CUET_Novice@DravidianLangTech 2025: A Bi-GRU Approach for Multiclass Political Sentiment Analysis of Tamil Twitter (X) Comments

Arupa Barua, Md Osama and **Ashim Dey**
Department of Computer Science and Engineering
Chittagong University of Engineering & Technology, Chattogram-4349, Bangladesh
{u2004089, u1804039}@student.cuet.ac.bd, ashim@cuet.ac.bd

Abstract

Multilingual political sentiment analysis faces challenges in capturing subtle variations, especially in complex and low-resourced languages. Identifying sentiments correctly is crucial to understanding public discourse. A shared task on Political Multiclass Sentiment Analysis of Tamil X (Twitter) Comments, organized by DravidianLangTech@NAACL 2025, provided an opportunity to tackle these challenges. For this task, we implemented two data augmentation techniques, which are synonym replacement and back translation, and then explored various machine learning (ML) algorithms. We experimented with deep learning (DL) models including GRU, BiLSTM, BiGRU, hybrid CNN-GRU and CNN-BiLSTM to capture the semantic meanings more efficiently using FastText and CBOW embedding. The Bidirectional Gated Recurrent Unit (BiGRU) achieved the best macro-F1 (MF1) score of 0.33, securing the 17th position in the shared task. These findings underscore the challenges of political sentiment analysis in low-resource languages and the need for advanced language-specific models for improved classification.

1 Introduction

Political sentiments are the views and feelings expressed by individuals or groups about political issues. Classifying political sentiments is crucial to understand public perspectives and addressing a variety of points of view. In multilingual contexts, sentiment analysis in Tamil is especially crucial due to the linguistic and cultural nuances that shape sentiment expression. The shared task on Political Multiclass Sentiment Analysis of Tamil X(Twitter) Comments organized by DravidianLangTech@NAACL 2025 aimed to address this challenge by identifying the types of political sentiments into seven classes: Substantiated, Sarcastic, Opinionated, Positive, Negative, Neutral and None of the above. Their workshop paper (Hegde et al., 2023) provided us an opportunity to engage with these challenges in processing South Asian languages and to leverage our work on political multiclass sentiment analysis.

In our participation, we focus on addressing the challenges of political sentiment classification through two primary contributions:

- We implement data augmentation in two steps to bring more diversity and balance to the training data.
- We leverage different machine learning and deep learning approaches to better capture contextual nuances and improve the overall accuracy of political sentiment classification.

Our code, developed for this shared task can be accessed at `https://github.com/ArupaBarua/DravidianLangTech-NAACL-Sentiment`.

2 Related Work

The complexity of understanding and categorizing political sentiments has driven extensive research employing various languages, datasets, and methodologies. Research has been done to advance sentiment analysis in under-resourced code-mixed languages (Sambath Kumar et al., 2024). Different machine learning models have been employed to classify sentiments in highly under-resourced, code-mixed languages like Tulu and Tamil (Shetty, 2023), (Shanmugavadivel et al., 2022a), (Kanta and Sidorov, 2023), (Ponnusamy et al., 2023), (Thavareesan and Mahesan, 2021). A grid search approach has been explored for analyzing sentiments in code-mixed Tamil and Telulu (B et al., 2024). ML models like SVMs and VSMs have been explored to analyze multiclass sentiments on short texts (K. Suresh Kumar and Moshayedi, 2024). Due to the inefficiency of machine learning models in extracting contextual meanings, their works lack the ability to fully capture nuanced expressions and complex sentiment patterns. Deep learning

Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 574–579
May 3, 2025 ©2025 Association for Computational Linguistics

models, CNN, and LSTM are particularly significant in this case as they excel at capturing complex patterns and contextual nuances in code-mixed languages (Rajasekar and Geetha, 2023), (Nithya et al., 2022), (Mandalam and Sharma, 2021). For multiclass sentiment analysis, Bidirectional Recurrent Neural Network (BiRNN) and its variations like BiLSTM have also been experimented (Krosuri and Aravapalli, 2024), (Roy and Kumar, 2021). To enhance the performance of ML and DL algorithms, different hybrid models have also been explored for Tamil sentiment analysis (Ramesh Babu, 2022), (Gandhi et al., 2021), (Shanmugavadivel et al., 2022b). Recently, transformer-based models like m-BERT, MiniLM, and Indic-BERT have been applied to hate speech detection, demonstrating improved contextual understanding and classification accuracy (Tofa et al., 2025). Multilingual transformers have been explored for multiclass sentiment analysis in code-mixed data, effectively capturing contextual nuances in low-resource languages (Nazir et al., 2025).

3 Task and Dataset Description

This shared task focuses on Political Multiclass Sentiment Analysis of Tamil X(Twitter) Comments. The task organizers provided a dataset comprising X (Twitter) comments in Tamil language, which are annotated with seven categories (Chakravarthi et al., 2025). The objective is to classify these sentiments into seven labels, which are as follows:

- **Substantiated** – Sentiment backed by evidence, reference or logical reasoning.
- **Sarcastic** – Sentiment expressed in a mocking or ironic tone.
- **Opinionated** – Sentiment based on personal beliefs or viewpoints.
- **Positive** – Sentiment expressing approval or good feeling towards a political entity.
- **Negative** – Sentiment expressing criticism.
- **Neutral** – Sentiment that is impartial or does not express a strong emotion.
- **None of the above** – Sentiment that does not fit into any of the specified categories.

The distribution of the political sentiment classes across the training, development and test datasets is shown in Table 1. The training dataset exhibits a noticeable class imbalance, with the "Opinionated" category having the highest representation, significantly outnumbering other classes. In contrast, categories like "None of the above" and "Substantiated" have relatively fewer samples, which is illustrated in Figure 1.

Table 1: Class distribution across datasets.

Class	Train	Dev	Test
Opinionated	1361	153	171
Sarcastic	790	115	106
Neutral	637	84	70
Positive	575	69	75
Substantiated	412	52	51
Negative	406	51	46
None of the above	171	20	25

Figure 1: Class distribution in the train set

4 Methodology

The methods and strategies applied to predict the classes of political sentiments are discussed in this section. Through thorough analysis, we propose a Bidirectional Gated Recurrent Unit (BiGRU) network to estimate the multiclasses of political sentiments. Figure 2 provides a visualization of our methodology, outlining the key steps involved.

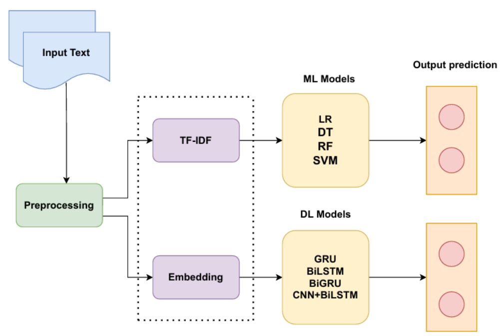

Figure 2: An abstract view of our methodology.

4.1 Preprocessing

In this step, several techniques were applied to refine the X comments. We cleaned the text by removing URLs, emojis, HTML tags, punctuation and special characters, normalized white spaces, and converted all text to lowercase. To address the class imbalance, we implemented data augmentation on the training data in two steps. First, we applied synonym replacement with the FastText Tamil model. For each word, we retrieved its nearest synonym from the pre-trained model and replaced it accordingly to enhance diversity. Then, to improve predictions for minority classes, we applied back-translation to the underrepresented categories 'None of the above', 'Negative', and 'Substantiated' using the mBART model (Tang et al., 2020) and then implemented RandomOverSampling. Finally, we applied tokenization and padding to the text sequences.

