Robust Spam Email Classification Using an Ensemble of Traditional ML and BERT-Based Deep Learning Models

Mushtaq Talib Mahdi*

University of Babylon, Hilla, Iraq
*Corresponding author, email: mushtaq.talib@uobabylon.edu.iq

https://doi.org/10.17977/um065.v7.i4.2027.12

Article history
Submitted: 22 August 2026
Revised: 24 September 2026
Accepted: 12 October 2026
Published: 14 October 2026

Keywords
BERT
Machine learning
Phishing email classification
TF-IDF

Abstract
Phishing and spam emails have become enormous cybersecurity threats that call for accurate and efficient identification processes. A comparative study of classical machine-learning and deep-learning approaches is developed in this paper, where we also present a BiLSTM-based model. The textual data is preprocessed (lowercased, punctuation, non-alphanumeric characters and stop words removed) followed by TF-IDF feature extraction. The obtained features are passed to MNB and SVM classifiers, whose probabilistic predictions are combined using soft voting. Along these lines, fine-tuned BERT and BiLSTM models are used to exploit the sequential and contextual dependencies in the text data. Performance of the model is evaluated using Accuracy, Precision, Recall and F1 score. All models are trained and evaluated on a single, publicly available spam-classification dataset (spam.csv, 5,573 labeled messages categorized as spam or ham) using a held-out test split. The proposed BiLSTM-based model proved its competitive classification performance with moderate computational complexity: the accuracy of 98.92% and F1-score of 96.03%. These results suggest that BiLSTM can accommodate contextual dependencies in spam and phishing-related text, while providing a trade-off to handle predictive performance and computational resources effectively. Hence, the proposed method provide a good foundation for text-wise spam and phishing detection that is both effective and scalable.

1. Introduction

The impressive growth of information technology (IT) has highly accelerated the expansion of online activities, and numerous digital platforms have emerged as a crucial part of modern life in cyberspace (Jang-Jaccard & Nepal, 2014). These activities cover a wide range of important sectors, including education and finance, as well as popular applications such as e-health systems and social networks (Adewole et al., 2019). The universal availability of social media, online gaming, and digital financial services has become humanity's recent reality (Elijah et al., 2019). Although this digital transformation has greatly increased accessibility and convenience, the diverse and increasingly interconnected nature of cyberspace has also increased the attack surface and exposure to cyber threats (Alsariera et al., 2020). Among these cybercrime threats, phishing is one of the most widespread and can affect anyone, from ordinary citizens to organizations and public institutions, regardless of their geographical location. While phishing remains one of the biggest sources of financial loss, identity theft, credential compromise, and reputational harm, email phishing remains high on criminals' list. The evolution and increasing sophistication of such attacks will continue to challenge organizations around the world as they attempt to protect users and ensure that digital communication remains safe and secure. Modern email phishing attacks continue to evolve as attackers advance their tactics to avoid detection and other security measures (Vrbančič et al., 2018). Phishing involves attempts to obtain sensitive information, such as usernames, passwords, and credit card details, by impersonating a trustworthy entity in electronic communication (Zamir et al., 2020). These impersonated links make it difficult for users to distinguish legitimate sources from fraudulent ones, increasing risks for individuals and organizations.

Recent data from the Anti-Phishing Working Group (APWG, n.d.) indicated that 1,097,811 phishing attacks were recorded in Q2 2022 alone, followed by a record-breaking high of 1,270,883 attacks during Q3 2022. This represents three times the level recorded in early 2020, when monthly attacks ranged between 68,000 and 94,000 (Figure 1).

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

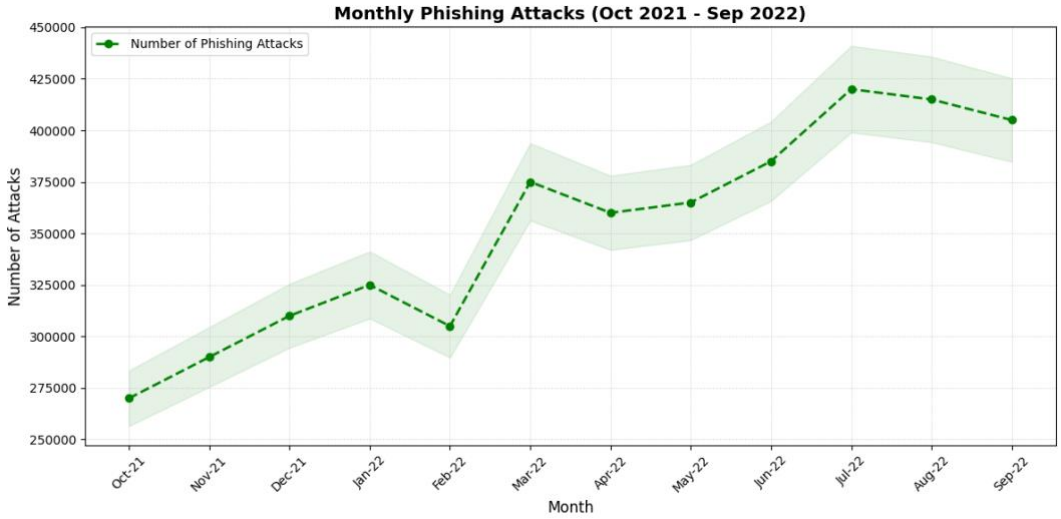


Figure 1. Monthly Distribution of Phishing Attacks from October 2021 to September 2022

