Findings of the Shared Task on Abusive Tamil and Malayalam Text Targeting Women on Social Media: DravidianLangTech@NAACL 2025

Saranya Rajiakodi[1], **Bharathi Raja Chakravarthi**[2], **Shunmuga Priya Muthusamy Chinnan**[2], **Ruba Priyadharshini**[3], **Rajameenakshi J**[4], **Kathiravan P**[1], **Rahul Ponnusamy**[5], **Bhuvaneswari Sivagnanam**[1], **Paul Buitelaar**[5], **Bhavanimeena K**[1], **Jananayagam V**[1], **Kishore Kumar Ponnusamy**[6]

[1]Department of Computer Science, Central University of Tamil Nadu, India.
[2]School of Computer Science, University of Galway, Ireland
[3]Department of Mathematics, Gandhigram Rural Institute -Deemed to be university, India
[4]Department of Social Sciences, Vellore Institute of Technology, Vellore, India
[5]Data Science Institute, University of Galway, Ireland. [6]Digital University of Kerala, Kerala, India.

Abstract

This overview paper presents the findings of the Shared Task on Abusive Tamil and Malayalam Text Targeting Women on Social Media, organized as part of DravidianLangTech @NAACL 2025. The task aimed to encourage the development of robust systems to detect abusive content targeting women in Tamil and Malayalam, two low-resource Dravidian languages. Participants were provided with annotated datasets containing abusive and nonabusive text curated from YouTube comments. We present an overview of the approaches and analyse the results of the shared task submissions. We believe the findings presented in this paper will be useful to researchers working in Dravidian language technology.

Disclaimer: This research paper contains offensive/harmful content for research purposes. Viewer discretion is advised.

1 Introduction

As per United Nations, one in three of all women experience Gender Based Violence (GBV) at least once in their lives[1]. Exposure to abusive content on social networks significantly affects people's emotional states (Soral et al., 2023). However, systems for detecting abusive content targeting women have been well developed and widely deployed for English, there is a significant gap in resources and models for Dravidian languages such as Tamil and Malayalam (Caselli et al., 2020; Park and Fung, 2017). The motivation behind conducting this shared task is to address the growing issue of gender-based online harassment, particularly targeting women, in this digital era of social media platforms (Pandey, 2024; Battisti et al., 2024).

This task focuses on identifying abusive text directed at women in YouTube comments. Given a sentence in Tamil or Malayalam, the goal is to contextually analyze the text and determine whether it contains abusive text that specifically targets women.

Tamil and Malayalam, both Dravidian languages rich in literary heritage and agglutinative structures, are widely spoken in Tamil Nadu, Sri Lanka, Kerala, Singapore and Malaysia. This shared task of identifying abusive text in Tamil and Malayalam languages poses challenges such as handling grammatical and spelling errors, managing code-switching between languages, and detecting specific text patterns that target women. This shared task provides an avenue for researchers to classify abusive content in Tamil and Malayalam languages.

Recent growth in multilingual Large Language Models(LLMs), such as LLaMA, mBERT, XLM-R, and IndicBERT, has significantly improved the accuracy of systems for Indian languages (Touvron et al., 2023; Devlin et al., 2019; Kakwani et al., 2020). However, the effectiveness of abusive content detection is highly dependent on the quality of training data. This task utilizes a dataset that is carefully annotated with guidelines in Tamil and Malayalam languages.

2 Related Works

Recent advances in NLP, driven by open source language models, have significantly improved performance on a variety of tasks. However, the detection of abusive content, particularly in Indian languages, is less explored (Mohan et al., 2025; Shunmuga Priya et al., 2022). Languages like Tamil and Malayalam present unique challenges due to their agglutinative nature, frequent code mixing, and the limited availability of large annotated datasets. However, few works have analyzed the ex-

[1]https://www.unwomen.org/en/articles/facts-and-figures/facts-and-figures-ending-violence-against-women

Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 75–85
May 3, 2025

isting challenges in hate speech detection for Indian languages (Mandl et al., 2021). Ponnusamy et al. (2024) has contributed significantly by providing annotated datasets for offensive content detection, laying the foundation for improving NLP models in Dravidian languages.

Subramanian et al. (2022) conducted a comparative study on three variants of transformer models for the detection of hate speech in Tamil. In 2022, Chakravarthi et al. (2022) organized a shared task on the detection of hope speech in Tamil and Malayalam, evaluating system performance using the F1 score. In 2024, another shared task on detecting homophobia and transphobia in social media comments has gained significant attention, addressing the challenges of identifying hateful content in various Indian languages, including Tamil and Malayalam (Chakravarthi et al., 2024). Another critical area of research is code mixing in Dravidian languages, B et al. (2024) work focused on decoding YouTube comments in code mixed Tamil-English and Malayalam-English.

