Jurnal Masyarakat Informatika, 17(1), 2025
DOI:10.14710/jmasif.17.1.81053
ISSN: 2777-0648

Parameter-Efficient Few-Shot Sentiment Analysis Using LoRA-Enhanced Transformers

Nurudeen Jibrin *1), Gilbert Aimufua2), Okorie Sunday Onyedikachi3), Alegbe Adesola Anthony2), Ugbai Solomon Chukwunwike2), Fadila Dantalle Aliyu2)

1)Department of Computer Science and Information Technology, Baze University, Abuja, Nigeria
2)Faculty of Natural and Applied Science, Nasarawa State University, Keffi, Nigeria
3)Economic Community of West African States, Abuja, Nigeria
4)Department of Computer Science and Engineering, GD Goenka University, Gurugram, India
** Corresponding author: nurudeen.jibrin@bazeuniversity.edu.ng*

Abstract

Sentiment analysis in low-resource languages is often limited by scarce annotated data and the high computational cost of fine-tuning large language models. This study proposes a parameter-efficient framework that integrates Low-Rank Adaptation (LoRA) with lightweight transformer architectures, including AfriBERTa, DistilBERT, and MiniLMv2, for Hausa sentiment analysis using the NaijaSenti dataset. The framework is designed to address three key challenges: effective few-shot learning, robustness under extreme data scarcity, and mitigation of language-specific linguistic errors. Experimental results demonstrate that AfriBERTa-LoRA achieves 69.0% accuracy, only 4.8 percentage points below a fully fine-tuned XLM-RoBERTa baseline, while utilizing just 1.06% of trainable parameters and reducing GPU memory consumption by approximately 50%. Performance improves consistently with increasing data, indicating strong scalability under few-shot conditions. Linguistic error analysis reveals four dominant Hausa-specific failure modes accounting for 71.5% of misclassifications. Targeted mitigation strategies yield an 8.7 percentage point reduction in error rate (28% relative reduction, $p < 0.01$), with each individual strategy demonstrating statistical significance. These findings establish LoRA as an effective and efficient paradigm for low-resource natural language processing, providing a scalable and reproducible framework for sentiment analysis in underrepresented African languages.

Keywords: *Low-Rank Adaptation, Sentiment Analysis, Hausa Language, Natural Language Processing, Few-Shot Learning*

1 Introduction

Sentiment analysis, a core task in NLP, automates the extraction of opinions and emotional tone from text, yielding critical insights for applications in public opinion monitoring, consumer feedback, and socio-political discourse [1], [2]. Recent advances driven by large pre-trained language models have pushed performance benchmarks above 90% F1-score in high-resource languages such as English [3]. However, these gains remain largely inaccessible to low-resource languages due to severe data scarcity, linguistic complexity, and prohibitive computational demands [4].

Hausa, a major Chadic language spoken by over 80 million people primarily in Nigeria and Niger, exemplifies this disparity. Notwithstanding its growing digital presence on platforms like X, Hausa remains severely underrepresented in NLP research [5], [6]. Existing sentiment classifiers for Hausa typically achieve Macro-F1 scores below 70%, far trailing high-resource benchmarks [7], [8]. This performance gap stems from Hausa's unique linguistic properties, including tonal morphology, discontinuous negation structures, dialectal variation, and frequent code-mixing with English and Arabic, which challenge standard multilingual models [9]. The recent release of NaijaSenti, a large-scale corpus of 30,000 annotated Hausa tweets capturing real-world variation, provides a valuable resource but has yet to be fully exploited under constrained settings [10], [11].

Received: 28 December 2025 ; Revised: 01 March 2026; Accepted: 06 April 2026; Published: 13 April 2026

Parameter-Efficient Fine-Tuning (PEFT) methods [12], particularly LoRA [13], have emerged as promising solutions for low-resource scenarios by drastically reducing trainable parameters, up to 99% while preserving performance [14]. Extensions such as LoRA+ and adaptive variants have further improved efficiency [15]. Combined with few-shot learning paradigms, PEFT techniques show strong potential in data-scarce domains [16]. While these methods have demonstrated impressive results on European and Asian languages, their evaluation remains limited for African languages with distinct morphological and syntactic structures, leaving open questions about their effectiveness in such contexts.

However, their application to African languages remains limited, with few studies exploring LoRA-based sentiment analysis for Hausa or similar morphologically rich, low-resource languages. Existing Hausa sentiment models predominantly rely on resource-intensive full fine-tuning or on smaller architectures, without systematic efficiency analysis or linguistically targeted adaptations. This study addresses these gaps by introducing a parameter-efficient, few-shot sentiment analysis framework specifically tailored for Hausa. We apply LoRA to lightweight transformer backbones (DistilBERT, AfriBERTa, and MiniLMv2) and evaluate their performance on the NaijaSenti Hausa subset under stratified few-shot conditions with 100-1,000 training samples. The primary contributions are threefold:

1. A comprehensive benchmarking of LoRA against full fine-tuning baselines in terms of accuracy, data-efficiency, and computational cost,
2. A detailed linguistic error analysis identifying Hausa-specific failure modes,
3. Targeted mitigation strategies that significantly reduce error rates.

This work thus provides one of the first systematic explorations of parameter-efficient few-shot learning for Hausa sentiment analysis, offering not only performance insights but also a replicable methodological framework for advancing NLP in underrepresented African languages. To the best of our knowledge, this study is among the first systematic investigations of LoRA-based few-shot sentiment analysis for Hausa, combining parameter efficiency, statistical validation, and linguistically informed error mitigation.

2 Research Methods

2.1 Experimental Framework

The study employs a five-phase pipeline as shown in Figure 1. Integrating parameter-efficient fine-tuning via LoRA, stratified few-shot evaluation, and linguistically-informed error analysis to achieve robust Hausa sentiment analysis under data and computational constraints. All experiments were conducted with five different random seeds: 42, 123, 456, 789, and 2024 to ensure statistical robustness, with results reported as mean ± standard deviation throughout.

