FUDMA Journal of Engineering and Technology
ISSN: 3092-9385
https://fjet.fudutsinma.edu.ng/index.php/fjet/
Volume 2, Issue 2, 2026, 663-670

A Hybrid BioBERT-BiLSTM Model for Sentiment Analysis of Malaria Drug Reviews in Pharmacovigilance

Adedeji O. ADEBARE[1*], Olayinka O. OLUSANYA[2], Ademola A. OMILABU[3], Iyinoluwa T. IDOWU[4], Jímọ̀h À. HASSAN[5], Peter A. IDOWU[6]

[1*,3]Department of Computer and Information Sciences, Tai Solarin Federal University of Education, Ijagun, Nigeria
[2]Department of Computer Science, Tai Solarin Federal University of Education, Ijagun, Nigeria
[4]Department of Nursing Science, Elizade University, Ilara-Mokin, Ondo State, Nigeria
[5]Department of Intelligent Systems Engineering, Obafemi Awolowo University, Ife, Nigeria
[6]Department of Computer Science and Cybersecurity, Obafemi Awolowo University, Ife, Nigeria

[1*]adebareao@tasued.edu.ng, [2]olusanya_oo@tasued.edu.ng, [3]omilabuaa@tasued.edu.ng, [4]idowuiyinoluwa056@gmail.com, [5]hassanja@oauife.edu.ng, [6]paidowu@oauife.edu.ng

Abstract

Malaria remains a significant public health challenge in Nigeria, with widespread use of antimalarial drugs leading to frequent adverse drug reactions (ADRs). Current pharmacovigilance systems, such as the Pharmacovigilance Rapid Alert System for Consumer Reporting (PRASCOR), rely on SMS-based reporting and lack comprehensive data collection and analysis capabilities. This study develops a hybrid sentiment analysis model combining BioBERT (Bidirectional Encoder Representations from Transformers for Biomedical Text) and Bidirectional Long Short-Term Memory (BiLSTM) with an attention mechanism to classify malaria drug reviews for enhanced pharmacovigilance. A dataset of 3,194 malaria drug reviews was collected from patients in South-West Nigeria and preprocessed using natural language processing techniques. The proposed BioBERT-BiLSTM model leverages domain-specific contextual embeddings from BioBERT and captures sequential dependencies through BiLSTM, while the attention mechanism focuses on the most informative words for sentiment classification. The model was implemented in Python and evaluated using an 80-20 train-test split, achieving an accuracy of 60.90%, with balanced precision (0.54), recall (0.54), and F1-score (0.54). Comparative analysis against BiLSTM, Random Forest, Support Vector Machine, and Logistic Regression demonstrated the superior performance of the hybrid model across all metrics. The findings show that integrating domain-specific transfer learning with sequential deep learning architectures provides a scalable, accessible, and effective tool for automated pharmacovigilance, improving drug safety monitoring in resource-limited settings.

Keywords: *Sentiment Analysis, BioBERT, BiLSTM, Pharmacovigilance, Malaria Drugs, Hybrid Model.*

1.0 Introduction

Malaria represents a major health and socioeconomic concern in Nigeria, which contributes approximately 27% of malaria cases and fatalities worldwide [1]. Its effects are particularly pronounced among pregnant women and children below five years of age [2]. The disease also generates substantial economic pressure through direct healthcare expenditure and losses in productivity [3]. Antimalarial medicines account for a large proportion of reported Adverse Drug Reactions (ADRs), partly because these medicines are widely consumed and readily obtainable without prescription [4]. The Nigerian National Pharmacovigilance Policy and Implementation Framework (NNPP, 2020) records an ADR incidence of about 16.2%, with 6.5% of reported cases associated with hospitalisation, permanent disability, or death [5].

Within Nigeria's post-marketing surveillance (PMS) structure, the National Agency for Food and Drug Administration and Control (NAFDAC) monitors medicine safety through the Pharmacovigilance Rapid Alert System for Consumer Reporting (PRASCOR) [4]. The SMS-centred nature of PRASCOR constrains the amount of information that can be submitted and limits the depth of subsequent analysis [4]. A computational approach capable of processing patient-generated reviews at scale could therefore complement existing reporting arrangements and provide an additional source of information for malaria-drug safety monitoring.

