JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(3), SEPTEMBER, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

A Hierarchical Attention Network for Explainable Phishing URL Detection on Nigerian Financial Platforms

Faridah Adamu Hobon, Ridwan Kolapo, Ibrahim Anka Salihu
Department of Software Engineering, Faculty of Computing, Nile University of Nigeria, Abuja, Nigeria

ABSTRACT

Phishing remains the major channel of cyber-fraud in Nigeria, where fraudulent uniform resource locators (URLs) are used to pretend to be legitimate banks, mobile money services, and digital payments channels in order to steal users' credentials, Bank Verification Number (BVN) and card details. Current detection solutions used by Nigerian financial organisations depend heavily on blacklists, heuristics, and traditional machine learning classifiers working as black boxes, which do not give any explanation of why a particular URL was classified as malicious and show poor generalization to emerging phishing schemes. In this paper, we propose a Hierarchical Attention Network (HAN) that views a URL as a two-level hierarchy of domain and path, where each of them is segmented into tokens and apply attention both at token and segment levels to detect phishing URLs and highlight the specific parts of a URL contributing to this classification. Our model is based on 164,548 Nigerian bank-related URLs extracted from the 822,010-row Kaggle Phishing and Legitimate URLs dataset with 80/20 stratified train/test split. The baseline is a flattened BiLSTM network without attention layers. The baseline gave 94.37% accuracy (phishing F1 score = 0.9421), while our HAN scored 94.05% accuracy (phishing F1 score = 0.9373) along with higher stability of validation accuracy curve and smaller false-positive rate, meaning that slight drop in accuracy is compensated by increased precision and interpretability. The visualisation of the attention weights showed that the model always gives the most weight to the structural anomalies like abnormal subdomain tokens, combined or hyphenated strings, and non-standard prefixes like 'secure' or 'mail', but not to brand names of the banks themselves.

ARTICLE INFO

Article History
Received: May, 2026
Received in revised form: May, 2026
Accepted: August, 2026
Published online: September, 2026

KEYWORDS

Phishing URL Detection; Hierarchical Attention Network; Explainable Artificial Intelligence; Deep Learning; Cybersecurity; Nigerian Financial Platforms

INTRODUCTION

The advent of digital banking, mobile money, and other forms of financial technology solutions in Nigeria has brought an increase in cyber-enabled financial fraud activities, phishing being the most prominent among them. Phishing involves the use of deception by pretending to be someone or something trustworthy via a deceptive website to extract valuable and confidential information from a victim such as their usernames and passwords, Bank Verification Number (BVN), and card details (Ayorinde, 2025). The fraudsters send out these malicious websites via SMS, WhatsApp, emails, and social media channels using the naivety of Nigerians' trust in familiar website addresses. The Nigeria Inter Bank Settlement System (NIBSS) has reported a consistent yearly increase in losses arising from fraud activities in the banks, with phishing being one of the top sources of such activities (NIBSS, 2023; Fatokun et al., 2025).

The financial institutions as well as fintech firms operating in Nigeria presently use a combination of domain blacklisting, rule-based heuristics, and conventional machine learning algorithms like Random Forest and SVM to detect

Corresponding author: Faridah Adamu Hobon
faridahadamuhobon@gmail.com
Department of Software Engineering, Faculty of Computing, Nile University of Nigeria, Abuja, Nigeria.

harmful URLs. Such approaches are easily deployable but suffer from two shortcomings. First, blacklists and static rules are reactive systems since they can only detect URLs that have been previously reported, thereby creating a period of vulnerability for new registration of the phishing domains (George et al., 2025). Second, the machine learning based systems also behave like opaque black boxes since their decisions are not accompanied by any information regarding the features of the URLs responsible for the classification outcome (Fajar et al., 2024).

There are two gaps in knowledge that form the basis of this research. The rule based and blacklist-based systems lack the ability to generalize beyond threats that have already been seen before, and conventional classifiers in machine learning are unable to understand the sequential nature and structure that differentiate a phishing URL from a real URL (George et al., 2025). Where deep learning models have been used, they have mainly been assessed based on global benchmark datasets which are not representative of the lexicons and structure of Nigerian banking sites such as the combination of the .com and .com.ng versions and thus their models are not interpretable (Fajar et al., 2024). This paper attempts to fill these gaps through the development of a Hierarchical Attention Network (HAN) in detecting phishing URLs in the context of Nigerian financial websites.

Specifically, the objective will be to create a two-level hierarchy of URLs' domain and path elements, develop the HAN model along with a non hierarchical BiLSTM model, and compare both models on accuracy, precision, recall, F1 score and the confusion matrix as well as evaluate the interpretability of both models. The study is restricted to URL based features; webpage content, network-based information, and user interaction data are not used.

