Journal of Basics and Applied Sciences Research (JOBASR)
ISSN (print): 3026-9091, ISSN (online): 1597-9962
Volume 4(2) March 2026
DOI: https://dx.doi.org/10.4314/jobasr.v4i2.24

A Multi-Task Learning Framework for Sentiment Analysis and News Classification for Low-Resource Language

Salim Abubakar[1*], Abdulkadir Abubakar Bichi[2], Salisu Mamman Abdulrahman[3] & Ibrahim Said Ahmad[4]
[1,2]Northwest University, Kano Nigeria, Department of Computer Science
[3]Aliko Dangote University of Science and Technology, Wudil, Nigeria, Department of Computer Science,
[4]Bayero University, Kano, Nigeria, Department of Computer Science,
[*]Corresponding Author Email: saleemabubakar10@gmail.com

ABSTRACT

Despite the growing progress in Natural Language Processing (NLP), low-resource languages such as Hausa remain underrepresented in model development and evaluation. This paper presents a shared-encoder multi-task learning (MTL) framework based on AfriBERTa for joint Hausa sentiment analysis and news topic classification. The model learns both tasks simultaneously using Hausa subsets of the NaijaSenti and MasakhaNEWS datasets. Seven structured experiments were conducted, systematically evaluating optimization strategies including class weighting, encoder freezing, gradual unfreezing, and dynamic loss weighting. The best sentiment result was achieved in Experiment 5 (F1 = 80.7%) through encoder freezing and class rebalancing, while the best news classification result was achieved in Experiment 7 (F1 = 89.0%) through combined optimization with seed variation. The final shared-encoder model achieves a Macro-F1 of 74.0% and a Harmonic-F1 of 72.3%, substantially outperforming the FonMTL dual-encoder baseline in task balance (+27.2 Macro-F1 points; +64.1 Harmonic-F1 points). Although task interference was observed, the results confirm the viability of shared-encoder MTL for heterogeneous tasks in low-resource African language NLP and provide practical guidance for adapting multilingual models in such settings.

Keywords:
Multi-Task Learning;
AfriBERTa; Hausa;
Sentiment Analysis;
News Classification;
Low-Resource NLP;
Hard Parameter Sharing;
Task Interference

INTRODUCTION

Hausa is spoken by over 120 million people across Nigeria, Niger, and the broader West African Sahel region, making it one of the most widely spoken languages in Africa (Simons & Fennig, 2025). Despite its demographic significance, Hausa remains severely underrepresented in NLP research and model development, with a persistent shortage of annotated datasets, pre-trained language models, and benchmarked systems (Adelani et al., 2023; Muhammad et al., 2022). This resource gap limits the deployment of NLP-powered applications in Hausa-speaking communities, where demand for tools such as sentiment monitoring, news aggregation, and content classification is substantial.

Multi-Task Learning (MTL) offers a principled approach to addressing resource limitations by enabling a single model to learn multiple tasks simultaneously, leveraging shared representations to improve generalization and reduce the annotation burden for individual tasks (Caruana, 1997; Ruder, 2017). Rather than training separate models for each task, an MTL framework trains a unified model that shares parameters across tasks, exploiting complementary information in the training signals to develop richer, more generalizable representations.

AfriBERTa (Ogueji et al., 2021) is a multilingual pre-trained language model designed specifically for African languages, including Hausa. The NaijaSenti fine-tuned variant of AfriBERTa has achieved competitive results on Hausa sentiment analysis (Muhammad et al., 2022), demonstrating strong domain-specific representations. However, its potential as a shared encoder for multi-task learning in Hausa, particularly for heterogeneous tasks that combine social media sentiment with formal news classification, has not been empirically evaluated.

Therefore, this study aims to bridge this gap by developing and systematically evaluating a shared-encoder MTL framework that uses a NaijaSenti-fine-tuned AfriBERTa model as the backbone encoder for

How to cite this article: Salim A., Abdulkadir A. B., Salisu M. A. & Ibrahim S. A. (2026). A Multi-Task Learning Framework for Sentiment Analysis and News Classification for Low-Resource Language. *Journal of Basics and Applied Sciences Research*, *4*(2), 233-242. https://dx.doi.org/10.4314/jobasr.v4i2.24

joint Hausa sentiment analysis and news topic classification. The framework is evaluated against published single-task baselines and a re-implemented FonMTL dual-encoder baseline. Seven structured experiments assess the effects of encoder freezing, gradual unfreezing, class weighting, and dynamic loss weighting on task performance and task balance.

