HausaNLP at SemEval-2025 Task 11: Advancing Hausa Text-based Emotion Detection

Sani Abdullahi Sani[1,2], **Salim Abubakar**[1,3], **Falalu Ibrahim Lawan**[1,4], **Abdulhamid Abubakar**[1], **Maryam Bala**[1]

[1]HausaNLP, [2]University of the Witwatersrand, [3]Federal Polytechnic Daura, [4]Kaduna State University

correspondence: saniabdullahisani1@students.wits.ac.za

Abstract

This paper presents our approach to multi-label emotion detection in Hausa, a low-resource African language, as part of SemEval Track A. We fine-tuned AfriBERTa, a transformer-based model pre-trained on African languages, to classify Hausa text into six emotions: anger, disgust, fear, joy, sadness, and surprise. Our methodology involved data preprocessing, tokenization, and model fine-tuning using the Hugging Face Trainer API. The system achieved a validation accuracy of 74.00%, with an F1-score of 73.50%, demonstrating the effectiveness of transformer-based models for emotion detection in low-resource languages.

1 Introduction

Emotion detection in text is a fundamental Natural Language Processing (NLP) task with far-reaching applications in sentiment analysis, social media monitoring, and mental health support. While significant progress has been made in high-resource languages, low-resource languages like Hausa remain underrepresented in the NLP landscape. One of the key challenges is the limited availability of annotated Hausa datasets for emotion detection, which hinders the development of robust models. This work aims to bridge this gap by leveraging a newly available dataset called BRIGHTER by (Muhammad et al., 2025a) and fine-tuning a transformer-based model for multi-label emotion detection in Hausa.

The scarcity of comprehensive emotion lexicons and annotated corpora makes it difficult to develop and evaluate emotion detection systems for low-resource languages (Kabir et al., 2023; Al-Wesabi et al., 2023; Raychawdhary et al., 2023a). Additionally, these languages exhibit rich morphological variations, syntax, and semantic differences that are not well-captured by models trained on high-resource languages (Marreddy et al., 2022; V R et al., 2023). The phenomenon of code-mixing, where multiple languages are used within the same text, adds another layer of complexity to emotion detection in linguistically diverse contexts (Raychawdhary et al., 2023a; Sonu et al., 2022).

Our approach leverages AfriBERTa, a transformer model specifically trained on African languages, fine-tuned for multi-class emotion classification. The text data is preprocessed, tokenized using the AfriBERTa tokenizer, and the model is fine-tuned on the BRIGHTER Hausa dataset. This work contributes to the growing body of research on emotion detection in low-resource languages by demonstrating effective methods for adapting transformer models to Hausa. Our findings highlight the potential of leveraging pre-trained models like AfriBERTa (Ogueji et al., 2021) for emotion detection tasks in low-resource African languages.

2 Background

2.1 Task Setup

The SemEval-2024 Track A, which is the Multi-label Emotion Detection task under Task 11: Bridging the Gap in Text-Based Emotion Detection (Muhammad et al., 2025b) focuses on classifying text from multiple languages into six emotion categories: anger, disgust, fear, joy, sadness, and surprise. While the task encompasses a wide range of languages, our team (HausaNLP) chose to focus on Hausa, a Chadic language widely spoken across West and Central Africa by approximately 88 million people, including 54 million native speakers and 34 million second-language users primarily in Nigeria, Niger, and neighboring countries (Wolff, 2024; Eberhard et al., 2024). The input to the system is a text sample in Hausa, and the output is a one-hot encoded vector indicating the presence or absence of each emotion. This task is particularly challenging due to the nuances of emotion expression in Hausa, as well as the limited availability of annotated data for low-resource languages.

2.2 Datasets

The Hausa dataset is divided into three splits: train, validation, and test. The training set contains approximately 2,145 samples, the validation set con-

Proceedings of the The 19th International Workshop on Semantic Evaluation (SemEval-2025), pages 2014–2019
July 31 - August 1, 2025 ©2025 Association for Computational Linguistics

tains 356 samples, and the test set contains 1,080 samples. The complete distribution of the dataset is shown in Figure 1. Each sample is annotated with one or more emotion labels, represented as a one-hot encoded vector as shown in 1. The dataset exhibits some class imbalance, with certain emotions (e.g., joy and sadness) being more prevalent than others (e.g., fear and disgust). This imbalance posed a challenge during model training and evaluation.

