# cuetRaptors@DravidianLangTech 2025: Transformer-Based Approaches for Detecting Abusive Tamil Text Targeting Women on Social Media

**Md. Mubasshir Naib[a], Md. Saikat Hossain Shohag[b], Alamgir Hossain[c]**
**Jawad Hossain[d] and Mohammed Moshiul Hoque[e]**
Department of Computer Science and Engineering
Chittagong University of Engineering and Technology
{u1904089[a], u1904088[b], u1704039[d]}@student.cuet.ac.bd
alamgir.hossain.cs@gmail.com[c], moshiul_240@cuet.ac.bd[e]

## Abstract

With the exponential growth of social media usage, the prevalence of abusive language targeting women has become a pressing issue, particularly in low-resource languages (LRLs) like Tamil and Malayalam. This study is part of the shared task at DravidianLangTech@NAACL 2025, which focuses on detecting abusive comments in Tamil social media content. The provided dataset consists of binary-labeled comments (Abusive or Non-Abusive), gathered from YouTube, reflecting explicit abuse, implicit bias, stereotypes, and coded language. We developed and evaluated multiple models for this task, including traditional machine learning algorithms (Logistic Regression, Support Vector Machine, Random Forest Classifier, and Multinomial Naive Bayes), deep learning models (CNN, BiLSTM, and CNN+BiLSTM), and transformer-based architectures (DistilBERT, Multilingual BERT, XLM-RoBERTa), and fine-tuned variants of these models. Our best-performing model, Multilingual BERT, achieved a weighted F1-score of 0.7203, ranking 19$^{th}$ in the competition.

## 1 Introduction

The rapid expansion of social media has transformed communication, but it has also amplified the spread of abusive content, particularly targeting women and other marginalized groups (Priyadharshini et al., 2022b; Ghanghor et al., 2021b). In low-resource languages like Tamil, this issue is exacerbated by the lack of linguistic tools and datasets, making automated detection of abusive text a critical yet underexplored challenge (Chakravarthi et al., 2020; Priyadharshini et al., 2020). Tamil, a Dravidian language spoken by over 80 million people in South Asia, faces unique complexities due to its rich morphology, code-mixing tendencies, and the prevalence of implicit bias, stereotypes, and coded language in online discourse (Anita and Subalalitha, 2019; Subalalitha and Poovammal, 2018). While prior work has addressed abusive language detection in Tamil, most studies focus on broad categories (e.g., hate speech (Hossan et al., 2025), misogyny) or coarse-grained binary classification (Sharif et al., 2021b; Chakravarthi et al., 2022), with limited emphasis on nuanced abuse targeting women specifically.

Social media platforms like YouTube, Facebook, and Twitter have struggled to manually filter such content due to its sheer volume and linguistic diversity (Ghanghor et al., 2021a). Existing solutions for high-resource languages like English rely heavily on transformer-based models (Kumar et al., 2020; Sampath et al., 2022), but their efficacy in Tamil remains understudied. Recent initiatives like the DravidianLangTech shared tasks have spurred progress in abusive text detection (Chakravarthi et al., 2021; B et al., 2022), yet gaps persist in addressing gender-targeted abuse with computational efficiency and cultural sensitivity.

This work, part of the DravidianLangTech@NAACL 2025 shared task, focuses on detecting abusive Tamil social media comments directed at women. Besides using various ML and Dl models, We leverage transformer-based architectures—DistilBERT, Multilingual BERT (mBERT), and XLM-RoBERTa—to tackle binary classification on a dataset of YouTube comments labeled as *Abusive* or *Non-Abusive*. Our contributions include:

- A comparative analysis of traditional machine learning, deep learning, and lightweight transformer models for Tamil abuse detection.

- An evaluation of multilingual, language-specific pre-trained models and deep learning architectures (CNN, BiLSTM, CNN+BiLSTM) in capturing contextual and cultural nuances.



*Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 756–762
May 3, 2025 ©2025 Association for Computational Linguistics

## 2 Related Task

The detection of abusive language in low-resource languages has gained traction in recent years, driven by the proliferation of harmful content on social media platforms.

