MUCS@Dravidianlangtech@ACL2026: Hope Speech Detection in Code-Mixed Tulu Language Using Multiple Features

Hosahalli Lakshmaiah Shashirekha and Rachana A
Department of Computer Science
Mangalore University
Mangalore, India
`hlsrekha@mangaloreuniversity.ac.in`, `rachanabanger0514@gmail.com`

Abstract

Hope speech refers to online expressions that promote positivity, encouragement, and social harmony. It fosters inclusivity and resilience, making it particularly valuable in culturally diverse and code-mixed communities. Detecting hope speech is an emerging area in computational linguistics, aimed at supporting healthier digital interactions and improving accessibility for vulnerable groups. While most of the hope speech detection work has been focused on high-resource languages, low- resource languages such as Tulu remains unexplored. In this paper, we - Team MUCS, describe our proposed system submitted to the first shared task on Hope Speech Detection in Code-Mixed Tulu, organized by DravidianLangTech@ACL 2026. As there are no pretrained language models for Tulu, we explored multiple hand crafted features - word n-grams (n = 1, 3), character n-grams (n = 1, 3), syllable n-grams (n = 1, 3) and sub-words, to train ensemble of classical Machine Learning (ML) models: i) Multinomial Naive Bayes (MNB) and Logistic Regression (LR) classifiers and ii) k Nearest Neighbor (kNN) and Decision Tree (DT) classifiers, both with soft-voting. Experimental results demonstrate that feature integration effectively captures lexical, sub-lexical, and phonological cues in noisy code-mixed text. The system achieves competitive performance on both development and test datasets, highlighting the effectiveness of feature-based approaches for hope speech detection in code-mixed Tulu. An ablation study is also conducted to evaluate the contribution of multiple feature sets for hope speech detection.

1 Introduction

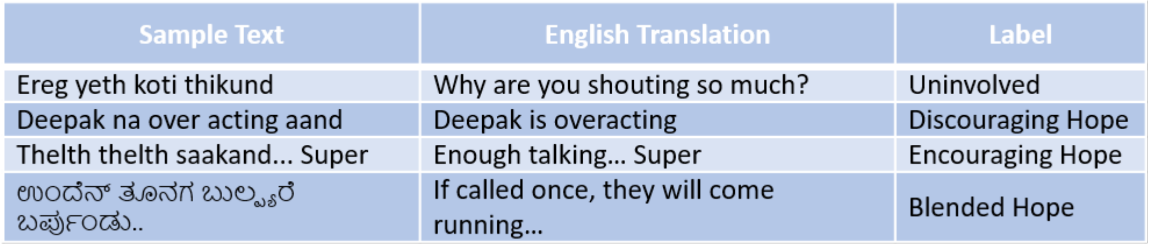
The rapid growth of social media platforms has transformed the way individuals communicate, express opinions, and share emotions online. Among various forms of online expression, hope speech plays a crucial role in promoting positivity, encouragement, and social harmony, particularly in multilingual and culturally diverse communities. In recent years, computational approaches for detecting harmful content such as hate speech and offensive language have received significant attention. However, comparatively less focus has been given to identifying and promoting hopeful and supportive content. This has led to the emergence of hope speech detection as an important research area within Natural Language Processing (NLP) (Silva et al., 2025).

Hope speech detection becomes more challenging in code-mixed settings, where users blend multiple languages within a single sentence or conversation. In the Indian context, code-mixing between regional languages and English is highly prevalent across social media platforms. Tulu code-mixed text, for instance, often contains transliterated words, spelling variations, informal expressions, and non-standard grammatical structures. These linguistic characteristics increase the complexity of automatic text classification tasks. Addressing these challenges requires robust preprocessing techniques and feature representations capable of capturing lexical, morphological, phonetic, and sub-word patterns (Hegde et al., 2023).

Tulu is a Dravidian language predominantly spoken in the coastal districts of Karnataka, such as Dakshina Kannada and Udupi, and in parts of northern Kerala, particularly Kasaragod region. Tulu belongs to Dravidian language family, which also includes Kannada, Tamil, Telugu, and Malayalam. It has a rich cultural and oral heritage reflected in traditional art forms such as Yakshagana and Bhuta Kola, as well as in folk songs and proverbs. Despite its cultural significance, Tulu is considered as an extremely low-resource language due to the absence of large-scale annotated corpora, pretrained language models, standardized lexical resources, and benchmark datasets. These limitations make computational research in Tulu very much challenging (Hegde and Shashirekha, 2023).