4.2 Feature Extraction

To capture meaningful features we used Term Frequency-Inverse Document Frequency (TF-IDF) for ML models And for DL models, we performed two types of embeddings: CBOW Word2Vec embedding and pre-trained FastText Tamil embedding. These embeddings were used to transform input tokens into dense vector representations, capturing semantic word relationships. The pre-trained embeddings were fine-tuned during model training to enhance the model's understanding of the text.

4.3 Model Building

In our research, we explored a variety of ML and DL models.

4.3.1 ML models

We trained traditional ML models such as Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines and Multinomial Naive Bayes on TF-IDF features. The models identify patterns statistically but struggle to extract the complex contextual meanings of the sentiments.

4.3.2 DL models

The deep learning models implemented for this task include GRU (Sachin et al., 2020), BiLSTM (Xu et al., 2019), BiGRU (Xu et al., 2024), hybrid CNN-GRU (Adam and Setiawan, 2023) and CNN-BiLSTM (Liu et al., 2020). Each DL model was trained for 8 epochs with a batch size of 64. We also applied layer normalization, dropout and the Adam optimizer (Kingma and Ba, 2014) for better generalization and more balanced learning. These models learn contextual word meanings based on how words appear in the training data via the embedding layer. The embedded vector is then passed into the network layers which capture contextual dependencies by learning the order and relationships between words and sentiment patterns.

5 Results and Discussion

In this section, we compare the performance achieved by different ML and DL models. The effectiveness of the models is primarily assessed based on the macro F1-score. The hyperparameters of the DL models were manually fine-tuned based on their performance on the validation data. The final hyperparameter values are as shown in Table 2. A summary of the precision (P), recall (R), and macro-F1 (MF1) scores for each model on the test set is presented in Table 3. Through our analysis, we found that the Bidirectional Gated Recurrent Unit (BiGRU) achieves the highest macro-F1 score of 0.33 on the test dataset using the FastText embedding, outperforming other ML and DL models. By processing both preceding and succeeding words, BiGRU enhances feature extraction for better sentiment classification than GRU. While CNN-GRU and CNN-BiLSTM benefit from convolutional feature extraction, the CNN component processes text with fixed-size receptive fields, which may restrict the recurrent layers' ability to

Table 2: The hyperparameters in BiGRU model

Hyperparameters	Values
Embedding Dimension	300
Units	300
Dropout Rate	0.2
Learning Rate	0.001
Optimizer	Adam
Loss Function	Sparse CCE
Batch Size	64
Epochs	8

Table 3: Results of various models on the test dataset

Classifier	P	R	MF1
LR	0.17	0.22	0.16
DT	0.21	0.23	0.21
RF	0.30	0.26	0.26
SVM	0.18	0.23	0.17
MNB	0.16	0.24	0.18
CBOW Embedding			
GRU	0.36	0.29	0.30
BiLSTM	**0.37**	0.29	0.30
CNN-BiLSTM	0.31	0.26	0.28
CNN-GRU	0.28	0.27	0.26
BiGRU	0.35	0.29	0.30
FastText Embedding			
GRU	0.32	0.29	0.30
BiLSTM	0.31	0.27	0.29
CNN-BiLSTM	0.34	0.31	0.31
CNN-GRU	0.27	0.27	0.26
BiGRU	0.35	**0.32**	**0.33**

capture long-range dependencies effectively. Moreover, these hybrid models have a higher number of parameters, making them more prone to overfitting. BiLSTM, though similar to BiGRU, has higher computational complexity and may overfit on smaller datasets, whereas BiGRU achieves a balance between performance and efficiency.