Deep learning, as a subtype of machine learning based on artificial neural networks (ANNs), serves as a powerful tool for analyzing and classifying high-dimensional data, making it well suited to text classification tasks (LeCun et al., 2015). While prior work has generally evaluated classical machine-learning and deep-learning spam/phishing detectors in isolation from one another, few studies directly benchmark a soft-voting ensemble of classical classifiers against transformer-based (BERT) and recurrent (BiLSTM) models on the same corpus under an identical preprocessing pipeline and evaluation protocol. This study addresses that gap by comparing an MNB+SVM soft-voting ensemble with fine-tuned BERT and a BiLSTM-based classifier for text-based spam and phishing email detection on a single public benchmark dataset.

The contributions of this research are threefold:

1. Directly comparing classical machine-learning classifiers (Multinomial Naïve Bayes and linear-kernel SVM, combined via soft voting) against deep-learning models (fine-tuned BERT and a BiLSTM-based architecture) for text-based spam and phishing email classification, under an identical preprocessing pipeline.

2. Proposing a BiLSTM-based classifier that achieves accuracy and F1-score competitive with fine-tuned BERT while requiring substantially lower computational complexity, offering a favorable performance-efficiency trade-off.

3. Providing a standardized evaluation protocol (TF-IDF feature extraction; Accuracy, Precision, Recall and F1-score) that enables a like-for-like comparison of classical and deep-learning approaches on the same dataset.

The remainder of this paper is organized as follows: Section 2 provides background on deep learning applications in cybersecurity; Section 3 reviews existing deep learning approaches for phishing detection; Section 4 presents the proposed model; Section 5 evaluates its performance; and Section 6 concludes the study.

1.1. Related Works

The increase in awareness about combating false emails and fraudulent messages has led many researchers to use machine-learning and deep-learning technologies to classify different types of emails, such as spam and ham. For example, AbdulNabi and Yaseen (2021) used a combination of word embedding techniques and fine-tuning of BERT models to classify emails as spam or ham, achieving an accuracy of 98.67% and an F1-score of 98.66% compared with models such as deep neural networks based on bidirectional LSTM, k-nearest neighbors, and Naïve Bayes. In addition, studies evaluating spam and phishing detection using machine-learning algorithms, including logistic regression, Naïve Bayes, decision trees, boosted trees, and Random Forest, yielded lower overall performance. Random Forest achieved a maximum email classification accuracy of 96.8%; however, deep neural networks, such as Bidirectional Long Short-Term Memory (BiLSTM), outperformed traditional machine-learning algorithms by learning contextual relationships and achieved more than 99% accuracy on training data (Malhotra & Malik, 2022).

In addition to email datasets, several studies have used transformer-based architectures to improve spam detection accuracy by applying attention mechanisms to extract more complete textual representations (AbdulNabi & Yaseen, 2021). These methods have extended their application to social media by incorporating information about the users who generated the content, including spam detection on Twitter (Madisetty &

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

Desarkar, 2018). An attention-based BiLSTM approach has also been used to extract significant semantic features from text and improve deceptive opinion spam classification (Salunkhe, 2021). Hybrid approaches have combined classifiers with optimization techniques, such as Horse Herd Optimization, to improve model accuracy (Hosseinalipour & Ghanbarzadeh, 2022). Another example is the combination of a Naïve Bayes classifier and Particle Swarm Optimization to improve classification accuracy through model optimization (Agarwal & Kumar, 2018). Feature reduction techniques using WordNet semantic analysis, Principal Component Analysis, and correlation-based feature selection have been applied to reduce redundant textual information while maintaining semantic relevance (Bahgat et al., 2018). Recent work has employed Graph Neural Networks and graph-based representations to structurally analyze email content and improve spam detection (Pan et al., 2022).

For phishing detection, deep-learning models have achieved promising results using features from URLs and HTML. BERT and LSTM models have achieved more than 99% accuracy, with BERT reaching 99.61%, when compared with traditional models such as Random Forest and SVM on a limited dataset. Other studies have demonstrated the importance of persuasion cues, attention-based RCNN models for detecting email headers and bodies, the combination of graph convolutional networks and natural language processing, and transformer models such as DistilBERT. A dual-layer deep-learning architecture combining ANN, RNN, and CNN achieved an accuracy of 99.51%. Specialized datasets for deep-learning-based phishing detection remain limited, although reported validation accuracy has exceeded 98%. CNN approaches to URL classification have also achieved substantial results, including an accuracy of 98.77%. Overall, these studies suggest an evolution from traditional machine learning to deep learning and transformer-based architectures, hybrid optimization techniques, and graph-based models. This development highlights the need for reliable, explainable, and adaptive systems to address increasingly sophisticated spam and phishing attacks (Alhogail & Alsabih, 2021; Atawneh & Aljehani, 2023; Benavides-Astudillo et al., 2023; Doshi et al., 2023; Fang et al., 2019; Gogoi & Ahmed, 2022; Pan et al., 2022; Valecha et al., 2021; Yao et al., 2019; Zamir et al., 2020).

