FUDMA Journal of Engineering and Technology
ISSN: 3092-9385
https://fjet.fudutsinma.edu.ng/index.php/fjet/
Volume 2, Issue 2, 2026, 71-79

Graph-Augmented Transformers for Hausa Sentiment Analysis: When Does Graph Structure Help?

Hafsa K. AHMAD

Department of Computer Science, Bayero University, Kano, Nigeria

hkahmad.cs@buk.edu.ng

Abstract

Hausa is still under-represented within natural language processing (NLP) research, despite having one of the largest speaker populations on the African continent. This study asks whether graph-augmented transformer architectures deliver measurable gains for sentiment analysis under those conditions. A hybrid model was developed and evaluated that combines a pre-trained transformer encoder with a Graph Attention Network (GAT) built over a document-word corpus graph, whose node features are obtained from Term Frequency–Inverse Document Frequency (TF-IDF) representations and fused via a learned scalar attention gate. Through systematic experiments across three transformer backbones, namely AfriBERTa (Hausa-inclusive pretraining), XLM-RoBERTa (general multilingual), and mDeBERTa-v3 (general multilingual), across four labeled data regimes (10%, 25%, 50%, 100%), the results show that the value of graph augmentation depends critically on the transformer's language-specific pretraining. When paired with AfriBERTa, the hybrid offers no significant benefit ($\Delta F1 = -0.003$), as language-specific pretraining already captures sufficient lexical and structural information. When paired with XLM-RoBERTa, the hybrid achieves a consistent gain of +0.017 Macro F1, with the advantage growing with data availability. The findings further demonstrate that TF-IDF node feature initialization is essential for enabling genuine transformer-GAT fusion, and that random initialization causes the attention gate to collapse to transformer-only behavior. These results suggest that for Hausa and other comparable low-resource languages, a well-adapted language-specific model is often enough, and that graph augmentation becomes truly useful as a complementary technique when only general multilingual transformers are available. Taken together, these findings offer concrete guidance for researchers developing NLP systems for Hausa and similar low-resource languages.

Keywords: *Low-resource languages, sentiment analysis, graph attention networks, transformer models, hybrid architectures, African languages.*

1.0 Introduction

Automatically detecting the sentiment expressed in text has advanced considerably since large pre-trained transformer models became widely available [1], [2]. That progress, however, has not reached all languages equally: English and Chinese continue to benefit from abundant pretraining corpora and mature benchmark suites, whereas most African languages still lack comparable resources [3], [4].

Hausa is spoken by approximately 200 million people worldwide, including both native (L1) speakers and second-language (L2) speakers [5], yet Hausa NLP remains in its early stages, with limited labeled data, few pretrained models, and a scarcity of systematic evaluations [6], [3]. Reliable sentiment analysis tools for the language could support several practical use cases, such as tracking public sentiment on social media, informing public health communication, and studying political discourse [5], [7].

One direction that has shown potential for improving low-resource performance is incorporating graph structure into the model, since corpus-level co-occurrence signals can supply information that sequence-only architectures tend to overlook, particularly when labeled examples are scarce [8], [9], [10]. Section 2.2 reviews this line of work in detail.

This paper presents an empirical study that investigates a practical but still underexplored question: Does graph augmentation actually help for Hausa sentiment analysis, and if so, under what conditions? A hybrid model was developed that combines a pretrained transformer encoder with a Graph Attention Network (GAT) built on a document-word graph. The two representations are fused using a learned scalar attention gate. Importantly, the GAT node features are initialized with TF-IDF vectors instead of random noise, giving the graph meaningful semantic information from the very beginning.

This study compares this hybrid architecture against strong transformer-only and GAT-only baselines using three different transformer backbones that vary in their exposure to Hausa during pretraining. Experiments are conducted across four levels of labeled data (10%, 25%, 50%, and 100%). The main findings are as follows:

1. Graph augmentation adds no meaningful benefit when paired with AfriBERTa, a model already pretrained on Hausa.

2. It provides consistent gains (+0.017 Macro F1) when paired with XLM-RoBERTa, a general multilingual model without Hausa-specific pretraining.

Received: 08/07/2026; Revised: 22/07/2026; Accepted: 27/07/2026; Published: 14/08/2026

3. The hybrid advantage becomes more pronounced as more labeled data becomes available, appearing clearly at 50% and full data, but remaining limited in extremely low-resource settings (10–25%).

4. Using TF-IDF for node feature initialization is crucial; random initialization causes the attention gate to collapse, effectively ignoring the graph component.

