Findings of the Shared Task on Caste and Migration Hate Speech Detection

Saranya Rajiakodi[1], **Bharathi Raja Chakravarthi**[2], **Rahul Ponnusamy**[2], **Shunmuga Priya MC**[2], **Prasanna Kumar Kumaresan**[2], **Sathiyaraj Thangasamy**[3], **Bhuvaneswari Sivagnanam**[1], **Balasubramanian Palani**[4], **Kogilavani Shanmugavadivel**[5], **Abirami Murugappan**[6], **Charmathi Rajkumar**[7]

[1] Central University of Tamil Nadu, India

[2] School of Computer Science, University of Galway, Ireland

[3]Department of Tamil, Sri Krishna Adithya College of Arts and Science, Tamil Nadu, India

[4]Indian Institute of Information Technology Kottayam, Kerala, India [5]Kongu Engineering College, Tamil Nadu, India

[6]Department of Information Science and Technology, Anna University, India [7]The American College, Tamil Nadu, India

Abstract

Hate speech targeting caste and migration communities is a growing concern in online platforms, particularly in linguistically diverse regions. By focusing on Tamil language text content, this task provides a unique opportunity to tackle caste or migration related hate speech detection in a low resource language Tamil, contributing to a safer digital space. We present the results and main findings of the shared task caste and migration hate speech detection. The task is a binary classification determining whether a text is caste/migration related hate speech or not. The task attracted 17 participating teams, experimenting with a wide range of methodologies from traditional machine learning to advanced multilingual transformers. The top performing system achieved a macro F1-score of 0.88105, enhancing an ensemble of fine-tuned transformer models including XLM-R and MuRIL. Our analysis highlights the effectiveness of multilingual transformers in low resource, ensemble learning, and culturally informed socio political context based techniques.

1 Introduction

In the current digital world, the online social network platforms such as Twitter, YouTube, LinkedIn, Facebook and WhatsApp are widely used by the individuals from the different parts of the country (Kruse et al., 2018; Kubin and von Sikorski and, 2021). Due to the high-speed internet facility, the news like an offensive speech (Sreelakshmi et al., 2024), fake news (Subramanian et al., 2025), hate speech and its curated forms against the caste or people are disseminated across the globe in a fraction of minutes (Rajiakodi et al., 2024; Chakravarthi et al., 2025). Hence, it leads to lot of issues such as protests, inflation in the economy of the country and majorly affects the healthy environment in the society (Jost et al., 2018). The hate speech and offensive words are more important to eradicate from the society to preserve the equality among the citizens. Apart from the global issues, the discrimination and bias against the certain communities and its people will affect directly their communities and their people in terms of their mental health (Abubakar et al., 2022; Matamoros-Fernández and Farkas, 2021). Hence, the hate speech against any caste/migration is an important issue and it should be detected to save the society and nation's peace.

India is a democratic country and follows the "Unity in Diversity". According to the Constitution of India [1], Articles 14-18 deal with the right to equality. Hence if any people discriminate others in terms of language, caste or any variants, the government will take immediate actions against them with these articles. With the proliferation of social media platforms, public discourse has become increasingly dynamic, often marked by a surge of user-generated content in response to specific events (Shanmugavelan, 2022). This environment has made the manual identification of caste-based hate speech both time-consuming and resource-intensive. To address this challenge, there has been a growing adoption of Natural Language Processing (NLP) based systems capable of automatically detecting hate speech across multiple languages and platforms, thereby enhancing the efficiency and scalability of monitoring efforts (Roy et al., 2022; Sharma et al., 2025).

Researchers are developed various machine learning (ML) and deep learning (DL) model to identify offensive language and hate speech from the various regional low-resource languages such as Hindi (Rani et al., 2020; Rajak and Baruah, 2025), Telugu, Malayalam and Tamil (G et al., 2025). A novel dataset which comprises of hate speech by castes and migration collected from

[1]https://bit.ly/44tqpwg

Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity, Inclusion, pages 214–220

YouTube and published in LT-EDI@EACL 2024 (Rajiakodi et al., 2024). In this shared task, the team of participants have developed an automatic hate speech detection against the caste and migration model.

