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

Explainable Sequential Deep Learning for Detecting Botnet Command-and-Control Communications in Transport Layer Security-Encrypted Network Traffic

[1]Darlington Nwanebu, [2]E. H. Gadzama, [3]Shamsu Garba
[1&2]Department of Cyber Security, Nigerian Defence Academy Kaduna
[3]Department of Civil Engineering, Nigerian Defence Academy Kaduna

ABSTRACT
The increasing use of Transport Layer Security (TLS) encryption has improved communication privacy but has also created challenges for detecting botnet Command-and-Control (C2) communications. This study presents an explainable sequential deep-learning framework combining Bidirectional Long Short-Term Memory (BiLSTM) and Transformer models for detecting botnet C2 communications in TLS-encrypted network traffic. Random Forest was used as a baseline. The study considered CTU-13, CIC-Encrypted-Traffic-2022, and a synthetically generated TLS-BotFlow2025 dataset. Five-fold cross-validation produced F1-scores of 0.74, 0.84, and 0.87 for Random Forest, BiLSTM, and Transformer, respectively, with ROC-AUC values of 0.81, 0.88, and 0.92. The hybrid model achieved a best validation F1-score of 0.9640 and ROC-AUC of 0.9870 at epoch 47. These are validation results, not independent test-set results. External evaluation on unseen CTU-13 traffic produced F1-scores of 0.67, 0.75, and 0.80 for Random Forest, BiLSTM, and Transformer, respectively. SHAP and attention analysis were used to support interpretation of model decisions.

ARTICLE INFO
Article History
Received: January, 2026
Received in revised form: May, 2026
Accepted: June, 2026
Published online: September, 2026

KEYWORDS
Botnet Detection; Command and Control; TLS encryption; BiLSTM; Transformer; Explainable Artificial Intelligence; SHAP; Encrypted Traffic; Network Security; Deep Learning

INTRODUCTION

Encrypted communication limits the usefulness of payload-based intrusion detection because the plaintext content of Transport Layer Security (TLS) traffic is not directly visible. Nevertheless, encrypted network traffic can retain observable behavioural characteristics, including packet-size distributions, inter-packet intervals, session duration, traffic directionality, connection frequency, endpoint interactions, and communication regularity. These characteristics provide opportunities for detecting malicious activity without decrypting packet payloads.

Recent studies illustrate complementary approaches to payload-independent network security analysis. Yin et al. (2022) demonstrated the application of deep learning to botnet detection using a dilated convolution-based model, while Kumar et al. (2023) proposed machine-learning models for detecting phishing activity from TLS 1.2 and TLS 1.3 traffic without decrypting the underlying communication. These studies demonstrate the increasing importance of analysing observable characteristics of network traffic when conventional plaintext inspection is unavailable.

Encrypted-traffic classification has also benefited from deep-learning architectures capable of learning both local and long-range relationships. Wang et al. (2023) proposed a multi-task encrypted-traffic classification model combining Transformer and one-dimensional convolutional neural network architectures. The Transformer component captures long-range dependencies within network traffic sequences, while the 1D-CNN component extracts local characteristics. Such approaches demonstrate that useful security and traffic information can be learned from encrypted communication without relying on access to plaintext payloads.

Corresponding author: Darlington Nwanebu
darlingtonnwanebu@gmail.com
Department of Cyber Security, Nigerian Defence Academy Kaduna.

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

The increasing sophistication of encrypted malicious traffic has therefore created a need for detection approaches that move beyond static signatures towards behavioural, temporal, and sequence-aware analysis. This study investigates an explainable sequential deep-learning framework combining Bidirectional Long Short-Term Memory (BiLSTM) and Transformer architectures for detecting botnet Command-and-Control (C2) communication in TLS-encrypted network traffic. Random Forest is used as a conventional baseline, while SHAP and attention analysis are incorporated to support interpretation of model decisions.

