deal with the complex relationships and not even semantics because it was linear and only used TF-IDF features.

Essentially, all of these models demonstrated positive results, the hybrid DistilBERT model is the best because it has an in-depth contextual understanding and a smart feature addition. This technical discussion highlights the fact that hybrid deep learning architectures are better suited towards detection of phishing attempts due to the requirements of detecting phishing in dynamic deployments where the attackers are continually adapting their methods by obfuscating intent by exploiting not only the textual entropy but also creating deceptive URLs.

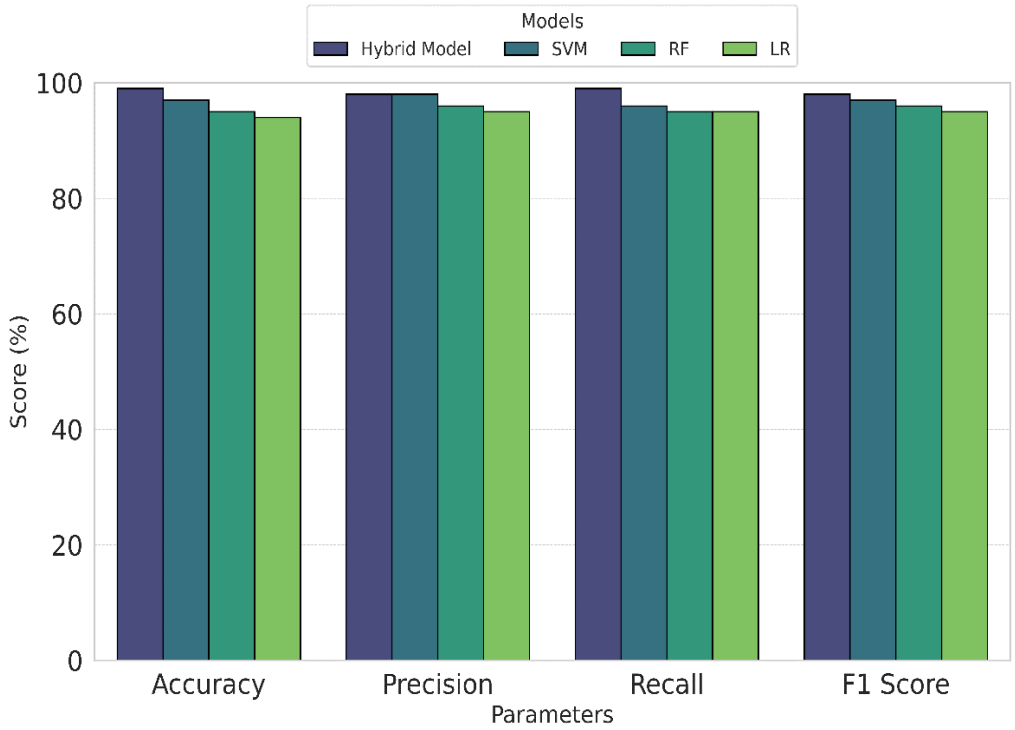

Figure 3. Performance Evaluation Graph.

5. Conclusion

By delving into both traditional machine learning methods and an innovative hybrid deep learning strategy, we were able to tackle the pressing issue of phishing email detection in this research. Since the phishing attacks are getting more sophisticated, in particular, those ones that use textual content and URLs embedded in the email messages, the models which can track several data modalities are acutely required. We have employed a complete phishing set that has been downloadable where the email repositories are several containing more than 82,000 emails comprising of phishing and genuine data full of deposits. In our preprocessing pipeline we would clean the text of the emails, code the labels and apply TF-IDF to transform the textual data into a numerical vectors which could be used in machine learning models. Along with it, it kept the features of URL, and then it used to add text insights to activations based on the URL in the form of a feature-level hybrid configuration. A number of conventional ML algorithms, SVM, RF, and LR, were trained using the data transformed with TF-IDF. These models have been chosen because of their high results in previous phishing detection studies and also as a basis of comparison. We used the same 80:20 train-test split on each model and accuracy, precision, recall, and F1 score measures were used in evaluation.

We sought to take the performance to the next level and therefore presented a Hybrid DistilBERT model within which we use DistilBERT transformer to generate the contextual embeddings of the textual element within the email contents. These were combined with handcrafted numerical features (features created using URLs) and formed a hybrid input to a fully connected neural layer used in completing a final classification. This conjunction enabled the model to take advantage of both semantic patterns and numerical cues that resulted in the improved detection performance. The revealed results of the experimental testing indirectly proved the better performance of our proposed approach. Hybrid DistilBERT model recorded the best percent of accuracy 99.1%, precision of 98%, recall of 99%, and F1 score of 98%. On the conversely, SVM, the most successful classical model, attained an accuracy of 97%. Such gains justify the success of integrating deep contextual knowledge with ordered feature facts. The oxymoronic performance of the hybrid model is explained with its ability to model subtle language pattern of email text along with assimilating essential metadata features that are commonly used and contain a phishing intent.