Overall, these results suggest that for African languages like Hausa, language-specific pretraining is usually sufficient when available, consistent with prior evidence that compact, language-adapted pretrained models can match or exceed larger general-purpose multilingual models in low-resource settings [6]. Graph augmentation appears most useful as a complementary technique when working with general multilingual models, consistent with prior evidence that graph-based augmentation yields consistent gains when layered on top of general multilingual pretrained encoders across text classification and multilingual disaster-response tasks [11], [12], and that knowledge-graph-informed adapters improve sentiment analysis performance for low-resource languages when paired with general multilingual large language models (LLMs) [13]. To the authors' knowledge, however, no prior study has tested whether this benefit persists when the backbone already has language-specific pretraining, which is the comparison addressed here.

2.0 Related Work

2.1 Hausa NLP

Among Africa's languages, Hausa has one of the largest speaker populations [5]. Despite this, Hausa Natural Language Processing has only recently begun to attract significant research interest and remains severely underserved relative to its speaker population [5]. Much of the progress so far traces back to grassroots collaborations, above all the Masakhane initiative, whose contributions span translation, entity recognition, and both document- and sentiment-level classification [3], [14], [5], and, more recently, text generation across hundreds of African languages [15]. Within this ecosystem, the Masakhane-Afrisenti effort has specifically targeted sentiment analysis for African languages using Afro-centric pretrained models and adapters [16].

Sentiment analysis stands out as one of the most actively studied tasks for the language. Early work by Abubakar et al. [17] relied on lexicon- and rule-based approaches for analyzing political tweets, delivering moderate results on relatively small datasets. A turning point was the release of NaijaSenti [7]: with roughly 30,000 annotated tweets, it was the first Hausa sentiment resource at scale, and it enabled substantially stronger supervised models, with multilingual transformers reaching weighted F1 scores near 0.81. Subsequent work explored shallow hybrid architectures that paired convolutional and recurrent neural networks (CNNs and RNNs) with hierarchical attention layers on top of Hausa-specific sentiment lexicons [18], while a separate line of work built manually curated lexicons for standalone lexicon-based systems [19].

More recently, research has shifted toward fine-tuning pretrained transformer models. Sani et al. [20] explored language-adaptive fine-tuning of AfriBERTa, and Zandam et al. [21] proposed HauBERT for aspect-based sentiment analysis of Hausa movie reviews. Despite this growing body of work, the majority of Hausa sentiment research continues to rely on standard transformer fine-tuning, traditional machine learning with hand-crafted features, or relatively shallow deep learning models. As far as the authors are aware, no earlier study has investigated hybrid transformer–graph architectures for Hausa sentiment analysis.

2.2 Graph-Based Text Classification

Graph neural networks (GNNs) offer a way to capture relationships and co-occurrence patterns spanning an entire corpus, something purely local, token-by-token processing cannot easily do, which makes them attractive for text classification. The seminal TextGCN model [8] pioneered this idea for text by constructing a single heterogeneous graph containing both document and word nodes: document-word connections carry TF-IDF weights, while word-word connections are weighted by pointwise mutual information (PMI). Document embeddings are then obtained through a Graph Convolutional Network (GCN) [22].

Later variants improved upon this foundation by replacing standard GCN layers with Graph Attention Networks (GAT) [23] or with inductive sampling-and-aggregation schemes such as GraphSAGE [24], letting the model assign different importance to each neighbor on the heterogeneous graph rather than treating them uniformly.

Recent survey work confirms that graph-based methods keep gaining ground in text classification [9]. Hybrid architectures that combine pretrained transformers with GNNs have gained traction in recent years. These models aim to merge the rich contextual representations of transformers with the corpus-level co-occurrence signals captured by graph structures [25], [10], [26]. Such hybrids have shown particular promise in low-resource or semi-supervised settings, where graph propagation can leverage unlabeled data effectively. Prior work has combined transformer and graph representations through strategies such as concatenation, cross-attention, or joint end-to-end training; the present study instead adopts a learned scalar attention gate that produces a single interpretable mixing weight between the two representations, as detailed in Section 3.5.

Received: 08/07/2026; Revised: 22/07/2026; Accepted: 27/07/2026; Published: 14/08/2026

The evidence base for hybrid transformer–GNN models, however, rests almost entirely on languages with abundant resources, English and Chinese in particular. Their applicability to low-resource African languages, which present unique challenges such as morphological complexity, code-switching, and limited pretraining data, remains largely unexamined.

2.3 Low-Resource NLP

Progress on low-resource NLP has largely come from three directions: pretraining on many languages jointly, transferring knowledge across languages, and designing techniques that make efficient use of limited data [27], [3]. Within the African-language space, compact models such as AfriBERTa [6], pretrained on 11 African languages including Hausa, have demonstrated strong downstream performance despite modest parameter counts. Larger multilingual models such as XLM-RoBERTa [27] and mDeBERTa-v3 [28] serve as competitive baselines when language-specific pretraining is unavailable.