Sentiment analysis provides one such computational route because it enables opinions expressed in textual material to be grouped into positive, negative, or neutral categories [6]. Early pharmaceutical sentiment studies commonly used rule-based procedures, but such methods can be sensitive to manually defined rules and may perform poorly when language becomes less predictable [7]. Classical machine-learning techniques subsequently became common [8–10]. Although optimisation studies have reported accuracies of up to 94% with Support Vector Machine (SVM) and Random Forest models [11, 12], these methods do not readily capture aspect-specific conditions or the specialised meanings found in biomedical narratives. Deep-learning research later examined LSTM, BiLSTM, and GRU models using more than 215,000 drug reviews, with a semi-supervised GAN reporting 87.5% accuracy [13]. Other investigations combined GloVe representations with BiLSTM [6] or CNN

Received: 01/09/2026; Revised: 12/09/2026; Accepted: 14/09/2026; Published: 25/09/2026

architectures [14]. Such developments improved contextual modelling, but the use of general-purpose word embeddings remained a limitation for biomedical language.

Hybrid approaches have also been investigated. One example combined domain-oriented information with transfer learning and reported 88.96% accuracy using XGBoost, although aspect-based drug conditions were not explicitly modelled [15]. Comparative evaluations involving BERT, RoBERTa, XLNet, and BioBERT identified BioBERT as the strongest model in terms of F1-score for drug-rating classification [16]. Another study combined an ontology with XLNet for ADR classification [17]. Taken together, the literature points to three unresolved issues relevant to this study. First, malaria-specific treatment aspects have received limited attention. Second, many existing datasets originate outside Nigeria, which reduces their direct contextual relevance to Nigerian patients. Third, linguistic patterns characteristic of Nigerian users remain insufficiently represented in existing sentiment-analysis resources.

The present study responds to these limitations by developing a BioBERT-BiLSTM model with attention for malaria drug reviews originating from Nigeria. The architecture uses BioBERT to obtain biomedical contextual information, BiLSTM to represent relationships across the sequence of a patient's narrative, and attention to identify tokens that contribute most to the sentiment decision. The resulting pipeline is intended as a pharmacovigilance-oriented classification aid for analysing patient-generated drug experiences.

2.0 Methods

2.1 Data Identification and Collection

The review corpus was developed with contributions from three physicians and two pharmacists and was cross-checked against more than 3,000 patient records to establish the relevant attributes. The final dataset comprises 3,194 malaria drug review records described by 10 features: reviewer age, gender, condition, drug name, ease of use, effectiveness, review text, satisfaction, side effects, and useful count. The observations were maintained in Comma-Separated Values (CSV) format and hosted on Google Drive for access within the modelling environment.

2.2 Data Preprocessing and Exploratory Data Analysis

Data preparation was undertaken in Google Colaboratory. Pandas was used for data organisation, NumPy for numerical operations, and Matplotlib/Seaborn for visualisation. Review text was processed with the Natural Language Toolkit (NLTK), including tokenisation, stop-word removal, and Porter stemming. BioBERT, a BERT-based model trained with biomedical material from PubMed and PMC [18], was used for the representation-learning and fine-tuning stages.

The exploratory analysis considered several characteristics of the dataset. These included descriptive statistics, word-frequency patterns, n-gram distributions, review-length characteristics, sentiment in relation to effectiveness ratings and age groups, word-cloud representations, and relationships among variables. This combination of numerical and textual exploration provided a basis for understanding the structure of the reviews before model training.

2.3 Model Formulation: BioBERT-BiLSTM with Attention Mechanism

The proposed classifier links biomedical language representation, bidirectional sequence learning, and attention-based weighting within a single pipeline. BioBERT first converts the input review into contextual token representations. The resulting sequence is processed by a BiLSTM so that information from both directions of the text contributes to each token representation. An attention layer then assigns different weights to these representations, allowing the model to concentrate on information that is more useful for sentiment prediction. The final context representation is supplied to a dense softmax layer to obtain the predicted sentiment probabilities.

(a) BioBERT embedding.

Let $X = \{x_1, x_2, \ldots, x_n\}$ denote the token sequence of an input review. Each token x_i is mapped to a contextual embedding by BioBERT:

$$H^0 = BioBERT(X) = \{h_1^0, h_2^0, \ldots, h_n^0\} \quad (1)$$

Here, h_i^0 denotes the contextual representation generated by BioBERT for token i.

(b) BiLSTM layer.

