

Leakage-Aware Fake-News Screening for Nigerian Political Claims: Temporal and Domain Evaluation of a Context-Aware Ensemble

¹Ndudi-Wali Kinikanwo Wisdom, ²Prof. Daniel Matthais, ³Dr. Victor T Emmah

^{1,2,3}Computer Science Department, Rivers State University, Nigeria

E-mail: ¹wali.wisdom@rsu.edu.ng, ²matthias.daniel@ust.edu.ng, ³victor.emmah@ust.edu.ng

Abstract: Automatic fake-news classifiers can learn publisher, recurring topics, and other factors unrelated to factual veracity. In this paper, a context-aware classifier is evaluated under temporal and event grouping for Nigerian political and election-related claims. The global training set included a corpus of 29,269 non-synthetic fact-check claims descriptions and a manually verified supplement of 135 Nigerian fact-check claims collected from 46 fact-check web pages and 75 event groups. Verdicts were classified as REAL if they were unequivocally TRUE, CORRECT, or ACCURATE and classified as FAKE if they were FALSE, INCORRECT, or FAKE. The global backbone used a combination of word and character TF-IDF feature sets with Logistic Regression classifier, calibrated Linear Support Vector Machine, Random Forest, Gradient Boosting, and validation-weighted soft voting. A separate context head using a Logistic Regression model was trained from grouped out-of-fold predictions on 101 claims with dates of 2023-2025 and added to the global ensemble as a blend. All 34 claims from 2026 were reserved for final testing. Contextual blend achieved 0.6765 accuracy, 0.6825 balanced accuracy, 0.6762 macro F1-score, and 0.7368 ROC-AUC, and the REAL and FAKE recall of 0.6316 and 0.7333, respectively. The global ensemble achieved 0.6947 balanced accuracy and 0.6964 macro F1-score in the same test. The difference in balanced accuracies was -0.0123 , and its 95% event-group bootstrap confidence interval was $(-0.1399; 0.0947)$, with an exact p-value of 1.000.

Keywords: Fake-news screening; Nigerian political claims; temporal validation; domain shift; ensemble learning.

I. INTRODUCTION

Politically deceptive claims spread through social media, traditional media, campaigning, and individual-level interaction in Nigeria. The effect of these claims depends on the fault lines existing along political, ethnic, and religious lines [1]; the 2023 Nigerian presidential campaign showed that the information disorder does not necessarily have to occur at the margins but could also be part of mainstream campaigning [2]. Automated processes may aid journalists, observers, and fact checkers in filtering the high number of claims; however, it is important to remember that a classifier does not replace facts; it only finds statistical correlations unless it gets and evaluates evidence independently.

Nigeria represents a difficult NLP environment. Nigerian English alongside Nigerian Pidgin, other indigenous languages, code switching, abbreviated forms, and unpredictable spellings coexist within Nigeria [3]. This has seen transfer learning being used for detection of Hausa language [4]; Nigerian LSTM models perform differently based on class imbalance [5], and

culturally appropriate BERT models focus on multilingualism [6]. These contributions represent technical viability but make it difficult to compare scores when corpora, labels, time periods, sources, and validation procedures differ.

Methodologically, fake-news detection covers a broad spectrum from sparse lexical classifiers to neural, transformer, graph, and explainable systems. While classical machine learning approaches such as logistic regression, support vector machines, tree-based classifiers, and ensemble methods perform well on specific datasets, BERT and RoBERTa use contextual information that cannot be captured by sparse vectors [7]. Reviews of deep learning approaches show an overall trend towards models that combine textual, social, and knowledge signals [8] in addition to graph neural network approaches that consider propagation structures and relations between users, posts, and news [9]. Explainability has become an important feature since prediction performance is not enough for making classification decisions that can have political or social consequences [10]. Studies in Nigeria show the potential of machine learning approaches to automate the detection of fake

news [5], [11], however, fewer studies exist regarding leakage-aware approaches in which different types of evidence (lexical, stylistic, pragmatic, and meta-information) are separated before combination.

The research asks three questions: 2024-2025: Can a lightweight text ensemble tell real from fake claim descriptions? How relevant is the frozen model to the Nigerian claims, especially those relating to politics and elections, which are not part of all development decisions? Model comparisons, feature coefficients, source-stratified errors, and bootstrap intervals offer insights into ensemble value and shortcut risk.

The research provides a replicable validated-claim framework, a secure temporal/domain assessment, and a clear explanation of the circumstances in which model integration is beneficial. It does not argue for ensemble complexity as inherently superior or linguistic prediction establishes truth.

II. RELATED WORK AND RESEARCH GAP

2.1 Classical, Ensemble, and Contextual Models

Linear models applied to sparse TF-IDF features remain well-suited owing to their efficiency and discriminative capacity. At the same time, the comparative analysis shows high variability in performance orders of Logistic Regression, SVMs, tree-based approaches, deep learning, and transformer architectures across datasets [7]. The analysis of deep learning shows growing emphasis on the combination of textual, social, and knowledge information [8] while graph-based approaches involve propagation and relations between users and content not available to text-based models [9]. The problem of interpretability becomes especially important in case predictions refer to politically relevant speech acts [10].

