Pruning for Performance: Efficient Idiom and Metaphor Classification in Low-Resource Konkani Using mBERT

Anonymous ACL submission

Abstract

In this paper, we address the persistent challenges that figurative language expressions pose for natural language processing (NLP) systems, particularly in low-resource languages such as Konkani. We present a hybrid model that integrates a pre-trained Multilingual BERT (mBERT) with a bidirectional LSTM and a linear classifier. This architecture is fine-tuned on a newly introduced annotated dataset for metaphor classification, developed as part of this work. To improve the model's efficiency, we implement a gradient-based attention head pruning strategy. For metaphor classification, the pruned model achieves an accuracy of 78%. We also applied our pruning approach to expand on an existing idiom classification task, achieving 83% accuracy. These results demonstrate the effectiveness of attention head pruning for building efficient NLP tools in underrepresented languages.

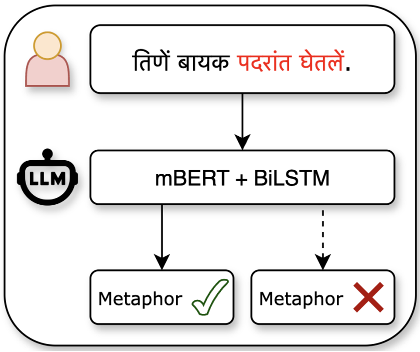

Figure 1: Processing of Konkani metaphorical expressions using mBERT+BiLSTM. The phrase highlighted in red is analyzed for metaphorical content, with contrasting classification outcomes shown.

1 Introduction

Understanding figurative language is crucial for building NLP systems that can accurately interpret meaning, support effective communication, and preserve cultural nuance (Shutova, 2015). This is especially important for low-resource languages like Konkani (Gaonkar and Fernandes, 2019). Improving NLP for Konkani not only advances linguistic research but also contributes to equitable technology access and the safeguarding of linguistic heritage (Gaonkar and Fernandes, 2019). Figurative language expressions like idioms and metaphors are common in Konkani but remain challenging for computational models (Shaikh et al., 2024). While such tasks have been explored in major languages, research on Konkani is still emerging (Naik et al., 2024; Shaikh et al., 2024). Recent work has introduced the first idiom-annotated corpus and neural models for idiom classification (Shaikh et al., 2024; Shaikh and Pawar, 2024), but these efforts are limited. They focus solely on idioms, neglect metaphor classification, and do not consider model efficiency improvements.

We present a hybrid model that integrates a pre-trained Multilingual BERT (mBERT) (Devlin et al., 2019) with a bidirectional LSTM and a linear classifier, as shown in Figure 1. This architecture is fine-tuned on an adapted version of the Konidioms corpus (Shaikh et al., 2024), which we extend to include metaphor annotations. To improve efficiency, we apply gradient-based attention head pruning. Our results show that pruning significantly reduces model complexity, with one experiment maintaining performance and the other showing a small decline. These findings demonstrate the effectiveness of pruning for building efficient, high-performing NLP models in low-resource settings.

2 Related Work

Research on low-resource languages has underscored challenges such as limited annotated data, script diversity, and dialectal variation (Rajan et al., 2020; Nigatu et al., 2024; Gaonkar and Fernandes, 2019). Konkani reflects these issues through its use of multiple scripts, dialectal fragmentation, and a

Id	Sentence instance identifier
Expression	The expression in Konkani
Sentence	Konkani sentence with the expression
Idiom	Identification tag for Idioms (Yes/No)
Metaphor	Identification tag for Metaphors (Yes/No)
Split	Data split assignment (train or test)

Table 1: Data schema for modified Konidioms Corpus.

shrinking speaker population. Prior work has addressed tasks like text summarization using a small folk tale dataset and language-independent features with pre-trained embeddings (D'Silva and Sharma, 2022), but figurative language remains largely unexplored.