EVALUATION OF EXISTING LITERATURE

Phishing detection has shifted from using static rules to employing learning approaches that can respond to emerging trends in attacks. Blacklisting is the approach used in early detection which involved blocking phishing websites; however, this is an elementary approach since new registered phishing domains will not be detected unless they have been reported (Lim et al., 2020). The use of heuristics as a method for phishing detection is another step forward in that it involved the verification of HTTPS, WHOIS, and search engine rankings. This is however becoming ineffective since attackers can now get HTTPS certificates for their websites and also host phishing websites on legitimate compromised domains (Banik et al., 2022).

The development of machine learning based approaches was aimed at learning discriminative patterns from labeled URL and page features (Ngo et al., 2024). As demonstrated in Gupta et al. (2021), just nine lexical features like URL length, domain length, and counts of tokens, digits, and delimiters were sufficient to develop a Random Forest classifier capable of reaching 99.57% accuracy, while Korkmaz et al. (2020) developed classifiers using 58 lexical features resulting in accuracies ranging from 90.5% to 94.6% in three datasets. Zamir et al. (2020) utilized a number of feature selection methods on an 11,055 URLs data set and achieved 97.3% accuracy using Random Forests, while Rashid et al. (2020) applied Principal Component Analysis on 30 features reducing them to five components and achieved 95.66% accuracy using Support Vector Machines.

Karim et al. (2023) used feature selection together with cross-validation and grid search to achieve 98.12% accuracy, one of the highest results obtained in the application of classical machine learning algorithms to this task. Nevertheless, machine learning pipelines require manual designing of feature sets and after being trained provide only a simple ranking of feature importance but no explanation on how the tokens in the particular URL result in a certain prediction. Deep learning approaches have rendered manual feature engineering obsolete by extracting the features directly from raw URL strings (Aljofey et al., 2020). Aljofey et al. (2020) suggested a character level convolutional neural network approach that obtained accuracies ranging from 95.0% to 98.6% over four datasets with a latency

Corresponding author: Faridah Adamu Hobon
faridahadamuhobon@gmail.com
Department of Software Engineering, Faculty of Computing, Nile University of Nigeria, Abuja, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(3), SEPTEMBER, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

of 0.47 milliseconds per URL and real-time usability but had poor results for short URLs.

Tajaddodianfar et al. (2020) introduced Texception, a dual character level and word level model which yielded a true positive rate of 47.95% at a very strict false positive rate of 0.01%, indicating the importance of analyzing the detectors at the right operating point. Ozcan et al. (2021) used hand-crafted features from natural language processing alongside LSTM and BiLSTM branches to obtain 99.21% accuracy on the PhishTank dataset, whereas Adebowale et al. (2023) used CNN images and LSTM text features from webpages along with webpage frame information to achieve 93.28% accuracy.

Many data sources have a natural hierarchical structure, and URLs are no exception: they comprise domain and path segments that are themselves composed of tokens (Asiri et al., 2023). Hierarchical Attention Networks exploit this structure by applying attention at two nested levels and were first shown to improve both accuracy and interpretability in document classification by highlighting the words and sentences that most influenced a prediction (Chanaa, 2021). This attention-based interpretability has since been extended to emotion cause extraction (Hu et al., 2021), context aware machine translation (Jaziriyan & Ghaderi, 2023), and multimodal analysis of phishing webpages (Chai et al., 2021). The additive attention formulation used in this study, in which a learned context vector produces normalised importance weights over hidden states, was first introduced for machine translation (Bahdanau et al., 2014).

On balance, there have been improvements in terms of phishing detection accuracy in the current literature, but there are two key areas where there have been limitations noted. First, most of the highly accurate models (Aljofey et al., 2020; Ozcan et al., 2021) have been limited to only reporting their results as aggregate measures without revealing which features of the URL were used in reaching decisions, and this limitation will severely hamper the efforts of those trying to understand the reasons for the generation of an alert. Secondly, the majority of the studies reviewed have used the globally known datasets for training their algorithms without considering the peculiarities of phishing attempts in the Nigerian financial context (NIBSS, 2023; Fatokun et al., 2025).

MATERIALS AND METHOD

Two models of deep learning architectures were designed and compared – namely, a flattened BiLSTM model and a Hierarchical Attention Network. This research was conducted using a filtered dataset of Nigeria's bank URLs and is described in the subsections presented below.

Dataset

The dataset for this study is derived from the publicly available dataset titled "Phishing and Legitimate URLs" available on Kaggle, accessed using the KaggleHub API. The data comprises of 822,010 instances with a column for URL strings along with a binary label; 0 stands for phishing and 1 stand for legitimate. Out of the total, 394,982 (48.05%) are phishing while 427,028 (51.95%) are legitimate, as depicted in Figure 1 below.

