

Phishing website detection model based on Tabular Multi-Head Attention (Tabmha)**Mohammad A. Alsharaiah^a, Mohammed Amin^a, Amer Alqutaish^{b*} and Ghada Alradwan^b**^aKing Abdullah II School for Information Technology, The University of Jordan, Amman, Jordan^bDeanship of Development and Quality Assurance, King Faisal University, 31982, Al-Ahsa, Saudi Arabia**CHRONICLE**

Received October 14, 2025
 Received in revised format
 December 28, 2025
 Accepted February 3 2026
 Available online
 February 4 2026

Keywords:

Phishing Detection
Deep learning
Classification
Tabular Multi-Head Attention

ABSTRACT

The vast usage and development of web technology generate numerous types of web pages. Besides, not all these types are legitimate webpages. Phishing sites mislead web page users into taking harmful actions. However, there is a need for a tool to address this type of problem. Deep learning models are used in dealing with web technology to detect whether the webpage is either legitimate or phishing. Herein, a novel Tabular Multi-Head Attention (TabMHA) model is presented to perform a binary classification task. The main task is to classify whether the webpages are phishing or not. The proposed model is trained and tested on a benchmark dataset related to phishing detection. It contains 5000 legitimate web pages and 5000 phishing ones; the overall is 10,000. Also, the feature numbers in the dataset are out of 48 features. As a consequence, the proposed model achieved a powerful performance compared with other models in the literature; the model achieved an accuracy level of 99.6%. This result is considered a promising result and can be integrated into real-world detection models.

© 2026 by the authors; licensee Growing Science, Canada.

1. Introduction

The proliferation of internet usage and increasing dependency on web-based services have significantly heightened users' vulnerability to malicious cyber attacks (Aslan et al., 2023). Aslan and colleagues document how expanded digital footprints and online activities create larger attack surfaces, with particular concern for users lacking cybersecurity awareness. Their 2023 study quantifies this correlation across different demographic groups, showing that increased internet engagement directly correlates with higher phishing susceptibility rates. Among various cyber threats, phishing attacks represent a particularly insidious form of cybercrime that jeopardizes user confidentiality through sophisticated deception techniques. These attacks frequently serve as gateways for malware distribution and systematic theft of sensitive personal and financial information (Alkhalil et al., 2021; Asiri, 2023). Alkhalil's comprehensive 2021 taxonomy of phishing methodologies reveals how attackers continuously evolve their strategies, with their work establishing foundational understanding of phishing's technical and psychological dimensions. Subsequent research by Vacalares et al. (2024) documenting how fraudulent emails meticulously mimic legitimate organizational communications, particularly targeting financial institutions. Their longitudinal study of bank-themed phishing emails reveals increasing sophistication in spoofing techniques, including improved visual design, domain name manipulation, and psychological triggers. Similarly, Alzoubi and Ghazal (2022) investigate payment website phishing, demonstrating how attackers create convincing replicas of banking login portals. Their 2022 analysis of 500 fraudulent banking websites identifies common design patterns and technical implementations used to deceive users, with particular focus on mobile banking interfaces where screen size limitations may reduce users' ability to identify fraudulent elements. The fundamental mechanism of website phishing involves creating deceptive facsimiles of legitimate websites. Lim et al. (2024) conceptualize this as "digital identity spoofing," where malicious actors systematically replicate visual elements, branding, and interaction patterns of trusted websites. Their framework distinguishes between complete replication (mirror sites) and partial imitation strategies, with empirical data showing varying effectiveness rates. Once users interact with these fraudulent

* Corresponding author

E-mail address: aalqutish@kfu.edu.sa (A. Alqutaish)

platforms, attackers successfully harvest sensitive information through various means. Kalla et al. (2023) document the technical infrastructure supporting these operations, while subsequent research by Almarshood and Rahman (2025) analyzes the post-compromise phase where users are redirected to additional malicious resources.

In response to these evolving threats, researchers have developed diverse protective methodologies spanning multiple technical approaches. Heuristic-based detection systems represent one established approach, with Rao et al. (2023) proposing a comprehensive heuristic framework that evaluates website characteristics against known phishing patterns. Their work establishes 35 distinct heuristics derived from URL structure, content analysis, and behavioral indicators. Building upon this foundation, Bawany et al. (2017) introduce adaptive heuristic mechanisms that evolve based on emerging threat patterns, while Ferdous et al. (2025) extend heuristic approaches to mobile and IoT environments respectively, addressing platform-specific vulnerabilities. Rule-based methodologies provide another detection avenue, particularly in specialized domains like internet banking. Vaghela's 2023 contribution develops a feature-based rule system specifically optimized for financial services, incorporating transaction patterns, session characteristics, and behavioral biometrics into a comprehensive detection framework. Their approach demonstrates particular effectiveness against banking-specific phishing variants but highlights challenges in generalization to other domains. Machine learning approaches have significantly advanced phishing detection capabilities through automated pattern recognition. Their comparative analysis shows Random Forest and Gradient Boosting machines achieving superior performance on imbalanced datasets. Lertampaiporn et al. (2013) focus on feature engineering aspects, identifying the most discriminative characteristics for ML-based classification, while Ren et al. (2018) address the critical challenge of real-time detection, optimizing algorithms for computational efficiency without sacrificing accuracy. Deep learning represents the current frontier in phishing detection research, leveraging neural architectures for enhanced pattern recognition. Tang and Mahmoud (2021) pioneer convolutional neural network applications to visual website analysis, establishing baseline performance metrics for DL approaches. Khaleel et al. (2024) innovate further with hybrid architectures combining multiple neural approaches, demonstrating state-of-the-art performance on benchmark datasets while addressing challenges like adversarial attacks and concept drift.