Shaikh et al. (2024) introduced the first idiom-annotated corpus of 6,520 Devanagari-script sentences, and Shaikh and Pawar (2024) developed a neural classifier. Yayavaram et al. (2024) further improved idiom classification using a BERT-based model with custom loss functions. To improve model efficiency, recent studies have explored pruning redundant attention heads. Feng et al. (2018) showed that gradients can assess feature importance, and Ma et al. (2021) extended this to cross-lingual attention head pruning, a method we adopt for our multilingual, low-resource Konkani setting.

2.1 Konkani Language

Konkani is an Indo-Aryan language spoken along India's western coast, classified within the Southern Indo-Aryan Outer Languages branch alongside Marathi (Figure 3) (Rajan et al., 2020; Gaonkar and Fernandes, 2019). With approximately 2.5 million speakers (Encyclopedia Britannica, 2025) concentrated in the coastal regions of western India (Figure 4), the language faces endangerment due to dialectal fragmentation and limited digital resources, despite ongoing corpus development efforts (Gaonkar and Fernandes, 2019). This precarious situation underscores the urgency of preserving Konkani not only as a medium of communication but also as a vessel of cultural identity, as echoed by native speakers' reflections and personal narratives (Appendix B).

3 Metaphor Classification

To our knowledge, **this is the first work to introduce and utilize a metaphor-annotated dataset for the Konkani language within the NLP domain.** We extend the existing Konidioms Corpus (Shaikh et al., 2024) by manually annotating 500 sentences with binary labels indicating the presence or absence of metaphor. All annotations were verified by a native Konkani speaker to ensure linguistic accuracy. Table 1 illustrates the structure of an entry in our annotated corpus.

For our experiments, we selected a balanced subset of 200 sentences with an equal distribution of metaphorical and non-metaphorical instances (50/50 split). This was done to mitigate class imbalance and support stable and interpretable model training. We fine-tuned a multilingual BERT (mBERT) model (Devlin et al., 2019) combined with a two-layer BiLSTM (128 hidden units) using standard fine-tuning settings. Training was performed with the AdamW optimizer, a learning rate of 2×10^{-5}, batch size of 16, and a maximum input length of 128 tokens. A sigmoid-activated linear layer followed the BiLSTM to produce the final output. The model was trained using binary cross-entropy loss for up to 20 epochs, with early stopping applied if validation loss did not improve for 10 consecutive epochs. The best-performing model, selected based on minimum validation loss, balances computational efficiency and representational capacity for detecting idioms and metaphors.

We build on prior work in attention head pruning and transformer-based models, introducing key innovations for figurative language understanding in low-resource settings. We introduce the first application of attention head pruning to Konkani metaphor classification. As an additional experiment, we apply the same pruning technique to idiom classification, a task previously addressed in earlier work, to demonstrate the broader applicability of our method. A high-level overview of the methodology is illustrated in Figure 5 in Appendix C.

4 Results

The comparison between original and pruned models reveals differential impacts across the two figurative language classification tasks, as shown in Table 2. For idiom classification, pruning resulted in remarkably stable performance. The model maintained nearly identical precision and F1-score, with a slight improvement in recall and accuracy. This stability extended to macro and weighted averages across all metrics, with minimal changes observed (0.01-0.02 point differences), demonstrating that removed attention heads contributed minimally to idiom detection capabilities.

