

Findings of the Shared Task on Hope Speech Detection for Equality, Diversity, and Inclusion

Bharathi Raja Chakravarthi[♥] and Vigneshwaran Muralidaran^{*}

[♥]Data Science Institute, National University of Ireland Galway

^{*}Cardiff University, Cardiff, United Kingdom

bharathi.raja@insight-centre.org

Abstract

Hope plays a crucial role in the well-being, recuperation and restoration of human life. Hope speech reflects the belief that one can discover pathways to their desired objectives and become motivated to utilise those pathways. We created a hope speech detection dataset to encourage research in Natural Language Processing (NLP) towards reinforcement of positivity rather than minimising negativity (through control of hate speech etc.). In this paper, we report the findings of the shared task of hope speech detection for Tamil, English, and Malayalam languages conducted as a part of the EACL 2021 workshop on Language Technology for Equality, Diversity, and Inclusion (LT-EDI-2021). We also present an overview of the methods and results of the competing systems. The datasets for this challenge are openly available¹. To the best of our knowledge, this is the first shared task to conduct hope speech detection.

1 Introduction

In the recent years, there has been an exponential rise in the number of studies focusing on the detection and management of hate speech and offensive language in social media (Kumar et al., 2018; Mandl et al., 2020; Zampieri et al., 2020). However, this has led to controlling user expression instead of improving user experience. (Ghanghor et al., 2021a; Hegde et al., 2021; Yasaswini et al., 2021). This has also resulted in putting barriers on modes of expression of different groups of people resulting in the violation of the principles of Equality, Diversity, and Inclusion (EDI). For instance, some NLP systems have classified the comments made in African American English as offensive language without accommodating the linguistic features peculiar to the dialect. Hence there is a need for a

¹<https://competitions.codalab.org/competitions/27653>

shift in the approaches taken to handle hate speech without compromising on the principles of EDI.

We propose to shift the prevailing research direction in Natural Language Processing from controlling negativity (curbing hate speech etc.) to encouraging positivity (promoting hope speech). Hope speech is any expression that is positive, encouraging, supportive, and / or inspires promise of the future (Chakravarthi, 2020a). For the shared task organised in this connection, participants were provided with development, training and test dataset in English as well as in two under-resourced languages - Tamil and Malayalam. Tamil (ISO 639-3: tam) belongs to the Dravidian language family and is widely spoken in the southern state of Tamil Nadu in India, Sri Lanka, Malaysia and Singapore (Krishnamurti, 2003; Kolipakam et al., 2018; Chakravarthi, 2020b; Thavareesan and Mahesan, 2019, 2020a,b). Malayalam (ISO 639-3: mal) also belongs to the Dravidian language family and is spoken in the Indian state of Kerala and the Union Territories of Lakshdweep and Puducherry (Chakravarthi et al., 2020). Both Tamil and Malayalam have their own scripts which are alphasyllabaries like other Indic scripts i.e. they are partially alphabetic and partially syllabic. The Tamil language was written using Tamili, Vattezhuthu, Chola, Pallava and Chola-Pallava scripts at different points in history. The modern Tamil script descended from the Chola-Pallava script that was conceived around the 6th century CE (Srinivasan, 2019). Malayalam was first written with Vattezhuthu script that evolved from Tamili script around 4-5th century CE (Mahadevan, 2003). Modern Malayalam is written using the Vattezhuttu alphabets extended with symbols from the Grantha script to accommodate non-native Sanskrit sounds (Krishnamurti, 2003). Although they have their own scripts, the social media comments in these languages are often written in Latin script as it is

easy to input.

Our dataset for the hope speech for EDI shared task was created from user-generated content in Tamil, Malayalam, and English (Chakravarthi and Muralidaran, 2021). The user-generated comments in our dataset for Tamil and Malayalam were code-mixed (Chakravarthi, 2020a). We proposed a comment/post level classification task. The goal for the participants was to classify a given Youtube comment into either ‘Hope_speech’, ‘Non_hope_speech’ or ‘Not_Tamil / Not_Malayalam / Not_English’. Our CodaLab website will remain open to allow researchers to access the data and build upon this work.

2 Task Description

We define the hope speech for our problem as “YouTube comments/posts that offer support, reassurance, suggestions, inspiration and insight”. A comment/post within the corpus may contain more than one sentence, but the average sentence length of the corpus is one. The annotations in the corpus are made at a comment/post level. The participants were provided with development, training and test dataset in English, Tamil, and Malayalam. A few examples from the dataset are given below along with their translations and hope speech class annotations.