Gaps and Research Opportunities Identified:

Despite these advancements, several research gaps remain evident. First, most existing studies focus on specific phishing vectors in isolation, with limited investigation of integrated multi-vector attacks that increasingly characterize sophisticated campaigns. Second, while detection accuracy has improved significantly, real-time deployment challenges—particularly concerning computational efficiency and resource constraints—require further attention. Third, the psychological dimensions of phishing susceptibility remain underexplored in technical detection literature, suggesting opportunities for interdisciplinary approaches combining cybersecurity with behavioral science. Finally, emerging technologies like decentralized web architectures and progressive web applications introduce new attack surfaces that existing detection systems may not adequately address, indicating fertile ground for future research innovation. This comprehensive review establishes the evolutionary trajectory of phishing detection research while identifying critical areas requiring further investigation. The progression from heuristic and rule-based approaches through machine learning to contemporary deep learning solutions reflects both technological advancement and the escalating sophistication of phishing attacks, creating an ongoing arms race between detection systems and adversarial techniques.

This study introduces a novel work for classifying websites based on an influential deep learning method, such as multihued attention. The contribution to this research study is as follows: Feature embedded and positional encoding. Assets the performance for the proposed model based on Tabular Multi-Head Attention (TabMHA) network, which is structurally based on the Transformer Encoder block. Comparison of the proposed model performance via other literature models such as KNN, Neural Network, SVM, and Attention model. Performance evaluation metrics, like accuracy, f-measure, recall, and precision, are involved to evaluate the performance of all classifiers.

The other part of this research paper is organized in the following structure: the available methods and techniques that would be available in the literature review to detect phishing websites. Detailed information on the methodology of this research is presented in Section 3. Section 4 investigates the results achieved. Section 5 shows the conclusions for this novel work and future works.

2. Related Works

Recent years have witnessed a proliferation of methodological advancements in phishing detection research, with numerous novel approaches documented across the literature. Addula et al. (2025) provide a systematic meta-analysis of detection methodologies developed between 2020-2024, identifying key trends in algorithm evolution and performance benchmarks across different application domains. Their comprehensive review establishes methodological taxonomies and identifies recurring limitations in current approaches.

Foundational methodological frameworks have established core detection paradigms that continue to influence contemporary research. Zuraiq and Alkasassbeh (2019) provide a seminal classification of phishing detection approaches into three primary categories: heuristic-based methods, content-based techniques, and fuzzy rule-based systems. Their heuristic approach establishes rule-based systems that evaluate websites against known phishing indicators, while their content-based methodology analyzes webpage structure, text, and visual elements. Most innovatively, their fuzzy rule-based system introduces uncertainty handling mechanisms that address the ambiguous nature of phishing indicators, creating more robust detection frameworks that better reflect real-world complexity. This tripartite classification has served as a foundational reference for subsequent methodological development.

Deep learning methodologies have significantly advanced detection capabilities, with Mohammad et al. (2020) establishing important benchmarks through their implementation of Deep Belief Networks (DBN). Their work, further elaborated by Alamgir (2020), represents one of the earliest comprehensive applications of deep architectures to phishing detection, achieving a noteworthy 94.43% accuracy on standardized datasets. This performance established a baseline for deep learning approaches while revealing challenges related to computational requirements and training data demands. The research demonstrated DBN's particular strength in identifying complex, non-linear patterns in website characteristics but also highlighted limitations in real-time deployment due to inference speed constraints. The pursuit of improved accuracy has driven methodological refinement, with Ubung et al. (2021) explicitly targeting performance enhancement beyond existing technological capabilities. Their systematic approach achieved 95% accuracy, representing a meaningful improvement from the 92.52% range reported in prior works. As noted by Yusuf et al. (2024), this 2.48% improvement, while statistically significant, remained below the 99% threshold that represents current state-of-the-art aspirations. This performance plateau at the mid-90% range has established what Yusuf et al. term the "baseline performance ceiling" for generalized or less optimized detection approaches, highlighting the increasing difficulty of marginal accuracy gains as performance approaches theoretical maximums.

Feature engineering and selection have emerged as critical determinants of detection system performance. Rashid et al. (2021) provide comprehensive methodologies for feature extraction from website structure, content, and behavioral patterns, while Nath et al. (2022) develop systematic approaches to feature relevance assessment and selection. Their work demonstrates how judicious feature selection can dramatically impact model performance, training efficiency, and interpretability. Kashkool et al. (2024) exemplify this approach by achieving 95.66% accuracy using Support Vector Machines (SVM) combined with Principal Component Analysis (PCA) for dimensionality reduction. Their work illustrates the dual benefit of feature space reduction: improved computational efficiency and enhanced model generalization through elimination of redundant or noisy features. This research establishes that effective feature engineering can compensate for limitations in base algorithms, particularly for traditional machine learning approaches. Random Forest methodologies have demonstrated consistent effectiveness in phishing detection, with Owa and Adewole (2025) dedicated to optimizing this ensemble approach. Their systematic hyperparameter tuning and feature selection process achieved 97.069% accuracy, as corroborated by Pathak and Shrivastava (2024). This research reveals the trade-off between comprehensive feature sets and optimized feature subsets, demonstrating that performance optimization often involves balancing feature richness against model complexity and interpretability. Gupta et al. (2023) extend this work through Adaptive Random Forest (ARF) algorithms that achieve 97.1% accuracy while introducing online learning capabilities that address concept drift in evolving phishing tactics.