Ensembling can help avoid being dependent on one decision boundary only if the individual models are complementary in their errors. Voting is a method that takes into account probabilities of the individual models either using static or validation weights, and stacking involves a meta-model that works on out-of-fold predictions. Both do not guarantee generalization. When the individual models use the same cue for a shortcut, the ensembling can perpetuate the shortcut rather than correct it. Therefore, the appropriate test for models would be through comparison to robust single models in a distribution-shift setting.

2.2 Nigerian Fake-News Detection

Nigerian computational research includes political communication, imbalance in class, low-resource language, and culture-specific representation. Esan et al. [5] showed that results of LSTM depend on whether the dataset is balanced or not balanced. Combinations of methods, such as machine learning and topological data analysis, have been tried as well [11]. Results from the small-data research show that the topological information might prove useful for the cases with a limited number of samples [12]. Recent research using BERT widened the linguistic horizon [6]. However, heterogeneity of sources, labeling, languages, and evaluation prevents simple leaderboard understanding.

2.3 Gap Addressed by the Revised Study

The unresolved gap is credible transfer evaluation for a Nigerian political use case. Higher same-corpus accuracy is not enough when claims are repeated, publishers use similar editorial styles, fact-checkers describe claims differently, and the class prevalence shifts over time. The updated research addresses this deficiency by extracting a stringent binary subset from the ecosystem-scale FACTors corpus [13], ensuring that Nigerian entries do not interfere with development, setting thresholds on an intermediate year, evaluating subsequent years, and reviewing records from publishers not included in the training. The collection is kept as a hypothesis about additional decision boundaries and its value is assessed independently of the complexity.

III. MATERIALS AND METHODS

3.1 Research Design and Unit of Analysis

This research is a supervised binary classification experiment with the objective of understanding, preparing, modelling and evaluating data, following the CRISP-DM reproducibility focus [14]. The analytical unit is a claim description written by a fact-checker, not the whole news article. Descriptions are rephrased original social media posts or public statements. The model does not receive the evidence from the fact-checker or the text of the verdict. So its output is a linguistic screening score, not proof of factual truth.

3.2 Global Verified Corpus

The global framework was derived from FACTors [13], which consists of 118,112 English claims in 117,993 fact-checking reports produced by 39 fact-checking organisations.

The deterministic filtering retained only unqualified judgements with uniform upstream ratings, discarded source fields that could potentially lead to claim-conclusion inversion, ignored visual-only accusations and editorial counterarguments, reconciled label discrepancies, and removed redundant claim identifiers or standardised precise text. The resulting corpus consists of 29,269 authentic, source-referenced claim descriptions: 24,518 FAKE and 4,751 REAL.

Table 1: Corpus construction and quality-control audit

Stage	Records	Decision rule
FACTors source	118,112	All available source rows
Verdict/rating exclusion	70,800	Not one of six direct verdicts or normalized rating disagreed
Source inversion-risk exclusion	3,419	Africa Check claim field was excluded after audit found conclusion/claim inversion risk
Unusable claim exclusion	14,032	Visual-only (12,393), editorial/rebuttal (1,237), or length/URL (402)
Verdict conflict exclusion	6	Same claim identifier or normalized text carried opposing labels
Duplicate exclusion	586	Repeated claim identifier or normalized exact text
Retained strict corpus	29,269	Source-linked binary text claims satisfying all rules

Global records were grouped based on the fact-check report. Fact-check reports with a Nigeria-tagged entry were not used for global training. The rest of the fact-check reports were then divided based on the year of their most recent report: up to 2022 for model fitting, 2023 for validation, and 2024 to 2025 for time-based evaluation. This grouping avoids any report duplicates in the same partitioning while making sure that Nigeria's entries are out of all the stages.

3.3 Strict Nigerian Political and Election Supplement

The Nigerian dataset consists of 135 claim descriptions that are associated with 46 FactCheckHub public pages and 75 events/reports. In all, there are 87 documents which have been gathered for this research. Key sources are checks from candidate speeches in the 2023 election [15], [16], live checks during the election period in Ekiti and Osun states [17], [18], and government performance checks [19]. There are 75 REALs and 60 FAKES in this corpus. Every document retains a consistent record ID, claim description, publisher, fact check URL, precise verdict, publish date, mapped class, event group, category, and evidence note.

Only exact, unqualified TRUE, CORRECT, or ACCURATE verdicts map to REAL. Exact FALSE, INCORRECT, or FAKE verdicts map to FAKE. Verdicts like MOSTLY TRUE, PARTLY CORRECT, MISLEADING, UNPROVEN, and NOT DIRECTLY RATED were not considered. The authors did not derive any information about the truthfulness of the statements from the story contained in the article. There were no generated stories or verdicts. The row text provides a standardised form of the observed claim, and the publisher verdict gives the class. Agreement between annotators is irrelevant here since the study does not involve re-evaluating the publisher's verdicts.

