KCRL@DravidianLangTech 2025: Multi-View Feature Fusion with XLM-R for Tamil Political Sentiment Analysis

Md. Ayon Mia, Fariha Haq, Md. Tanvir Ahammed Shawon, Golam Sarwar Md. Mursalin, Muhammad Ibrahim Khan
Department of Computer Science and Engineering
Chittagong University of Engineering and Technology, Bangladesh
{u1804128, u1904051, u1904077}@student.cuet.ac.bd,
sarwarmursalin1015@gmail.com, muhammad_ikhan@cuet.ac.bd

Abstract

Political discourse on social media platforms significantly influences public opinion, necessitating accurate sentiment analysis for understanding societal perspectives. This paper presents a system developed for the shared task of Political Multiclass Sentiment Analysis in Tamil tweets. The task aims to classify tweets into seven distinct sentiment categories: Substantiated, Sarcastic, Opinionated, Positive, Negative, Neutral, and None of the above. We propose a Multi-View Feature Fusion (MVFF) architecture that leverages XLM-R with a CLS-Attention-Mean mechanism for sentiment classification. Our experimental results demonstrate the effectiveness of our approach, achieving a macro-average F1-score of 0.37 on the test set and securing the 2nd position in the shared task. Through comprehensive error analysis, we identify specific classification challenges and demonstrate how our model effectively navigates the linguistic complexities of Tamil political discourse while maintaining robust classification performance across multiple sentiment categories.

1 Introduction

Social media platforms have evolved into primary channels for expressing political opinions, generating massive volumes of data that demand sophisticated analysis techniques (Aqlan et al., 2019). While traditional sentiment analysis often emphasizes binary positive-negative classification, contemporary approaches must interpret nuanced evaluative meanings, particularly in political discourse (Alemayehu et al., 2023), (Katta and Hegde, 2019). The growing influence of social media on public opinion formation has highlighted the critical need for advanced sentiment analysis in diverse linguistic contexts, especially for low-resource languages like Tamil. This research addresses this challenge by focusing on Political Multiclass Sentiment Analysis of Tamil tweets, classifying them into seven distinct categories: Substantiated, Sarcastic, Opinionated, Positive, Negative, Neutral, and None of the above. The task presents unique challenges due to Tamil's linguistic complexity, the contextual nuances of political discourse, and the inherent informality of social media communication. The key contributions of this work are as follows:

- Development of a Multi-View Feature Fusion (MVFF) architecture incorporating XLM-R with CLS-Attention-Mean mechanism for robust Tamil political sentiment analysis
- Extensive experimental evaluation across multiple sentiment categories, demonstrating the model's effectiveness through rigorous performance metrics and comparative analysis

The following GitHub repository contains the complete implementation details: `https://github.com/Ayon128/Shared-Task/tree/main/Political%20Task`

2 Related Works

Research in sentiment analysis has demonstrated significant progress across various methodologies. (Nandi and Agrawal, 2016) enhanced sentiment analysis by combining lexical approaches with Linear SVC, achieving 93% accuracy. (Attia et al., 2018) proposed a language-independent CNN model, achieving accuracies of 78.3%, 75.45%, and 67.93% for English, German, and Arabic respectively.(Rao et al., 2020) explored traditional machine learning approaches, where linear kernel SVM reached 80% accuracy. (Derbentsev et al., 2022) compared deep neural networks using Word2vec and Glove vectorization, with CNN achieving 90.1% on IMDb reviews and BiLSTM-CNN reaching 82.1% on Sentiment140. (Alemayehu et al., 2023) compared neural architectures, with CNN-Bi-LSTM achieving 91.60% accuracy. (Dehghani and Yazdanparast, 2023) com-

Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 682–687

bined CNN and LSTM architectures, reaching 89% accuracy on their primary dataset. Recent advancements continue to show promising results, with (Ebabu and Chalie, 2024) evaluating models for code-mixed text analysis, where CNN demonstrated superior performance. (Rahman et al., 2024) introduced RoBERTa-BiLSTM, which effectively combines transformer capabilities with bidirectional LSTM networks to capture both contextual embeddings and sequential dependencies, achieving state-of-the-art results on multiple benchmark datasets. These studies demonstrate the effectiveness of hybrid approaches and the importance of appropriate model selection for specific language contexts.