The specific problem addressed in this study is that, despite the availability of AfriBERTa as a multilingual African language model and the existence of annotated Hausa datasets for sentiment analysis and news classification, no prior work has examined whether a shared-encoder MTL framework can effectively learn both tasks jointly without catastrophic task interference. This gap limits our understanding of how pre-trained African language models behave under multi-task fine-tuning constraints. The research objectives of this study are therefore: (i) to develop a shared-encoder MTL framework using AfriBERTa for joint Hausa sentiment analysis and news classification; (ii) to systematically evaluate optimization strategies, including encoder freezing, gradual unfreezing, class weighting, and dynamic loss weighting , for balancing task performance; (iii) to compare the shared-encoder MTL framework against a re-implemented dual-encoder baseline (FonMTL) using task-balanced evaluation metrics; and (iv) to derive practical guidelines for adapting pre-trained multilingual encoders to heterogeneous low-resource African language tasks. The two tasks were jointly modelled because they represent complementary NLP needs in Hausa-speaking communities, sentiment understanding and topic categorization, and because their combination stress-tests encoder plasticity in ways that single-task fine-tuning cannot reveal.
Prior work on Hausa NLP has focused on morphological analysis, machine translation, and word embeddings (Abdulmumin & Galadanci, 2019; Akinfaderin, 2020), with NaijaSenti (Muhammad et al., 2022) and MasakhaNEWS (Adelani et al., 2023) providing the principal annotated resources for sentiment analysis and news classification respectively. AfriBERTa (Ogueji et al., 2021), pre-trained on 11 African languages, has demonstrated competitive performance on Hausa NLP tasks through language-adaptive fine-tuning (Sani et al., 2025) and multilingual adaptive fine-tuning (Alabi et al., 2022). In the broader MTL literature, Caruana (1997) established the theoretical foundation for joint task training, extended to deep neural networks by Collobert and Weston (2008) and later to transformers through systems such as MT-DNN (Liu et al., 2019) and adapter-based approaches (Pfeiffer et al., 2020). Hard parameter sharing remains the most widely adopted MTL architecture (Ruder, 2017), though it is susceptible to task interference when tasks have conflicting gradient signals (Zhang et al., 2021), with mitigation strategies including gradient surgery (Yu et al., 2020) and GradNorm (Chen et al., 2018). For African languages specifically, Adebara and Elmadany (2023) demonstrated multi-task benefits across 50+ languages, Dossou et al. (2023) proposed the dual-encoder FonMTL framework for the Fon language, and Otoibhi and David (2025) introduced SabiYarn for large-scale MTL across four Nigerian languages. The current study extends this line of work with the first focused evaluation of shared-encoder MTL for joint Hausa sentiment analysis and news classification.
The contributions of this paper are: (1) the first empirical study of shared-encoder MTL for joint Hausa sentiment analysis and news classification; (2) a systematic evaluation of optimization strategies for balancing heterogeneous MTL tasks in a low-resource setting; (3) evidence that shared-encoder MTL outperforms a dual-encoder baseline in task-balanced metrics; and (4) practical guidance for adapting pre-trained multilingual encoders to low-resource, multi-task African language NLP.

MATERIALS AND METHODS

Datasets
Two Hausa datasets were used. NaijaSenti (Muhammad et al., 2022) is a Twitter-based sentiment corpus with three classes: Negative, Neutral, and Positive. The Hausa subset contains 14,172 training samples, 1,680 validation samples, and 1,680 test samples. MasakhaNEWS (Adelani et al., 2023) is a news topic classification dataset covering seven categories: Business, Entertainment, Health, Politics, Religion, Sports, and Technology. The Hausa subset contains 3,690 training samples, 968 validation samples, and 2,307 test samples. A 5:1 class imbalance exists between the two datasets, with NaijaSenti providing substantially more training samples, which was explicitly managed through loss weighting and sampling strategies.

Table 1. Dataset Summary

Dataset	Task	Classes	Train	Dev / Test
NaijaSenti (Hausa)	Sentiment Analysis	3	14,172	1,680 / 1,680

MasakhaNEWS (Hausa)	News Classification	7	3,690	968 / 2,307

Model Architecture

The proposed framework uses a hard parameter sharing architecture. A single AfriBERTa encoder (Davlan/naija-twitter-sentiment-afriberta-large, 10 transformer layers, hidden size 1024) serves as the shared backbone, initialized from the NaijaSenti fine-tuned checkpoint to leverage existing sentiment-domain representations. Two task-specific linear classification heads are attached to the shared encoder: a sentiment head (1024→3) and a news head (1024→ 7), each followed by a Softmax activation. During training, both tasks are processed sequentially within each batch, and their losses are combined to update the shared encoder through a single backward pass. Figure 1 illustrates the proposed architecture alongside the original AfriBERTa model and the NaijaSenti single-task baseline for comparison.