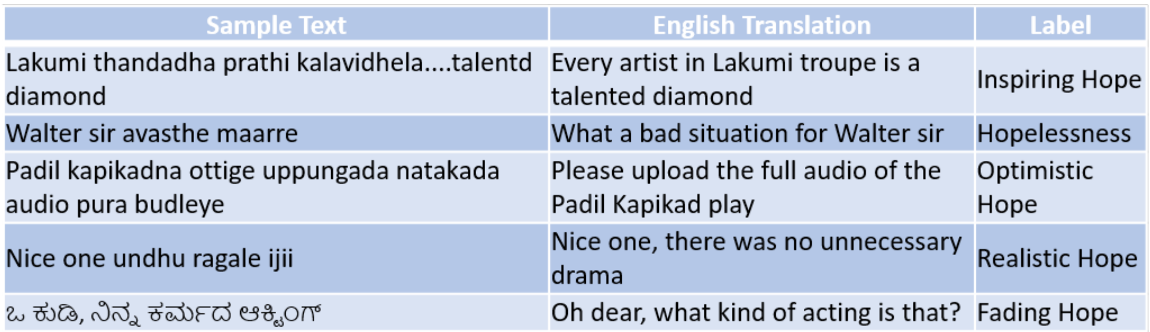
Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 306–315
July 3, 2026 ©2026 Association for Computational Linguistics

Table 1: Code-mixed Tulu Samples from Dataset for Task 1



Sample Text	English Translation	Label
Ereg yeth koti thikund	Why are you shouting so much?	Uninvolved
Deepak na over acting aand	Deepak is overacting	Discouraging Hope
Thelth thelth saakand... Super	Enough talking... Super	Encouraging Hope
ಉಂದೆನ್ ತೂನಗ ಬುಲ್ಪ್ಯರೆ ಬರ್ಪುಂಡು..	If called once, they will come running...	Blended Hope

Table 2: Code-mixed Tulu Samples from Dataset for Task 2



Sample Text	English Translation	Label
Lakumi thandadha prathi kalavidhela....talentd diamond	Every artist in Lakumi troupe is a talented diamond	Inspiring Hope
Walter sir avasthe maarre	What a bad situation for Walter sir	Hopelessness
Padil kapikadna ottige uppungada natakada audio pura budleye	Please upload the full audio of the Padil Kapikad play	Optimistic Hope
Nice one undhu ragale ijii	Nice one, there was no unnecessary drama	Realistic Hope
ಒ ಕುಡಿ, ನಿನ್ನ ಕರ್ಮದ ಆಕ್ಟಿಂಗ್	Oh dear, what kind of acting is that?	Fading Hope

Historically, Tulu was written using the Tigalari script; however, in contemporary usage, it is primarily written in Kannada script. On digital platforms and social media, Tulu is frequently written in Roman script, leading to multiple script representations of the same language. One major difficulty in processing code-mixed Tulu text is script variability, as the language may appear in Kannada or Roman script, complicating normalization and tokenization processes. Another key challenge is the prevalence of code-mixing and code-switching, especially with Kannada, and English. Further, code-mixing introduces transliteration inconsistencies, hybrid grammatical constructions, and irregular lexical patterns, thereby increasing the complexity of automatic text processing. Tulu is morphologically rich and agglutinative in nature, generating numerous inflected word forms that increase vocabulary sparsity. The lack of basic NLP tools such as tokenizers, lemmatizers, and part-of-speech taggers further restricts the development of robust computational systems for Tulu text (Shetty, 2024). Additionally, the lack of standardized orthography in Romanized Tulu results in inconsistent spellings, where the same word may appear in multiple forms.

To tackle the challenges of building applications for low-resource languages, Hope Speech Detection in Code-Mixed Tulu Language shared task[1] (Thenmozhi et al., 2026), organized by DravidianLangTech@ACL 2026, invites the research community to develop learning models for hope speech detection in code-mixed Tulu. The shared task consists of two subtasks: Task 1 – Coarse-Grained Hope Tone Classification and Task 2 – Fine-Grained Hope Type Classification. While Task 1 focuses on identifying whether a given text expresses a general tone of hope, Task 2 requires identifying specific types of hope expressed in the content. Coarse-grained classification involves broader category prediction, whereas fine-grained classification demands more precise semantic discrimination among closely related classes. Code-mixed Tulu text samples from the dataset for both Task 1 and 2 are shown in Tables 1 and 2 respectively.

To address the challenges of detecting hope speech in code-mixed Tulu text, in this paper, we team - MUCS, describe the proposed methodology submitted to the shared task. As there are no pretrained language models for Tulu, we explored multiple hand crafted features - word n-grams (n = 1, 3), character n-grams (n = 1, 3), syllable n-grams (n = 1, 3) and sub-words, to train ensemble of classical ML models: i) MNB and LR and ii) kNN and DT classifiers, both with soft-voting. The code to reproduce the proposed models is available

---

[1] https://www.codabench.org/competitions/11328/

in Github[2].

The subsequent sections of this paper details the Related work (Section 2), Methodology (Section 3), Experiments and Results (Section 4), followed by Conclusion and Future Work (Section 5).

