Bridging Linguistic Complexity: Sentiment Analysis of Tamil Code-Mixed Text Using Meta-Model

Anusha M D[1], Deepthi Vikram [1], Parameshwar R Hegde[1],

[1]Department of Computer Science,
Yenepoya Institute of Arts Science Commerce and Management,
Yenepoya (Deemed to be University),Balmata, Mangalore
Correspondence: param1000@yahoo.com

Abstract

Sentiment analysis in code-mixed languages poses significant challenges due to the complex nature of mixed-language text. This study explores sentiment analysis on Tamil code-mixed text using deep learning models such as Long Short-Term Memory (LSTM), hybrid models like Convolutional Neural Network (CNN) + Gated Recurrent Unit (GRU) and LSTM + GRU, along with meta-models including Logistic Regression, Random Forest, and Decision Tree. The LSTM+GRU hybrid model achieved an accuracy of 0.31, while the CNN+GRU hybrid model reached 0.28. The Random Forest meta-model demonstrated exceptional performance on the development set with an accuracy of 0.99. However, its performance dropped significantly on the test set, achieving an accuracy of 0.1333. The study results emphasize the potential of meta-model-based classification for improving performance in NLP tasks.

Keywords: Code-mixed, Dravidian Languages, Multi-class, Meta-model, Sentiment Analysis

1 Introduction

In recent years, the analysis of data from social networks and microblogging platforms has garnered significant attention. These platforms are widely used for discussions on a variety of topics, ranging from daily activities and plans to feedback on services and products(Bouazizi and Ohtsuki, 2016). Consequently, businesses and organizations are leveraging such data to extract valuable insights, including user interest in specific topics, satisfaction levels with products and services, and even their intentions and expectations concerning upcoming events like elections or sports competitions.

Another prominent area of research focuses on identifying the attitudes or opinions expressed by users in their posts on specific topics, a process known as "sentiment analysis"(Liu, 2022). The distinctive features of Twitter, a popular microblogging site, make it perfect for sentiment analysis (Memiş et al., 2024). Twitter users can follow others unilaterally, which makes information gathering easier than on many other social networks that require reciprocal connections. It is especially useful for sentiment analysis because of its open format, 140 character limit, and heavy hashtag usage(Bouazizi and Ohtsuki, 2017). Because of the character limit, posts are kept concise and targeted, making it simple to extract insights. Hashtags make Twitter a valuable tool for sentiment analysis by assisting companies in keeping an eye on tweets about their goods or services.

This study investigates hybrid and meta-model techniques for sentiment analysis of Tamil code-mixed text, focusing on multi-class classification. Key contributions include:

- Development of hybrid models (LSTM+GRU, CNN+GRU) for handling code-mixed sentiment data
- Implementation of meta-models (Random Forest, Logistic Regression, Decision Tree) to improve classification accuracy
- Performance evaluation, emphasizing Random Forest's 0.99 accuracy on the development set and generalization issues on the test set

2 Literature Review

This literature review examined sentiment analysis using hybrid models that integrate deep learning architectures to enhance feature extraction and classification, along with meta-models that optimize performance through ensemble methodologies.

2.1 Sentiment Analysis with Hybrid Model

Sentiment analysis plays a critical role in understanding public opinions, particularly from social

Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 31–36
May 3, 2025 ©2025 Association for Computational Linguistics

media data (Liu, 2022). However, the challenges posed by diverse linguistic structures, code-mixed text, and imbalanced datasets have necessitated the development of innovative hybrid models and optimization techniques.

Tan et al. (2022) addressed these challenges by proposing a hybrid model, RoBERTa-LSTM, which integrates the powerful text encoding capabilities of the pre-trained RoBERTa architecture with the ability of LSTM to capture long-term dependencies. To further improve the model's performance, data augmentation techniques utilizing GloVe embeddings were applied to balance the datasets by oversampling underrepresented classes. This hybrid approach achieved impressive F1-scores of 93%, 91%, and 90% on the IMDb, Twitter US Airline Sentiment, and Sentiment 140 datasets, respectively, highlighting its effectiveness in processing diverse textual data In their exploration of hybrid approaches.

A Bi-LSTM-GRU model combined with a Fuzzy Emotion Extractor (FEE) and the Enhanced Aquila Optimizer (EAQ) proposed by Sherin et al. (2024). This approach improved accuracy by 5.6%, 9.8%, and 8.3% on Sentiment 140, T4SA, and Airline Twitter datasets, respectively. However, limitations like few emotion categories and scalability issues were addressed by incorporating BERT and other optimization techniques

2.2 Sentiment Analysis with Meta-Models

Meta-models, commonly used in stacking ensembles, combine predictions from multiple base models to enhance overall performance (Mekala et al., 2020). In stacking, a meta-model is trained on the outputs of base models to determine the best way to integrate them for more accurate predictions.