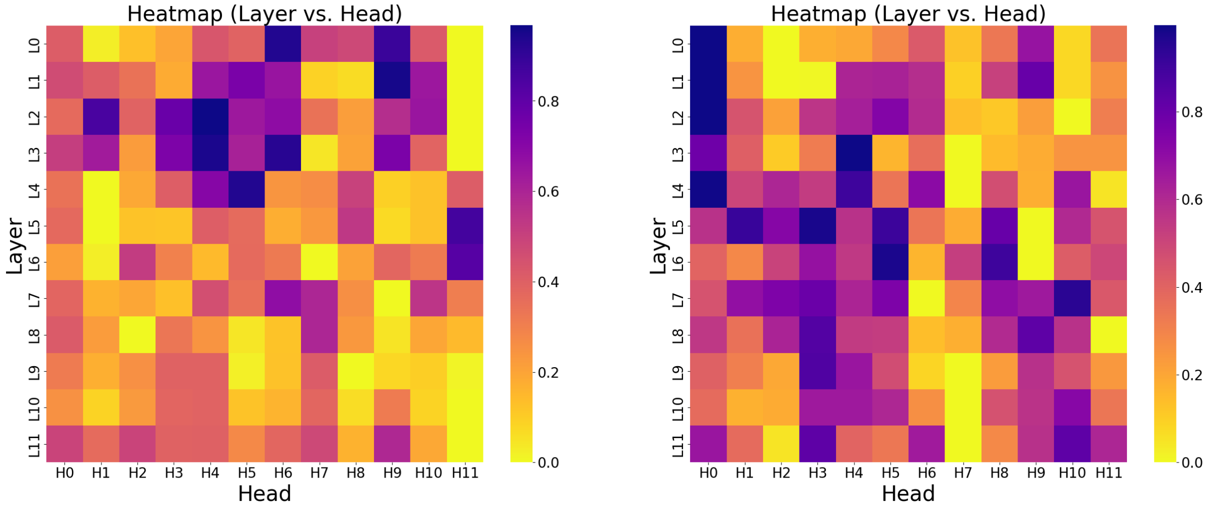

Figure 2: Heatmaps showing attention head importance scores across layers for idiom (left) and metaphor (right) classification. Idiom classification shows higher importance values in earlier layers compared to later ones, while metaphor classification exhibits a higher importance score around the center of the heatmap.

In contrast, metaphor classification exhibited greater sensitivity to pruning. The model experienced more significant decreases across evaluation metrics, with precision, recall, F1-score, and accuracy all showing a noticeable decrease. The drop in precision and recall contributed to a lower F1-score, while overall classification accuracy also declined. Despite this reduction, the pruned model maintained relatively balanced precision and recall values, indicating consistent behavior across figurative and literal classes even after pruning. The macro and weighted average metrics showed similar declines of approximately 0.10 points. This performance pattern is particularly notable given the low-resource nature of the metaphor dataset.

5 Attention Head Analysis

We prune attention heads in the mBERT component of the mBERT+BiLSTM model using a gradient-based importance metric (Michel et al., 2019). This metric quantifies each head's contribution by calculating the expected sensitivity of the model loss to the head's removal, expressed as $I_h = \mathbb{E}_{(x,y)\sim D}\left|\frac{\partial L}{\partial \mathbf{h}^{(h)}}\right|$, where I_h is the importance score for head h, (x, y) represents input-output pairs from dataset D, L is the loss, and $\mathbf{h}^{(h)}$ is the output of attention head h. For each of the 144 heads (12 layers $\times$ 12 heads), we compute the average absolute gradient of the loss with respect to the head's output. Heads with scores of zero were pruned post hoc, with no changes to the BiLSTM.

For both idiom and metaphor classification tasks, we pruned all attention heads that had an importance score of zero, resulting in 132 of 144 heads being retained (12 heads pruned) for both tasks. The attention head maps can be seen in Figure 2. By eliminating these attention heads with zero importance scores across both tasks, we create two pruned variants of the original model. These pruned models are evaluated and compared against the baseline. These results are presented in Table 2.

5.1 Head-Level Performance

Figure 2 visualizes the distribution of attention head importance for both idiom and metaphor classification tasks. For idiom classification, importance tends to cluster in the lower layers (**L0–L6**), with particularly prominent heads such as **L0-H6** and **L1-H9** standing out as key contributors. These heads likely encode lexical or syntactic patterns crucial for identifying idiomatic usage. In contrast, metaphor classification exhibits a more diffuse pattern of importance, with salient heads scattered across all layers. This broader distribution suggests that metaphor detection may require integrating cues from multiple linguistic levels. Despite some variation, both tasks reveal consistent retention of highly informative heads, supporting the effectiveness of selective pruning in reducing model complexity without compromising performance.