- **kashtam thaan. irundhaalum muyarchi seivom** - *It is indeed difficult. Let us try it out though.* Hope speech.
- **uff. China mon vannallo-** *Phew! Here comes the Chinese guy* Non-hope speech
- **paambu kari saappittu namma uyirai vaanguranunga-** *These guys (Chinese) eat snake meat and make our lives miserable* Non-hope speech
- **Enthinennariyilla karan-jupoy.sooryamma.you. are grate-** *I do not know what is the reason but I cried. Dear Soorya, you are great.* Hope speech
- **Plz talk slowly-** *Please talk slowly* Not Tamil
- **Gd msg.. thanks dr.-** *It is a good message. Thanks Doctor.* Not Malayalam

2.1 Datasets

We have used the HopeEDI datasets from Chakravarthi (2020a) for this shared task. Our

dataset contained user-generated comments from the social media platform YouTube with 28,451, 20,198 and 10,705 comments in English, Tamil and Malayalam, respectively, manually labelled as hope speech or not. The inter-annotator agreement of the dataset labels was verified using Krippendorff’s alpha. We will now briefly describe the annotation process and the statistics of the collected dataset. For detailed explanation of the datasets, please refer to the works of Chakravarthi (2020a). The data was collected on a wide range of socially relevant topics related to EDI, including LGBTIQ issues, COVID-19, women in STEM, Black Lives Matter, Dravidian languages, and the Indo-China war. The data for English was collected from English-speaking regions across the world. For Tamil and Malayalam, we collected the data from the states of Tamil Nadu and Kerala in India. Full consideration was given to minimize the risk associated with breach of privacy by removing the information related to individual identity from a comment before it was sent for annotation. The annotators were educated about EDI and a minimum of three annotators labelled each comment. Our diverse annotator base was from Australia, the Republic of Ireland, the United Kingdom, the United States of America, Tamil Nadu, Kerala (in India) and Sri Lanka. The inter-annotator agreement in terms of Krippendorff’s alpha (α) was 0.63, 0.76, and 0.85 for English, Tamil and Malayalam.

In total, HopeEDI dataset contains 59,354 comments with 28,451 comments in English, 20,198 comments in Tamil, and 10,705 comments in Malayalam. Table 1 and Table 2 show the distribution of HopeEDI dataset by language and annotation label respectively. To calculate corpus statistics *nltk* tool (Bird et al., 2009) was used to tokenise the sentences in the comments. The vocabulary size of the Tamil and Malayalam dataset was high due to different types of code-mixing in the user-generated content.

Most of the comments fell under the ‘Not_hope’ category because of how we defined hope speech and hence the distribution of classes in the dataset was skewed. It is typical for user-generated content from online social media platforms to be skewed and the systems should be able to handle imbalanced data common in real-life setting. ‘not_Tamil’ and ‘not_Malayalam’ labels are found in considerably large amount due to high code-mixing. The fully annotated dataset was split into training, de-

| Language | English | Tamil | Malayalam |
|---------------------------------------|---------|---------|-----------|
| Number of Words | 522,717 | 191,242 | 122,917 |
| Vocabulary Size | 29,383 | 46,237 | 40,893 |
| Number of Comments/Posts | 28,451 | 20,198 | 10,705 |
| Number of Sentences | 46,974 | 22935 | 13,643 |
| Average number of words per sentences | 18 | 9 | 11 |
| Average number of sentences per post | 1 | 1 | 1 |

Table 1: Corpus statistics

| Class | English | Tamil | Malayalam |
|------------|---------|--------|-----------|
| Hope | 2,484 | 7,899 | 2,052 |
| Not Hope | 25,940 | 9,816 | 7,765 |
| Other lang | 27 | 2,483 | 888 |
| Total | 28,451 | 20,198 | 10,705 |

Table 2: Classwise Data Distribution

| | English | Tamil | Malayalam |
|-------------|---------|--------|-----------|
| Training | 22,762 | 16,160 | 8564 |
| Development | 2,843 | 2,018 | 1070 |
| Test | 2,846 | 2,020 | 1071 |
| Total | 28,451 | 20,198 | 10,705 |

Table 3: Train-Development-Test Data Distribution

velopment and test sets. The training, development and test sets respectively contain 80%, 10% and 10% of the data as shown in Table 3.

2.2 Training Phase

In this first phase, training, validation and development data were released for the participants to train and develop hope speech detection for one or more of the three languages. The participants could choose to perform cross-validation on the training data or use the validation dataset for preliminary evaluations and use the development set for hyperparameter sharing. The goal of this phase was to ensure that the systems developed by the participants be ready for evaluation before the release of the test data. In total, 137 participants registered for all the three languages and downloaded the data.

2.3 Testing Phase

In this phase, the test dataset was released in Co-daLab without the gold labels. Participants were instructed to submit their predictions using Google Forms. They were allowed to submit their results as many times as they wanted to and their best

submission was chosen for evaluation and preparation of rank list. The predictions were evaluated against the gold standard labels. The performance of the classification system was measured in terms of weighted averaged Precision, Recall and F-Score across all the classes. Weighted averaged scores calculate the support-weighted mean per label. The metric used for the preparing the rank list was the weighted F1 score. Participants were encouraged to check their system with Sklearn classification report ². For Tamil, Malayalam and English languages, the number of participants were 30, 31 and 31 respectively for the final test.

3 Systems

We first describe the baseline systems from the HopeEDI paper. Next, we briefly describe the general approach taken by each participating team. We encourage the readers to refer to the individual papers of each team for more details.