Deep learning is a subtype of machine learning built on ANNs to analyze and classify high-dimensional data, making it suitable for text and image classification tasks (LeCun et al., 2015). This study proposes a framework for spam and phishing email detection by comparing natural language processing, classical machine learning, and deep learning. The workflow begins with dataset collection and text preprocessing, including lowercasing, punctuation removal, noise reduction, and stop-word filtering. TF-IDF transforms the preprocessed text into numerical representations. Multinomial Naïve Bayes (MNB) and Support Vector Machine (SVM) classifiers are trained on the extracted features, and their probabilistic outputs are combined through soft voting. BERT and the proposed BiLSTM-based model with dense layers are also used to capture contextual and sequential dependencies in textual data. The resulting models are compared using accuracy, precision, recall, and F1-score to determine which approach offers optimal spam and phishing detection performance while maintaining low computational cost.

1.2. Dataset

This study used a Kaggle data file called spam.csv to study spam detection; spam.csv contains 5573 individual e-mails that fall into one of two categories: spam (bulk/promo e-mails) or ham (non-spam). The two primary fields in the dataset are message (containing the text of the e-mail - subject and body) and category (indicating whether the e-mail is spam or ham). The dataset was obtained from Kaggle and was split into training and test subsets. With the varying types of e-mails that people send and receive daily and the presence of both types of messages in the same inbox, this data set offers an excellent benchmark for training and testing any models developed to detect spam. All experiments reported in this study, including those in the Results and Discussion section, were conducted on this single dataset; this work did not evaluate any independent, held-out external dataset, a limitation acknowledged in the Conclusion.

2. Method

Figure 2 illustrates the proposed spam email detection methodology, which consists of natural language processing, classical machine-learning algorithms, and a deep-learning architecture for robust and accurate classification. The workflow begins with labeled email data consisting of spam and ham messages. The text is then preprocessed by converting words to lowercase and removing noise symbols, such as non-alphabetic characters, excessive numbers or spaces, and stop words. These preprocessing steps eliminate textual irregularities and extraneous noise, enhance data quality, and prepare the text for feature extraction and model training.

After preprocessing, textual features are converted into numerical representations using Term Frequency–Inverse Document Frequency (TF-IDF), which down-weights overly common words while emphasizing less frequent but more discriminative ones (AlShammari, 2023; Mohammed & Rashid, 2023; Nafis & Awang, 2021). Multinomial Naïve Bayes and SVM with a linear kernel are then trained on the extracted features. These classical supervised learning algorithms are efficient for high-dimensional sparse data. Their outputs are combined in a soft-voting ensemble to generate probabilistic predictions for final classification.

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

As an extension to the classical machine-learning framework, a deep-learning classification model uses Bidirectional Encoder Representations from Transformers (BERT) to classify spam through contextual relationships between words learned using an attention mechanism (Shreyashree et al., 2022). To provide a basis for additional benchmarking, a BiLSTM-based neural architecture followed by dense layers is also designed to model sequential dependencies in textual data (Amadi et al., 2023; Ghasemlounia et al., 2021; Jamshidzadeh et al., 2024; Michael et al., 2024).

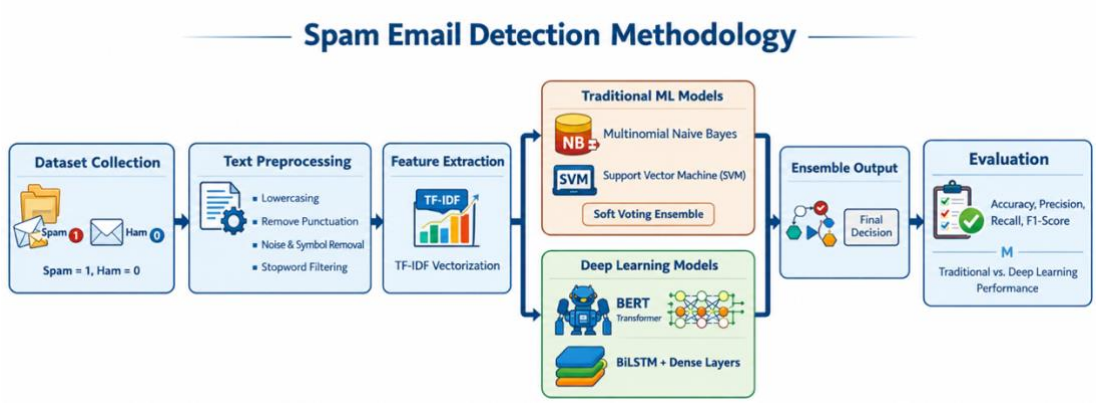


Figure 2. Block Diagram for the Proposed System

Finally, all models are evaluated using accuracy (Yacouby & Axman, 2020), precision (Sujon et al., 2025), recall (Ravi Kumar et al., 2024), and F1-score (Diallo et al., 2024), which are commonly used performance measures in the literature. This methodology provides a reproducible approach for comparing classical machine learning and deep-learning algorithms for spam email detection.