RELATED WORK

Botnet and intrusion detection has progressively moved from signature-based approaches toward machine-learning, deep-learning, graph-based, and explainable methods. This shift is important for encrypted traffic because payload inspection becomes less effective when communication is protected by TLS. Recent approaches therefore focus more on statistical, temporal, behavioural, and relational characteristics that remain observable. BiLSTM models are useful for learning bidirectional temporal dependencies, Transformers capture long-range relationships through self-attention, while SHAP and related explainability techniques support interpretation of model decisions.

Wang et al. (2023) proposed MTC, a multi-task model combining Transformer and 1D-CNN architectures for encrypted-traffic classification. The Transformer captures long-range dependencies, while the 1D-CNN extracts local traffic characteristics. Using the ISCX VPN-nonVPN dataset, the model achieved average F1-scores of 98.25% for application identification and 97.94% for traffic characterization. The study demonstrates the effectiveness of combining local and global traffic representations, although its main focus was encrypted-traffic classification rather than botnet C2 detection.

Hossain and Islam (2023) developed a botnet-detection framework that combined hybrid feature selection with ensemble machine learning. Their approach used categorical analysis, mutual information, and principal component analysis, with Extra Trees producing the strongest reported performance at approximately 99.99% accuracy. The study shows the value of selecting informative traffic features for robust botnet detection. However, it did not explicitly model sequential or temporal relationships within encrypted traffic.

Fu et al. (2023) proposed HyperVision, an unsupervised framework for detecting unknown encrypted malicious traffic through flow-interaction graph analysis. The framework was evaluated across 92 datasets containing 48 attacks and achieved at least 0.92 AUC and 0.86 F1. Its findings demonstrate that relationships among network flows can reveal malicious behaviour even when payloads are encrypted. However, the approach focuses primarily on graph interaction patterns rather than bidirectional temporal sequence modelling.

Ferrag et al. (2024) introduced SecurityBERT, a lightweight BERT-based model for cyber-threat detection in IoT and IIoT environments. Using the Edge-IIoTset dataset, the model achieved 98.2% accuracy, with a size of approximately 16.7 MB and CPU inference below 0.15 seconds. The study demonstrates that Transformer-based detection can be adapted to resource-constrained environments. Nevertheless, its focus was broader cyber-threat classification rather than TLS-encrypted botnet C2 communication.

Yang et al. (2024) developed MTSecurity, which combines a Transformer with a Graph Neural Network for privacy-preserving malicious-traffic classification. The framework achieved 99.46% accuracy and 99.40% F1 on MCFP, and 99.48% accuracy with 99.34% F1 on USTC-TFC. The study highlights the value of integrating traffic representation learning with relational graph information. However, graph construction and processing introduce additional preprocessing and computational complexity.

Hashmi et al. (2024) proposed a hybrid intrusion-detection model combining convolutional processing, feature-weighted attention, BiLSTM, and Random Forest. The model achieved above 99% accuracy on both NSL-KDD and UNSW-NB15 datasets. The study supports the

Corresponding author: Darlington Nwanebu
darlingtonnwanebu@gmail.com
Department of Cyber Security, Nigerian Defence Academy Kaduna.

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

effectiveness of combining bidirectional sequential learning, attention mechanisms, and conventional classifiers. However, the datasets used were general intrusion-detection benchmarks rather than TLS-specific botnet C2 datasets.

Ahmed et al. (2024) investigated explainable intrusion detection using hybrid bagging and boosting techniques with SHAP-based feature selection. The proposed framework achieved a peak accuracy of 98.47% and an F1-score of 96.19%. The study demonstrates the value of SHAP for improving transparency and identifying influential features in intrusion-detection models. However, the framework relied mainly on ensemble machine learning rather than sequential deep-learning architectures.

Overall, these studies demonstrate strong progress in encrypted-traffic classification, botnet detection, graph-based modelling, sequence learning, lightweight Transformers, hybrid architectures, and explainable AI. However, limited attention has been given to frameworks that jointly combine BiLSTM and Transformer architectures specifically for TLS-encrypted botnet C2 detection while also providing interpretable outputs through SHAP and attention mechanisms. This gap provides the motivation for the present study.