Table 2: Temporal roles in the Nigerian supplement

Role	Years	REAL	FAKE	Total
Adaptation training	2023	35	39	74
Model development	2024-2025	21	6	27
Prospective test	2026	19	15	34
Total	2023-2026	75	60	135

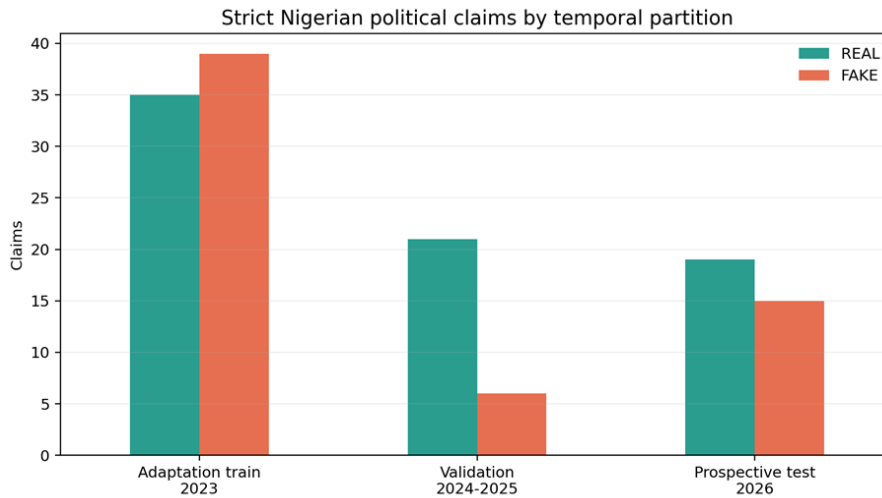


Figure 1: Nigerian strict-verdict claims by temporal role and class

3.4 Leakage Control and Development Protocol

The Nigerian split is organised in chronological order. The data recorded in 2023 form the adaptation training; data recorded in 2024 and 2025 form the development; and the data recorded in 2026 form the future assessment. No event type goes over the partitions. The 34 test data items have not been used for fitting the TF-IDF features, estimating the model parameters, calculating the out-of-fold development scores, setting the blend weights, deciding the decision thresholds, or selecting the model families.

The development set is class-imbalanced, with 21 REAL and 6 FAKE entries. As a result, accuracy was not the primary criterion for choosing a classifier, since balanced accuracy and macro-F1 were considered instead. In turn, the future test data set is more balanced, having 19 REAL and 15 FAKE cases, but still relatively small in size and local to fact checks.

3.5 Text Processing and Feature Representation

The processing steps included URL, username, and retweet indicator removal; hashtag term retention; conversion to Unicode case-folding; removal of non-word noise characters; and normalisation of white spaces. A conservative mapping of some Nigerian expressions, including “dis,” “una,” “dem,” “wetin,” and “dey,” was done. No stop-word deletion and aggressive stemming was performed since negation and functional word patterns may affect claim meaning. Upon audit, it was found that none of the five expressions mapped was in the 219 Nigerian political claims, hence no performance gain from this dictionary.

The selected sparse representation included word unigrams and bigrams (up to 50,000 features) and character 2–6-grams (up to 100,000 features) with a minimum document frequency of two. The integrated sparse matrix was employed by the calibrated Linear Support Vector Machine and Logistic Regression. Random Forest and histogram Gradient Boosting used a 200-component truncated-SVD representation, trained on the training dataset only. Character n-grams were intended to capture spelling and subword diversity; they do not substitute for evidence or local-language representation.

3.6 Classifiers and Context Adaptation

The global model consists of four probability generators, namely: Logistic Regression with balancing of classes; Linear Support Vector Machine with threefold sigmoid calibration with balancing of classes; Random Forest with 300 trees and class balancing; histogram Gradient Boosting with 180 estimators. Each of them produces the probability of being FAKE for the respective claim. The validation-balanced accuracy serves as the criterion for non-negative weights assignment in the ensemble. The weighted global score is then thresholded to produce the class using the 2023 global validation split.

Table 3: Implemented model configuration

Component	Configuration
Logistic Regression	C=1.0; liblinear; balanced classes; sparse fused TF-IDF
Support Vector Machine	Linear SVM, C=1.0; balanced classes; sigmoid calibration
Random Forest	300 trees; minimum leaf 3; sqrt features; balanced subsamples
Gradient Boosting	Learning rate 0.08; 180 iterations; 31 leaves; L2=1.0
Global soft vote	Validation weights 0.2781 LR, 0.2495 SVM, 0.2458 RF, 0.2267 GB
Nigerian context head	Balanced Logistic Regression; five-fold stratified group OOF
Context blend	0.70 local head + 0.30 global ensemble; threshold 0.544557

For the context head, five-fold StratifiedGroupKFold was used to compute out-of-fold probabilities of all 101 Nigerian observations for 2023–2025, where the event group is the grouping variable. Local candidate weights from zero to one with a step of 0.05 were chosen based only on the out-of-fold probabilities. For the obtained blend, 0.70 weight was attributed to the Nigerian head, 0.30 to the global ensemble, and the threshold is 0.544557. Then, the Nigerian head was fitted again to all 101 observations of the development set. The separate direct-augmentation experiment included the appending of 74 Nigerian 2023 observations to the global training data set, keeping the four models structure. No candidate weights of 1, 2, 4, 8, and 12 improved the development criterion; hence, the smallest tied value one was chosen.