3.1 Baseline Classifiers

We used the prior published work as the baseline for hope speech detection from social media. More details about the baseline system description and results can be seen in Chakravarthi (2020a). In this prior work, we described in detail our dataset and presented our baseline results by employing a wide range of standard classifiers on the unbalanced settings of the dataset. The experiment was applied on the token frequency-inverse document frequency (Tf-Idf) of tokens. We used sklearn ³ library to create baseline classifiers. We used a grid search for the k-nearest neighbours (KNN), support vector machine (SVM), decision tree and logistic regression. For the multinomial Naive Bayes, we set alpha = 0.7. More details about the parameters

²https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html

³<https://scikit-learn.org/stable/>

of the classifier will be published in the code.

3.2 System Descriptions

In this section we summarise the systems implemented by the participants for the shared task. For more details, please refer to the shared task papers submitted by the authors.

- [Chinnappa \(2021\)](#) participated in identifying hope speech classes in English, Tamil, and Malayalam datasets. They presented a two-phase mechanism to detect hope speech. In the first phase, they built a classifier to identify the language of the text. In the second phase, they created a classifier to identify the class labels. The author used language Models SBERT, FNN and BERT inference. They achieved 3rd, 4th, and 2nd ranks in Tamil, Malayalam and English respectively.
- [M K and A P \(2021\)](#) used context-aware string embeddings for word representations and Recurrent Neural Network (RNN) and pooled document embeddings for text representation. Their proposed methodology achieved higher performance than baseline results. The highest weighted average F-scores of 0.93, 0.56, and 0.84 for English, Tamil, and Malayalam were reported on the final evaluation test set. The proposed models outperformed baselines by 3%, 2% and 11% in absolute terms for English, Tamil, and Malayalam.
- [Dowlagar and Mamidi \(2021\)](#) performed experiments by taking advantage of pre-processing and transfer-learning models. They have shown that the pre-trained multilingual-BERT model with convolution neural networks gave the best results. Their model ranked 1st, 3rd, and 4th on English, Malayalam-English, and Tamil-English code-mixed datasets respectively.
- [S et al. \(2021a\)](#) trained the data using transformer models, specifically mBERT for Tamil and Malayalam and BERT for English, and achieved weighted average F1-scores of 0.38, 0.81, 0.92 for Tamil, Malayalam, and English respectively. They achieved the rank of 14, 4, 2 for Tamil, Malayalam, English respectively.
- [Puranik et al. \(2021\)](#) experimented with several transformer-based models including BERT, ALBERT, DistilBERT, XLM-RoBERTa, MuRIL to classify dataset in English, Malayalam and Tamil languages. ULM-FiT achieved a weighted average F1-score of 0.91 on English data, mBERT achieved 0.57 on Malayalam data, and distilmBERT achieved 0.37 on Tamil data. They secured 15th, 12th, and 3rd ranks for predictions on Tamil, Malayalam, and English datasets respectively.
- [Saumya and Mishra \(2021\)](#) used various machine learning and deep learning-based models (support vector machine, logistics regression, convolutional neural network, recurrent neural network) to identify the hope speech in the given YouTube comments. The best performing model on English data was 2-parallel CNN-LSTM with GloVe and Word2Vec embeddings, and it reported a weighted average F1 score 0.91 and 0.90 for development and test sets respectively. Similarly, the best performing model of Tamil and Malayalam was obtained from 3-parallel Bi-LSTM. For Tamil, the reported F1 scores were 0.56 and 0.54 on development and test dataset respectively. Similarly, for Malayalam, the reported weighted F1 was 0.78 and 0.79 on development and test datasets.
- [Dave et al. \(2021\)](#) used TF-IDF character n-grams and pretrained MuRIL embeddings for text representation and Logistic Regression and Linear SVM for classification. Their best approach achieved 2nd, 8th and 5th rank with a weighted F1 scores of 0.92, 0.75 and 0.57 in English, Malayalam-English and Tamil-English on test dataset.
- [Zhou \(2021\)](#) fine-tuned the Roberta pre-training model based on three data sets English, Tamil, and Malayalam. The F1 scores of their models in Tamil and Malayalam sub-tasks reached 0.56 and 0.78, respectively, and the F1 score in English sub-task reached 0.93, achieving the first rank.
- [Zhao \(2021\)](#) used attention mechanism to adjust the weights of all the output layers of XLM-RoBERTa to make full use of the information extracted from each layer, and used the weighted sum of all the output layers to complete the classification task. They used