Overall, the proposed approach presents a powerful, efficient, and scalable high-performance framework of phishing email detection that is significantly better than conventional ML-based methods.

The trend of a hybrid feature-level combination approach, especially the transformer-based model, such as DistilBERT, can be a good way of future phishing detection frameworks. In further research, this environment could be extended to accommodate the use of attachments and email header and visual elements, as well as to be compared to live phishing attacks to determine the applicability of deployment.

References

1. Nguyen NT, Childress FD, Yin Y. Debate-Driven Multi-Agent LLMs for Phishing Email Detection. In2025 13th International Symposium on Digital Forensics and Security (ISDFS) 2025; (pp. 1-5). IEEE.
2. Kulal D, Shiferaw L, Niyaz Q. Phishing Email Detection through Machine Learning and Word Error Correction. In2025 17th International Conference on COMmunication Systems and NETworks (COMSNETS) 2025; (pp. 1299-1304). IEEE.
3. Yasin A, Abuhasan A. An intelligent classification model for phishing email detection. arXiv preprint arXiv:1608.02196. 2016.
4. Birthriya SK, Ahlawat P, Jain AK. Detection and prevention of spear phishing attacks: A comprehensive survey. Computers & Security. 2025; 104317.
5. Harikrishnan NB, Vinayakumar R, Soman KP. A machine learning approach towards phishing email detection. InProceedings of the anti-phishing pilot at ACM international workshop on security and privacy analytics (IWSPA AP) 2018; (Vol. 2013, pp. 455-468).
6. Meléndez R, Ptaszynski M, Masui F. Comparative Investigation of Traditional Machine-Learning Models and Transformer Models for Phishing Email Detection. Electronics. 2024; 13(24):4877.
7. Rashed S, Ozcan C. A comprehensive review of machine and deep learning approaches for cyber security phishing email detection. Al-Iraqia Journal for Scientific Engineering Research. 2024; 3(3):1-2.
8. Altwaijry N, Al-Turaiki I, Alotaibi R, Alakeel F. Advancing phishing email detection: A comparative study of deep learning models. Sensors. 2024; 24(7):2077.
9. Vergelis M, Shcherbakova T, Sidorina T, Kulikova T. Spam and phishing in 2018. Secure List. 2019.
10. Chanis I, Arampatzis A. Enhancing phishing email detection with stylometric features and classifier stacking. International Journal of Information Security. 2025; 24(1):15.
11. Koide T, Fukushi N, Nakano H, Chiba D. Chatspamdetector: Leveraging large language models for effective phishing email detection. arXiv preprint arXiv:2402.18093. 2024.
12. Sturman D, Auton JC, Morrison BW. Security awareness, decision style, knowledge, and phishing email detection: Moderated mediation analyses. Computers & Security. 2025; 148:104129.
13. Zhang J, Wu P, London J, Tenney D. Benchmarking and Evaluating Large Language Models in Phishing Detection for Small and Midsize Enterprises: A Comprehensive Analysis. IEEE Access. 2025.
14. Beaman C, Isah H. Anomaly detection in emails using machine learning and header information. arXiv preprint arXiv:2203.10408. 2022.
15. Mughaid A, AlZu'bi S, Hnaif A, Taamneh S, Alnajjar A, Elsoud EA. An intelligent cyber security phishing detection system using deep learning techniques. Cluster Computing. 2022; (6):3819-28.
16. Nabeel M, Altinisik E, Sun H, Khalil I, Wang H, Yu T. CADUE: Content-agnostic detection of unwanted Emails for enterprise security. InProceedings of the 24th International Symposium on Research in Attacks, Intrusions and Defenses 2021; (pp. 205-219).
17. Sonowal G. Phishing email detection based on binary search feature selection. SN Computer Science. 2020; 1(4):191.
18. Rasool S, Hanif I, Rasool Q, Afzal H, Mufti MR. A Text Mining Approach for Automated Case Classification of Judicial Judgment. Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences. 2025; 62(1):41-51.
19. Prasad R. Phishing email detection using machine learning: A critical review. In2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT) 2024; (Vol. 5, pp. 1176-1180). IEEE.
20. Sharma AK. A Systematic Review on Phishing Attacks Detection Techniques based on Machine Learning. In2025 3rd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS) 2025; (pp. 930-937). IEEE.
21. Opara C, Modesti P, Golightly L. Evaluating spam filters and Stylometric Detection of AI-generated phishing emails. Expert Systems with Applications. 2025; 276:127044.
22. Chinta PC, Moore CS, Karaka LM, Sakuru M, Bodepudi V, Maka SR. Building an Intelligent Phishing Email Detection System Using Machine Learning and Feature Engineering. European Journal of Applied Science, Engineering and Technology. 2025; 3(2):41-54.
23. Casido SO, Ballaho JC, Arip EI. Tree-Based Machine Learning Models for Phishing Email Detection. In2025 12th International Conference on Information Technology (ICIT) 2025; (pp. 381-386). IEEE.
24. Alhuzali A, Alloqmani A, Aljabri M, Alharbi F. In-Depth Analysis of Phishing Email Detection: Evaluating the Performance of Machine Learning and Deep Learning Models Across Multiple Datasets. Applied Sciences. 2025; 15(6):3396.