3 Related Work

Several researchers have proposed innovative approaches to address the challenges of emotion detection in low-resource settings. Efforts to create and annotate datasets for low-resource languages include the AfriSenti-SemEval 2023 Shared Task, which provided annotated datasets for languages like Hausa and Igbo (Raychawdhary et al., 2023a). Techniques such as few-shot and zero-shot learning have been employed to augment datasets and improve model performance in data-scarce environments (Agarwal and Abbas, 2025; Hasan et al., 2024).

Pre-trained models like mBERT, XLM-R, and BanglaBERT have been fine-tuned for specific low-resource languages, showing improved performance in emotion detection tasks (Raychawdhary et al., 2023b; Kabir et al., 2023; Raychawdhary et al., 2024). Hybrid models combining lexicon features with transformers have also been explored to enhance emotion classification accuracy (Kabir et al., 2023). Advanced methods such as the use of deep learning models (e.g., CNNs, DNNs) and hybrid approaches (e.g., TCSO-DGNN) have been proposed to better capture the emotional nuances in text (Jadon and Kumar, 2023; Bao and Su, 0).

AfriBERTa, a transformer model pre-trained on a diverse corpus of African languages, has shown promising results for tasks like text classification and named entity recognition in Hausa and other African languages (Ogueji et al., 2021). Our work builds on these advancements by fine-tuning AfriBERTa for emotion detection in Hausa, addressing the unique challenges of low-resource language processing.

Future directions in this field include multimodal emotion detection that incorporates additional data modalities (e.g., audio, video) to complement textual analysis (Sarbazi-Azad et al., 2021; Wang and Zhang, 2025), developing cross-lingual and multilingual models that can generalize across multiple low-resource languages (Raychawdhary et al., 2023b, 2024), and addressing ethical considerations related to data privacy, bias, and the deployment of emotion detection systems in diverse cultural contexts (Agarwal and Abbas, 2025).

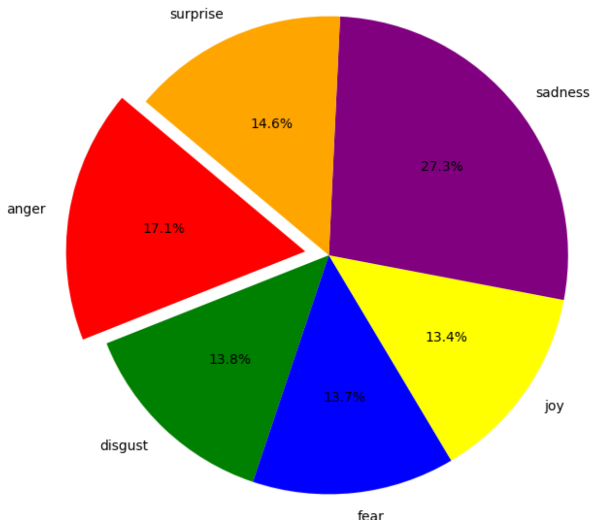

Figure 1: Distribution of Emotions in the Hausa Dataset

4 Methodology - Experimental Setup

4.1 Data Preparation and Preprocessing

To prepare the data for training, a preprocessing pipeline was implemented. First, the text data was cleaned by converting it to lowercase and stripping extra spaces. This step ensured consistency in the text format and improved tokenization. Next, the one-hot encoded labels were mapped to a single integer label representing the dominant emotion for each text sample. This transformation simplified the classification task into a multi-class problem with six classes.

4.2 Tokenization and Dataset Formatting

The preprocessed text data was tokenized using the AfriBERTa tokenizer, a pre-trained tokenizer specifically designed for African languages, including Hausa. The tokenizer was configured to truncate sequences longer than 128 tokens and pad shorter sequences to ensure uniform input sizes. The tokenized dataset was then formatted into PyTorch tensors, containing the input_ids, attention_mask, and label columns. This step prepared the data for input into the transformer model.

Table 1: Sample Entries from the Hausa Emotion Detection Dataset

ID	Text	Anger	Disgust	Fear	Joy	Sadness	Surprise
1	Kotu Ta Yi Hukunci Kan Shari'ar Zaben Dan Majalisar PDP, Ta Yi Hukuncin Bazata	0	0	0	0	0	1
2	Toh fah inji 'yan magana suka ce "ana wata ga wata"	0	0	0	0	0	1
3	Bincike ya nuna yan Najeriya sun fi damuwa da rashin tsaro da talauci fiye da korona	0	0	1	0	1	0