3.7 Evaluation and Statistical Analysis

Evaluation measures consist of accuracy, balanced accuracy, macro F1, class-specific precision, recall and F1 score, ROC-AUC, average precision, and confusion matrix. Balanced accuracy measure calculates the average of REAL and FAKE recall values, while macro-F1 considers both F1 values equally. These measures provide better information compared to accuracy when there is an imbalance between the classes [20], [21]. The confusion matrix is structured such that rows correspond to the real values of REAL and FAKE, while columns represent the predicted values of REAL and FAKE.

This selected blend was compared to the unmodified global set using the same test examples. The discrepancies in terms of balanced accuracy and macro F1 were assessed by drawing 5,000 bootstrap samples pairs in an event group manner. The exact two-sided paired test for discrepancy in disagreement in correctness was calculated. The uncertainty analysis is still descriptive because the test only contains 34 cases.

IV. RESULTS

4.1 Development Selection

Out-of-fold development prediction in groups used a 0.70 local and 0.30 global mix. For the cutoff value of 0.544557, the development balance accuracy and the macro-f1 scores were equal to 0.6683 and 0.6686, respectively. These parameters were constant before analysing the 34 statements for 2026. No direct boost could be achieved by raising the local sample weight, which means that repeating the Nigerian training samples had no more value.

4.2 Prospective 2026 System Performance

Table 4: System performance on the untouched 2026 Nigerian test

System	Accuracy	Balanced accuracy	Macro-F1	ROC-AUC	FAKE recall
Global weighted ensemble	0.7059	0.6947	0.6964	0.7719	0.6000
Direct augmentation	0.6765	0.6544	0.6521	0.7579	0.4667
Nigerian context blend	0.6765	0.6825	0.6762	0.7368	0.7333

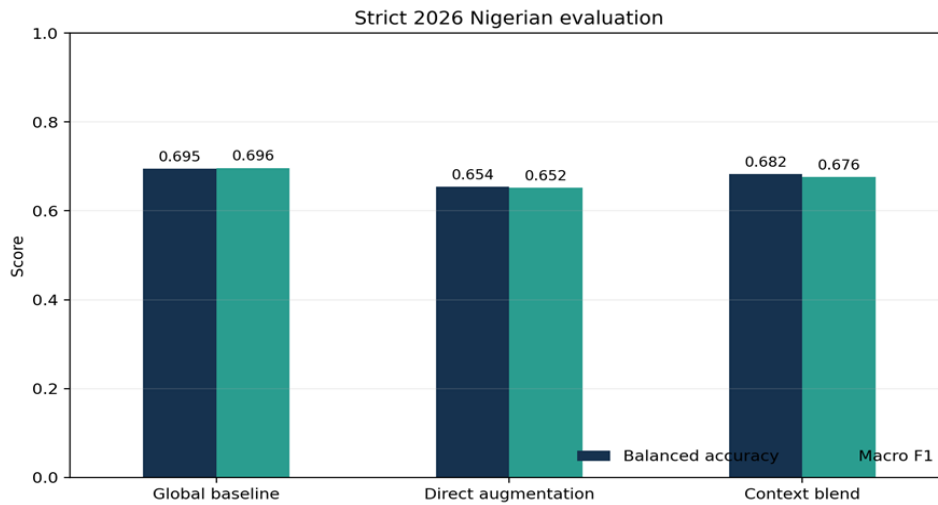


Figure 2: Performance of the three systems on strict 2026 Nigerian claims

The ensemble built on the basis of the Nigerian context gave the following results: accuracy equals 0.6765, balanced accuracy equals 0.6825, macro-F1 equals 0.6762, ROC-AUC equals 0.7368. This ensemble had the best FAKE recall (0.7333), while its REAL recall was worse than the REAL recall of the global ensemble. Global weighted ensemble turned out to be the best in terms of balanced accuracy, macro-F1, and ROC-AUC. Direct augmentation gave the worst balanced accuracy and macro-F1, therefore, it means that adding local rows did not improve the transfer when there was no additional context representation.

4.3 Individual Model Results

Table 5: Global model family on strict 2026 Nigerian claims

Model	Balanced accuracy	Macro-F1	ROC-AUC	TN/FP/FN/TP
Logistic Regression	0.7281	0.7294	0.7123	15/4/5/10
Support Vector Machine	0.5561	0.5554	0.6386	11/8/7/8
Random Forest	0.6281	0.6264	0.7368	15/4/8/7
Gradient Boosting	0.7088	0.7049	0.7509	13/6/4/11
Weighted ensemble	0.6947	0.6964	0.7719	15/4/6/9

Logistic Regression generated the best observed balanced accuracy (0.7281) and macro F1 (0.7294) scores, whereas the weighted ensemble generated the best ROC-AUC score (0.7719). Gradient Boosting produced a FAKE recall score of 0.7333 and balanced accuracy of 0.7088. Since these ranks were determined after the unlocking of the test data set, the results are merely exploratory and do not override the pre-selected system. However, it is clear from these ranks that increasing model complexity does not guarantee improved performance at the chosen operating point in the prospective small sample size.