The contrasting patterns observed in the two classification tasks, suggests fundamental differences in how these separate linguistic classification problems are processed within the transformer's

Metric	Idiom Classification		Metaphor Classification	
	Original Model	Pruned Model	Original Model	Pruned Model
Precision	0.87	0.86	1.00	0.87
Recall	0.89	0.91	0.75	0.65
F1-Score	0.88	0.88	0.86	0.74
Accuracy	0.82	0.83	0.88	0.78
Macro Avg Precision	0.78	0.79	0.90	0.79
Macro Avg Recall	0.77	0.77	0.88	0.78
Weighted Avg Precision	0.82	0.82	0.90	0.79
Weighted Avg Recall	0.82	0.83	0.88	0.78

Table 2: Comparison of original and pruned mBERT+BiLSTM models on idiom and metaphor classification. Idiom performance remains stable post-pruning, while metaphor classification shows metric drops, reflecting its reliance on a broader set of attention heads and the need for task-specific pruning strategies.

attention mechanism. Full detailed heatmaps for idiom and metaphor classification can be found in Appendix C (Figure 6 and Figure 7 respectively).

6 Discussion

The heatmaps in Figure 2 provide critical insights into why pruning affects idiom and metaphor classification so differently. Idiom classification shows higher importance values concentrated in earlier layers, creating natural redundancy that allows the model to maintain performance even when less important heads are removed. In contrast, metaphor classification exhibits a more distributed pattern with importance centered in the middle layers, making it more vulnerable to pruning operations.

This architectural difference explains the divergent responses observed in our experiments. While idiom classification maintained stable metrics after pruning, with some measures even showing slight improvement, **metaphor classification experienced substantial performance degradation across all evaluation metrics**. This suggests that metaphor detection relies on a more complex, interconnected network of attention heads that cannot be easily reduced without compromising functionality.

These findings have significant implications for deploying transformer models in resource-constrained environments. They indicate that pruning strategies should be task-specific rather than universal. For idiom classification, pruning appears viable without significant performance costs, while metaphor classification requires a more conservative approach that preserves the distributed processing network.

Future work should explore adaptive pruning methodologies that account for these task-specific architectural requirements, potentially enabling more efficient deployment for figurative language processing across diverse linguistic contexts. In particular, future work could also involve experiments testing different thresholds for pruning attention heads to better understand their impact on task performance and model efficiency. A critical direction would be expanding the dataset. Larger datasets would reduce overfitting risks and improve the model's ability to handle real-world variability. This combined approach of improved pruning strategies and expanded data resources would support the development of more efficient, task-specific compression techniques for figurative language processing.

7 Conclusion

We introduce the first metaphor-annotated dataset for Konkani and apply a unified framework for idiom and metaphor classification in a low-resource setting. By extending the Konidioms corpus and fine-tuning a hybrid mBERT+BiLSTM model, we establish strong baselines for figurative language understanding. Gradient-based attention head pruning reveals structural differences: idioms rely on localized, lower-layer heads, while metaphors engage a more diffuse attention profile. As a result, idiom classification remains robust under pruning, whereas metaphor performance is more sensitive to head removal. Our work advances interpretable NLP for underrepresented languages. We release our dataset and pruning framework to support future research in figurative language modeling, model compression, and multilingual generalization.

Limitations

This study is limited by several key factors. Although the metaphor classification dataset includes 500 newly annotated data points, our experiment utilized only 200 balanced sentences, which limits the generalizability of our results and highlights the need for broader evaluation in future work. Although we verified annotations with a native Konkani speaker, the small number of validators introduces potential subjective bias in the labeling process. The corpus itself may not capture the full range of figurative expressions or dialectal variations present in Konkani, affecting model performance across different speaker communities. Our pruning approach, while effective for our experiments, employed fixed thresholds that may not transfer optimally to other tasks or datasets. Finally, evaluation on a single test split necessitates further validation with more diverse data to confirm the robustness of our findings across different contexts.