| Team-Name | Precision | Recall | F1 Score | Rank |
|--|-----------|--------|----------|------|
| spartans (Sharma and Arora, 2021) | 0.62 | 0.62 | 0.61 | 1 |
| TeamUNCC (Mahajan et al., 2021) | 0.61 | 0.61 | 0.61 | 1 |
| NLP@CUET (Hossain et al., 2021) | 0.61 | 0.61 | 0.6 | 2 |
| res - si sun | 0.61 | 0.6 | 0.6 | 2 |
| team-hub (Huang and Bai, 2021) | 0.61 | 0.61 | 0.59 | 3 |
| MUCS (Balouchzahi et al., 2021) | 0.59 | 0.59 | 0.59 | 3 |
| ZYJ (Zhao, 2021) | 0.59 | 0.59 | 0.59 | 3 |
| dhivya-hope-detection (Chinnappa, 2021) | 0.59 | 0.59 | 0.59 | 3 |
| GCDH (Ziehe et al., 2021) | 0.62 | 0.6 | 0.58 | 4 |
| e8ijs | 0.59 | 0.59 | 0.58 | 4 |
| EDIOne - suman (Dowlagar and Mamidi, 2021) | 0.58 | 0.58 | 0.58 | 4 |
| IITK (Ghanghor et al., 2021b) | 0.58 | 0.58 | 0.58 | 4 |
| HopeIsAllYouNeed | 0.59 | 0.59 | 0.57 | 5 |
| IRNLP-DAIICT-LR (Dave et al., 2021) | 0.59 | 0.59 | 0.57 | 5 |
| KBCNMUJAL | 0.59 | 0.59 | 0.57 | 5 |
| KU-NLP (M K and A P, 2021) | 0.62 | 0.6 | 0.56 | 6 |
| Zeus (Zhou, 2021) | 0.59 | 0.59 | 0.56 | 6 |
| CFILT-IITB-Submission | 0.55 | 0.55 | 0.55 | 7 |
| IIT-DWD (Saumya and Mishra, 2021) | 0.54 | 0.54 | 0.54 | 8 |
| hopeful-nlp (Awatramani, 2021) | 0.57 | 0.56 | 0.53 | 9 |
| MUM | 0.53 | 0.53 | 0.53 | 9 |
| snehan-coursera | 0.53 | 0.55 | 0.52 | 10 |
| TeamX - Olawale Onabola | 0.55 | 0.55 | 0.52 | 10 |
| Hopeful-Men (Upadhyay et al., 2021) | 0.52 | 0.55 | 0.49 | 11 |
| SIMON (Que, 2021) | 0.63 | 0.56 | 0.49 | 11 |
| result | 0.63 | 0.56 | 0.49 | 11 |
| Amrita-CEN-NLP (S et al., 2021b) | 0.48 | 0.49 | 0.47 | 12 |
| mIGeng | 0.42 | 0.42 | 0.42 | 13 |
| ssn-diBERTsity (S et al., 2021a) | 0.43 | 0.44 | 0.38 | 14 |
| IITTT - Karthik Puranik (Puranik et al., 2021) | 0.38 | 0.39 | 0.37 | 15 |

Table 4: Rank list based on F1-score along with other evaluation metrics (Precision and Recall) for Tamil language

| Team-Name | Precision | Recall | F1 Score | Rank |
|---|-----------|--------|----------|------|
| NLP@CUET(Hossain et al., 2021) | 0.86 | 0.85 | 0.85 | 1 |
| MUCS (Balouchzahi et al., 2021) | 0.85 | 0.85 | 0.85 | 1 |
| GCDH (Ziehe et al., 2021) | 0.84 | 0.85 | 0.85 | 1 |
| ZYJ (Zhao, 2021) | 0.84 | 0.84 | 0.84 | 2 |
| team-hub (Huang and Bai, 2021) | 0.84 | 0.85 | 0.84 | 2 |
| res - si sun | 0.84 | 0.85 | 0.84 | 2 |
| KU-NLP (M K and A P, 2021) | 0.84 | 0.85 | 0.84 | 2 |
| CFILT-IITB-Submission | 0.84 | 0.85 | 0.84 | 2 |
| TeamUNCC (Mahajan et al., 2021) | 0.83 | 0.83 | 0.83 | 3 |
| IITK (Ghanghor et al., 2021b) | 0.83 | 0.84 | 0.83 | 3 |
| HopeIsAllYouNeed | 0.83 | 0.83 | 0.83 | 3 |
| EDIOne - suman t (Dowlagar and Mamidi, 2021) | 0.83 | 0.83 | 0.83 | 3 |
| e8ijs | 0.83 | 0.84 | 0.83 | 3 |
| ssn-diBERTsity (S et al., 2021a) | 0.82 | 0.81 | 0.81 | 4 |
| snehan-coursera | 0.82 | 0.81 | 0.81 | 4 |
| KBCNMUJAL | 0.81 | 0.82 | 0.81 | 4 |
| hopeful-nlp (Awatramani, 2021) | 0.82 | 0.81 | 0.81 | 4 |
| dhivya-hope-detection (Chinnappa, 2021) | 0.81 | 0.82 | 0.81 | 4 |
| IIT-DWD (Saumya and Mishra, 2021) | 0.79 | 0.79 | 0.79 | 5 |
| Zeus (Zhou, 2021) | 0.79 | 0.81 | 0.78 | 6 |
| MUM | 0.76 | 0.78 | 0.77 | 7 |
| TeamX - Olawale Onabola | 0.77 | 0.74 | 0.75 | 8 |
| IRNLP-DAIICT-LR (Dave et al., 2021) | 0.78 | 0.79 | 0.75 | 8 |
| Hopeful-Men (Upadhyay et al., 2021) | 0.76 | 0.79 | 0.75 | 8 |
| Amrita-CEN-NLP (S et al., 2021b) | 0.78 | 0.73 | 0.75 | 8 |
| Amrita (S et al., 2021b) | 0.76 | 0.72 | 0.73 | 9 |
| spartans (Sharma and Arora, 2021) | 0.62 | 0.62 | 0.61 | 10 |
| mIGeng | 0.58 | 0.61 | 0.59 | 11 |
| IITT - Karthik Puranik (Puranik et al., 2021) | 0.57 | 0.57 | 0.57 | 12 |
| SIMON (Que, 2021) | 0.63 | 0.56 | 0.49 | 13 |
| result | 0.63 | 0.56 | 0.49 | 13 |

