ERROR_500@DravidianLangTech2026: Automatic Prompt Style Classification in Telugu Using Transformer-Based Language Models

Mahashweta Manjari Barua, Tasnia Khanam, Nuzha Saifa Rahmat, Shiti Chowdhury, Hasan Murad
Department of Computer Science and Engineering
Chittagong University of Engineering and Technology, Bangladesh
{u2104042, u2104065, u2104098, u2004027}@student.cuet.ac.bd
hasanmurad@cuet.ac.bd

Abstract

Prompt style recovery in low-resource languages has been daunting due to diverse morphology, culturally specific language patterns and scarce annotated data. Prior works have largely focused on binary sentiment or single-attribute transfer, multi-class style classification for languages like Telugu has remained underexplored. We have addressed this chasm here through the Telugu Prompt-Style Recovery Shared Task at DravidianLangTech@ACL 2026 (Premjith et al. (2026)), framing it as a nine class classification problem. We have evaluated three input configurations–Change Style, Original Transcripts and Merged input style and also trained three transformer-based models-MuRIL, XLM-RoBERTa and IndicBERT v2 under identical conditions. Our best model, IndicBERT v2 with partial layer freezing and weighted cross-entropy loss, has obtained a macro-F1 of 0.2987 and accuracy of 0.299. Change Style has significantly outperformed other inputs, indicating that explicit style changes enhance tonal and meaning cues. These results have highlighted the value of language-specific pretraining and careful input design for style-sensitive NLP in low-resource settings, securing 1st rank on the shared task.

1 Introduction

Text style has conveyed meaning beyond literal content through tone, register and rhetorical intent. In low-resource languages like Telugu, modeling such variation has been taxing due to rich morphology and culture-specific communication patterns (Mukherjee et al. (2024a)). Although previous work on prompt recovery and style reconstruction has shown that pragmatic intent can be inferred from text (Liu et al. (2024)), most studies have focused on binary or single-attribute transformations, falling short on fully exploring multi-class style classification for Telugu due to limited resources.

In this study, we have proposed a transformer-based framework for Telugu writing style prompt recovery. Using IndicBERT v2 with Change Style as input, our system has obtained a macro-F1 of 0.3648 and an accuracy of 0.3700, performing better than XLM-RoBERTa and MuRIL across all three input configurations. Our contributions are as follows:

- We have developed an IndicBERT v2-based framework that achieves state-of-the-art performance.
- We have analyzed and shown that Change Style text provides the strongest signal.
- We have compared multilingual transformers for low-resource Telugu style classification.

Implementation details and code have been made available at: https://github.com/Tasni-a/Prompt_Recovery_LLM_ST

2 Related Work

Prompt recovery techniques have been essential for reverse-engineering stylistic instructions from LLM outputs. Liu et al. (2024) introduced StyleRec, a benchmark for reconstructing prompts in style transformation tasks, modeling tones relevant to Telugu style classification. Li and Klabjan (2025) proposed a training-free Reverse Prompt Engineering framework using semantic similarity and chain-of-thought decoding for recovering intents like authoritative and optimistic framing. Telugu-focused pretrained models by Niyogi et al. (2026) have achieved strong zero-shot performance, though modeling Dravidian rhetorical nuances remain challenging. DravidianLangTech shared tasks and Blevins et al. (2023) has helped to advance Telugu stylistic analysis, while Mukherjee et al. (2024b) has shown

Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 253–257

LLM inconsistencies in multi-attribute style control, motivating the use of macro-F1 across nine categories. Current existing work largely focuses on binary or sentiment-level styles, leaving multi-category Telegu benchmarks overlooked. The DravidianLangTech@ACL 2026 shared task (Premjith et al. (2026)) has addressed this gap with curated datasets, cross-domain evaluation and strong standardized benchmarks.

3 Data Description

The dataset contains 3601 instances split into Train(3000), Development(300) and Test(301) sets, each having an ID, Original Transcript, Change Style text and one of nine style labels - provided only for Train and Development sets.