Ensemble approaches represent a significant advancement in detection methodology, with Alsharaiah et al. (2023) proposing a sophisticated ensemble model that integrates multiple algorithms to achieve 98.64% accuracy. Their approach combines complementary strengths of different classifiers while mitigating individual weaknesses, representing a methodological shift toward hybrid systems that leverage algorithmic diversity. This ensemble paradigm addresses the "no free lunch" theorem in machine learning by acknowledging that no single algorithm performs optimally across all phishing variants and attack scenarios.

Deep learning architectures have pushed performance boundaries further, with Convolutional Neural Networks (CNN) demonstrating particular effectiveness in visual and structural analysis of websites. Li et al. (2019) pioneered CNN applications to phishing detection, achieving 98.60% accuracy by treating website screenshots and structural representations as image-like data. Their approach leverages CNN's spatial pattern recognition capabilities to identify visual similarities between legitimate and phishing sites that may elude feature-based approaches. This represents a paradigm shift toward end-to-end learning systems that automatically derive relevant features from raw data rather than relying on handcrafted feature engineering.

Table 1

Phishing website classification models.

Reference (Year)	Core Methodology	Key Features Used	Classification Algorithms	Highlight/Best Performance (<99%)
------------------	------------------	-------------------	---------------------------	-----------------------------------

Mohammad et al. (2020)	Functional Tree (FT) based Meta-learners and Deep Belief Network (DBN)	Not explicitly detailed, assumed standard content/URL features	Functional Tree (FT) variants, DBN, Decision Tree, RF	DBN model achieved 94.43% accuracy; FT-based models reached up to 98.51% accuracy.
Ubing et al. (2021)	Feature Selection combined with Ensemble Learning	Standard URL and Content features	Random Forest, Logistic Regression, etc.	Current technologies discussed were between 70% and 92.52. The proposed model achieved 95% accuracy.
Rashid et al. (2021)	Feature Selection (PCA) and Machine Learning Comparison	48 features reduced to 5 relevant features	SVM, Logistic Regression, other ML models	SVM (best-performing model) achieved 95.66% accuracy after feature selection.
Nath et al. (2022)	Improved Random Forest (RF) using feature selection and hyper parameter optimization	URL and Domain Name features	Random Forest (RF)	Improved RF model yielded 97.069 % accuracy.
Mohan et al. (2023)	Multi-modal Comparative Analysis and Incremental Execution	URL and HTML/Content Attributes	Adaptive Random Forest (ARF) , RF, SVM, etc.	Adaptive Random Forest (ARF) provided 97.1% accuracy in real-time/incremental mode.
Alsharaiah et al. (2023)	Ensemble Model	Website phishing	Ensemble Model	accuracy of 98.64%
(Y. Li, Yang, Chen, Yuan, & Liu, 2019)	Convolutional Neural Network (CNN)	Website phishing	Convolutional Neural Network (CNN)	An accuracy of 98.60%.

3. Methodology

3.1 The Phishing dataset

This study used a viable benchmark dataset explored by Chiew et al. (2019). It contains 48 features and 10,000 samples, 5,000 legitimate ones and 5,000 phishing websites. The data set is balanced as shown in Fig. 1. The samples were gathered during the period January to May 2015 and May to June 2017.

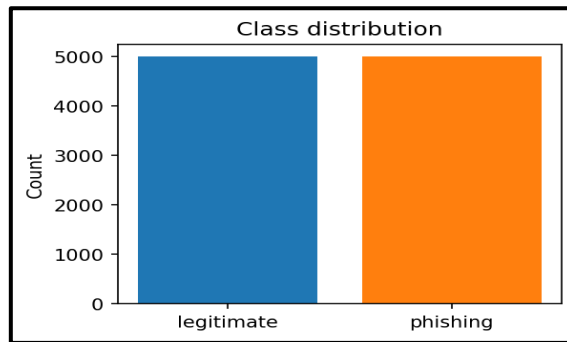


Fig. 1. Class distribution for the Phishing websites’ Dataset

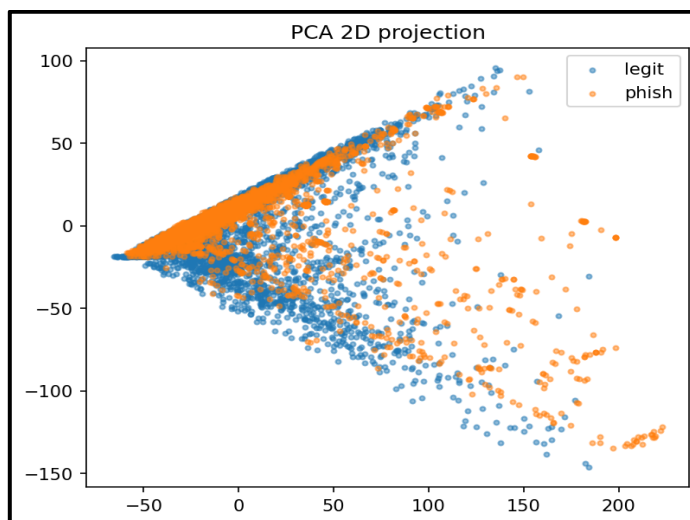


Fig. 2. The PCA 2D projection for the dataset feature space

The nature of the dataset is very complex to get an accurate binary distinction. As shown in Fig. 2, the PCA 2D projection exposes the significant overlap between the feature space of legitimate blue color and phishing websites in the orange color. In this study, we examined the discriminative power and consistency of the features. Fig. 3 provides a visual summary of the distribution and spread for numerous features used to classify websites either as legitimate or phishing. The Y-axis: Lists the features (e.g., PctExtHyperlinks, URLlength, SubDomainLen).