Historically, multi-class text classification, especially in sentiment analysis, relied on traditional machine learning methods like Naive Bayes, SVM, and Random Forests. These models used feature extraction techniques such as bag-of-words, TF-IDF, and n-grams but faced challenges in capturing complex semantics and managing large-scale data (Yenter and Verma, 2017). Additionally, Support Vector Regression (SVR) was occasionally applied for text similarity, though its utility was limited when handling diverse text sources (Li et al., 2014). In Yenter and Verma (2017), sentiment analysis was performed using the Doc2Vec embedding model with seven classifiers, including KNN, AdaBoost, and SVM, on the US airline services dataset. The AdaBoost classifier achieved 84.5% accuracy, but the small dataset raised concerns about potential underfitting. Jiang et al. (2019) used Word2Vec and GloVe embeddings with LSTM networks for sentiment analysis on US airline tweets, achieving 75% accuracy in classifying sentiments into positive, neutral, and negative categories.

3 Methodology

The proposed study includes text preprocessing, tokenization, and padding to prepare the data for effective analysis and model training.

3.1 Pre-processing

To improve the quality of the text data, preprocessing techniques were used, such as eliminating user mentions, Tamil stopwords, punctuation, and numerical values. For multiclass classification tasks, the text was tokenized into words and sentiment labels (positive, negative, and neutral) were transformed into binary vectors using one-hot encoding.

3.2 Feature Extraction

Words were converted into numerical indices and padding sequences to 100 tokens using a Keras tokenizer. Word vectors for feature extraction were supplied by Tamil FastText embeddings; zero vectors were allocated to missing words, and the word vectors from a pre-trained model formed an embedding matrix. These procedures set up the data so that machine learning models could process it efficiently for sentiment analysis.

3.3 Model Building

Three models were developed for multi-class classification and integrated through a hybrid approach with a meta-model to enhance performance, as shown in Figures 1 and 2. The hybrid model combines individual strengths for a more robust solution to classification challenges.

Hybrid Models

- LSTM: a type of RNN(Graves and Graves, 2012), model sequences and captures long-term dependencies, making it effective for text data. The model used a pre-trained embedding layer, followed by an LSTM layer with 128 units. To prevent overfitting, dropout layers were added, along with a dense layer of 64 units and ReLU activation. The final output

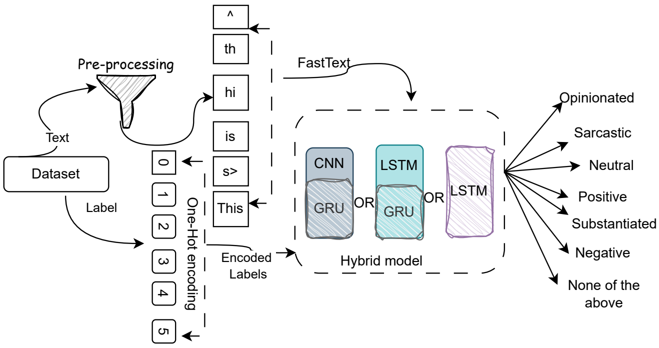

Figure 1: Framework of the Hybrid Sentiment Analysis Model

layer had 7 units with softmax activation for multi-class classification.

- LSTM+GRU Hybrid: combines the strengths of both architectures, with the LSTM layer followed by a GRU layer to capture different aspects of sequential data (MARCELLINA, 2022). GRU units, similar to LSTM, have a simplified structure and fewer parameters. The model uses a pre-trained embedding layer, followed by an LSTM layer (128 units) and a GRU layer (64 units). A dropout layer (rate 0.2) regularizes the network, and the final dense layer (64 units, ReLU activation) is followed by a softmax output layer for class prediction.

- CNN+GRU Hybrid: allows the model to learn both spatial and temporal features by combining CNNs for local feature extraction with GRU layers for processing sequential dependencies (Wu et al., 2020). A max-pooling layer, a 1D convolutional layer (128 filters, kernel size 5), and an embedding layer come first in the model. Sequential dependencies are captured by a 128-unit GRU layer, and the network is regularized by a dropout layer. Finally, a dense layer (7 units, softmax activation) is applied to the output.

Meta-Models(Stacking Ensembles)

To enhance classification performance, a meta-modeling approach was employed, where predictions from base models serve as inputs to a higher-level model that combines these predictions for the final decision (Mekala et al., 2020). This study tested three distinct meta-models:

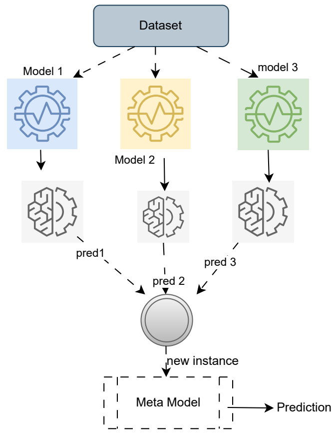

Figure 2: Framework of the proposed Meta-Model, consisting of three models. `pred1`, `pred2`, and `pred3` represent the predictions generated by the three individual models in the hybrid system. The final output is obtained by aggregating these predictions that is new instance.