4.4 Class-Specific Errors

Table 6: Class-specific performance and confusion counts

System	REAL P/R/F1	FAKE P/R/F1	Confusion matrix
Global weighted ensemble	0.7143 / 0.7895 / 0.7500	0.6923 / 0.6000 / 0.6429	[[15, 4], [6, 9]]
Nigerian context blend	0.7500 / 0.6316 / 0.6857	0.6111 / 0.7333 / 0.6667	[[12, 7], [4, 11]]

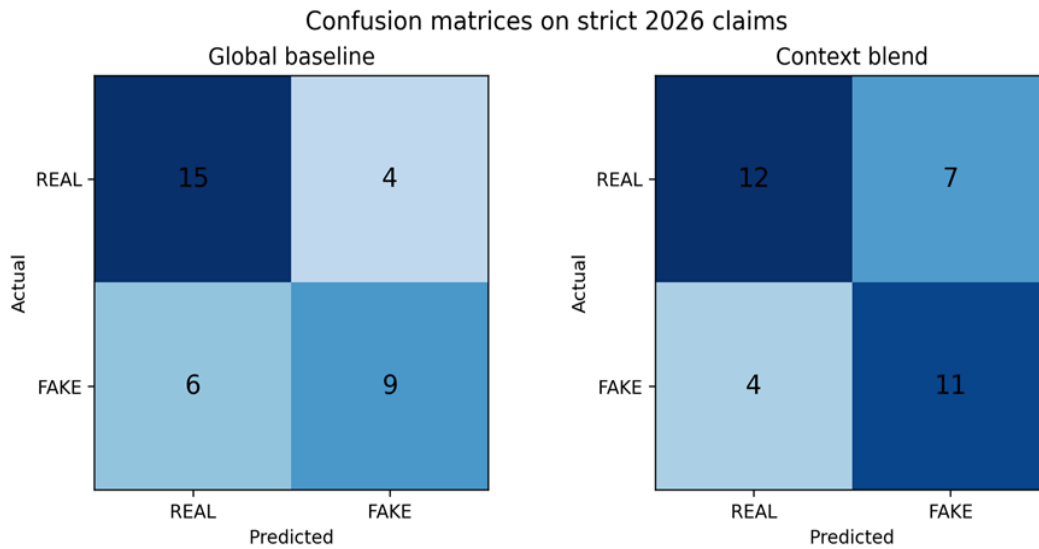


Figure 3: Confusion matrices for the global ensemble and Nigerian context blend

The contextual blend was able to correctly identify 12 out of 19 REAL claims and 11 out of 15 FAKE claims. Compared to the ensemble of classifiers, the contextual blend was able to correctly classify two more FAKE claims but was false on three more REAL claims. It could be beneficial in a triage setting, where suspicious political claims receive priority, as long as users are aware that being more sensitive to FAKE means a higher false rate for REAL claims that would have to be examined.

4.5 Ranking Performance and Paired Uncertainty

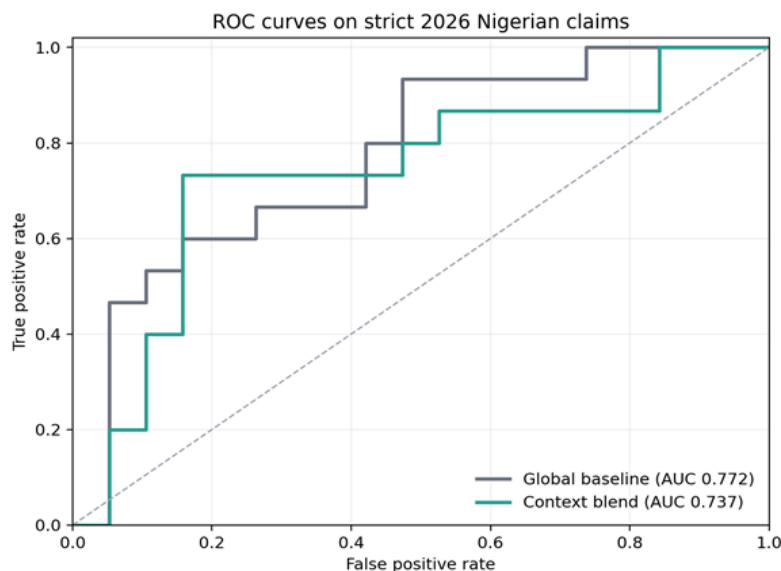


Figure 4: ROC curves on the strict 2026 Nigerian test

The global model generated an ROC-AUC of 0.7719 compared to 0.7368 for the context ensemble. Thus, the higher FAKE recall for the context ensemble depends on the threshold selection, not its superiority in ranking in comparison with other thresholds. Both metrics with and without thresholds support the same view that the local head impacts the distribution of errors but fails to improve discrimination.

