MUCIC@LT-EDI-ACL2022: Hope Speech Detection using Data Re-Sampling and 1D Conv-LSTM

M. D. Anusha [1, a], **F. Balouchzahi** [2, b], **H. L. Shashirekha** [1, c], **G. Sidorov** [2, d]
[1]Department of Computer Science, Mangalore University, Mangalore, India
[2]Instituto Politécnico Nacional, Centro de Investigación en Computación, CDMX, Mexico
{[a]anugowda251, [c]hlsrekha}@gmail.com,
{[b]fbalouchzahi2021, [d]sidorov}@cic.ipn.mx

Abstract

Spreading positive vibes or hope content on social media may help many people to get motivated in their life. To address Hope Speech detection in YouTube comments, this paper presents the description of the models submitted by our team - MUCIC, to the Hope Speech Detection for Equality, Diversity, and Inclusion (HopeEDI) shared task at Association for Computational Linguistics (ACL) 2022. This shared task consists of texts in five languages, namely: English, Spanish (in Latin scripts), and Tamil, Malayalam, and Kannada (in code-mixed native and Roman scripts) with the aim of classifying the YouTube comment into "Hope", "Not-Hope" or "Not-Intended" categories. The proposed methodology uses the re-sampling technique to deal with imbalanced data in the corpus and obtained 1st rank for English language with a macro-averaged F1-score of 0.550 and weighted-averaged F1-score of 0.860. The code to reproduce this work is available in GitHub[1].

1 Introduction

Hope is vital for human health, recovery, and restoration, according to health professionals. One of the goals of Hope Speech is to express the belief that someone can get motivated to move on in life to achieve the desired goals (Chakravarthi, 2020a). Positive vibes and hope content push human beings to take steps to create a better tomorrow through sustaining optimism and resilience during hardships. The advent of social media has enabled people from all over the world to connect with each other and to express their feelings or opinions in a positive, negative, or neutral manner (Chakravarthi et al., 2021, 2022b; Sampath et al., 2022; Ravikiran et al., 2022; Bharathi et al., 2022; Priyadharshini et al., 2022). In social media, hope resides in positive and motivating content which helps to maintain healthy social media ecosystems.

The recent advancements in social media have changed the lifestyle of many people and their daily life is extended with the virtual territory of the internet and social networks. Social media platforms are influencing users' daily lives in a very large way. Users may share positive vibes and hope or motivating content with the intention of positive suggestions for peace or to overcome situations like COVID-19, war, election and etc., (Balouchzahi et al., 2021; Chakravarthi, 2020b). Several internet forums have also become popular for giving aid, advice, or support. Further, when users are going through a difficult or unfavorable moment, in addition to seeking emotional support from family, friends and relatives, they may also knock the virtual platforms to get through the situation (Ghanghor et al., 2021a,b; Yasaswini et al., 2021).

The freedom and anonymity in social media have provided users an opportunity to share their opinions and comments without revealing their identity (Balouchzahi, Fazlourrahman and Aparna, BK and Shashirekha, HL, 2021). This allows the users to share any kind of content including negative content such as abusive or hate speech and fake news. The majority of the social media analysis tasks deal with identifying the negative content such as Hate Speech, Abusive Language, Fake News, etc. (Chakravarthi, 2020b), with the aim of avoiding such content and having healthy social media. However, very few works have focused on social media analysis for positive vibes such as supportive, motivative, and hope content.

In general, Hope Speech includes words of encouragement, motivation, promise, and advice (Hossain et al., 2021). Identifying such a content and promoting them in social media can be an alternative solution for having healthy and promising social media. In this direction, HopeEDI[2] Chakravarthi et al. (2022a) shared task calls the researchers to address the challenges of Hope Speech

[1]https://github.com/anushamdgowda/Hope-speech

[2]https://sites.google.com/view/lt-edi-2022/home

Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, pages 161 - 166
May 27, 2022 ©2022 Association for Computational Linguistics

detection in YouTube comments. The objective of the shared task is to classify the YouTube comments in five languages, namely: English, Spanish in Latin scripts and Tamil, Kannada and Malayalam in code-mixed native and Roman scripts, into "Hope", "Not-Hope" or "Not- Intended" categories.