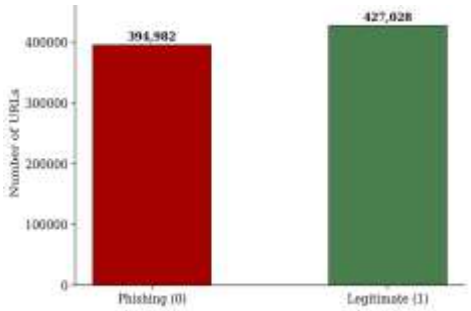

Figure 1: Class distribution of the source Phishing and Legitimate URLs dataset (n = 822,010)

As the objective was to generate threat models based on phishing attacks targeting the Nigerian finance industry, a relevance filter for Nigerian banks was employed. This involved creating a list of genuine domains of top Nigerian commercial banks, including the com and com.ng variants, and retaining URLs that had an identifiable Nigerian bank domain, regardless of whether it was the actual domain or a variant

Corresponding author: Faridah Adamu Hobon
faridahadamuhobon@gmail.com
Department of Software Engineering, Faculty of Computing, Nile University of Nigeria, Abuja, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(3), SEPTEMBER, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

domain. URL character length was also examined and showed a visible difference between the two classes, supporting its indirect use as a signal through the segment and token structure described next. The filtered dataset was split into training and test sets using a stratified 80:20 split with a fixed random seed of 42.

Preprocessing and Hierarchical URL Representation

Each retained URL was decomposed into a two-level hierarchy that mirrors the structure of a web address. The tldextract library separated the domain component into subdomain, domain, and top-level domain parts, since domain level irregularities such as misspellings, lookalike characters, or unexpected top-level domains are a strong signal of phishing intent. The remaining part of the URL was segmented at the path level with a regular expression tokenizer that splits on dots, slashes, hyphens, underscores, question marks, and ampersands, giving a two-segment hierarchy of domain tokens and path tokens for every URL. A global tokenizer was then fitted on the training tokens, with an out of vocabulary token to handle novel or randomised strings. A maximum of two segments and a fixed number of tokens per segment were used to give the model a consistent input shape; shorter sequences were padded and longer ones truncated. Figure 2 summarises this pipeline.

Figure 2: Data preprocessing pipeline, from raw dataset filtering to URL segmentation, tokenization, and train and test splitting

Model Development

A non-hierarchical baseline was built by reshaping the hierarchical token tensor into a flat sequence, discarding the explicit boundary between domain and path tokens. The baseline is a Keras and TensorFlow model with an embedding layer that maps each token to a 128-dimensional vector, a single bidirectional LSTM layer with 64 hidden units per direction, a dropout layer at rate 0.5, and a dense sigmoid output producing the phishing probability.

The proposed HAN keeps the hierarchical structure and applies attention at both the token and segment level, using a word level encoder followed by a segment level encoder, as shown in Figure 3. Given hidden states H = {h1, h2, ..., hT} from a recurrent encoder, an attention layer computes an importance score for each state and normalises it with a softmax function, then forms a context vector as the weighted sum of the hidden states:

$$e_t = \tanh(W h_t + b) \quad (1)$$
$$\alpha_t = \exp(e_t) / \Sigma_k \exp(e_k) \quad (2)$$
$$v = \Sigma_t \alpha_t h_t \quad (3)$$

At the word level, each URL segment is treated as a sequence of tokens. Every token is embedded in 128 dimensions and passed through a bidirectional LSTM, and the attention formulation above is applied over the tokens in the segment to produce a single segment vector. Empty, zero padded segments are excluded from this step. At the segment level, the resulting segment vectors are passed through a second bidirectional LSTM, and the same attention formulation is applied across segments to produce one attention weighted representation of the whole URL. By comparing the segment level weights, the model shows whether the domain segment or the path segment contributed more strongly to a given classification. The final representation is passed through a dense layer with sigmoid activation to give the phishing probability. Figure 4 shows the generic attention sub module used at both levels.

Corresponding author: Faridah Adamu Hobon
faridahadamuhobon@gmail.com
Department of Software Engineering, Faculty of Computing, Nile University of Nigeria, Abuja, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(3), SEPTEMBER, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