3 Task and Dataset Description

This shared task (Chakravarthi et al., 2025) was organized to analyze political discourse in Tamil language content from X (Twitter) by classifying posts into seven sentiment categories. The objective is to classify Tamil tweets into seven distinct sentiment categories: Substantiated, Sarcastic, Opinionated, Positive, Negative, Neutral, and None of the above. Table 1 shows the category-wise distribution of tweets in the training, validation, and test sets. The dataset comprises 5,440 Tamil tweets, distributed across three sets: 4,352 tweets for training, 544 for validation, and 544 for testing. This task aims to advance sentiment analysis capabilities in Tamil, addressing the growing importance of understanding political discourse in low-resource languages on social media platforms.

4 Methodology

Our framework presents an efficient way of carrying out multiclass sentiment analysis of Tamil political comments from X (formerly Twitter). Figure 1 illustrates an abstract overview of the whole system.

4.1 Preprocessing

For standardizing Tamil political tweets, we implement a systematic preprocessing pipeline. The raw text undergoes multiple cleaning operations: URL removal, emoji handling, hashtag and mention elimination, and consecutive punctuation normalization. We also lowercase English text while preserving Tamil script integrity, ensuring consistent text representation for model processing.

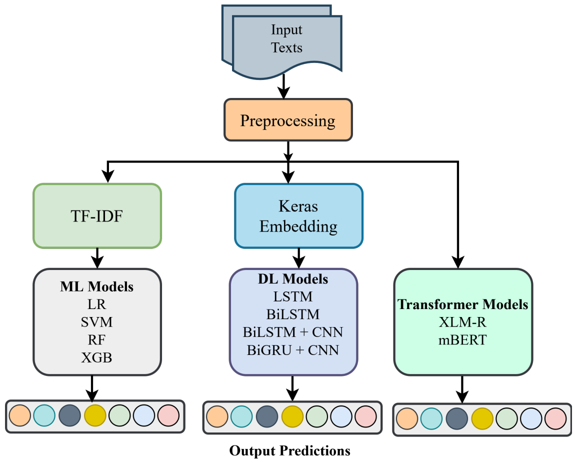

Figure 1: Abstract process of Tamil political sentiment classification using traditional ML models with TF-IDF features, Deep Learning architectures with Keras embeddings, and Transformer models for seven-class sentiment analysis.

4.2 Augmentation

To address the significant class imbalance discovered in Table 1, we employed random oversampling techniques by using scikit-learn. This was crucial due to substantial disparity between the majority classes (Opinionated: 1361 samples) and the minority classes (None: 171 samples). We resampled minority classes (Neutral, Substantiated, Positive, Negative, and None) with replacement specifically to get 200 instances per class. This balanced sampling was essential for preventing model bias towards majority classes and promoting fair learning for all sentiment categories.Without addressing this imbalance, models would likely develop significant bias toward the "Opinionated" class, which constitutes approximately 31% of the training data, while potentially neglecting the "None" category, which represents less than 4% of samples.

4.3 ML-based Approach

For the sentiment classification of Tamil political comments, we implemented traditional ML-based methods: Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and XGBoost (XGB). We utilized TF-IDF (Takenobu, 1994) vectorizer to extract features from the preprocessed dataset. The Random Forest classifier was configured with 200 estimators and a minimum split threshold of 10 samples. For XGBoost, we employed a learning rate of 0.1, 100 estimators, maximum depth of 5, and sampling parameters of 0.8 for both instances and features. The SVM

Table 1: Distribution of Tamil tweets across different sentiment categories in the training, validation, and test sets.