25. Arthy S. A Hybrid Machine Learning Approach for Securing Emails: Phishing Detection and Prevention. In2025 3rd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA) 2025; (pp. 1-6). IEEE.
26. Al-Subaiey A, Al-Thani M, Alam NA, Antora KF, Khandakar A, Zaman SA. Novel interpretable and robust web-based AI platform for phishing email detection. Computers and Electrical Engineering. 2024; 120:109625.
27. Bandahala T, Suhaili NS, Monabi K, Suhuri H, Iboh S, Jaujali M, Bagindah N, Musin M, Shaik NA, Adjaraini K, Tahil SK. The Role of Artificial Intelligence in Detecting and Preventing Phishing Emails. International Journal of Innovative Science and Research Technology. 2025.
28. Bauskar SR, Madhavaram CR, Galla EP, Sunkara JR, Gollangi HK. AI-driven phishing email detection: Leveraging big data analytics for enhanced cybersecurity. Library Progress International. 2024; 44(3):7211-24.
29. Paul M, Bartlett G, Mirkovic J, Freedman M. Phishing email detection using inputs from artificial intelligence. arXiv preprint arXiv:2405.12494. 2024.
30. Meléndez R, Ptaszynski M, Masui F. Comparative Investigation of Traditional Machine-Learning Models and Transformer Models for Phishing Email Detection. Electronics. 2024; 13(24):4877.
31. Chanis I, Arampatzis A. Enhancing phishing email detection with stylometric features and classifier stacking. International Journal of Information Security. 2025; 24(1):15.
32. Atawneh S, Aljehani H. Phishing email detection model using deep learning. Electronics. 2023; 12(20):4261.
33. Murti YS, Naveen P. Machine learning algorithms for phishing email detection. Journal of Logistics, Informatics and Service Science. 2023; 10(2):249-61.
34. Mehdi Gholampour P, Verma RM. Adversarial robustness of phishing email detection models. InProceedings of the 9th ACM international workshop on security and privacy analytics 2023 Apr 26 (pp. 67-76).
35. Brindha R, Nandagopal S, Azath H, Sathana V, Joshi GP. Intelligent Deep Learning Based Cybersecurity Phishing Email Detection and Classification. Computers, Materials & Continua. 2023; 74(3).
36. Zhao P, Jin S. Fewshing: A few-shot learning approach to phishing email detection. In2024 IEEE 4th International Conference on Software Engineering and Artificial Intelligence (SEAI) 2024; (pp. 371-375). IEEE.
37. Uyyala PR. Phishing email detection using CNN. J. Eng. Technol. Manag. 2024; 72:1046-51.
38. Hilani M, Nassih B, Lmati I, Balouki Y, Amine A. Phishing Email Detection Using NLP and CNN Model. InInternational Conference on intelligent systems and digital applications 2025; (pp. 203-212). Cham: Springer Nature Switzerland.
39. Sukanya G, Priyadarshini J. Analysis on word embedding and classifier models in legal analytics. InAIP Conference Proceedings 2024; (Vol. 2802, No. 1, p. 140001). AIP Publishing LLC.
40. Meléndez R, Ptaszynski M, Masui F. Comparative Investigation of Traditional Machine-Learning Models and Transformer Models for Phishing Email Detection. Electronics. 2024; 13(24):4877.
41. Basile A, Dwyer G, Medvedeva M, Rawee J, Haagsma H, Nissim M. Simply the best: minimalist system trumps complex models in author profiling. InInternational Conference of the Cross-Language Evaluation Forum for European Languages 2018; (pp. 143-156). Cham: Springer International Publishing.