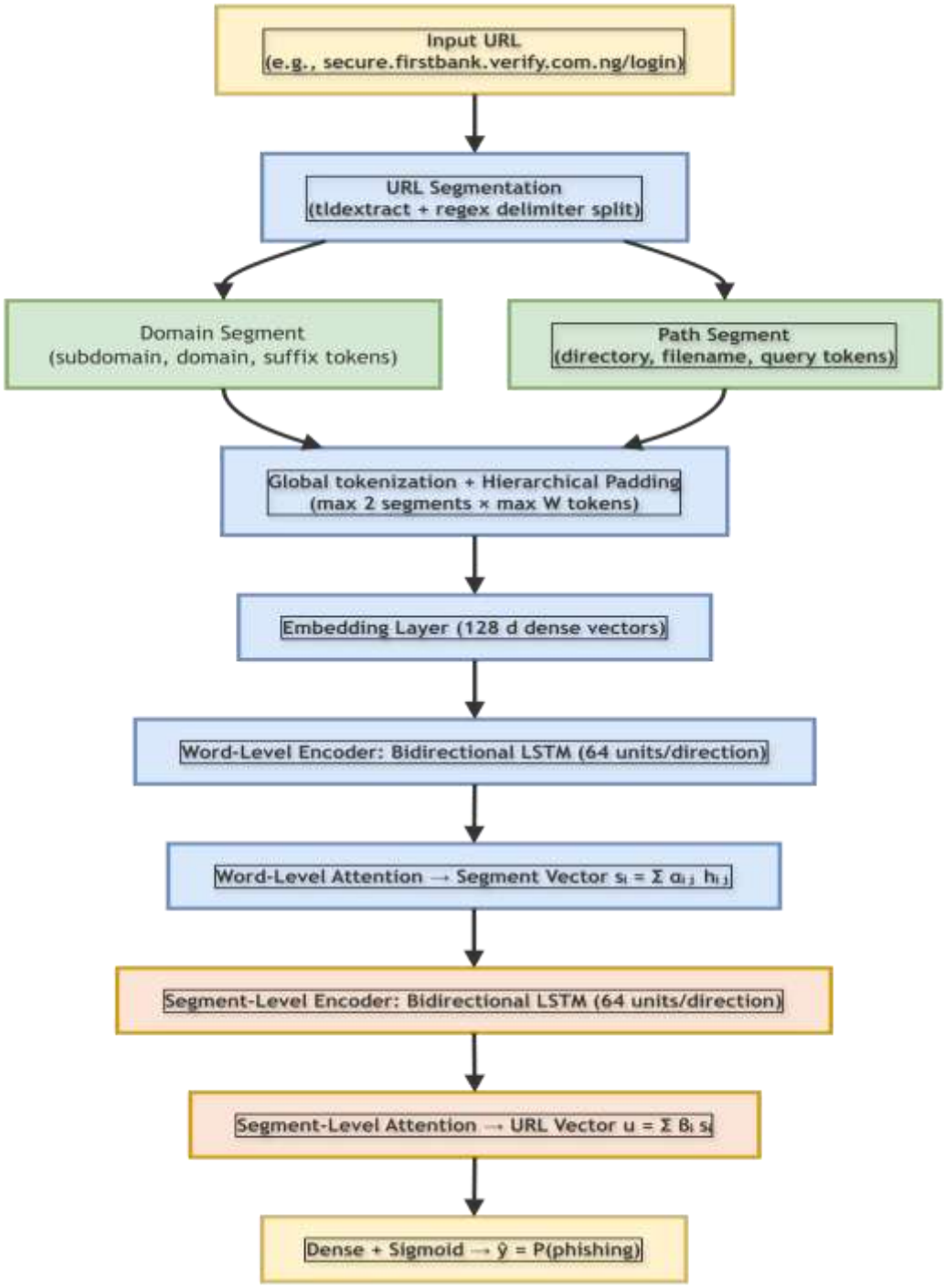

Figure 3: Architecture of the proposed Hierarchical Attention Network, from raw URL input to the phishing probability output

Corresponding author: Faridah Adamu Hobon
faridahadamuhobon@gmail.com
Department of Software Engineering, Faculty of Computing, Nile University of Nigeria, Abuja, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(3), SEPTEMBER, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