5.1 Quantitative Discussion

The performance of the BiGRU model for Political Multiclass Sentiment Analysis of Tamil X Comments is evaluated using a confusion matrix and ROC curve, as illustrated in Figure 3 and 4. The confusion matrix shows the model correctly classifies Opinionated comments with high accuracy (104 instances). However, Negative (16 misclassified as Opinionated), Neutral (25 misclassified as Opinionated), and Sarcastic (39 misclassified as Opinionated) sentiments exhibit significant misclassification, suggesting the model struggles to differentiate these classes. None of the above classes achieves a high correct classification rate (18 instances out of 23). The ROC curve highlights the model's varying discrimination ability. None of the above has the highest AUC (0.980), indicating strong separability, while Negative has the lowest (0.479), suggesting frequent misclassification. Other classes fall within 0.552–0.665, reflecting moderate distinction. The micro-average AUC of 0.700 suggests overall moderate performance, with challenges in handling nuanced sentiments.

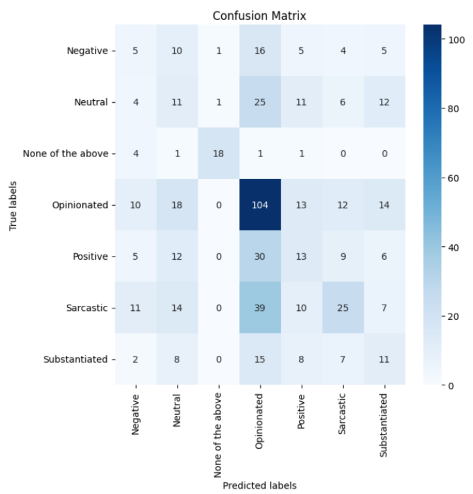

Figure 3: Confusion matrix of BiGRU model using FastText embedding

5.2 Qualitative Discussion

The BiGRU model's performance highlights the challenges posed by dataset imbalance, with minority classes like Negative, Substantiated, and Sarcastic often misclassified as Opinionated, reflecting a bias toward dominant sentiment patterns. Data augmentation played a crucial role in improving the model's performance by introducing more representative samples for the minority classes. Synonym replacement enhanced lexical diversity, while back-translation helped the model better capture variations in Negative, Substantiated, and None of the above sentiments. This led to a more balanced learning process, reducing extreme misclassification. However, the model still struggled with subtle sentiment distinctions, particularly implicit negativity and sarcasm, due to its reliance on sequential dependencies, which limited its ability to fully capture complex political context.

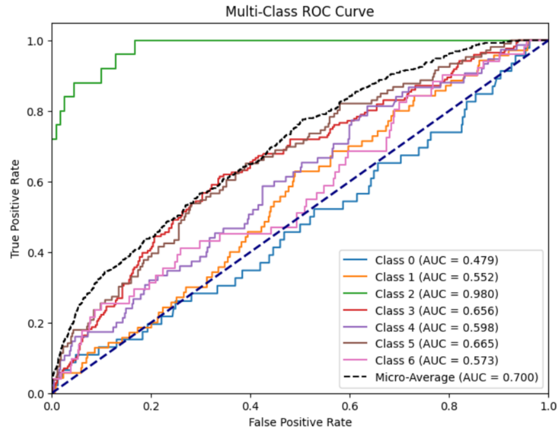

Figure 4: ROC curve of BiGRU model using FastText embedding

6 Conclusion

This study explored Political Multiclass Sentiment Analysis of Tamil X Comments as part of the DravidianLangTech@NAACL 2025 shared task. Key challenges included dataset imbalance and a test set that was not a strong representative of the embeddings. To address these issues, we employed synonym replacement to expand the dataset, improving the representation of embeddings and back-translation augmentation for underrepresented classes to enhance model robustness. Among various ML and DL models, the BiGRU model demonstrated the best performance, achieving an MF1 score of 0.33, a precision of 0.35, and recall of 0.32 using FastText embedding. Future work should explore domain-adaptive transformer models tailored for low-resource languages to further improve sentiment classification performance. Models such as mBERT, IndicBERT, and Tamil-BERT could be fine-tuned to political discourse data to enhance sentiment classification accuracy.