The sequence of BioBERT representations is supplied to two LSTM pathways, one processing the sequence from the beginning and the other from the end:

$$H\rightarrow = LSTM\rightarrow(H^0), \quad H\leftarrow = LSTM\leftarrow(H^0) \quad (2)$$

For each position, the two directional hidden states are joined to form a combined representation:

Received: 01/09/2026; Revised: 12/09/2026; Accepted: 14/09/2026; Published: 25/09/2026

$$H' = \{[h_1\rightarrow, h_1\leftarrow], [h_2\rightarrow, h_2\leftarrow], \ldots, [h_n\rightarrow, h_n\leftarrow]\} \quad (3)$$

(c) Attention mechanism.

The attention component evaluates the BiLSTM representations and assigns larger weights to tokens that are more informative for the classification task:

$$u_i = tanh(W_h h_i + b_h) \quad (4)$$

$$a_i = exp(u_i^T u\_w) / \sum_t exp(u_t^T u\_w) \quad (5)$$

In these expressions, W_h and b_h are trainable parameters, whereas u_w represents the word-level context vector. The review-level representation C is obtained by combining the BiLSTM outputs according to their attention weights:

$$C = \sum_i a_i h_i \quad (6)$$

(d) Dense layer and output.

The resulting context representation is then passed to a dense softmax output layer, which converts the learned representation into sentiment probabilities:

$$\hat{y} = softmax(W\_dC + b\_d) \quad (7)$$

In Equation (7), W_d and b_d are the parameters of the dense output layer, while ŷ denotes the predicted probability distribution for the positive and negative sentiment categories.

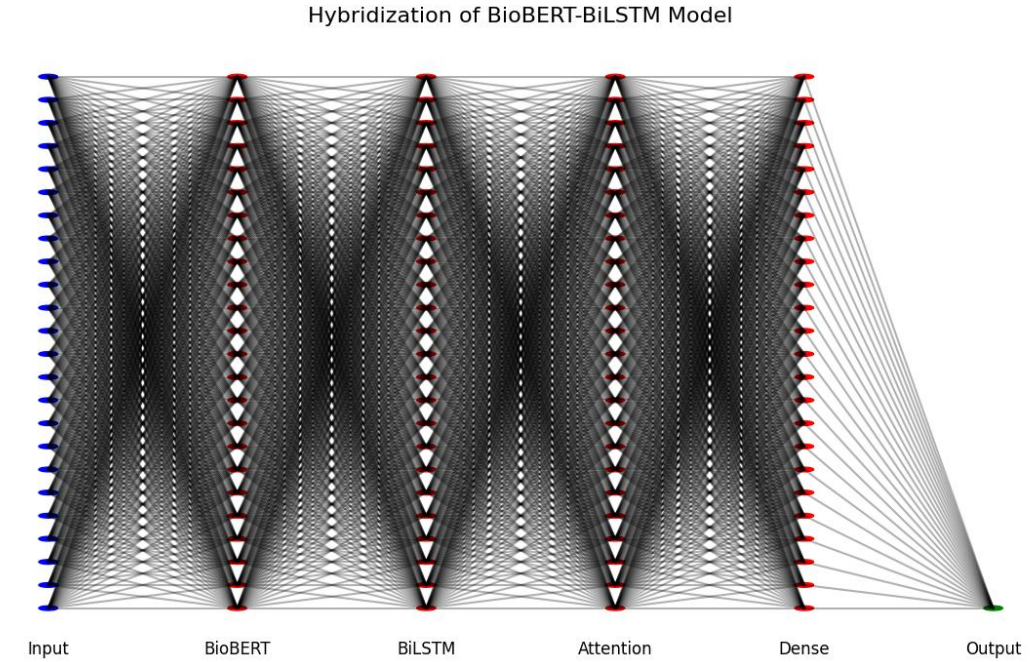

Figure 1: Architecture of the Hybrid BioBERT-BiLSTM Model for Malaria Drug Review Classification
Source: Researchers' Computation (2024)

Figure 1 provides the structural view of the proposed classifier. Five processing stages can be distinguished in the diagram: the input stage receives preprocessed and tokenised reviews; BioBERT transforms those tokens into biomedical contextual representations; the BiLSTM processes the representations in both directions; the attention layer assigns relative importance to the resulting token features; and the dense softmax layer produces the final positive or negative prediction. The key implication of the figure is that the model does not depend on a single representation technique. Instead, biomedical semantics, sequential context, and selective feature weighting are combined before classification. This arrangement provides the architectural basis for the hybrid model evaluated in the study.