Table 7: Paired context-minus-global uncertainty

Metric	Difference	95% event-group bootstrap interval	Paired p
Balanced accuracy	-0.0123	-0.1399 to 0.0947	1.000
Macro-F1	-0.0202	-0.1461 to 0.0926	-

The difference in context minus global was found to be -0.0123 and -0.0202 for balanced accuracy and macro-F1, respectively. In both cases, 95% confidence intervals contained the value of zero, while the exact probability of discordant correctness results was 1.000. As such, there is no statistically significant prediction advantage shown by the Nigerian context mix in the experiment. The wide range of intervals demonstrates that an independent Nigerian sample should be obtained first.

V. DISCUSSION

5.1 Principal Findings

The experiment resulted in three main findings. A Nigerian supplement with source references can be useful initially for a very promising assessment, and further assertion is robust to any model-selection choices. Second, the Nigerian context head increases FAKE recall, but not balanced accuracy, macro-F1 or ROC-AUC, compared with the global ensemble on the same records. Third, Logistic Regression still remains a strong benchmark, with the highest exploratory balanced accuracy, suggesting that a more complex ensemble should only be kept if it clearly offers operational advantages.

The main contribution is therefore based on evidence rather than promotional. The 135-row supplement enhances Nigerian representation, preserves accurate fact-check lineage, filters out qualified judgements, and organises events chronologically. However, a better quality of dataset does not always lead to a better classifier. Results are consistent with domain adaptation theory: a localised model can overfit limited publisher or event patterns when only 101 pre-test samples are available, particularly if the distribution of development classes differs from that of the expected test.

5.2 Comparison to Nigerian Studies

Table 8: Relationship to published fake-news detection studies

Study	Data and evaluation	Reported result	Relationship to this study
Esan <i>et al.</i> [5]	Nigerian news; LSTM; SMOTE-balanced and imbalanced experiments	92.86% accuracy and 93% F1 after SMOTE; minority failure when imbalanced	Higher reported scores, but article-level data and synthetic oversampling differ from prospective strict claims
Ibrahim <i>et al.</i> [4]	More than 6,600 Hausa articles; cross-lingual transfer learning	High classification performance reported across standard metrics	Broader Hausa focus; this study provides stricter event-time separation but mostly English claim descriptions
Ogunrinde and Yusuf [6]	10,000 multilingual Nigerian articles; culturally aware BERT	91.3% accuracy; 89.7% F1	Much broader language coverage; task unit and validation design differ
Augenstein <i>et al.</i> [12]	MultiFC claims with evidence and metadata from 26 sites	49.2% best macro-F1	Harder evidence-aware multi-domain task; illustrates why text-only scores are not verification scores
This study	135 strict Nigerian claims; event-grouped development; untouched 2026 test	67.65% accuracy; 67.62% macro-F1	Lower headline score, but direct provenance and prospective local evaluation

Table 8 is not a ranking list, but rather a methodological comparison. Esan *et al.* [7] show the substantial accuracy gains from class balancing. However, the fact that the imbalanced model ignored the FAKE minority highlights the current interest in class-specific recall. Ogunrinde and Yusuf [9] have demonstrated better headline performance and significantly wider multilingual coverage than the current claim supplement. Their work demonstrates the power of language-adaptive transformers, while the present results highlight the unique need for event-disjoint prospective evaluation. Ibrahim *et al.* [8] works directly on Hausa. The present corpus is mainly characterised by English descriptions and lacks validation from the indigenous languages.

The distinction with MultiFC [11] makes interpretation even more difficult. MultiFC connects claims with evidence to back them up, reaching a macro-F1 score of 49.2%. The current context blend gets a macro-F1 score of 67.62% without evidence. The quantitative difference does not equate to superiority, as the tasks are not comparable; retrieving evidence and cross-source validation are more difficult than the prediction of a publisher's decision based on claim wording.

5.3 Contribution to Knowledge

The first contribution is on provenance. The Nigerian supplement adds fact-check URL, date, event category, and definitive verdict to each statement, enabling row-level auditing and replicable label mapping. This addresses the general problem raised in ecosystem studies that claims, judgements and publishing practices differ between organisations [13].

The second contribution tackles the problem of evaluation leakage. The temporal divisions and the discrete nature of the sets of events imply that claims made later and multi-claim reports are irrelevant to progress. The 2026 assessment thus evaluates the transfer to downstream fact-checking events more directly than a random article split. The third contribution is on evidence of adaptation. This experiment shows that local coverage and local predictive gain are different results. We test the global ensemble, direct augmentation, and an independent context head on the same test records. The fourth contribution is uncertainty-aware reporting: group bootstrap intervals and paired correctness analysis explicitly make the weaknesses of small samples apparent, rather than hiding them behind a single point estimate.

