CUET_SYNTHETICA@DravidianLangTech 2026: Multilingual Transformer Based Hope Speech Detection for Coarse and Fine-Grained Classification in Tulu

Sumaiya Zaman, Miftahul Jannat Rishta, Shiti Chowdhury, Hasan Murad
Department of Computer Science and Engineering,
Chittagong University of Engineering and Technology, Bangladesh
{u2104110, u2104019, u2004027}@student.cuet.ac.bd, hasanmurad@cuet.ac.bd

Abstract

Hope speech has played a vital role in online communities, yet most NLP work has focused on English and a few high-resource languages, leaving code-mixed varieties like Tulu largely unexplored. In the Shared Task on Hope Speech Detection in Code-Mixed Tulu at DravidianLangTech@ACL 2026, we have tackled two subtasks: (i) coarse-grained classification into Encouraging, Discouraging, Uninvolved and Blended categories (Task 1) and (ii) fine-grained classification into Optimistic, Realistic, Inspiring, Fading and Hopelessness (Task 2). We have fine-tuned three multilingual transformer encoders XLM-RoBERTa-base, MuRIL and mBERT on the official training splits. In Task 1, a three-way soft-voting ensemble of all three models has yielded the best performance with a macro F1 of 0.58, securing 1st place. In Task 2, XLM-RoBERTa-base alone has outperformed both MuRIL and mBERT, achieving a macro F1 of 0.42 and also securing 1st place.

1 Introduction

Social media has made it easier to share support, but most NLP work has focused on detecting negative language rather than positive. Hope speech has received far less attention, particularly in low-resource code-mixed languages (Chakravarthi, 2020; Chakravarthi and Muralidaran, 2021). Tulu, a Dravidian language spoken in coastal Karnataka, has been especially underrepresented, with users writing it heavily mixed with English and Kannada on YouTube (Chakravarthi, 2020; Sundar et al., 2021; Chakravarthi et al., 2022).

The DravidianLangTech@ACL 2026 Shared Task has introduced the first gold-standard Tulu dataset for two subtasks (Thenmozhi et al., 2026; Chakravarthi and Muralidaran, 2021; Chakravarthi et al., 2022): coarse-grained classification into Encouraging, Discouraging, Uninvolved and Blended in Task 1 and fine-grained classification into Optimistic, Realistic, Inspiring, Fading and Hopelessness in Task 2. We have fine-tuned XLM-RoBERTa-base, MuRIL and mBERT, achieving macro F1 of 0.58 (Task 1) and 0.42 (Task 2), both securing 1st place. Our main contributions are:

- We have implemented an ensemble of pretrained models, including XLM-R, MuRIL and mBERT for coarse-grained hope tone classification, demonstrating the power of combining models for superior performance.
- We have utilized XLM-RoBERTa-base, a pretrained transformer model, for fine-grained classification of code-mixed Tulu text, showing the efficacy of fine-tuning a single model for detecting nuanced emotional expressions.

Detailed implementation information is available in the GitHub repository: https://github.com/sumaiyaa110/Hope-Speech-Detection

2 Related Work

Hope speech detection has evolved from monolingual datasets to multilingual resources through LT-EDI workshops (Chakravarthi, 2020; Chakravarthi and Muralidaran, 2021; Chakravarthi, 2022). Early work on English, Tamil and Malayalam has shown that cross-lingual transformers outperform traditional classifiers on code-mixed Dravidian text (Sundar et al., 2021; Malik et al., 2023). Subsequent systems have demonstrated that context-aware embeddings improve multilingual hope speech detection (Aggarwal et al., 2022; Das et al., 2023), while PolyHope has extended this to English tweets (Balouchzahi et al., 2023; Sidorov et al., 2025) and recent work has explored hierarchical attention and capsule-based models (Rehman et al., 2025). However, regional code-mixed languages like Tulu remain underexplored, a gap addressed by the DravidianLangTech@ACL 2026 Shared Task (Thenmozhi et al., 2026).

Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 217–221

3 Data Description

The Hope Speech Detection in Code-Mixed Tulu dataset has been derived from YouTube comments annotated for Task 1 (Coarse-Grained Hope Tone Classification) and Task 2 (Fine-Grained Hope Type Classification), split into Training, Development and Test sets. Task 1 includes four hope tone categories: Encouraging, Discouraging, Uninvolved and Blended, while Task 2 contains five fine-grained types: Optimistic, Realistic, Inspiring, Fading Hope and Hopelessness. The data distribution is shown in Table 1.