2.4 Model Simulation and Validation

Implementation was performed in Python with TensorFlow and Keras in Google Colaboratory. The dataset was divided into training and testing portions at an 80% to 20% ratio. The hybrid network was trained for 10 epochs with the Adam optimiser, a learning rate of 0.0005, categorical cross-entropy loss, and a batch size of 32. BiLSTM, Random Forest, Support Vector Machine, and Logistic Regression were also trained using the same data split. Accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC) were used to assess all models.

Received: 01/09/2026; Revised: 12/09/2026; Accepted: 14/09/2026; Published: 25/09/2026

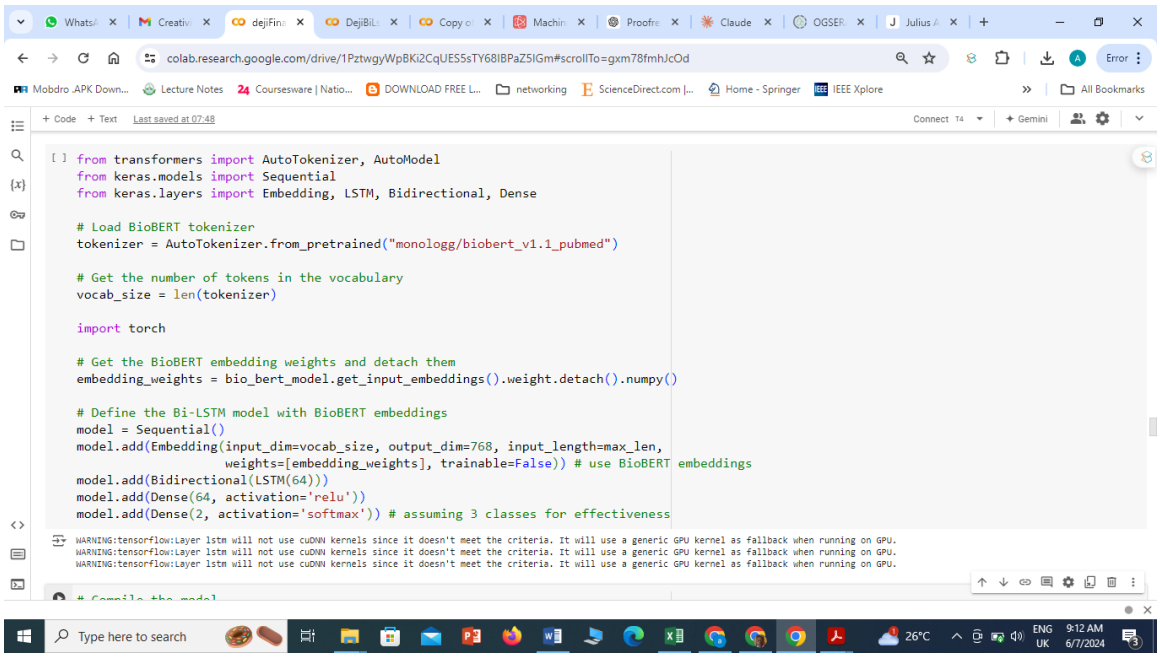

Figure 2: Implementation of the BioBERT Tokenizer, BiLSTM Model, and Training Routine
Source: Field Work (2024)

Figure 2 documents the implementation workflow used for model construction and training. The displayed routine connects BioBERT tokenisation and embedding extraction with the BiLSTM sequence-processing stage and the subsequent training operations. Its main significance is methodological: whereas Figure 1 explains the conceptual flow of the classifier, Figure 2 shows how that design was translated into an executable programme within Google Colaboratory. The figure therefore provides implementation-level support for the experimental procedure used to obtain the reported results.

3.0 Results and Discussion

3.1 Training and Validation Performance

Across the 10 training epochs, the loss on the training data fell from 0.6794 to 0.6633, whereas the validation loss remained approximately 0.6727. Training accuracy increased from 0.5935 to 0.6000, while validation accuracy stayed close to 0.5984. The relatively similar training and validation behaviour provides no clear indication of substantial overfitting during this period. However, the near-0.60 accuracy plateau also suggests that the selected training configuration had reached limited improvement by the end of the 10 epochs. This behaviour is important when interpreting the test-set results because the eventual performance reflects a model that learned without a large separation between training and validation behaviour, but also without achieving strong predictive discrimination.