Using classifiers like Logistic Regression, Support Vector Machines (SVM), and ensemble approaches, early attempts at abusive language detection concentrated on high-resource languages like English (Oswal, 2021). Traditional machine learning approaches have been the main method used in studies for low-resource languages like Bengali and Tamil. For example, (Eshan and Hasan, 2017) used SVM with tri-gram features to classify abusive Tamil texts with 95% accuracy. A weighted ensemble of BERT variants was also proposed by (Sharif and Hoque, 2021), who created a dataset of hostile Bengali text and achieved 93% weighted F1-scores. However, these studies rarely examine gender-specific abuse, instead concentrating on broad categories like hate speech and aggressiveness (Sharif et al., 2021a; Aurpa et al., 2021) or coarse-grained binary classification (e.g., abusive/non-abusive).

NLP jobs have been transformed by recent developments in transformer-based models, especially for high-resource languages. (Kumar et al., 2020), for instance, showed how effective BERT is in identifying implicit hate speech in English. However, morphological complexity, code-mixing, and cultural context make it difficult to adapt these models to low-resource languages like Tamil (Anita and Subalalitha, 2019). Although multilingual transformers (such as mBERT and XLM-RoBERTa) have demonstrated promise in cross-lingual tasks (Chakravarthi et al., 2021), nothing is known about how well they function in fine-grained abusive language detection, particularly when focused on women. Previous research in Devanagari script languages, including (Jha et al., 2020), used FastText to detect hate speech in Hindi with 92% accuracy, and (Chopra et al., 2023) used transformers to detect hate speech that was code-mixed between Hindi and English. These studies highlight the potential of hybrid and transformer-based approaches but underscore the need for language-specific adaptations.

Existing research on Tamil abusive language detection lacks focus on gender-targeted abuse and relies heavily on traditional ML methods (Priyadharshini et al., 2020; Chakravarthi et al., 2022). While (Sharif and Hoque, 2022) advanced Bengali aggression detection using BERT variants, similar efforts for Tamil are scarce. Our work bridges these gaps by:

Investigating transformer models (DistilBERT, mBERT, XLM-R) for detecting abusive Tamil text *targeting women*, a fine-grained and culturally sensitive task. Benchmarking against traditional ML baselines (Logistic Regression, Random Forest Classifier, and Multinomial Naive Bayes) and deep learning architectures (CNN, BiLSTM, CNN+BiLSTM) to quantify the benefits of lightweight transformers in low-resource settings. Addressing implicit bias and coded language through contextual embeddings, a challenge highlighted in prior Devanagari script research (Parihar et al., 2021; Nandi et al., 2024).

## 3 Task and Dataset Description

This shared task was organized to detect abusive Tamil and Malayalam texts targeting women on social media (Rajiakodi et al., 2025). The task focused on binary classification, categorizing texts as *Abusive* or *Non-Abusive*. We utilized the corpus provided by the organizers of DravidianLangTech@NAACL 2025 (Priyadharshini et al., 2023, 2022a), which comprises Tamil social media comments annotated for gender-specific abusive content. The dataset includes comments collected from YouTube, reflecting explicit abuse, implicit bias, stereotypes, and coded language targeting women. Table 1 summarizes the distribution of the dataset across training, validation, and test splits. While the dataset exhibits slight class imbalance, this reflects real-world social media data where abusive content often appears less frequently than non-abusive interactions.

| **Class** | **Train** | **Validation** | **Test** | **$W_T$** | **$UW_T$** |
|---|---|---|---|---|---|
| Abusive | 1236 | 129 | 305 | 25,585 | 13,181 |
| Non-Abusive | 1274 | 150 | 293 | 23,475 | 12,105 |
| **Total** | 2510 | 279 | 598 | 49,060 | 18,394 |

Table 1: Class distribution across training, validation, and test splits, where $W_T$ represents total words and $UW_T$ represents total unique words.

