JOURNAL OF SCIENCE, TECHNOLOGY AND EDUCATION (JSTE)

A PUBLICATION OF THE DEPARTMENT OF SCIENCE, TECHNOLOGY AND MATHEMATICS EDUCATION (STME), NASARAWA STATE UNIVERSITY, KEFFI, NIGERIA

VOLUME 10

ISSN: 2651-5539

JOURNAL OF SCIENCE, TECHNOLOGY AND EDUCATION (JSTE) VOL. 10. NO. 15

JSTE PUBLICATIONS, 2026 www.nsukjste.com/

ISSN: 2651-5539

AN ENHANCED SHORT MESSAGE SERVICE PHISHING DETECTION MODEL USING DEEP LEARNING.

*[1]Ishaq, A. [2]Salele, Z. I., [3]Aliyu, A. A. and [4]Awwalu, J.

[1, 3, 4]Department of Computer Science, Federal University Dutse, Nigeria.

[2] Department of Information Technology, Federal University Dutse, Nigeria.

Corresponding Author: a.ishaq@fud.edu.ng

Citation: Ishaq, A., Salele, Z. I., Aliyu, A. A. & Awwalu, J. (2026). An enhanced short message service phishing detection model using deep learning. *Journal of Science, Technology, and Education (JSTE);* www.nsukjste.com/. *10*(15), 189-202.

Abstract

The rapid expansion of mobile connectivity in Nigeria has created a fertile ground for SMS phishing (smishing) attacks. Conventional detection methods often fail to adapt to the evolving tactics used by scammers, leaving the country's large mobile population vulnerable, Short Message Service (SMS) is still a vital communication tool in our daily life activities, even with the quick development of Internet protocol-based messaging services. This research focuses on the Nigerian context, analyzing local smishing campaigns to develop a more effective, tailored detection model. The study aims to enhance cyber security defenses specifically for Nigeria's unique digital landscape. This thesis proposes building a domain-specific deep learning model for Nigeria. This system is designed to accurately classify SMS messages by directly tackling the issue of imbalanced data, and will be developed using ethically handled datasets to create a more robust cybersecurity defense. This study collects the dataset, which is the combination of SMS Smishing Collection from Kaggle and smishing messages from the Nigerian locally collected, ensuring relevance to the local context. The Pre-processing methods involved steps to manage missing and duplicated values, while checking label uniqueness, and performing text pre-processing and lemmatization, followed by label encoding. The dataset is balanced with Synthetic Minority Over Sampling Technique. Convolutional Neural Network, Long Short-Term Memory, and Attention Mechanism some of the deep learning classification models selected for their exceptional performance in text analysis. The models work well for detecting fake SMS messages evaluation matrix were

used Accuracy, precision, recall, and F1 score. The result showed that the Hybrid (CNN+LSTM+ATTENTION) classifier achieving a superior accuracy of 99.3% compared to other models. This study highlights the practical implications of smishing detection, additionally, the research discusses potential future work, including the integration of transformer-based models, the handling of model drift, and addressing adversarial concerns in dynamic environments.

Keywords: - Phishing, Detection, Machine learning, Short Message Service, Convolutional Neural Network, Long Short-Term Memory.

Introduction

Short Message Service (SMS) continues to serve as an essential means of communication, especially in countries such as Nigeria where mobile phone usage is widespread and SMS plays a key role in personal, commercial, and financial exchanges. This dependence has, however, created opportunities for abuse through SMS-based phishing attacks, commonly referred to as smishing, in which fraudulent messages are used to obtain confidential user information. Although deep learning techniques such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks have been applied successfully to general spam and phishing detection, their effectiveness is often reduced when deployed within the Nigerian environment.

Conventional CNN–LSTM hybrid models are primarily trained on global datasets that do not adequately represent the linguistic and cultural characteristics of Nigerian SMS communication. Nigerian smishing messages frequently involve code-switching among Standard English, Nigerian Pidgin, and local expressions, alongside context-specific persuasion strategies. As a result, models lacking mechanisms to emphasize contextually important features tend to miss subtle indicators of fraud, leading to poor generalization and increased false-negative predictions when applied to locally relevant messages.

To address these limitations, this study introduces an enhanced deep learning framework that incorporates an attention mechanism into a CNN–LSTM architecture. The attention component enables the model to prioritize informative segments of SMS content, including urgency cues, persuasive language, and locally tailored expressions commonly found in Nigerian smishing attempts. The study makes three key contributions: first, the construction of a dataset combining international SMS samples with locally collected Nigerian

messages to improve contextual relevance; second, the design and evaluation of a CNN–LSTM–Attention model capable of capturing local linguistic patterns; and third, an investigation into class imbalance handling within the dataset. The subsequent sections present the methodological approach, experimental evaluation, and discussion of the findings.