2 Related Work

Hope speech detection has gained significant attention in recent years due to its potential to promote positive interaction and reduce online toxicity across social media platforms. Several studies have explored ML, Deep Learning (DL), and Transfer Learning (TL) approaches for detecting hope speech across multilingual and domain-specific datasets in several high-resource languages.

One of the earliest structured efforts was the *HopeEDI* dataset introduced by Chakravarthi (Chakravarthi, 2020). This dataset focused on Equality, Diversity, and Inclusion (EDI) from multilingual YouTube comments in English, Tamil, and Malayalam. The dataset was annotated into two primary classes: Hope and Not Hope. Traditional ML classifiers such as kNN, Support Vector Machine (SVM), DT, and LR, were trained using Term Frequency-Inverse Docuemtn Frequency (TF-IDF) of word features. Among these, DT and LR classifers showed competitive performance across languages, establishing strong baselines for future research. Balouchzahi et al. (Balouchzahi et al., 2021a) extended this line of research by developing binary and multiclass hope speech detection datasets in English tweets and their study benchmarked multiple ML, DL, and TL techniques. Classical ML models (SVM, DT, RF, LR, XGBoost, MLP, CatBoost) were trained using TF-IDF features, while DL models utilized GloVe and FastText embeddings with Long Short-Term Memory (LSTM), Bidirection LSTM (BiLSTM), and Convolutional Neural Networks (CNN) architectures. Transformer-based models such as BERT, RoBERTa, and multilingual BERT (mBERT) were fine-tuned for improved contextual representation. Their findings indicated that LR and CatBoost achieved strong macro F1-scores for binary classification, while performance declined slightly in multiclass settings.

Sidorov et al. (Sidorov et al., 2021) proposed transformer-based architectures for regret and hope speech detection using BERT, ALBERT, RoBERTa, DistilBERT, XLNet, and ELECTRA. Their experiments on Reddit and PolyHope datasets demonstrated that contextual transformer models significantly outperform traditional approaches, particularly in multiclass emotional categorization tasks. For Spanish and English datasets, Shahiki-Tash et al. (Shahiki-Tash et al., 2022) introduced a 5-layer CNN model trained using Keras embeddings for binary classification of Hope vs. Non-Hope. Their results showed a better performance for English compared to Spanish, highlighting the challenges associated with multilingual modeling. Ensemble strategies explored by Balouchzahi et al. (Balouchzahi et al., 2021b) used soft-voting combining TF-IDF of word and character n-grams with neural networks. These ensemble methods achieved improved weighted F1-scores across Malayalam, English, and Tamil datasets, demonstrating the effectiveness of hybrid architectures.

Puranik et al. (Puranik et al., 2023) experimented with CNN and BiLSTM architectures trained on various BERT variants - ULMFiT, mBERT, XLM-RoBERTa, and MuRIL, for code-mixed texts. Their CNN model trained with ULMFiT achieved high weighted F1-scores for English, while BiLSTM with mBERT performed strongly for Malayalam. However, Tamil code-mixed text remained comparatively challenging. Aggarwal et al. compared classical ML models (NB, LR, SVM) with BERT-based architectures for social media hope speech detection. Their results showed that contextual transformer models significantly outperform traditional ML models, achieving higher macro F1-scores due to better contextual understanding.(Aggarwal et al., 2023)

Several studies have demonstrated that traditional ML models using TF-IDF features, such as LR, SVM, and DT, provide strong baselines for hope speech detection, while transformer-based models like BERT and RoBERTa improve contextual understanding in complex settings. But most of these works are for English and other languages leaving behind Tulu language. This has motivated us to explore multiple features to design an ensemble framework with soft-voting for robust and efficient classification of code-mixed Tulu hope speech.

3 Methodology

Tulu is an extremely low-resource language, and no pretrained language models support Tulu. Hence, we focused on developing a framework exploring

---

[2] https://github.com/Rachu2k03/hope-speech-detection-

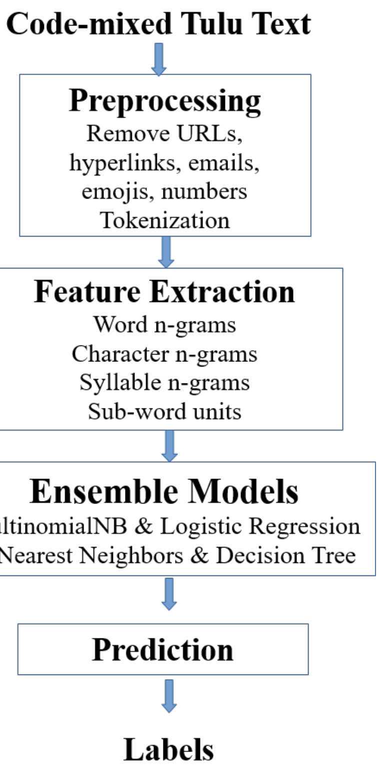


Figure 1: Flowchart of the Proposed Hope Speech Detection Framework

multiple hand crafted features to build models to detect hope speech in code-mixed Tulu. The proposed methodology includes: Pre-processing, Feature Extraction and Model Building. The framework of this proposed methodology is shown in Figure 1 and the steps are explained below:

3.1 Pre-processing

The give text data is pre-processed using a normalization pipeline specifically designed for noisy and code-mixed data. The pre-processing steps include removal of special characters, URLs and email addresses, converting emoji to text, and performing Indic-aware tokenization, which refers to tokenizing text while handling Indian language scripts like Kannada and Tulu particularly in code-mixed environment.