- Logistic Regression Meta-Model:A linear model that integrates the predictions from base models into a single vector and learns the decision boundary (Taha and Malebary, 2022). Logistic Regression was trained on the stacked predictions from the LSTM+GRU, CNN+GRU, and standalone LSTM models to generate the final classification outcomes.

- Random Forest Meta-Model: An ensemble technique that aggregates the predictions of multiple decision trees (Bjerre et al., 2022). The predictions from the LSTM+GRU, CNN+GRU, and standalone LSTM models served as features for a Random Forest classifier, which was then trained to produce the final sentiment prediction.

- Decision Tree Meta-Model:A tree-based model that makes decisions using input features (Pavel and Soares, 2002). The stacked predictions from LSTM+GRU, CNN+GRU, and standalone LSTM models were used as input to the Decision Tree classifier.

The performance of all three meta-models was evaluated using standard classification metrics includes accuracy, precision, recall, and F1-score. The Random Forest Classifier outperformed the others, delivering the best results on the classification task. The implementation code is available on GitHub

4 Experimental Results

4.1 Dataset

The dataset (Chakravarthi et al., 2025) used in this study is derived from the Political Multiclass Sentiment Analysis of Tamil X (Twitter) Comments, shared during the Dravidian-LangTech@NAACL2025 competition, as outlined in DravidianLangTech@NAACL2025. Specifically curated for multi-class classification tasks, it includes Tamil social media comments with a range of sentiment categories. The dataset is divided into training (4,352 samples) and testing (544 samples) subsets. The sentiment labels span various emotional expressions, and a summary of the label distribution for both the training and testing datasets is presented in Table 1.

Label	Train Count	Test Count
Opinionated	1,361	153
Sarcastic	790	115
Neutral	637	84
Positive	575	69
Substantiated	412	52
Negative	406	51
None of the Above	17	20

Table 1: Label Distribution in Training and Testing Datasets

4.2 Classification

This study experimented with three meta-models (Random Forest, Logistic Regression, and Decision Tree) and three hybrid deep learning models (LSTM, LSTM+GRU, and CNN+GRU) to evaluate sentiment analysis performance on Tamil code-mixed text, focusing on both development and test sets.

The Random Forest meta-model, GRU Hybrid, and CNN+GRU Hybrid models performed best on the development set, leading to their submission in the DravidianLangTech@NAACL 2025 competition. The Random Forest model achieved an impressive 0.99 Macro Average accuracy on the development set but struggled on the test set, with accuracy dropping to 0.1333, suggesting overfitting. The LSTM+GRU and CNN+GRU Hybrid models recorded accuracies of 0.31 and 0.28, respectively, showing potential but not surpassing the Random Forest meta-model.

Class Label	Precision	Recall	F1-Score
Meta Model			
Opinionated	1.00	0.98	0.99
Sarcastic	1.00	0.96	0.98
Neutral	1.00	1.00	1.00
Positive	0.99	0.99	0.99
Substantiated	0.99	1.00	0.99
Negative	0.97	1.00	0.98
None of the Above	1.00	0.98	0.99
Accuracy		**0.99**	
Macro Avg	0.99	0.99	0.99
Weighted Avg	0.99	0.99	0.99
LSTM+GRU Hybrid Model			
Opinionated	0.00	0.00	0.00
Sarcastic	0.14	0.06	0.08
Neutral	0.47	0.80	0.59
Positive	0.32	0.59	0.42
Substantiated	0.26	0.33	0.29
Negative	0.31	0.30	0.30
None of the Above	0.00	0.00	0.00
Accuracy		**0.31**	
Macro Avg	0.22	0.30	0.24
Weighted Avg	0.23	0.31	0.25
CNN + GRU Hybrid Model			
Opinionated	0.02	0.12	0.03
Sarcastic	0.11	0.21	0.14
Neutral	0.80	0.43	0.56
Positive	0.61	0.30	0.40
Substantiated	0.14	0.19	0.16
Negative	0.19	0.24	0.21
None of the Above	0.04	0.50	0.07
Accuracy		**0.28**	
Macro Avg	0.27	0.29	0.23
Weighted Avg	0.45	0.28	0.33

Table 2: Performance metrics for Meta-Model, GRU Hybrid, and CNN + GRU Hybrid.