3. Results and Discussion

The proposed hybrid framework is empirically validated using standard spam detection datasets for traditional machine learning as well as deep learning paradigms. The class distribution of the dataset is shown in Figure 3, and we can see that it suffers from class imbalance with a dominant majority of legitimate emails (ham). All models achieved good classification accuracy despite this imbalance, indicating they are robust to skewed data.

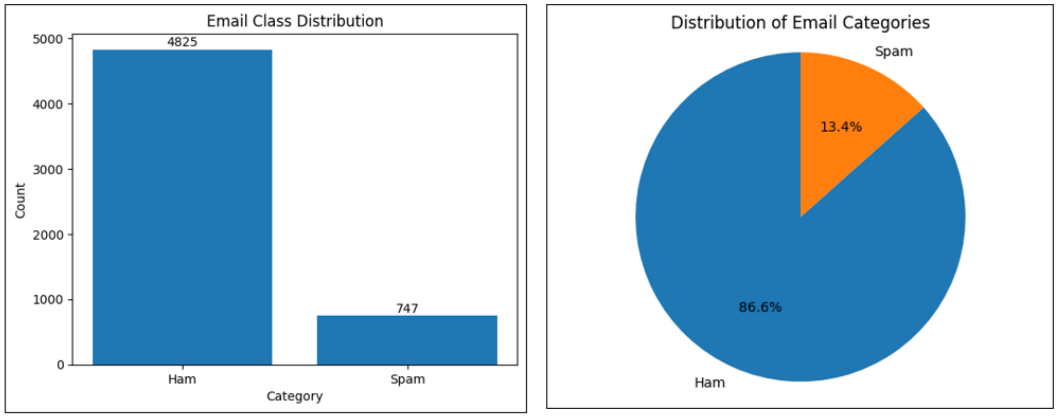


Figure3. Dataset Description

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

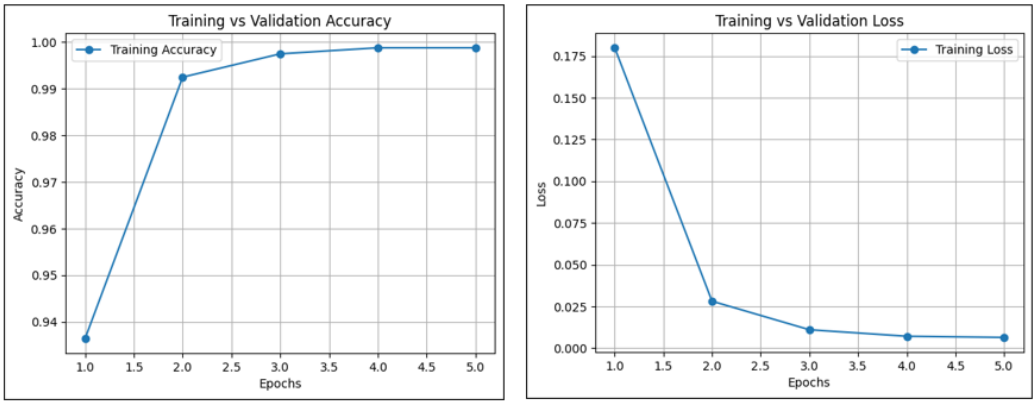


Figure4. Training Curve for Spam Dataset

The results for traditional machinelearning models is shown in Figure 4. Figure 5 shows the Multinomial Naïve Bayes classifier which has an accuracy of 97.48% and a precision that is way better than its recall as we can see, the precision is very high (99.23%) but it misses some spam with a score of F1 equal to 88.67%. On the other hand, Support Vector Machine (SVM) with linear kernel showed a better balance of precision(96.62%) and recall(91.08%) that performed much higher in the terms of F1-score(93.77%) and overall accuracy 98.29%. Data of the Support Vector Machine (SVM) model performance is shown in Figure 6.