Understanding abusive content involves the analysis of its grammatical structure. Syntactic and semantic ambiguities play an important role in the identification of abusive language (Waseem and Hovy, 2016). This linguistic complexity also complicates the generation of effective word embeddings (Miaschi and Dell'Orletta, 2020), making it challenging to capture the underlying meanings and relationships, especially in abusive language directed at women.

Most of the existing work focuses on general abusive content, but there is a significant gap in detecting gender-specific abuse, especially directed toward women. This shared task aims to fill this gap by encouraging participants to develop models focused on detecting such targeted abuse.

3 Task Description

The shared task challenge was hosted in Codalab [2]. The task's goal is to classify YouTube comments into two categories: Abusive and Non-Abusive in Tamil and Malayalam languages. Participants were provided with:

- Training and Validation 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.

The availability of pre-trained models can help participants address the challenge of linguistic, contextual, and cultural variations by generating meaningful feature representations.

3.1 Dataset Statistics

The dataset was prepared in Tamil and Malayalam, which are low-resource languages. YouTube comments were scraped using targeted queries focused on controversial and sensitive topics where abuse text against women is commonly found. The dataset collection and annotation process has been illustrated in the Figure 1.

The Inter-Annotator Agreement(IAA) was analyzed for both Tamil and Malayalam datasets on Abusive text targeting Women in Social Media. The annotation process involved six annotators, including four Computer Science students and two Social Work students, comprising four females and two males. For Tamil annotation, two Computer Science Students and a Social Work student took part and likewise the same proportion of Malayalam speaking students were involved in the Malayalam dataset annotation process. Majority voting was used to determine the final labels for each instance. The Krippendorff's Alpha value for Tamil annotations was 0.6474, indicating moderate agreement, while Malayalam annotations achieved a Krippendorff's Alpha value of 0.9573, reflecting near-perfect agreement.

Table 1 shows the dataset statistics. In Tamil, the dataset consists of a total of 2,790 samples in training, with 1,424 non abusive samples and 1,366 abusive samples. The average number of words per sample is approximately 14.5 for Tamil.
The Malayalam training corpus consists of a total of 2,933 samples, with 1,531 abusive samples and 1,402 non-abusive samples. The average number of words per sample is approximately 16 for Malayalam. The test set in both languages was used for final evaluation and ranking, providing insights into how well the model generalizes to unseen data and its overall performance. Figures 2 and 3 present sample sentences from the Tamil and Malayalam corpora, respectively.

[2]https://codalab.lisn.upsaclay.fr/competitions/20701

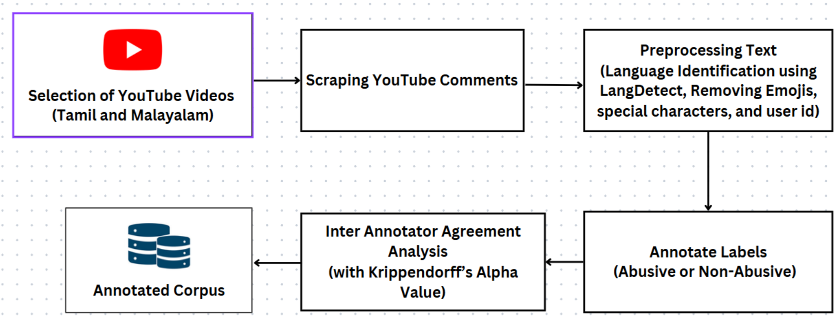

Figure 1: Dataset collection and Annotation Process

Language	Dataset	No of samples	Vocab size
Tamil	Train	2,790	15,863
	Dev/Val	598	4517
	Test	598	4841
Malayalam	Train	2,933	16,344
	Dev/Val	629	4893
	Test	629	4829