MATERIALS AND METHODS

The study considered CTU-13, CIC-Encrypted-Traffic-2022, and TLS-BotFlow2025, with TLS-BotFlow2025 representing a synthetic dataset generated specifically for this study to model contemporary TLS-encrypted botnet C2 behaviour. The datasets comprised approximately 2.8 million records in total, with TLS-BotFlow2025 accounting for approximately 1.2 million records. Features were grouped into temporal, statistical, and behavioural categories. Temporal features included inter-packet intervals and session duration. Statistical features included packet counts, byte counts, and packet-size distributions. Behavioural features included endpoint interaction sequences and beaconing regularity. The dataset split was specified as 70% training, 15% validation, and 15% testing with stratification.

Model Development

Random Forest was used as a conventional baseline. BiLSTM was used to capture temporal dependencies in encrypted traffic sequences. The Transformer was used to capture longer-range relationships through self-attention. The hybrid architecture combines BiLSTM temporal encoding with Transformer attention before classification.

Evaluation Metrics

Precision, recall, F1-score, and ROC-AUC were prioritised because of class-imbalance concerns. Accuracy was not treated as the principal metric. Precision measures the proportion of predicted malicious instances that are actually malicious. Recall measures the proportion of actual malicious instances detected. F1-score combines precision and recall, while ROC-AUC measures discrimination across classification thresholds.

Comparative Five-Fold Cross-Validation

The reported five-fold cross-validation results are shown below.

Model	Precision	Recall	F1-score	ROC-AUC
Random Forest	0.78	0.71	0.74	0.81
BiLSTM	0.82	0.87	0.84	0.88
Transformer	0.89	0.85	0.87	0.92

Random Forest achieved precision 0.78, recall 0.71, F1-score 0.74, and ROC-AUC 0.81. BiLSTM achieved precision 0.82, recall 0.87, F1-score 0.84, and ROC-AUC 0.88. Transformer achieved precision 0.89, recall 0.85, F1-score 0.87, and ROC-AUC 0.92. These are standalone model cross-validation results and should not be described as hybrid-model results.

Corresponding author: Darlington Nwanebu
darlingtonnwanebu@gmail.com
Department of Cyber Security, Nigerian Defence Academy Kaduna.

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

Hybrid Model Validation

The best hybrid checkpoint was identified as models/hybrid_BEST_epoch_047.pth. The validation results at epoch 47 were:

Metric	Hybrid BiLSTM–Transformer
Validation Accuracy	96.50%
Validation Precision	0.9667
Validation Recall	0.9613
Validation F1-score	0.9640
Validation ROC-AUC	0.9870
Best Validation Loss	0.0923

The best hybrid checkpoint was identified as models / hybrid_BEST_epoch_047.pth. At epoch 47, the hybrid model achieved validation accuracy of 96.50%, validation precision of 0.9667, validation recall of 0.9613, validation F1-score of 0.9640, validation ROC-AUC of 0.9870, and best validation loss of 0.0923. The thesis subsequently indicates that the process was proceeding to test-set evaluation. Therefore, these values must be reported as validation results rather than independent test results.

Ablation Analysis

The feature-group ablation results are presented below.

Model	Feature Group	Precision	Recall	F1	ROC-AUC
BiLSTM	Temporal	0.78	0.82	0.80	0.84
BiLSTM	Statistical	0.74	0.75	0.74	0.79
BiLSTM	Behavioural	0.80	0.81	0.80	0.85
BiLSTM	All features	0.82	0.87	0.84	0.88
Transformer	Temporal	0.84	0.81	0.82	0.86
Transformer	Statistical	0.79	0.77	0.78	0.82
Transformer	Behavioural	0.86	0.83	0.84	0.88
Transformer	All features	0.89	0.85	0.87	0.92
Random Forest	All features	0.78	0.71	0.74	0.81

The ablation results indicate that combining temporal, statistical, and behavioural features improved the reported performance of both BiLSTM and Transformer models. BiLSTM achieved F1 = 0.80 with temporal features and F1 = 0.84 with all features. Transformer achieved F1 = 0.82 with temporal features and F1 = 0.87 with all features. Behavioural features also produced relatively strong individual results.