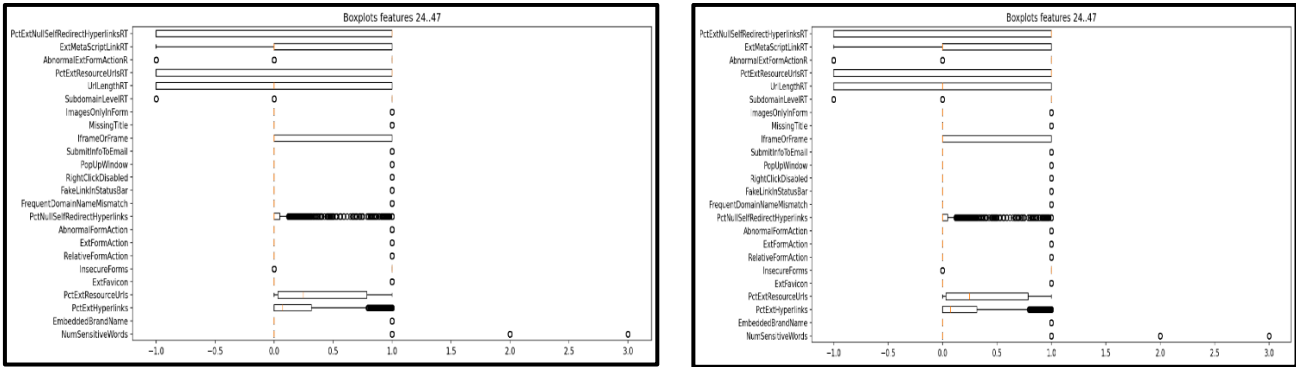


Fig. 3. box plot visualization for the discriminative power and consistency of the features.

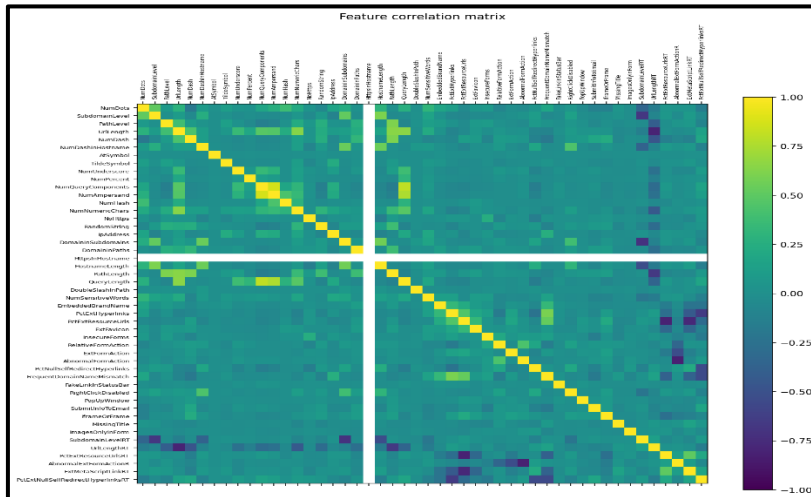


Fig. 4. Heatmap correlation for the feature sets.

These features are extracted from websites (URL, content, etc.) to train the classification models. X-axis: Represents the standardized or normalized values of the features (ranging from approximately -1.5 to 3.0). Every horizontal boxplot summarizes the data for one feature, such as the box itself, which shows the Interquartile Range (IQR), comprising the middle 50% of the data. The vertical line inside the box is the median (50th percentile). The whiskers extend to show the variability outside the IQR. The dots/circles outside the whiskers represent outliers. In fact, researchers use the heat matrix to understand the dataset structure (Owa & Adewole, 2025), then decide which redundant features to remove before training the phishing website classification model. Fig. 4 explores the pairwise correlation coefficients among all the features extracted from the websites, such as external links and domain age. These features will be used to train the classification model. Mainly, the goal is to identify the highly correlated features that are shown in (Red/Yellow) features that deliver redundant information. On the other hand, Uncorrelated features (Green/Teal/Blue) features that offer independent information. Furthermore, his study involves Data transformation and normalization. The nominal and ordinal data are converted into numerical data via the OneHot encoder technique (Majgave & Gavankar, 2024). As given in Eq. (1). Besides, the Z-score is applied as given in Eq. (2) (Majgave & Gavankar, 2024).

$$Z = \frac{x - \min}{\max - \min} \tag{1}$$

$$Z = \frac{x - m}{6} \tag{2}$$

3.2 Evaluation metrics

The proposed model is assessed by a set of metrics, such as the accuracy and the confusion matrix metrics (Ubing et al., 2019). Further, precision, recall, and F-score measures are similarly used for the assessment. The confusion matrix is also known as the error matrix. It is a statistical classification to visualize the model’s performance, as shown in Fig. 5.

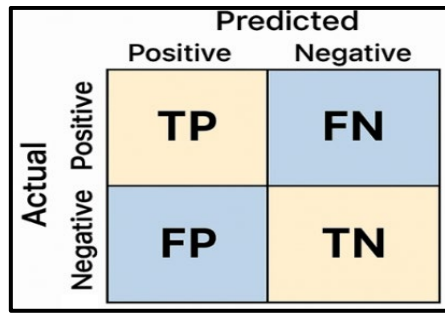


Fig. 5. The confusion matrix

The confusion matrix has several indicators. For example, the True Positive named with (TP) signifies the true prediction count of the positive data points, which specifies the predicted value as positive for samples with actual values that are also positive. The False Positive named with (FP) determines the overall number of negative values incorrectly classified as positive. Besides, True Negative named with (TN) signifies the sum of actually predicted negative data points in which the forecast value is negative, and the actual value is negative. Lastly, a False Negative has name (FN) signifies the sum of positive values that are incorrectly categorized as negative.