## 4 Dataset Visualization

Figure 1 represents the most common words in abusive texts in the training set, potentially indicating offensive or harmful language patterns. In contrast, Figure 2 highlights the frequent words in non-abusive texts, reflecting more neutral and factual vocabulary. This analysis generated word clouds to visualize the linguistic characteristics of both categories, using a maximum of 200 words for each cloud, with word sizes proportional to their frequency.


Figure 1: Word Cloud distribution of *Abusive* class

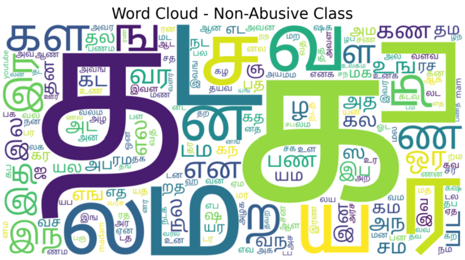


Figure 2: Word Cloud distribution of *Non-Abusive* class

## 5 Methodology

Various machine learning, deep learning, and transformer-based models were explored to establish baselines for detecting abusive language in Tamil comments in Figure 3. The implementation details of the models have been open-sourced to ensure reproducibility[1] .

### 5.1 Data Preprocessing

The dataset was preprocessed to ensure quality and consistency by removing rows with missing or invalid values in the "Class" column, mapping binary labels ( "Abusive" to 1, "Non-Abusive" to 0), and verifying the absence of NaN values. It was then split into training and validation sets using a stratified split to maintain label distribution, resulting in a clean and balanced dataset ready for modeling.

### 5.2 Feature Extraction

To represent text data numerically, TF-IDF, Bag of Words (BoW), and FastText embeddings were used.

[1] https://github.com/MubasshirNaib/Detecting-Abusive-Tamil-Text

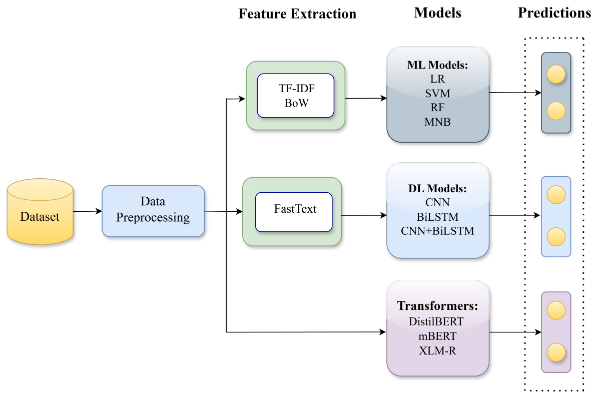


Figure 3: Schematic process for detecting abusive comments in Tamil social media content.

TF-IDF and BoW extracted the top 5,000 features to capture word importance and occurrences. FastText embeddings, trained on the tokenized training data with 100-dimensional vectors, provided context-aware representations, enhancing the ability of models to capture linguistic nuances. These methods ensured diverse and effective feature representations for model evaluation.

### 5.3 Machine Learning Models

For this task, we evaluated multiple machine learning models to detect abusive comments, including Logistic Regression, Support Vector Machine (SVM), Random Forest Classifier, and Multinomial Naive Bayes (MNB). By training and evaluating these models, we compared their capabilities comprehensively, enabling a deeper understanding of their strengths and suitability for the classification task.

### 5.4 Deep Learning Models

For this task, we implemented three deep learning models—CNN, BiLSTM, and a hybrid CNN+BiLSTM—for this classification. The CNN model includes a 1D convolution layer with 128 filters and ReLU activation, followed by max-pooling and fully connected layers. The BiLSTM model uses a bidirectional LSTM layer with 128 units to capture long-term dependencies from both directions. The combined CNN+BiLSTM model integrates two convolutional layers followed by a max-pooling and BiLSTM layer, leveraging both local feature extraction and sequential context learning. A dropout rate of 0.5 is applied to reduce overfitting, and models are trained using the Adam optimizer with binary cross-entropy loss for five epochs, ensuring robust prediction performance.