Cross-Dataset Evaluation

External evaluation was performed using unseen CTU-13 traffic.

Model	Precision	Recall	F1-score	ROC-AUC
Random Forest	0.70	0.65	0.67	0.73
BiLSTM	0.77	0.74	0.75	0.79
Transformer	0.82	0.78	0.80	0.84

On unseen CTU-13 traffic, Random Forest achieved precision 0.70, recall 0.65, F1-score 0.67, and ROC-AUC 0.73. BiLSTM achieved precision 0.77, recall 0.74, F1-score 0.75, and ROC-AUC 0.79. Transformer achieved precision 0.82, recall 0.78, F1-score 0.80, and ROC-AUC 0.84. The reduction compared with development-stage validation demonstrates the importance of evaluating models across different traffic distributions.

Corresponding author: Darlington Nwanebu
darlingtonnwanebu@gmail.com
Department of Cyber Security, Nigerian Defence Academy Kaduna.

Explainability Analysis

SHAP was used for feature-level interpretation of model predictions. Important feature groups included inter-packet intervals, session duration, packet and byte distributions, endpoint interactions, beaconing regularity, and communication sequences. Attention analysis was used to examine which parts of a traffic sequence received greater attention in the Transformer.

DISCUSSION OF FINDINGS

The sequential deep-learning models produced higher reported cross-validation performance than the Random Forest baseline. The Transformer obtained F1 = 0.87 and ROC-AUC = 0.92, while BiLSTM obtained F1 = 0.84 and ROC-AUC = 0.88. The hybrid model's F1 = 0.9640 and ROC-AUC = 0.9870 are best validation results. External CTU-13 evaluation produced lower performance, showing that changes in traffic distributions can affect generalisation.

Reconciliation of Results

The 0.87 F1-score and 0.92 ROC-AUC values refer to the standalone Transformer in the five-fold cross-validation table. The 0.9640 F1-score and 0.9870 ROC-AUC values refer to the best documented hybrid validation checkpoint. The external CTU-13 results are separate evaluation results. The thesis does not provide final independent test-set metrics for the hybrid model in the supplied training logs.

Implications for Cybersecurity

The framework supports payload-independent encrypted-traffic detection, sequential behavioural analysis, and analyst-oriented interpretation. Observable timing patterns, endpoint interactions, and traffic distributions can provide useful information even when packet contents are encrypted.

Limitations

The final independent test-set metrics for the hybrid model are not documented in the supplied thesis. Dataset distribution differences affect external performance. Additional testing using contemporary real-world encrypted traffic would strengthen evidence of operational generalisation. SHAP and attention provide interpretation but should not automatically be treated as causal explanations.

Contribution to Knowledge

The study contributes an explainable sequential-learning framework for encrypted botnet C2 detection; combines BiLSTM and Transformer architectures; integrates temporal, statistical, and behavioural characteristics; applies SHAP and attention analysis; evaluates external transferability; and separates validation evidence from external evaluation.

CONCLUSION

This study investigated botnet C2 detection in TLS-encrypted traffic using Random Forest, BiLSTM, Transformer, and a hybrid BiLSTM–Transformer framework. Five-fold cross-validation produced F1-scores of 0.74, 0.84, and 0.87 for Random Forest, BiLSTM, and Transformer, respectively. The hybrid model achieved a best validation F1-score of 0.9640 and ROC-AUC of 0.9870. External evaluation on unseen CTU-13 traffic produced F1 = 0.80 and ROC-AUC = 0.84 for the Transformer. The results indicate that observable temporal and behavioural characteristics can support encrypted C2 detection, while external validation remains essential.