Specifically, the top ranking teams used an ensemble model which consists of transformer based multilingual models such as XLM-RoEERTa and MuRIL for contextual embedding of sentences to predict hate speech that achieves higher macro F1 score among other models. The work by (Alam et al., 2024) developed six different ML models and three DL models including Bi-LSTM, Attention, Bi-LSTM-CNN to identify the hate or not-hate speech. Finally, they designed transformer-based models such as M-BERT, XLM-R and Tamil-BERT for effective contextual word embedding, which achieves better performance than the other models.

In the subsequent sections, we have discussed the task description, dataset statistics, methodologies used by the participants to detect caste/migration hate speech in Tamil, and their results and ranking.

2 Tasks and Dataset

2.1 Task Description

The task's goal is to classify YouTube text comments in Tamil into two categories: hate speech on caste/migration, non-hate speech on caste/migration. Participants were provided with:

- **Training and Development Sets**: These sets were annotated with labels to allow participants to train and fine tune their models effectively.

- **Testing Set**: This set was unlabeled, requiring participants to generate predictions without the aid of ground truth labels, which were reserved for evaluation purposes.

2.2 Dataset Description

The dataset has been carefully curated from the YouTube platform by collecting comments from videos discussing caste and migration related issues. To identify relevant videos, we utilized a combination of hashtags such as *'vadakan'* (North Indian), *'devar'* (a caste), alongside manual keyword searches including terms like *'Agamudiyar'* (a caste group) and *'Melpathi temple issue'* (a specific caste related controversy).

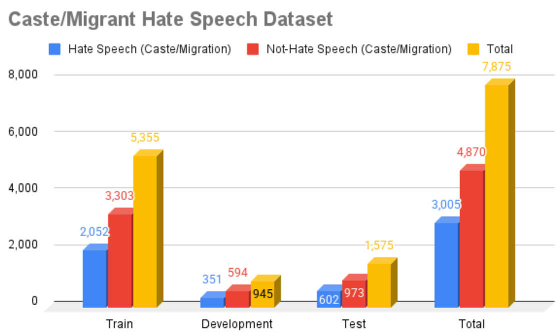

Figure 1: Dataset Statistics

After gathering the comments, we applied preprocessing steps to remove unrelated or noisy content. This included removing usernames, URLs, and comments containing fewer than three words to ensure data quality and relevance.

Each comment was then annotated independently by three annotators, all proficient in the Tamil language. The annotation team was deliberately heterogeneous, composed of individuals from diverse caste backgrounds—including historically marginalized and dominant groups various age ranges from young adults (20–29 years) to middle aged adults (30–45 years), as well as gender and professional diversity. This diversity in the annotator pool was intended to reduce bias and enhance the reliability of the annotations.

The annotation labels are defined as follows:

- **Hate speech on caste/migration:** Comments containing abusive, disrespectful, or discriminatory language, including ridicule or mockery, and content aimed at delegitimizing specific caste or migrant groups.

- **Not-Hate speech on caste/migration:** Comments that do not contain any discriminatory or derogatory remarks towards individuals based on their caste or migration status.

The reliability of the annotations was measured using Krippendorff's alpha, which yielded a high agreement score of 0.83441, indicating strong consistency among annotators. The detailed statistics of the dataset is shown in Figure 1.

3 Results

The evaluation metric used was the macro-averaged F1-score, calculated using the scikit-learn library[2].

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

3.1 Participant Statistics

Our competition was created on Codalab [3], and has attracted registrations and a total of 17 submissions. Team Samsung Research CUET_N317 won first place in the task, holding a significant lead over the second place team. Teams CUET's_white_walkers and Wise respectively captured the second and third places. The official leaderboards for this task is shown in Table 1.

3.2 Participant Methodology

In this section, we first summarize common features for all teams based on the information they provided in the Google Documents. Then, we delve into the methods employed by the top 5 teams, accompanied by brief descriptions of the approaches utilized. The approaches of all other teams are presented briefly in table 1.