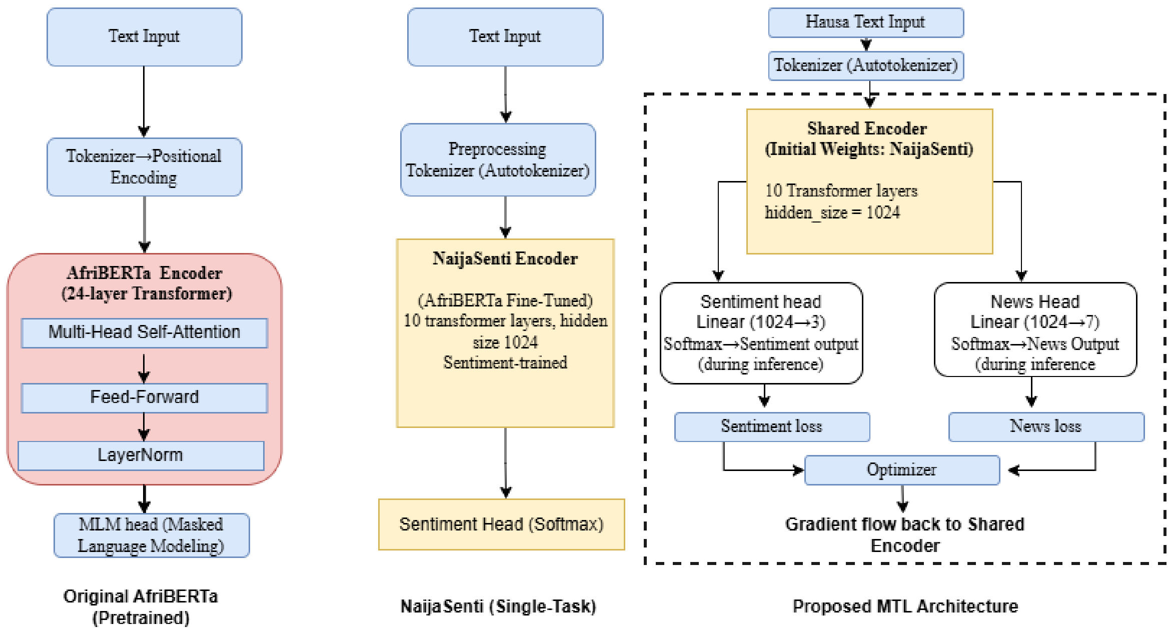

Figure 1. Proposed Shared-Encoder MTL Architecture Compared with Original AfriBERTa and NaijaSenti Single-Task Baseline

Experimental Setup

Seven structured experiments were conducted, each building on the previous one to address specific MTL challenges. All experiments used the AdamW optimizer with CrossEntropyLoss applied to both tasks. Table 2 summarizes the configuration of each experiment.

Table 2. Experimental Configurations

Exp.	Description	Encoder	LR	Epochs	Key Strategy
1	Shared encoder, two heads	Active	$5x10^{-5}$	3	Baseline MTL
2	Sentiment head only (ablation)	Active	$5x10^{-5}$	4	Single-task diagnostic
3	Fine-tuning from Exp. 1 checkpoint	Active	$5x10^{-6}$	1	Loss weights 1.5/1.0

4	Error analysis and misclassification	Active	N/A	N/A	Diagnostic
5	Encoder freezing + class rebalancing	Frozen	$5x10^{-5}$	4	Class weights [1,3,1]
6	Gradual unfreezing + dynamic loss	Gradual	Dynamic	8	1 layer/2 epochs
7	Combined optimization + seed variation	Gradual	Dynamic	15	Seeds 42, 24

Optimization Strategies

Four strategies were employed to address MTL challenges in the low-resource Hausa setting. (i) **Class Weighting** assigns weighted cross-entropy penalties to minority classes: sentiment weights [1.0, 3.0, 1.0] priorities the neutral class; news category weights address underrepresented topics. (ii) **Dynamic loss weighting** adapts task weights per Equation 1. (iii) **Encoder freezing** preserves pre-trained sentiment representations in early training; **gradual unfreezing** then progressively exposes one encoder layer per two epochs to joint gradient updates. (iv) **Seed variation** repeats key experiments with seeds 42 and 24 to assess result stability.

$$\text{Loss}_{task} = \alpha + (1 - \text{F1}_{best}) / \text{F1}_{target} \quad (1)$$

Where α is a base weight constant, F1_{best} is the best validation F1 achieved for the task, and F1_{target} is a preset target F1 value. The weight increases as a task falls behind its target, directing more gradient signal toward the underperforming task.