Limitations

The BiGRU model exhibited several limitations due to its reliance on sequential dependencies, which limited its ability to capture complex contextual nuances, leading to frequent misclassification of minority classes despite data augmentation efforts. Synonym replacement was applied to bring diversity to the training data, but it could not fully address the intricacies of sentiment variations. Back translation was implemented to improve prediction for minority classes, but this technique also struggled with the challenge of handling Out-of-Vocabulary (OOV) words, especially in informal social media text with transliterations, code-mixing, and spelling variations. Another key limitation was the imbalance between the training and test datasets, with the test dataset not being a strong representative of the training data, affecting the model's generalization ability. A more balanced dataset and transformer-based models could enhance contextual understanding and improve accuracy, particularly in handling nuanced sentiments and linguistic variations.

References

Ahmad Zahri Ruhban Adam and Erwin Budi Setiawan. 2023. Social media sentiment analysis using convolutional neural network (cnn) dan gated recurrent unit (gru). *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, 9(1):119–131.

Prathvi B, Manavi K, Subrahmanyapoojary K, Asha Hegde, Kavya G, and Hosahalli Shashirekha. 2024. MUCS@DravidianLangTech-2024: A grid search approach to explore sentiment analysis in code-mixed Tamil and Tulu. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 257–261, St. Julian's, Malta. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Saranya Rajiakodi, Elizabeth Sherly, Thenmozhi Durairaj, Sathiyaraj Thangasamy, Ratnasingam Sakuntharaj, Prasanna Kumar Kumaresan, Kishore Kumar Ponnusamy, Arunaggiri Pandian Karunanidhi, and Rohan R. 2025. Overview of the Shared Task on Political Multiclass Sentiment Analysis of Tamil X(Twitter) Comments: DravidianLangTech@NAACL 2025. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Usha Devi Gandhi, Priyan Malarvizhi Kumar, Gokulnath Chandra Babu, and Gayathri Karthick. 2021. Sentiment analysis on twitter data by using convolutional neural network (cnn) and long short term memory (lstm). *Wireless Personal Communications*, pages 1–10.

Asha Hegde, Bharathi Raja Chakravarthi, Hosahalli Lakshmaiah Shashirekha, Rahul Ponnusamy, Subalalitha Cn, SK Lavanya, Durairaj Thenmozhi, Martha Karunakar, Shreya Shreeram, and Sarah Aymen. 2023. Findings of the shared task on sentiment analysis in tamil and tulu code-mixed text. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, pages 64–71.

T. Ananth Kumar Ahmad Jalili Mehdi Gheisari Yasir Malik Hsing-Chung Chen K. Suresh Kumar, A.S. Radha Mani and Ata Jahangir Moshayedi. 2024. Sentiment analysis of short texts using svms and vsms-based multiclass semantic classification. *Applied Artificial Intelligence*, 38(1):2321555.

Selam Kanta and Grigori Sidorov. 2023. Selam@ dravidianlangtech: Sentiment analysis of code-mixed dravidian texts using svm classification. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, pages 176–179.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lakshmi Revathi Krosuri and Rama Satish Aravapalli. 2024. Novel heuristic bidirectional-recurrent neural network framework for multiclass sentiment analysis classification using coot optimization. *Multimedia Tools and Applications*, 83(5):13637–13657.

Zi-xian Liu, De-gan Zhang, Gu-zhao Luo, Ming Lian, and Bing Liu. 2020. A new method of emotional analysis based on cnn–bilstm hybrid neural network. *Cluster Computing*, 23:2901–2913.

Asrita Venkata Mandalam and Yashvardhan Sharma. 2021. Sentiment analysis of dravidian code mixed data. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 46–54.

Muhammad Kashif Nazir, CM Nadeem Faisal, Muhammad Asif Habib, and Haseeb Ahmad. 2025. Leveraging multilingual transformer for multiclass sentiment analysis in code-mixed data of low-resource languages. *IEEE Access*.