3.2 Model Evaluation Metrics

Table 1 reports the classifier's performance on the held-out test data for the two sentiment categories. Precision is 0.55 for both negative and positive reviews. Recall differs slightly, reaching 0.49 for Class 0 (Negative) and 0.59 for Class 1 (Positive), while the respective F1-scores are 0.52 and 0.55. Both the macro and weighted averages are 0.54 for precision, recall, and F1-score. The principal result is therefore a broadly even performance across the two classes rather than a strong preference for one category. At the same time, the lower negative-class recall indicates that a proportion of negative reviews may be missed. In a pharmacovigilance setting, this is an important limitation because negative experiences may contain information relevant to adverse drug reactions and medicine safety.

Table 1: Classifier's performance on the held-out test data for the two sentiment categories

Metric	Class 0 (Negative)	Class 1 (Positive)	Macro Avg	Weighted Avg
Precision	0.55	0.55	0.54	0.54
Recall	0.49	0.59	0.54	0.54
F1-score	0.52	0.55	0.54	0.54

.

Received: 01/09/2026; Revised: 12/09/2026; Accepted: 14/09/2026; Published: 25/09/2026

3.3 Comparative Analysis

The comparison shows that the BioBERT-BiLSTM model performs best in overall accuracy, although its advantage over the baseline methods is relatively small, ranging from 2 to 5 percentage points. AUC values for all five models remain close to 0.50, indicating that none of the classifiers demonstrates strong separation between positive and negative reviews. Another useful distinction concerns class recall. SVM records a recall of 0.93 for negative reviews but only 0.07 for positive reviews, whereas the proposed model produces the more even values of 0.49 and 0.59. For pharmacovigilance, this balance is useful because both favourable treatment experiences and reports of adverse experiences can contribute to safety assessment. A classifier that heavily favours one sentiment category could consequently overlook relevant patient information.

Table 2: Model Performance Comparison for Malaria Drug Reviews

Model	Accuracy (%)	Precision (Avg)	Recall (Avg)	F1-score	AUC
BioBERT-BiLSTM	60.90	0.54	0.54	0.54	0.54
Logistic Regression	58.84	0.54	0.59	0.49	0.51
Support Vector Machine	58.06	0.51	0.58	0.48	0.50
Random Forest	56.49	0.51	0.56	0.50	0.51
BiLSTM	55.40	0.51	0.55	0.51	0.50

Table 2 brings the five classifiers together for a direct performance comparison. BioBERT-BiLSTM records the highest accuracy at 60.90%, with precision, recall, F1-score, and AUC all at 0.54. Logistic Regression records 58.84% accuracy, precision of 0.54, recall of 0.59, F1-score of 0.49, and AUC of 0.51. SVM obtains 58.06% accuracy, precision of 0.51, recall of 0.58, F1-score of 0.48, and AUC of 0.50. Random Forest records 56.49% accuracy, precision of 0.51, recall of 0.56, F1-score of 0.50, and AUC of 0.51, while BiLSTM records 55.40% accuracy, precision of 0.51, recall of 0.55, F1-score of 0.51, and AUC of 0.50. Thus, the proposed architecture provides the highest accuracy, precision, F1-score, and AUC in the comparison, although its recall is lower than those of the four baselines. The accuracy difference relative to the comparison models ranges from 2.06 to 5.50 percentage points. Because the AUC values remain close to 0.50, these results should be interpreted as a modest comparative advantage rather than evidence of strong sentiment discrimination.

3.4 Web and Mobile Prototype Implementation

After training, the classifier was exposed as a back-end service through an API linked to a database containing patient drug reports. Two user interfaces were developed around this service: a Flask/Jinja2 web application and a Java-based Android application. Healthcare professionals, patients, and regulators can use the interfaces to submit review text and obtain real-time sentiment classifications. This deployment arrangement separates the predictive model from the user interface and provides a practical route for integrating the classifier into a larger reporting workflow.