Further, the accuracy (Foody, 2023) ratio can be measured based on the usage of a set of indicators, such as TN and TP indicators; it is given in Eq. (3). Besides, the metric precision (St-Aubin & Agard, 2022) is calculated as the indicators of TP divided by the count of positively considered samples by the operated classifier; this is presented in Eq. (4). While the recall metric (Sykes et al., 2024) is computed as the count of TP divided by the count of positive samples in the dataset, this is specified in Eq. (5). Finally, Eq. (6) shows the F-score (Krasnodębska et al., 2025), it is computed in is utilized for averaging the precision and recall.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{F-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

3.3 Model Main methodology and architecture

The main method and the proposed model architecture are shown in Fig. 6.

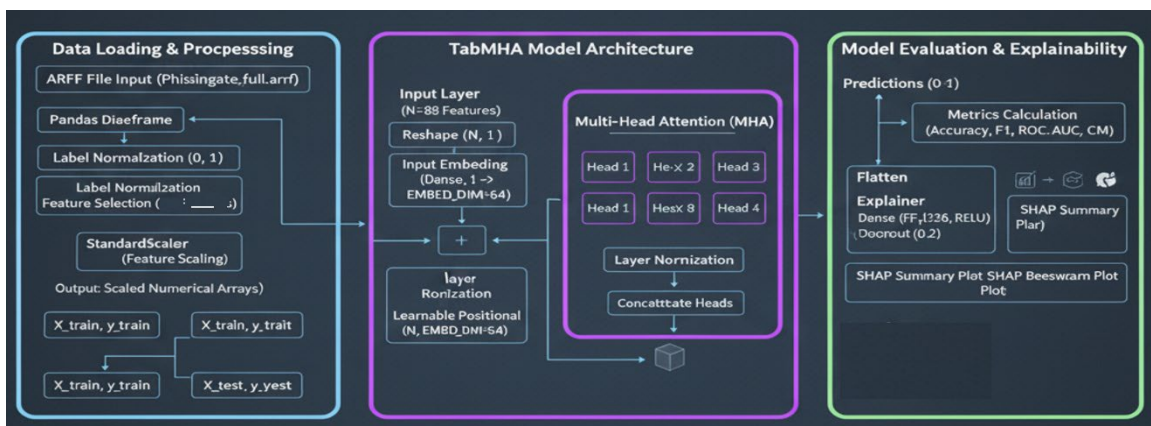


Fig. 6. The proposed model architecture.

The core of this pipeline is the Tabular Multi-Head Attention (TabMHA) Model Architecture. It is designed to handle the 48 selected and scaled features from the pre-processing phase for binary classification. The process starts in the Input Layer,

which reforms the N features into a sequence of length 1, formulating the tabular data for sequential treating typical in attention mechanisms. This is enclosed by Input Embedding, this step gives the data a new shape and projects it into a higher dimension; dense vector space (EMBED_DIM=64). Following the embedding, Layer Normalization and Learnable Positional Embedding are employed. The positional embedding, though usually used for sequences, benefits the model in distinguishing between feature locations, enhancing its ability to capture feature importance. The subsequent vector is then fed into the Multi-Head Attention (MHA) block, which comprises multiple Heads (Heads 1 through 8). Each head independently handles the input, allowing the model to pay attention to diverse subsets of features or relationships simultaneously. The outputs from these eight heads are concatenated, followed by additional Layer Normalization, which creates the final representation utilized for prediction. This architecture influences the power of Transformers to weigh the significance of diverse features dynamically, making it highly operational for complex classification tasks such as phishing detection. The fundamental component for the proposed model is the Multi-Head Attention (MHA) Block (Weng et al., 2023). Figure 7 explores the Internal Architecture of the (MHA), which takes the input X, and it could be considered as the initial input embedding or output of the previous layer. The Initial Projections (Q, K, V). The input X is first used to produce 3 key tensors: Query (Q), Key (K), and Value (V). These are all imitative of the input embeddings. Then, these key sensors passed through separate Linear Projections. (Parameterized by weight matrices W_Q , W_K , and W_V). As a consequence, it projects them into spaces proper for attention calculation. The projected matrices are split into h Heads, where $h=4$. The embedded dimension $D=64$ is split among the heads. Besides, every head i get Q_i , K_i and V_i with a smaller dimension equal to 16. Then, each head of the Scaled Dot-Product Attention function must be calculated. As represented in Eq. (7).

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^t}{\sqrt{d_k}} \right) V \tag{7}$$

The output from Eq. (7) represents the attention matrix, and the $\sqrt{d_k}$ is a scaling factor to inhibit the dot products from becoming too large and pushing the softmax function into regions with very small gradients. The outcome from all the h individual attention heads such as H1, H2, H3, and H4, is concatenated back together: $H1 \parallel H2 \parallel H3 \parallel H4$. As a result, a single tensor with the original embedding dimension, and this tensor is forwarded to the final layer W^o . This produces the final block named Z with an Embedded dimension equal to 64. This final projection lets the heads combine their collectively learned representations.

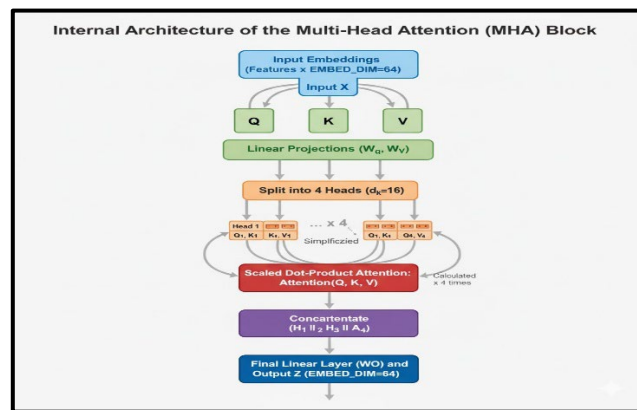


Fig. 7. The core of Multi Head Attention – MHA.