### 5.5 Transformer-Based Models

Transformer-Based models are particularly well-suited for multilingual and cross-lingual tasks, making them ideal for addressing abusive language detection in low-resource languages like Tamil. To tackle the shared task, we experimented with various transformer-based architectures, including DistilBERT (Sanh et al., 2020), Multilingual BERT (m-BERT) (Pires et al., 2019), and XLM-RoBERTa (XLM-R) (Conneau et al., 2020). Each model was fine-tuned on the binary classification task of identifying abusive and non-abusive comments in Tamil social media data. Here, the multilingual BERT was fine-tuned using the following hyperparameters shown in the Table 2. These hyperparameter choices ensured a balance between convergence and regularization, enabling the model to achieve a weighted F1-score of 0.7203. This demonstrates Multilingual BERT's ability to effectively capture nuanced patterns of abusive language while maintaining computational efficiency.

| Parameter | Value |
|---|---|
| Batch Size | 16 |
| Epochs | 7 |
| Weight Decay | 0.003 |
| Learning Rate | 5e-5 |

Table 2: Hyperparameters used in the best model

## 6 Results and Analysis

The performance of the various methods is presented in Table 3 The macro F1-score is used to evaluate and compare the overall performance of the models. Among the traditional machine learning models, Logistic Regression (LR) achieved the highest performance with an F1-score of 0.6933, an accuracy of 0.6935, and a G1-Score of 0.6833, outperforming both SVM and RF. The SVM model, while showing competitive results, lagged behind LR with an F1-score of 0.6746, an accuracy of 0.6756, and a G1-Score of 0.6764. Random Forest (RF) showed consistent performance but did not surpass LR or SVM, achieving an F1-score of 0.6738, an accuracy of 0.6738, and a G1-Score of 0.6739.

Deep learning models such as CNN and CNN+BiLSTM showed moderate performance, with an F1-score of 0.5679 for CNN and 0.5680 for CNN+BiLSTM, both having a G1-Score of 0.5681. BiLSTM, on the other hand, had a significantly lower performance with an F1-score of 0.3294, an accuracy of 0.4964, and a G1-Score of 0.3497.

Among the transformer models, m-BERT achieved the highest F1-score of 0.7203, an accuracy of 0.6404, and a G1-Score of 0.7233, followed by DistilBERT with an F1-score of 0.7068, an accuracy of 0.6164, and a G1-Score of 0.7183. XLM-R demonstrated a strong recall of 1.0000 but delivered lower overall performance with an F1-score of 0.6521, an accuracy of 0.4838, and a G1-Score of 0.6656.

| Classifier | P | R | F1 | A | G1 |
|---|---|---|---|---|---|
| LR | 0.68 | 0.68 | 0.68 | 0.68 | 0.68 |
| SVM | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| RF | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| MNB | 0.69 | 0.69 | 0.69 | 0.69 | 0.69 |
| CNN | 0.56 | 0.56 | 0.56 | 0.56 | 0.56 |
| BiLSTM | 0.24 | 0.49 | 0.32 | 0.49 | 0.35 |
| CNN+BiLSTM | 0.56 | 0.56 | 0.56 | 0.56 | 0.56 |
| **m-BERT** | **0.59** | **0.96** | **0.72** | **0.64** | **0.72** |
| DistilBERT | 0.56 | 0.93 | 0.70 | 0.61 | 0.71 |
| XLM-R | 0.48 | 1.00 | 0.65 | 0.48 | 0.66 |

Table 3: Performance of various models, where P, R, F1, A and G1 denote precision, recall, macro F1-score, accuracy and G1-Score respectively.

Overall, transformer-based models, particularly Multilingual BERT (m-BERT), excelled due to its pretraining on a multilingual corpus, including Tamil, enabling it to grasp contextual nuances of abusive language. Its self-attention mechanism outperforms traditional models (e.g., Logistic Regression, SVM), which miss subtleties, and deep learning models (e.g., CNN, BiLSTM), which struggle with limited data or long-range dependencies. While m-BERT uses generalized embeddings rather than Tamil-specific ones, its Tamil exposure was enough for strong performance (F1: 0.7203). Tamil-specific embeddings might enhance results but this is not explored in this work.