Table 1: Dataset Description for Tamil and Malayalam

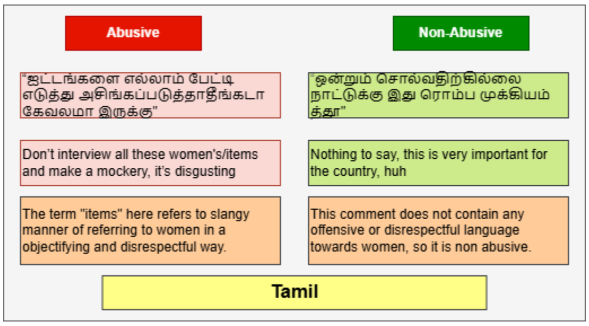

Figure 2: Samples from the corpus for Tamil language

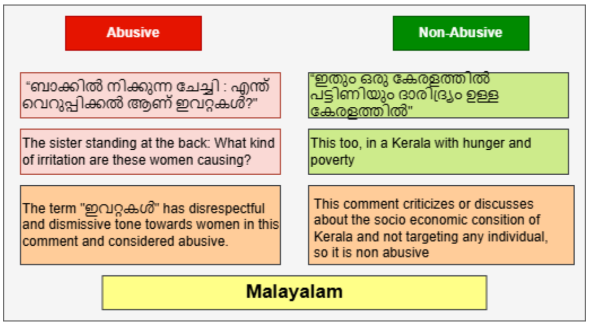

Figure 3: Samples from the corpus for Malayalam language

4 Participant's Methodology

A total of 157 teams registered to participate in this shared task. However, only 37 teams submitted their results for the Tamil, while 35 teams submitted for the Malayalam. This shows a wide variety of methodologies for detecting abusive content targeting women. Transformer-based models such as BERT, mBERT, MuRIL, and XLM-RoBERTa were widely used, with many teams fine-tuning these models to improve their performance on the given dataset. Few teams combined transformer embeddings with traditional machine learning classifiers. For instance, HTMS and KECEmPower used embeddings from models like BERT or TF-IDF and applied Random Forest, SVM, or Logistic Regression to make predictions. Below is a detailed description of each team methodology:

- **ANSR** (Nishanth et al., 2025): Utilized XLM-RoBERTa-XL to extract contextual embeddings from input text. These embeddings were fed into Random Forest (Run 1) and XGBoost (Run 2) classifiers to categorize text as "Abusive" or "Non-Abusive."
- **ARINDASCI**: Employed transformer models, BERT and mBERT, fine-tuning them on Tamil and Malayalam text with tokenization and class label encoding. They applied data augmentation and hyperparameter tuning to address class imbalance problem.
- **Byte-Sized LLM** (Kodali et al., 2025): Hybrid approach was developed, combining attention BiLSTM network with a fine-tuned XLM-RoBERTa base model.
- **Code Crafters**: The team used feature extraction techniques such as Word2Vec, GloVe, and BERT embeddings. Models like Random Forest, LSTM, and pre-trained transformers such as DistilBERT were employed.
- **Courfour IITK** (S et al., 2025): The dataset was normalized, preprocessed to remove incomplete entries, and cleaned to eliminate

punctuation, special characters, and redundant words. The team used NLTK for tokenization and TfidfVectorizer for feature extraction, followed by Random Forest, SVM, and Logistic Regression models.

- **CUET Agile** (Hanif and Rahman, 2025): Tamil BERT and Malayalam BERT were fine-tuned for their respective languages, while IndicBERTv2 was utilized for both. Models were also fine-tuned on unprocessed texts. Each model was trained for 5 epochs using AdamW with a learning rate of 5e-5, with the best validation F1 score determining the optimal epoch.

- **CUET Ignite** (Rahman et al., 2025b): Implemented multilingual BERT with mixed-precision training for faster convergence. The model was fine-tuned using the AdamW optimizer, cross-entropy loss and dynamic learning rate adjustment throughout 15 epochs.

- **CUET NLP FiniteInfinity**: Employed Sarvam-1, Tamil LLaMA 7B Base, and Gemma-2-2B models for fine tuning the model.

- **CUET Novice** (Sayma et al., 2025): Used l3cube-pune/malayalam-bert. To address class imbalance, the team calculated class weights and incorporated them into the loss function to ensure fair training across all classes.

- **CUET Raptors** (Naib et al., 2025): Fine tuning was performed using the PyTorch-based Hugging Face Transformers library, optimizing a single linear classification layer for binary classification. Training was conducted over 5 epochs with a batch size of 16, using the Adam optimizer with weight decay for regularization and binary cross entropy loss.

- **CVF@NITT**: Devleoped IndicBERT and LaBSE embeddings with Bi-GRU(Bidirectional Gated Recurrent Unit), incorporating external knowledge bases.

- **Cyber Protectors** (Rohit et al., 2025): FastText embeddings were generated to create vector representations of the text. Transformed based architecture is utilized for training the model.

- **Falcons**: Fine-tuned MuRIL(Multilingual Representations for Indian Languages) with Adam and binary cross-entropy loss.