2.2 Dataset Description and Pre-processing

Experiments utilize the Hausa subset of NaijaSenti [7], comprising 30,000 manually annotated tweets sourced from X from January 2020 until December 2022 via the Twitter API v2. The dataset is nearly balanced: 10,142 positive (33.8%), 10,103 negative (33.7%), and 9,755 neutral (32.5%), with an average tweet length of 18.4 tokens (SD=7.2). A fixed test set of 5,303 tweets was held out. For few-shot experiments, stratified random sampling from the remaining 24,697 tweets generated training subsets of 100, 500, and 1,000 instances, preserving class proportions. Table 1 presents the complete data distribution scheme across all experimental conditions.

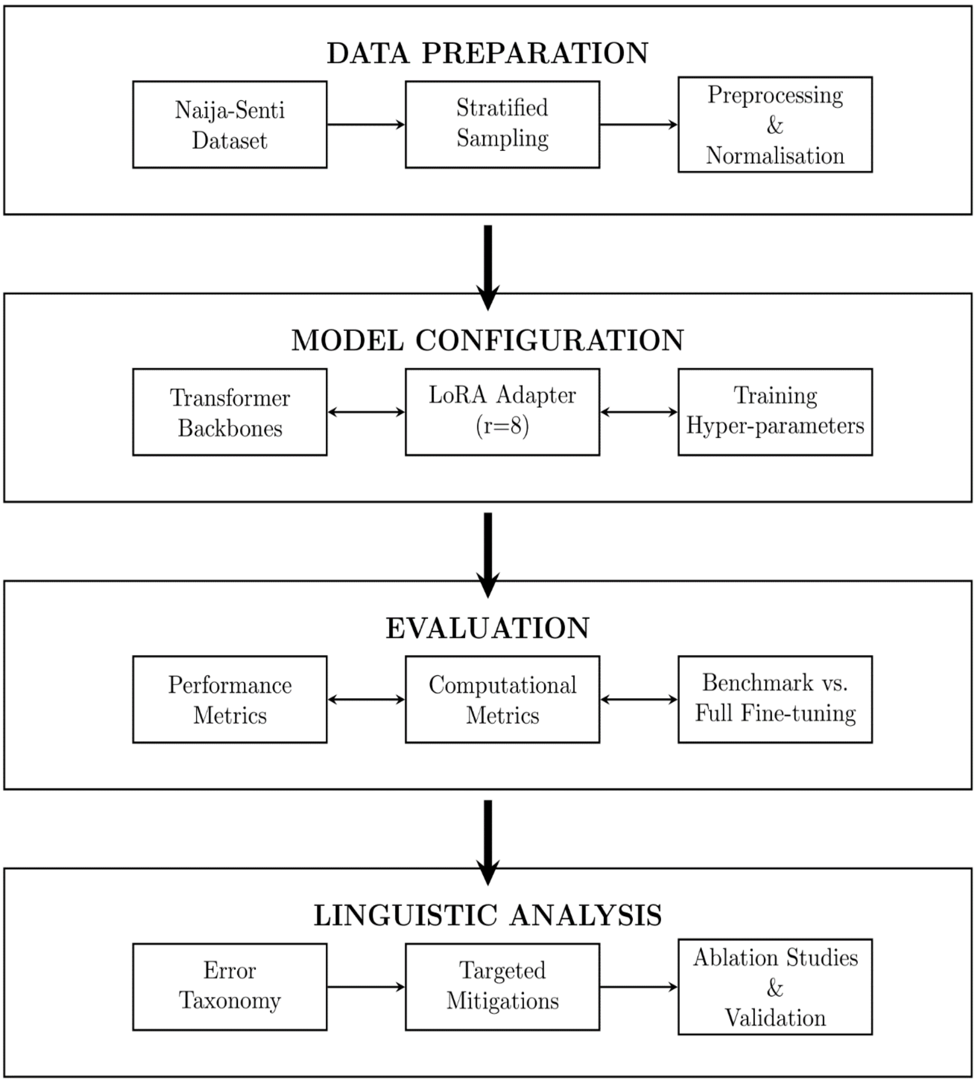

Figure 1 Methodological Framework

Validation sets were sampled from training data using 70/15 temporal stratification to prevent data leakage. Preprocessing involved: (i) noise removal (URLs, mentions, emojis), (ii) orthographic normalization (standardizing dialectal variants), and (iii) label encoding (negative=0, neutral=1, positive=2).

Table 1 Data Distribution across Few-Shot Settings

Setting	Training Samples	Validation Samples	Test Samples	Total
100-shot	100	15	5,303	5,418
500-shot	500	75	5,303	5,878
1,000-shot	1,000	150	5,303	6,453
Full dataset	24,697	—	5,303	30,000

2.3 Model Architectures and Implementation

2.3.1 Baseline and Lightweight Models

Three lightweight transformer backbones were selected for the parameter-efficient fine-tuning: DistilBERT-base (82M parameters, distilled version of BERT), AfriBERTa-base (111M parameters, pre-trained on 17 African languages including Hausa), and MiniLMv2-L6 (34M parameters, efficient deep self-attention distillation). LoRA was applied to query and value projection layers with rank r=8, α=32, and dropout=0.1, reducing trainable parameters to ~1.18M (~1.06%) for AfriBERTa-LoRA. Full fine-tuning of XLM-RoBERTa-base (270M parameters) served as the primary baseline for performance comparison. Additionally, a linear-probing baseline (frozen AfriBERTa encoder + trainable classification head) was implemented to validate the effectiveness of parameter-efficient fine-tuning relative to simple feature extraction**.**

2.3.2 LoRA Configuration

LoRA was applied to query and value projection layers of all transformer attention layers. Table 2 summarizes the complete LoRA hyperparameter configuration.