4. Analysis and Results

The model was trained and tested by the benchmark dataset aforementioned above; the computational environments are represented in Table 2.

Table 2

The infrastructure for the experimental environment

Operating System	Windows 11
Processor (CPU)	Intel Core i7-12700H (14 cores, 24MB cache)
Graphics Card (GPU)	NVIDIA RTX 3060 (6GB VRAM)
The Memory (RAM)	16GB DDR5 (4800 MHz)
The Environment	Anaconda, Python
The Frameworks	TensorFlow

Precisely, the model achieved promising results. For instance, the training and validation curves are shown in Fig. 8. The training accuracy and validation accuracy in the first 10 epochs. Also, the curves are tracked very carefully, showing the minimum gap between them. This indicates no significant overfitting.

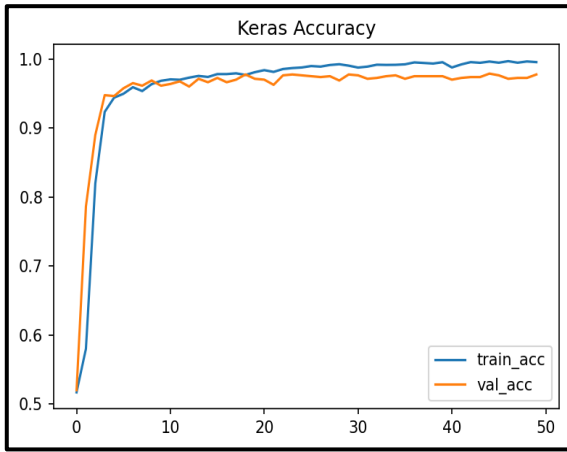


Fig. 8. Training and validation curves

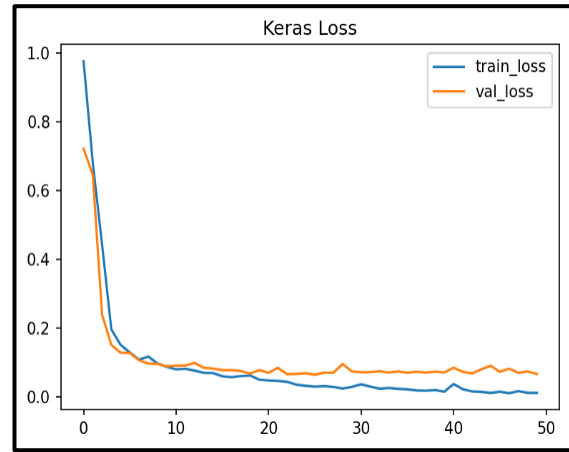


Fig. 9. Training and validation loss curves

The relation between the model and the loss curves is shown in Fig. 9. Both validation and training loss have decreased sharply, and the values are very low values near 0.0 to 0.1. Overall, the model is generalizing well and has no high variance. The confusion matrix for the model is shown in Fig. 10. It illustrates the confusion matrix that reflects a 99.6% accuracy for the 2000 samples. The False Positives (Class 0 incorrectly predicted as 1) equal 4, and the False Negatives (Class 1 incorrectly predicted as 0) equal 4.

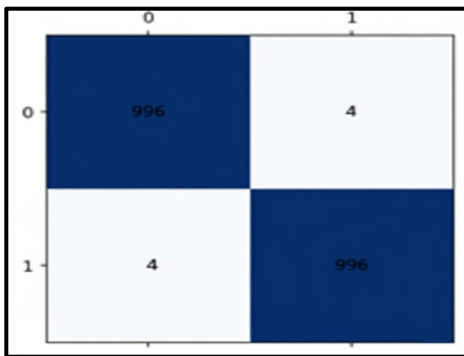


Fig. 10. Model Confusion Matrix

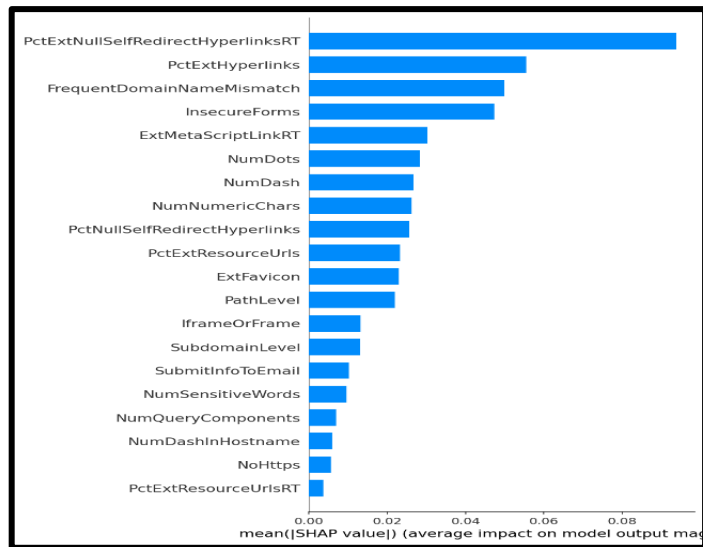


Fig. 11. SHAP mean values and impacts on the model.

Furthermore, a SHAP method (Li, 2022) was used to understand the interrelationship between the model components and to get more insight into the period of learning and testing. Fig. 11 shows the average magnitude of impact each feature has on the model's output prediction. For example, the features are ranked in descending order of their importance, with the most influential factor, the top feature, PctExtNullSelfRedirectHyperlinksRT, having the largest mean (SHAP value). While the PctExtResourceUrlsRT feature has the smallest average on the system impact. Finally, a comparative analysis has been done, since we compared the proposed model's performance with the most popular reference models that can be found in the literature review.