Task	Train	Development	Test	Total
Task 1	5991	1284	1126	8401
Task 2	3185	682	683	4550

Table 1: Dataset Distribution for Task 1 and Task 2 (Coarse-Grained and Fine-Grained) Classification

4 Methodology

4.1 Problem Formulation

Task 1: The task is to classify code-mixed Tulu text into one of the four categories. Given a text t, the objective is to learn a mapping function $f(t) \rightarrow \{0, 1, 2, 3\}$, where 0 indicates "Encouraging Hope", 1 indicates "Discouraging Hope", 2 indicates "Uninvolved" and 3 indicates "Blended Hope".
Task 2: The task is to classify code-mixed Tulu text into five specific categories. Given a text t, the objective is to learn a mapping function $f(t) \rightarrow \{0, 1, 2, 3, 4\}$, where 0 indicates "Optimistic Hope", 1 indicates "Realistic Hope", 2 indicates "Inspiring Hope", 3 indicates "Fading Hope" and 4 indicates "Hopelessness."

4.2 Data Preprocessing

The text data for both tasks has been processed by removing extra spaces, standardizing column names and applying padding and truncation to ensure a consistent input length of 128 tokens. Tokenization has been done using XLM-R, MuRIL and mBERT for Task 1 and XLM-Roberta for Task 2. Task 2 additionally has involved normalizing punctuation, handling character repetitions and adjusting class weights during training to address class imbalance. This preprocessing has ensured clean, well-organized text for model input.

4.3 Transformer Models and Ensemble

For Task 1, we have experimented with XLM-RoBERTa-base, MuRIL and mBERT, averaging their softmax probabilities via soft-voting. Each model contributes complementary strengths: cross-lingual transfer, code-mixed Indian language knowledge and broad multilingual representations respectively, reducing overconfidence on ambiguous classes. For Task 2, XLM-RoBERTa-base alone has outperformed both MuRIL and mBERT for fine-grained classification and has been selected as the final model. The methodology is illustrated in Figure 1 and Figure 2.

4.4 Evaluation Metrics

The models have been evaluated using macro-F1 score, precision and recall to ensure balanced performance and accurate identification of the different hope speech categories.

5 Results and Analysis

5.1 Model Performance

5.1.1 Task 1: Coarse-Grained Hope Tone Classification

Model	Acc	P	R	F1
Dev Set				
MuRIL	0.6931	0.5938	0.5905	0.5914
XLM-Roberta-base	0.6682	0.6151	0.6187	0.6509
mBERT	0.6752	0.5860	0.5864	0.5859
XLM-R-base + mBert	0.7138	0.6272	0.6299	0.5978
All 3 Models	**0.7156**	**0.6200**	**0.6167**	**0.6179**
Test Set (Official Submission)				
All 3 Models (Ensemble)	**0.6700**	**0.5800**	**0.5800**	**0.5800**

Table 2: Dev set performance for model selection and official test set macro F1 of the submitted ensemble.

Table 2 has shown the performance of MuRIL, XLM-Roberta-base and mBERT in Coarse-Grained Hope Tone Classification. MuRIL has achieved 0.5914, XLM-Roberta-base has achieved 0.6509 and mBERT has achieved 0.5859. The three-model ensemble has achieved the best dev set F1 of 0.6179 and a macro F1 of 0.58 on the official test set, securing 1st place in Task 1.

Figure 3 has shown that the ensemble has performed well on the majority classes, correctly classifying 309 “Encouraging Hope” and 447 “Uninvolved” instances. However, it has struggled with “Discouraging Hope” and “Blended Hope,” mainly due to semantic overlap, mixed sentiment and class imbalance. “Blended Hope” is the most confused class, reflecting its ambiguous nature.