Table 5: Rank list based on F1-score along with other evaluation metrics (Precision and Recall) for Malayalam language

| Team-Name | Precision | Recall | F1 Score | Rank |
|---|-----------|--------|----------|------|
| Zeus (Zhou, 2021) | 0.93 | 0.94 | 0.93 | 1 |
| TeamUNCC (Mahajan et al., 2021) | 0.93 | 0.94 | 0.93 | 1 |
| team-hub (Huang and Bai, 2021) | 0.93 | 0.93 | 0.93 | 1 |
| res - si sun | 0.93 | 0.93 | 0.93 | 1 |
| NLP@CUET(Hossain et al., 2021) | 0.93 | 0.93 | 0.93 | 1 |
| KU-NLP (M K and A P, 2021) | 0.92 | 0.93 | 0.93 | 1 |
| Hopeful-men (Upadhyay et al., 2021) | 0.93 | 0.93 | 0.93 | 1 |
| GCDH | 0.93 | 0.93 | 0.93 | 1 |
| EDIOne - suman t (Dowlagar and Mamidi, 2021) | 0.93 | 0.94 | 0.93 | 1 |
| cs-english (Chen and Kong, 2021) | 0.93 | 0.94 | 0.93 | 1 |
| Autobots (Gundapu and Mamidi, 2021) | 0.93 | 0.93 | 0.93 | 1 |
| Hopeful-nlp (Awatramani, 2021) | 0.93 | 0.94 | 0.93 | 1 |
| ZYJ (Zhao, 2021) | 0.92 | 0.93 | 0.92 | 2 |
| ssn-diBERTsity (S et al., 2021a) | 0.91 | 0.93 | 0.92 | 2 |
| MUCS (Balouchzahi et al., 2021) | 0.92 | 0.93 | 0.92 | 2 |
| IRNLP-DAIICT-LR (Dave et al., 2021) | 0.92 | 0.93 | 0.92 | 2 |
| IITK (Ghanghor et al., 2021b) | 0.92 | 0.92 | 0.92 | 2 |
| HopeIsAllYouNeed | 0.92 | 0.93 | 0.92 | 2 |
| dhivya-hope-detection (Chinnappa, 2021) | 0.92 | 0.92 | 0.92 | 2 |
| CFILT-IITB-Submission | 0.92 | 0.93 | 0.92 | 2 |
| snehan-coursera | 0.92 | 0.91 | 0.91 | 3 |
| IIIT - Karthik Puranik (Puranik et al., 2021) | 0.92 | 0.91 | 0.91 | 3 |
| MUM | 0.89 | 0.91 | 0.9 | 4 |
| IIIT-DWD (Saumya and Mishra, 2021) | 0.9 | 0.91 | 0.9 | 4 |
| e8ijs | 0.91 | 0.92 | 0.9 | 4 |
| wrecking-crew | 0.9 | 0.91 | 0.87 | 5 |
| HopeFighters | 0.83 | 0.91 | 0.87 | 5 |
| Amrita-CEN-NLP (S et al., 2021b) | 0.83 | 0.91 | 0.87 | 5 |
| mlGeng | 0.86 | 0.85 | 0.85 | 6 |
| TeamX - Olawale Onabola | 0.9 | 0.77 | 0.81 | 7 |
| KBCNМУJAL | 0.88 | 0.5 | 0.61 | 8 |

Table 6: Rank list based on F1-score along with other evaluation metrics (Precision and Recall) for English language

the Stratified-K-Fold method to address class imbalance. They achieved a weighted average F1-score of 0.59, 0.84, and 0.92 for Tamil, Malayalam, and English languages, which ranked 3rd, 2nd, and 2nd respectively.