3.2 Feature Extraction

We extracted multiple features - word n-grams, character n-grams, syllable n-grams, and subwords, to capture different patterns at different levels of granularity:

- Word n-grams (n = 1, 3) model lexical patterns and contextual word usage.

- Character n-grams (n = 1, 3) capture orthographic and morphological patterns which includes spelling variations, transliteration inconsistencies, and mixed-script writing commonly observed in code-mixed social media text.

- Syllable n-grams (n = 1, 3) helps to capture phonological and pronunciation-level patterns within a word. They are generated using a bilingual syllabification strategy applied to both Tulu and English words.

- Sub-word features capture morphological variations, rare word formations, and unseen vocabulary. They are data-driven segments created by iteratively merging frequent character sequences, enabling robust handling of spelling variations and low-frequency terms in code-mixed social media text. Sub-words are captured using a Byte Pair Encoding[3] (BPE) tokenizer trained exclusively on the given training set.

This resulted in 71,752 features for Task 1 and 45,729 features for Task 2. Features and few examples of these features are shown in Table 3. These features are combined to obtain lexical, sub-lexical, and morphological information, enabling the models to capture diverse linguistic signals, and are vectorized using TfidfVectorizer[4] to train the MLmodels. Term Frequency-Inverse Document Frequency (TF-IDF) offers a simple yet powerful way to represent text by balancing the importance of terms across a corpus. It reduces the weight of frequent but uninformative words while emphasizing rarer, more discriminative ones, making it particularly effective for classification tasks.

3.3 Model Training

The ensembling technique enhances prediction stability by overcoming the weakness of one classifier with the strength of the other. We experimented different combinations of baseline models for ensembling with soft-voting and considered the following models which gave good results on Development set:

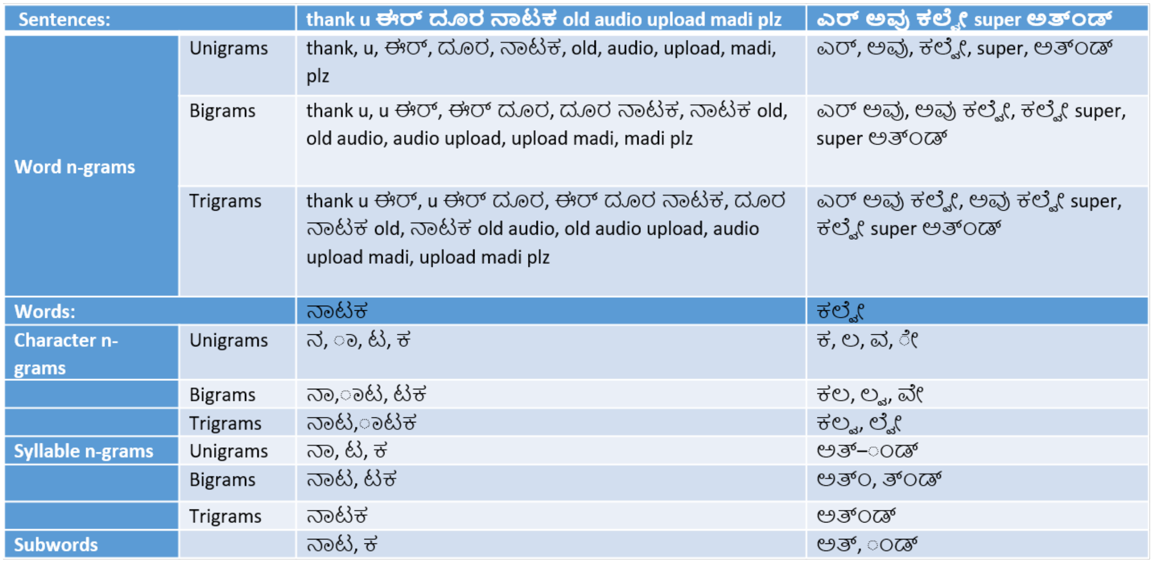
- **Model 1** : Ensembling MNB and LR classifiers

