Still Loading@DravidianLangTech 2026: Telugu Prompt-Style Recovery using Multilingual Transformers

Samonwita Sarker, Isnat Mehrin Sami, Priyontee Mojumder
Arpita Mallik, Hasan Murad
Department of Computer Science and Engineering
Chittagong University of Engineering and Technology, Bangladesh
sarker.samonwita@gmail.com, u2104038@student.cuet.ac.bd,
u2104010@student.cuet.ac.bd, u2104012@student.cuet.ac.bd,
u2004023@student.cuet.ac.bd, hasanmurad@cuet.ac.bd

Abstract

This paper describes the system that our *Still Loading* team designed to run the Telugu Prompt-Style Recovery shared task at DravidianLangTech@ACL 2026. The purpose of the given task is categorizing Telugu transcript passages as belonging to one of 9 communicative styles: *Formal*, *Informal*, *Optimistic*, *Pessimistic*, *Humorous*, *Serious*, *Inspiring*, *Authoritative*, and *Persuasive*. We compared several multilingual Transformer-based models, i.e. MuRIL, XLM-RoBERTa-Large, mBERT, and IndicBERTv2. We chose a "Turbo Sandwich" preprocessing strategy which helps to give more emphasis to lexical deltas, in addition to Focal Loss. Our system based on the MuRIL was rated at the 7th place in the official leaderboard with a Macro-F1 rating of 0.1703, while post-submission experiments show that IndicBERTv2 achieves 0.3000 under identical conditions. The source code to reproduce our experiments is publicly available on Github[1]

1 Introduction

The most recent engineering advancement in Transformer architecture platforms has brought about incredible strides in natural language understanding (NLU). The languages with low resources, such as the Telugu language, are often paralyzed by the lack of properly annotated data on more complex tasks, such as style classification, which makes it difficult to develop an effective computational model. The Telugu Prompt-Style Recovery shared task, planned by DravidianLangTech@ACL2026, aims to fill this gap, providing a set of standard data and a reference point when it comes to the systematic determination of communicative styles in Telugu text.(DravidianLangTech Organizing Committee, 2024).

The categorization of the stylistic overtones requires a representation of subtle tonal heuristics and pragmatic expectations, which are invaluable in personalized conversational agents and tone-uniform generation of prompts, to content regulation. The assignment outlines nine stylistic classes including *Formal*, *Informal*, *Optimistic*, *Pessimistic*, *Humorous*, *Serious*, *Inspiring*, *Authoritative* and *Persuasive*, each portraying a different tonal or pragmatic focus.

Label ambiguity between similar styles (e.g., *Pessimistic* and *Serious*), the small corpus of annotations ($3,000$ training instances) and the agglutinative morphology of Telugu and code-mixing of the language are some of the fundamental problems involved.

To find our way through these barriers, we do a comparative analysis of multilingual Transformer architectures, which finally leads to the choice of the leading model, which is MuRIL. The main achievements of this research project are as follows:

- A comparative study of multilingual Transformer-based models on Telugu prompt-style classification, and finding IndicBERTv2 are the best for this task.
- A new "Turbo Sandwich" input approach that exploits sequence matching to identify and evaluate lexical differences between original and modified text, allowing the encoder to focus on stylistic edits.
- Focal Loss with square-root weight dampening to encourage equal learning of the nine stylistic categories, dampening the impact of confusable class pairs.

2 Related Work

Telugu NLP has progressed from rule-based systems to Transformer-based architectures. Vemula et al. (2022) introduced TeQuAD, highlighting data

[1] Still-Loading-Prompt-Recovery-for-LLM-in-Telugu:https://github.com/Priyontee1713/Still-Loading-Prompt-Recovery-for-LLM-in-Telugu

Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 371–375

scarcity in Dravidian languages, while sub-word modelling has shown value for morphologically rich languages (Bhattacharyya and Bhattacharya, 2025) and multilingual pretraining has benefited offensive language detection (B et al., 2024). Models such as XLM-RoBERTa (Conneau et al., 2020), MuRIL (Khanuja et al., 2021), and IndicBERTv2 (Kakwani et al., 2020) bridge the resource gap through cross-lingual and Indic-specific pretraining, with prompt-based fine-tuning showing further promise (Ullah et al., 2025).

Prompt engineering and recovery have become active research areas. Carlini et al. (2021) showed that inputs can be recovered from model outputs, while Lester et al. (2021) demonstrated prompt tuning as an effective low-resource strategy. Surveys by Sahoo et al. (2024) and Vatsal et al. (2025) catalogue techniques across multilingual settings, and Lewis et al. (2020) illustrates how attending to specific input segments improves task-focused classification, a principle central to our Turbo Sandwich strategy.