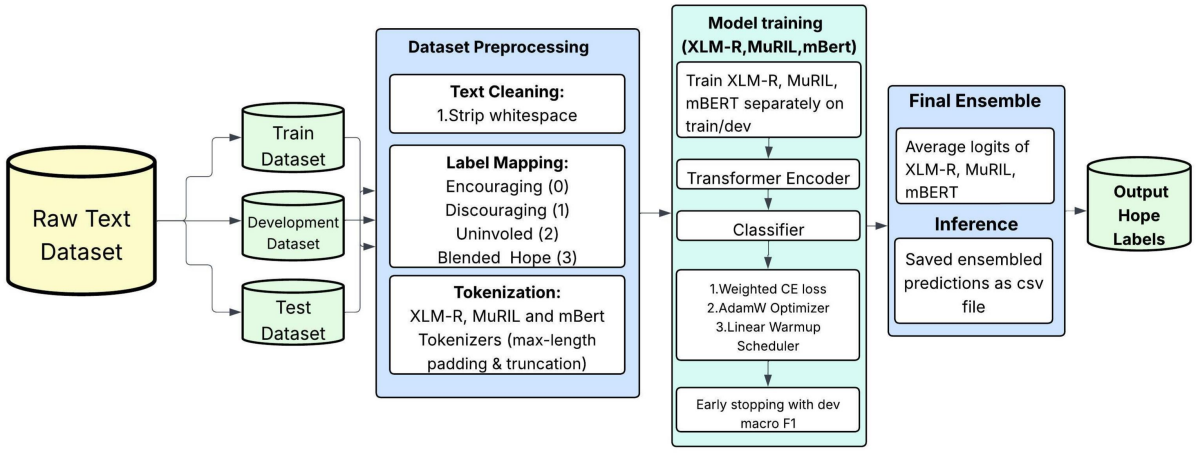

Figure 1: Process Flow Framework for Coarse-Grained Hope Tone Classification

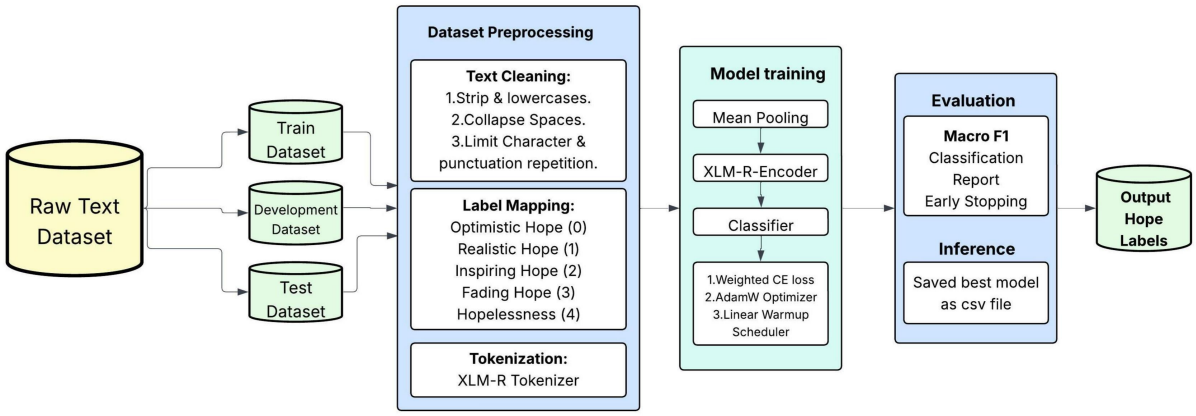

Figure 2: Process Flow Framework for Fine-Grained Hope Type Classification

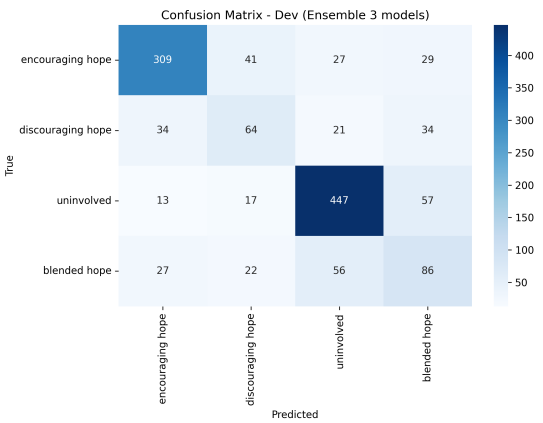

Figure 3: Confusion Matrix for Coarse-Grained Hope Tone Classification

5.1.2 Task 2: Fine-Grained Hope Tone Classification

Table 3 has shown the performance of MuRIL, XLM-Roberta-base and mBERT in Fine-Grained Hope Tone Classification. XLM-Roberta-base leads with a dev set F1 of 0.4218, followed by mBERT at 0.4414 and MuRIL at 0.3319. On the official test set, XLM-Roberta-base has achieved a macro F1 of 0.42, securing 1st place in Task 2.