Data Availability Statement

The study uses CTU-13, CIC-Encrypted-Traffic-2022, and a synthetic TLS-BotFlow2025 dataset generated specifically for this research. Access to the synthetic dataset may be provided subject to the study's data-sharing conditions, while the public datasets remain subject to their respective licences and repository access conditions.

Ethics Statement

This study involves analysis of network-traffic datasets and does not involve direct experimentation on human participants. Dataset-

Corresponding author: Darlington Nwanebu
darlingtonnwanebu@gmail.com
Department of Cyber Security, Nigerian Defence Academy Kaduna.

specific ethical and usage requirements should be observed.

Author Contributions

Darlington Nwanebu: Conceptualisation, methodology, data analysis, model development, investigation, interpretation of results, writing - original draft, and writing - review and editing.

Conflict of Interest

The author declares that there is no conflict of interest associated with this study.

REFERENCES

Ahmed, U., Jiangbin, Z., Almogren, A., Sadiq, M., Rehman, A. U., Sadiq, M. T., & Choi, J. (2024). Hybrid bagging and boosting with SHAP based feature selection for enhanced predictive modeling in intrusion detection systems. *Scientific Reports, 14*, Article 30532. https://doi.org/10.1038/s41598-024-81151-1

Ferrag, M. A., Ndhlovu, M., Tihanyi, N., Cordeiro, L. C., Debbah, M., Lestable, T., & Thandi, N. S. (2024). Revolutionizing cyber threat detection with large language models: A privacy-preserving BERT-based lightweight model for IoT/IIoT devices. *IEEE Access, 12*, 23733–23750. https://doi.org/10.1109/ACCESS.2024.3363469

Fu, C., Li, Q., & Xu, K. (2023). Detecting unknown encrypted malicious traffic in real time via flow interaction graph analysis. In *30th Annual Network and Distributed System Security Symposium (NDSS 2023)*. Internet Society. https://doi.org/10.14722/ndss.2023.23080

Hashmi, A., Barukab, O. M., & Hamza Osman, A. (2024). A hybrid feature weighted attention based deep learning approach for an intrusion detection system using the random forest algorithm. *PLOS ONE, 19*(5), e0302294. https://doi.org/10.1371/journal.pone.0302294

Hossain, M. A., & Islam, M. S. (2023). A novel hybrid feature selection and ensemble-based machine learning approach for botnet detection. *Scientific Reports, 13*, Article 21207. https://doi.org/10.1038/s41598-023-48230-1

Kumar, M., Kondaiah, C., Pais, A. R., & Rao, R. S. (2023). Machine learning models for phishing detection from TLS traffic. *Cluster Computing, 26*(5), 3263–3277. https://doi.org/10.1007/s10586-023-04042-6

Wang, K., Gao, J., & Lei, X. (2023). MTC: A multi-task model for encrypted network traffic classification based on Transformer and 1D-CNN. *Intelligent Automation & Soft Computing, 37*(1), 619–638. https://doi.org/10.32604/iasc.2023.036701

Yang, J., Jiang, X., Lei, Y., Liang, W., Ma, Z., & Li, S. (2024). MTSecurity: Privacy-preserving malicious traffic classification using graph neural network and Transformer. *IEEE Transactions on Network and Service Management, 21*(3), 3583–3597. https://doi.org/10.1109/TNSM.2024.3383851

Yin, Z., Qin, R., Ye, C., Li, Y., He, Y., Shu, Y., & Jiang, R. (2022). Dilated convolution-based botnet detection model. In C. Zhao & H. Imane (Eds.), *Third International Conference on Computer Communication and Network Security (CCNS 2022)* (Vol. 12453, Article 124531D). SPIE. https://doi.org/10.1117/12.2659107.

Corresponding author: Darlington Nwanebu
darlingtonnwanebu@gmail.com
Department of Cyber Security, Nigerian Defence Academy Kaduna.