Style	Training	Development	Test
Formal	327	36	38
Informal	321	47	35
Optimistic	331	29	33
Pessimistic	347	29	22
Humorous	344	27	31
Serious	324	35	47
Inspiring	332	33	37
Authoritative	338	33	30
Persuasive	336	31	28
Total	**3,000**	**300**	**301**

Table 1: Distribution of Training, Development and Test Data

4 Methodology

4.1 Problem Formulation

The task is to recover in which style a given Telugu text has been written. Given a Change Style text x, the goal is to predict its Style label. Given a stylized Telugu input x, the goal is to predict its style label $\hat{y} = f_\theta(x)$, where $f_\theta : \mathcal{X} \to S$ maps inputs to K discrete styles $S = \{s_1, \ldots, s_K\}$ over a dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$.

4.2 Data Preprocessing

Style labels have been encoded using LabelEncoder. Inputs have been tokenized, truncated to 256 tokens and dynamically padded via DataCollatorWithPadding. 256 tokens limit has been chosen to prevent overfitting, reduce GPU consumption and improve training efficiency. Preliminary analysis confirmed that most Change Style text fit within 256 tokens, preserving full context. The slight class imbalance has been addressed using per-class weights computed via compute_class_weight and applied in cross-entropy loss.

4.3 Model Architecture

4.3.1 Encoder

IndicBERT v2 (`ai4bharat/IndicBERTv2-MLM-only`) has been used as the encoder which has been pretrained on 23 Indian languages including Telugu. The first 6 layers have been frozen and the next 6 layers remaining trainable. The final [CLS] token (768-dim) has been extracted for the classification head.

4.3.2 Classification Head and Training Objective

The [CLS] representation has been fed into a linear classification head to produce raw scores over all style categories. The model has been trained using weighted cross-entropy loss, per-class weights to address class imbalance and model parameters have been optimized with AdamW.

4.4 Input Feature Selection

Three configurations have been evaluated: (1) Change Style, (2) Original Transcripts and (3) Merged input and have selected Change Style as the input due to itsclass separating features, as shown in Figure 1.

4.5 Model Selection

Three multilingual transformers have been compared using Change Style input: IndicBERT v2, MuRIL (`google/muril-base-cased`) and XLM-RoBERTa (`xlm-roberta-base`). All models have shared identical architectures, training settings and hyperparameters to guarantee a fair evaluation.

4.6 Evaluation Metrics

All the models have been evaluated using macro-F1, accuracy, precision and recall.

5 Results and Analysis

5.1 Input Feature Comparison

Table 2 shows the macro-averaged performance of IndicBERT v2 across different input configurations. As Change Style has surpassed Original Transcripts with a macro-F1 of 0.0715 and Merged input with 0.0656, it has been used as input in all experiments.

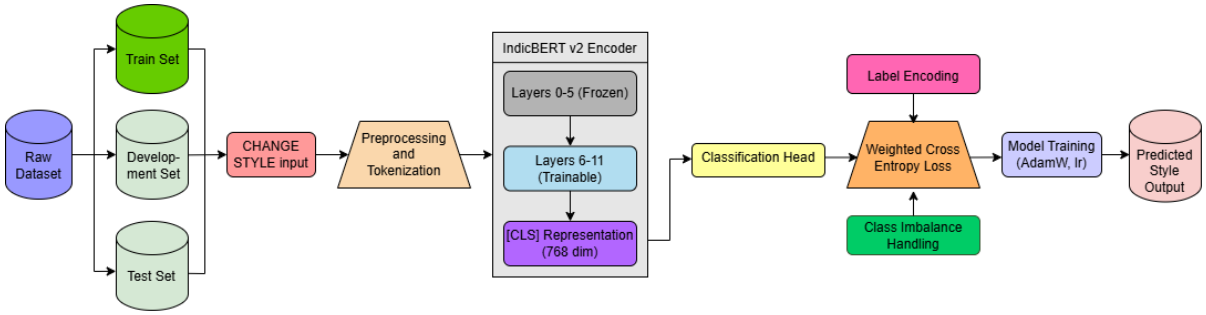

Figure 1: System architecture for Telugu prompt-style recovery.