---

[3] https://huggingface.co/docs/tokenizers/api/models

[4] https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

Table 3: Features along with their Examples



Sentences:		thank u ಈರ್ ದೂರ ನಾಟಕ old audio upload madi plz	ಎರ್ ಅವು ಕಲ್ವೇ super ಅತ್ಂಡ್
Word n-grams	Unigrams	thank, u, ಈರ್, ದೂರ, ನಾಟಕ, old, audio, upload, madi, plz	ಎರ್, ಅವು, ಕಲ್ವೇ, super, ಅತ್ಂಡ್
	Bigrams	thank u, u ಈರ್, ಈರ್ ದೂರ, ದೂರ ನಾಟಕ, ನಾಟಕ old, old audio, audio upload, upload madi, madi plz	ಎರ್ ಅವು, ಅವು ಕಲ್ವೇ, ಕಲ್ವೇ super, super ಅತ್ಂಡ್
	Trigrams	thank u ಈರ್, u ಈರ್ ದೂರ, ಈರ್ ದೂರ ನಾಟಕ, ದೂರ ನಾಟಕ old, ನಾಟಕ old audio, old audio upload, audio upload madi, upload madi plz	ಎರ್ ಅವು ಕಲ್ವೇ, ಅವು ಕಲ್ವೇ super, ಕಲ್ವೇ super ಅತ್ಂಡ್
Words:		ನಾಟಕ	ಕಲ್ವೇ
Character n-grams	Unigrams	ನ, ◌ಾ, ಟ, ಕ	ಕ, ಲ, ವ, ◌ೇ
	Bigrams	ನಾ,◌ಾಟ, ಟಕ	ಕಲ, ಲ್ವ, ವೇ
	Trigrams	ನಾಟ,◌ಾಟಕ	ಕಲ್ವ, ಲ್ವೇ
Syllable n-grams	Unigrams	ನಾ, ಟ, ಕ	ಅತ್–◌ಂಡ್
	Bigrams	ನಾಟ, ಟಕ	ಅತ್ಂ, ತ್ಂಡ್
	Trigrams	ನಾಟಕ	ಅತ್ಂಡ್
Subwords		ನಾಟ, ಕ	ಅತ್, ◌ಂಡ್

Table 4: Statistics of Datasets for Task 1 and Task 2

Task 1 – Coarse-Grained Hope Tone		
Class	**Train Set**	**Development Set**
Uninvolved	2,490	534
Encouraging	1,895	406
Blended Hope	895	191
Discouraging	711	153
Total	5,991	1,284
Task 2 – Fine-Grained Hope Type		
Class	**Train Set**	**Development Set**
Inspiring Hope	1,129	242
Hopelessness	937	200
Realistic Hope	503	108
Optimistic Hope	380	81
Fading Hope	236	51
Total	3,185	682

- **Model 2** : Ensembling kNN and DT classifiers

MNB is a probabilistic model well-suited for text classification due to its effectiveness with word frequency features, while LR classifier is a linear discriminative classifier that performs well for high-dimensional sparse feature spaces. kNN is an instance-based method that predicts labels based on the majority class among the nearest samples and we have set the value of k to 5. DT classifier is a rule-based model that creates hierarchical splits to learn decision boundaries. The parameter random_state was set to 42 to obtain consistent results across runs, while all the other hyperparameters were retained at their default values.

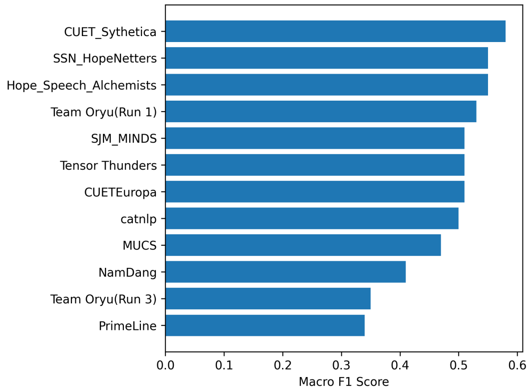


Figure 2: Comparison of Macro F1-scores of the Participating Teams on Test Set for Task 1

4 Experiments and Results

The statistics of the datasets used to build the models for Task 1 and 2 are shown in Table 4. It can be observed that the given data is imbalanced and we have not put any efforts to balance the dataset. The Test sets for Task 1 and 2 consists of 1,284 and 682 samples respectively. Performances of the models were evaluated by the organizers of the shared task based on macro F1-score. The class-wise performances of the ensemble models for Development and Test sets for both Task 1 and 2 are shown in Tables 5 and 6 respectively. Model 1 achieved stable and reliable behavior while Model 2 achieved complementary decision patterns. The comparison of macro F1-scores of the participating teams for Task 1 and 2 are shown in Figures 2 and 3 respectively.