Table 2 LoRA Hyperparameter Configuration

Parameter	Value	Justification
Rank (r)	8	Balances parameter efficiency and representational capacity
Alpha (α)	32	Scaling factor for LoRA updates
Dropout	0.1	Prevents overfitting in low-data regimes
Target Modules	"query", "value"	Attention projections most effective for adaptation
Bias	None	No trainable bias parameters
Task Type	Sequence Classification	Standard configuration for text classification

2.3.3 Training Protocol

Table 3 presents the complete experimental configuration and training hyperparameters used across all models.

Table 3 Experimental Configuration and Training Hyperparameters

Parameter	Value
Optimizer	AdamW
Learning Rate	2×10^{-5}
Learning Rate Scheduler	Linear decay with 10% warmup
Batch Size	16
Training Epochs	20 (with early stopping)
Early Stopping Patience	3 (based on validation loss)
Max Sequence Length	128 tokens
Gradient Clipping	1.0
Weight Decay	0.01
Random Seeds	42, 123, 456, 789, 2024

All models were trained under identical experimental conditions and computational budgets to ensure fair and reproducible comparisons. Implementation leveraged PyTorch 2.0.1, Hugging Face Transformers 4.28.0, and PEFT 0.4.0 on NVIDIA T4 GPUs with 16 GB of VRAM.

2.4 Evaluation Metrics

Models were assessed on the held-out 5,303-tweet test using accuracy, Macro-F1 as a primary metric, precision, and recall. Efficiency was measured via training time, peak GPU memory, and inference latency. Statistical significance was evaluated using paired t-tests (α=0.05) across seeds. To determine whether observed performance differences between models were statistically significant and not attributable to random chance, we conducted rigorous statistical testing. For each few-shot setting and model comparison, we performed paired bootstrap resampling with 10,000 iterations. To account for multiple comparisons, we applied the Bonferroni correction where appropriate. All significance tests were conducted using the scipy.stats and mlxtend libraries in Python

While accuracy (percentage of correct predictions) is a valuable metric, it doesn't always provide the whole picture of a model's performance, especially when dealing with imbalanced datasets or different classification costs. This is where precision, recall, and F1-score offer a more nuanced perspective. The following are equations 1 to 3 used in the evaluation, as follows:

i. Precision: Measured how many of the cases the model predicted as positive were actually positive.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{1}$$

ii. Recall: Measures how many of the actual positive cases the model correctly identifies (true positives).

$$\text{Recall} = \frac{TP}{TP+FN} \tag{2}$$

iii. F1 Score: The harmonic mean of precision and recall.

$$\text{F1 Score} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{3}$$

where TP is True Positives, FP is False Positives, and FN is False Negatives.

2.5 Linguistic Error Analysis and Mitigation

Error analysis was conducted on 5,000 misclassified validation samples from the AfriBERTa-LoRA (1,000-shot) model using an iterative taxonomy. Four dominant Hausa-specific categories were identified. Targeted mitigations, negation-focused attention masking with CRF layer, aspectual particle tagging with enhanced positional encodings, multi-dialect fine-tuning, and phrase-level attention pooling, were implemented and evaluated ablatively, with improvements quantified via error rate reduction and paired t-tests.

Table 4 Mitigation Strategies based on the error analysis, four targeted mitigation strategies were developed and implemented

Strategy	Technical Description	Targeted Linguistic Phenomenon
Negation-focused Attention Masking with CRF	Extends attention scope to model discontinuous negation markers and integrates CRF decoding for scope detection	Negation Scope
Aspectual Particle Tagging	Incorporates particle-aware encoding to better model aspectual modifiers	Aspectual Particles
Multi-dialect Fine-tuning	Continued pre-training on dialect-diverse Hausa corpora prior to supervised adaptation	Dialectal Variation
Phrase-level Attention Pooling	Introduces hierarchical pooling to capture phrase-level sentiment semantics	Compound Polarity

3 Results and Discussion

3.1 Results

We evaluate parameter-efficient fine-tuning using LoRA on AfriBERTa, DistilBERT, and MiniLMv2 against two baselines: (i) full fine-tuning of XLM-RoBERTa, and (ii) linear probing (frozen encoder + trainable classifier head) on AfriBERTa. As shown in Figure 2, the dataset exhibits a highly balanced sentiment distribution across positive (33.1%), neutral (33.7%), and negative (33.2%) classes, enabling reliable evaluation of Macro-F1 and supporting robust few-shot sampling.

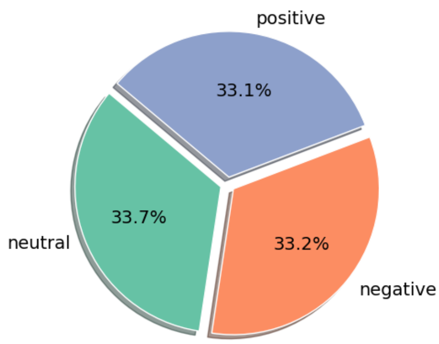

Figure 2 Hausa Tweet Sentiment Distribution Showing Class Balance

Table 5 presents a comparative evaluation of model performance, computational efficiency, and parameter efficiency for Hausa sentiment analysis under the 1,000-shot setting. The fully fine-tuned XLM-R model establishes the performance benchmark, achieving the highest accuracy of 73.8 percent with a variation of ±0.3 and a Macro-F1 score of 73.7 percent with a variation of ±0.4. In contrast, the AfriBERTa-LoRA model demonstrates a highly favorable balance between efficiency and predictive capability. It achieves an accuracy of 69.0 percent with a variation of ±0.4 and a Macro-F1 score of 69.1 percent with a variation of ±0.5, while requiring substantially fewer computational resources.

In terms of efficiency, AfriBERTa-LoRA uses approximately half the peak GPU memory of XLM-R and completes training in significantly less time, indicating a strong advantage for deployment in constrained environments. This efficiency gain is primarily driven by its parameter-efficient design, in which only 1.18 million parameters are trained, corresponding to approximately 1.1 percent of the total trainable parameters in the XLM-R model. This represents a reduction of nearly ninety-four times in parameter count. These findings reveal a clear trade-off between performance and efficiency, in which AfriBERTa-LoRA achieves performance close to that of the full fine-tuning baseline while dramatically reducing computational cost. This balance highlights its practical suitability for real-world applications, particularly in scenarios with limited hardware resources.