- [Huang and Bai \(2021\)](#) used the method and model that combines the XLM-RoBERTa pre-training language model and the TF-IDF algorithm. They got 1st, 2nd, and 3rd ranks on the English, Malayalam, and Tamil dataset respectively.
- [Chen and Kong \(2021\)](#) used fine-tuned BERT and K fold cross-validation to accomplish classification on English dataset. They achieved a final F1 score of 0.93, and got 1st rank for English language.
- [Ziehe et al. \(2021\)](#) demonstrated that even very simple baseline algorithms perform reasonably well on this task if provided with enough training data. However, their best performing algorithm is a cross-lingual transfer learning approach in which they fine-tuned XLM-RoBERTa. The model achieved the 1st rank for Malayalam and English and the 4th rank for Tamil.
- [Mahajan et al. \(2021\)](#), in their paper, described their approach of fine-tuning RoBERTa for Hope Speech detection in English and fine-tuning XLM-RoBERTa for Hope Speech detection in Tamil and Malayalam languages. They ranked 1st in English (F1 = 0.93), 1st in Tamil (F1 = 0.61) and 3rd in Malayalam (F1 = 0.83).
- [Gundapu and Mamidi \(2021\)](#) describe a Transformer-based BERT model for Hope speech detection. Their model achieved a weighted averaged F1-score of 0.93 on the test set for English. They showed that BERT model helped for better contextual representation of words in a comment, and the language identification model- assisted in detecting 'Other language' comments. They also explored with other transformer models like RoBERTa, XLNet, Albert, FLAIR, and ELMo for a superior hope speech detection.
- [S et al. \(2021b\)](#) proposed a BiLSTM with attention based approach in solving hope speech detection and using this approach they achieved an F1 score of 0.73 (9th rank) in Malayalam-English dataset.
- [Upadhyay et al. \(2021\)](#) experimented with two approaches. In the first approach, they used contextual embeddings to train classifiers using logistic regression, random forest, SVM, and LSTM based models. The second approach involved using a majority voting ensemble of 11 models which were obtained by fine-tuning pre-trained transformer models (BERT, AL-BERT, RoBERTa, IndicBERT) after adding an output layer. They found that the second approach was superior for English, Tamil and Malayalam. They got a weighted F1 score of 0.93, 0.75 and 0.49 for English, Malayalam and Tamil respectively. They ranked 1st in English, 8th in Malayalam and 11th in Tamil.
- [Awatramani \(2021\)](#) achieved an F-score of 0.93, ranking 1st on the leaderboard for English comments. The paper used pre-trained transformers and Paraphrasing Generation for Data Augmentation.
- [Hossain et al. \(2021\)](#) employed various machine learning (SVM, LR, ensemble), deep learning (CNN+BiLSTM) and transformer (m-BERT, Indic-BERT, XLNet, XLM-R) based methods. They showed that XLM-R outperformed all other techniques by gaining a weighted F1-score of 0.93, 0.60 and 0.85 respectively for English, Tamil and Malayalam language. Their team achieved 1st, 2nd and 1st rank in these three tasks respectively.
- [Que \(2021\)](#) used XLM- Roberta model and proposed an excellent multilingual model to achieve the classification task.
- [Balouchzahi et al. \(2021\)](#) created three models namely, CoHope-ML, CoHope-NN, and CoHope-TL based on Ensemble of classifiers, Keras Neural Network (NN) and BiLSTM with Conv1D model. CoHope-ML, CoHope-NN models were trained on a feature set comprising of char sequences extracted from sentences combined with words for Ma-En and Ta-En code-mixed texts and a combination of word and char ngrams along with syntactic word ngrams for English text. CoHope-TL model consisted of three major parts: training tokenizer, BERT Language Model (LM)

training and then using pre-trained BERT LM as weights in BiLSTM- Conv1d model. Out of the three proposed models, CoHope-ML model (best one among the models proposed) obtained 1st, 2nd, and 3rd ranks with weighted F1-scores of 0.85, 0.92, and 0.59 for Ma-En, English and Ta-En texts respectively.

- [Sharma and Arora \(2021\)](#) extends that of ([Arora, 2020a](#)) as they used their strategy to synthetically generate code-mixed data for training a transformer-based model RoBERTa and used it in an ensemble along with their pre-trained ULMFiT. They presented RoBERTa language model for code-mixed Tamil which they pre-trained from scratch. Using transfer learning they fine-tune RoBERTa and ULMFiT language models on down-stream tasks of OLI and HSD. They got Rank 4 in the former task using an ensemble of classifiers trained on RoBERTa and ULMFiT and Rank 1 in the latter task using classifier based on ULMFiT.

4 Results and Discussion

Overall, we received a total of 31,31 and 30 submissions for English, Malayalam and Tamil tasks. It is interesting to note that the top performing teams in all the three languages predominantly used XLM-Roberta to complete the shared task. One of the top ranking teams for English used context-aware string embeddings for word representations and Recurrent Neural Networks and pooled document embeddings for text representation. Among the other submissions, although Bi-LSTM was popular, there were other machine learning and deep learning models that were used. However, they did not achieve good results compared to the Roberta based models.