### 6.1 Error Analysis

We conducted both quantitative and qualitative error analyses to gain comprehensive insights into the performance of the proposed model.

#### 6.1.1 Quantitative Analysis:

The classifier demonstrated notable performance in identifying abusive and non-abusive content. However, a closer examination of the confusion matrix, Figure 4 reveals key areas of error, providing insights into the model's behavior across the different classes.

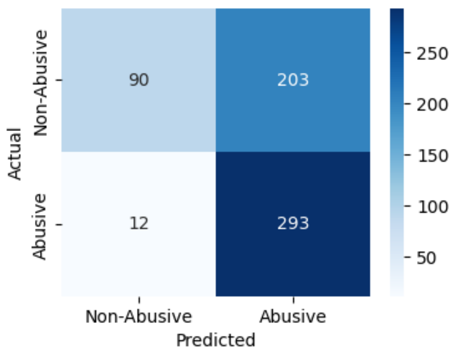


Figure 4: Confusion matrix of m-Bert

The classifier performed well in identifying abusive content, achieving a high True Positive Rate (TPR) of 96.05% for the abusive class, with minimal misclassification. However, the non-abusive class had a significantly lower TPR of 30.72%, with a large number of non-abusive instances being incorrectly classified as abusive. This suggests an overprediction of the abusive class, potentially caused by class imbalance, ambiguous features, or limited representation of non-abusive examples in the training data. To improve performance, the issues of class imbalance and feature ambiguity need to be addressed by refining the dataset, enhancing feature representation, and employing better modeling techniques to improve the classification of non-abusive content while maintaining high recall for the abusive class.

#### 6.1.2 Qualitative Analysis:

Figure 5 illustrates a qualitative analysis of the m-BERT model's predictions for the abusive language detection task. The model correctly classified samples 1 and 5 as *Abusive* and samples 3 and 4 as *Non-Abusive*, demonstrating its effectiveness in distinguishing between different language tones. However, sample 2 was misclassified as *Abusive* instead of *Non-Abusive*, likely due to contextual ambiguity or overlapping linguistic patterns in the dataset. This misclassification highlights a potential area for improvement in capturing subtle differences in expression.

## 7 Conclusion

This study explored a range of machine learning, deep learning, and transformer-based models for detecting abusive language in Tamil social media content. Among these, m-BERT emerged as the best-performing model, achieving a macro F1-score of 0.7203, showcasing its effectiveness in capturing nuanced patterns in text. Transformer-based models demonstrated clear advantages over traditional and deep learning approaches, highlighting their ability to manage complex tasks like abusive language detection. This study underscores the importance of leveraging advanced models and fine-tuning strategies to improve the detection of abusive content in low-resource, code-mixed languages.

| Sample Text | Actual Label | Predicted Label |
|---|---|---|
| **Sample Text 1:** இவ ஒரு மானெங்கெட்ட பொறுக்கி. ஒரே ஒரு routine ஒர்க் அவளுக்கு இருக்குறது தண்ணிய போட்டுட்டு அசிங்கமா பேசுறது. | **Abusive** | **Abusive** |
| **Sample Text 2:** இப்டியே பேசிக்கிட்டே இருந்தா எப்டி...... யாரு பெருசுனு அடிசிக்காட்டு ... | **Non- Abusive** | **Abusive** |
| **Sample Text 3:** அடக் கடவுளே இது என்னக் கொடுமையை ஊருல உலகத்துல எவ்வளவு பிரச்சினை இருக்கு இது என்னக்கொடுமை அடேய் கார்த்திக் நீ எங்கு இருந்தாலும் வந்துவிடு உன் காளில் விழுகின்றேன் | **Non- Abusive** | **Non- Abusive** |
| **Sample Text 4:** இதற்கு ஒரு தீர்வு இருக்கு. அவன் அவன் வேலை அவன் அவன் பார்த்தால் எந்த பாதிப்பும் ஏற்படாது. | **Non- Abusive** | **Non- Abusive** |
| **Sample Text 5:** தம்பி போய் நல்லவங்களை பேட்டிஎடு அவ சொல்வது அத்தனையும் பொய் தெரியாதா உனக்கு | **Abusive** | **Abusive** |

Figure 5: Some outputs predicted by the best model(m-Bert).