- **CUET_N317 (Md. Nur Siddik Ruman, 2025)**: This team employed three distinct strategies: traditional machine learning models with TF-IDF features, individual fine-tuned multilingual transformers (e.g., XLM-R, MuRIL, IndicBERT), and an ensemble of five fine-tuned multilingual transformer models. The ensemble model achieved the best performance. Extensive hyperparameter tuning and early stopping techniques were used to optimize results.

- **CUET's_White_Walkers (Jidan Al Abrar, 2025)**: The team used a fine-tuned TamilBERT model, incorporating techniques such as cosine annealing and the AdamW optimizer. They emphasized training stability and leveraged visualizations to analyze performance. The model was optimized for Tamil text without relying on transliteration or external datasets.

- **Wise (Ganesh Sundhar S, 2025)**: They explored both transliterated and original Tamil text preprocessing. A hybrid approach combined TF-IDF, Truncated SVD, MLP, and multilingual transformer embeddings. Feature fusion and late decision fusion were used to aggregate predictions. The team emphasized the comparative value of transliteration-aware vs. native Tamil models.

- **CUET_blitz_aces**: This team adopted a three-phase pipeline: initial training with traditional ML (TF-IDF + classifiers), fine-tuning multilingual transformers, and a final ensemble voting mechanism. They conducted thorough evaluation and analysis at each stage to select optimal model configurations.

- **hinterwelt (MD AL AMIN, 2025)**: The team experimented with XLM-R Large, MuRIL-large-cased, and IndicBERT models. Training included advanced strategies like learning rate schedulers, gradient clipping, and early stopping. The final system combined multiple transformer models, optimized with rigorous experimentation for multilingual hate speech detection.

- **ItsAllGoodMan (Amritha Nandini KL, 2025)**: Combined train/dev data, used back translation for data augmentation, segmented hashtags, replaced mentions, and converted emojis. Tried TF-IDF with Random Forest, soft voting ensembles with various embeddings and ML models, and ensembles of fine-tuned transformers.

- **NS**: Transliterated Tamil text to English and vice versa for consistent embeddings. Trained six ML models on different embedding sets and selected top three (XGBoost, Logistic Regression and MLP) for final predictions.

- **CUET_perceptrons** Trained several transformer-based models and combined their predictions using a weighted ensemble, where better-performing models had higher influence on the final label.

- **KCRL**: Combined TamilBERT and distilbert-base-multilingual-cased with k-fold cross-validation. Used a classification architecture that concatenated CLS token, mean pooling, and max pooling.

- **Cuet_try_NLP**: Used a traditional ML pipeline with Bag of Words (unigrams, bigrams) and Random Forest for classification. Focused on computational efficiency but lacked semantic understanding.

- **Wictory**: Hybrid approach: fine-tuned MuRIL-based transformer with discriminative learning rates and dense layers, plus multilingual embeddings fed into an SVM. Combined