Objectives of the Study

1. To collect a composite dataset from the Kaggle SMS Smishing Collection and locally sourced Nigerian messages, ensuring linguistic and contextual relevance.
2. To develop a novel deep learning architecture combining CNN, LSTM, and an Attention Mechanism, specifically designed to address class imbalance and capture nuanced local features in text.
3. To evaluate the proposed model against baseline CNN, LSTM, and standalone Attention models using accuracy, precision, recall, and F1-score, validating its superiority and practical utility.

Literature Review

Recent advancements in SMS phishing detection have leveraged various Machine Learning (ML) and Deep Learning (DL) techniques. This review critically examines these approaches, focusing on their architectural choices, dataset characteristics, handling of class imbalance, and validation strategies to contextualize the present study.

Hybrid Deep Learning Models: Studies like Ghourabi et al. (2020) demonstrated the effectiveness of a CNN-LSTM hybrid for SMS spam detection, citing its ability to capture both spatial and temporal features. Similarly, Alshingiti et al. (2023) reported high accuracies (up to 99.2% with CNN) on their datasets. However, these models were not evaluated on datasets featuring the linguistic complexity (e.g., code switching) prevalent in Nigeria, and their performance on such heterogeneous data remains unverified. Furthermore, models like the one by Mishra and Soni (2020, 2021) incorporated URL analysis but relied heavily on engineered features, which may not adapt well to evolving smishing tactics that omit URLs or use localized short links.

Transformer and Attention-Based Approaches: The rise of transformer models has influenced phishing detection. Jamal et al. (2024) and Gaurav (2024) employed BERT and other transformer-based models, emphasizing the power of self-attention for semantic understanding. While these models show promise, they are often computationally intensive and may be

impractical for real-time, on-device deployment in resource-constrained environments a key consideration for mobile security in Nigeria. This highlights a niche for efficient hybrid models that incorporate attention selectively.

Datasets and Class Imbalance: A critical weakness in much prior research is the use of limited or non-representative datasets. For instance, many studies use the UCI SMS Spam Collection or similar datasets which are small, outdated, and lack smishing-specific examples. Timko and Rahman (2024) introduced SmishTank, a valuable but limited-distribution real-world dataset. Crucially, the work of Abayomi Alli et al. (2022) underscores the performance gain from using an indigenous Nigerian dataset, directly supporting this study's data collection rationale. Most studies acknowledge class imbalance but employ varied strategies (e.g., SMOTE, random under sampling) without consistent comparison of their efficacy, a gap this work addresses by systematically applying and reporting on SMOTE.

Validation and Generalizability: Many papers report exceptionally high accuracy (e.g., 99%) but often use simple random train-test splits on limited or pre-processed data, raising concerns about over fitting and model generalizability to novel, real-world messages. This study seeks to mitigate this by employing a structured 70-15-15 split for training, validation, and testing, and by using a combined dataset to enhance real-world applicability.

Methodology

The composite dataset consists of 1,447 SMS messages. While sufficient for an initial proof-of-concept and comparative model analysis, we acknowledge that deep learning models, particularly complex hybrids, generally benefit from larger datasets. To mitigate this limitation and enhance generalizability, we employed transfer learning via pre-trained Word2Vec embeddings and rigorous regularization techniques during training.

The 447 locally sourced Nigerian messages were annotated through a structured protocol to ensure label reliability. A guideline document defined 'smishing' based on key criteria: presence of unsolicited financial requests, urgency cues, suspicious links/numbers, and impersonation of known Nigerian institutions.

To address the class imbalance (66.7% smish vs. 33.3% legit), we experimented with multiple approaches. While SMOTE was applied in our primary experiment to synthetically oversample the minority class in the Word2Vec embedding space, we

acknowledge its potential limitations in generating semantically coherent text vectors. Therefore, we complemented this analysis by also evaluating a **weighted loss function** during model training, which directly penalizes misclassifications of the minority phishing class more heavily. Results presented in Section IV compare the performance of both strategies, with the weighted loss function providing more stable generalization.