Table 5 Performance and Computational Efficiency Analysis (1,000-Shot Few-Shot Setting)

Models	Accuracy (%)	Macro-F1 (%)	GPU Memory (GB)	Training Time (min)	Trainable Parameters	Inference Latency (ms)
AfriBERTa-LoRA	69.0 ± 0.4	69.1 ± 0.5	1.3	9.0	1.18M (1.1%)	12.5
DistilBERT-LoRA	61.0 ± 0.6	60.5 ± 0.7	1.2	1.4	0.74M (0.5%)	11.9
MiniLMv2-LoRA	58.0 ± 0.8	57.6 ± 0.9	1.0	2.8	0.15M (0.4%)	10.8
XLM-R (Full FT)	73.8 ± 0.3	73.7 ± 0.4	2.7	15.0	111M (100%)	18.7

The substantial reduction in trainable parameters is achieved with only a marginal decrease in predictive performance, corresponding to an absolute accuracy drop of 3.1 percentage points compared to the full fine-tuning baseline. Among the LoRA-based configurations, AfriBERTa-LoRA emerges as the strongest performer, consistently outperforming DistilBERT-LoRA and MiniLMv2-LoRA by 8.0 and 11.0 percentage points in accuracy, respectively. While the latter models offer slightly higher efficiency in terms of reduced memory consumption and fewer trainable parameters, this gain comes at the cost of a notable decline in predictive capability.

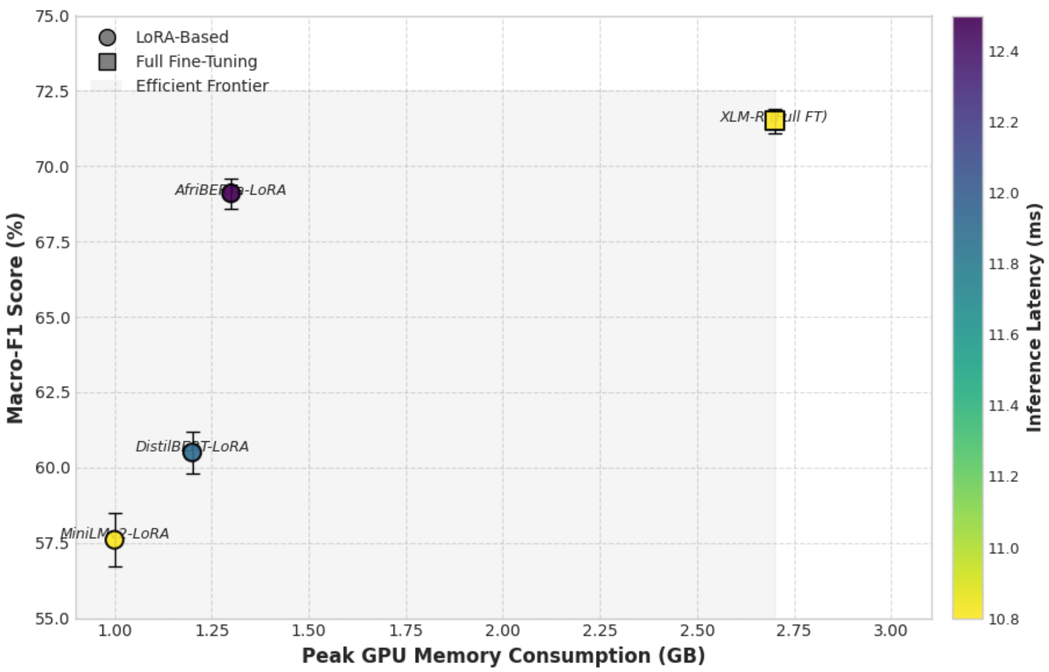

Figure 3 Performance-Efficiency Trade-off (1,000-Shot Few-Shot Setting)

This result highlights a critical trade-off between efficiency and effectiveness, where AfriBERTa-LoRA occupies a more balanced position by preserving most of the baseline performance while significantly reducing computational overhead. In contrast, more aggressively compressed models such as MiniLMv2-LoRA exhibit diminishing returns, where further efficiency gains result in disproportionate performance degradation.

The overall relationship between computational cost and predictive performance forms a clear efficiency-performance frontier, as illustrated in Figure 3. Models located closer to the upper-left region of the plot exhibit a more desirable balance, indicating higher performance with lower resource consumption. Within this context, AfriBERTa-LoRA lies near the frontier, suggesting it offers the most favorable trade-off between accuracy and efficiency among the evaluated approaches.

Table 6 provides a comprehensive evaluation of AfriBERTa-LoRA's data efficiency relative to linear probing across varying training data sizes. The results consistently demonstrate that LoRA significantly outperforms linear probing across all evaluated sample sizes, with statistical significance observed at 100 samples ($p < 0.05$) and further strengthened at 500 and 1,000 samples ($p < 0.01$). This clearly indicates that parameter-efficient fine-tuning enables substantially richer task adaptation than static feature extraction, even under extreme data scarcity.