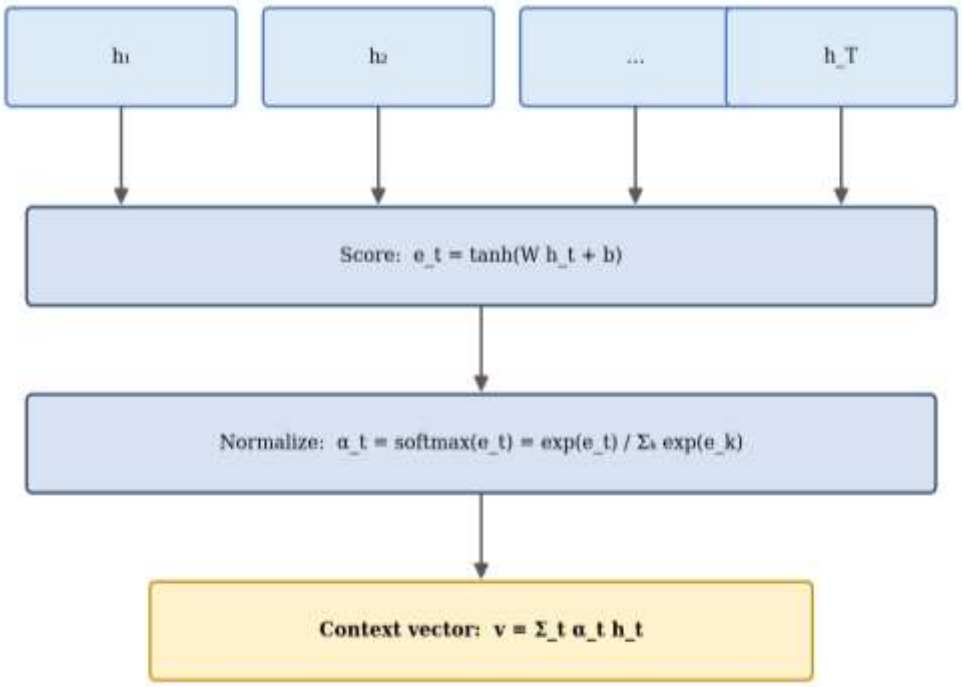

Figure 4: Attention sub module: score computation, softmax normalisation, and context vector aggregation, applied at the word level and the segment level

The embedding dimension for the HAN was set to 128 to match the baseline, both LSTM layers used 64 hidden units, and a dropout rate of 0.5 was applied, so that any difference in performance can be attributed to the hierarchical attention structure rather than a hyperparameter advantage. Both models were trained with the Adam optimiser and binary cross entropy loss for five epochs with a batch size of 256, holding out 10% of the training data for validation. Table 1 summarises both architectures.

Table 1: Summary of the baseline and proposed model architectures. V = vocabulary size; S = segments per URL (2); W = tokens per segment

Model	Layer	Description	Output shape
Baseline BiLSTM	Embedding	Token to 128 d dense vector	(None, S·W, 128)
Baseline BiLSTM	BiLSTM (64 units)	Forward and backward context	(None, 128)
Baseline BiLSTM	Dropout + Dense	Rate 0.5; sigmoid output	(None, 1)
Proposed HAN	Embedding + word BiLSTM	128 d vectors; per segment context	(None, S, W, 128)
Proposed HAN	Word attention	Token importance to segment vector	(None, S, 128)
Proposed HAN	Segment BiLSTM + attention	Dependencies and importance across segments	(None, 128)
Proposed HAN	Dense (sigmoid)	Phishing probability	(None, 1)

Corresponding author: Faridah Adamu Hobon
faridahadamuhobon@gmail.com
Department of Software Engineering, Faculty of Computing, Nile University of Nigeria, Abuja, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(3), SEPTEMBER, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

Model Evaluation

Both models were evaluated on the same 80:20 stratified split using accuracy, precision, recall, specificity, and F1 score, computed from the confusion matrix counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN):

Accuracy = (TP + TN) / (TP + FP + TN + FN) (4)

Recall = TP / (TP + FN), Precision = TP / (TP + FP) (5)

F1 score = 2 × (Precision × Recall) / (Precision + Recall) (6)

Recall is the operationally critical metric in this setting, since a false negative, a phishing URL wrongly labelled legitimate, carries far greater financial risk than a false positive that only inconveniences a legitimate user. Token and segment level attention weights were also extracted and visualised for representative test set URLs to assess the explainability of the proposed model.

RESULTS

This section presents the results obtained for the performances of the baseline and proposed approaches for the Nigerian bank filtered dataset consisting of 164,548 URLs.

Baseline Model Performance

Accuracy obtained by using the flattened BiLSTM baseline was 94.37%. The precision, recall, and F1 score for each class of the flat BiLSTM baseline is shown in Table 2. If the phishing precision is 0.9313, then it indicates that most of the URLs classified as phishing were indeed phishing. Similarly, phishing recall is 0.9531.

Table 2: Baseline (flattened BiLSTM) classification report

Class	Precision	Recall	F1 score
Phishing	0.9313	0.9531	0.9421
Legitimate	0.9557	0.9350	0.9452

Proposed HAN Model Performance

The developed HAN model achieved an accuracy of 94.05% in testing, very near to the accuracy achieved by the baseline. The per-class scores of the HAN have been presented in Table 3. The HAN model's phishing precision of 0.9500 is higher than the baseline, which implies that there will be fewer false phishing detections, while its phishing recall of 0.9249 is lower than the baseline.

Table 3: Proposed HAN classification report

Class	Precision	Recall	F1 score
Phishing	0.9500	0.9249	0.9373
Legitimate	0.9322	0.9550	0.9434

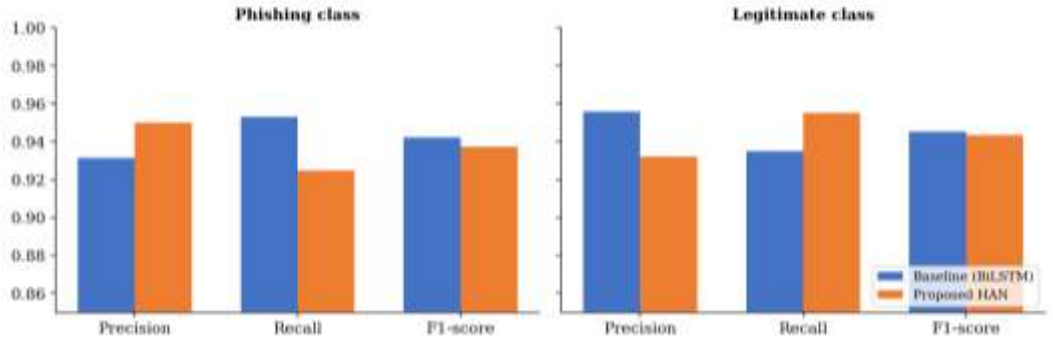

Figure 5: Per class precision, recall, and F1 score for the baseline and proposed HAN models