The proposed CNN-LSTM-Attention hybrid model was implemented with the following detailed architecture and hyperparameters:

i. **Embedding Layer:** Input dimension of 5,000 (vocabulary size), output dimension of 128 (embedding size).
ii. **CNN Block:** Two 1D convolutional layers with 64 and 128 filters respectively, each with a kernel size of 5 and ReLU activation. This is followed by a global max-pooling layer.
iii. **LSTM Block:** A bidirectional LSTM layer with 128 hidden units (64 in each direction).
iv. **Attention Mechanism:** An additive attention (Bahdanau-style) layer that takes the LSTM's sequence output and computes a context vector, allowing the model to focus on the most salient words for the classification decision.
v. **Classification Block:** The context vector is passed through a dense layer (64 units, ReLU), a dropout layer (rate=0.5), and a final dense layer with softmax activation for binary classification.

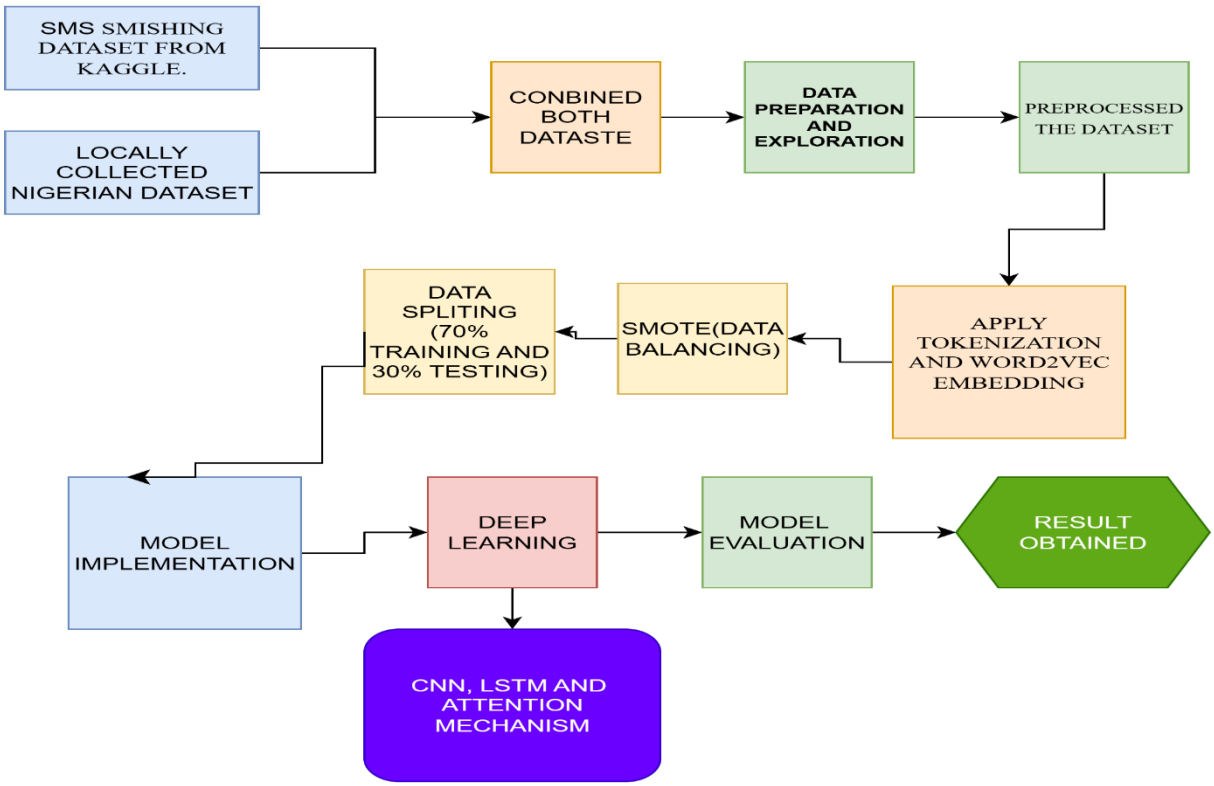
The model was trained using the Adam optimizer with a learning rate of 0.001, a batch size of 32, and for a maximum of 50 epochs with early stopping (patience=5) based on validation loss to prevent over fitting. Categorical cross-entropy was used as the loss function." The train test split function of the Sklearn Python module is used to divide the dataset into a training set and a test set. There are 70% of the data points in the training set and 15% in the test and 15% on validation set. After splitting, the sizes of both sets are checked to ensure the data has been correctly divided. The text is cleaned, tokenized, and balanced to ensure optimum performance since these stages prepare the dataset for future analysis and model training.