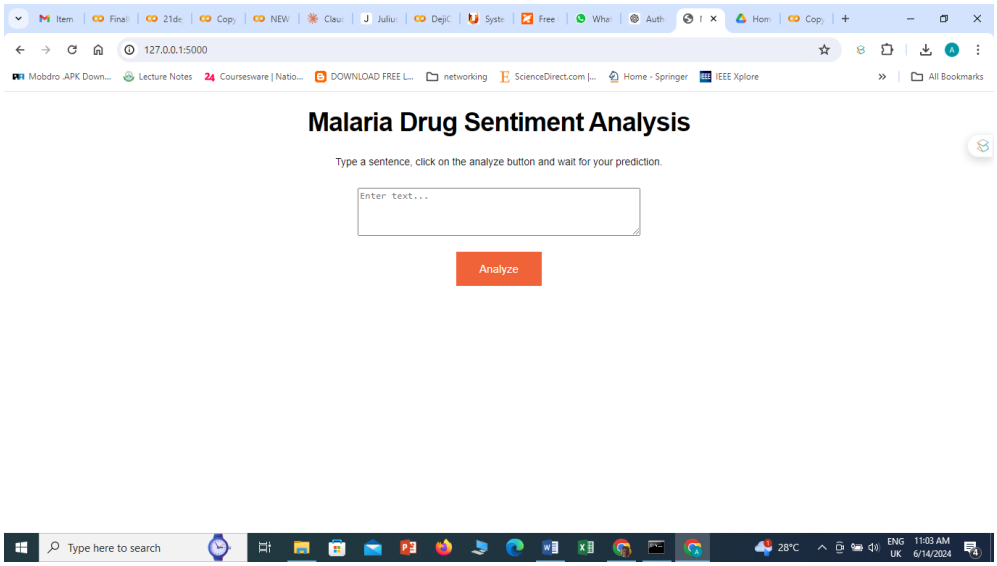

Figure 3: Web-based Prototype for Malaria Drug Sentiment Analysis
Source: Field Work (2024)

Figure 3 shows the web implementation used to access the sentiment-analysis service. The interface provides a text box for entering a malaria drug review and an analysis control for submitting the text to the model. Its key contribution is to demonstrate how the classifier can be made accessible to users without exposing them to the

underlying machine-learning implementation. In a pharmacovigilance workflow, such an interface could serve as an initial point for processing patient feedback before the information is subjected to further review.

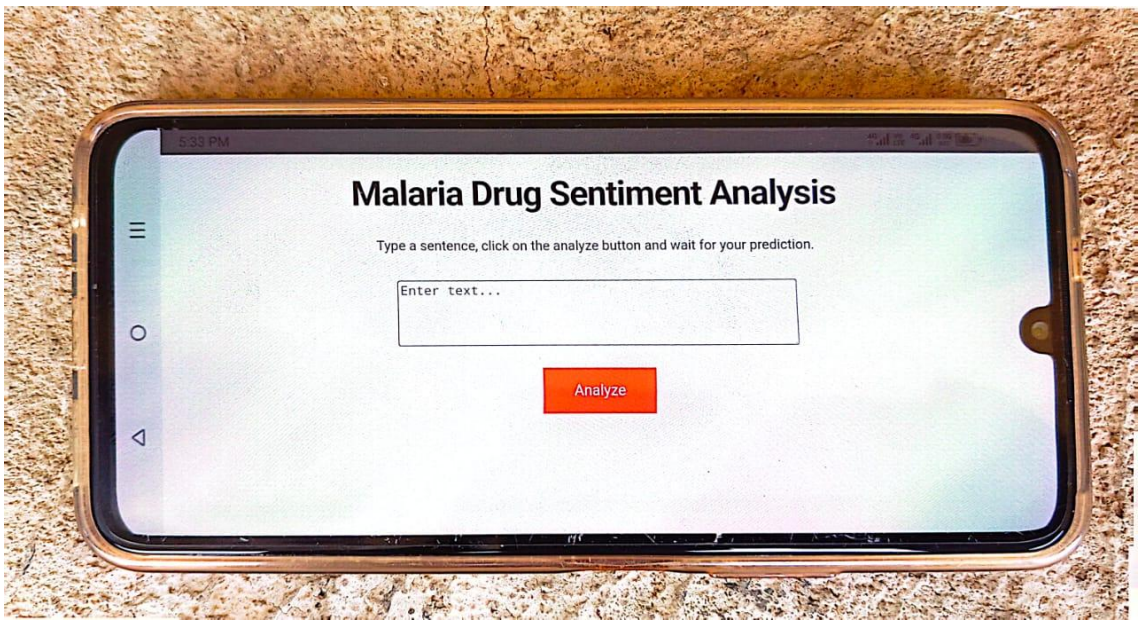

Figure 4: Android Mobile-Based Prototype for Malaria Drug Sentiment Analysis
Source: Field Work (2024)