Ethics Statement

Our research addresses the technological gap between high and low-resource languages while recognizing the ethical responsibilities inherent in working with Konkani, an endangered language. We engaged native speakers throughout the annotation and verification process to ensure linguistic accuracy and cultural sensitivity. This work contributes to preserving Konkani's cultural heritage by documenting and enabling computational processing of its figurative expressions. The resources we have developed are intended to serve both the Konkani-speaking community and researchers working on low-resource language technologies. We have maintained transparency about our limitations to prevent misrepresentation of capabilities, and our pruning approach specifically addresses accessibility in resource-constrained environments. By balancing our dataset and committing to continued community engagement, we aim to support linguistic diversity and ensure all languages receive technological support that preserves their unique characteristics in digital spaces. In the spirit of transparency, our code is made publicly available in an anonymous repository at https://anonymous.4open.science/r/KonkaniNLP.

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Jovi D'Silva and Uzzal Sharma. 2022. Automatic text summarization of konkani texts using pre-trained word embeddings and deep learning. *International Journal of Electrical and Computer Engineering (IJECE)*, 12:1990.

Encyclopedia Britannica. 2025. Konkani language. Encyclopedia Britannica. Accessed May 11, 2025.

Shi Feng, Eric Wallace, Alvin Grissom II, Mohit Iyyer, Pedro Rodriguez, and Jordan Boyd-Graber. 2018. Pathologies of neural models make interpretations difficult. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Palia Tukaram Gaonkar and Andre Rafael Fernandes. 2019. Digitization of Konkani Texts, and their Transliteration: An Initiative towards Preservation of a Language Culture. *CEUR Workshop Proceedings*, 2364:110–117.

Weicheng Ma, Kai Zhang, Renze Lou, Lili Wang, and Soroush Vosoughi. 2021. Contributions of transformer attention heads in multi- and cross-lingual tasks. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, page 1956–1966. Association for Computational Linguistics.

Paul Michel, Omer Levy, and Graham Neubig. 2019. Are sixteen heads really better than one? In *Advances in Neural Information Processing Systems*, volume 32, pages 14014–14024. Curran Associates, Inc.

Pratik Naik, Nilesh Kamat, Shweta Naik, Prashant Naik, and Rajesh Kamat. 2024. Konidioms corpus: A dataset of idioms in konkani language. In *Proceedings of the 2024 International Conference on Language Resources and Evaluation (LREC)*, pages 7857–7866.

Y. Nigatu, I. D. Raji, M. Choudhury, S. Diddee, G. Le Ferrand, J. Dearden, and A. Tucker. 2024. The zeno's paradox of 'low-resource' languages. ArXiv preprint arXiv:2410.20817.

Annie Rajan, Ambuja Salgaonkar, and Ramprasad Joshi. 2020. A survey of konkani nlp resources. *Computer Science Review*, 38:100299.

Naziya Mahamdul Shaikh and Jyoti Pawar. 2024. Identification of idiomatic expressions in Konkani language using neural networks. In *Proceedings of the 21st International Conference on Natural Language Processing (ICON)*, pages 54–58, AU-KBC Research Centre, Chennai, India. NLP Association of India (NLPAI).

Naziya Mahamdul Shaikh, Jyoti D. Pawar, and Mubarak Banu Sayed. 2024. Konidioms corpus: A dataset of idioms in Konkani language. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 9932–9940, Torino, Italia. ELRA and ICCL.

Ekaterina Shutova. 2015. Design and evaluation of metaphor processing systems. *Computational Linguistics*, 41(4):579–623.