Graph-based methods are particularly attractive in low-resource scenarios because the document-word graph can be constructed from the entire (labeled + unlabeled) corpus, providing inductive bias from global statistics. Nevertheless, the interaction between language-specific pretraining quality and graph augmentation has received little attention. It remains unclear whether graph structures offer complementary value when strong language-specific transformers are already available, or whether they primarily benefit general multilingual backbones.

2.4 Research Gap and Contributions

Although Hausa sentiment analysis has progressed through transformer fine-tuning and lexicon-based methods, and graph-augmented models have shown value in high-resource settings, a critical gap persists: no prior work has systematically investigated whether graph neural networks provide complementary benefits when combined with pretrained transformers for Hausa language. Specifically, the field lacks understanding of when graph augmentation helps, specifically, whether its utility depends on the presence of language-specific pretraining, the amount of labeled data, or the architectural characteristics of the transformer backbone.

To address this gap, the present study makes the following contributions:

1. A document-word heterogeneous graph is constructed for Hausa sentiment analysis using TF-IDF and PMI edges.

2. A hybrid architecture is proposed that fuses pretrained transformer representations with GAT document embeddings via a learned scalar attention gate.

3. A controlled evaluation is conducted across three transformer backbones (AfriBERTa, XLM-RoBERTa, mDeBERTa-v3) with varying degrees of Hausa-specific pretraining and across four labeled data regimes (10%, 25%, 50%, 100%).

4. The critical role of TF-IDF node feature initialization and the dynamics of the attention fusion gate are analyzed.

This work provides practical guidance for when graph augmentation is worthwhile in low-resource African language settings.

3.0 Methodology

3.1 Document-Word Graph Construction

Following TextGCN, a heterogeneous graph is constructed $G = (V, E)$ over the full corpus (train + validation + test). The node set V contains one node per document and one node per vocabulary word (minimum frequency 2), for a total of $|V| = N_{\text{docs}} + N_{\text{words}}$ nodes. The edge set E contains two types of edges:

Each document-word pair receives an edge weighted by its TF-IDF score. Concretely, for a document d and a word w that co-occur, an undirected edge is created whose weight equals the TF-IDF score of w within d. Separately, word-to-word edges are assigned pointwise mutual information (PMI) scores calculated over a size-10 sliding window within each document, and only pairs with a positive PMI value are kept in the graph.

3.2 Node Feature Initialization

A critical contribution of this work is the initialization of node features. Rather than the identity-based node initialization adopted in TextGCN, node features are initialized using TF-IDF representations projected to a compact dimension via a fixed random Johnson–Lindenstrauss projection, which preserves approximate pairwise distances during dimensionality reduction [29]. Document nodes receive their TF-IDF document vector projected from vocabulary size to 128 dimensions. Word nodes receive the transpose TF-IDF column vector (capturing each word's distribution across documents) projected to 128 dimensions. This initialization preserves approximate cosine similarity between documents and words before the GAT has processed any information, providing the GNN with meaningful semantic signal from the start.

Received: 08/07/2026; Revised: 22/07/2026; Accepted: 27/07/2026; Published: 14/08/2026

Empirical results demonstrate that this initialization is critical: with random node features, the learned attention gate collapses to $\alpha \approx 1.0$ (transformer-only behavior) from the first epoch, while TF-IDF initialization enables genuine fusion with α stabilizing in the range 0.3–0.5.

3.3 Graph Attention Network Encoder

A two-layer Graph Attention Network operates over the constructed graph (8 attention heads per layer; hidden dimension 32; output dimension 128), producing one embedding $\boldsymbol{h}_g \in \mathbb{R}^{128}$ for each document node. During hybrid training, the GAT weights are frozen (train_gat=False) and document embeddings are pre-computed once per epoch, significantly reducing computational cost.

3.4 Transformer Encoder

Three transformer backbones are evaluated: AfriBERTa-small (castorini/afriberta_small), XLM-RoBERTa-base, and mDeBERTa-v3-base. For each model, the classification token ([CLS]) embedding is extracted from the final hidden layer as the document representation $\boldsymbol{h}_t$. The bottom 10 layers of each transformer are frozen during fine-tuning, following standard practice for low-resource settings. All three models have a hidden dimension of 768.