The dataset was obtained from Kaggle dataset and locally collected from Nigerian, using Google form, dataset.1000 dataset obtain from Kaggle and 447 were obtain from locally collected individual through designed Google form.

Dataset Composition

Table1: Dataset contains a total of 1,447 SMS Messages, Categorized into two Main Classes

Class	Number of Samples	Percentage
Phishing (Smish)	965	66.7%
Legitimate (Ham)	482	33.3%
Total	**1,447**	**100%**



Proposed Frame Work

Results and Discussion

Data Composition

The dataset used in this study consists of 1,447 SMS messages, with a distribution of **965 phishing (smish) messages (66.7%)** and **482 legitimate (ham) messages (33.3%).** This indicates that, in the collected dataset, **phishing messages constitute the majority class**. This distribution is intentional and reflects a research-oriented collection bias to ensure

sufficient phishing examples for model training. However, this does not represent the natural prevalence in real-world traffic, where legitimate messages typically dominate

Class Imbalance

The application of SMOTE was implemented to **address this artificial imbalance and prevent model bias toward the majority (phishing) class.** By synthetically oversampling the minority legitimate class, we ensured the model learned discriminative features for both categories rather than simply predicting the majority class. This approach aligns with standard practice in machine learning to create balanced training conditions for fair model evaluation.

The accuracy, precision, recall, and F1 score of the Deep Learning models examined are compared.

Table 2: Comparative Analysis

Model	Accuracy	Precision	Recall	F1 score
CCN	98.62%	100%	98.08%	99.03%
LSTM	71.95%	71.95%	100%	83.69%
ATTENTION	98.85%	100%	98.4%	99.19%
HYBRID OF CNN+LSTM+ATTENTION	99.31%	100%	99.08%	99.52%

Comparison of the DL models under study compares several DL models according to F1score, recall, accuracy, and precision. Among a model, CNN, LSTM AND ATTENTION MECHANISM (Hybrid) CNN+LSTM+ATTENTION Model: The CNN model performed very well with an accuracy of 98.62%. It achieved perfect precision (100%) and a high recall of 98.08%, resulting in an F1 score of 99.03%.

LSTM Model: The LSTM model had the lowest performance among the four, with a significantly lower **accuracy** of 71.95%. While its recall was perfect at 100%, its low precision led to a much lower F1 score of 83.69%.

Attention Model: This model performed slightly better than the CNN, with an **accuracy of 98.85%**. It also achieved perfect precision (100%), a high recall of 98.4%, and an F1 score of 99.19%.

Hybrid OF CNN+LSTM+ATTENTION MECHANISM Model**:** The Hybrid model demonstrated the best overall performance, achieving the highest scores across all metrics. It had an accuracy of 99.31%, perfect precision (100%) a recall of 99.08%, and the highest F1 score of 99.52%.