Figure 4 depicts the Android version of the prototype. The mobile interface offers the same essential operation: a user enters review text and requests a sentiment prediction from the deployed model. The significance of the mobile implementation is its potential to extend access to the analysis service through handheld devices. Considered alongside Figure 3, the two prototypes demonstrate that the same back-end classification service can support both browser-based and mobile interaction, although their practical pharmacovigilance value ultimately depends on the predictive reliability of the underlying model.

3.5 Discussion

The performance pattern can be related to the different roles assigned to the three principal components of the hybrid architecture. BioBERT has been trained on PubMed and PMC material, giving its representations exposure to biomedical terminology and expressions that can occur in descriptions of symptoms and side effects [18]. BiLSTM adds sequence-sensitive modelling by considering information in both directions of a review. This is useful when a single patient narrative contains favourable and unfavourable statements in different parts of the text. Attention provides a further selection mechanism by assigning greater influence to words that carry useful sentiment information, including expressions associated with effectiveness, recovery, nausea, or headache.

The observed pattern also relates to previous findings. Shiju and He (2021) reported the strongest F1-score for BioBERT among the transformer models examined for drug-rating classification [16]. Sweidan et al. (2021) similarly demonstrated the usefulness of combining domain knowledge with transformer-based methods for ADR classification [17]. The present work applies these ideas to malaria drug reviews in Nigeria and adds bidirectional recurrent processing with attention to the biomedical representation. The contribution is therefore contextual as well as architectural, since the model is evaluated using patient-generated reviews from the Nigerian setting.

The web and Android implementations provide an operational demonstration of how the classification pipeline could be incorporated into a pharmacovigilance environment such as PRASCOR. Nevertheless, the performance obtained in Sections 3.2 and 3.3 places a clear boundary on this interpretation. With an accuracy of 60.90% and an AUC of 0.54, the system should presently be considered a proof-of-concept triage aid. It should not be treated as a validated diagnostic instrument or as an autonomous mechanism for making pharmacovigilance decisions.

The findings should also be considered alongside several study limitations. The corpus contains 3,194 reviews, which is relatively small for fine-tuning a deep-learning model and may restrict the diversity of linguistic patterns available during training. The analysis also covers English-language reviews only, limiting immediate transfer to multilingual environments. In addition, the experiment relies on a single 80/20 split and does not report cross-validation or significance testing. Consequently, the accuracy differences observed across the models, from 55.4% to 60.9%, should be interpreted cautiously rather than regarded as statistically established performance gaps.

Received: 01/09/2026; Revised: 12/09/2026; Accepted: 14/09/2026; Published: 25/09/2026

Finally, changes in medical terminology and the ways patients describe treatment experiences may necessitate periodic retraining.

4.0 Conclusion

This study developed a hybrid BioBERT-BiLSTM classifier with attention for analysing malaria drug reviews in support of pharmacovigilance in Nigeria. The approach combines biomedical transfer learning, bidirectional sequence modelling, and attention-based feature weighting. On the reported test set, the hybrid model produced stronger overall results than the traditional machine-learning and standalone deep-learning comparators. Its performance across the two sentiment classes was also relatively balanced, which is relevant because pharmacovigilance depends on information from both favourable and unfavourable treatment experiences.

The study also demonstrated a practical deployment pathway through web and Android interfaces. The trained classifier operates as a back-end service linked by an API to a database of patient drug reports. New reports can be retrieved, processed, classified, and returned to the application interface for presentation to the user. This arrangement illustrates how automated text analysis could complement existing approaches to drug-safety monitoring, particularly in settings where large volumes of patient feedback are difficult to review manually.

The results reinforce the value of using biomedical pre-training when analysing healthcare-related text and suggest that combining transfer learning with sequential modelling can provide a useful basis for patient-review classification. Future investigations should enlarge and diversify the dataset, explore alternative architectures and hyperparameter configurations, and assess the model with stronger validation procedures before operational deployment. Integration with national pharmacovigilance infrastructure could then be examined under appropriate human oversight. The same methodological approach may also be adapted to other drug categories and medical conditions, extending its potential contribution to automated pharmacovigilance and patient-safety monitoring.