3.5 Attention Fusion Gate

The transformer representation $\boldsymbol{h}_t \in \mathbb{R}^{768}$ and the GAT representation $\boldsymbol{h}_g \in \mathbb{R}^{128}$ are fused via a gated linear combination mechanism, specifically a scalar attention gate. The two hidden vectors are first concatenated and passed through a linear layer with a sigmoid activation function to calculate the gating scalar α, as defined in Equation (1):

$$\alpha = \sigma(\boldsymbol{W}_\alpha[\boldsymbol{h}_t ; \boldsymbol{h}_g] + b_\alpha) \tag{1}$$

The fused representation is then computed as shown in Equation (2):

$$\boldsymbol{h}_{fused} = \alpha \cdot \text{proj}(\boldsymbol{h}_t) + (1 - \alpha) \cdot \boldsymbol{h}_g \tag{2}$$

where proj(·) is a linear projection from 768 to 128 dimensions, and σ is the sigmoid function. The scalar $\alpha \in (0,1)$ controls the relative contribution of the transformer and GAT representations. A value of α near 1 indicates transformer-dominant fusion; a value near 0 indicates GAT-dominant fusion. To prevent the gate from collapsing to $\alpha = 1$ (transformer-only), a penalty term is added to the training objective as given in Equation (3):

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda \cdot \mathcal{F}[\alpha], \ \lambda = 0.05 \tag{3}$$

This penalty is applied after a 2-epoch warmup to allow the transformer to stabilize before the regularization introduces pressure toward lower α. The final fused representation $\boldsymbol{h}_{fused}$ feeds into a two-layer MLP classification head with dropout ($p = 0.3$).

3.6 Training Setup

All models are trained with AdamW (weight decay 0.01). The transformer encoder uses learning rate 2e-5; all other components use 1e-3. Gradient clipping is applied at 1.0. Training stops early once validation Macro F1 fails to improve for 5 consecutive epochs, and the best-performing checkpoint is kept for test evaluation. Batches contain 16 examples. All experiments are run on Apple Metal Performance Shaders (MPS) hardware.

4.0 Experimental Setup

4.1 Dataset

This study draws its Hausa sentiment data from the AfriSenti benchmark [30], a large multilingual sentiment collection for African languages built mainly from social media posts. Once neutral-labeled examples are discarded, the Hausa portion consists of 9,260 training samples (4,573 negative and 4,687 positive), 1,781 validation samples, and 3,514 test samples — 14,555 documents in total feeding into the graph construction step. Only two sentiment labels are used, positive and negative. Before training, each text is lowercased, stripped of URLs and user mentions, and cleaned of extra whitespace.

Received: 08/07/2026; Revised: 22/07/2026; Accepted: 27/07/2026; Published: 14/08/2026

4.2 Models

Seven models are evaluated in total, spanning transformer-only baselines, graph-only baselines, and hybrid combinations:

i. AfriBERTa-only. A fine-tuned AfriBERTa-small transformer used as a strong language-specific baseline. AfriBERTa was pretrained on 11 African languages including Hausa, making it the most linguistically tailored backbone in this study. Its [CLS] token embedding feeds straight into a two-layer MLP classifier, with no graph component involved.

ii. AfriBERTa + GAT (Hybrid). The hybrid model pairing AfriBERTa with the document-word GAT encoder. The transformer and GAT representations are fused via the learned scalar attention gate described in Equations (1) and (2). This configuration tests whether graph-level corpus structure adds complementary signal on top of strong language-specific pretraining.

iii. XLM-R-only. A fine-tuned XLM-RoBERTa-base transformer, chosen as a broad-coverage multilingual model trained across 100 languages with no dedicated Hausa data. It acts as the main multilingual point of comparison for measuring hybrid gains.

iv. XLM-R + GAT (Hybrid). The hybrid model pairing XLM-RoBERTa with the GAT encoder, fused via the attention gate. This configuration is the principal test of the study's hypothesis: that graph augmentation provides the most benefit when the transformer lacks Hausa-specific pretraining.

v. mDeBERTa-only. A fine-tuned mDeBERTa-v3-base transformer, an mBERT variant built around a disentangled attention mechanism that keeps token content and relative position separate. This baseline probes whether architectural sophistication alone can substitute for language-specific pretraining.

vi. mDeBERTa + GAT (Hybrid). This variant combines mDeBERTa-v3 with the GAT encoder. It enables a test of whether the graph-based component can provide additional value on top of mDeBERTa's already sophisticated disentangled attention mechanism.

vii. GAT-only. This is a standalone two-layer Graph Attention Network that operates directly on the document-word graph, using TF-IDF vectors as node features. It serves as a lower-bound baseline, indicating how much the graph structure alone can achieve without any contextual knowledge from a pretrained transformer.

4.3 Low-Resource Evaluation

To evaluate model performance under data-scarce conditions, smaller training subsets were created by randomly sampling 10%, 25%, and 50% of the original training data, using stratified sampling so that class balance is preserved in each subset. Every experiment was repeated with three separate random seeds, and the mean Macro F1 score is reported along with the standard deviation. The validation and test sets were kept at their full original size across all experiments.