Model	Acc	P	R	F1
Dev Set				
MuRIL	0.4399	0.4758	0.4432	0.3319
mBERT	0.5156	0.4394	0.4493	0.4414
XLM-Roberta-base	**0.4697**	**0.4202**	**0.4460**	**0.4218**
Test Set (Official Submission)				
XLM-Roberta-base	**0.5000**	**0.4200**	**0.4200**	**0.4200**

Table 3: Dev set performance for Task 2 model selection and official test set macro F1 of the submitted model.

Figure 4 has shown that the model performs best on "Inspiring Hope" and "Hopelessness," with 153 and 71 correct predictions. However, it has struggled with "Optimistic Hope," "Realistic Hope" and "Fading Hope" due to semantic overlap and subtle emotional differences. "Realistic Hope" is the most difficult class to distinguish, as it lies between optimism and hopelessness. These confusions are further affected by the severe class imbalance in

the Task 2 training data.

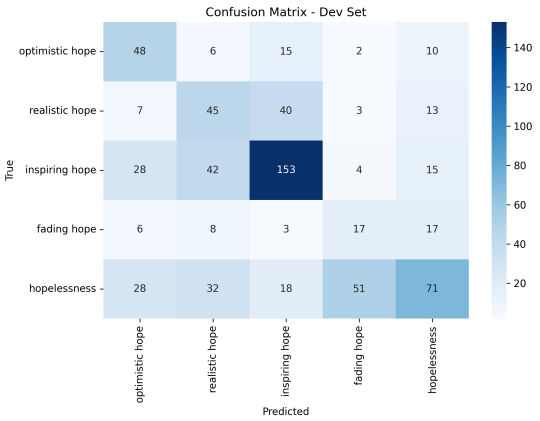

Figure 4: Confusion Matrix for Fine-Grained Hope Tone Classification

5.2 Hyperparameter Settings

Table 4 has presented the hyperparameter settings for Task 1 models, while Table 5 has shown the settings for Task 2 models.

Setting	XLM-Roberta-Base	MuRIL	mBERT
LR	2e-5	2e-5	2e-5
Batch	16	16	16
Epochs	8	8	8
Max Len	128	128	128
Optimizer	AdamW	AdamW	AdamW

Table 4: Training configuration for Task 1 models.

Setting	MuRIL	XLM-RoBerta-Base	mBERT
LR	2e-5	1.5e-5	2e-5
Batch	16	16	16
Epochs	4	10	10
Max Len	96	128	128
Optimizer	AdamW	AdamW	AdamW

Table 5: Training configuration for the three models used in Task 2.

All models have been trained on a single NVIDIA T4 GPU via Kaggle Notebooks, with per-model training time ranging from 30 to 60 minutes and the Task 1 ensemble requiring approximately 3 hours in total.

6 Conclusion

The Hope Speech Detection shared task at DravidianLangTech@ACL 2026 has shown that ensemble models perform well in detecting hope speech in code-mixed Tulu. The ensemble of XLM-R, MuRIL and mBERT has effectively captured different aspects of hope speech. XLM-RoBERTa-base has demonstrated strong performance in Task 2, but has struggled with identifying minority classes. Class weighting, early stopping and logits averaging have improved performance. Future work should focus on expanding the Tulu dataset and integrating real-time detection for better moderation of positive discourse.

Error Analysis

In Task 1, the ensemble has struggled most with *Blended Hope*, which has shared vocabulary with both *Encouraging* and *Discouraging* classes, making boundary detection difficult. In Task 2, *Realistic Hope* has proven the hardest class, as it has linguistically overlapped with both *Optimistic Hope* and *Hopelessness*. Additionally, heavy English-Tulu code-mixing has caused inconsistent tokenization across models, particularly for culturally specific expressions. Class imbalance in Task 2 has further degraded performance on minority classes such as *Fading Hope* and *Optimistic Hope*, despite class weighting during training.

Limitations

The model has struggled with overfitting and understanding the intricate details of Tulu. Issues such as class imbalance and the detection of small emotional changes persist. Limited data have made prediction difficult and the model has struggled with informal variations and culturally distinctive expressions. Future work incorporating larger Tulu corpora and language-specific pre-training is expected to improve performance.