Sets	Classes							Total
	Substantiated	Sarcastic	Opinionated	Positive	Negative	Neutral	None	
Train	412	790	1361	575	406	637	171	4352
Val	52	115	153	69	51	84	20	544
Test	51	106	171	75	46	70	25	544

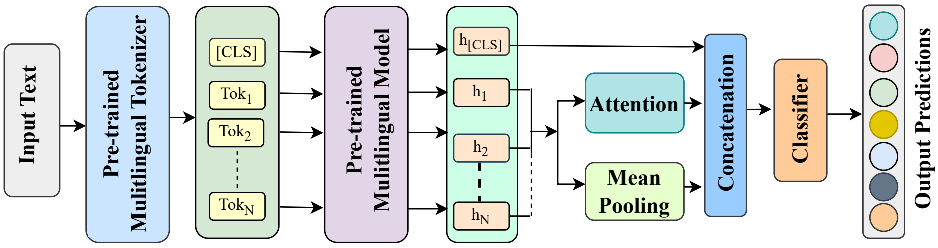

Figure 2: Architecture of XLM-R-based Tamil political sentiment classifier with Multi-View Feature Fusion, combining [CLS], attention, and mean pooling operations for enhanced sentiment classification.

classifier used a linear kernel, while Logistic Regression was implemented with default parameters.

4.4 DL-based Approach

We implemented three different deep learning architectures for the analysis of Tamil political sentiment. A Keras embedding layer was used with a dimensionality of 300 for text representations. The first architecture consists of two LSTM layers (64 and 32 units, respectively) followed by dropout (0.3), batch normalization. The second one proposes a Bidirectional LSTM model that enhances sequential processing. Our third approach implements a hybrid BiLSTM-CNN architecture that combines a Bidirectional LSTM with 128 units, dual-stage CNN with 128 and 64 filters, and kernel size=3. We also developed another variant where LSTM was replaced with GRU units for computational efficiency. All models were trained up to 15 epochs using the categorical cross-entropy loss function with monitoring of validation loss for optimal model selection.

4.5 Transformer-based Approach

We also explored transformer-based approaches by utilizing two multilingual models mBERT (Devlin, 2018) and XLM-R (Conneau, 2019). These pretrained transformer models were imported from Hugging Face (Wolf, 2019) and implemented using Pytorch library. We fine-tuned both models on the dataset using AdamW optimizer with batch size 32 for 15 epochs, implementing early stopping to prevent overfitting and enhance classification performance.

4.6 Multi-View Feature Fusion (MVFF)

Our methodology introduces a Multi-View Feature Fusion (CLS-Attention-Mean) architecture based on XLM-R for Tamil political sentiment classification. The system processes input Tamil text through a specialized tokenizer, generating a sequence $T = \{[\text{CLS}], \text{tok}_1, \text{tok}_2, \ldots, \text{tok}_n\}$, which is embedded into initial representations $E = \{E_{[\text{CLS}]}, E_1, E_2, \ldots, E_n\}$. The XLM-R transformer processes these embeddings through self-attention mechanisms:

$$H = \text{XLM-R}(E)$$

producing contextualized representations $H = \{h_{[\text{CLS}]}, h_1, h_2, \ldots, h_n\}$. Our architecture implements parallel operations: CLS token extraction $h_{[\text{CLS}]} = H[0]$, mean pooling $h_{\text{mean}} = \text{MeanPool}(h_1 : h_n)$, and the attention mechanism. These features are combined through concatenation:

$$H_{\text{fused}} = [h_{[\text{CLS}]}, h_{\text{att}}, h_{\text{mean}}] \,.$$

The final classification output is computed through a feed-forward layer with softmax activation:

$$y = \text{softmax}(W \cdot H_{\text{fused}} + b)$$

where y represents the probability distribution over sentiment classes. The complete architecture is illustrated in Figure 2.

Table 2: Performance comparison across different model architectures on the test set, where Pr, Re, and F1 denote macro-averaged precision, recall, and F1-score respectively.