4.4 Evaluation Metric

Macro F1 is used as the sole evaluation metric. It is computed as the unweighted average of the F1 scores for the positive and negative sentiment classes, thereby assigning equal importance to both classes. This choice is particularly appropriate for the present binary sentiment classification task because it provides a class-balanced assessment of performance and prevents the performance of one sentiment class from being obscured by the other. Furthermore, using Macro F1 maintains comparability with prior sentiment analysis research in Hausa and other African languages, including NaijaSenti [7] and AfriSenti [30]. Given the near-balanced distribution of positive and negative samples in the dataset, accuracy and support-weighted F1 would be expected to provide broadly similar overall assessments; therefore, reporting them in addition to Macro F1 would provide limited additional information for the primary comparison in this study.

5.0 Results

5.1 Full-Data Comparison

Table 1 presents results on the full training set. AfriBERTa-only achieves the highest performance (0.8933), confirming that language-specific pretraining is highly effective for Hausa sentiment analysis. The GAT-only baseline (0.7251) demonstrates that graph structure alone, without strong semantic representations, is insufficient for this task.

Table 1: Full-data test Macro F1 results (Δ shows gain of hybrid over corresponding transformer-only baseline)

Model	Macro F1	Δ vs Baseline
AfriBERTa-only	0.8933	N/A
AfriBERTa + GAT (Hybrid)	0.8907	−0.0026
XLM-R-only	0.8527	N/A
XLM-R + GAT (Hybrid)	0.8697	+0.0170

Received: 08/07/2026; Revised: 22/07/2026; Accepted: 27/07/2026; Published: 14/08/2026

Model	Macro F1	Δ vs Baseline
mDeBERTa-only	0.8694	N/A
mDeBERTa + GAT (Hybrid)	0.8674	−0.0020
GAT-only	0.7251	N/A

Note: "N/A" indicates the delta column is not applicable, since transformer-only and GAT-only rows are baselines with no hybrid counterpart to compare against.

The key finding is that the value of graph augmentation depends strongly on the transformer backbone. The AfriBERTa hybrid shows a negligible difference (ΔF1 = −0.003), while the XLM-R hybrid achieves a meaningful gain of +0.017. The mDeBERTa hybrid shows neither benefit nor harm. This pattern is consistent with the study's hypothesis: graph augmentation compensates for weak language-specific pretraining.

Notably, the mean attention gate α for the AfriBERTa hybrid stabilized at 0.384, while the XLM-R hybrid stabilized at 0.542. This indicates that the XLM-R hybrid achieves a more balanced fusion, with the GAT contributing approximately equally to the transformer, whereas the AfriBERTa hybrid is more transformer-dominant.

5.2 Low-Resource Results

The low-resource results for AfriBERTa and XLM-R appear in Tables 2 and 3, respectively.

Table 2: Low-resource results for AfriBERTa (Mean Macro F1 ± std across 3 seeds)

Data Fraction	N Samples	Hybrid F1 (mean±std)	Transformer F1	Δ (Hybrid Gain)
10%	926	0.8376 ± 0.0040	0.8412	−0.0036
25%	2,315	0.8524 ± 0.0011	0.8588	−0.0065
50%	4,630	0.8721 ± 0.0060	0.8740	−0.0020
100%	9,260	0.8907	0.8933	−0.0026

Table 3: Low-resource results for XLM-R (Mean Macro F1 ± std across 3 seeds)

Data Fraction	N Samples	Hybrid F1 (mean±std)	Transformer F1	Δ (Hybrid Gain)
10%	926	0.7811 ± 0.0065	0.7853	−0.0042
25%	2,315	0.7915 ± 0.0081	0.8141	−0.0226
50%	4,630	0.8418 ± 0.0046	0.8400	+0.0018
100%	9,260	0.8697	0.8527	+0.0170

For AfriBERTa, the hybrid consistently underperforms the transformer-only baseline across all data fractions, with small but stable negative deltas (std ≤ 0.006). This confirms that AfriBERTa's Hausa-specific pretraining is robust even at 10% labeled data, and graph augmentation provides no complementary benefit regardless of data availability.

For XLM-R, the pattern is more nuanced. At 10% and 25% data, the hybrid underperforms, with the largest negative delta occurring at 25% (−0.0226). However, the gap closes at 50% (+0.0018) and becomes a clear positive at full data (+0.0170). This points to the hybrid needing a certain volume of labeled data before it can learn a useful fusion. When labels are very scarce, the transformer's own task-specific representations may simply be too weak to give the gate a reliable signal to weight against.