## Limitations

Despite the success of m-BERT, the system exhibited an overprediction tendency for the abusive class, struggling to accurately classify non-abusive content. This imbalance reflects challenges related to skewed class distribution, feature ambiguity, and limited representation of non-abusive data in the training set. Additionally, the reliance on pre-trained transformer models restricted opportunities for domain-specific optimization. Addressing these limitations will require balancing datasets, employing data augmentation strategies, and exploring innovative model architectures tailored to the complexities of low-resource, code-mixed languages like Tamil.

## Acknowledgments

We thank the DravidianLangTech 2025 shared task organizers for running this task. This work was supported by the Directorate of Research & Extension (DRE), Chittagong University of Engineering & Technology (CUET).

## References


R. Anita and C.N. Subalalitha. 2019. An approach to cluster tamil literatures using discourse connectives. In *2019 IEEE 1st International Conference on Energy, Systems and Information Processing (ICESIP)*, pages 1–4.

Tanjim Taharat Aurpa, Rifat Sadik, and Md Shoaib Ahmed. 2021. Abusive bangla comments detection on facebook using transformer-based deep learning models. *Social Network Analysis and Mining*, 12(1):24.

Bharathi B, Bharathi Raja Chakravarthi, Subalalitha Cn, Sripriya N, 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*, pages 339–345, Dublin, Ireland. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Vigneshwaran Muralidaran, Ruba Priyadharshini, and John Philip McCrae. 2020. Corpus creation for sentiment analysis in code-mixed Tamil-English text. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pages 202–210, Marseille, France. European Language Resources association.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Thenmozhi Durairaj, John McCrae, Paul Buitelaar, Prasanna Kumaresan, and Rahul Ponnusamy. 2022. Overview of the shared task on homophobia and transphobia detection in social media comments. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 369–377, Dublin, Ireland. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Kayalvizhi Sampath, Durairaj Thenmozhi, Sathiyaraj Thangasamy, Rajendran Nallathambi, and John Philip McCrae. 2021. Dataset for identification of homophobia and transophobia in multilingual youtube comments. *CoRR*, abs/2109.00227.

Abhishek Chopra, Deepak Kumar Sharma, Aashna Jha, and Uttam Ghosh. 2023. A framework for online hate speech detection on code-mixed hindi-english text and hindi text in devanagari. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 22(5).

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Shahnoor C. Eshan and Mohammad S. Hasan. 2017. An application of machine learning to detect abusive bengali text. In *2017 20th International Conference of Computer and Information Technology (ICCIT)*, pages 1–6.

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.

Md. Refaj Hossan, Nazmus Sakib, Md. Alam Miah, Jawad Hossain, and Mohammed Moshiul Hoque. 2025. CUET_Big_O@NLU of Devanagari script languages 2025: Identifying script language and detecting hate speech using deep learning and transformer model. In *Proceedings of the First Workshop on Challenges in Processing South Asian Languages (CHiPSAL 2025)*, pages 253–259, Abu Dhabi, UAE. International Committee on Computational Linguistics.

Vikas Kumar Jha, Hrudya P, Vinu P N, Vishnu Vijayan, and Prabaharan P. 2020. Dhot-repository and classification of offensive tweets in the hindi language. *Procedia Computer Science*, 171:2324–2333. Third International Conference on Computing and Network Communications (CoCoNet'19).

Ritesh Kumar, Atul Kr. Ojha, Bornini Lahiri, Marcos Zampieri, Shervin Malmasi, Vanessa Murdock, and Daniel Kadar, editors. 2020. *Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying*. European Language Resources Association (ELRA), Marseille, France.

Arpan Nandi, Kamal Sarkar, Arjun Mallick, and Arkadeep De. 2024. A survey of hate speech detection in indian languages. *Social Network Analysis and Mining*, 14(1):70.

Nikhil Oswal. 2021. Identifying and categorizing offensive language in social media. *CoRR*, abs/2104.04871.