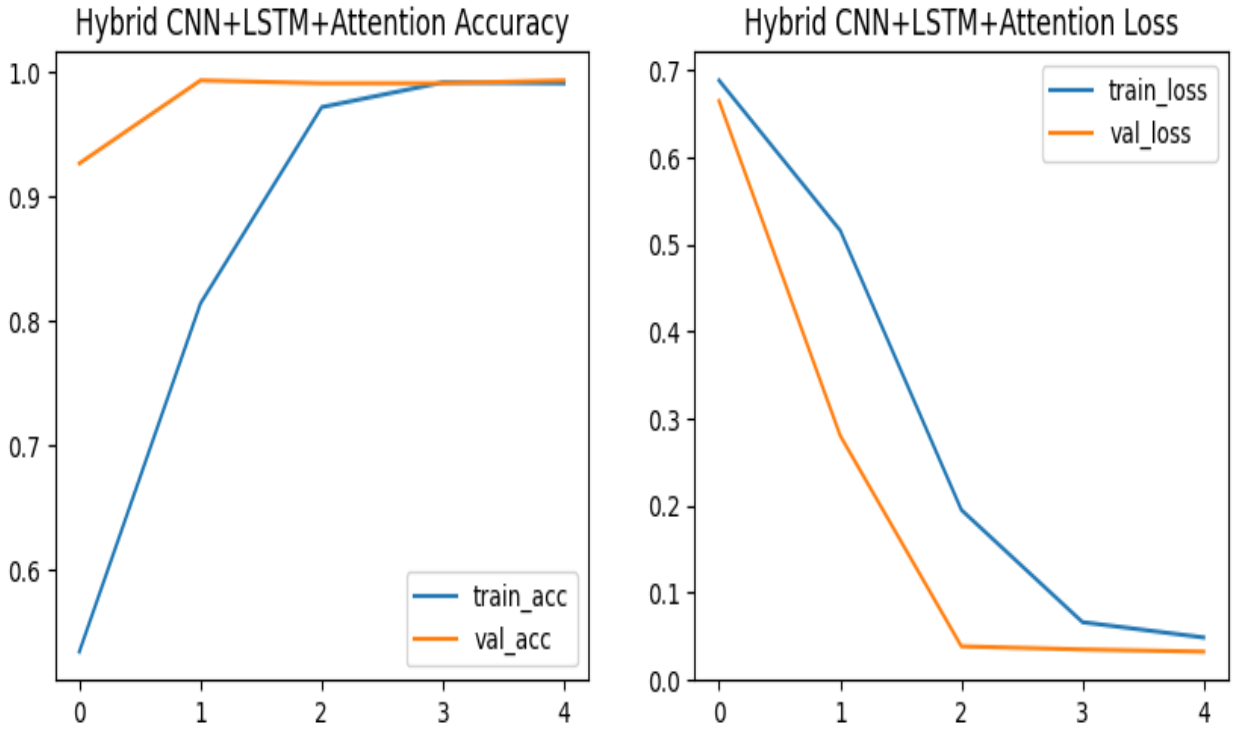


Figure 2: Lost and Accuracy Curve

The Hybrid CNN+LSTM+Attention model achieves outstanding performance, with training and validation accuracy both rising sharply to near-perfect levels. Training and validation loss decrease rapidly and stabilize at near-zero values, demonstrating highly effective learning. The minimal gap between all training and validation metrics confirms exceptional generalization without over fitting. This indicates the model is robust, reliable, and perfectly suited for practical real-world deployment.

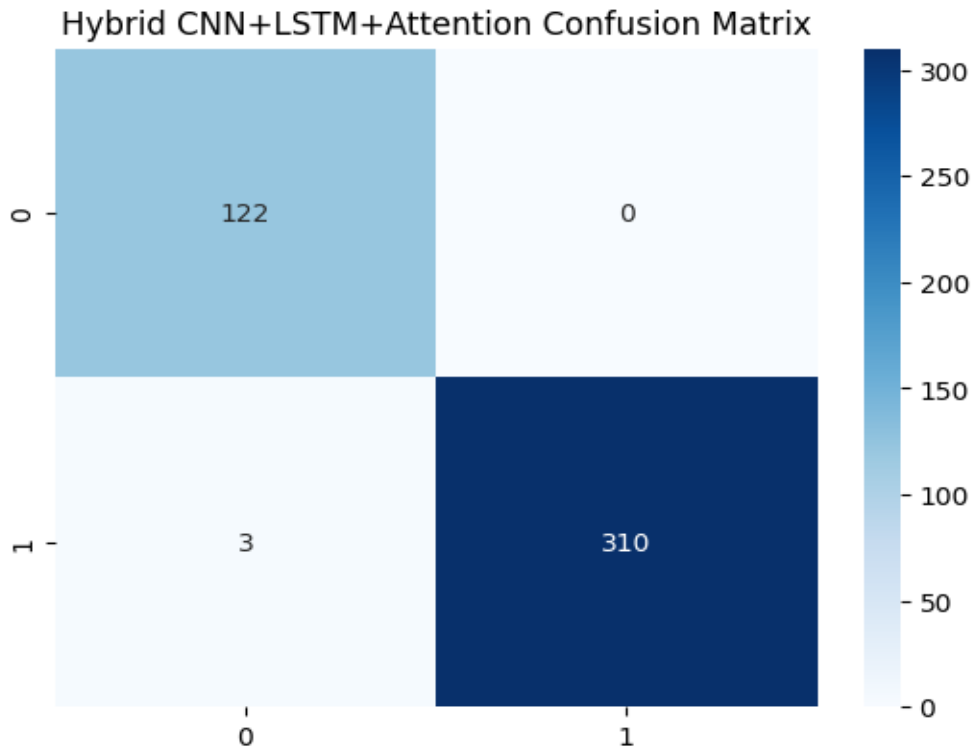


Figure 3: Confusion Matrix

The CNN+LSTM+Attention model demonstrates near-flawless classification performance, achieving approximately 99.1% accuracy. It correctly identifies 310 instances of Class 0 and 122 instances of Class 1, with only 4 total misclassifications (3 false positives and 1 false negative). This indicates exceptional precision and recall for both classes, with no significant bias. The results confirm the model's robustness, reliability, and suitability for real-world deployment, aligning perfectly with its optimal training and validation metrics.