html

[3]https://codalab.lisn.upsaclay.fr/competitions/21884

Rank	Team Name	Run	Macro F1 Score	Highlighted Methodology
1	**CUET_N317** (Md. Nur Siddik Ruman, 2025)	2	**0.88105**	Fine-tuned multilingual transformer model
2	CUET's_white_walkers (Jidan Al Abrar, 2025)	1	0.86289	Fine tuned **Tamil-BERT** with cosine annealing, AdamW, and performance visualization.
3	Wise (Ganesh Sundhar S, 2025)	1	0.81827	Transliteration vs Non-transliteration preprocessing, TF-IDF + Truncated SVD + MLP + Transformer fusion.
4	CUET_blitz_aces (Shahriar Farhan Karim, 2025)	3	0.81682	3-phase: ML (TF-IDF), transformer fine-tuning, and ensemble voting of 5 models.
5	hinterwelt (MD AL AMIN, 2025)	2	0.80916	Fine-tuned **XLM-R Large**, MuRIL-large-cased, IndicBERT with LR scheduler, gradient clipping, early stopping.
6	ItsAllGoodMan (Amritha Nandini KL, 2025)	3	0.80364	TF-IDF + Random Forest (best), back translation for augmentation, MuRIL + voting ensemble.
7	NS (Nishanth S, 2025)	1	0.80095	Tamil English transliteration embeddings + XGBoost, Logistic Regression, MLP.
8	CUET_perceptrons	2	0.79812	Weighted ensemble of multiple transformer models.
9	KCRL	1	0.79081	Triple embedding (CLS, mean, max) + TamilBERT and DistilBERT with cross-validation.
10	Cuet_try_NLP	2	0.78175	Bag-of-Words + n-grams + Random Forest.
11	Wictory	1	0.76630	MuRIL + SVM hybrid with focal loss, dense layers, class weighting.
12	Solvers (Mohanapriya K T, 2025)	2	0.76518	CNN-based model + BERT fine-tuning.
13	DravLang	1	0.76182	BiLSTM + Attention + Voting and Stacking Classifiers (XGBoost, SVM, NB).
14	girlsTeam (Towshin Hossain Tushi, 2025)	1	0.74522	MuRIL/mBERT/IndicBERT + BiLSTM + ML classifiers + multilingual embeddings.
15	ScalarLab	1	0.73308	M-BERT and XLM-R; ensemble predictions from both.
16	SSN_IT (Maria Nancy C, 2025)	1	0.72462	BERT multilingual model + preprocessing Tamil and Romanized Tamil.
17	EM-26 (Tewodros Achamaleh, 2025)	1	0.65672	XLM-Roberta + AdamW

Table 1: Leader board Results with Methodologies

outputs for interpretability and deep contextual understanding.

- **Solvers (Mohanapriya K T, 2025)**: Explored deep learning models starting with CNNs, then modified for multi-class with softmax. Also fine-tuned a BERT-based model for contextual understanding and compared their effectiveness.
- **DravLang**: Hybrid approach: preprocessed text with tokenization and TF-IDF, trained a BiLSTM with Attention, an ensemble Voting Classifier (XGBoost, SVM, Naïve Bayes), and a Stacking Classifier (XGBoost, SVM, Logistic Regression meta-learner)
- **girlsTeam (Towshin Hossain Tushi, 2025)**: Comprehensive hybrid: transformer-based embeddings (MuRIL, mBERT, IndicBERT), enhanced with BiLSTM, and traditional ML classifiers (Random Forest, SVM, XGBoost) using TF-IDF, FastText, and Word2Vec. Included tokenization, normalization, transliteration, and class balancing.
- **ScalaraLab**: Used Multilingual BERT and XLM-R models, then combined their outputs for predictions.
- **SSN_IT (Maria Nancy C, 2025)**: Fine-tuned bert-base-multilingual-cased on Tamil and Romanized Tamil. Preprocessing included lowercasing, URL/special character removal, and BERT tokenization. Used AdamW optimizer, cross-entropy loss, and evaluated with standard metrics.
- **EM-26 (Tewodros Achamaleh, 2025)**: Finetuned XLM-RoBERTa , with class weighting and augmentation (word swapping/dropping). Used AdamW, learning rate scheduler, and early stopping based on F1 score.

The leaderboard results and methodological descriptions reveal a clear trend toward the effectiveness of multilingual transformer-based models for caste and migration-related hate speech detection. Teams that employed fine-tuned transformer architectures such as XLM-R, MuRIL, and IndicBERT often in ensemble settings consistently outperformed traditional machine learning approaches. The top-performing submissions integrated advanced training strategies, including learning rate schedulers, gradient clipping, early stopping, and ensemble voting, to boost performance and ensure robustness. Additionally, teams that conducted rigorous ablation studies, language-specific preprocessing, and hybrid model experimentation (eg: combining TF-IDF with deep models) demonstrated a deeper understanding of the multilingual and culturally nuanced nature of the task. Overall, the results highlight that a blend of domain-specific linguistic preprocessing and transformer-based modeling yields the most competitive outcomes in multilingual hate speech detection tasks.