The top scores were 0.61, 0.85 and 0.93 for Tamil, Malayalam and English respectively. The range of scores was between 0.37 to 0.61, 0.49 to 0.85 and 0.61 to 0.93 for Tamil, Malayalam and English datasets respectively. It can be seen that the F1 scores of all the submissions on the Tamil dataset were considerably lower than those of Malayalam and English. It is not surprising that the English scores were better because many approaches used variations of pretrained transformer based models trained on English data. Due to code-mixing at various levels the scores are naturally lower for Malayalam and Tamil datasets. Among these two, the systems submitted performed badly

on Tamil data. The identification of the exact reasons for the bad performance in Tamil requires further research. However, one possible explanation for this could be that the distribution of ‘Hope_speech’, ‘Non_hope_speech’ and ‘not-Tamil’ classes is starkly different from those of English and Malayalam. In the English dataset there were just 27 out of 28,451 comments which were labelled as not-English by the annotators. In the remaining two classes, the number of non-hope speech comments were significantly higher than hope speech comments. Similarly in Malayalam there were only 888 out of 10,705 comments that were not Malayalam and overwhelmingly higher proportion of non-hope speech in the dataset. However in Tamil dataset, there were 2,483 comments out of 20,198 comments that were not Tamil. Moreover the proportion of Non-hope speech comments in Tamil (9,816 comments) is close enough to the number of hope speech labels (7,899 comments). This made the classifications in English and Malayalam a binary prediction task rather than a three class classification and naturally the former systems performed better than the latter.

The participants rarely faced challenges during the training, testing and submission phase. However, there were minor inconsistencies in the labelling of classes that was duly rectified on Codalab immediately. The initial round of peer review offered significant suggestions to the participants in terms of rewriting paper in a more presentable way. These suggestions of the reviewers were duly incorporated before the final submissions.

5 Conclusion

In this paper, we have presented the first shared task results on hope speech detection for equality, diversity, and inclusion. We received a wide range of entries that satisfied the goals of the shared task. We hope that HopeEDI shared task makes a lasting contribution to the NLP field.

References

- Vasudev Awatramani. 2021. Hopeful NLP@LT-EDI-EACL2021: Finding Hope in YouTube Comment Section. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Fazlourrahman Balouchzahi, Aparna B K, and H L Shashirekha. 2021. MUCS@LT-EDI-

- EACL2021:CoHope-Hope Speech Detection for Equality, Diversity, and Inclusion in Code-Mixed Texts. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*. ” O’Reilly Media, Inc.”.
- Bharathi Raja Chakravarthi. 2020a. [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. 2020b. *Leveraging orthographic information to improve machine translation of under-resourced languages*. Ph.D. thesis, NUI Galway.
- Bharathi Raja Chakravarthi, Navya Jose, Shardul Suryawanshi, Elizabeth Sherly, and John P McCrae. 2020. A sentiment analysis dataset for code-mixed Malayalam-English. In *Proceedings of the 1st Joint Workshop of SLTU (Spoken Language Technologies for Under-resourced languages) and CCURL (Collaboration and Computing for Under-Resourced Languages) (SLTU-CCURL 2020)*, Marseille, France. European Language Resources Association (ELRA).
- 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*. Association for Computational Linguistics.
- Shi Chen and Bing Kong. 2021. [cs-english@LT-EDI-EACL2021: Hope Speech Detection Based On Fine-tuning AIBERT Model](#). In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Dhivya Chinnappa. 2021. Multilingual Hope Speech Detection for Code-mixed and Transliterated Texts. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Bhargav Dave, Shripad Bhat, and Prasenjit Majumder. 2021. [IRNLP-DAIICT@LT-EDI-EACL2021: Hope Speech detection in Code Mixed text using TF-IDF Char N-grams and MuRIL](#). In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Suman Dowlagar and Radhika Mamidi. 2021. [EDIOne@LT-EDI-EACL2021: Pre-trained Transformers with Convolutional Neural Networks for Hope Speech Detection](#). In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Nikhil Kumar Ghanghor, Parameswari Krishnamurthy, Sajeetha Thavareesan, Ruba Priyadharshini, and Bharathi Raja Chakravarthi. 2021a. [IITK@DravidianLangTech-EACL2021: Offensive Language Identification and Meme Classification in Tamil, Malayalam and Kannada](#). In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, Online. Association for Computational Linguistics.
- Nikhil Kumar Ghanghor, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021b. [IITK@LT-EDI-EACL2021: Hope Speech Detection for Equality, Diversity, and Inclusion in Tamil, Malayalam and English](#). In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online.
- Sunil Gundapu and Radhika Mamidi. 2021. [Autobots@LT-EDI-EACL2021: All Lives Matter! Hope Speech Detection with BERT Transformer Model](#). In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Siddhanth U Hegde, Adeep Hande, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. [UVCE-IITTT@DravidianLangTech-EACL2021: Tamil Troll Meme Classification: You need to Pay more Attention](#). In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.
- Eftekhar Hossain, Omar Sharif, and Mohammed Moshiul Hoque. 2021. [NLP-CUET@LT-EDI-EACL2021: Multilingual Code-Mixed Hope Speech Detection using Cross-lingual Representation Learner](#). In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Bo Huang and Yang Bai. 2021. [TEAM HUB@LT-EDI-EACL2021: Hope Speech Detection Based On Pre-trained Language Model](#). In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Vishnupriya Kolipakam, Fiona M Jordan, Michael Dunn, Simon J Greenhill, Remco Bouckaert, Russell D Gray, and Annemarie Verkerk. 2018. A