Table 3
Comparative Model Performance and References

Ref.	Model Architecture	Accuracy Range (Approx.)	Key Input Data
(Li, 2022)	Self-Attention CNN (with MHSA)	~95.6%	Phishing URL strings
(Xiao et al., 2020)	PHISH ATTENTION (VAE + MHSA)	~98.57%	Advanced URL features, balanced data
(Prabhakaran et al., 2025)	Multi-View Transformer (with MHSA)	>97%	URL, attributes, content, behavioral info
(Wang et al., 2023)	Hybrid CNN-LSTM/DNN	99.1%	Multidimensional features, large datasets
(Alshingiti et al., 2023)	LSTM (Long Short-Term Memory)	~96.8%	Character-level URL features (Sequential Data)
(Zakaria et al., 2024)	Random Forest (Traditional ML)	~98.2%	Lexical, Host-based, and Content-based URL Features
(Abdelaziz et al., 2020)	Naive Bayes (Traditional ML)	~94.1%	Email Content Features (Bag-of-Words or TF-IDF)
	The proposed model	99.6%	Phishing websites dataset.

Table 3 summarizes the comparative analysis. The typical ML models, like Naive Bayes, provide 94.1% and even the more influential Random Forest provides 98.2% serve as robust baselines. The performance increased with DL architectures, mainly those integrating Multi-Head Self-Attention (MHSA). For instance, the PHISH_ATTENTION provides 98.57% and the Multi-View Transformer provides as well 97%, which leverage richer input data and contextual understanding. The highest accuracy is accomplished by Hybrid models, such as the CNN-LSTM/DNN, 99.1%. On the other hand, the proposed model provides a highly accurate result with 99.6%, confirming that the proposed model yields the most robust and state-of-the-art results for current phishing threats.

5. Conclusion

This research introduces a novel approach based on Multi-Head Attention (MHA) for classifying websites as either legitimate or phishing. We demonstrate that MHA, previously employed successfully in natural language processing, can also deliver highly accurate results with website classification datasets. The proposed model achieved exceptional accuracy, reaching 99.6%, which underscores its reliability and effectiveness in phishing detection. These findings position the model as a promising solution for enhancing real-world web security. Despite these encouraging results, further research is required to improve the model's adaptability and robustness. Future work should emphasize deploying the model in real-time environments to evaluate its performance under dynamic conditions and diverse attack scenarios. Additionally, incorporating feature selection and feature extraction techniques, rather than using all available features, could reduce computational complexity and improve efficiency.

Acknowledgment

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant No. KFU260655).

References

- Abdelaziz, O., Deb, S., Hodhod, R., & Ray, L. (2020). A novel phishing email detection algorithm based on multinomial naïve Bayes classifier and natural language processing. In *Proceedings of the International Conference on Computing and Emerging Sciences*.
- Abu Zuraiq, A. M., & Alkasassbeh, M. (2019). Phishing detection based on machine learning and feature selection methods. *International Journal of Interactive Mobile Technologies*, 13(12).
- Addula, S. R., Norozpour, S., & Amin, M. (2025). Risk Assessment for Identifying Threats, vulnerabilities and countermeasures in Cloud Computing. *Jordanian Journal of Informatics and Computing*, 2025(1), 38–48. <https://doi.org/10.63180/jjic.thestap.2025.1.5>
- Alamgir, N. (2020). *Computer vision-based smoke and fire detection for outdoor environments* (Doctoral dissertation, Queensland University of Technology).
- Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing attacks: A recent comprehensive study and a new anatomy. *Frontiers in Computer Science*, 3, 563060.
- Alsharaiah, M. A., Abu-Shareha, A. A., Abualhaj, M., Baniata, L. H., Adwan, O., Al-Saaidah, A., & Oraiqat, M. (2023). A new phishing website detection framework using ensemble classification and clustering. *International Journal of Data and Network Science*, 7, 857–864.
- Alshingiti, Z., Alaql, R., Al-Muhtadi, J., Haq, Q. E. U., Saleem, K., & Faheem, M. H. (2023). A deep learning-based phishing detection system using CNN, LSTM, and LSTM-CNN. *Electronics*, 12(1), 232.
- Alzoubi, H. M., & Ghazal, T. M. (2022). The effect of e-payment and online shopping on sales growth: Evidence from the banking industry. *International Journal of Data and Network Science*, 6(4), 1369–1380.
- Asiri, S., Xiao, Y., Alzahrani, S., Li, S., & Li, T. (2023). A survey of intelligent detection designs of HTML URL phishing attacks. *IEEE Access*, 11, 6421–6443.
- Aslan, Ö., Aktuğ, S. S., Özkan-Okay, M., Yılmaz, A. A., & Akın, E. (2023). A comprehensive review of cybersecurity vulnerabilities, threats, attacks, and solutions. *Electronics*, 12(6), 1333.
- Bawany, N. Z., Shamsi, J. A., & Salah, K. (2017). DDoS attack detection and mitigation using SDN: methods, practices, and solutions. *Arabian Journal for Science and Engineering*, 42(2), 425–441.
- Fadheel, W., Abusharkh, M., & Abdel-Qader, I. (2017). On feature selection for the prediction of phishing websites. In *IEEE 15th International Conference on Dependable, Autonomic and Secure Computing* (pp. 1–6).
- Ferdous, J., Islam, R., Mahboubi, A., & Islam, M. Z. (2025). A Survey on ML Techniques for Multi-Platform Malware Detection: Securing PC, Mobile Devices, IoT, and Cloud Environments. *Sensors (Basel, Switzerland)*, 25(4), 1153.
- Foody, G. M. (2023). Challenges in the real-world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient. *PLOS ONE*, 18(10), e0291908.
- Gupta, D., Gandotra, E., Mohan, Y., & Singh, S. (2023). Analysis of ensemble methods for phishing detection. In *Intelligent Multimedia Signal Processing for Smart Ecosystems* (pp. 85–100). Springer.