To tackle the challenges of Hope Speech detection in English and code-mixed Kannada and Malayalam texts, we - team MUCIC, present a methodology based on re-sampling the minority class ("Hope" class) and using 1D Convolutional Neural Network with Long Short-Term Memory (1D Conv-LSTM) for classification. The proposed methodology obtained **first** rank in the shared task for **English** texts with a weighted-averaged F1-score of 0.860, while the same methodology for code-mixed Malayalam and Kannada texts did not perform well to our expectations.

The rest of paper is organized as follows: Section 2 gives a brief description of the best performing teams in (LT-EDI-2021)[3] Chakravarthi and Muralidaran (2021) and Section 3 presents the proposed methodology followed by the results in Section 4. The paper concludes with the future work in Section 5.

2 Related Work

Researchers are attempting to create computational models to identify positive and supportive text on social media. Despite the various Machine learning (ML) and Deep Learning (DL) approaches for Text Classification (TC), transformers also have grown in prominence in recent years due to their ability to handle dependencies between input and output with both attention and recurrence. As a result, several Natural Language Processing (NLP) tasks such as Sentiment Analysis, Hope and Hate Speech detection etc., modeled as TC are using the transformer based models to achieve cutting-edge performance.

This section presents a summary of the models submitted to HopeEDI[4] shared task Chakravarthi (2020b). A multilingual dataset of YouTube comments in English, and code-mixed Tamil and Malayalam languages was released for public access. The dataset containing 28,451, 20,198, and 10,705 comments in English, Tamil and Malayalam languages respectively are distributed into two main categories, namely: “Hope” and “Not-Hope” (and an extra category for texts in Not-Intended language). Term Frequency-Inverse Document Frequency (TF-IDF) features were used to train k-Nearest Neighbors (kNN), Support Vector Machines (SVM), Decision Trees (DT), and Logistic Regression (LR) classifiers. Among all the classifiers, DT classifier obtained a weighted-averaged F1-scores of 0.46, 0.51, and 0.56 for English, Tamil, and Malayalam texts respectively.

Balouchzahi et al. (2021) proposed a method that utilizes a combination of TF-IDF vectors of words, char sequences, and syntactic n-grams to train: (i) a voting classifier of three estimators, namely: LR, eXtreme Gradient Boosting (XGB), and Multi-Layer Perceptron (MLP) and (ii) Keras Neural Network-based model. They also trained a Bidirectional Encoder Representations from Transformers (BERT) language model from scratch using the given dataset and then used it for Hope Speech detection. For the voting classifier, the authors obtained 1st, 2nd, and 3rd ranks with weighted-averaged F1-scores of 0.85, 0.92, and 0.59 for Malayalam, English, and Tamil texts respectively.

Dowlagar and Mamidi (2021) preprocessed texts by removing punctuation symbols, emotions, and hashtags and then transliterated Tamil and Malayalam texts back to their native scripts. Using multilingual BERT (mBERT) embedding as weights for Convolutional Neural Network (CNN) classifier, they secured 1st, 3rd, and 4th ranks for English, Malayalam, and Tamil texts. Similarly, fine-tuning mBERT for Malayalam and Tamil and using BERT for English, Arunima et al. (2021) obtained weighted-averaged F1-scores of 0.46, 0.81, 0.92 for Tamil, Malayalam, and English texts respectively. Upadhyay et al. (2021) tried two different approaches to detect Hope Speech in the HopeEDI dataset. In the first method they used contextual embeddings to train LR, Random Forest (RF), SVM, and LSTM classifiers. Using a majority voting ensemble of BERT, RoBERTa, ALBERT and LSTM models in their second model, they obtained weighted-averaged F1-scores of 0.93, 0.75, and 0.49 for English, Malayalam, and Tamil texts respectively.

In another transformer-based method M K and A P (2021), the authors used Bidirectional Long Short-Term Memory (BiLSTM), Universal Language Model Fine-tuning (ULMFiT), BERT, ALBERT, DistilBERT, Roberta, and CharBERT for English and multilingual language versions of the

[3]https://sites.google.com/view/lt-edi-2021/home

[4]https://competitions.codalab.org/competitions/27653

mentioned transformers for Tamil and Malayalam. mBERT for Malayalam and multilingual Distilbert for Tamil obtained weighted-averaged F1-scores of 0.85 and 0.59 respectively. ULMFiT model for English texts secured the 2nd rank with a weighted-averaged F1-score of 0.92.