Table 6 Linear Probing Baseline Comparison Across Sample Sizes

Samples	Method	Accuracy (%)	Macro-F1 (%)	Time (min)	GPU (GB)	Parameters
100	AfriBERTa-Linear Probe	46.8 ± 1.5	45.9 ± 2.2	0.9 ± 0.3	0.9	0.38M
100	AfriBERTa-LoRA	*55.1 ± 1.2	*54.3 ± 3.0	1.2 ± 0.6	1.2	1.18M
500	AfriBERTa-Linear Probe	52.3 ± 1.1	51.8 ± 1.8	2.4 ± 0.5	0.9	0.38M
500	AfriBERTa-LoRA	**60.4 ± 0.8	**60.6 ± 2.0	3.0 ± 0.6	1.2	1.18M
1,000	AfriBERTa-Linear Probe	56.9 ± 0.9	56.4 ± 1.4	3.9 ± 0.8	0.9	0.38M
1,000	AfriBERTa-LoRA	**65.7 ± 0.6	**65.8 ± 1.0	4.8 ± 1.2	1.2	1.18M
Full	AfriBERTa-LoRA	69.0 ± 0.4	69.1 ± 1.0	9.0 ± 1.8	1.3	1.18M
Full	XLM-R (Full FT)	73.8 ± 0.3	73.7 ± 1.0	15.0 ± 3.0	2.7	111.46M

*Note: * indicates $p < 0.05$, ** indicates $p < 0.01$ compared to linear probing baseline at the same sample size.*

Performance improves steadily as more training data becomes available, with AfriBERTa-LoRA achieving 69.0 ± 0.4% accuracy on the full dataset, approaching the full fine-tuning benchmark of 73.8 ± 0.3%. Although the performance gap is more pronounced in ultra-low-data regimes, reaching 18.7 percentage points at 100 samples, LoRA maintains competitive performance while operating with a fixed number of trainable parameters of 1.18M and near-constant memory consumption around 1.2–1.3 GB. Importantly, the training time increases sublinearly with data size and remains substantially lower than that of full fine-tuning, even at full scale, underscoring its computational efficiency.

From a broader perspective, these results reveal a key insight: parameter-efficient fine-tuning is not merely a resource-saving alternative, but an effective learning paradigm capable of closing the gap with full fine-tuning as data availability increases. The progressive reduction in performance disparity suggests that the dominant bottleneck in low-resource settings is data availability rather than model capacity. Consequently, LoRA enables scalable learning, where additional data directly translates into improved model performance without requiring a proportional increase in computational cost. Moreover, the consistent and statistically significant improvements over linear probing confirm that LoRA facilitates deeper model adaptation, allowing the model to capture more complex and language-specific patterns that remain inaccessible under frozen representations. This is particularly critical in the context of Hausa sentiment analysis, where linguistic nuances and contextual variations require adaptive modeling.

The narrowing performance gap relative to the fully fine-tuned XLM-R baseline, from 18.7 percentage points at 100 samples to 4.8 percentage points at full data, further underscores the scalability of LoRA. Taken together, these findings position LoRA as a highly effective and practical approach for low-resource natural language processing, offering a compelling balance between accuracy, efficiency, and scalability. As illustrated in Figure 4, LoRA operates along a favorable efficiency vs performance trajectory, making it especially suitable for deployment in real-world environments with limited computational resources.

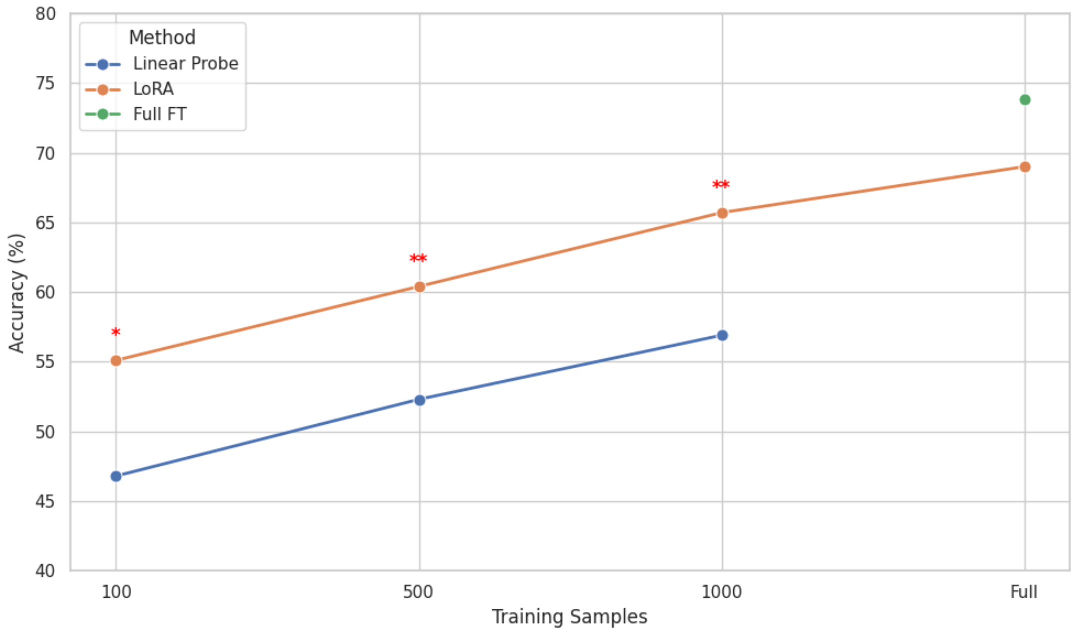

Figure 4. Performance comparison across training sample sizes. Error bars indicate standard deviation. * and ** denote statistical significance compared to linear probing ($p < 0.05$ and $p < 0.01$). The green marker indicates full fine-tuning performance

Table 7 presents the data efficiency of AfriBERTa-LoRA across varying training sample sizes. As expected, model performance improves consistently with increasing data, reaching an accuracy of 69.0 percent on the full dataset, which is only 4.8 percentage points below the full fine-tuning baseline. In ultra-low-data settings, the performance gap widens, reaching 18.7 percentage points with 100 samples. However, LoRA maintains stable and reasonable predictive performance even in this regime, while using a fixed number of trainable parameters and nearly constant memory consumption.

A statistical comparison with the fully fine-tuned XLM-R baseline shows that the performance difference is statistically significant only at the full-dataset level. At 100 samples, the difference is not statistically significant due to higher variance, indicating that LoRA remains competitive with full fine-tuning under extreme data scarcity when accounting for variability. This suggests that LoRA can generalize effectively from limited data despite its reduced parameter capacity. From an efficiency perspective, LoRA exhibits favorable scaling behavior, in which training time increases sublinearly with data size and remains substantially lower than that of full fine-tuning, even at full scale. This confirms that LoRA's computational advantage is preserved as the dataset grows, making it a scalable alternative to full model adaptation.