Model	Pooling Strategy	Performance Metric		
		Pr	**Re**	**F1**
	ML Models			
LR	-	0.36	0.32	0.33
SVM	-	0.36	0.33	0.33
RF	-	0.38	0.33	0.32
XGB	-	0.29	0.31	0.27
	DL Models			
LSTM	-	0.35	0.34	0.35
BiLSTM	-	0.32	0.31	0.32
BiLSTM + CNN	-	0.31	0.31	0.31
BiGRU + CNN	-	0.35	0.34	0.34
	Transformer Models			
mBERT	[CLS]	0.32	0.31	0.31
	Mean	0.36	0.34	0.34
	Attention	0.34	0.35	0.34
	[CLS] + Mean	0.35	0.32	0.32
	[CLS] + Attention	0.35	0.36	0.35
	[CLS] + Mean + Attention	0.36	0.32	0.33
XLM-R	[CLS]	0.36	0.34	0.33
	Mean	0.37	0.36	0.36
	Attention	0.36	0.35	0.35
	[CLS] + Mean	0.35	0.36	0.35
	[CLS] + Attention	0.35	0.34	0.33
	[CLS] + Mean + Attention	**0.38**	**0.37**	**0.37**

5 Experiments and Results

Table 2 presents the comparative results of different models using macro-averaged precision (Pr), recall (Re), and F1-score (F1). Among ML approaches, RF achieved the highest F1-score of 0.32, slightly outperforming LR and SVM with 0.33, while XGB showed lower effectiveness with 0.27 F1-score. The DL architectures, particularly LSTM and BiGRU+CNN, demonstrated stronger performance with F1-scores of 0.35 and 0.34 respectively. This suggests that the integration of convolutional layers with recurrent architectures enhances feature extraction for Tamil sentiment analysis. Our proposed Multi-View Feature Fusion approach using XLM-R with [CLS]+Mean+Attention strategy achieved the best overall performance with precision of 0.38, recall of 0.37, and F1-score of 0.37, significantly outperforming all baselines. The consistent improvement across different pooling strategies validates the effectiveness of feature fusion for capturing diverse aspects of Tamil political sentiment. The mBERT variants also showed competitive performance, with attention-based pooling providing consistent improvements in F1-scores ranging from 0.31 to 0.35. These results demonstrate that transformer-based architectures with sophisticated pooling strategies are more effective at capturing the nuanced sentiments in Tamil political discourse compared to traditional approaches.

6 Error Analysis

The confusion matrix shown in Figure 3, reveals several key misclassification patterns, providing valuable insights into model behavior and limitations. Our Multi-View Feature Fusion model demonstrates strongest performance in classifying "None of the above" (88%) and "Opinionated" (51%) categories. However, there is a notable tendency to misclassify most categories as "Opinionated," particularly evident in "Neutral" (40%) and "Negative" (50%) content. "Substantiated" content shows dispersed misclassification across categories, primarily confused with "Neutral" (25%) and "Opinionated" (22%). This indicates difficulty in identifying factual content with supporting evidence, possibly due to the complex linguistic markers used in Tamil to denote substantiation. While "Sarcastic" content achieves reasonable classification accuracy (41%), it faces confusion with "Opinionated" (24%), highlighting the challenge of detecting subtle sarcastic cues in text-only political communication. The relatively lower performance on "Neutral" and "Positive" categories (14% and 17% true positive rates respectively) suggests a model bias toward more distinctly marked categories, which could be addressed through balanced training data.

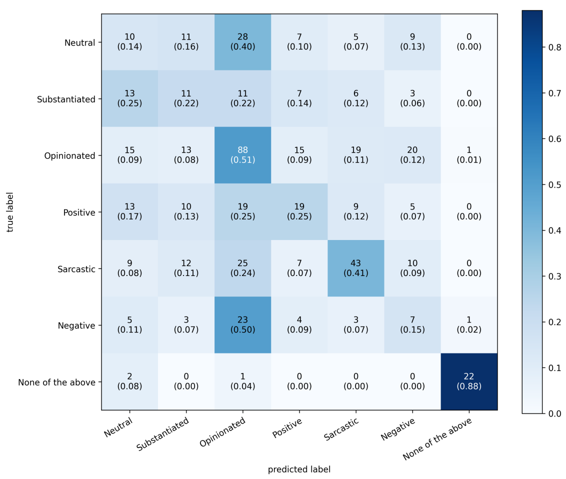

Figure 3: Confusion matrix demonstrating the proposed model's classification performance across seven categories.