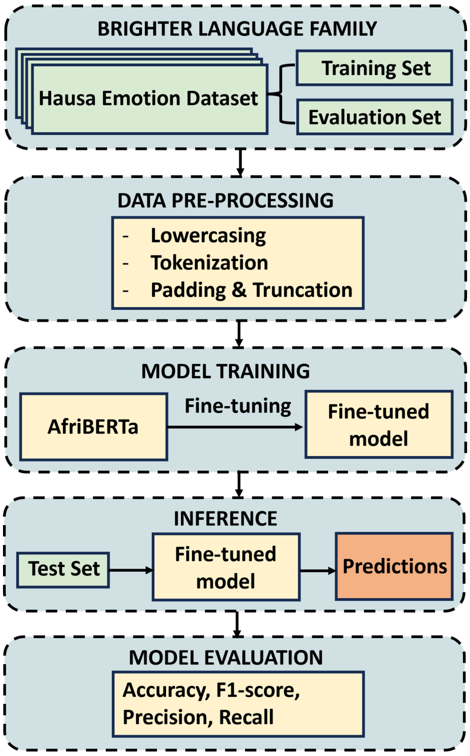

Figure 2: Multi-label Emotion Detection - System Overview

4.3 Model Selection and Fine-Tuning

The AfriBERTa Small model, a compact version of the AfriBERTa architecture, was selected for this task mainly for compute limitation. The model was fine-tuned for sequence classification with six output labels corresponding to the six emotions.

The fine-tuning process was conducted using the Hugging Face Trainer API. The training arguments were configured with a learning rate of 2e-5, a batch size of 8, and 5 epochs. To optimize training, mixed precision (fp16) was enabled, and a warmup schedule with 500 steps was applied to gradually increase the learning rate at the start of training. The model was evaluated on the validation set after each epoch, and the best model was saved based on validation accuracy.

4.4 Evaluation Metrics

During training, the model's performance was evaluated using standard classification metrics: accuracy, precision, recall, and F1-score. These metrics were computed using the classification_report function from the sklearn.metrics library. The evaluation was performed on the validation set after each epoch to monitor the model's progress and ensure it was learning effectively.

4.5 Model Saving and Inference

After fine-tuning, the best-performing model was saved to disk for future use. The model and tokenizer were saved. For inference, the model was loaded using the Hugging Face pipeline API, which simplified the process of making predictions on new text samples. The model was tested on a sample Hausa text, and the predicted emotion was displayed.

4.6 Test Set Predictions

To evaluate the model's performance on unseen data and submit to the SemEval 2025 Task 11, predictions were made on the test set. The test set was loaded from a CSV file, and the model predicted the dominant emotion for each text sample. The predictions were converted into one-hot encoded labels and saved back to a new CSV file. This file contained the original text samples along with the predicted emotion labels, allowing for further analysis and evaluation.

Table 2: Training and Validation Metrics for the Fine-Tuned AfriBERTa Model

	Accuracy	Precision	Recall	F1-Score
Train	0.7416 ± 0.02	0.7515 ± 0.02	0.7348 ± 0.02	0.7393 ± 0.02
Val	0.7400 ± 0.02	0.7500 ± 0.02	0.7300 ± 0.02	0.7350 ± 0.02

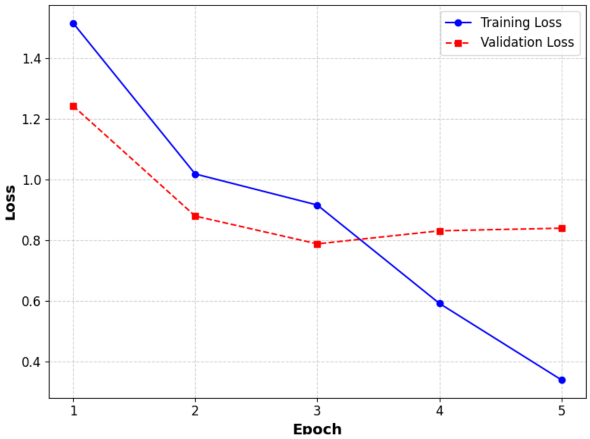

Figure 3: Comparison of Training and Validation Loss Across Epochs: The plot illustrates the progression of training and validation loss over five epochs. The training loss steadily decreases, indicating that the model is learning from the data. However, the validation loss initially decreases but later fluctuates, suggesting potential overfitting or variations in generalization performance.

5 Results and Discussion

The fine-tuned AfriBERTa model demonstrated consistent improvement over the training epochs, as shown in Table 2. The model achieved its best performance in the fifth epoch, with a training accuracy of **74.16%** and an F1-score of **73.93%**. Precision and recall scores were also strong, at **75.15%** and **73.48%**, respectively. These results indicate that the model effectively learned to classify emotions in Hausa text, despite the challenges posed by class imbalance and noisy data.