Table 5: Results of Proposed Models on Development Sets

Task 1: Coarse-Grained Hope Tone Classification

Class	**Model 1 - Ensemble of MNB and LR**			**Model 2 - Ensemble of kNN and DT**		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
Blended Hope	0.64	0.04	0.07	0.23	0.24	0.23
Discouraging Hope	0.88	0.05	0.09	0.28	0.28	0.28
Encouraging Hope	0.70	0.81	0.75	0.68	0.65	0.66
Uninvolved	0.63	0.94	0.76	0.71	0.71	0.71
Accuracy	0.66			0.57		
Macro F1-Score	0.42			0.47		
Macro Avg F1-score	0.42			0.47		
Weighted Avg F1-score	0.57			0.57		

Task 2: Fine-Grained Hope Type Classification

Class	**Model 1 - Ensemble of MNB and LR**			**Model 2 - Ensemble of kNN and DT**		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
Fading Hope	0.00	0.00	0.00	0.18	0.12	0.14
Hopelessness	0.58	0.68	0.63	0.42	0.43	0.43
Inspiring Hope	0.52	0.93	0.67	0.50	0.61	0.55
Optimistic Hope	0.70	0.09	0.15	0.29	0.23	0.26
Realistic Hope	0.50	0.02	0.04	0.18	0.14	0.16
Accuracy	0.54			0.40		
Macro F1-Score	0.30			0.31		
Macro Avg F1-score	0.30			0.31		
Weighted Avg F1-score	0.44			0.39		

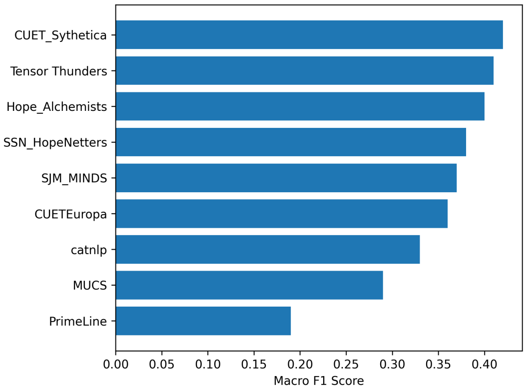


Figure 3: Comparison of Macro F1-scores of the Participating Teams on Test Set for Task 2

4.1 Ablation Study

An ablation study was conducted to evaluate the contribution of different features to the performance of the hope speech detection system. The methodology involved training the classifiers with all the features (All Features) and then systematically removing one feature type at a time, retraining the models, and comparing the results. This approach highlights the relative importance of each feature type in both Task 1 and Task 2 of the shared task. The results of ablation study for Task 1 and 2 are shown in Tables 7 and 8 respectively and the findings are given below:

- **Task 1: Coarse-Grained Hope Tone Classification**
    - **Model 1: Ensemble of MNB and LR** - the system with all the features achieved 0.66 accuracy and 0.41 macro F1-score. Removing word n-grams reduced performance, confirming their importance. Excluding character n-grams or syllable n-grams slightly improved macro F1-score, suggesting they add complementary signals but may introduce noise. Sub-word features had minimal impact, with negligible changes when removed.
    - **Model 2: Ensemble of kNN and DT** - the system with all the features achieved lower accuracy of 0.57 and macro F1-score of 0.46). Removing character n-grams improved macro F1-score to 0.48, showing their strong influence in non-linear models. Word and syllable

Table 6: Results of Proposed Models on Test Sets

Task 1: Coarse-Grained Hope Tone Classification

Class	**Model 1 - Ensemble of MNB and LR**			**Model 2 - Ensemble of kNN and DT**		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
Blended Hope	0.55	0.03	0.06	0.23	0.22	0.22
Discouraging Hope	0.67	0.03	0.05	0.25	0.26	0.25
Encouraging Hope	0.69	0.84	0.76	0.69	0.69	0.69
Uninvolved	0.64	0.93	0.76	0.68	0.69	0.69
Accuracy	0.65			0.56		
Macro F1-Score	0.40			0.46		
Macro Avg F1-score	0.41			0.46		
Weighted Avg F1-score	0.57			0.57		

Task 2: Fine-Grained Hope Type Classification

Class	**Model 1 - Ensemble of MNB and LR**			**Model 2 - Ensemble of kNN and DT**		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
Fading Hope	0.00	0.00	0.00	0.10	0.08	0.09
Hopelessness	0.55	0.66	0.60	0.42	0.46	0.44
Inspiring Hope	0.51	0.92	0.66	0.49	0.54	0.51
Optimistic Hope	0.75	0.07	0.13	0.25	0.20	0.22
Realistic Hope	0.00	0.00	0.00	0.20	0.17	0.18
Accuracy	0.53			0.38		
Macro F1-Score	0.28			0.29		
Macro Avg F1-score	0.28			0.29		
Weighted Avg F1-score	0.42			0.37		

n-grams had smaller effects, while sub-word features contributed the least.