- **GS**: Utilized a BERT model for fine tuning.

- **Habiba A, G Aghila** (Habiba A, 2025): Team Habiba A, G Aghila employed a Recurrent Neural Network (RNN) architecture.

- **HTMS** (Harini et al., 2025): The team conducted three runs for the task using machine and deep learning techniques. In Run 1, BERT embeddings were used with a Random Forest classifier, utilizing 5-fold cross-validation. Run 2 combined TF-IDF and BERT embeddings, with dimensionality reduction and fusion, followed by training a Random Forest classifier. In Run 3, TF-IDF embeddings were used with Logistic Regression, employing 5-fold cross-validation for robust evaluation.

- **Hydrangea** (Thirumoorthy et al., 2025): The team used BERT, XLM-RoBERTa, and DistilBERT for three runs on Tamil and Malayalam datasets. Each model was trained for two epochs.

- **Incepto** (Thavarasa et al., 2025): Combined XLM-RoBERTa-base model with a four multi-head attention heads. The extracted features are processed through a deep feed-forward network with layer normalization, and ReLU activation.

- **KEC Tech Titans** (Subramanian et al., 2025a): The team utilized GRU, FastText, and XGBoost models for Tamil, and LSTM, BiLSTM, and XGBoost models for Malayalam in detecting. XGBoost complemented the deep learning models by handling non-linear relationships

- **KECEmPower** (Subramanian et al., 2025b): Applied Logistic Regression, Random Forest, and SVM with TF-IDF embeddings.

- **LinguAIsts** (G et al., 2025): Experimented with SVM, Naïve Bayes, and Logistic Regression using TF-IDF features. Hyperparameter tuning was performed using GridSearchCV.

- **Lexi Logic** (M et al., 2025): The data was cleaned to remove noise, and duplicates. The dataset was balanced using techniques like oversampling underrepresented classes. The BERT model was then fine tuned on the data.

- **MSM-CUET** (Rahman et al., 2025a): Incorporated MuRIL transformer for malayalam and XLM-RoBERTa for Tamil. Hyperparameter tuning was applied to optimize the training process, and early stopping was introduced to prevent overfitting. Team JAS also fine-tuned the MuRIL model.

- **Necto** (Dhasan, 2025): Fine-tuned a multilingual SBERT model (microsoft/Multilingual-MiniLM-L12-H384) using Tamil and Malayalam data jointly.

- **NLP Goats** (Vaidyanathan et al., 2025): The BERTbase model was fine-tuned with hyperparameters, including a batch size of 16 and a learning rate of 2e-5.

- **NLPopsCIOL** (Nahian et al., 2025): Used custom models specifically trained on Malayalam and Tamil hate speech data to encode the text and extract embeddings. These embeddings were passed through a Multi-Layer Perceptron (MLP) for model training and. A search-based modeling approach was adopted, where the best-performing model was tracked throughout the training process based on the highest F1 score.

- **NomoreHate**: The team employed a fusion model for Malayalam, combining mBERT and Indic-BERT to generate larger and diverse embeddings. For Tamil, the team used a BiLSTM model combined with Indic-BERT, to capture sequential dependencies.

- **ParsePros**: Explored XLM-RoBERTa embeddings integrated with a BiLSTM-based autoencoder.

- **RMKmavericks** (Johnson et al., 2025): Adopted BiLSTM and SVM with TF-IDF for Tamil, and Decision Tree, Random Forest, Multinomial Naïve Bayes for Malayalam.

- **SSN IT NLP** (Maria Nancy et al., 2025): The mBERT model was fine-tuned using cross-entropy loss, with periodic evaluations during training to monitor performance. To address data imbalance, class weights were applied during training to ensure effective learning from both classes.

- **SSN Trio** (T T et al., 2025): Leveraged mBERT and MuRIL for multilingual classification.

- **SSN SQUAD**: This team fine-tuned the mBERT model separately for Malayalam and Tamil using the Hugging Face Trainer. The training process was optimized with hyperparameters, including a learning rate of 2e-5 and a batch size of 16, over five epochs. This team attained a macro F1 score of 0.751 for Tamil and 0.667 for Malayalam.

- **Syndicate IIITK**: Utilized TF-IDF embeddings with an optimized SVM classifier after tokenization and normalization.

- **Techbusters**: Addressed class imbalance with SMOTE(Synthetic Minority Over-sampling Technique). Multiple classifiers, including Random Forest, Naive Bayes, and Decision Tree, were evaluated. Random Forest and Naive Bayes outperformed others.