Additionally, the model showed steady progress, with the training loss decreasing from **1.5168** in the first epoch to **0.3385** in the fifth epoch. The validation loss also decreased initially but stabilized around **0.8393** in the final epoch, suggesting that the model reached a point of convergence. The model's performance on the validation set was consistent with the training results, achieving an accuracy of **74.00%** and an F1-score of **73.50%**, demonstrating its ability to generalize to unseen data.

The model performed particularly well on emotions like **joy** and **sadness**, which were more prevalent in the dataset. However, it struggled slightly with underrepresented emotions such as **fear** and **disgust**, likely due to their limited representation in the training data. This highlights the importance of addressing class imbalance in future work, potentially through techniques like data augmentation or weighted loss functions. Overall, the results underscore the effectiveness of AfriBERTa for emotion detection in low-resource languages and provide a strong baseline for future research in this domain.

6 Conclusion

In this work, we fine-tuned AfriBERTa for multi-label emotion detection in Hausa text, achieving an accuracy of 74.00% and an F1-score of 73.50%. The model performed well on prevalent emotions like joy and sadness but struggled with underrepresented emotions such as fear and disgust, highlighting the challenge of class imbalance. Our approach demonstrates the effectiveness of transformer-based models for low-resource language tasks. Future work could address class imbalance through data augmentation or weighted loss functions and extend this work to other African languages. This study provides a strong baseline for emotion detection in Hausa and contributes to the development of inclusive NLP systems.

7 Limitation

One major limitation of this work is the variability in Hausa dialects. The dataset primarily represents standard Hausa, which may not generalize well to regional dialects or informal variations commonly used in social media and conversational settings. This could lead to misclassification of emotions when processing text from underrepresented dialects.

Another challenge is the generality of the dataset. While the dataset is carefully curated, it may not fully capture the diversity of emotions expressed in different contexts, such as sarcasm, code-mixing with English or Arabic, and cultural-specific expressions. This limits the robustness of the model when applied to unseen, real-world data.

Additionally, computational constraints influenced the choice of AfriBERTa-small instead of larger transformer models. While this ensures efficiency, it may come at the cost of lower accuracy compared to models with higher capacity. The trade-off between computational efficiency and model performance is a key consideration for practical deployment.

Lastly, deep learning models often lack interpretability, making it difficult to explain why a particular emotion was predicted. This could pose challenges in high-stakes applications, such as mental health monitoring or crisis detection, where transparency is crucial.

Acknowledgments

The authors acknowledge Google DeepMind for a study grant.

References

Ritika Agarwal and Noorhan Abbas. 2025. Emotion detection in hindi language using gpt and bert. In *Artificial Intelligence XLI*, pages 105–118, Cham. Springer Nature Switzerland.

Fahd N. Al-Wesabi, Hala J. Alshahrani, Azza Elneil Osman, and Elmouez Samir Abd Elhameed. 2023. Low-resource language processing using improved deep learning with hunter–prey optimization algorithm. *Mathematics*, 11(21).

Dianqing Bao and Wen Su. 0. Optimizing deep learning-based natural language processing for sentiment analysis. *International Journal of High Speed Electronics and Systems*, 0(0):2540304.

David M. Eberhard, Gary F. Simons, and Charles D. Fennig. 2024. *Ethnologue: Languages of the World*, twenty-seventh edition. SIL International, Dallas, Texas. Online version: http://www.ethnologue.com.

Md. Arid Hasan, Shudipta Das, Afiyat Anjum, Firoj Alam, Anika Anjum, Avijit Sarker, and Sheak Rashed Haider Noori. 2024. Zero- and few-shot prompting with LLMs: A comparative study with fine-tuned models for Bangla sentiment analysis. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 17808–17818, Torino, Italia. ELRA and ICCL.

Anil Kumar Jadon and Suresh Kumar. 2023. A comparative study of cnns and dnns for emotion detection from text using tf-idf. In *2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)*, pages 1329–1334.

Ahasan Kabir, Animesh Roy, and Zaima Taheri. 2023. BEmoLexBERT: A hybrid model for multilabel textual emotion classification in Bangla by combining transformers with lexicon features. In *Proceedings of the First Workshop on Bangla Language Processing (BLP-2023)*, pages 56–61, Singapore. Association for Computational Linguistics.

Mounika Marreddy, Subba Reddy Oota, Lakshmi Sireesha Vakada, Venkata Charan Chinni, and Radhika Mamidi. 2022. Am i a resource-poor language? data sets, embeddings, models and analysis for four different nlp tasks in telugu language. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 22(1).