Preprocessing and Implementation Details

All experiments were implemented using PyTorch with the HuggingFace Transformers and Datasets libraries. Raw Hausa text was tokenized using the AfriBERTa tokenizer with truncation and padding applied to a fixed maximum sequence length of 128 tokens. The [CLS] token representation was used as the sentence embedding input to both classification heads. A batch size of 8 and the AdamW optimizer with a learning rate of 5×10^{-5} were used throughout, except in Experiment 3 where the learning rate was reduced to 5×10^{-6} for fine-tuning. In Experiment 1, tasks were trained sequentially with independent CrossEntropyLoss functions and separate backward passes. From Experiment 2 onwards, task losses were combined into a single weighted total loss with one backward pass and gradient accumulation over 2 steps, with a learning rate scheduler applied. Model checkpoints were saved at the epoch achieving the best validation F1 score, and final test evaluation was performed on the saved checkpoint.

Evaluation Metrics

Task-specific weighted F1-score and accuracy are reported for each task. F1-score is a widely adopted performance metric for classification systems, including in applied machine learning contexts (Sani, 2025). The primary cross-task metrics are Macro-F1, the simple average of per-task F1-scores, and Harmonic-F1, the harmonic mean of per-task F1-scores, which penalizes task imbalance non-linearly per Equation 2. These metrics were computed using the scikit-learn library.

$$\text{Harmonic-F1} = 2 \times (\text{F1}_{sent} \times \text{F1}_{news}) / (\text{F1}_{sent} + \text{F1}_{news}) \quad (2)$$

RESULTS AND DISCUSSION

Validation Performance

Table 3 presents validation F1-scores across the five main experiments (Experiments 2 and 4 were diagnostic and are therefore excluded from validation comparison).

Table 3. Validation F1-Scores across Experiments

Exp.	Description	Sentiment F1 (%)	News F1 (%)
1	Shared Encoder with Two Task-Specific Heads	55.7	86.1
3	Fine-Tuning with Adjusted Loss Weights	52.8	82.8

5	Layer Freezing and Class Rebalancing	78.9	8.2
6	Gradual Unfreezing with Dynamic Loss Weighting	30.0	86.9
7	Combined Optimization and Seed Variation	21.8	90.3

A clear performance tension is evident across experiments. News classification consistently achieved higher validation F1-scores than sentiment analysis in multi-task configurations. This asymmetry reflects the differing natures of the two tasks: news classification involves topically distinct categories with strong lexical cues, while sentiment analysis on Twitter data involves ambiguous, noisy text and a difficult neutral class. Experiment 5 is an outlier: freezing the encoder preserved sentiment-domain representations (F1 = 78.9%) but prevented the encoder from adapting to the news domain, causing a sharp drop in news validation performance (F1 = 8.2%).

Test Performance

Table 4 presents test results across experiments. Where no optimal checkpoint was saved during validation (i.e., experiments that were diagnostic in nature or that did not produce a stable checkpoint), test evaluation was not conducted; these entries are marked accordingly.

Table 4. Test Performance across Experiments

Exp.	Description	Sentiment F1 (%)	Sentiment Accuracy (%)	News F1 (%)	News Accuracy (%)
1	Shared Encoder, Two Heads	62.8	63.8	85.3	85.6
2	Sentiment Head Only (ablation)	31.6			
3	Fine-Tuning, Adjusted Weights	*	*	*	*
4	Error Analysis (diagnostic)	*	*	*	*
5	Encoder Freezing + Class Rebalancing	80.7	80.7	*	*
6	Gradual Unfreezing + Dynamic Loss	*	*	87.0	87.0
7	Combined Optimization, Seed Variation	*	*	89.0	89.0

* Test evaluation was not conducted for this experiment as no optimal checkpoint was saved during validation, or the experiment was diagnostic in nature and not intended for full test evaluation.

Baseline Comparisons

Single-Task Baselines

Table 5 presents published single-task benchmark results drawn from the original dataset papers. These serve as upper-bound references for task-specific performance under independent fine-tuning. AfriBERTa-large achieved a weighted F1 of 81.0% on Hausa sentiment analysis (Muhammad et al., 2022), while AfroXLMR-large achieved 92.2% on Hausa news classification (Adelani et al., 2023). Since these are independently fine-tuned single-task systems, their scores cannot be directly combined with multi-task results. The goal of this study is not to match single-task ceilings but to achieve well-balanced joint performance across both tasks.