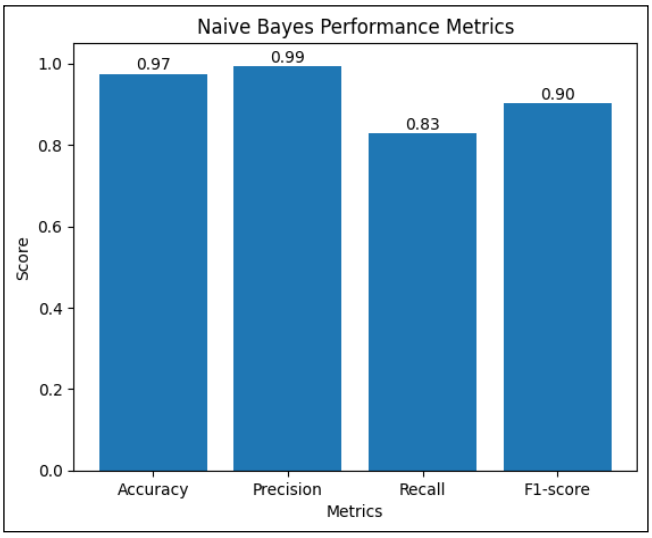


Figure 5. Multinomial Naïve Bayes Classifier Performance Metrics

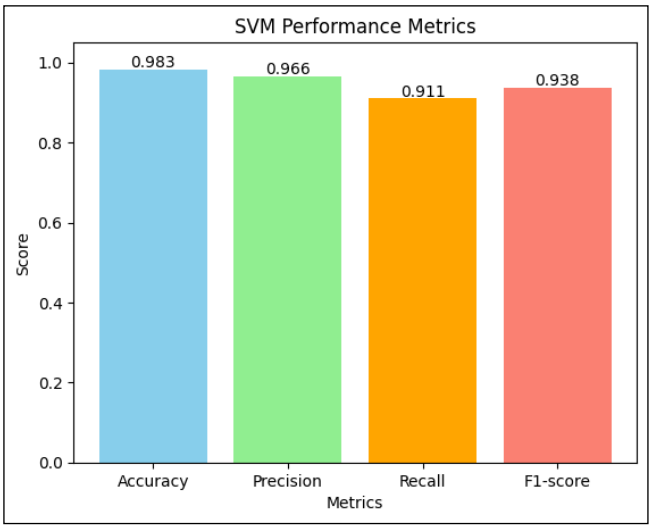


Figure 6. SVM Performance Metrics

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

To achieve even better classification performance, we propose a soft voting ensemble (Jabbar, 2024) that combines the probabilistic output from both classifiers, as illustrated in Figure 7. The ensemble outperformed the Naïve Bayes with an accuracy of 98.29% and F1-score of 93.81% confirming that increasing recall (91.72%) gives advantage in grouping complementary strengths of models.

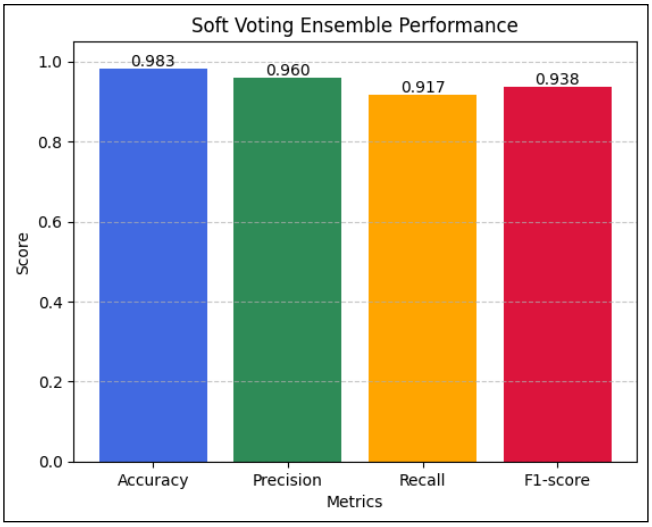


Figure 7. Soft voting ensemble performance.

As shown in Figure 8, deep learning models performed better. BiLSTM model resulted in an accuracy of 98.92% with a precision of (100%) Recall: (92.36%) and F1-score respectively as 96.03%. Referring to the train and val curves as seen in Figure 9, training loss shows rapid decline while for accuracy it stays at a higher zone thus capturing proof of strong generalization with next-to-no overfitting issues.

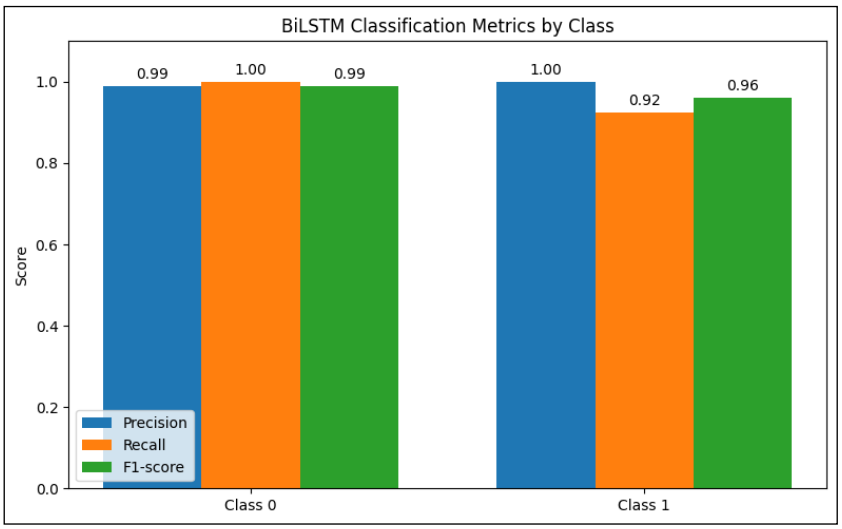


Figure 8. BiLSTM classification metrics

In the same context, our fine-tuned BERT also performed excellently with over 98% validation accuracy within only two epochs. The continuous decrease in loss and high accuracy in validation show that, indeed, representation learning at the context level is properly done with transformers. The oscillation of validation loss through epochs indicates a small sensitivity to training iterations but not a serious loss in performance. In summary, the results validate that deep learning approaches, in particular BiLSTM and BERT, outperform traditional ML models due to their capability of handling contextual and sequential dependencies inherent in the textual data. The ensemble method gives a competitive and computationally optimal alternative however. These findings emphasize the need for a synergy of classical and modern techniques to build efficient and scalable spam detection systems that can be deployed in practice or serve as high-quality baseline code for scientific research.