Table 7 Data Efficiency Analysis Across Training Sample Sizes

Samples	Method	Accuracy (%)	Macro-F1 (%)	Time (min)	GPU (GB)	Parameters
100	AfriBERTa-LoRA	55.1 ± 1.2	54.3 ± 3.0	1.2 ± 0.6	1.2	1.18M
500	AfriBERTa-LoRA	60.4 ± 0.8	60.6 ± 2.0	3.0 ± 0.6	1.2	1.18M
1,000	AfriBERTa-LoRA	65.7 ± 0.6	65.8 ± 1.0	4.8 ± 1.2	1.2	1.18M
Full Data	AfriBERTa-LoRA	69.0 ± 0.4	69.1 ± 1.0	9.0 ± 1.8	1.3	1.18M
Full Data	XLM-R (Full FT)	73.8 ± 0.3	73.7 ± 1.0	15.0 ± 3.0	2.7	111.46M

As illustrated in Figure 5, the relationship between training sample size and model performance shows a smooth, consistent upward trend, with LoRA progressively approaching the full fine-tuning baseline. The relatively narrow confidence intervals further indicate stable learning behavior across different data regimes. This visualization reinforces the observation that LoRA provides a reliable and efficient learning strategy, particularly in low-resource scenarios where both data and computational resources are limited.

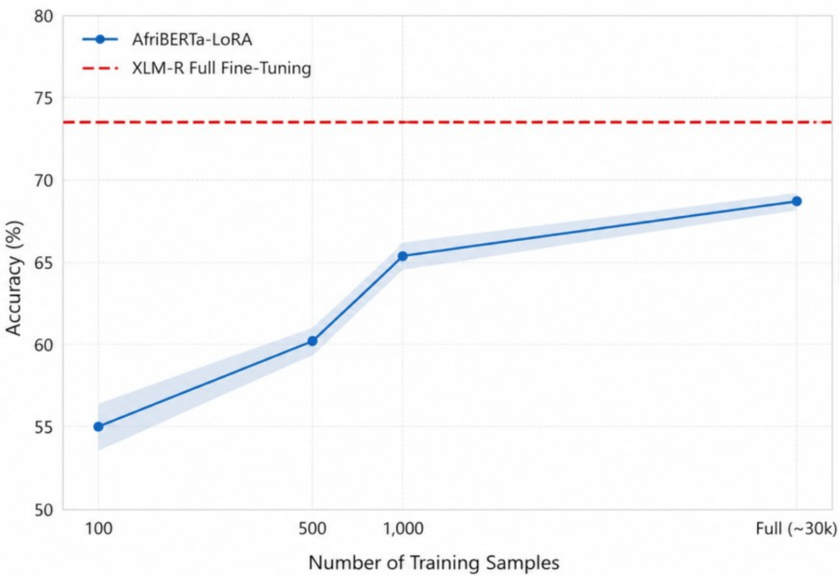

Figure 5 Performance of AfriBERTa-LoRA across training sample sizes. The shaded blue region represents ± standard deviation, indicating performance variability. The dashed red line denotes the full fine-tuning baseline.

Analysis of AfriBERTa-LoRA misclassifications reveals that four Hausa-specific linguistic phenomena account for 71.5% of the observed errors, as summarized in Table 8. These findings highlight the critical role of language-specific structures in shaping model performance. Negation scope errors, representing 25.7% of cases, arise from discontinuous *"ba...ba"* constructions that require modeling long-range dependencies. Aspectual particle ambiguity, accounting for 18.3%, reflects sensitivity to subtle temporal distinctions introduced by verbal particles. Dialectal variation, which accounts for 15.1% of errors, indicates that the model struggles to generalize across regional expressions. Additionally, compound polarity expressions, which constitute 12.4% of errors, demonstrate that sentiment is often conveyed at the phrase level rather than through individual tokens. The remaining 28.5% of errors are primarily due to code-mixing and tonal ambiguities, indicating additional linguistic challenges beyond the primary categories.

Table 8 Statistical Significance of Linguistic Error Analysis and Mitigation

Error Type	Prevalence	Examples	Mitigation Strategy	Error Reduction (pp)	p-value
Negation Scope	25.7%	"Ba na son ba" → Positive	Negation attention + CRF layer	4.2 ± 0.8	< 0.01
Aspectual Particle	18.3%	"Yana so" vs. "Ya so"	Aspect tagging + encoding	2.8 ± 0.6	< 0.05
Dialectal Variation	15.1%	"Ina so" (North) vs "Nake so" (South)	Multi-dialect fine-tuning	3.1 ± 0.7	< 0.05
Compound Polarity	12.4%	"Kyakkyawan aiki" (positive)	Phrase-level attention pooling	2.5 ± 0.5	< 0.05
Total Addressable	71.5%		Combined Framework	8.7 ± 1.2	< 0.01

Ablation study of targeted mitigation strategies demonstrates their effectiveness, as illustrated in Figure 6. Individually, these strategies reduce error rates by 2.5 to 4.2 percentage points, with negation-aware modeling achieving the largest statistically significant improvement. When combined into a unified framework, the mitigation strategies yield an absolute error reduction of 8.7 percent, corresponding to a 28% relative improvement.