- Kalla, D., Samaah, F., Kuraku, S., & Smith, N. (2023). Phishing detection implementation using Databricks and artificial intelligence. *International Journal of Computer Applications*, 185(11), 1–11.
- Kashkool, H. J. M., Farhan, H. M., Naseri, R. A. S., & Kurnaz, S. (2024). Enhancing facial recognition accuracy and efficiency through integrated CNN, PCA, and SVM techniques. In *International Conference on Forthcoming Networks and Sustainability* (pp. 57–70).
- Khaleel, Y. L., Habeeb, M. A., Albahri, A. S., Al-Quraishi, T., Albahri, O. S., & Alamoodi, A. H. (2024). Network and cybersecurity applications of defense in adversarial attacks: A state-of-the-art using machine learning and deep learning methods. *Journal of Intelligent Systems*, 33(1), 20240153.
- Krasnodębska, K., Goch, W., Uhl, J. H., Verstegen, J. A., & Pesaresi, M. (2025). Advancing precision, recall, F-score, and Jaccard index: An approach for continuous, ratio-scale measurements. *Environmental Modelling & Software*, 193, 106614.
- Lertampaiporn, S., Thammarongtham, C., Nukoolkit, C., Kaewkamnerdpong, B., & Ruengitchatchawalya, M. (2013). Heterogeneous ensemble approach with discriminative features and modified-SMOTEbagging for pre-miRNA classification. *Nucleic acids research*, 41(1), e21–e21.
- Li, Z. (2022). Extracting spatial effects from machine learning models using local interpretation methods: An example of SHAP and XGBoost. *Computers, Environment and Urban Systems*, 96, 101845.
- Lim, K., Park, J., & Kim, D. (2024). Phishing vs. legit: Comparative analysis of client-side resources of phishing and target brand websites. In *Proceedings of the ACM Web Conference* (pp. 1756–1767).
- Majgave, A. B., & Gavankar, N. L. (2024). Automatic phishing website detection and prevention model using transformer deep belief network. *Computers & Security*, 147, 104071.
- Owa, K., & Adewole, O. (2025). Benchmarking machine learning techniques for phishing detection and secure URL classification. *International Journal of Computer Science and Mobile Computing*, 14(1), 20–37.
- Pathak, P., & Shrivastava, A. K. (2024). Development of a proposed model using random forest with optimization technique for classification of phishing websites. *SN Computer Science*, 5(8), 1059.
- Prabhakaran, M. K., Chandrasekar, A. D., & Meenakshi Sundaram, P. (2025). PHISH_ATTENTION: Achieving robust phishing website detection with balanced datasets and advanced URL features. *The Computer Journal*.
- Rao, R. S., Pais, A. R., & Anand, P. (2023). A heuristic technique to detect phishing websites using TWSVM classifier. *Neural Computing and Applications*, 33(11), 5733–5752.
- Ren, J., Guo, Y., Zhang, D., Liu, Q., & Zhang, Y. (2018). Distributed and efficient object detection in edge computing: Challenges and solutions. *IEEE Network*, 32(6), 137–143.
- St-Aubin, P., & Agard, B. (2022). Precision and reliability of forecast performance metrics. *Forecasting*, 4(4), 882–903.
- Sykes, B., Simon, L., & Rabin, J. (2024). Unifying and extending precision–recall metrics for assessing generative models. *arXiv preprint arXiv:2405.01611*.
- Tang, L., & Mahmoud, Q. H. (2021). A deep learning-based framework for phishing website detection. *IEEE Access*, 10, 1509–1521.
- Ubung, A. A., Jasmi, S. K. B., Abdullah, A., Jhanjhi, N. Z., & Supramaniam, M. (2019). Phishing website detection: Improved accuracy through feature selection and ensemble learning. *International Journal of Advanced Computer Science and Applications*, 10(1).
- Vacalares, S. T., Ana, B. P. E. S., Dranto, D. Q., & Gallano, J. S. (2024). Bank emails: The language of legit and scam. *International Journal of Research and Review*.
- Vaghela, R. S. (2023). Exploring feature importance in phishing URL detection models. *Journal of Cyber Security and Digital Forensics*, 2(2), 27–34.
- Wang, Y., Ma, W., Xu, H., Liu, Y., & Yin, P. (2023). A lightweight multi-view learning approach for phishing attack detection using transformer with mixture of experts. *Applied Sciences*, 13(13), 7429.
- Weng, J., Jia, X., & Liu, Y. (2023). Study on deep learning-based natural scene text recognition. *Academic Journal of Computing & Information Science*, 6(2), 44–52.
- Xiao, X., Zhang, D., Hu, G., Jiang, Y., & Xia, S. (2020). CNN–MHSA: A Convolutional Neural Network and multi-head self-attention combined approach for detecting phishing websites. *Neural Networks*, 125, 303–312.
- Yusuf, M., Kana, A. F. D., Bagiwa, M. A., & Abdullahi, M. (2024). Multi-classification of breast cancer histopathological images using enhanced shallow convolutional neural network. *Journal of Engineering and Applied Science*, 72(1), 24.
- Zakaria, N. H., Mansor, N. S., Husni, H., & Mohammed, F. (2024). *Computing and Informatics*.
- Zurairq, A. A., & Alkasassbeh, M. (2019). Phishing detection approaches. In *2019 2nd International Conference on New Trends in Computing Sciences (ICTCS)* (pp. 1–6). IEEE.