4 Conclusion

This shared task on caste and migration hate speech detection focused on detecting hate speech against people based on their caste or migration status, especially in the Tamil language. Since this type of hate is often ignored in existing AI systems, this task helped researchers and developers create models that are more aware of these social issues. Participants used many different methods, including machine learning and advanced models like BERT and XLM-R, to analyze the text. The best teams used a mix of good text cleaning, careful model training, and techniques like ensemble voting and data balancing to get better results. The top system scored a high macro F1 of 0.88105, showing that with the right tools and approaches, it is possible to detect even subtle and hidden forms of hate speech. This task showed that we need models that understand regional languages and cultural context to handle online hate more effectively. It also highlights the need for more work in areas like explaining model predictions, creating respectful counter-speech, and labeling more detailed types of hate.

Acknowledgments

Author Bharathi Raja Chakravarthi was funded by a research grant from Research Ireland under grant number SFI/12/RC/2289_P2 (Insight). This work is also supported by the Centre for Research Training in Artificial Intelligence grant number SFI/18/CRT/6223, as well as a grant from the College of Science and Engineering, University of Galway, Ireland.

References

Ibrahim Abubakar, Lu Gram, Sarah Lasoye, E. Tendayi Achiume, Laia Becares, Gurpreet Kaur Bola, Rageshri Dhairyawan, Gideon Lasco, Martin McKee, Yin Paradies, Nidhi S. Sabharwal, Sujitha Selvarajah, Geordan Shannon, and Delan Devakumar. 2022. Confronting the consequences of racism, xenophobia, and discrimination on health and health-care systems. *The Lancet*, 400(10368):2137–2146. PMID: 36502851.

Md Alam, Hasan Mesbaul Ali Taher, Jawad Hossain, Shawly Ahsan, and Mohammed Moshiul Hoque. 2024. CUET_NLP_Manning@LT-EDI 2024: Transformer-based approach on caste and migration hate speech detection. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion*, pages 238–243, St. Julian's, Malta. Association for Computational Linguistics.

Giri Prasath R Anerud Thiyagarajan Sachin Kumar S Amritha Nandini KL, Vishal S. 2025. Itsallgoodman@lt-edi-2025: Fusing tf-idf and muril embeddings for detecting caste and migration hate speech. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

B. R. Chakravarthi, S. Rajiakodi, R. Ponnusamy, Bhuvaneswari Sivagnanam, Sara Yogesh Thakare, and Sathiyaraj Thangasamy. 2025. Detecting caste and migration hate speech in low-resource tamil language. *Language Resources and Evaluation*.

Jyothish Lal G, Premjith B, Bharathi Raja Chakravarthi, Saranya Rajiakodi, Bharathi B, Rajeswari Natarajan, and Ratnavel Rajalakshmi. 2025. Overview of the shared task on multimodal hate speech detection in Dravidian languages: DravidianLangTech@NAACL 2025. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 114–122, Acoma, The Albuquerque Convention Center, Albuquerque, New Mexico. Association for Computational Linguistics.

Gnanasabesan G Hari Krishnan N Dhanush MC Ganesh Sundhar S, Durai Singh K. 2025. Wise@ltedi2025: Combining classical and neural representations with multiscale ensemble learning for codemixed hate speech detection. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Ariful Islam Md Mehedi Hasan Md Mubasshir Naib Mohammad Shamsul Arefin Jidan Al Abrar, Md Mizanur Rahman. 2025. Cuets_white_walkers@lt-edi 2025: Transformerbased model for the detection of caste and migration hate speech. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

John T. Jost, Pablo Barbera, Richard Bonneau, Max Langer, Miriam Metzger, Jonathan Nagler, Josh Sterlin, and Joshua A. Tucker. 2018. How social media facilitates political protest: Information, motivation, and social networks. *Advances in Political Psychology*, 39(S1):85–118.

Lisa M. Kruse, Dawn R. Norris, and Jonathan R. Flinchum and. 2018. Social media as a public sphere? politics on social media. *The Sociological Quarterly*, 59(1):62–84.

Emily Kubin and Christian von Sikorski and. 2021. The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association*, 45(3):188–206.

Swathika R Maria Nancy C, Radha N. 2025. Ssn_it_hate@ltedi2025: Caste and migration hate speech detection. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Ariadna Matamoros-Fernández and Johan Farkas. 2021. Racism, hate speech, and social media: A systematic review and critique. *Television & New Media*, 22(2):205–224.