Acknowledgement

We are grateful to the DravidianLangTech@ACL 2026 shared task organizers for having provided the dataset and evaluation framework(Thenmozhi et al., 2026). We also thank the developers of XLM-RoBERTa-base(Conneau et al., 2019), MuRIL(Khanuja et al., 2021) and mBERT(Devlin et al., 2019) whose open-source models have made this work possible.

Ethical Statement

All data processing and modeling have adhered to ethical guidelines to ensure users' privacy and fairness. The study has sought to improve the detection of positive, supportive language while fostering digital virtual spaces. Efforts have been made to address biases and limitations in the dataset for fair hope speech detection.

References

Pranjal Aggarwal, Pasupuleti Chandana, Jagrut Nemade, Shubham Sharma, Sunil Saumya, and Shankar Biradar. 2022. Hope speech detection on social media platforms. *arXiv preprint arXiv:2212.07424*.

Fazlourrahman Balouchzahi, Grigori Sidorov, and Alexander Gelbukh. 2023. Polyhope: Two-level hope speech detection from tweets. *Expert Systems with Applications*, 222:120078.

Bharathi Raja Chakravarthi. 2020. HopeEDI: A multilingual hope speech detection dataset for equality, diversity, and inclusion. In *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*, pages 41–53, Barcelona, Spain (Online). Association for Computational Linguistics.

Bharathi Raja Chakravarthi. 2022. Multilingual hope speech detection in english and dravidian languages. *BMC Bioinformatics*, 23(303).

Bharathi Raja Chakravarthi and Vigneshwaran Muralidaran. 2021. Findings of the shared task on hope speech detection for equality, diversity, and inclusion. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 61–72, Kyiv. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Vigneshwaran Muralidaran, Ruba Priyadharshini, Subalalitha Chinnaudayar Navaneethakrishnan, John Philip McCrae, Miguel Ángel García-Cumbreras, Salud María Jiménez-Zafra, Rafael Valencia-García, Prasanna Kumar Kumaresan, Rahul Ponnusamy, Daniel García-Baena, and José Antonio García-Díaz. 2022. Overview of the shared task on hope speech detection for equality, diversity, and inclusion. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion (LT-EDI)*, Dublin, Ireland. Association for Computational Linguistics.

Alexis Conneau, Guillaume Lample, Sebastian Ruder, et al. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Mithun Das, Shubhankar Barman, and Subhadeep Chatterjee. 2023. hate-alert@lt-edi-2023: Hope speech detection using transformer-based models. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 250–256, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Sanchit Khanuja et al. 2021. Muril: Multilingual representations for indian languages. *arXiv preprint arXiv:2103.10730*.

Muhammad Shahid Iqbal Malik, Anna Nazarova, Mona Mamdouh Jamjoom, and Dmitry I. Ignatov. 2023. Multilingual hope speech detection: A robust framework using transfer learning of fine-tuning roberta model. *Journal of King Saud University – Computer and Information Sciences*, 35(8):101736.

Mohammad Zia Ur Rehman, Devraj Raghuvanshi, Harshit Pachar, Chandravardhan Singh Raghaw, and Nagendra Kumar. 2025. Hierarchical attention-enhanced contextual capsulenet for multilingual hope speech detection. *Expert Systems with Applications*, 268:126285.

Grigori Sidorov, Fazlourrahman Balouchzahi, Luis Ramos, Helena Gómez-Adorno, and Alexander Gelbukh. 2025. Multilingual identification of nuanced dimensions of hope speech in social media. *Journal of Biomedical Informatics*, 150:104512.

Arunima Sundar, Akshay Ramakrishnan, Avantika Balaji, and Thenmozhi Durairaj. 2021. Hope speech detection for dravidian languages using cross-lingual embeddings with stacked encoder architecture. *SN Computer Science*, 2(1):Article 27.

Durairaj Thenmozhi, Rathnakar Shetty P, Parameshwar R. Hegde, Anusha M D, Raksha Adyanthaya, Mohammed Fadhel Aljunid, Prasanna Kumar Kumaresan, and Bharathi Raja Chakravarthi. 2026. Findings of the Shared Task on Hope Speech Detection in Tulu. In *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.