- **Tewodros**: Implemented Logistic Regression using TF-IDF vectorizer, and hyperparameter tuning using GridSearchCV.

- **Trix**: The preprocessing involved Unicode normalization using the Indic NLP Library, tokenization with Stanza and custom text cleaning to remove non-native characters, stopwords, and standardize spoken-to-written variations. To address class imbalances, upsampling techniques were applied using Scikit-learn's resample function. For feature extraction, CountVectorizer with n-grams was used. Logistic Regression, Multinomial Naive Bayes, and Decision Tree Classifier, were trained and optimized using GridSearchCV.

- **Yadu**: Applied MuRIL with focal loss, label smoothing, to handle class imbalance and a custom multilingual trainer. The model is trained with a batch size of 16, gradient accumulation, a cosine learning rate scheduler and early stopping.

- **YenCS**: ELMo (Embeddings from Language Models) embeddings are then used for feature extraction. The extracted features are fed into a deep learning based classifier.

Table 2: Rank List of Tamil Language

Team Name	F1 Score	RANK
CUET_Agile (Hanif and Rahman, 2025)	0.7883	1
MSM_CUET (Rahman et al., 2025a)	0.7873	2
Incepto (Thavarasa et al., 2025)	0.7864	3
Lexi Logic (M et al., 2025)	0.7824	4
Necto (Dhasan, 2025)	0.7821	5
byteSizedLLM (Kodali et al., 2025)	0.7820	6
CUETNLP FiniteInfinity	0.7767	7
techbusters	0.7721	8
Hydrangea (Thirumoorthy et al., 2025)	0.7708	9
JAS	0.7687	10
SSNTrio (T T et al., 2025)	0.7668	11
Code_Crafters	0.7587	12
NLP_goats (Vaidyanathan et al., 2025)	0.7504	13
KEC TECH TITANS (Subramanian et al., 2025a)	0.7447	14
Cyber_Protectors (Rohit et al., 2025)	0.7356	15
GS	0.7293	16
CUET_Ignite (Rahman et al., 2025b)	0.7224	17
Habiba A ,G Agila (Habiba A, 2025)	0.7207	18
cuetRaptors (Naib et al., 2025)	0.7203	19
ANSR (Nishanth et al., 2025)	0.7201	20
NLPOPSCIOL (Nahian et al., 2025)	0.7039	21
PARSPROSE	0.6998	22
KECEmpower (Subramanian et al., 2025b)	0.6903	23
CoreFour_IIITK (S et al., 2025)	0.6901	24
Syndicate_IIITK	0.6872	25
SSN_IT_NLP (Maria Nancy et al., 2025)	0.6519	26
nomorehate	0.6517	27
YenCS	0.6381	28
Falcons	0.6255	29
LinguAIsts (G et al., 2025)	0.6251	30
RMKMavericks (Johnson et al., 2025)	0.6196	31
CVF@NITT	0.6174	32
TRIX	0.6001	33
VSS	0.5881	34
Yadu	0.5099	35
HTMS (Harini et al., 2025)	0.5007	36
Tewodros	0.3378	37

Table 3: Rank List of Malayalam Language

Team Name	**F1 Score**	**RANK**
Habiba A ,G Agila (Habiba A, 2025)	0.7571	1
CUET_Agile (Hanif and Rahman, 2025)	0.7234	2
CUET_Novice (Sayma et al., 2025)	0.7083	3
Incepto (Thavarasa et al., 2025)	0.7058	4
Lexi Logic (M et al., 2025)	0.7001	5
byteSizedLLM (Kodali et al., 2025)	0.6964	6
Necto (Dhasan, 2025)	0.6915	7
ANSR (Nishanth et al., 2025)	0.6901	8
NLP_goats (Vaidyanathan et al., 2025)	0.6843	9
MSM_CUET (Rahman et al., 2025a)	0.6812	10
LinguAIsts (G et al., 2025)	0.6779	11
Hydrangea (Thirumoorthy et al., 2025)	0.6769	12
VSS	0.6757	13
CVF@NITT	0.6701	14
CUETNLP FiniteInfinity	0.6645	15
SSN_IT_NLP (Maria Nancy et al., 2025)	0.6601	16
Cyber_Protectors (Rohit et al., 2025)	0.6518	17
TRIX	0.6501	18
RMKMaveriks (Johnson et al., 2025)	0.6484	19
KECEmpower (Subramanian et al., 2025b)	0.6454	20
techbusters	0.6452	21
NLPopsCIOL (Nahian et al., 2025)	0.6402	22
nomorehate	0.6401	23
Syndicate_IIITK	0.6295	24
ParsePros	0.6201	25
KEC Tech Titans (Subramanian et al., 2025a)	0.6174	26
CoreFour_IIITK (S et al., 2025)	0.6101	27
YenCS	0.5701	28
HTMS (Harini et al., 2025)	0.4947	29
Yadu	0.4801	30
Falcons	0.4772	31
Tewodros	0.3396	32
SSNTrio (T T et al., 2025)	0.3094	33
ARINDASCI	0.2201	34
GS	0.2147	35