3 Data

This paper uses the Telugu language dataset provided by the Shared Task on Prompt Recovery at DravidianLangTech@ACL 2026 (DravidianLangTech Organizing Committee, 2024). The dataset includes Telugu transcript excerpts consisting of an original transcript and its stylistically modified version, each annotated with a target communicative style. The data are partitioned into training, development and test sets with $3,000$, 300 and 301 samples respectively, with near-uniform class balance across all nine styles (321–347 instances per class; see Table 1).

Style	Train	Dev	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: Label distribution across the training, development, and test splits.

4 Methodology

We build our system upon the MuRIL (Multilingual Representations for Indian Languages) model for sequence classification. An overview of hyperparameters is provided in Table 2.

Hyperparameter	Value
Base Model	MuRIL (*base-cased*)
Max Sequence Length	256
Batch Size	8
Learning Rate	4e-5
Optimizer	AdamW
Scheduler	Cosine
Warmup Ratio	0.1
Weight Decay	0.05
Epochs	15
Focal Loss (γ)	3
FP16 Precision	True

Table 2: Hyperparameter configuration for the MuRIL training pipeline.

4.1 Data Pre-processing: Turbo Sandwich Strategy

We **propose** the "Turbo Sandwich" strategy as a **novel** preprocessing formulation specifically designed to isolate stylistic changes between an original and a modified text. Rather than feeding the full modified text directly to the encoder, our pipeline applies a word-level sequence-matching approach using the `difflib` library to identify added or substituted tokens. As illustrated in Figure 1, these lexical deltas are then formatted alongside their immediate context, constructing the input as:

```
Delta: {added_text} | Context: {context_text}
```

This representation forces the Transformer encoder to attend specifically to the stylistic edits rather than treating the full sequence uniformly. To validate the contribution of this strategy, we compare it against a plain baseline (concatenating original and modified text with a separator), as reported in Table 4.

4.2 MuRIL Architecture

We optimize the MuRIL (`google/muril-base-cased`) model for sequence classification. Pre-trained on 17 Indian languages with millions of translations and transliterations, MuRIL captures phonetic, morphological, and script-level peculiarities of Telugu, enabling a linear classification head over the pooled representation to learn the 9 styles.

4.3 Training Configuration and Objectives

The training hyperparameters are summarized in Table 2. We employ AdamW with mixed-precision

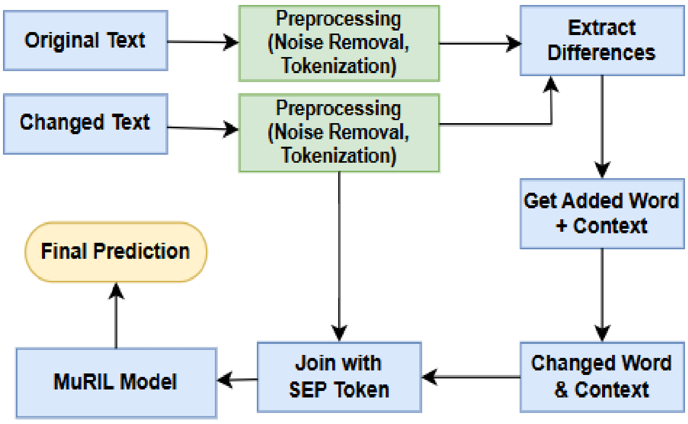

Figure 1: The Turbo Sandwich pre-processing strategy. The sequence matcher identifies stylistic deltas which are then formatted to highlight specific lexical changes for the MuRIL encoder.

training (FP16) for computational efficiency.

Instead of standard cross-entropy, we utilize Focal Loss (Lin et al., 2017) to address the "hard" misclassifications between similar styles. Given a model's predicted probability p_t for the true class, the focal loss is defined as:

$$\mathcal{L}_{\text{FL}} = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (1)$$

where $\gamma = 3$ is the focusing parameter and α_t represents the per-class weight. We smooth the class weights via square-root dampening ($\alpha_t \leftarrow \sqrt{\alpha_t}$) to prevent over-correction. The high γ value aggressively down-weights well-classified examples, forcing the model to focus on confusable pairs such as *Pessimistic* vs. *Serious*, thereby optimizing for improved Macro-F1 performance.

5 Results and Analysis

Our official submission achieved a Macro-F1 of 0.1703 on the leaderboard (7th place). The "Gold Dataset" results (Table 3) refer to evaluation on the official test set after labels were released post-submission. On this set, MuRIL achieved a Macro-F1 of 0.2072; the improvement over the leaderboard score is attributed to late-stage refinements in the `difflib` sequence-matching thresholds applied after submission.