Emojis, punctuation marks, mentions, hashtags, etc., are removed from the dataset in the study conducted by Thara et al. (2021). After cleaning the dataset, Word2Vec and FastText were used to build the feature vectors which were fed to BiLSTM using an attention-based technique. This approach obtained an weighted-averaged F1-score of 0.73 and 9th rank for Malayalam dataset.

3 Methodology

The proposed methodology consists of Pre-processing and Model construction steps to classify the given text into "Hope", "Not Hope" or "Not-Intended" categories and the framework of the proposed methodology is shown in Figure 1. Description of Pre-processing and Model construction steps are given below:

3.1 Pre-processing

Pre-processing is the task of cleaning data to remove noise to improve the quality of data for better performance. (Shashirekha et al., 2020). All the punctuation symbols, numerical data, frequently occurring words, stopwords and uninformative phrases (names that begin with @) are removed as they do not contribute to the classification task and the upper-case characters in Latin script are converted to lower-case to reduce the number of unique words.

The dataset provided by the organizers for the shared task has an uneven distribution of the target classes. This imbalanced distribution of labels over the dataset makes the classification task more challenging and ignoring this may result in the lower performance of the classification models. Hence, the data imbalance problem is addressed by using the Synthetic Minority Oversampling Technique (SMOTE)[5] (Chawla et al., 2002) technique for only English and Kannada texts. This technique increases the samples of the minority class by generating the synthetic data between each sample of the minority class based on "k" nearest neighbors and the default value of "k" (=3) is used in this work.

Languages	Labels	Datasets	Original Data	Re-Sampled Data
English	HS	Train	1962	**20778**
		Dev	272	272
		Test	–	–
	NHS	Train	20778	20778
		Dev	2569	2569
		Test	–	–
Kannada	HS	Train	1699	**3241**
		Dev	210	210
		Test	200	-
	NHS	Train	3241	3241
		Dev	408	408
		Test	413	-
	NK	Train	0	-
		Dev	0	-
		Test	5	-
Malayalam	HS	Train	1668	-
		Dev	190	-
		Test	194	-
	NHS	Train	6205	-
		Dev	784	-
		Test	776	-
	NM	Train	0	-
		Dev	0	-
		Test	101	-

Table 1: Distribution of labels in the dataset before and after re-sampling (Dev: Development, HS: Hope Speech, NHS: Not Hope Speech, NM: Not Malayalam, NK: Not Kannada)

3.2 Model Construction

The texts are tokenized and converted to sequences using TensorFlow Keras[6] tokenizer API and "texts_to_sequences" function. The vocabulary size and the maximum length of the sequences has been set to 15,000 and 50 respectively. The "pad_sequences" is used to ensure that all sequences in a list have the same length. After creating a padded sequence for text, data is passed as input to the Keras embedding layer. An embedding matrix derived from Keras embedding layer and a one-hot representation of the labels are fed into a 1D CNN-LSTM architecture. The parameters "input_dim", "output_dim" and "input_length" in embedding layers are set to 15,000 (vocabulary size), 1,000 (length of the word vector), and 50 (maximum length of a sequence) respectively. The convolutional layers with 64 filters, two pooling layers, and a relu activation function are used for the Conv1D layer, along with 100 fully connected LSTM layers, and a soft-max output layer.

4 Experiments and Results

Several experiments were conducted to classify a YouTube comment into "Hope", "Not-Hope" or "Not-Intended" categories for each language and

[5]https://pypi.org/project/imbalanced-learn/

[6]https://www.tensorflow.org/api_docs/python/tf/keras

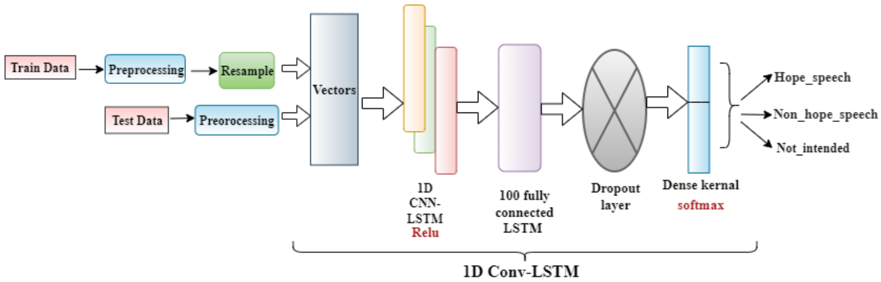


Figure 1: Framework of the proposed 1D Conv-LSTM model

the models that performed well on the Development sets were applied to the Test sets for the evaluation.