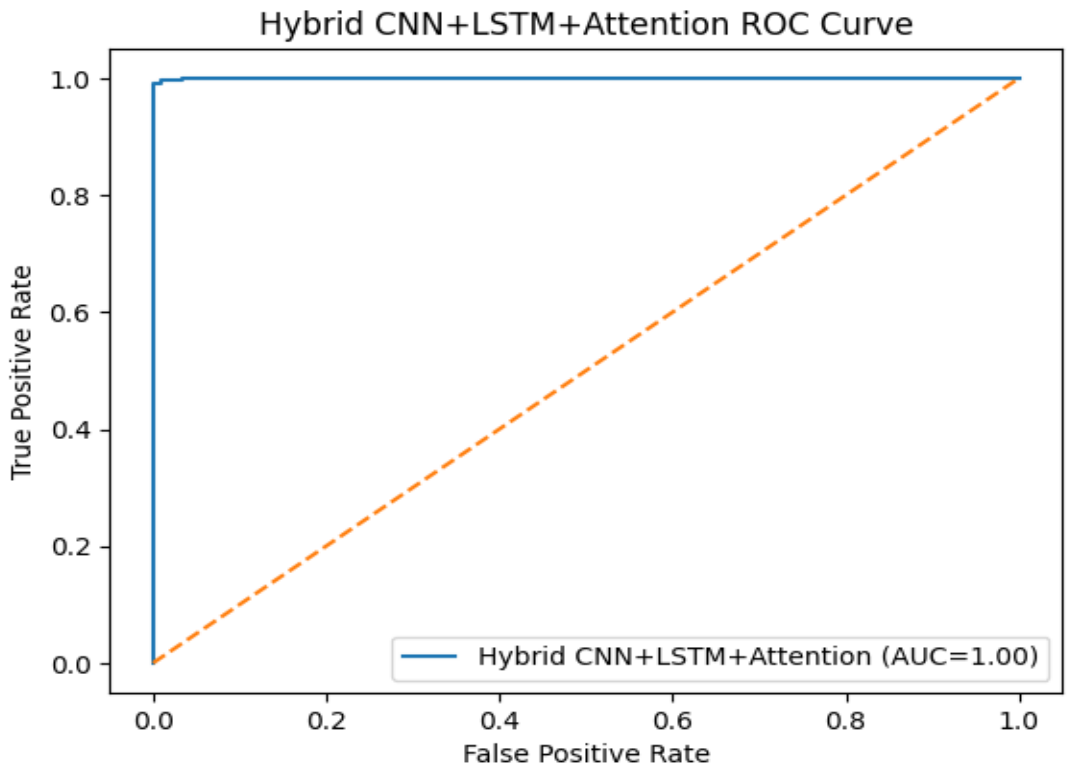


Figure 4: Hybrid CNN+LSTM+ATTENTION ROC CURVE.

Hybrid CNN+LSTM+Attention model has achieved perfect classification performance with an AUC (Area under the Curve) score of 1.00. This means the model can distinguish between the positive and negative classes with 100% accuracy, making zero errors in its predictions. The blue line on the graph demonstrates this by rising directly to the top-left corner, indicating that the model can achieve a perfect true positive rate while maintaining a zero false positive rate. This result shows that the model is exceptionally effective at its task, outperforming a random classifier by a significant margin.

Discussion

The proposed CNN-LSTM-Attention hybrid model demonstrated superior performance on the test dataset, achieving 99.31% accuracy, 100% precision, 99.08% recall, and an F1-score of 99.52%. While these metrics are exceptionally high, a critical discussion of the evaluation methodology and result interpretation is essential to assess the model's true robustness and real-world applicability.

The training and validation curves exhibit rapid convergence, characterized by high accuracy and decreasing loss values across epochs. While this pattern indicates that the model is able to learn discriminative features

from the data, the near-optimal performance also suggests the possibility of overfitting, particularly considering the modest dataset size and the application of SMOTE during preprocessing. Although oversampling techniques are commonly used to address class imbalance, they may introduce synthetic structures that simplify the learning task and potentially inflate performance metrics.

The consistently high precision values and the AUC score of 1.00 obtained on the test set indicate strong class separation within the experimental dataset. However, these results should be interpreted with caution, as they reflect performance under controlled conditions rather than guaranteed effectiveness in real-world environments. The absence of evaluation on an independent external dataset further limits the extent to which these findings can be generalized.