Corresponding author: Faridah Adamu Hobon
faridahadamuhobon@gmail.com
Department of Software Engineering, Faculty of Computing, Nile University of Nigeria, Abuja, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(3), SEPTEMBER, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

Comparative Analysis

Table 4 summarizes both overall accuracy and confusion matrix totals for both models. The baseline model is winning by a mere 0.32 percentage point, which isn't enough to be considered an advantage in itself. The HAN has lower false positive rates (3,845 compared to 5,552), corresponding to its greater precision for phishing, while the baseline has lower false negative rates (3,701 compared to 5,934). These figures must be taken as approximations, because there appears to be some sort of rounding problem when comparing them to the recall scores in the classification report.

Table 4: Overall accuracy and confusion matrix counts for both models (phishing = positive class)

Model	Accuracy	TP	TN	FP	FN
Baseline	0.9437	80,000	75,295	5,552	3,701
HAN	0.9405	73,062	81,561	3,845	5,934

The validation accuracy of the baseline rose quickly to about 94.95% within two epochs, then dropped to about 93.75% at epoch three before recovering to 94.25% by epoch five, a pattern consistent with mild overfitting. The HAN, however, had an easier trajectory, beginning near 94.85%, then declining to 93.25% at epoch three and rising to 93.85% by epoch five. This is in line with the hypothesis that attention tends to suppress less relevant elements. Figures 6 and 7 illustrate these trends.

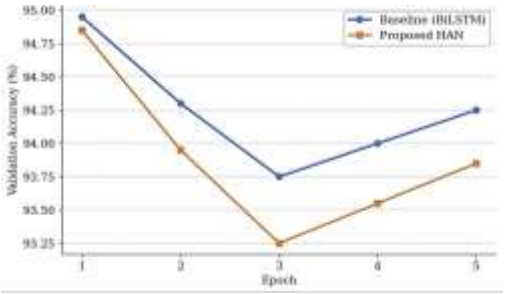

Figure 6: Validation accuracy trajectories for the baseline and proposed HAN models across five training epochs

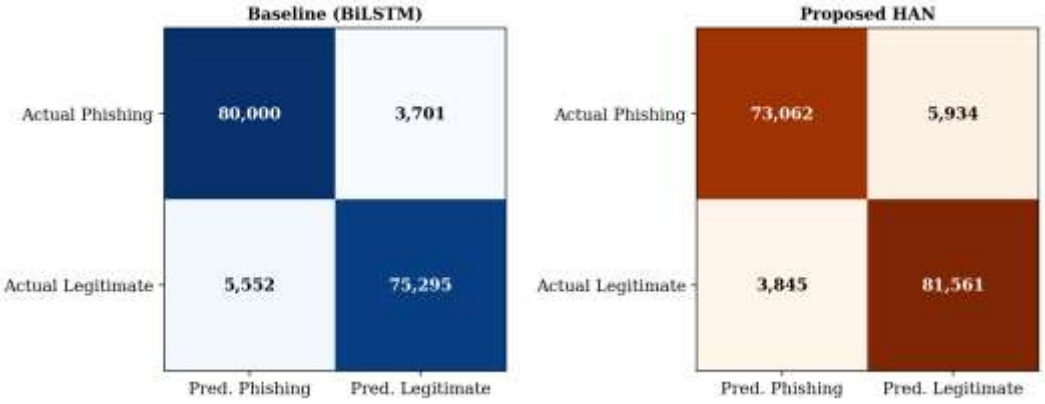

Figure 7: Confusion matrix heatmaps for the baseline and proposed HAN models

Explainability via Attention Weight Visualisation

Token level and segment level attention weights were gathered from representative phishing and benign URLs and presented as heat maps, where the more attention is being paid to a certain token or segment, the darker its color will be. Three major patterns have been detected regularly. An unusually high amount of subdomains, strange separators, and abnormal

Corresponding author: Faridah Adamu Hobon
faridahadamuhobon@gmail.com
Department of Software Engineering, Faculty of Computing, Nile University of Nigeria, Abuja, Nigeria.

prefixes get higher attention than just the presence of a bank word, which means that the model reacts on structural anomaly, not just on the presence of a word. Hyphenated tokens and composite tokens, created by connecting two random words together, in order to imitate a legitimate website structure, received increased attention as well. The tokens positioned before the primary domain like 'mail', 'secure', or any random alphanumeric sequences got higher attention levels than bank name tokens, meaning that the model values structure positioning more than the presence of a brand name, which is important since the bank names are present in both genuine and fraudulent URLs.