7 Conclusions

In this study, we conducted a comprehensive analysis of Tamil political sentiment classification using various machine learning, deep learning, and transformer-based approaches. Our experimental results demonstrated that the Multi-View Feature Fusion approach with XLM-R achieved superior performance through effective integration of [CLS], mean, and attention-based features, obtaining a macro-average F1 score of 0.37. Deep learning architectures like LSTM and BiGRU+CNN showed promising results (F1: 0.35, 0.34), outperforming traditional machine learning approaches, while mBERT variants demonstrated competitive performance with attention-based pooling strategies. The results highlight the importance of combining multiple feature views when analyzing the complex linguistic patterns in Tamil political content, providing a foundation for future sentiment analysis research in low-resource languages.

8 Limitations

Several limitations can be noted in our work. First, the relatively modest F1-scores across all models indicate inherent challenges in Tamil political sentiment analysis. The dataset size constraints and class imbalance issues significantly impacted model development and performance, as evident in our results section. Secondly, our employed models showed limitations in effectively capturing nuanced political sentiments in Tamil text, particularly for complex expressions. The models also underperformed when analyzing tweets without considering broader conversational context or cultural references crucial for accurate classification. Future work should explore advanced techniques for handling class imbalance, larger Tamil political datasets, and enhanced architectures for better sentiment understanding.

References

Fikirte Alemayehu, Million Meshesha, and Jemal Abate. 2023. Amharic political sentiment analysis using deep learning approaches. *Scientific Reports*, 13(1):17982.

Ameen Aqlan, Dr. Manjula Bairam, and R Lakshman Naik. 2019. *A Study of Sentiment Analysis: Concepts, Techniques, and Challenges*, pages 147–162.

Mohammed Attia, Younes Samih, Ali Elkahky, and Laura Kallmeyer. 2018. Multilingual multi-class sentiment classification using convolutional neural networks. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

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.

A Conneau. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Mohammad Dehghani and Zahra Yazdanparast. 2023. Political sentiment analysis of persian tweets using cnn-lstm model. *arXiv preprint arXiv:2307.07740*.

Vasily D Derbentsev, Vitalii S Bezkorovainyi, Andriy V Matviychuk, Oksana M Pomazun, Andrii V Hrabariev, and Alexey M Hostryk. 2022. A comparative study of deep learning models for sentiment analysis of social media texts. In *M3E2-MLPEED*, pages 168–188.

Jacob Devlin. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Yitayew Ebabu and Minalu Chalie. 2024. Sentiment analysis for amharic-english code-mixed sociopolitical posts using deep learning.

Padmaja Katta and Nagaratna Parameshwar Hegde. 2019. A hybrid adaptive neuro-fuzzy interface and support vector machine based sentiment analysis on political twitter data. *International Journal of Intelligent Engineering & Systems*, 12(1).

Vikash Nandi and Suyash Agrawal. 2016. Political sentiment analysis using hybrid approach. *International Research Journal of Engineering and Technology*, 3(5):1621–1627.

Md Mostafizer Rahman, Ariful Islam Shiplu, Yutaka Watanobe, and Md Ashad Alam. 2024. Roberta-bilstm: A context-aware hybrid model for sentiment analysis. *arXiv preprint arXiv:2406.00367*.

Dr D Rajeswara Rao, S Usha, S Krishna, M Sai Ramya, G Charan, and U Jeevan. 2020. Result prediction for political parties using twitter sentiment analysis. *International Journal of Computer Engineering and Technology*, 11(4).

Tokunaga Takenobu. 1994. Text categorization based on weighted inverse document frequency. *Information Processing Society of Japan, SIGNL*, 94(100):33–40.

T Wolf. 2019. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.