Table 5. Single-Task Baseline Performance for Hausa

Task	Best Model	F1 (%)	Reference
Sentiment Analysis	AfriBERTa-large	81.0	Muhammad et al. (2022)
News Classification	AfroXLMR-large	92.2	Adelani et al. (2023)

FonMTL Baseline Comparison

To provide a controlled architectural comparison, FonMTL (Dossou et al., 2023), a dual-encoder MTL framework originally developed for Fon language part-of-speech tagging and named entity recognition, was re-implemented and adapted to the current classification tasks. Task-specific heads were replaced with linear layers for Hausa sentiment analysis (3 classes) and news classification (7 classes). The adapted FonMTL was trained from scratch on the same datasets, splits, and evaluation protocol as the proposed model. Table 6 presents the comparison.

Table 6. Shared-Encoder MTL vs. FonMTL Baseline

Model	Sent. F1 (%)	News F1 (%)	Macro-F1 (%)	Harmonic-F1 (%)
FonMTL (dual-encoder)	89.4	4.3	46.8	8.2
Shared-Encoder MTL (ours)	62.8	85.3	74.0	72.3

Although FonMTL achieved higher sentiment F1 (89.4%), it failed almost entirely on news classification (4.3%), yielding a Harmonic-F1 of only 8.2%. The proposed shared-encoder model achieved substantially better task balance with Macro-F1 of 74.0% (+27.2 points) and Harmonic-F1 of 72.3% (+64.1 points). This confirms that dual-encoder architectures, while capable of excelling on individual tasks, do not generalize effectively across heterogeneous tasks in low-resource settings. The shared-encoder approach enforces representational overlap between tasks, providing the cross-task regularization necessary for balanced performance.

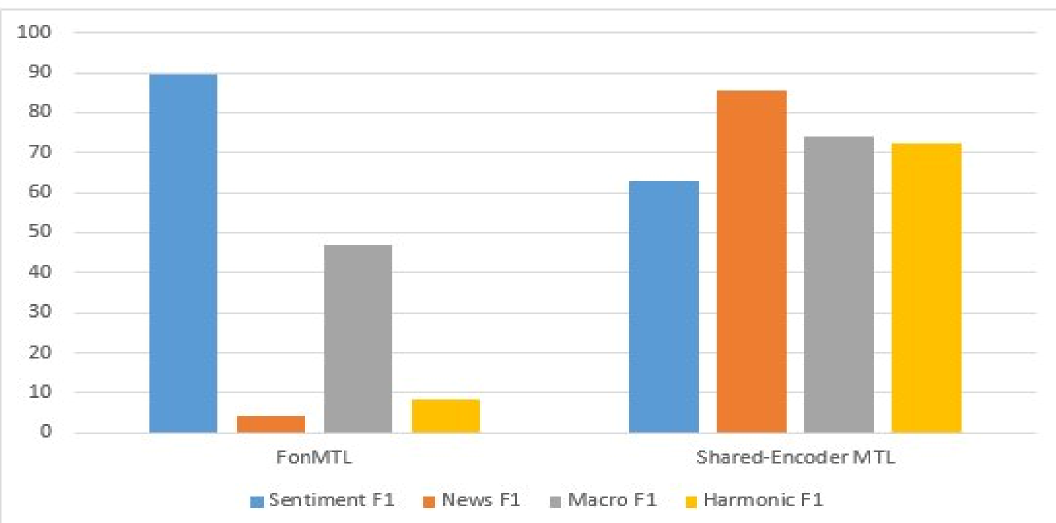

Figure 2. Sentiment F1, News F1, Macro-F1, and Harmonic-F1 Comparison: FonMTL vs. Shared-Encoder MTL

Qualitative Error Analysis

A qualitative examination of misclassified samples from Experiment 1 reveals systematic patterns that illuminate the nature of task interference. For sentiment analysis, the most frequent error type was misclassification of Neutral samples as Negative (463 instances). Inspection of these samples reveals that many contain negative lexical items used in non-evaluative contexts, such as news reporting of political events or health crises, which the shared encoder conflates with affective negativity. A representative misclassified example is: " An yi zabe a jihar Kano " (Elections were held in Kano State), a factual statement classified as Negative, likely because electoral content co-occurs with negatively-framed political discourse in the training data. These error patterns confirm that the shared encoder learns a feature space that is partially but imperfectly separated between affective and topical dimensions, and that future work targeting domain adaptation or task-specific adapter layers could address these systematic confusions.