Anil Singh Parihar, Surendrabikram Thapa, and Sushruti Mishra. 2021. Hate speech detection using natural language processing: Applications and challenges. In *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 1302–1308.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual BERT? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4996–5001, Florence, Italy. Association for Computational Linguistics.

Ruba Priyadharshini, Bharathi Raja Chakravarthi, Subalalitha Chinnaudayar Navaneethakrishnan, Malliga Subramanian, Kogilavani Shanmugavadivel, Premjith B, Abirami Murugappan, Sai Prashanth Karnati, Rishith, Chandu Janakiram, and Prasanna Kumar Kumaresan. 2023. Findings of the shared task on Abusive Comment Detection in Tamil and Telugu. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages DravidianLangTech 2023*. Recent Advances in Natural Language Processing.

Ruba Priyadharshini, Bharathi Raja Chakravarthi, Subalalitha Cn, Thenmozhi Durairaj, Malliga Subramanian, Kogilavani Shanmugavadivel, Siddhanth U Hegde, and Prasanna Kumaresan. 2022a. Overview of abusive comment detection in Tamil-ACL 2022. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*, pages 292–298, Dublin, Ireland. Association for Computational Linguistics.

Ruba Priyadharshini, Bharathi raja Chakravarthi, Sajeetha Thavareesan, Dhivya Chinnappa, Durairaj Thenmozhi, and Rahul Ponnusamy. 2022b. Overview of the dravidiancodemix 2021 shared task on sentiment detection in tamil, malayalam, and kannada. In *Proceedings of the 13th Annual Meeting of the Forum for Information Retrieval Evaluation*, FIRE '21, page 4–6, New York, NY, USA. Association for Computing Machinery.

Ruba Priyadharshini, Bharathi Raja Chakravarthi, Mani Vegupatti, and John P. McCrae. 2020. Named entity recognition for code-mixed indian corpus using meta embedding. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pages 68–72.

Saranya Rajiakodi, Bharathi Raja Chakravarthi, Shunmuga Priya Muthusamy Chinnan, Ruba Priyadharshini, Rajameenakshi J, Kathiravan Pannerselvam, Rahul Ponnusamy, Bhuvaneswari Sivagnanam, Paul Buitelaar, Bhavanimeena K, Jananayagam V, and Kishore Kumar Ponnusamy. 2025. Findings of the Shared Task on Abusive Tamil and Malayalam Text Targeting Women on Social Media: DravidianLangTech@NAACL 2025. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Anbukkarasi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, Ruba Priyadharshini, Subalalitha Cn, Kogilavani Shanmugavadivel, Sajeetha Thavareesan, Sathiyaraj Thangasamy, Parameswari Krishnamurthy, Adeep Hande, Sean Benhur, Kishore 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*, pages 279–285, Dublin, Ireland. Association for Computational Linguistics.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2020. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *Preprint*, arXiv:1910.01108.

Omar Sharif and Mohammed Moshiul Hoque. 2021. Identification and classification of textual aggression in social media: Resource creation and evaluation. In *Combating Online Hostile Posts in Regional Languages during Emergency Situation*, pages 9–20, Cham. Springer International Publishing.

Omar Sharif and Mohammed Moshiul Hoque. 2022. Tackling cyber-aggression: Identification and fine-grained categorization of aggressive texts on social media using weighted ensemble of transformers. *Neurocomputing*, 490:462–481.

Omar Sharif, Eftekhar Hossain, and Mohammed Moshiul Hoque. 2021a. Combating hostility: Covid-19 fake news and hostile post detection in social media. *CoRR*, abs/2101.03291.

Omar Sharif, Eftekhar Hossain, and Mohammed Moshiul Hoque. 2021b. NLP-CUET@DravidianLangTech-EACL2021: Offensive language detection from multilingual code-mixed text using transformers. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 255–261, Kyiv. Association for Computational Linguistics.

C.N Subalalitha and E. Poovammal. 2018. Automatic bilingual dictionary construction for tirukural. *Applied Artificial Intelligence*, 32(6):558–567.