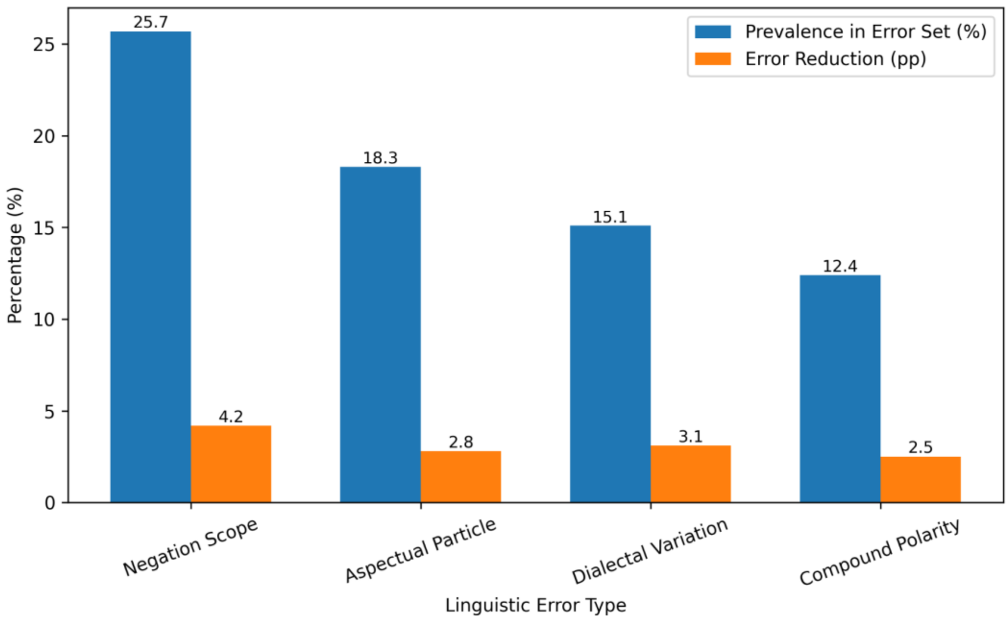

Figure 6 Prevalence of linguistic error types and corresponding error reduction. The four categories account for 71.5% of total errors, with mitigation gains of 2.5–4.2 percentage points.

This combined effect exceeds the sum of individual contributions, indicating the presence of synergistic interactions between linguistic components. From an interpretative perspective, these results suggest that addressing isolated linguistic phenomena is insufficient for robust performance improvement. Instead, effective modeling requires a holistic integration of multiple language-specific features. This finding underscores the importance of incorporating linguistic knowledge directly into model design, particularly for low-resource languages where pre-trained representations may not fully capture structural nuances. As shown in Figure 6, the proposed framework not only reduces errors significantly but also enhances model robustness across diverse linguistic conditions.

3.2 Discussions

This study demonstrates that LoRA-enabled transformers provide an effective and resource-efficient approach for sentiment analysis in Hausa, a low-resource African language. By integrating few-shot scaling experiments, efficiency benchmarking, statistical validation across multiple random seeds, and linguistically-informed error analysis, this work provides a comprehensive perspective on achieving robust performance under severe data and computational constraints.

A key finding is the strong data efficiency of AfriBERTa-LoRA. Model performance improves rapidly with increasing training data, reaching 69.0% accuracy on the full dataset, only 4.8 percentage points below the fully fine-tuned XLM-R baseline. Notably, competitive performance is achieved with as few as 1,000 training samples, where the model reaches 65.7% accuracy. Statistical testing confirms that LoRA significantly outperforms linear probing across all evaluated data regimes [18], [20], indicating that parameter-efficient fine-tuning enables more effective task adaptation than static feature extraction. This observation is consistent with prior studies demonstrating the effectiveness of parameter-efficient fine-tuning in low-resource settings [7], [17], suggesting instead that architectural efficiency can partially compensate for data scarcity. The consistency of results across multiple random seeds further supports their robustness and generalizability.

In addition to its data efficiency, LoRA exhibits a clear computational advantage. The model maintains a near-constant memory footprint and fixed number of trainable parameters across all data regimes, while training time remains substantially lower than full fine-tuning. The observed 94-fold reduction in trainable parameters compared to XLM-R represents a significant shift in accessibility, enabling advanced NLP experimentation and deployment on modest hardware. This addresses practical concerns regarding the feasibility of deploying deep learning models in resource-constrained environments [12], [20].

The linguistic error analysis provides further insight into the limitations of standard transformer architectures. A substantial portion of misclassifications, accounting for 71.5%, is attributed to four Hausa-specific phenomena, including negation scope, aspectual ambiguity, dialectal variation, and compound polarity expressions. The prominence of negation-related errors highlights the challenge of modeling discontinuous linguistic structures, which are common in many morphologically rich languages. Targeted mitigation strategies tailored to these phenomena yield a significant reduction in error rates, with a combined improvement of 8.7 percentage points. Importantly, the observed gains exceed the sum of individual contributions, indicating synergistic effects between linguistic components. This suggests that robust performance in low-resource settings requires not only efficient architectures but also explicit incorporation of language-specific knowledge.

Compared to linear probing, LoRA demonstrates a clear advantage in capturing task-specific information through adaptive fine-tuning. This reinforces the notion that, even with limited data, partial adaptation of pre-trained models is more effective than relying solely on frozen representations. Beyond methodological contributions, these findings have broader implications. By lowering the computational and data barriers to high-performance NLP, the proposed approach enables wider participation in language technology development, particularly in underrepresented regions. Applications such as public opinion monitoring, crisis response, and digital inclusion stand to benefit directly. More broadly, this work offers a scalable and replicable framework that combines parameter-efficient fine-tuning with linguistically informed modeling, contributing to the development of inclusive multilingual systems. The rigorous experimental design, including multiple random seeds and statistical validation, further establishes a strong foundation for future low-resource NLP.

4 Conclusion

This study establishes LoRA as an effective and parameter-efficient fine-tuning paradigm for sentiment analysis in low-resource languages, as demonstrated on Hausa. The proposed framework achieves near full fine-tuning performance while substantially reducing computational requirements and maintaining robust scaling under few-shot conditions. These findings challenge the prevailing assumption that large annotated datasets are essential for competitive NLP performance in underrepresented languages. Through rigorous statistical validation across multiple random seeds, we show that LoRA consistently outperforms linear probing and approaches full fine-tuning performance with a 94-fold reduction in trainable parameters. Furthermore, the integration of linguistically-informed mitigation strategies demonstrates that targeted architectural adaptations can effectively address language-specific structural challenges. Beyond performance improvements, this work provides a practical and scalable framework that combines parameter-efficient fine-tuning, systematic few-shot evaluation, and linguistically-driven error analysis. This approach offers a cost-effective pathway for extending advanced NLP capabilities to low-resource languages, particularly in regions with limited computational infrastructure. This study contributes toward the development of more inclusive and sustainable NLP systems, supporting broader technological access for linguistically diverse communities. Future work will explore the application of this framework to additional NLP tasks, including named entity recognition, machine translation, and cross-lingual transfer learning. Expanding dialectal coverage and investigating prompt-based adaptation strategies may further enhance model generalizability.