5 Results and Discussion

The evaluation metric used was the macro-averaged F1-score, calculated using the scikit-learn library[3].

The Tamil dataset results are shown in table 2. The top performing team, CUET_Agile, achieved a macro F1-score of 0.7883. They have used fine-tuned Tamil BERT model combined with effective optimization strategies. MSM_CUET (0.7873) and Incepto (0.7864) followed closely, utilized multilingual transformer models with fine-tuning techniques. Other notable performers included Lowes, Necto, and ByteSizedLLM, all achieving macro F1-scores above 0.78. These teams demonstrated the effectiveness of pretrained transformer-based models, including MuRIL, XLM-RoBERTa, and BiLSTM.

The Malayalam dataset results are shown in table 3. The top-performing team, Habiba A, G Agila, achieved a macro F1-score of 0.7571, using a Recurrent Neural Network (RNN) approach. Their methodology highlights the potential of traditional deep learning models when combined with effective preprocessing. The second and third positions were claimed by CUET_Agile (0.7234) and CUET_Novice (0.7083), who used fine tuned transformer-based models. Other teams, such as Incepto (0.7058) and Lowes (0.7001), performed well using multilingual pretrained models. The top-performing teams used transformer-based models like BERT, mBERT, MuRIL, and XLM-RoBERTa generally outperformed others. The use of the macro-averaged F1-score and a detailed classification report enabled a fair and comprehensive evaluation of this shared task.

6 Conclusion

This shared task has provided pivotal insights for addressing abuse content targeting women in Dravidian languages. It is evident from the results that the usage of transformer-based models like mBERT and XLM-RoBERTa have out performed the other traditional approaches. In the future, we plan to enhance this task by multiclass problem such as stereotype, bias detection, and gender neutral term analysis. This will enable more contextual analysis for understanding abusive content.

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

Acknowledgments

This work was conducted with the financial support from Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2(Insight_2), supported in part of Science Foundation Ireland Centre for Research Training in Artificial Intelligence under Grant No. 18/CRT/6223.

References

Premjith B, Bharathi Raja Chakravarthi, Prasanna Kumar Kumaresan, Saranya Rajiakodi, Sai Karnati, Sai Mangamuru, and Chandu Janakiram. 2024. Findings of the shared task on hate and offensive language detection in Telugu codemixed text (HOLD-Telugu)@DravidianLangTech 2024. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 49–55, St. Julian's, Malta. Association for Computational Linguistics.

Michele Battisti, Ilpo Kauppinen, and Britta Rude. 2024. Breaking the silence: The effects of online social movements on gender-based violence. *European Journal of Political Economy*, 85.

Tommaso Caselli, Valerio Basile, Jelena Mitrović, Inga Kartoziya, and Michael Granitzer. 2020. I feel offended, don't be abusive! implicit/explicit messages in offensive and abusive language. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6193–6202, Marseille, France. European Language Resources Association.

Bharathi Raja Chakravarthi, Prasanna Kumaresan, Ruba Priyadharshini, Paul Buitelaar, Asha Hegde, Hosahalli Shashirekha, Saranya Rajiakodi, Miguel Ángel García, Salud María Jiménez-Zafra, José García-Díaz, Rafael Valencia-García, Kishore Ponnusamy, Poorvi Shetty, and Daniel García-Baena. 2024. Overview of third shared task on homophobia and transphobia detection in social media comments. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion*, pages 124–132, St. Julian's, Malta. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Vigneshwaran Muralidaran, Ruba Priyadharshini, Subalalitha Cn, John McCrae, Miguel Ángel García, Salud María Jiménez-Zafra, Rafael Valencia-García, Prasanna Kumaresan, Rahul Ponnusamy, Daniel García-Baena, and José García-Díaz. 2022. Overview of the shared task on hope speech detection for equality, diversity, and inclusion. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 378–388, Dublin, Ireland. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Livin Nector Dhasan. 2025. Necto@DravidianLangTech: Fine-tuning Multilingual MiniLM for Text Classification in Dravidian Languages. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Dhanyashree G, Kalpana K, Lekhashree A, Arivuchudar K, ARTHI R, Bommineni Sahitya, Pavithra J, and SANDRA JOHNSON. 2025. LinguAIsts@DravidianLangTech 2025: Abusive Tamil and Malayalam Text targeting Women on Social Media. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