5.3 Analysis of Attention Gate

The behavior of the attention gate α provides insight into the fusion dynamics. With random node features, α collapsed to approximately 1.0 (pure transformer) from the first epoch regardless of the transformer used, indicating that the GAT produced no useful signal. With TF-IDF node features, α stabilized in the range 0.25–0.55 across epochs, with values differing between models: AfriBERTa hybrid ($\alpha \approx 0.38$) versus XLM-R hybrid ($\alpha \approx 0.54$). Recalling that higher α indicates greater reliance on the transformer component (Equation 2), the AfriBERTa hybrid's lower α reflects the model assigning a relatively larger share of the fusion to the GAT, a pattern consistent with AfriBERTa's stronger Hausa-specific representations providing a more stable base from which the gate can afford to draw on the graph signal. By contrast, the XLM-R hybrid's higher α suggests that, despite the regularisation pressure encouraging lower values, the gate has not fully learned to exploit the GAT signal when paired with a weaker transformer backbone. This may partly explain why the hybrid advantage over XLM-R only emerges clearly at higher data fractions: sufficient labeled data appears necessary before the model can learn a fusion weighting that meaningfully leverages the GAT's corpus-level representations.

Received: 08/07/2026; Revised: 22/07/2026; Accepted: 27/07/2026; Published: 14/08/2026

6.0 Discussion

6.1 When Does Graph Augmentation Help?

The results paint a nuanced picture. Graph augmentation is not universally beneficial; its value depends on the interplay between the transformer's language-specific knowledge and the GAT's corpus-level co-occurrence signal. When the transformer has strong Hausa-specific pretraining (AfriBERTa), it already implicitly captures lexical co-occurrence patterns through its attention mechanisms, making the explicit graph structure redundant. When the transformer is a general multilingual model (XLM-R), the graph provides complementary information that the transformer lacks.

The mDeBERTa result is instructive. Despite having no Hausa-specific pretraining, mDeBERTa-only (0.8694) performs nearly as well as AfriBERTa-only (0.8933), and the hybrid provides no benefit. This suggests that DeBERTa's disentangled attention mechanism [28], which explicitly models relative position and content separately, may already capture the co-occurrence structure that the graph would otherwise provide. Architecture matters as much as pretraining language coverage.

6.2 The Role of Node Feature Quality

One of the most important methodological findings of this study is the critical importance of node feature initialization. Using TF-IDF vectors for GAT node features is essential for successful transformer-GAT fusion. When random initialization was used instead, the attention gate quickly learned to ignore the graph component entirely, regardless of regularization strength, effectively reducing the model to a transformer-only system.

This has clear practical implications: many existing TextGCN-style hybrid implementations that rely on random or learned-from-scratch node features may be significantly underperforming. Simply reusing the same TF-IDF representations that define the graph edges turns out to be a simple yet highly effective initialization strategy.

6.3 Practical Recommendations

For researchers and practitioners working on Hausa sentiment analysis or similar low-resource African languages, these results offer the following practical guidance: when a strong language-specific pretrained model (such as AfriBERTa for Hausa) is available, it is usually sufficient on its own, even with as little as 10% of the labeled data. Additionally, when only general multilingual models are available, adding graph augmentation can deliver meaningful gains. However, the hybrid approach generally needs a reasonable amount of labeled data (at least 50% of the training set) before it consistently outperforms the transformer baseline. Moreover, in extremely low-resource settings (less than 25% labeled data), the simpler transformer-only approach remains preferable, as the hybrid does not show a clear advantage.

6.4 Limitations

This study has several limitations. First, all experiments were conducted on a single dataset (AfriSenti Hausa). It is not yet known whether these results carry over to other Hausa tasks, such as named entity recognition or topic classification, or to other African languages. Second, the GAT weights were frozen during hybrid training for computational efficiency. Full end-to-end joint training might produce different results. Third, even in the low-resource experiments, the document-word graph was built using the entire corpus, which represents an optimistic scenario for the hybrid model. Finally, the statistical significance testing is limited by the relatively small number of random seeds (n = 3). Larger-scale evaluations would strengthen the conclusions.

7.0 Conclusion

This work presents an investigation of graph-augmented transformers for Hausa sentiment analysis, examining how the benefit of adding graph structure varies across different transformer backbones and different amounts of labeled data.

The central finding is that graph augmentation provides meaningful improvements (+0.017 Macro F1) when paired with general multilingual models like XLM-RoBERTa that lack Hausa-specific pretraining. However, it offers little to no benefit when combined with strong language-specific models (AfriBERTa) or architecturally advanced multilingual models (mDeBERTa). Results also showed that proper TF-IDF initialization of node features is crucial for effective fusion and that the hybrid model requires sufficient labeled data before its advantage becomes evident.