- Bayesian phylogenetic study of the Dravidian language family. *Royal Society open science*, 5(3):171504.
- Bhadriraju Krishnamurti. 2003. *The Dravidian languages*. Cambridge University Press.
- Ritesh Kumar, Atul Kr. Ojha, Shervin Malmasi, and Marcos Zampieri. 2018. [Benchmarking aggression identification in social media](#). In *Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018)*, pages 1–11, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Junaida M K and Ajees A P. 2021. KU-NLP@LT-EDI-EACL2021: A Multilingual Hope Speech Detection for Equality, Diversity, and Inclusion using Context Aware Embeddings. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Iravatham Mahadevan. 2003. Early tamil epigraphy. *From the earliest times to the sixth century ad*.
- Khyati Mahajan, Erfan Al-Hossami, and Samira Shaikh. 2021. TeamUNCC@LT-EDI-EACL2021: Hope Speech Detection using Transfer Learning with Transformers. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Thomas Mandl, Sandip Modha, Anand Kumar M, and Bharathi Raja Chakravarthi. 2020. [Overview of the HASOC Track at FIRE 2020: Hate Speech and Offensive Language Identification in Tamil, Malayalam, Hindi, English and German](#). In *Forum for Information Retrieval Evaluation, FIRE 2020*, page 29–32, New York, NY, USA. Association for Computing Machinery.
- Karthik Puranik, Adeep Hande, Ruba Priyadarshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. IIIT@LT-EDI-EACL2021: Hope Speech Detection: There is always hope in Transformers. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Qinyu Que. 2021. Simon @ LT-EDI-EACL2021: Detecting Hope Speech with BERT. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Arunima S, Akshay Ramakrishnan, Avantika Balaji, Thenmozhi D, and Senthil Kumar B. 2021a. ssn-diBERTcity@LT-EDI-EACL2021: Hope Speech Detection on multilingual YouTube comments via transformer based approach. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Thara S, Ravi Teja Tasubilli, and Kothamasu Sai Rahul. 2021b. Amrita@LT-EDI-EACL2021: Hope Speech Detection on Multilingual Text. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Sunil Saumya and Ankit Kumar Mishra. 2021. IIIT-DWD@LT-EDI-EACL2021: Hope Speech Detection in YouTube multilingual comments. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Megha Sharma and Gaurav Arora. 2021. Spartans@LT-EDI-EACL2021: Inclusive Speech Detection using Pretrained Language Models. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Ramya Srinivasan. 2019. Reinventing tamil script.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2019. [Sentiment Analysis in Tamil Texts: A Study on Machine Learning Techniques and Feature Representation](#). In *2019 14th Conference on Industrial and Information Systems (ICIIS)*, pages 320–325.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2020a. [Sentiment Lexicon Expansion using Word2vec and fastText for Sentiment Prediction in Tamil texts](#). In *2020 Moratuwa Engineering Research Conference (MERCon)*, pages 272–276.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2020b. [Word embedding-based Part of Speech tagging in Tamil texts](#). In *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, pages 478–482.
- Ishan Sanjeev Upadhyay, Nikhil E, Anshul Wadhawan, and Radhika Mamidi. 2021. Hopeful Men@LT-EDI-EACL2021: Hope Speech Detection Using Indic Transliteration and Transformers. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.
- Konthala Yaraswini, Karthik Puranik, Adeep Hande, Ruba Priyadarshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. IIIT@DravidianLangTech-EACL2021: Transfer Learning for Offensive Language Detection in Dravidian Languages. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.
- Marcos Zampieri, Preslav Nakov, Sara Rosenthal, Pepa Atanasova, Georgi Karadzhov, Hamdy Mubarak, Leon Derczynski, Zeses Pitenis, and Çağrı Çöltekin. 2020. [SemEval-2020 task 12: Multilingual offensive language identification in social media \(OffenseEval 2020\)](#). In *Proceedings of the Fourteenth*

Workshop on Semantic Evaluation, pages 1425–1447, Barcelona (online). International Committee for Computational Linguistics.

Yingjia Zhao. 2021. ZYJ@LT-EDI-EACL2021:XLM-RoBERTa-Based Model with Attention for Hope Speech Detection. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.

Siyao Zhou. 2021. Zeus@LT-EDI-EACL2021: Hope speech detection based on Pre-training Mode. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.

Stefan Ziehe, Franziska Pannach, and Aravind Krishnan. 2021. cs-english@GCDH@LT-EDI-EACL2021: XLM-RoBERTa for Hope Speech Detection in English, Malayalam, and Tamil. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online. Association for Computational Linguistics.