Task Interference and Encoder Plasticity

The results reveal a fundamental tension in shared-encoder MTL when applied to heterogeneous tasks. Sentiment analysis on Twitter text requires sensitivity to affective markers, negation, and informal language, while news classification requires sensitivity to formal topical vocabulary and genre conventions. Forcing a single encoder to represent both feature spaces simultaneously introduces task interference, manifesting as competitive gradient signals that prevent either task from being optimally represented (Zhang et al., 2021).

The encoder freezing experiments (Experiment 5) clearly isolate this tension clearly. When the encoder was frozen, sentiment performance improved substantially (F1 = 80.7%) because the pre-trained NaijaSenti representations were preserved without disruption from news classification gradients. However, news classification performance collapsed in validation (F1 = 8.2%) because the static encoder could not adapt to news-domain features. The inverse pattern in gradual unfreezing experiments (Experiments 6 and 7), where news classification improved but sentiment declined progressively, confirms that encoder flexibility is a limited (effectively zero-sum) resource in this heterogeneous setup.

Figure 3 presents the confusion matrix for sentiment classification under the active encoder configuration (Experiment 1), illustrating the pattern of misclassification across the three sentiment classes. The model performs well on Negative samples (859 correctly classified) but shows considerable confusion in the Neutral and Positive classes, with 463 Neutral samples misclassified as Negative and 346 Positive samples similarly misclassified. This pattern confirms that the shared encoder, when exposed to joint gradient updates, develops a bias toward the Negative class, likely due to its stronger lexical distinctiveness in the training data. The Neutral class remains the most difficult to classify accurately, consistent with the class rebalancing motivation for Experiment 5.

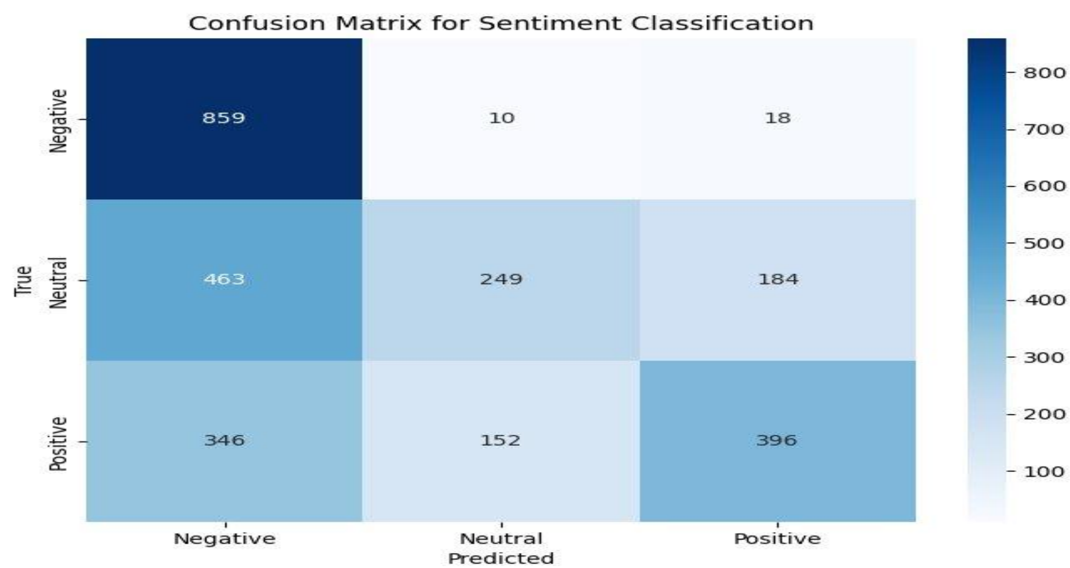

Figure 3. Confusion Matrix for Sentiment Classification - Active Encoder Configuration (Experiment 1)

Optimization Strategy Effects

Dynamic loss weighting (Equation 1) and class rebalancing partially mitigated task imbalance but did not fully resolve task interference. In Experiment 7, news classification achieved its best result across all experiments (F1 = 89.0%, consistent across seeds 42 and 24), demonstrating that combined optimization strategies can produce stable and replicable improvements for the stronger task. However, sentiment analysis continued to decline throughout Experiments 6 and 7, reaching validation F1 of only 21.8% in the final experiment, indicating that the training dynamics progressively favored news classification and that the applied strategies were insufficient to protect sentiment performance in a shared encoder under gradual unfreezing.