K Nithya, S Sathyapriya, M Sulochana, S Thaarini, and CR Dhivyaa. 2022. Deep learning based analysis on code-mixed tamil text for sentiment classification with pre-trained ulmfit. In *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, pages 1112–1116. IEEE.

Kishore Kumar Ponnusamy, Charmathi Rajkumar, Prasanna Kumar Kumaresan, Elizabeth Sherly, and Ruba Priyadharshini. 2023. Vel@ dravidianlangtech: Sentiment analysis of tamil and tulu. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, pages 211–216.

M Rajasekar and Angelina Geetha. 2023. Sentiment analysis of tamil tweets using deep convolution neural networks. In *2023 First International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI)*, pages 1–5. IEEE.

Suba Sri Ramesh Babu. 2022. *Sentiment Analysis In Tamil Language Using Hybrid Deep Learning Approach*. Ph.D. thesis, Dublin, National College of Ireland.

Pradeep Kumar Roy and Abhinav Kumar. 2021. Sentiment analysis on tamil code-mixed text using bi-lstm. In *FIRE (Working Notes)*, pages 1044–1050.

Sharat Sachin, Abha Tripathi, Navya Mahajan, Shivani Aggarwal, and Preeti Nagrath. 2020. Sentiment analysis using gated recurrent neural networks. *SN Computer Science*, 1:1–13.

Lavanya Sambath Kumar, Asha Hegde, Bharathi Raja Chakravarthi, Hosahalli Shashirekha, Rajeswari Natarajan, Sajeetha Thavareesan, Ratnasingam Sakuntharaj, Thenmozhi Durairaj, Prasanna Kumar Kumaresan, and Charmathi Rajkumar. 2024. Overview of second shared task on sentiment analysis in code-mixed Tamil and Tulu. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 62–70, St. Julian's, Malta. Association for Computational Linguistics.

Kogilavani Shanmugavadivel, Sai Haritha Sampath, Pramod Nandhakumar, Prasath Mahalingam, Malliga Subramanian, Prasanna Kumar Kumaresan, and Ruba Priyadharshini. 2022a. An analysis of machine learning models for sentiment analysis of tamil code-mixed data. *Computer Speech & Language*, 76:101407.

Kogilavani Shanmugavadivel, VE Sathishkumar, Sandhiya Raja, T Bheema Lingaiah, S Neelakandan, and Malliga Subramanian. 2022b. Deep learning based sentiment analysis and offensive language identification on multilingual code-mixed data. *Scientific Reports*, 12(1):21557.

Poorvi Shetty. 2023. Poorvi@ dravidianlangtech: Sentiment analysis on code-mixed tulu and tamil corpus. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, pages 124–132.

Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, Jiatao Gu, and Angela Fan. 2020. Multilingual translation with extensible multilingual pretraining and finetuning.

Sajeetha Thavareesan and Sinnathamby Mahesan. 2021. Sentiment analysis in tamil texts using k-means and k-nearest neighbour. In *2021 10th International Conference on Information and Automation for Sustainability (ICIAfS)*, pages 48–53. IEEE.

Farjana Alam Tofa, Lorin Tasnim Zeba, Md Osama, and Ashim Dey. 2025. CUET_INSights@NLU of Devanagari script languages 2025: Leveraging transformer-based models for target identification in hate speech. In *Proceedings of the First Workshop on Challenges in Processing South Asian Languages (CHiPSAL 2025)*, pages 267–272, Abu Dhabi, UAE. International Committee on Computational Linguistics.

Guixian Xu, Yueting Meng, Xiaoyu Qiu, Ziheng Yu, and Xu Wu. 2019. Sentiment analysis of comment texts based on bilstm. *Ieee Access*, 7:51522–51532.

Wei Xu, Jianlong Chen, Zhicheng Ding, and Jinyin Wang. 2024. Text sentiment analysis and classification based on bidirectional gated recurrent units (grus) model. *arXiv preprint arXiv:2404.17123*.