Input Feature	F1	R	P	A
Change Style	**0.3648**	**0.3667**	**0.3691**	**0.3700**
Original Transcripts	0.0715	0.1088	0.1032	0.0967
Merged (OT+CS)	0.0656	0.1089	0.0541	0.1000

Table 2: IndicBERT v2 performance across different input features using 256 tokens

The Original Transcripts perform below the random baseline, likely due to truncation of long texts, distributed stylistic patterns and Telugu–-English code-switching noise. Merged Input further degrades performance, proving that the original text introduces noise.

The better performance of Change Style input suggests that lingual transformations enhance tonal and communicative markers, enabling more distinct style separation by displaying clear lexical and grammatical patterns. The Merged input has degraded results by introducing noise from the original text. Highlighting stylistic features has strengthened model discrimination, especially for low-resource language like Telugu. Post competition experiments with 512 tokens instead of 256 tokens validated this claim as only Change Style has shown marginal improvement, increasing macro-F1 to 0.3831. These results have confirmed that 256 tokens were a better choice for this task.

5.2 Model Comparison

Table 3 compares three transformer models with Change Style as input. IndicBERT v2 has achieved the highest macro-F1 of 0.3648, followed by XLM-R at 0.3414, while MuRIL yields the lowest scores at 0.2276. XLM-R has remained competitive, whereas MuRIL has failed on Optimistic and Serious. IndicBERT v2 has performed its best on Pessimistic and Informal, whereas Humorous, Serious and Persuasive remain the most challenging style classes.

Model	F1	R	P	A
IndicBERT v2	**0.3648**	**0.3667**	**0.3691**	**0.3700**
XLM-RoBERTa	0.3414	0.3433	0.3411	0.3400
MuRIL	0.2276	0.2742	0.2426	0.2800

Table 3: Comparison of three transformer models-IndicBERT, XLM-RoBERTa, MuRIL on Change Style as input

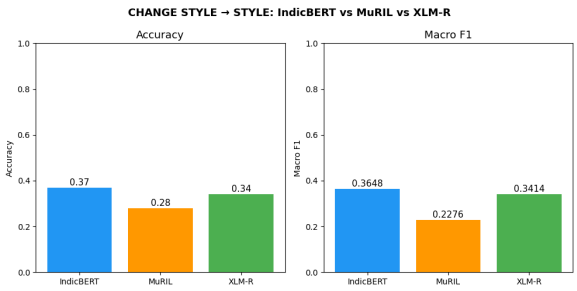

Figure 2: Macro F1 and Accuracy of three transformer models-IndicBERT v2, MuRIL, XLM-RoBERTa

The confusion matrix in Figure 3 has shown that Pessimistic and Informal styles have been most reliably classified, while Serious, Persuasive, Formal and Authoritative have shown higher confusion, due to the overlapping stylistic traits in Telugu.

5.3 Parameter Settings

All experiments have used hyperparameters as shown in Table 4:

5.4 Reproducibility Note

The submitted csv (Run 3) has achieved Rank 1 without any fixed seed, so results in Tables 2 and 3 may slightly differ from the official leaderboard scores. The results can be reproduced using the hyperparamters mentioned in Table 4 and a seed of 42.

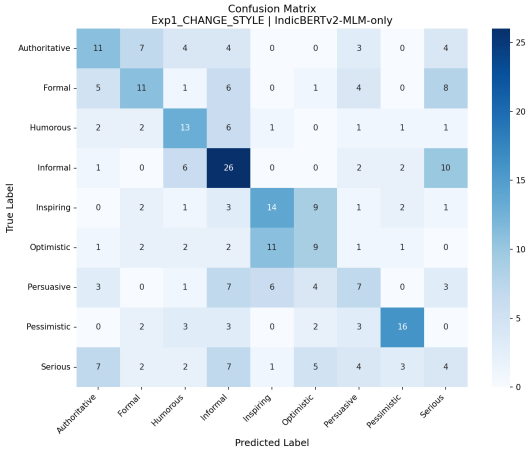

Figure 3: Confusion Matrix of IndicBERT v2

Training Hyperparameter	Configured Settings
Optimizer	AdamW
Learning rate	1.5e-5
Weight decay	0.01
Epochs	12
Batch size	16
Max sequence length	256
Frozen layers	0–5
Selection metric	Macro-F1
Random seed	42

Table 4: Training configuration.