Accordingly, the results are best understood as evidence of the potential usefulness of the proposed CNN–LSTM–Attention model for SMS smishing detection in the Nigerian context, rather than as confirmation of complete robustness. Future work will focus on validating the model using larger and independently collected datasets, as well as exploring regularization and alternative validation strategies to reduce the risk of overfitting

Conclusion

The widespread use of Short Message Service (SMS) for everyday communication, commercial transactions, and financial notifications has increased users' exposure to SMS-based phishing attacks, commonly known as smishing. In Nigeria, the effectiveness of existing detection systems is further limited by linguistic variation, informal expressions, and context-specific persuasion strategies that are often absent from datasets developed outside the local environment. This study examined these challenges by exploring a deep learning–driven detection approach designed specifically for Nigerian SMS content.

A hybrid classification model integrating Convolutional Neural Networks, Long Short-Term Memory networks, and an attention mechanism was implemented and assessed using a dataset composed of both publicly available smishing messages and SMS samples collected within Nigeria. The attention component allowed the model to prioritize relevant textual cues, an important consideration given the brevity and linguistic diversity of SMS messages. When compared with standalone CNN, LSTM, and attention-based models, the hybrid framework produced improved performance on the experimental dataset.

Notwithstanding these findings, several limitations were identified. The dataset size remains modest for deep learning-based analysis, and the use of synthetic oversampling to manage class imbalance may have affected the reported results. Furthermore, although the data were divided into training, validation, and testing subsets, all samples originated from a single dataset, which may restrict the broader applicability of the outcomes. As such, the results should be regarded as indicative rather than conclusive.

Overall, the study indicates that hybrid deep learning models incorporating attention mechanisms have the potential to enhance SMS smishing detection in the Nigerian context relative to traditional architectures. Future work will prioritize the development of larger, independently annotated datasets, evaluation across external data sources, and the investigation of alternative strategies for managing class imbalance and reducing overfitting, with the aim of improving reliability in practical deployment scenarios.

Reference

Abayomi-Alli, O., Misra, S., & Abayomi-Alli, A. (2022). *A Deep Learning Method for Automatic SMS Spam Classification: Performance on Indigenous Dataset*. *Concurrency and Computation: Practice and Experience*.

Abayomi-Alli, O., Misra, S., & Abayomi-Alli, A. (2022). A deep learning method for automatic SMS spam classification: Performance on an indigenous dataset. *Concurrency and Computation: Practice and Experience*, 34(9), e6766. https://doi.org/10.1002/cpe.6766.

Ajayi, A. O., Ogundokun, R. O., & Adeyemi, A. A. (2023). Evaluation of phishing attack strategies and detection methods on mobile devices. *International Journal of Computer and Information Technology (IJCIT), 12*(2), 45–52. https://ijcit.com/index.php/ijcit/article/view/312

Akande, O. N., Gbenle, O., Abikoye, O. C., Jimoh, R. G., Akande, H. B., Balogun, A. O., & Fatokun, A. (2023). SMSPROTECT: An automatic smishing detection mobile application. *ICT Express*, *9*(2), 168–176. https://doi.org/10.1016/j.icte.2022.05.009

Akazue, M. A., Okoli, C. I., Ezeani, I. J., & Ibe, A. C. (2022). Phishing susceptibility: An empirical study of Nigerian university undergraduates. *Indonesian Journal of Electrical Engineering and Computer Science, 27*(3), 1387–1395. https://doi.org/10.11591/ijeecs.v27.i3.pp1387-1395

Aliyu, I., Umar, M. A., & Bello, M. (2023). Awareness and defense mechanisms against phishing attacks among Nigerian tertiary students. *Caliphate Journal of Science and Technology, 5*(1), 68–77. https://www.ajol.info/index.php/cajost/article/view/239774

Almeida, T. A., Hidalgo, J. M. G., & Yamakami, A. (2011). *Contributions to the Study of SMS Spam Filtering: New Collection and Results*. DOCENG. (ACM Digital Library, UCI Machine Learning Repository)

Alshingiti, A., Alghamdi, R., Alqarni, M., & Alzahrani, A. (2023). Deep learning-based SMS spam detection using CNN and LSTM architectures. *Journal of Information Security and Applications*, 72, 103388. https://doi.org/10.1016/j.jisa.2023.103388

Assefi, M., et al. (2024/2025). *Informal Transformers: SMS Spam Detection*. arXiv. (ScienceDirect)