Summary of Results

It can be observed from the experiments that a hierarchical structure in the attention mechanism can produce results comparable to the non-hierarchical baseline with higher stability during training and improved phishing detection precision along with token-level and segment-level justification for all predictions. The results are in line with other research studies regarding phishing detection algorithms, and the attention mechanism analysis further confirms that the model works using structural features rather than keywords.

DISCUSSION OF FINDINGS

Model Effectiveness and Interpretability

The key result is that HAN sacrifices minimal aggregate accuracy of 0.32 percentage points for smoothness in training process, high phishing precision, and explanations at token and segment level for individual predictions. In a financial security context, this is a favourable trade-off, as missing a phishing URL results in direct theft of credentials and money, whereas a false alert of a legitimate URL results only in a nuisance. An explanation for a prediction helps as well, as it allows quick verification of an alert by the analysts, updating the model to recognize new patterns, and explaining the block to affected clients.

Comparison with Existing Literature

The accuracy of 94.05% to 94.37% in this paper is generally consistent with, but a little less than, similar studies in deep learning phishing detectors using global benchmark datasets. Adebowale et al. (2023) obtained 93.28% accuracy using a hybrid CNN LSTM that also incorporates images of webpages, Aljofey et al. (2020) obtained 95.0% to 98.6% using a CNN at character level, and Ozcan et al. (2021) got an accuracy of up to 99.21% using hybrid DNN BiLSTM with handcrafted features. It is understandable that the slightly lower accuracy in this study could be attributed to the small-sized and specifically focused Nigerian bank filter dataset and the short five-epoch training process applied to both models.

Implications and Limitations

For Nigerian banks, mobile money operators, and digital payments services, an explainable phishing detector based on an understanding of structural and lexical characteristics could be integrated into the security process chain, where it would filter suspicious links and, at the same time, provide the tokens or other features that trigger the decision, eliminating the need for further inspection and decreasing reaction time. As the model was trained using data of Nigerian banks, it will also be better equipped than globally trained models to detect spoofing techniques such as.com.ng URLs.

There are a number of limitations to this study. Firstly, it only makes use of URL lexical and structural features and not webpage content, visual similarity, network-level and user behaviour data. The confusion matrix raw values differ from the recall values in the classification report and need to be aligned based on the predictions from the saved model. No hardware information and software versions were logged during the training and prediction processes, which reduces the possibility of deploying the system in real-time. The number of training epochs and the number of RNN layers per stage in both models were very limited (five and one, respectively), and the differences found are simply point estimates and

Corresponding author: Faridah Adamu Hobon
faridahadamuhobon@gmail.com
Department of Software Engineering, Faculty of Computing, Nile University of Nigeria, Abuja, Nigeria.

cannot be confirmed statistically. The list of Nigerian bank domains for filtering was not exhaustive.

Future Directions

The future research needs to test the performance of the HAN inference in terms of latency and memory requirements using actual hardware in order to answer the question of real-time deployment. There is potential to develop the model into a multimodal version which incorporates both webpage content and visual features in addition to the URL-based hierarchical attention mechanism that we used. The confusion matrix analysis needs to be redone using saved prediction values and combined with a test of statistical significance, for example McNemar's test. The Nigerian bank domain list needs to be expanded to include more financial technology and mobile money service sites, as well as longitudinal testing.

CONCLUSION

This study designed and analysed a Hierarchical Attention Network for interpretable detection of phishing URLs on the financial sector of Nigeria, solving issues of black-box nature of previous highly accurate classifiers and their lack of Nigerian-specific understanding of the phishing problem. On a subset of the public URL database filtered specifically for Nigerian banks, a two-layer hierarchical representation of domain and path was created, and a HAN architecture with attention mechanism at each of its layers was trained in parallel with BiLSTM baseline network.

The HAN achieved accuracy score of 94.05% – nearly as good as baseline's 94.37% while displaying better convergence pattern, greater phishing precision and capability of highlighting tokens and parts of URL contributing to a particular classification decision. Visualization of attention weights revealed that HAN prefers to pay attention to structural inconsistencies, hyphenated compound tokens, and unusual subdomains prefixes rather than mere mention of bank brand name, providing transparent and granular explanations missing in conventional classifiers. This research suggests that when selecting a model for phishing URL detection, one must consider not only accuracy but minimize missed detections and improve interpretability as well, which is especially important in financial sector of Nigeria.

Future Work

Further development of this research would include multimodal inputs that integrate webpage information and visual characteristics in addition to the URL-based hierarchical attention employed herein, benchmarks of latency and throughput on actual hardware, as well as an extended set of Nigerian bank domains for continual updates.