6 Conclusion

The DravidianLangTech@ACL 2026 Shared Task highlights the challenges of detecting prompt style written in low-resource languages like Telugu. Changed style text has proved to be more effective in capturing tonal and contextual signals. IndicBERT v2 has performed exceptionally over other multilingual models like XLM-R and MuRIL. Weighted cross-entropy and partial layer freezing have addressed the slight class imbalance and overfitting. These results show the significance of language-specific pretraining and careful input selection.

Error Analysis

Figure 3 shows Formal, Authoritative and Serious as the most frequently confused pairs. Table 5 shows representative misclassified examples. These styles share similar features — formal greetings (హలో, నమస్కారం), directive phrasing (గమనించండి, చూడాలి) and formal vocabulary — making them difficult to distinguish. The model has shown bias towards predicting the Authoritative class, showing strong overlap with Formal and Serious styles in the training data

Truncated Change Style text	True	Predicted
అందరికీ నమస్కారం. ఈరోజు, టెక్ ట్రావెల్ తెలుగు ద్వారా ఒక సమగ్ర నివేదికను...	Formal	Authoritative
శ్రద్ధ వహించండి. డబ్బు సంపాదించాలనే నిజమైన ఆకాంక్ష ఉన్నవారు మాత్రమే ఈ...	Serious	Authoritative
హలో, వెల్కమ్ టు తెలుగు వన్ అకాడమీ. ఈరోజు మనం జాగ్రఫీలో సహజ వృక్ష...	Authoritative	Serious

Table 5: Representative misclassified examples from IndicBERT v2 on the validation set.

Limitations

We have fixed the learning rate at 1.5e-5 and frozen first 6 layers without extensive tuning, so better configurations may exist. Evaluation of single 300 split instances and limited metrics have restricted vigor and generalization.

Acknowledgement

We thank the organizers of Prompt Recovery for LLM in Telugu – DravidianLangTech@ACL 2026 Premjith et al. (2026) for the dataset and evaluation framework. We also acknowledge multilingual pretrained models - IndicBERT v2, MuRIL and XLM-R. We used Claude and ChatGPT as AI-assisted writing tools for refining portions of this manuscript. All technical content, experiments and results were conducted and verified by us.

Ethical Statement

The dataset has been provided by the shared task organizers. As style annotation may reflect cultural assumptions, our model's decisions may not generalize equally across all Telugu dialects.

References

Terra Blevins, Hila Gonen, and Luke Zettlemoyer. 2023. Prompting language models for linguistic structure. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6649–6663, Toronto, Canada. Association for Computational Linguistics.

Hanqing Li and Diego Klabjan. 2025. Reverse prompt engineering. *Preprint*, arXiv:2411.06729.

Shenyang Liu, Yang Gao, Shaoyan Zhai, and Liqiang Wang. 2024. Stylerec: A benchmark dataset for prompt recovery in writing style transformation. In *2024 IEEE International Conference on Big Data (BigData)*, pages 1678–1685.

Sourabrata Mukherjee, Atul Kr. Ojha, Akanksha Bansal, Deepak Alok, John P. McCrae, and Ondřej Dušek. 2024a. Multilingual text style transfer: Datasets & models for Indian languages. In *Proceedings of the 17th International Natural Language Generation Conference*, pages 494–522, Tokyo, Japan. Association for Computational Linguistics.

Sourabrata Mukherjee, Atul Kr. Ojha, and Ondrej Dusek. 2024b. Are large language models actually good at text style transfer? In *Proceedings of the 17th International Natural Language Generation Conference*, pages 523–539, Tokyo, Japan. Association for Computational Linguistics.

Mitodru Niyogi, Eric Gaussier, and Arnab Bhattacharya. 2026. Paramanu: Compact and competitive monolingual language models for low-resource morphologically rich indian languages. *Preprint*, arXiv:2401.18034.

B. Premjith, G. Jyothish Lal, Bharathi Raja Chakravarthi, Saranya Rajiakodi, Thenmozhi Durairaj, Ratnavel Rajalakshmi, Rahul Ponnusamy, and Bhuvanesh Chinthala. 2026. Shared task on prompt recovery for llms in telugu. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.