These findings contribute to a broader understanding of when architectural complexity adds real value in low-resource NLP. For the Hausa NLP community, these results highlight the power of language-specific pretraining while offering clear guidance on when graph-based methods can serve as a useful complementary tool. Future work should explore whether these patterns hold across other African languages, different NLP tasks, and alternative graph construction strategies.

Received: 08/07/2026; Revised: 22/07/2026; Accepted: 27/07/2026; Published: 14/08/2026

References

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. 2019 Conf. North American Chapter Assoc. Comput. Linguistics: Human Language Technologies (NAACL-HLT), vol. 1, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.

[2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017, pp. 5998–6008.

[3] W. Nekoto, V. Marivate, T. Matsila, T. Fasubaa, T. Fagbohungbe, S. O. Akinola, S. H. Muhammad, S. Kabongo Kabenamualu, S. Osei, F. Sackey, A. Niyongabo Rubungo, J. Mukiibi, A. Mnyakeni, C. Mbohwa, and J. Abbott, "Participatory research for low-resourced machine translation: A case study in African languages," in Findings of the Assoc. Comput. Linguistics: EMNLP 2020, 2020, pp. 2144–2160, doi: 10.18653/v1/2020.findings-emnlp.195.

[4] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, et al., "On the opportunities and risks of foundation models," arXiv preprint arXiv:2108.07258, 2021.

[5] S. H. Muhammad, I. S. Ahmad, I. Abdulmumin, F. I. Lawan, S. H. Imam, Y. Aliyu, S. A. Sani, A. U. Umar, T. Gwadabe, K. Church, and V. Marivate, "HausaNLP: Current status, challenges and future directions for Hausa natural language processing," in Proc. 6th Workshop on African Natural Language Processing (AfricaNLP 2025), 2025, pp. 176–191, doi: 10.18653/v1/2025.africanlp-1.17.

[6] K. Ogueji, Y. Zhu, and J. Lin, "Small data? No problem! Exploring the viability of pretrained multilingual language models for low-resourced languages," in Proc. 1st Workshop on Multilingual Representation Learning, 2021, pp. 116–126, doi: 10.18653/v1/2021.mrl-1.11.

[7] S. H. Muhammad, D. I. Adelani, S. Ruder, I. S. Ahmad, I. Abdulmumin, B. S. Bello, M. Choudhury, C. C. Emezue, S. S. Abdullahi, A. Aremu, A. Jorge, and P. Brazdil, "NaijaSenti: A Nigerian Twitter sentiment corpus for multilingual sentiment analysis," in Proc. 13th Language Resources and Evaluation Conf. (LREC), 2022, pp. 590–602.

[8] L. Yao, C. Mao, and Y. Luo, "Graph convolutional networks for text classification," in Proc. AAAI Conf. Artif. Intell., vol. 33, no. 1, pp. 7370–7377, 2019, doi: 10.1609/aaai.v33i01.33017370.

[9] K. Wang, Y. Ding, and S. C. Han, "Graph neural networks for text classification: A survey," Artif. Intell. Rev., vol. 57, Art. no. 190, 2024, doi: 10.1007/s10462-024-10808-0.

[10] E. L. S. Perin, M. C. de Souza, J. A. Silva, and E. T. Matsubara, "DynGraph-BERT: Combining BERT and GNN using dynamic graphs for inductive semi-supervised text classification," Informatics, vol. 12, no. 1, Art. no. 20, 2025, doi: 10.3390/informatics12010020.

[11] Z. Wang, X. Liu, P. Yang, S. Liu, and Z. Wang, "Cross-lingual text classification with heterogeneous graph neural network," in Proc. 59th Annu. Meeting Assoc. Comput. Linguistics and 11th Int. Joint Conf. Natural Language Processing (ACL-IJCNLP), 2021, pp. 612–620, doi: 10.18653/v1/2021.acl-short.78.

[12] S. Ghosh, S. Maji, and M. S. Desarkar, "GNoM: Graph neural network enhanced language models for disaster related multilingual text classification," in Proc. 14th ACM Web Science Conf. (WebSci), 2022, pp. 55–65, doi: 10.1145/3501247.3531561.

[13] D. Gurgurov, M. Hartmann, and S. Ostermann, "Adapting multilingual LLMs to low-resource languages with knowledge graphs via adapters," in Proc. 1st Workshop on Knowledge Graphs and Large Language Models (KaLLM), 2024, pp. 63–74, doi: 10.18653/v1/2024.kallm-1.7.

[14] R. Y. Zakari, Z. K. Lawal, and I. Abdulmumin, "A systematic literature review of Hausa natural language processing," Int. J. Comput. Inf. Technol., vol. 10, no. 4, pp. 173–179, 2021, doi: 10.24203/ijcit.v10i4.86.