Arnav Yayavaram, Siddharth Yayavaram, Prajna Devi Upadhyay, and Apurba Das. 2024. BERT-based idiom identification using language translation and word cohesion. In *Proceedings of the Joint Workshop on Multiword Expressions and Universal Dependencies (MWE-UD) @ LREC-COLING 2024*, pages 220–230, Torino, Italia. ELRA and ICCL.

A Appendix A

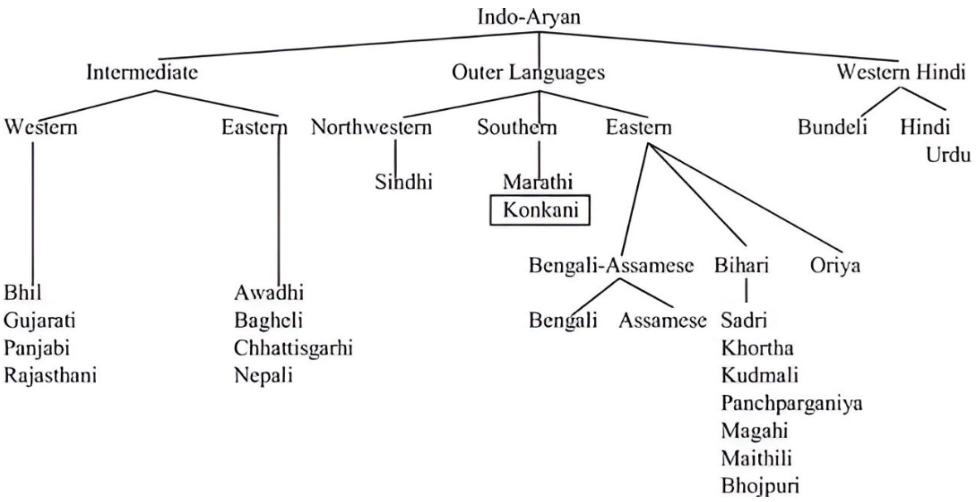

Figure 3: Linguistic tree showing Konkani's classification as a Southern language within the Indo-Aryan Outer Languages branch, alongside Marathi and distinct from other major Indo-Aryan language groups.