5.3 Practical Implications

The context model can be used as a queuing system by journalists, fact checkers and election monitoring groups. A safe interface would display the claim, source, time of entry, evaluation score, deviation of base models, and a space for abstention. Get original documents or trustworthy reports. Insist on a human judgement. The phrase predicted REAL should not be read as verified true. In the 2026 exam, the context blend continued to misclassify four FAKE statements as REAL, whereas the global ensemble misclassified six. False acceptance is therefore a known risk, not a theoretical one.

Thresholds for operational use need to be set depending on the reviewing capacity and the cost of any errors. In case where there is need for detection of more suspicious claims, the 0.7333 FAKE recall rate can be used since the context blend will provide this without too much concern for the increased false positives that result. In comparison, where the need is for balanced discrimination or ranking, then the global ensemble is the method to be used considering the existing data. The logistic regression method makes an ideal comparison due to its low cost.

5.4 Limitations and Validity Threats

The Nigerian prospective test contains only 34 claims, and multiple instances are taken from the same fact-check reports. Event-group bootstrap resampling captures pre-existing clustering but cannot generate missing diversity. The supplement depends primarily on a single fact-checking outfit. The language used by the publisher and the application of verdicts may still be associated with class. The claim descriptions are standardised summaries and are not guaranteed to be exact copies. These are not fake news articles but they may be employing editorial language that is not common on social media.

Publisher assessments are considered labels and were not subject to independent re-evaluation. Limiting the data to six unqualified verdict strings increases consistency at the cost of including nuanced cases. We are not entirely able to compare the various fact-checking organisations. The regional corpus is largely English and does not test for Nigerian Pidgin, Hausa, Yoruba, Igbo or code-switching capabilities. The rule-based classifications of Nigeria and its political categories might not identify implicit references or may include references that are merely incidental.

The 2026 assessment is prospective in terms of model development, but it covers a single calendar year and a limited

range of electoral events. The results obtained from the individual model observed in the test, in particular the superior balanced accuracy of Logistic Regression, are preliminary and need to be validated on a separate secured sample.

VI. CONCLUSION

This research developed and prospectively tested a context-aware screening model for Nigerian political and electoral assertions based on rigorous source-associated judgements. The Nigerian supplement is composed of 135 statements, of which 87 were recently obtained records while the global framework uses 29,269 validated claim descriptions. The selected context combination obtained a balanced accuracy of 0.6825, a macro-F1 of 0.6762 and a ROC-AUC of 0.7368 on 34 unchanged claims from 2026. It raised FAKE recall to 0.7333, but failed to outperform global ensemble on the same test, with paired uncertainty intervals included zero.

The research thus supports two inferences. Better Nigerian data improve coverage, auditability and reliability of assessments but do not automatically ensure predictive progress. The system should be used only to rank claims for human evaluation according to evidence. Future research could build a broader multi-publisher corpus of Nigerian English, keep the original claims verbatim, purposefully include Nigerian English and indigenous-language materials, assign complete events and publishers for subsequent assessment, and combine linguistic analysis with evidence gathering and clear adjudication.

ETHICAL CONSIDERATIONS

The research used publicly available secondary fact-checking data and did not recruit participants, interact with people or collect new sensitive personal information. Therefore, the computational experiment was not subject to ethical approval for human subjects or animals. Source URLs are preserved to allow for verification, and the model is restricted to assisting with research and oversight by humans.

CONFLICT OF INTEREST

The authors declare no known competing financial or personal interests that could have influenced this work.

REFERENCES

[1] I.Hassan, "Dissemination of disinformation on political and electoral processes in Nigeria: An exploratory study," *Cogent Arts & Humanities*, vol. 10, no. 1, Art. no. 2216983, 2023, doi: 10.1080/23311983.2023.2216983.

[2] A.A. Ibrahim and B. Ekdale, "Political campaign responses to information disorder: A case study of the 2023 Nigerian presidential elections," *The International Journal of Press/Politics*, vol. 31, no. 3, pp. 595–615, Jul. 2026, doi: 10.1177/19401612251324985.

[3] K. Garba, T. Kolajo, and J. B. Agbogun, "A transformer-based approach to Nigerian Pidgin text generation," *International Journal of Speech Technology*, vol. 27, pp. 1027–1037, 2024, doi: 10.1007/s10772-024-10136-2.

[4] A.A. Ibrahim, H. U. Ali, I. Z. Yakubu, and I. A. Lawal, "Fake news detection in Hausa language using transfer learning method," *International Journal of Innovative Science and Research Technology*, vol. 9, no. 10, pp. 2259–2269, Oct. 2024, doi: 10.38124/IJSRT/IJSRT24OCT1050.

[5] A.Esan, O. Abodunrin, A. Sobowale, I. Adeyanju, N. Okomba, B. Omodunbi, T. Adebisi, J. Jooda, T. Abdul-Hameed, and O. Asaolu, "Long-short-term memory model for fake news detection in Nigeria," *Ianna Journal of Interdisciplinary Studies*, vol. 5, no. 1, pp. 167–180, 2023. [Online]. Available: <https://iannajournalofinterdisciplinarystudies.com/index.php/1/article/view/149>

[6] M. A. Ogunrinde and Y. B. Yusuf, "A culturally aware NLP approach for fake-news detection in Nigerian online media using BERT," *African Scientific Reports*, vol. 5, no. 2, Art. no. 351, 2026, doi: 10.46481/asr.2026.5.2.351.