The ablation in Experiment 2 provides important evidence for the value of the multi-task setup itself. Training the shared encoder for sentiment analysis alone yielded F1 = 31.6%, substantially below the joint training result (Experiment 1, F1 = 62.8%), confirming that news classification provides regularization signals that improve sentiment representation, even in a shared encoder that is simultaneously being pulled toward the news domain.

Implications for Low-Resource African Language NLP

The findings have several practical implications for MTL in low-resource African language settings. First, shared-encoder architectures are viable for heterogeneous tasks, but careful encoder management is essential. Encoder freezing schedules and gradual unfreezing should be treated as primary hyperparameters, not afterthoughts, in multi-task fine-tuning. Second, task-balanced evaluation metrics such as Harmonic-F1 are necessary complements to per-task metrics: the FonMTL comparison demonstrates that strong per-task performance can conceal catastrophic failure on a secondary task. Third, the data imbalance between NaijaSenti and MasakhaNEWS (5:1 ratio) amplifies task interference, suggesting that data augmentation or curriculum sampling strategies could further improve balance in future work.

Limitations

This study has several limitations that should be acknowledged. First, the experiments were conducted on a single GPU with constrained computational resources, which limited the ability to perform large-scale hyperparameter searches or train models for extended epochs. Second, the study is restricted to two tasks and two datasets for a single language (Hausa); generalizability to other African languages or additional task combinations remains to be empirically established. Third, although the publicly available FonMTL code was used, it was originally designed for part-of-speech tagging and named entity recognition in the Fon language; adapting it to Hausa sentiment and news classification required replacing the original task-specific heads with new linear classification layers, meaning the adapted baseline differs from the original system and may not fully reflect its intended design. Fourth, statistical significance testing was not performed across experimental runs due to the computational cost of repeated trials; the seed variation in Experiment 7 provides partial evidence of result stability but is not a substitute for full significance testing. Fifth, the qualitative error analysis is limited to Experiment 1 and may not fully represent error patterns under other optimization configurations. These limitations define a clear agenda for future work, including multi-language evaluation, formal significance testing, and exploration of adapter-based architectures.

CONCLUSION

This study addressed the challenge of applying multi-task learning (MTL) to low-resource African languages, focusing on Hausa sentiment analysis and news classification. A shared-encoder framework was developed by adapting the pre-trained AfriBERTa model (Davlan/naija-twitter-sentiment-afriberta-large) as a single encoder to jointly handle both tasks, and evaluated through structured experiments to investigate optimization strategies and task balance in heterogeneous settings.

The key contributions of this work include: (1) the adaptation of Davlan/naija-twitter-sentiment-afriberta-large as a shared-encoder MTL framework for joint Hausa NLP tasks; (2) a systematic evaluation of optimization strategies, including encoder freezing, gradual unfreezing, class weighting, and dynamic loss weighting; (3) a comparative analysis demonstrating improved task balance over an adapted dual-encoder baseline (FonMTL); and (4) practical guidelines for applying MTL in low-resource African language contexts. All four objectives outlined in the introduction were successfully addressed.

Experimental results show that the proposed approach achieves balanced performance across tasks, with a Macro-F1 score of 74.0% and a Harmonic-F1 score of 72.3%, outperforming the baseline in task balance. Task interference remains a key challenge, with encoder plasticity identified as a central factor affecting performance, while the results provide a reproducible and replicable empirical foundation for future optimization.

Beyond methodological contributions, this work supports the development of inclusive NLP technologies for African languages. Applications such as media monitoring, public opinion analysis, and digital governance can benefit from improved Hausa language processing.

Future work should explore adapter-based architectures, meta-learning approaches for dynamic loss balancing, and the extension to additional tasks and languages. Overall, this study demonstrates that carefully adapting Davlan/naija-twitter-sentiment-afriberta-large within a shared-encoder MTL framework provides a practical and effective approach for advancing NLP in low-resource African language settings.