References

[1] World Health Organization. (2017). World Malaria Report. World Health Organisation.
[2] Woolf, H., Aron, L., Dubay, L., Simon, S., Zimmerman, E., & Luk, K. (2015). How are Income and Wealth linked to Health and Longevity? Urban Institute Centre on Society and Health, 1–10.
[3] Ahuru, R., & Omon, I. (2018). The Economic Burden of Malaria: Evidence from Nigeria Data. Amity Journal of Health Management, 3(1), 28–39.
[4] Ogar, C., Ibrahim, A., & Osakwe, A. (2018). Pharmacovigilance Rapid Alert System for Consumer Reporting (PRASCOR): A Look at Its Quantitative Contribution to Spontaneous Reporting in Nigeria from August 2012 to February 2017. Pharmaceutical Medicine, 32, 131–141.
[5] Nigerian National Pharmacovigilance Policy and Implementation Framework (NNPP). (2020). Abuja, Nigeria: Federal Ministry of Health.
[6] Obayes, H., Al-Turaihi, F., & Alhussayni, K. (2021). Sentiment Classification of Users' Reviews on Drugs based on Global Vectors for Word Representation and Bidirectional Long Short-Term Memory Recurrent Neural Network. Indonesian Journal of Electrical Engineering and Computer Science, 23(1), 345–353.
[7] Mahboob, K., & Ali, F. (2018). Sentiment Analysis of Pharmaceutical Products Evaluation based on Customer Review Mining. Journal of Computer Science and Systems Biology, 11(3), 190–194.
[8] Gawand, Y., & Varma, S. (2019). Enhanced Opinion Mining Approach for Product Reviews. International Journal of Innovative Science and Research Technology, 4(8), 612–615.
[9] Haque, R., Laskar, S. H., Khushbu, K. G., Hasan, M. J., & Uddin, J. (2023). Data-Driven Solution to Identify Sentiments from Online Drug Reviews. Computers (MDPI), 12(4), 1–22.
[10] Colón-Ruiz, C., & Segura-Bedmar, I. (2020). Comparing Deep Learning Architectures for Sentiment Analysis on Drug Reviews. Journal of Biomedical Informatics, 110, 1–11.
[11] Asri, A. M., Ahmad, S. R., & Yusop, N. M. (2023). Feature Selection using Particle Swarm Optimization for Sentiment Analysis of Drug Reviews. International Journal of Advanced Computer Science and Applications, 15(5), 286–295.
[12] Uddin, M. N., Hafiz, M. F., Hossain, S., & Islam, S. M. (2022). Drug Sentiment Analysis using Machine Learning Classifiers. International Journal of Advanced Computer Science and Applications, 13(1), 92–100.
[13] Colón-Ruiz, C. (2021). Semi-Supervised Generative Adversarial Network for Sentiment Analysis of Drug Reviews. TechRxiv, 1–10.
[14] Suhartono, D., Purwandari, K., Jeremy, N. H., Philip, S., Arisaputra, P., & Parmonangan, I. H. (2023). Deep Neural Networks and Weighted Word Embeddings for Sentiment Analysis of Drug Product Reviews. Procedia Computer Science, 216, 664–671.
[15] Rufai, K. I., Olusanya, O. O., Adebare, A. O., & Usman, O. L. (2022). The Development of a Framework for the Sentiment Analysis of Drug Review Dataset Using Gradient Boosting Algorithm. KWASU Journal of Information, Communication and Technology, 1(1), 17.

Received: 01/09/2026; Revised: 12/09/2026; Accepted: 14/09/2026; Published: 25/09/2026

[16] Shiju, A., & He, Z. (2021). Classifying Drug Ratings Using User Reviews with Transformer-Based Language Models. MedRxiv, 1–10.
[17] Sweidan, A., El-Bendary, H., & Al-Feel, H. (2021). Sentence-Level Aspect-based Sentiment Analysis for Classifying Adverse Drug Reactions (ADRs) using Hybrid Ontology-XLNet Transfer Learning. IEEE Access, 9, 1–19.
[18] Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2019). BioBERT: A Pre-trained Biomedical Language Representation Model for Biomedical Text Mining. Bioinformatics, 36(4), 1234–1240.

Received: 01/09/2026; Revised: 12/09/2026; Accepted: 14/09/2026; Published: 25/09/2026