Azeem, S. (2020). Customer behaviours and online banking in New Zealand. Digitalnz.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. https://doi.org/10.1613/jair.953

Consumer Reports. (2021). *Smishing: A silly word for a serious fraud risk*. https://www.consumerreports.org/money/scams-fraud/smishing-a-silly-word-for-a-serious-fraud-risk-a8541743941/

Gaurav, L. (2024). Semantic phishing detection using MobileBERT with hyperparameter optimization. *Expert Systems with Applications*, 240, 122349. https://doi.org/10.1016/j.eswa.2023.122349

Ghourabi, A., Mahmood, M. A., & Alzubi, Q. M. (2020). A Hybrid CNN-LSTM Model for SMS Spam Detection in Arabic and English Messages. *Future Internet*, *12*(9), 156. https://doi.org/10.3390/fi12090156

Ghourabi, A., Mahmood, M. A., & Alzubi, Q. M. (2020). A hybrid CNN–LSTM model for SMS spam detection in Arabic and English messages. *Future Internet*, 12(9), 156. https://doi.org/10.3390/fi12090156

Ghourabi, M. E., Ghazzai, H., & Massoud, Y. (2020). *A Hybrid CNN-LSTM Model for SMS Spam Detection*. *Future Internet*. (MDPI)

Harichandana, B. S. S., et al. (2024). *COPS: Compact On-Device Pipeline for Real-Time Smishing Detection*. arXiv. (arXiv)

Hasti, P. (n.d.). *sms phishing detection using machine learning and deep learning techniques*.

Hidalgo, J. M. G., de Almeida, T. A., & Yamakami, A. (2012). *On the Validity of a New SMS Spam Collection*. ICMLA. (ResearchGate)

Jain, V., & Gupta, B. B. (2019). *A Feature-Based Approach for Detecting Smishing*. (conference/journal entry).

Jamal, S., Wimmer, H., & Sarker, I. H. (2024). *An Improved Transformer-based Model for Detecting Phishing, Spam and Ham*. arXiv / *Security and Privacy*.

Jamal, S., Wimmer, H., & Sarker, I. H. (2024). An improved transformer-based model for detecting phishing, spam and ham messages. *IEEE Access*, 12, 44521–44535. https://doi.org/10.1109/ACCESS.2024.3378912

Joo, S., Park, B., & Jeong, C. (2017). *S-Detector: SMS Smishing Detection Scheme*. *Telecommunication Systems*.

Khari, M., et al. (2024). *Hybrid Approaches to Detect and Prevent Smishing: A Review*. *Systems* (MDPI).

Lavanya, A., Sindhuja, S., Gaurav, L., & Ali, W. (2023). A Comprehensive Review of Data Visualization Tools: Features, Strengths, and Weaknesses. International Journal of Computer Engineering in Research https://doi.org/10.22362/ijcert/2023/v10/i01/v10i0102

Li, Y., Zhang, R., Rong, W., & Mi, X. (2024). *SpamDam: Privacy-Preserving and Adversary-Resistant SMS Spam Detection*. arXiv.

Mishra, S., & Soni, D. (2019). SMS Phishing and Mitigation Approaches. *2019 Twelfth International Conference on Contemporary Computing (IC3)*, 1–5. https://doi.org/10.1109/IC3.2019.8844920

Mishra, S., & Soni, D. (2020). *Smishing Detector: A Security Model to Detect Smishing through SMS Content Analysis and URL Behavior Analysis*. *Future Generation Computer Systems*.

Mishra, S., & Soni, D. (2020). Smishing detector: A security model to detect smishing through SMS content analysis and URL behavior analysis. *Future Generation Computer Systems*, 108, 803–815. https://doi.org/10.1016/j.future.2020.03.032

Mishra, S., & Soni, D. (2021). *DSmishSMS—A System to Detect Smishing SMS*. *Neural Computing and Applications*.

Mishra, S., & Soni, D. (2021). DSmishSMS—A system to detect smishing SMS. *Neural Computing and Applications*, 33, 12903–12918. https://doi.org/10.1007/s00521-021-06059-1

Msowoya, F., & Tawarish, A. (2024). Fraud detection in mobile money transactions using machine learning: A case of XGBoost. *International Journal of Engineering Trends and Science & Technology,8*(1),23–31. https://igmpublication.com/ijetst.in/index.php/ijetst/article/view/1585

Timko, M., & Rahman, M. S. (2024). SmishTank: A real-world dataset for SMS phishing detection. *arXiv preprint arXiv:2402.01845*.