4.1 Dataset

The dataset provided by the shared task organizers includes English, Spanish, and code-mixed Tamil, Malayalam, and Kannada texts. However, the current work focuses solely on English, Tamil and Kannada texts and the distribution of labels across the Train and Development sets for these languages are shown in Table 1. The size of the re-sampled data using SMOTE technique for English and Kannada texts is also shown in Table 1. Further, the Test sets consists of 389, 1,070, 1,760 unlabeled samples for English, Malayalam, and Kannada languages respectively.

4.2 Results and Analysis

The proposed models are evaluated on the unlabeled Test set provided by the organizers and the predictions are graded based on macro-averaged F1-score (M_F1-score) and weighted-averaged F1-score (W_F1-score). The results on the Development set shown in Table 2 illustrates that the proposed methodology performed better for English and Kannada texts than for Malayalam texts. The results on final leaderboard revealed that the proposed methodology obtained 1^{st} rank for English with an M_F1-score of 0.550, but the results for Kannada and Malayalam texts are significantly lower than the expectation and the results on the Test sets are given in Table 3. The comparison of macro-averaged F1-scores of the proposed methodology with the top 5 macro-averaged F1-scores in the shared task is presented in Figure 2 (since the best performing teams for each language are different, the best scores obtained in each language are mentioned). The observation of datasets reveal that unlike Train and Development sets, the Test set provided by the organizers include a extra label called

Languages / Scores		English	Kannada	Malayalam
M_F1-scores	P	0.63	0.64	0.51
	R	0.60	0.64	0.64
	F1	0.61	0.64	0.53
W_F1-scores	P	0.87	0.68	0.73
	R	0.88	0.67	0.60
	F1	0.78	0.67	0.63

Table 2: The results on the Development sets (P: Precision, R: Recall, F1: F1-score)

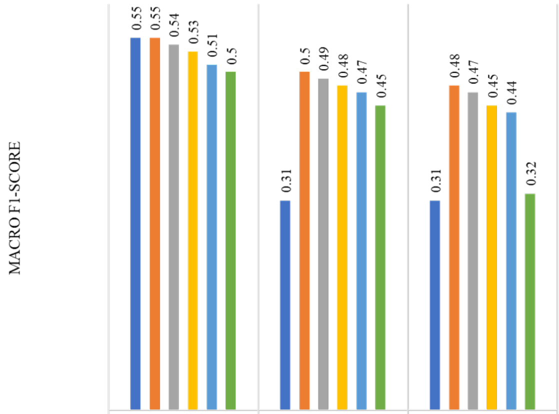
Languages / Scores		English	Kannada	Malayalam
M_F1-scores	P	0.540	0.310	0.310
	R	0.550	0.310	0.320
	F1	0.550	0.310	0.310
W_F1-scores	P	0.870	0.520	0.560
	R	0.850	0.530	0.580
	F1	0.860	0.520	0.570
Rank		1	6	7

Table 3: The results on the Test sets

"Not-Kannada" with 5 samples for Kannada and "Not-Malayalam" with 101 samples for Malayalam. As the Train set did not include these two labels, the proposed model failed to predict these labels and this is the main reason for low performance for Kannada and Malayalam texts.

5 Conclusion

This paper describes the models submitted to HopeEDI shared task at ACL 2022 to classify the YouTube comments in English and code-mixed Kannada and Malayalam texts into "Hope", "Not-Hope" or "Not-Intended" categories. The proposed methodology addresses the Hope Speech detection by using SMOTE technique to resolve the data imbalance problem and 1D Conv-LSTM model for classification. For English texts, the proposed methodology performed the best and achieved 1^{st} rank with a M_F1-score of 0.550 but did not perform well for Kannada and Malayalam texts. As future work, we would like to extend our experiments on various feature types such as stylistic and



	English	Malayalam	Kannada
MUCIC	0.55	0.31	0.31
First rank	0.55	0.5	0.48
Second rank	0.54	0.49	0.47
Thirs rank	0.53	0.48	0.45
Forth rank	0.51	0.47	0.44
Fifth rank	0.5	0.45	0.32

Figure 2: Comparison of macro-averaged F1-scores of the proposed methodology with 5 top macro-averaged F1-scores in the shared task

psychological and explore different word embeddings along with language models.