- **Task 2: Fine-Grained Hope Type Classification**
  - **Model 1: Ensemble of MNB and LR** - the system with all the features achieved 0.53 accuracy and 0.28 macro F1-score. Removing word n-grams slightly reduces performance (0.52 accuracy, 0.27 macro F1-score), confirming their importance for lexical cues. Excluding character n-grams yields the best performance (0.54 accuracy, 0.30 macro F1-score), showing that character-level signals add strong complementary information. Removing syllable n-grams or sub-words produces no meaningful change (both around 0.53 accuracy, 0.28 macro F1-score), suggesting limited contribution.
  - **Model 2: Ensemble of kNN and DT** - the system with all the features achieved lower performance (0.38 accuracy, 0.29 macro F1-score). Removing word n-grams improves results (0.40 accuracy, 0.32 macro F1-score), indicating that word-level features may introduce noise in the model. Excluding character n-grams or syllable n-grams gives the highest performance (0.41 accuracy, 0.32–0.33 macro F1-score), highlighting their importance in capturing fine-grained distinctions. Removing sub-words also improves performance slightly (0.41 accuracy, 0.32 macro F1-score).

While Word n-grams dominate linear models and character n-grams strengthen non-linear ones in Task 1, character and syllable n-grams become more important, reflecting the need for finer linguistic granularity in Task 2. Sub-word features consistently play a supplementary role, with limited impact across both tasks.

4.2 Error Analysis

To better understand the model's behavior, we performed a qualitative analysis of predictions along with class-wise evaluation metrics as shown in Table 6. The analysis reveals several key challenges:

Table 7: Ablation Study Results for Task 1: Coarse-Grained Hope Tone Classification

Feature Set	**No. of Features**	**Accuracy**	**Macro F1-score**
Ensemble of MNB and LR			
All Features	71752	0.66	0.41
Without Word n-grams	70355	0.65	0.38
Without Char n-grams	69049	0.66	0.43
Without Syllable n-Ngrams	15672	0.67	0.43
Without Sub-words	71262	0.66	0.40
Ensemble of kNN and DT			
All Featrues	71752	0.57	0.46
Without Word Ngrams	70355	0.56	0.46
Without Char Ngrams	69049	0.57	0.48
Without Syllable Ngrams	15672	0.57	0.47
Without Sub-words	71262	0.55	0.45

Table 8: Ablation Study Results for Task 2: Fine-Grained Hope Type Classification

Feature Set	**No. of Features**	**Accuracy**	**Macro F1-score**
Ensemble of MNB and LR			
All Features	45729	0.53	0.28
Without Word Ngrams	44791	0.52	0.27
Without Char Ngrams	43471	0.54	0.30
Without Syllable Ngrams	10843	0.53	0.28
Without Sub-words	45294	0.53	0.28
Ensemble of kNN and DT			
All Features	45729	0.38	0.29
Without Word Ngrams	44791	0.40	0.32
Without Char Ngrams	43471	0.41	0.32
Without Syllable Ngrams	10843	0.41	0.33
Without Sub-words	45294	0.41	0.32

- The models perform significantly better on majority classes while struggling with minority categories. Classes such as Blended Hope and Discouraging Hope in Task 1, and Fading Hope, Realistic Hope, and Optimistic Hope in Task 2 exhibit very low performance. This is primarily due to class imbalance, where these categories are underrepresented in the training data. As a result, the model becomes biased toward dominant classes and fails to generalize effectively for minority labels.

- There is considerable linguistic and semantic overlap between classes, especially in the fine-grained task. Many instances share similar lexical patterns and differ only in subtle contextual cues, making it difficult for feature-based models to distinguish between categories.

- Code-mixing and orthographic variation introduce additional complexity. Users frequently switch between Kannada and English and often use phonetic transliterations in Roman script. This leads to inconsistent word forms, increasing ambiguity and reducing model robustness. While character and syllable features partially mitigate this issue, they are not sufficient to fully resolve it.

- The limitations of feature-based representations also contribute to misclassifications. Unlike contextual models, the current approach does not capture deeper semantic relationships, which are often necessary for identifying nuanced expressions of hope.

These observations highlight the inherent challenges of hope speech detection in code-mixed, low-resource settings.

5 Conclusion

This paper presents the ensemble of ML models for Hope Speech detection in code-mixed Tulu text to tackle the challenges of Hope Speech Detection in Code- Mixed Tulu Language shared task organized by DravidianLangTech@ACL 2026. To address the linguistic variability, orthographic inconsistencies, and morphological richness of code-mixed social media content, we designed a multiple feature framework incorporating word n-grams, character n-grams, syllable n-grams, and sub-words representations. Ensemble of traditional ML classifiers with soft-voting demonstrated that integrating multiple feature representations significantly improves classification robustness and stability. Future directions include exploring class balancing techniques such as data augmentation and re-sampling to improve performance on minority classes. In addition, integrating transformer-based or multilingual contextual models, will be explored to enhance the system's ability to capture deeper semantic relationships in code-mixed text.

Acknowledgments

1. The authors are grateful to the Vision Group on Science and Technology (VGST), Department of Science and Technology, Government of Karnataka, for providing the financial support under the GRE scheme (GRD No: 1114) for carrying out this research work.