DR G AGHILA Habiba A. 2025. DLTC-NITPY@DravidianLangTech 2025 Abusive Code-mixed Text Detection System Targeting Women for Tamil and Malayalam Languages using Deep Learning Technique. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Tareque Md Hanif and Md Rashadur Rahman. 2025. CUET_Agile@DravidianLangTech 2025: Fine-tuning Transformers for Detecting Abusive Text Targeting Women from Tamil and Malayalam Texts. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Bachu Naga Sri Harini, Kankipati Venkata Meghana, Kondakindi Supriya, Tara Samiksha, and Premjith B. 2025. HTMS@DravidianLangTech 2025: Fusing TF-IDF and BERT with Dimensionality Reduction for Abusive Language Detection in Tamil and Malayalam. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Sandra Johnson, Boomika E, and Lahari P. 2025. RMKMavericks@DravidianLangTech 2025: Tackling Abusive Tamil and Malayalam Text Targeting Women: A Linguistic Approach. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020. IndicNLPSuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for Indian languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4948–4961, Online. Association for Computational Linguistics.

Rohith Gowtham Kodali, Durga Prasad Manukonda, and Maharajan Pannakkaran. 2025. byteSizedLLM@DravidianLangTech 2025: Abusive Tamil and Malayalam Text targeting Women on Social Media Using XLM-RoBERTa and Attention-BiLSTM. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Niranjan Kumar M, Pranav Gupta, Billodal Roy, and Souvik Bhattacharyya. 2025. LexiLogic@Dravidian Lang Tech 2025: Detecting Misogynistic Memes and Abusive Tamil and Malayalam Text Targeting Women on Social Media. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Thomas Mandl, Sandip Modha, Anand Kumar M, and Bharathi Raja Chakravarthi. 2021. Overview of the HASOC Track at FIRE 2020: Hate Speech and Offensive Language Identification in Tamil, Malayalam, Hindi, English and German. In *Proceedings of the 12th Annual Meeting of the Forum for Information Retrieval Evaluation*, FIRE '20, page 29–32, New York, NY, USA. Association for Computing Machinery.

C Maria Nancy, Radha N, and Swathika R. 2025. SSN_IT_NLP@DravidianLangTech 2025: Abusive Tamil and Malayalam Text targeting Women on Social Media. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Alessio Miaschi and Felice Dell'Orletta. 2020. Contextual and non-contextual word embeddings: an in-depth linguistic investigation. In *Proceedings of the 5th Workshop on Representation Learning for NLP*, pages 110–119, Online. Association for Computational Linguistics.

Jayanth Mohan, Spandana Reddy Mekapati, Premjith B, Jyothish Lal G, and Bharathi Raja Chakravarthi. 2025. A multimodal approach for hate and offensive content detection in Tamil: From corpus creation to model development. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* Just Accepted.

Abdullah Al Nahian, Mst Rafia Islam, Azmine Toushik Wasi, and Md Manjurul Ahsan. 2025. NLPopsCIOL@DravidianLangTech 2025: Classification of Abusive Tamil and Malayalam Text Targeting Women Using Pre-trained Models. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Md. Mubasshir Naib, Md. Saikat Hossain Shohag, Alamgir Hossain, Jawad Hossain, and Mohammed Moshiul Hoque. 2025. cuetRaptors@DravidianLangTech 2025: Transformer-Based Approaches for Detecting Abusive Tamil Text Targeting Women on Social Media. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

S. Nishanth, Shruthi Rengarajan, S. Ananthasivan, Burugu Rahul, and S. Sachin Kumar. 2025. ANSR@DravidianLangTech 2025: Detection of abusive tamil and malayalam text targeting women on social media using RoBERTa and XGBoost. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Ayush Pandey. 2024. Gender and the legal discourse: Exploring indian laws and cases. *International Journal of Human Rights Law Review*, 3(4):1–41.

Ji Ho Park and Pascale Fung. 2017. One-step and two-step classification for abusive language detection on Twitter. In *Proceedings of the First Workshop on Abusive Language Online*, pages 41–45, Vancouver, BC, Canada. Association for Computational Linguistics.