References

S Arunima, Akshay Ramakrishnan, Avantika Balaji, Thenmozhi D., and Senthil Kumar B. 2021. ssn_diBERTsity@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*, pages 92–97, Kyiv. Association for Computational Linguistics.

Fazlourrahman Balouchzahi, BK Aparna, and HL 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*, pages 180–187.

Balouchzahi, Fazlourrahman and Aparna, BK and Shashirekha, HL. 2021. MUCS@DravidianLangTech-EACL2021: COOLI-Code-Mixing Offensive Language Identification. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 323–329.

B Bharathi, Bharathi Raja Chakravarthi, Subalalitha Chinnaudayar Navaneethakrishnan, N Sripriya, Arunaggiri Pandian, and Swetha Valli. 2022. Findings of the shared task on Speech Recognition for Vulnerable Individuals in Tamil. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.

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. 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.

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 Phillip 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. 2022a. Findings 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*. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Thenmozhi Durairaj, John Phillip McCrae, Paul Buitaleer, Prasanna Kumar Kumaresan, and Rahul Ponnusamy. 2022b. Findings of the shared task on Homophobia Transphobia Detection in Social Media Comments. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Kayalvizhi Sampath, Durairaj Thenmozhi, Sathiyaraj Thangasamy, Rajendran Nallathambi, and John Phillip McCrae. 2021. Dataset for identification of homophobia and transophobia in multilingual YouTube comments. *arXiv preprint arXiv:2109.00227*.

Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of artificial intelligence research*, 16:321–357.

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*, pages 86–91, Kyiv. Association for Computational Linguistics.

Nikhil Ghanghor, Parameswari Krishnamurthy, Sajeetha Thavareesan, Ruba Priyadharshini, and Bharathi Raja Chakravarthi. 2021a. IIITK@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*, pages 222–229, Kyiv. Association for Computational Linguistics.

Nikhil Ghanghor, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021b. IIITK@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*, pages 197–203, Kyiv. 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. *arXiv preprint arXiv:2103.00464*.

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*, pages 79–85, Kyiv. Association for Computational Linguistics.

Ruba Priyadharshini, Bharathi Raja Chakravarthi, Subalalitha Chinnaudayar Navaneethakrishnan, Thenmozhi Durairaj, Malliga Subramanian, Kogilavani Shanmugavadivel, Siddhanth U Hegde, and Prasanna Kumar Kumaresan. 2022. Findings of the shared task on Abusive Comment Detection in Tamil. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Manikandan Ravikiran, Bharathi Raja Chakravarthi, Anand Kumar Madasamy, Sangeetha Sivanesan, Ratnavel Rajalakshmi, Sajeetha Thavareesan, Rahul Ponnusamy, and Shankar Mahadevan. 2022. Findings of the shared task on Offensive Span Identification in code-mixed Tamil-English comments. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Anbukkarasi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, Ruba Priyadharshini, Subalalitha Chinnaudayar Navaneethakrishnan, Kogilavani Shanmugavadivel, Sajeetha Thavareesan, Sathiyaraj Thangasamy, Parameswari Krishnamurthy, Adeep Hande, Sean Benhur, Kishor Kumar Ponnusamy, and Santhiya Pandiyan. 2022. Findings of the shared task on Emotion Analysis in Tamil. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

HL Shashirekha, MD Anusha, and Nitin S Prakash. 2020. Ensemble Model for Profiling Fake News Spreaders on Twitter. In *CLEF (Working Notes)*.

S Thara, Ravi teja Tasubilli, et al. 2021. Amrita@ lt-edi-eacl2021: Hope Speech Detection on Multilingual Text. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 149–156.

Ishan Sanjeev Upadhyay, Anshul Wadhawan, Radhika Mamidi, et al. 2021. Hopeful_men@ lt-edi-eacl2021: Hope Speech Detection using Indic Transliteration and Transformers. *arXiv preprint arXiv:2102.12082*.

Konthala Yasaswini, Karthik Puranik, Adeep Hande, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. IIITT@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*, pages 187–194, Kyiv. Association for Computational Linguistics.