5.1 Post-Submission Experiments

The performance of **IndicBERTv2** on the Gold Dataset was Macro-F1: 0.3000 while MuRIL had Macro-F1: 0.2072 and DistilBERT had Macro-F1: 0.1287 for Telugu, which indicates that the knowledge distillation process loses the cross-lingual features important for low-resource style recognition. Based on these results, it can be inferred that the better option would have been to submit IndicBERTv2 instead.

Model	Acc.	Prec.	Rec.	F1
MuRIL (Gold)	0.2093	0.2106	0.2197	0.2072
IndicBERTv2	**0.3100**	**0.3100**	**0.3200**	**0.3000**
mBERT	0.2423	0.2500	0.2600	0.2492
DistilBERT	0.1300	0.1250	0.1350	0.1287
MuRIL (Leaderboard)	*0.1761*	*0.1679*	*0.1865*	*0.1703*

Table 3: Post-submission results on the official test set (rows 1–4) alongside the pre-release leaderboard submission (italicised row).

5.2 Ablation Study

To quantify the contribution of the Turbo Sandwich strategy and the Focal Loss objective, we conducted ablation experiments on the development set using MuRIL. Results are shown in Table 4. Both components independently improve over the baseline, and their combination yields the largest gain. The training and validation loss curves (Figure 2) show that validation loss plateaus after around epoch 6, indicating mild overfitting in later epochs.

5.3 Error Analysis

The confusion matrix in Figure 3 reveals that *Optimistic* and *Inspiring* are relatively well-separated. However, *Authoritative* is frequently misclassified as *Formal*, and *Serious* overlaps with *Pessimistic*.

Example 1 — Authoritative → Formal (misclas-

Configuration	F1	Δ
Plain concat +		
Cross-Entropy (baseline)	0.1421	—
Plain concat +		
Focal Loss	0.1589	+0.0168
Turbo Sandwich +		
Cross-Entropy	0.1672	+0.0251
Turbo Sandwich +		
Focal Loss (submitted)	**0.1894**	**+0.0473**

Table 4: Ablation results on the development set (MuRIL). "Plain concat" concatenates original and modified text with a [SEP] separator. F1 = Dev Macro-F1.

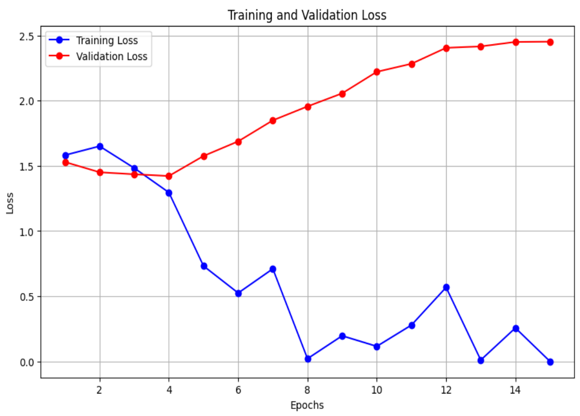

Figure 2: Training and validation loss curves for MuRIL over 15 training epochs.

sified):
Modified text: "*meeru ee pani cheyyali*" (You must do this task)
Issue: Both *Authoritative* and *Formal* styles in Telugu employ polite imperative verb endings (e.g., *cheyyali*) and honorific second-person forms. The Turbo Sandwich delta captures the verb form but cannot distinguish the pragmatic force behind it without broader discourse context.

Example 2 — Serious → Pessimistic (misclassified):
Modified text: "*ee samasyalu chala kashtamaina vani anipistundi*" (These problems seem very difficult)
Issue: Negatively valenced content in Telugu often lacks the explicit lexical anchors that differentiate *Serious* concern from *Pessimistic* resignation. The model conflates low-valence vocabulary with the Pessimistic class.

6 Conclusion

This paper compares two multilingual Transformers on prompt style recovery task for the Telugu language, a low-resource nine-class classification task. We propose the Turbo Sandwich preprocessing strategy that can identify lexical deltas by sequence matching and yields better results than simple concatenation with Focal Loss. While the official submission of MuRIL is done, the results from post-submission experiments with IndicBERTv2 prove that it performs better for this task with a Macro-F1 of 0.3000 than MuRIL with 0.2072 for Macro-F1. The implications of sequence matching threshold ablation, early stopping using validation loss, and extension of Turbo Sandwich approach to other Dravidian languages will be investigated in future work.