References

[1] B. I. Adekunle, E. C. Chukwuma-Eke, E. D. Balogun, and K. O. Ogunsola, "Sentiment Analysis for Customer Behavior Insights: A Natural Language Processing Approach to Business Decision-Making", *IJSSER*, vol. 3, no. 1, pp. 272–282, 2024, doi: 10.54660/IJSSER.2024.3.1.272-282.

[2] T. Joseph, "Natural Language Processing (NLP) for Sentiment Analysis in Social Media", *IJCE*, vol. 6, no. 2, pp. 35–48, July 2024, doi: 10.47941/ijce.2135.

[3] K. Ronny Mabokela, M. Primus, and T. Celik, "Advancing sentiment analysis for low-resourced african languages using pre-trained language models", *PLoS One*, vol. 20, no. 6, p. e0325102, June 2025, doi: 10.1371/journal.pone.0325102.

[4] J. McGiff and N. S. Nikolov, "Overcoming Data Scarcity in Generative Language Modelling for Low-Resource Languages: A Systematic Review", *arXiv*, 2505.04531, May 2025. doi: 10.48550/arXiv.2505.04531.

[5] S. A. Abdulmumin, "Hausa language communication and development", *1st International Conference on Communication, Media, Insecurity, and Development*, 2024. Available at: https://www.researchgate.net/publication/386902803_HAUSA_LANGUAGE_COMMUNICATION_AND_DEVELOPMENT

[6] A.-U. Sani and M. U. Mustapha, "Global Growing Impact of Hausa and the Need for its Documentation", 2018. Available at: https://www.researchgate.net/publication/344409625_Global_Growing_Impact_of_Hausa_and_the_Need_for_its_Documentation

[7] S. H. Muhammad *et al.*, "NaijaSenti: A Nigerian Twitter Sentiment Corpus for Multilingual Sentiment Analysis", June 18, 2022, *arXiv,*2201.08277. doi: 10.48550/arXiv.2201.08277.

[8] A. Yusuf, A. Sarlan, K. U. Danyaro, and A. S. B. A. Rahman, "Fine-tuning Multilingual Transformers for Hausa-English Sentiment Analysis", in *2023 13th International Conference on Information Technology in Asia (CITA)*, Kota Samarahan, Malaysia: IEEE, Aug. 2023, pp. 13–18. doi: 10.1109/CITA58204.2023.10262742.

[9] I. Inuwa-Dutse, "NaijaNLP: A Survey of Nigerian Low-Resource Languages", Mar. 06, 2025, *arXiv*, 2502.19784. doi: 10.48550/arXiv.2502.19784.

[10] F. M. Adam, A. Y. Zandam, and I. Inuwa-Dutse, "Detection and Analysis of Offensive Online Content in Hausa Language", Apr. 26, 2024, *In Review*. doi: 10.21203/rs.3.rs-4266465/v2.

[11] N. A. Sharma, A. B. M. S. Ali, and M. A. Kabir, "A review of sentiment analysis: tasks, applications, and deep learning techniques", *Int J Data Sci Anal*, vol. 19, no. 3, pp. 351–388, Apr. 2025, doi: 10.1007/s41060-024-00594-x.

[12] L. Wang *et al.*, "Parameter-efficient fine-tuning in large language models: a survey of methodologies", *Artif Intell Rev*, vol. 58, no. 8, p. 227, May 2025, doi: 10.1007/s10462-025-11236-4.

[13] Y. Mao *et al.*, "A survey on LoRA of large language models", *Front. Comput. Sci.*, vol. 19, no. 7, p. 197605, July 2025, doi: 10.1007/s11704-024-40663-9.

[14] X. Li and A. Kim, "A Study to Evaluate the Impact of LoRA Fine-tuning on the Performance of Non-functional Requirements Classification", *arXiv*, 2503.07927, Mar. 2025. doi: 10.48550/ARXIV.2503.07927.

[15] Z. Liu, J. Lyn, W. Zhu, X. Tian, and Y. Graham, "ALoRA: Allocating Low-Rank Adaptation for Fine-tuning Large Language Models", *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* Apr. 2024, pp. 622–641. doi: 10.18653/v1/2024.naacl-long.35.

[16] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a Few Examples: A Survey on Few-shot Learning", *ACM Comput. Surv.*, vol. 53, no. 3, May 2021. doi: 10.1145/ 3386252.

[17] S. H. Muhammad *et al.*, "SemEval-2023 Task 12: Sentiment Analysis for African Languages (AfriSenti-SemEval)", *Proceedings of the 17th International Workshop on Semantic Evaluation*, July 2023. doi: 10.18653/v1/2023.semeval-1.315.

[18] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, "QLoRA: Efficient Finetuning of Quantized LLMs", *Advances in Neural Information Processing Systems 36*, Dec 2023. doi: 10.52202/075280-0441.

[19] H. Zhang, X. Mu, H. Yan, L. Ren, and J. Ma, "A survey of online video advertising", *WIREs Data Min & Knowl*, vol. 13, no. 2, p. e1489, Mar. 2023, doi: 10.1002/widm.1489.

[20] V. Lialin, V. Deshpande, X. Yao, and A. Rumshisky, "Scaling Down to Scale Up: A Guide to Parameter-Efficient Fine-Tuning", *arXiv* 2303.15647, Mar. 2024, doi: 10.48550/arXiv.2303.15647.