[15] I. Adebara, A. Elmadany, and M. Abdul-Mageed, "Cheetah: Natural language generation for 517 African languages," arXiv preprint arXiv:2401.01053, 2024.

[16] I. A. Azime, S. S. Al-Azzawi, A. L. Tonja, I. Shode, J. Alabi, A. Awokoya, M. Oduwole, T. Adewumi, S. Fanijo, O. Awosan, and O. Yousuf, "Masakhane-Afrisenti at SemEval-2023 Task 12: Sentiment analysis using Afro-centric language models and adapters for low-resource African languages," in Proc. 17th Int. Workshop on Semantic Evaluation (SemEval-2023), 2023, pp. 1311–1316, doi: 10.18653/v1/2023.semeval-1.182.

[17] A. I. Abubakar, A. Roko, A. M. Bui, and I. Saidu, "An enhanced feature acquisition for sentiment analysis of English and Hausa tweets," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 9, pp. 102–110, 2021, doi: 10.14569/IJACSA.2021.0120913.

[18] H. A. Shehu, K. U. Majikumna, A. B. Suleiman, S. Luka, M. H. Sharif, R. A. Ramadan, and H. Kusetogullari, "Unveiling sentiments: A deep dive into sentiment analysis for low-resource languages—A case study on Hausa texts," IEEE Access, vol. 12, pp. 177437–177455, 2024, doi: 10.1109/ACCESS.2024.3502963.

Received: 08/07/2026; Revised: 22/07/2026; Accepted: 27/07/2026; Published: 14/08/2026

[19] I. Mohammed and R. Prasad, "Lexicon dataset for the Hausa language," Data in Brief, vol. 54, Art. no. 110124, 2024, doi: 10.1016/j.dib.2024.110124.

[20] S. A. Sani, S. Abubakar, F. I. Lawan, A. Abubakar, and M. Bala, "Investigating the impact of language-adaptive fine-tuning on sentiment analysis in Hausa language using AfriBERTa," arXiv preprint arXiv:2501.11023, 2025.

[21] A. Y. Zandam, F. M. Adam, A. Musa, and U. Ibrahim, "HauBERT: A transformer model for aspect-based sentiment analysis of Hausa-language movie reviews," Eng. Proc., vol. 87, no. 1, Art. no. 43, 2025, doi: 10.3390/engproc2025087043.

[22] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Int. Conf. Learning Representations (ICLR), 2017.

[23] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in Int. Conf. Learning Representations (ICLR), 2018.

[24] W. L. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.

[25] Y. Zhang, X. Yu, Z. Cui, S. Wu, Z. Wen, and L. Wang, "Every document owns its structure: Inductive text classification via graph neural networks," in Proc. 58th Annu. Meeting Assoc. Comput. Linguistics, 2020, pp. 334–339, doi: 10.18653/v1/2020.acl-main.31.

[26] [26] B. Liu, W. Guan, C. Yang, Z. Fang, and Z. Lu, "Transformer and graph convolutional network for text classification," Int. J. Comput. Intell. Syst., vol. 16, no. 1, Art. no. 161, 2023, doi: 10.1007/s44196-023-00337-z.

[27] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, et al., "Unsupervised cross-lingual representation learning at scale," in Proc. 58th Annu. Meeting Assoc. Comput. Linguistics, 2020, pp. 8440–8451, doi: 10.18653/v1/2020.acl-main.747.

[28] P. He, X. Liu, J. Gao, and W. Chen, "DeBERTa: Decoding-enhanced BERT with disentangled attention," in Int. Conf. Learning Representations (ICLR), 2021.

[29] D. Achlioptas, "Database-friendly random projections: Johnson–Lindenstrauss with binary coins," J. Comput. Syst. Sci., vol. 66, no. 4, pp. 671–687, 2003, doi: 10.1016/S0022-0000(03)00025-4.

[30] S. H. Muhammad, I. Abdulmumin, S. M. Yimam, D. I. Adelani, C. C. Emezue, C. Chukwuneke, E. Munkoh-Buabeng, D. Gebreyohannes, A. B. Tilaye, A. Niyongabo Rubungo, J. Mukiibi, J. Nakatumba-Nabende, M. A. Fale, B. Sibanda, A. Bukula, A. Mnyakeni, J. Nakatumba, V. Marivate, and I. Orife, "AfriSenti: A Twitter sentiment analysis benchmark for African languages," in Proc. 2023 Conf. Empirical Methods in Natural Language Processing (EMNLP), H. Bouamor, J. Pino, and K. Bali, Eds., 2023, pp. 12276–12301, doi: 10.18653/v1/2023.emnlp-main.862.

Received: 08/07/2026; Revised: 22/07/2026; Accepted: 27/07/2026; Published: 14/08/2026