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

3.1. Comparison with Related Works

able 1 compares the proposed BiLSTM model with selected machine-learning and deep-learning methods reported in the literature. The model achieves an accuracy of 98.92% and an F1-score of 96.03%. While the F1-score of fine-tuned BERT (98.66%) exceeds that of the proposed BiLSTM, its accuracy (98.67%) is slightly lower. More sophisticated hybrid architectures, such as BERT–LSTM and ANN–CNN–RNN, achieved higher accuracy than single classifiers, although their multi-component frameworks often increased architectural complexity.

The proposed BiLSTM demonstrated improved performance over Random Forest, which achieved an accuracy of 96.80%, highlighting the potential benefits of bidirectional sequence modeling in capturing preceding and subsequent context in email text. The proposed approach offers a favorable balance between predictive performance and architectural complexity. Its capacity to process bidirectional textual dependencies supports its application to spam and phishing detection.

However, numerical results reported across studies may differ because of variations in datasets, preprocessing procedures, experimental settings, and evaluation protocols; therefore, cross-study comparisons require caution (AbdulNabi & Yaseen, 2021). The results indicate that the proposed BiLSTM approach offers competitive performance and practical potential for spam and phishing email detection with moderate architectural complexity.

Table 1. Comparative Analysis of the Proposed BiLSTM Model with related works

Study	Model Used	Accuracy (%)	F1-score (%)	Complexity	Key Characteristics
Proposed	Hybrid BERT–BiLSTM Soft-Voting Ensemble	98.92	96.03	Medium	Efficient contextual representation with balanced performance
AbdulNabi and Yaseen (2021)	BERT (Fine-tuned)	98.67	98.66	High	Transformer-based deep contextual modeling
Malhotra and Malik (2022)	Random Forest	96.80	—	Low	Traditional machine learning approach
Malhotra and Malik (2022); Salunkhe (2021)	BiLSTM	~99 (train)	—	Medium	Context-aware sequence learning
Atawneh and Aljehani (2023)	BERT + LSTM	99.61	—	Very High	Hybrid deep learning architecture
Doshi et al. (2023)	ANN + CNN + RNN	99.51	—	Very High	Multi-layer hybrid framework

4. Conclusion

This approach is a BiLSTM-based model for spam and phishing email detection, with an accuracy of 98.92% and F1 (96.03%). The experimental results show that our model effectively captures the long-distance contextual and sequential dependencies inherent in text data, producing a favorable trade-off between predictive performance and architectural complexity. All experiments in this study were conducted on a single, publicly available benchmark dataset using only textual message content; the model's generalization to independent, out-of-distribution datasets and to complementary modalities (e.g., URLs, headers, metadata) was not evaluated here and remains an open question for future work. This approach fills an important gap in the existing literature, as it provides a simpler yet still efficient alternative to increasingly complicated transformer-based and transformer-hybrid architectures that may incur heavier computational burdens. However, the model is mostly text-based and lacks complementary features like URLs, email metadata or behavioral attributes. In the future, we will explore BiLSTM capable of learning highlighted attention for multimodal features and lightweight transformer architectures that are able to learn multimodal information fused with explainable artificial intelligence (XAI) techniques. Additional analysis on larger and more diverse real-world datasets, along with deployment-focused experimentation and evaluation on independent datasets, will be performed to evaluate and improve the robustness, interpretability, generalizability, and practical utility of the model.

Author Contributions

The sole author contributed to all aspects of this study and has read and approved the final manuscript.

Funding

No funding support was received.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

Data Availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declaration on AI Use

The author declare that no artificial intelligence (AI) or AI-assisted tools were used in the preparation of this manuscript.

References

Abdel-Jaber, H. (2025). Detecting spam and ham SMS messages using natural language processing and machine learning algorithms. *PeerJ Computer Science, 11*, Article e3232. https://doi.org/10.7717/peerj-cs.3232

AbdulNabi, I., & Yaseen, Q. (2021). Spam email detection using deep learning techniques. *Procedia Computer Science, 184*, 853–858. https://doi.org/10.1016/j.procs.2021.03.107

Adewole, K. S., Akintola, A. G., Salihu, S. A., Faruk, N., & Jimoh, R. G. (2019). Hybrid rule-based model for phishing URLs detection. In M. H. Miraz, P. S. Excell, A. Ware, S. Soomro, & M. Ali (Eds.), *Emerging technologies in computing: Second International Conference, iCETiC 2019, London, UK, August 19–20, 2019, proceedings* (pp. 119–135). Springer. https://doi.org/10.1007/978-3-030-23943-5_9 EUDL

Agarwal, K., & Kumar, T. (2018). Email spam detection using integrated approach of Naïve Bayes and Particle Swarm Optimization. In *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 685–690). IEEE. https://doi.org/10.1109/ICCONS.2018.8662957

Alhogail, A., & Alsabih, A. (2021). Applying machine learning and natural language processing to detect phishing email. *Computers & Security, 110*, Article 102414. https://doi.org/10.1016/j.cose.2021.102414