REFERENCES

Adebowale, M. A., Lwin, K. T., & Hossain, M. A. (2023). Intelligent phishing detection scheme using deep learning algorithms. Journal of Enterprise Information Management, 36(3), 747 to 766.

Aljofey, A., Jiang, Q., Qu, Q., Huang, M., & Niyigena, J. P. (2020). An effective phishing detection model based on character level convolutional neural network from URL. Electronics, 9(9), 1514.

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 to 6443.

Ayorinde, A. S. (2025). Explainable deep learning models for detecting sophisticated cyber enabled financial fraud across multi layered FinTech infrastructure. International Journal of Cybersecurity and Digital Forensics, 5(3), 241 to 263.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

Banik, S., Dandyala, S. S. M., & Nadimpalli, S. V. (2022). Heuristic based detection techniques. International Journal of

Corresponding author: Faridah Adamu Hobon
faridahadamuhobon@gmail.com
Department of Software Engineering, Faculty of Computing, Nile University of Nigeria, Abuja, Nigeria.

Advanced Engineering Technologies and Innovations, 1(2), 352 to 362.

Chai, Y., Zhou, Y., Li, W., & Jiang, Y. (2021). An explainable multi modal hierarchical attention model for developing phishing threat intelligence. IEEE Transactions on Dependable and Secure Computing, 19(2), 790 to 803.

Chanaa, A. (2021). E learning text sentiment classification using hierarchical attention network (HAN). International Journal of Emerging Technologies in Learning, 16(13), 157 to 167.

Fajar, A., Yazid, S., & Budi, I. (2024). Comparative analysis of black box and white box machine learning model in phishing detection. arXiv preprint arXiv:2412.02084.

Fatokun, J. O., Sikiru, M., Balogun, F., & Okorie, D. D. (2025). Fraud detection in Nigerian investment advisory sector using machine learning algorithms. FUDMA Journal of Sciences, 9(10), 5 to 11.

George, M. Z. H., Alam, M. K., & Hasan, M. T. (2025). Machine learning for fraud detection in digital banking: A systematic literature review. arXiv preprint arXiv:2510.05167.

Gupta, B. B., Yadav, K., Razzak, I., Psannis, K., Castiglione, A., & Chang, X. (2021). A novel approach for phishing URLs detection using lexical based machine learning in a real time environment. Computer Communications, 175, 47 to 57.

Hu, G., Lu, G., & Zhao, Y. (2021). Bidirectional hierarchical attention networks based on document level context for emotion cause extraction. Findings of the Association for Computational Linguistics: EMNLP 2021, 558 to 568.

Jain, A. K., Parashar, S., Katare, P., & Sharma, I. (2020). Phishskape: A content based approach to escape phishing attacks. Procedia Computer Science, 171, 1102 to 1109.

Jaziriyan, M. M., & Ghaderi, F. (2023). Automatic post editing of hierarchical attention networks for improved context aware neural machine translation. Journal of AI and Data Mining, 11(1), 95 to 102.

Karim, A., Shahroz, M., Mustofa, K., Belhaouari, S. B., & Joga, S. R. K. (2023). Phishing detection system through hybrid machine learning based on URL. IEEE Access, 11, 36805 to 36822.

Korkmaz, M., Sahingoz, O. K., & Diri, B. (2020). Detection of phishing websites by using machine learning based URL analysis. Proceedings of the 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 1 to 7.

Lim, W. H., Liew, W. F., Lum, C. Y., & Lee, S. F. (2020). Phishing security: Attack, detection, and prevention mechanisms. Proceedings of the International Conference on Digital Transformation and Applications (ICDXA).

Ngo, V. D., Vuong, T. C., Van Luong, T., & Tran, H. (2024). Machine learning based intrusion detection: Feature selection versus feature extraction. Cluster Computing, 27(3), 2365 to 2379.

Nigeria Inter Bank Settlement System (NIBSS). (2023). 2023 annual fraud landscape report. Nigeria Inter Bank Settlement System Plc.

Ozcan, A., Catal, C., Donmez, E., & Senturk, B. (2021). A hybrid DNN LSTM model for detecting phishing URLs. Neural Computing and Applications, 35(7), 4957 to 4973.

Rashid, J., Mahmood, T., Nisar, M. W., & Nazir, T. (2020). Phishing detection using machine learning technique. Proceedings of the 2020 1st International Conference of Smart Systems and Emerging Technologies (SMART TECH), 43 to 46.

Tajaddodianfar, F., Stokes, J. W., & Gururajan, A. (2020). Texception: A character or word level deep learning model for phishing URL detection. Proceedings of ICASSP 2020, 2857 to 2861.

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 to 80.

Corresponding author: Faridah Adamu Hobon
faridahadamuhobon@gmail.com
Department of Software Engineering, Faculty of Computing, Nile University of Nigeria, Abuja, Nigeria.