2. We acknowledge the use of generative AI systems in assisting with language refinement during manuscript preparation. All substantive contributions, including research design, experiments, and analysis, are carried out by the authors.

3. We thank Rachana N for her active participation and valuable contribution in this shared task. Her involvement and support played a meaningful role in the successful completion of this work.

Limitations

Despite its effectiveness, the proposed approach has several limitations.

- Shallow feature reliance: word, character, syllable n-grams and sub-words capture surface patterns but miss deeper semantic and contextual meaning. This limits performance on nuanced categories where subtle context matters.

- Sensitivity to noise: n-gram features are brittle in noisy, code-mixed, or transliterated text (common in Tulu social media). Spelling variations and inconsistent scripts reduce feature reliability.

- Curse of dimensionality: u-sing multiple n-gram and subword features creates very high-dimensional sparse vectors. Models like kNN and DT struggle with scalability and efficiency in such spaces.

- Limited generalization: Ensemble combinations (MNB+LR, kNN+DT) improve robustness but remain feature-based. They cannot generalize well to unseen contexts compared to contextual embeddings (e.g., BERT).

- Class imbalance vulnerability: probabilistic models like MNB and distance-based models like kNN are biased toward majority classes. Minority categories suffer disproportionately, even with ensemble voting.

- Overfitting risk: DT and kNN classifiers are prone to overfitting in small, imbalanced datasets. Without regularization or pruning, they memorize training patterns rather than learning general rules.

- Lack of semantic hierarchy: The models treat features independently and cannot capture hierarchical linguistic structures (syntax, discourse), which are often crucial in fine-grained hope speech detection.

The proposed methodology is strong enough for surface-level lexical cues, but limited in handling contextual semantics, noisy code-mixed text, and minority class generalization.

References

Nidhi Aggarwal, Ujjwal Singh, Venkatesh M. Reddy, and Aniket Puranik. 2023. Comparative Analysis of Machine Learning and BERT Models for Hope Speech Detection. *International Journal of Social Computing*.

Fazlourrahman Balouchzahi, Reza Farahbakhsh, and Noel Crespi. 2021a. Hope Speech Detection on Multilingual Social Media Texts Using Machine Learning, Deep Learning, and Transformer Models. In *Proceedings of the Workshop on Online Abuse and Harms*.

Fazlourrahman Balouchzahi, Reza Farahbakhsh, and Noel Crespi. 2021b. Using Ensemble Methods for Multilingual Hope Speech Detection. In *Proceedings of the Forum for Information Retrieval Evaluation (FIRE 2021) – Hope Speech Detection Shared Task*, pages –. ACM.

Bharathi Raja Chakravarthi. 2020. HopeEDI: A Multilingual Hope Speech Detection Dataset for Equality, Diversity, and Inclusion. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP)*, pages 247–256.

Asha Hegde, Bharathi Raja Chakravarthi, Hosahalli Lakshmaiah Shashirekha, Rahul Ponnusamy, Subalalitha Chinnaudayar Navaneethakrishnan, Lavanya Sambath Kumar, Durairaj Thenmozhi, Martha Karunakar, Shreya Sriram, and Sarah Aymen. 2023. Findings of the Shared Task on Sentiment Analysis in Tamil and Tulu Code-Mixed Text. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, pages 64–71, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Asha Hegde and Hosahalli Lakshmaiah Shashirekha. 2023. KT2: Kannada-Tulu Parallel Corpus Construction for Neural Machine Translation. In *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, pages 743–753, Goa University, Goa, India. NLP Association of India (NLPAI).

Aniket Puranik, Ujjwal Singh, and Venkatesh M. Reddy. 2023. Multilingual and Code-Mixed Hope Speech Detection using CNN and BiLSTM with BERT Variants. In *Proceedings of the IEEE International Conference on Natural Language Processing*.

A. Shahiki-Tash and 1 others. 2022. CNN-Based Hope Speech Detection for Spanish and English Social Media Text. *Journal of Language Technology and Applications*.

Poorvi Shetty. 2024. Natural Language Processing for Tulu: Challenges, Review and Future Scope. In *Proceedings of the International Conference on Speech and Language Technologies for Low-Resource Languages (SPELLL 2023)*, volume 2046 of *Communications in Computer and Information Science*, pages 93–109, Varna, Bulgaria. Springer Nature Switzerland.

Grigori Sidorov, Alexander Gelbukh, and Luis Chanona-Hernandez. 2021. Transformer-Based Models for Hope and Regret Speech Classification. In *International Conference on Computational Linguistics and Intelligent Text Processing*.

Tulio Ferreira Leite Da Silva, Gonzalo Freijedo Aduna, Farah Benamara, Alda Mari, Zongmin Li, Li Yue, and Jian Su. 2025. CDB: A Unified Framework for Hope Speech Detection through Counterfactual, Desire and Belief. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 4448–4463, Albuquerque, New Mexico, USA. Association for Computational Linguistics.

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.