Alsariera, Y. A., Elijah, A. V., & Balogun, A. O. (2020). Phishing website detection: Forest by penalizing attributes algorithm and its enhanced variations. *Arabian Journal for Science and Engineering, 45*(12), 10459–10470. https://doi.org/10.1007/s13369-020-04802-1

AlShammari, A. F. (2023). Implementation of keyword extraction using term frequency-inverse document frequency (TF-IDF) in Python. *International Journal of Computer Applications, 185*(35), 9–14. https://doi.org/10.5120/ijca2023923137 ijcaonline.org

Amadi, C. O., Odii, J. N., Okpalla, C. L., & Ofoegbu, C. I. (2023). Emotion detection using a bidirectional long-short term memory (BiLSTM) neural network. *International Journal of Research Publication and Reviews, 4*(11), 1718–1732. ijrpr.com

Anti-Phishing Working Group. (2022). *Phishing activity trends report: 3rd quarter 2022*. https://docs.apwg.org/reports/apwg_trends_report_q3_2022.pdf

Atawneh, S., & Aljehani, H. (2023). Phishing email detection model using deep learning. *Electronics, 12*(20), Article 4261. https://doi.org/10.3390/electronics12204261

Bahgat, E. M., Rady, S., Gad, W., & Moawad, I. F. (2018). Efficient email classification approach based on semantic methods. *Ain Shams Engineering Journal, 9*(4), 3259–3269. https://doi.org/10.1016/j.asej.2018.06.001

Benavides-Astudillo, E., Fuertes, W., Sanchez-Gordon, S., Nuñez-Agurto, D., & Rodríguez-Galán, G. (2023). A phishing-attack-detection model using natural language processing and deep learning. *Applied Sciences, 13*(9), Article 5275.

Castillo, E., Dhaduvai, S., Liu, P., Thakur, K.-S., Dalton, A., & Strzalkowski, T. (2020). Email threat detection using distinct neural network approaches. In A. Bhatia & S. Shaikh (Eds.), *Proceedings for the First International Workshop on Social Threats in Online Conversations: Understanding and Management* (pp. 48–55). European Language Resources Association. https://aclanthology.org/2020.stoc-1.8/ ACL Anthology

Chebil, W., Wedyan, M., Alazab, M., Alturki, R., & Elshaweesh, O. (2023). Improving semantic information retrieval using multinomial naive Bayes classifier and Bayesian networks. *Information, 14*(5), Article 272. https://doi.org/10.3390/info14050272

Danyal, M. M., Khan, S. S., Khan, M., Ghaffar, M. B., Khan, B., & Arshad, M. (2023). Sentiment analysis based on performance of linear support vector machine and multinomial naïve Bayes using movie reviews with baseline techniques. *Journal on Big Data, 5*(1), 1–18. https://doi.org/10.32604/jbd.2023.041319 techscience.com

Diallo, R., Edalo, C., & Awe, O. O. (2025). Machine learning evaluation of imbalanced health data: A comparative analysis of balanced accuracy, MCC, and F1 score. In O. O. Awe & E. A. Vance (Eds.), *Practical statistical learning and data science methods: Case studies from LISA 2020 Global Network, USA* (pp. 283–312). Springer. https://doi.org/10.1007/978-3-031-72215-8_12

Do, N. Q., Selamat, A., Krejcar, O., Herrera-Viedma, E., & Fujita, H. (2022). Deep learning for phishing detection: Taxonomy, current challenges and future directions. *IEEE Access, 10*, 36429–36463. https://doi.org/10.1109/ACCESS.2022.3151903

Doshi, J., Parmar, K., Sanghavi, R., & Shekokar, N. (2023). A comprehensive dual-layer architecture for phishing and spam email detection. *Computers & Security, 133*, Article 103378. https://doi.org/10.1016/j.cose.2023.103378

Elijah, A. V., Abdullah, A., JhanJhi, N. Z., Supramaniam, M., & Balogun, A. O. (2019). Ensemble and deep-learning methods for two-class and multi-attack anomaly intrusion detection: An empirical study. *International Journal of Advanced Computer Science and Applications, 10*(9), 520–528. https://doi.org/10.14569/IJACSA.2019.0100969thesai.org

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

Fang, Y., Zhang, C., Huang, C., Liu, L., & Yang, Y. (2019). Phishing email detection using improved RCNN model with multilevel vectors and attention mechanism. *IEEE Access, 7*, 56329–56340.

Ghasemlounia, R., Gharehbaghi, A., Ahmadi, F., & Saadatnejadgharahassanlou, H. (2021). Developing a novel framework for forecasting groundwater level fluctuations using bi-directional long short-term memory (BiLSTM) deep neural network. *Computers and Electronics in Agriculture, 191*, Article 106568.