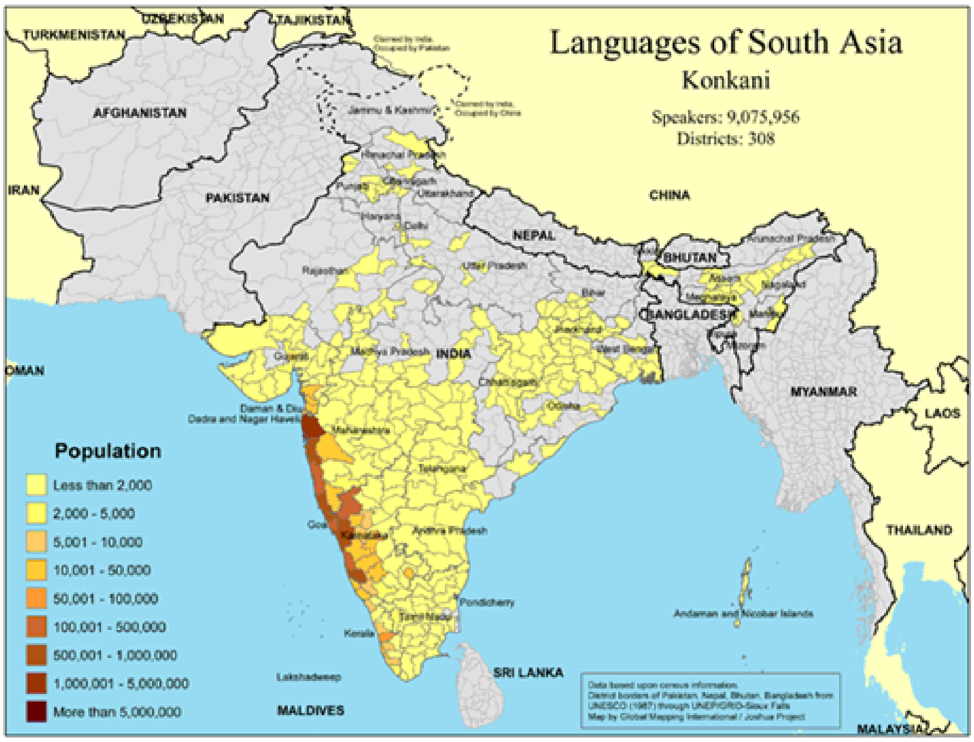

Figure 4: Geographic distribution of Konkani speakers across South Asia, concentrated along India's western coastal regions. As of 2018, approximately 9 million speakers were recorded across 308 districts. Source: https://www.missioninfobank.org/mib/index.php?main_page=product_info&products_id=6368

B Appendix B

B.1 Perspectives from a Native Konkani Speaker

As part of this work, we solicited reflections from a native Konkani speaker regarding the digital and computational underrepresentation of the language. The following excerpt is shared with permission and reflects the perspective of a native speaker from Goa:

> "As a native Konkani speaker from Goa, I find it deeply concerning that Konkani remains a low-resource language in the digital world today. Although spoken by hundreds of thousands and recognized as one of India's official languages, Konkani lacks the technological and academic investment that the more dominant languages receive. This underrepresentation threatens the long-term vitality of our language, culture, and identity.
>
> Languages like Konkani are not just modes of communication, they are carriers of unique histories, worldviews, and traditions. When they are ignored by major platforms, AI models, and digital tools, it sends the message that these voices matter less. But they do matter.
>
> I believe that it is our responsibility as speakers, researchers, and technologists to change that. Supporting Konkani through language research, resource development, and digital inclusion is not just about preserving a language. It's about empowering a community."

— Native Konkani speaker from Goa

B.2 In Memory of a Monolingual Konkani Speaker

This project is motivated in part by the memory of a monolingual speaker of Konkani whose life, conversations, and cultural expressions were deeply rooted in the language. His use of idioms and metaphors exemplified the richness and complexity of Konkani, elements that are often difficult to preserve or translate into other languages.

His recent passing highlights the urgency of documenting and understanding low resource languages like Konkani, not only from a linguistic perspective, but also as a means of preserving cultural and emotional heritage. This research, particularly its focus on idiomatic and metaphorical structures, reflects a commitment to honoring such speakers and the languages they embody.

We hope that advancements in AI models capable of capturing linguistic nuance may one day help reflect not just the syntax, but the soul of languages like Konkani.

C Appendix C

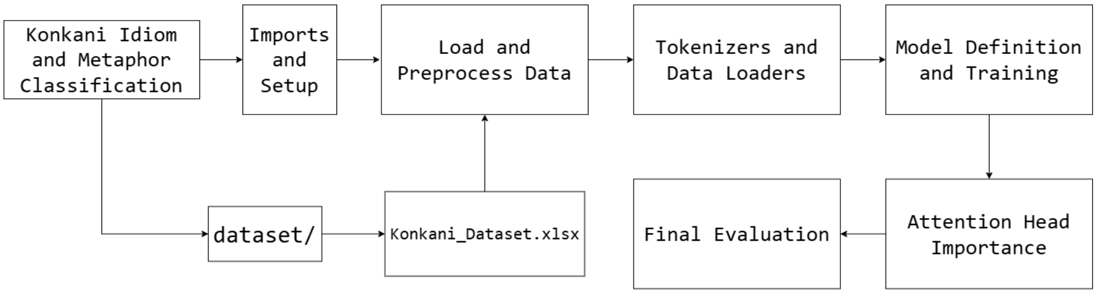

Figure 5: Flowchart outlining our experimental pipeline.