Rahul Ponnusamy, Kathiravan Pannerselvam, Saranya R, Prasanna Kumar Kumaresan, Sajeetha Thavareesan, Bhuvaneswari S, Anshid K.a, Susminu S Kumar, Paul Buitelaar, and Bharathi Raja Chakravarthi. 2024. From laughter to inequality: Annotated dataset for misogyny detection in Tamil and Malayalam memes. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 7480–7488, Torino, Italia. ELRA and ICCL.

Md Mizanur Rahman, Srijita Dhar, Md Mehedi Hasan, and Hasan Murad. 2025a. MSM_CUET@DravidianLangTech 2025: XLM-BERT and MuRIL Based Transformer Models for Detection of Abusive Tamil and Malayalam Text Targeting Women on Social Media. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

MD.Mahadi Rahman, Mohammad Minhaj Uddin, and Mohammad Shamsul Arefin. 2025b. CUET_Ignite@DravidianLangTech-NAACL2025: Detection of Abusive Comments in Tamil Text Using Transformer Models. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

VP Rohit, Madhav M, Ippatapu Venkata Srichandra, Neethu Mohan, and Sachin Kumar S. 2025. Cyber_Protectors@DravidianLangTech 2025: Abusive Tamil and Malayalam Text Targeting Women on Social Media using FastText. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Varun Balaji S, Bojja Revanth Reddy, Vyshnavi Reddy Battula, Suraj Nagunuri, and Balasubramanian Palani. 2025. CoreFour_IIITK@DravidianLangTech 2025: Abusive Content Detection Against Women Using Machine Learning And Deep Learning Models. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Khadiza Sultana Sayma, Farjana Alam Tofa, Md Osama, and Ashim Dey. 2025. CUET_Novice@DravidianLangTech-NAACL2025: Abusive Comment Detection in Malayalam Text Targeting Women on Social Media Using Transformer-Based Models. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

MC Shunmuga Priya, D Karthika Renuka, L Ashok Kumar, and Lovelyn Rose S. 2022. Multilingual Low Resource Indian Language Speech Recognition and Spell Correction using Indic BERT. *Sādhanā*, 47.

W. Soral, M. Bilewicz, and M. Winiewski. 2023. Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavious*, 44:136–146.

Malliga Subramanian, Rahul Ponnusamy, Sean Benhur, Kogilavani Shanmugavadivel, Adhithiya Ganesan, Deepti Ravi, Gowtham Krishnan Shanmugasundaram, Ruba Priyadharshini, and Bharathi Raja Chakravarthi. 2022. Offensive language detection in tamil youtube comments by adapters and cross-domain knowledge transfer. *Computer Speech & Language*, 76:101404.

Malliga Subramanian, Kogilavani Shanmugavadivel, Deepiga P, Dharshini S, Ananthakumar S, and Praveenkumar C. 2025a. KEC_TECH_TITANS@DravidianLangTech 2025: Abusive Text Detection in Tamil and Malayalam Social Media Comments Using Machine Learning. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Malliga Subramanian, Kogilavani Shanmugavadivel, Indhuja V S, Kowshik P, and Jayasurya S. 2025b. KECEmpower@DravidianLangTech 2025: Abusive Tamil and Malayalam Text targeting Women on Social Media. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Mirnalinee T T, J Bhuvana, Avaneesh Koushik, Diya Seshan, and Rohan R. 2025. SSNTrio@DravidianLangTech2025: LLM Based Techniques for Detection of Abusive Text Targeting

Women. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Luxshan Thavarasa, Sivasuthan Sukumar, and Jubeerathan Thevakumar. 2025. Incepto@DravidianLangTech-2025: Detecting Abusive Tamil and Malayalam Text Targeting Women on YouTube. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Shanmitha Thirumoorthy, Thenmozhi Durairaj, and Ratnavel Rajalakshmi. 2025. Hydrangea@DravidianLanTech2025: Abusive language Identification from Tamil and Malayalam Text using Transformer Models. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, et al. 2023. Llama 2: Open foundation and fine-tuned chat models.

Vijay Karthick Vaidyanathan, Srihari V K, and Thenmozhi Durairaj. 2025. NLP_goats@DravidianLangTech 2025: Towards Safer Social Media: Detecting Abusive Language Directed at Women in Dravidian Languages. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Zeerak Waseem and Dirk Hovy. 2016. Hateful symbols or hateful people? predictive features for hate speech detection on Twitter. In *Proceedings of the NAACL Student Research Workshop*, pages 88–93, San Diego, California. Association for Computational Linguistics.