Gogoi, B., & Ahmed, T. (2022). Phishing and fraudulent email detection through transfer learning using pretrained transformer models. In *2022 IEEE 19th India Council International Conference (INDICON)* (pp. 1–6). IEEE. https://doi.org/10.1109/INDICON56171.2022.10040097

Hosseinalipour, A., & Ghanbarzadeh, R. (2022). A novel approach for spam detection using horse herd optimization algorithm. *Neural Computing and Applications, 34*, 13091–13105. https://doi.org/10.1007/s00521-022-07148-xlink.springer.com

Jabbar, H. G. (2024). Advanced threat detection using soft and hard voting techniques in ensemble learning. *Journal of Robotics and Control, 5*(4), 1104–1116. https://doi.org/10.18196/jrc.v5i4.22005

Jamshidzadeh, Z., Ehteram, M., & Shabanian, H. (2024). Bidirectional long short-term memory (BiLSTM)-support vector machine: A new machine learning model for predicting water quality parameters. *Ain Shams Engineering Journal, 15*(3), Article 102510. https://doi.org/10.1016/j.asej.2023.102510

Jang-Jaccard, J., & Nepal, S. (2014). A survey of emerging threats in cybersecurity. *Journal of Computer and System Sciences, 80*(5), 973–993.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Madisetty, S., & Desarkar, M. S. (2018). A neural network-based ensemble approach for spam detection in Twitter. *IEEE Transactions on Computational Social Systems, 5*(4), 973–984. https://doi.org/10.1109/TCSS.2018.2878852

Malhotra, P., & Malik, S. (2022). *Spam email detection using machine learning and deep learning techniques*. SSRN. https://doi.org/10.2139/ssrn.4145123

Michael, N. E., Bansal, R. C., Ismail, A. A. A., Elnady, A., & Hasan, S. (2024). A cohesive structure of bi-directional long-short-term memory (BiLSTM)-GRU for predicting hourly solar radiation. *Renewable Energy, 222*, Article 119943. https://doi.org/10.1016/j.renene.2024.119943

Mohammed, M. T., & Rashid, O. F. (2023). Document retrieval using term frequency inverse sentence frequency weighting scheme. *Indonesian Journal of Electrical Engineering and Computer Science, 31*(3), 1478–1485. https://doi.org/10.11591/ijeecs.v31.i3.pp1478-1485

Nafis, N. S. M., & Awang, S. (2021). An enhanced hybrid feature selection technique using term frequency-inverse document frequency and support vector machine-recursive feature elimination for sentiment classification. *IEEE Access, 9*, 52177–52192.

Pan, W., Li, J., Gao, L., Yue, L., Yang, Y., Deng, L., & Deng, C. (2022). Semantic graph neural network: A conversion from spam email classification to graph classification. *Scientific Programming, 2022*, Article 6737080. https://doi.org/10.1155/2022/6737080 Wiley Online Library

Ravi Kumar, G., Venkataiah, V., Sivaiah, B., Kavitha Rani, B., Murali, K., & Swathi, G. (2026). An assessment on the performance of a composite learning system for prediction using precision, recall, accuracy, and F1-score for rainfall. In A. Kumar & S. Mozar (Eds.), *Proceedings of the 7th International Conference on Communications and Cyber Physical Engineering* (pp. 889–897). Springer. https://doi.org/10.1007/978-981-95-0269-1_101

Roy, P. K., Singh, J. P., & Banerjee, S. (2020). Deep learning to filter SMS spam. *Future Generation Computer Systems, 102*, 524–533. https://doi.org/10.1016/j.future.2019.09.001

Salunkhe, A. (2021). *Attention-based bidirectional LSTM for deceptive opinion spam classification* [Preprint]. arXiv. https://arxiv.org/abs/2112.14789

Shreyashree, S., Sunagar, P., Rajarajeswari, S., & Kanavalli, A. (2022). A literature review on bidirectional encoder representations from transformers. In S. Smys, V. E. Balas, & R. Palanisamy (Eds.), *Inventive computation and information technologies: Proceedings of ICICIT 2021* (pp. 305–320). Springer. https://doi.org/10.1007/978-981-16-6723-7_23 Springer Nature Link

Sujon, K. M., Hassan, R., Choi, K., & Samad, M. A. (2025). Accuracy, precision, recall, F1-score, or MCC? Empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models. *Journal of Big Data, 12*, Article 268. https://doi.org/10.1186/s40537-025-01313-4

Valecha, R., Mandaokar, P., & Rao, H. R. (2022). Phishing email detection using persuasion cues. *IEEE Transactions on Dependable and Secure Computing, 19*(2), 747–756. https://doi.org/10.1109/TDSC.2021.3118931

Vrbančič, G., Fister, I., Jr., & Podgorelec, V. (2018). Swarm intelligence approaches for parameter setting of deep learning neural network: Case study on phishing websites classification. In *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics* (pp. 1–8). Association for Computing Machinery.

Yacouby, R., & Axman, D. (2020). Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models. In S. Eger, Y. Gao, M. Peyrard, W. Zhao, & E. Hovy (Eds.), *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems* (pp. 79–91). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.eval4nlp-1.9 ACL Anthology

Yao, L., Mao, C., & Luo, Y. (2019). Graph convolutional networks for text classification. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(1), 7370–7377. https://doi.org/10.1609/aaai.v33i01.33017370

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174

Zamir, A., Khan, H. U., Iqbal, T., Yousaf, N., Aslam, F., Anjum, A., & Hamdani, M. (2020). Phishing web site detection using diverse machine learning algorithms. *The Electronic Library, 38*(1), 65–80. https://doi.org/10.1108/EL-05-2019-0118

© 2026 The Author(s)
This is an Open Access article under the Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 2797-3174