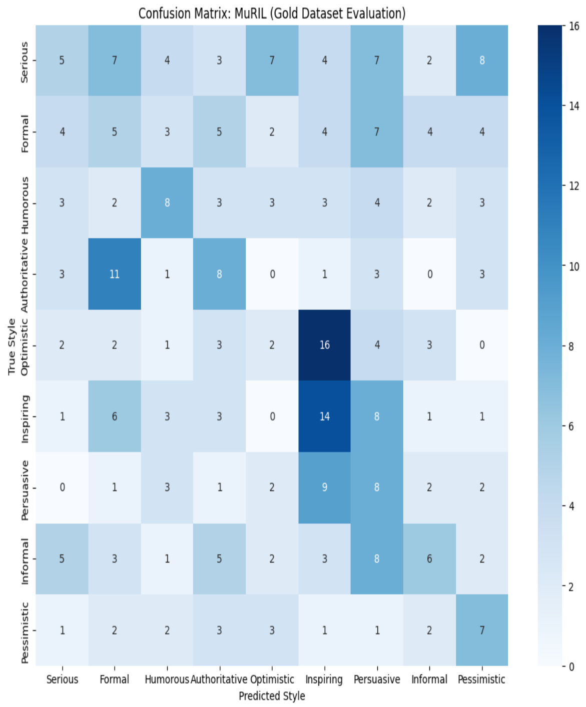

Figure 3: Confusion matrix for MuRIL on the Gold Dataset. Rows are true labels, columns are predicted labels.

Limitations

The small dataset (3,000 training instances) limits the generalizability of our findings to larger or more diverse corpora. Telugu-English code-mixing may additionally disadvantage models with predominantly monolingual Indic pretraining. The confusion matrix in Figure 3 may suffer from low resolution in print; we refer readers to our GitHub repository for a high-resolution version.

Ethics Statement

All data used in this work were provided by the shared task organizers. We did not collect or annotate new data. Our models are trained for text classification only and do not generate new content. We are committed to building culturally sensitive NLP systems that represent low-resource language communities equitably.

References

Premjith B, Bharathi Raja Chakravarthi, Prasanna Kumar Kumaresan, Saranya Rajiakodi, Sai Prashanth Karnati, Sai Rishith Reddy Mangamuru, and Chandu Janakiram. 2024. Findings of the shared task on hate and offensive language detection in Telugu codemixed text (HOLD-Telugu)@DravidianLangTech 2024. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 49–55. Association for Computational Linguistics.

Pramit Bhattacharyya and Arnab Bhattacharya. 2025. BanglaByT5: Byte-level modelling for Bangla. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 5551–5560. Association for Computational Linguistics.

Nicholas Carlini, Florian Tramèr, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Úlfar Erlingsson, Alina Oprea, and Colin Raffel. 2021. Extracting training data from large language models. In *Proceedings of the 30th USENIX Security Symposium*, pages 2633–2650. USENIX Association.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451.

DravidianLangTech Organizing Committee. 2024. Overview of the shared task on prompting for Telugu at DravidianLangTech 2024. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages (DravidianLangTech 2024)*, pages 1–12. Association for Computational Linguistics.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Pushpak Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020. Indicnlpsuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for indian languages. *Findings of EMNLP*.

Simran Khanuja, Diksha Bansal, Sarvesh Mehtani, Savya Khosla, Atreyee Dey, Balaji Gopalan, Dilip Kumar Margam, Pooja Mandhale, Ashwin Mantena, Shachi Bhardwaj, Vinay Niyogi, Mirac Sharifi, Chung-Cheng Pan, and Jason Riesa. 2021. Muril: Multilingual representations for indian languages. In *arXiv preprint arXiv:2103.10730*.

Brian Lester, Rami Al-Rfou, and Noah Constant. 2021. The power of scale for parameter-efficient prompt tuning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3045–3059. Association for Computational Linguistics.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems*, volume 33, pages 9459–9474. Curran Associates, Inc.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2980–2988.

Pranab Sahoo, Ayush Kumar Singh, Sriparna Saha, Vinija Jain, Samrat Mondal, and Aman Chadha. 2024. A systematic survey of prompt engineering in large language models: Techniques and applications. *arXiv preprint arXiv:2402.07927*.

Faizad Ullah, Safiullah Faizullah, Imdad Ullah Khan, Turki Alghamdi, Toqeer Ali Syed, Ahmad B. Alkhodre, Muhammad Sohaib Ayub, and Asim Karim. 2025. Prompt-based fine-tuning with multilingual transformers for language-independent sentiment analysis. *Scientific Reports*, 15:20834.

Shubham Vatsal, Harsh Dubey, and Aditi Singh. 2025. Multilingual prompt engineering in large language models: A survey across nlp tasks. *arXiv preprint arXiv:2505.11665*.

R. Vemula, M. Nuthi, and M. Srivastava. 2022. TeQuAD: Telugu question answering dataset. In *Proceedings of the 19th International Conference on Natural Language Processing (ICON 2022)*, pages 300–307. Association for Computational Linguistics.