Shamsuddeen Hassan Muhammad, Nedjma Ousidhoum, Idris Abdulmumin, Jan Philip Wahle, Terry Ruas, Meriem Beloucif, Christine de Kock, Nirmal Surange, Daniela Teodorescu, Ibrahim Said Ahmad, David Ifeoluwa Adelani, Alham Fikri Aji, Felermino D. M. A. Ali, Ilseyar Alimova, Vladimir Araujo, Nikolay Babakov, Naomi Baes, Ana-Maria Bucur, Andiswa Bukula, Guanqun Cao, Rodrigo Tufino Cardenas, Rendi Chevi, Chiamaka Ijeoma Chukwuneke, Alexandra Ciobotaru, Daryna Dementieva, Murja Sani Gadanya, Robert Geislinger, Bela Gipp, Oumaima Hourrane, Oana Ignat, Falalu Ibrahim Lawan, Rooweither Mabuya, Rahmad Mahendra, Vukosi Marivate, Andrew Piper, Alexander Panchenko, Charles Henrique Porto Ferreira, Vitaly Protasov, Samuel Rutunda, Manish Shrivastava, Aura Cristina Udrea, Lilian Diana Awuor Wanzare, Sophie Wu, Florian Valentin Wunderlich, Hanif Muhammad Zhafran, Tianhui Zhang, Yi Zhou, and Saif M. Mohammad. 2025a. Brighter: Bridging the gap in human-annotated textual emotion recognition datasets for 28 languages. *Preprint*, arXiv:2502.11926.

Shamsuddeen Hassan Muhammad, Nedjma Ousidhoum, Idris Abdulmumin, Seid Muhie Yimam, Jan Philip Wahle, Terry Ruas, Meriem Beloucif, Christine De Kock, Tadesse Destaw Belay, Ibrahim Said Ahmad, Nirmal Surange, Daniela Teodorescu, David Ifeoluwa Adelani, Alham Fikri Aji, Felermino Ali, Vladimir Araujo, Abinew Ali Ayele, Oana Ignat, Alexander Panchenko, Yi Zhou, and Saif M. Mohammad. 2025b. SemEval task 11: Bridging the gap in text-based emotion detection. In *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)*, Vienna, Austria. Association for Computational Linguistics.

Kelechi Ogueji, Yuxin Zhu, and Jimmy Lin. 2021. Small data? no problem! exploring the viability of pretrained multilingual language models for low-resourced languages. In *Proceedings of the 1st Workshop on Multilingual Representation Learning*, pages 116–126, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Nilanjana Raychawdhary, Amit Das, Sutanu Bhattacharya, Gerry Dozier, and Cheryl D. Seals. 2024. Optimizing multilingual sentiment analysis in low-resource languages with adaptive pretraining and strategic language selection. In *2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI)*, pages 1–5.

Nilanjana Raychawdhary, Amit Das, Gerry Dozier, and Cheryl D. Seals. 2023a. Seals_Lab at SemEval-2023 task 12: Sentiment analysis for low-resource African languages, Hausa and Igbo. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 1508–1517, Toronto, Canada. Association for Computational Linguistics.

Nilanjana Raychawdhary, Nathaniel Hughes, Sutanu Bhattacharya, Gerry Dozier, and Cheryl D. Seals. 2023b. A transformer-based language model for sentiment classification and cross-linguistic generalization: Empowering low-resource african languages. In *2023 IEEE International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings)*, pages 1–5.

Saeed Sarbazi-Azad, Ahmad Akbari, and Mohsen Khazeni. 2021. Exaaec: A new multi-label emotion classification corpus in arabic tweets. In *2021 11th International Conference on Computer Engineering and Knowledge (ICCKE)*, pages 465–470.

Sanket Sonu, Rejwanul Haque, Mohammed Hasanuzzaman, Paul Stynes, and Pramod Pathak. 2022. Identifying emotions in code mixed Hindi-English tweets. In *Proceedings of the WILDRE-6 Workshop within the 13th Language Resources and Evaluation Conference*, pages 35–41, Marseille, France. European Language Resources Association.

Girija V R, Sudha T, and Riboy Cheriyan. 2023. Analysis of sentiments in low resource languages: Challenges and solutions. In *2023 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE)*, pages 1–6.

Jing Wang and Ci Zhang. 2025. Cross-modality fusion with eeg and text for enhanced emotion detection in english writing. *Frontiers in Neurorobotics*, 18.

H. Ekkehard Wolff. 2024. *Hausa language*. Encyclopedia Britannica. Accessed: 2024-10-13.