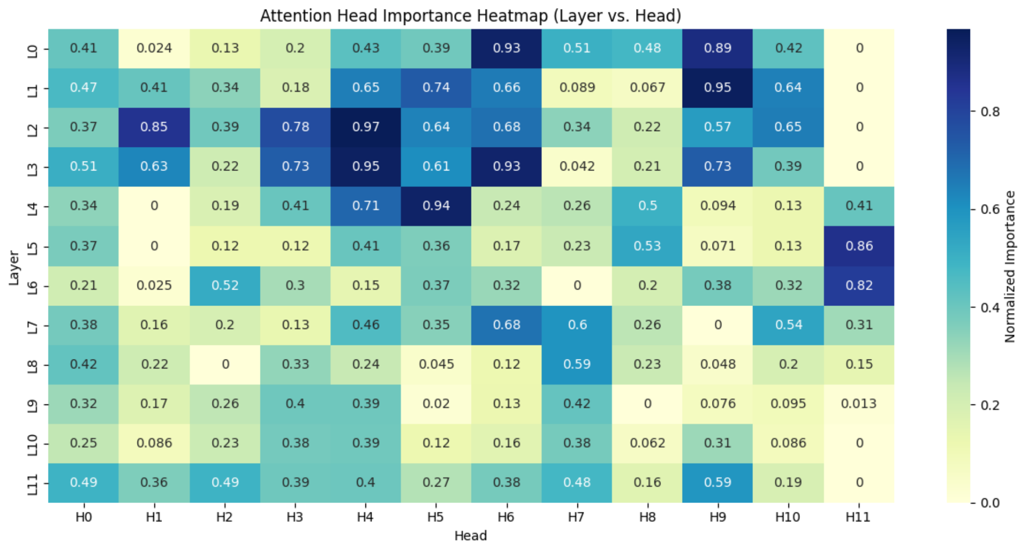

Figure 6: Heatmap visualization of attention head importance across model layers for idiom classification, with numerical decimal values displayed to facilitate detailed quantitative analysis.

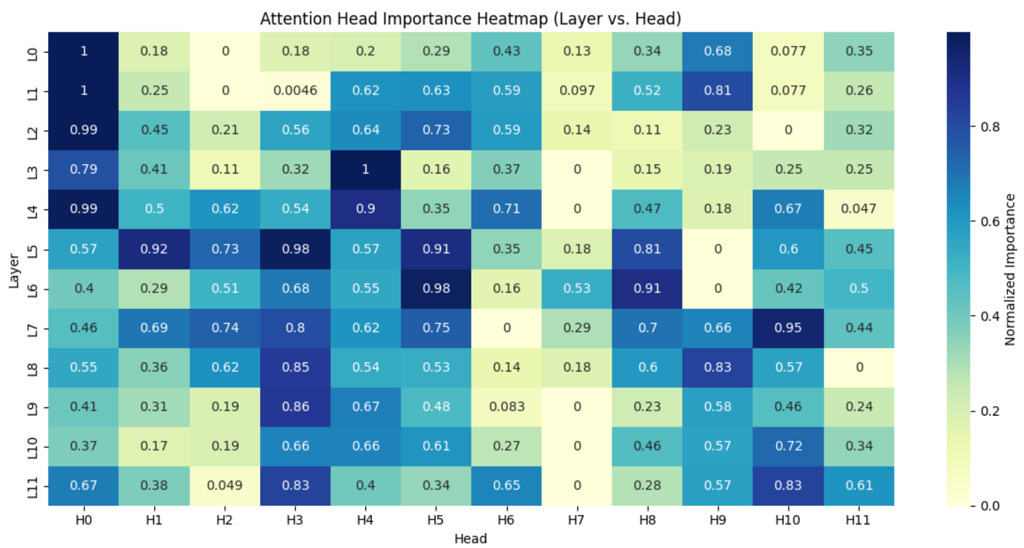

Figure 7: Heatmap visualization of attention head importance across model layers for metaphor classification, with numerical decimal values displayed to facilitate detailed quantitative analysis.