[7] J. Alghamdi, Y. Lin, and S. Luo, "A comparative study of machine learning and deep learning techniques for fake news detection," *Information*, vol. 13, no. 12, Art. no. 576, 2022, doi: 10.3390/info13120576.

[8] L. Hu, S. Wei, Z. Zhao, and B. Wu, "Deep learning for fake news detection: A comprehensive survey," *AI Open*, vol. 3, pp. 133–155, 2022, doi: 10.1016/j.aiopen.2022.09.001.

[9] H. T. Phan, N.-T. Nguyen, and D. Hwang, "Fake news detection: A survey of graph neural network methods," *Applied Soft Computing*, vol. 139, Art. no. 110235, 2023, doi: 10.1016/j.asoc.2023.110235.

[10] V. S. Pendyala and C. E. Hall, "Explaining misinformation detection using large language models," *Electronics*, vol. 13, no. 9, Art. no. 1673, 2024, doi: 10.3390/electronics13091673.

[11] V. C. Agbasonu, B. C. Amanze, and A. O. Agbakwuru, "Detecting online fake news using machine learning and topological data analysis," *International Journal of Engineering Applied Sciences and Technology*, vol. 7, no.

- 10, pp. 42–50, 2023, doi: 10.33564/IJEAST.2023.v07i10.006.
- [12] R. Deng and F. Duzhin, “Topological data analysis helps to improve accuracy of deep learning models for fake news detection trained on very small training sets,” *Big Data and Cognitive Computing*, vol. 6, no. 3, Art. no. 74, 2022, doi: 10.3390/bdcc6030074.
- [13] E. Altuncu, C. Başkent, S. Bhattacharjee, S. Li, and D. Roy, “FACTors: A new dataset for studying the fact-checking ecosystem,” in *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25), Padua, Italy, 2025*, pp. 3530–3539, doi: 10.1145/3726302.3730339.
- [14] C. Schröer, F. Kruse, and J. Marx Gómez, “A systematic literature review on applying CRISP-DM process model,” *Procedia Computer Science*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [15] N. Akewushola, F. Abeka, R. Olatunji, and F. Quadri, “2023 election: Fact-checking 12 claims made by Kwankwaso at Chatham House,” *FactCheckHub*, Feb. 3, 2023. [Online]. Available: <https://factcheckhub.com/2023-election-fact-checking-12-claims-made-by-kwankwaso-at-chatham-house/>. Accessed: Aug. 21, 2026.
- [16] N. Akewushola, F. Abeka, R. Olatunji, and F. Quadri, “2023 election: Fact-checking Peter Obi’s claims at Chatham House,” *FactCheckHub*, Jan. 20, 2023. [Online]. Available: <https://factcheckhub.com/2023-election-fact-checking-peter-obis-claims-at-chatham-house/>. Accessed: Aug. 26, 2026.
- [17] N. Akewushola, “Pre-election LIVE-CHECK for Ekiti 2026 governorship election,” *FactCheckHub*, Jun. 19, 2026. [Online]. Available: <https://factcheckhub.com/pre-election-live-check-for-ekiti-2026-governorship-election/>. Accessed: Aug. 26, 2026.
- [18] *FactCheckHub Editorial*, “LIVE-CHECK: Osun 2026 governorship election,” *FactCheckHub*, Aug. 15, 2026. [Online]. Available: <https://factcheckhub.com/live-check-osun-2026-governorship-election/>. Accessed: Aug. 20, 2026.
- [19] Z. Abdulrasaq and N. Akewushola, “Tinubu’s CNG bus promises: How has he fared?” *FactCheckHub*, Sep. 21, 2026. [Online]. Available: <https://factcheckhub.com/tinubus-cng-bus-promises-how-has-he-fared>. Accessed: Aug. 12, 2026.
- [20] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, “The balanced accuracy and its posterior distribution,” in *Proceedings of the 20th International Conference on Pattern Recognition, Istanbul, Turkey, 2010*, pp. 3121–3124, doi: 10.1109/ICPR.2010.764.
- [21] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.
- [22] I. Augenstein, C. Lioma, D. Wang, L. Chaves Lima, C. Hansen, C. Hansen, and J. G. Simonsen, “MultiFC: A real-world multi-domain dataset for evidence-based fact checking of claims,” in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 2019*, pp. 4685–4697, doi: 10.18653/v1/D19-1475.

Citation of this Article:

Ndudi-Wali Kinikanwo Wisdom, Prof. Daniel Matthais, & Dr. Victor T Emmah. (2026). Leakage-Aware Fake-News Screening for Nigerian Political Claims: Temporal and Domain Evaluation of a Context-Aware Ensemble. *Journal of Artificial Intelligence and Emerging Technologies (JAJET)*. 3(9), 57-67. Article DOI: <https://doi.org/10.47001/JAIET/2026.309007>

*** End of the Article ***