REFERENCE

Abdulmumin, I., & Galadanci, B. S. (2019). HauWE: Hausa words embedding for natural language processing. In Proceedings of the 2019 2nd International Conference of the IEEE Nigeria Computer Section (NigeriaComputConf). https://doi.org/10.1109/NigeriaComputConf45974.2019.8949674

Adebara, I., & Elmadany, A. (2023). Improving African language identification with multi-task learning. In Proceedings of the 4th Workshop on African Natural Language Processing (AfricaNLP 2023) at ICLR 2023. https://aclanthology.org/2023.africanlp-1.5

Adelani, D. I., Masiak, M., Azime, I. A., Alabi, J., Tonja, A. L., Adhiambo, W., Bamba, M., Ajayi, T., Emezue, C., Aremu, A., Ogbu, L., Adewumi, T., Eme-Nelson, A., & Aladesuru, O. (2023). MasakhaNEWS: News topic classification for African languages. In Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (IJCNLP-AACL 2023), 144-159. https://doi.org/10.18653/v1/2023.ijcnlp-main.10

Akinfaderin, A. (2020). HausaMT v1.0: Towards English-Hausa neural machine translation. In Proceedings of the 1st Workshop on NLP for Positive Impact (NLP4PI), 144-147. https://aclanthology.org/2020.nlp4posimpact-1.1

Alabi, J. O., Adelani, D. I., Mosbach, M., & Klakow, D. (2022). Adapting pre-trained language models to African languages via multilingual adaptive fine-tuning. In Proceedings of the 29th International Conference on Computational Linguistics (COLING 2022), 4336-4349. https://aclanthology.org/2022.coling-1.382

Caruana, R. (1997). Multitask learning. Machine Learning, 28(1), 41-75. https://doi.org/10.1007/BF00993277

Chen, Z., Badrinarayanan, V., Lee, C. Y., & Rabinovich, A. (2018). GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In Proceedings of the 35th International Conference on Machine Learning (ICML 2018), 794-803. https://doi.org/10.48550/arXiv.1711.02257

Collobert, R., & Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th International Conference on Machine Learning (ICML 2008), 160-167. https://doi.org/10.1145/1390156.1390177

Dossou, B. F. P., Emezue, C. C., Osei, S., Alabi, J., Etori, N. A., Traille, M. Y., Tossou, A., & Abbott, J. T. (2023). FonMTL: Towards multitask learning for the Fon language. In Proceedings of the 4th Workshop on African Natural Language Processing (AfricaNLP 2023). https://arxiv.org/abs/2308.14280

Liu, X., He, P., Chen, W., & Gao, J. (2019). Multi-task deep neural networks for natural language understanding. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019), 4487-4496. https://doi.org/10.18653/v1/p19-1441

Muhammad, S. H., Adelani, D. I., Ruder, S., Abdulmumin, I., Bashir, I., Alabi, J., Yimam, S. M., Abubakar, I., Muhammad, I., Sambo, F., & Rabiu, I. (2022). NaijaSenti: A Nigerian Twitter sentiment corpus for multilingual sentiment analysis. In Proceedings of the 13th Language Resources and Evaluation Conference (LREC 2022), 590-602. https://aclanthology.org/2022.lrec-1.63

Ogueji, K., Zhu, Y., & Lin, J. (2021). AfriBERTa: Leveraging African languages for self-supervised representation learning. In Proceedings of the 1st Workshop on Multilingual Representation Learning (MRL 2021), 116-126. https://aclanthology.org/2021.mrl-1.11

Otoibhi, J., & David, O. (2025). SabiYarn: Advancing low-resource languages with multitask NLP pretraining. In Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025), 95-107.

Pfeiffer, J., Rücklé, A., Poth, C., Geigle, C., Glockner, I., Both, T., Hermans, A., Schwartz, L., Reiter, N., & Gurevych, I. (2020). AdapterHub: A framework for adapting transformers. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (EMNLP 2020), 46-54. https://aclanthology.org/2020.emnlp-demos.7

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098. https://arxiv.org/abs/1706.05098

Sani, B. U. (2025). Optimizing video streaming quality with DBN-based ABR prediction models. Journal of Basics and Applied Sciences Research, 3(6), 77-81. https://doi.org/10.4314/jobasr.v3i6.9s

Sani, S. A., Muhammad, S. H., & Jarvis, D. (2025). Investigating the impact of language-adaptive fine-tuning on sentiment analysis in Hausa using AfriBERTa. arXiv preprint arXiv:2501.11023. https://arxiv.org/abs/2501.11023

Simons, G. F., & Fennig, C. D. (Eds.). (2025). Ethnologue: Languages of the world (28th ed.). SIL International. https://www.ethnologue.com/language/hau

Yu, T., Kumar, S., Gupta, A., Levine, S., Hausman, K., & Finn, C. (2020). Gradient surgery for multi-task learning. Advances in Neural Information Processing Systems (NeurIPS 2020), 33, 1174-1184.

Zhang, J., Yan, K., & Mo, Y. (2021). Multi-task learning for sentiment analysis with hard-sharing and task recognition mechanisms. Information, 12(5). https://doi.org/10.3390/info12050207