Md. Abdur Rahman Md Sajid Hossain Khan Md Ashiqur Rahman MD AL AMIN, Sabik Aftahee. 2025. Hinterwelt@lt-edi 2025: A transformer-based detection of caste and migration hate speech in tamil social media. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Hasan Murad Md. Nur Siddik Ruman, Md. Tahfim Juwel Chowdhury. 2025. Cuet_n317@lt-edi2025: Detecting hate speech related to caste and migration with transformer models. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Devasri A Bharath P Ananthakumar S Mohanapriya K T, Anirudh Sriram K S. 2025. Solvers@lt-edi-2025: Caste and migration hate speech detection in tamil-english code-mixed text. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Sachin Kumar S Nishanth S, Shruthi Rengarajan. 2025. Ns@ltedi2025 caste migration based hate speech detection. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Santosh Rajak and Ujwala Baruah. 2025. A comprehensive hindi hostile post detection dataset: Aannotated resource for fine-grained hostility analysis on twitter posts in the hindi language. *Expert Systems with Applications*, 289:128191.

Saranya Rajiakodi, Bharathi Raja Chakravarthi, Rahul Ponnusamy, Prasanna Kumaresan, Sathiyaraj Thangasamy, Bhuvaneswari Sivagnanam, and Charmathi Rajkumar. 2024. Overview of shared task on caste and migration hate speech detection. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion*, pages 145–151, St. Julian's, Malta. Association for Computational Linguistics.

Priya Rani, Shardul Suryawanshi, Koustava Goswami, Bharathi Raja Chakravarthi, Theodorus Fransen, and John Philip McCrae. 2020. A comparative study of different state-of-the-art hate speech detection methods in Hindi-English code-mixed data. In *Proceedings of the second workshop on trolling, aggression and cyberbullying*, pages 42–48.

Pradeep Kumar Roy, Snehaan Bhawal, and Chinnaudayar Navaneethakrishnan Subalalitha. 2022. Hate speech and offensive language detection in dravidian languages using deep ensemble framework. *Computer Speech Language*, 75:101386.

Hasan Murad Shahriar Farhan Karim, Anower Sha Shajalal Kashmary. 2025. Cuet_blitz_aces@lt-edi-2025: Leveraging transformer ensembles and majority voting for hate speech detection. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Murali Shanmugavelan. 2022. Caste-hate speech and digital media politics. *Journal of Digital Media amp; Policy*, 13(Special Issue: '(Re)Iterations, Transgressions, Recognition: Politics and Practices of Media Policies in South Asia'):41–55.

Deepawali Sharma, Vedika Gupta, Vivek Kumar Singh, and Bharathi Raja Chakravarthi. 2025. Stop the hate, spread the hope: An ensemble model for hope speech detection in english and dravidian languages. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* Just Accepted.

K. Sreelakshmi, B. Premjith, Bharathi Raja Chakravarthi, and K. P. Soman. 2024. Detection of Hate Speech and Offensive Language CodeMix Text in Dravidian Languages Using Cost-Sensitive Learning Approach. *IEEE Access*, 12:20064–20090.

Malliga Subramanian, Premjith B, Kogilavani Shanmugavadivel, Santhiya Pandiyan, Balasubramanian Palani, and Bharathi Raja Chakravarthi. 2025. Overview of the shared task on fake news detection in Dravidian languages-DravidianLangTech@NAACL 2025. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 759–767, Acoma, The Albuquerque Convention Center, Albuquerque, New Mexico. Association for Computational Linguistics.

Mikiyas Mebiratu Sara Getachew Grigori Sidorov Tewodros Achamaleh, Abiola T. O. 2025. Em-26@lt-edi 2025: Caste and migration hate speech detection in tamil-english code-mixed social media texts. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Rehenuma Ilman Samia Rahma Towshin Hossain Tushi, Walisa Alam. 2025. girlsteam@lt-edi-2025:

Caste/migration based hate speech detection. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).