4.3 Analysis

The literature review indicates that meta-learning techniques are underutilized in multi-class classification, presenting an opportunity for innovation in this area. This study also introduces a novel dataset, publicly available for the first time, which is a significant step in addressing political content classification challenges. By utilizing meta-learning approaches in a new research area, this study expands meta-learning applications and paves the way for future advancements in multi-class classification tasks.

The dataset exhibited class imbalance, leading to overfitting, where models learned biased patterns toward majority classes. This resulted in a sharp decline in accuracy from the development set to the test set. Additionally, while hybrid models are effective, they complicate computational processes. The study also overlooks transformer-based models, which have demonstrated potential for enhancing performance in similar tasks. To better investigate larger datasets, tackle class imbalance, and examine advanced architectures, further research is needed.

5 Conclusion

This study used hybrid deep learning models and sophisticated feature extraction techniques to examine sentiment analysis on Tamil code-mixed text. Due to its ability to capture temporal and spatial dependencies, the LSTM+GRU hybrid performed better than other models. Despite overfitting on the test set, the Random Forest meta-model demonstrated high accuracy during development.

Despite these encouraging outcomes, issues like overfitting and high processing requirements were found. The study results highlight the need for regularization strategies such as dropout, data augmentation, and cross-validation to mitigate overfitting. For increased accuracy, future studies can investigate transformer-based models like BERT, cross-validation, and sophisticated regularization. Addressing class imbalance through techniques like SMOTE or focal loss will further enhance the robustness of future models.

References

Elisa Bjerre, Michael N Fienen, Raphael Schneider, Julian Koch, and Anker L Højberg. 2022. Assessing spatial transferability of a random forest metamodel for predicting drainage fraction. *Journal of Hydrology*, 612:128177.

Mondher Bouazizi and Tomoaki Ohtsuki. 2016. Sentiment analysis: From binary to multi-class classification: A pattern-based approach for multi-class sentiment analysis in twitter. In *2016 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE.

Mondher Bouazizi and Tomoaki Ohtsuki. 2017. A pattern-based approach for multi-class sentiment analysis in twitter. *IEEE Access*, 5:20617–20639.

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.

Alex Graves and Alex Graves. 2012. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, pages 37–45.

Ming Jiang, Junlei Wu, Xiangrong Shi, and Min Zhang. 2019. Transformer based memory network for sentiment analysis of web comments. *IEEE Access*, 7:179942–179953.

Xiaodong Li, Haoran Xie, Li Chen, Jianping Wang, and Xiaotie Deng. 2014. News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, 69:14–23.

Bing Liu. 2022. *Sentiment analysis and opinion mining*. Springer Nature.

JESSLYN MARCELLINA. 2022. *Metode long short-term memory (LSTM), Gated recurrent unit (GRU), Dan convolutional long short-term memory (CONV-LSTM) untuk peramalan data runtun waktu (Studi Kasus: Jumlah Kasus Positif Harian COVID-19 di Indonesia)*. Ph.D. thesis, Universitas Gadjah Mada.

Dheeraj Mekala, Xinyang Zhang, and Jingbo Shang. 2020. Meta: Metadata-empowered weak supervision for text classification. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Erkut Memiş, Hilal Akarkamçı, Mustafa Yeniad, Javad Rahebi, and Jose Manuel Lopez-Guede. 2024. Comparative study for sentiment analysis of financial tweets with deep learning methods. *Applied Sciences*, 14(2):588.

YPPAF Pavel and Brazdil2 Carlos Soares. 2002. Decision tree-based data characterization for meta-learning. *IDDM-2002*, 111.

A Sherin, I Jasmine SelvakumariJeya, and SN Deepa. 2024. Enhanced aquila optimizer combined ensemble bi-lstm-gru with fuzzy emotion extractor for tweet sentiment analysis and classification. *IEEE Access*.

Altyeb Altaher Taha and Sharaf Jameel Malebary. 2022. A hybrid meta-classifier of fuzzy clustering and logistic regression for diabetes prediction. *Computers, Materials & Continua*, 71(3).

Kian Long Tan, Chin Poo Lee, Kalaiarasi Sonai Muthu Anbananthen, and Kian Ming Lim. 2022. Roberta-lstm: a hybrid model for sentiment analysis with transformer and recurrent neural network. *IEEE Access*, 10:21517–21525.

Lizhen Wu, Chun Kong, Xiaohong Hao, and Wei Chen. 2020. A short-term load forecasting method based on gru-cnn hybrid neural network model. *Mathematical problems in engineering*, 2020(1):1428104.

Alec Yenter and Abhishek Verma. 2017. Deep cnn-lstm with combined kernels from multiple branches for imdb review sentiment analysis. In *2017 IEEE 8th annual ubiquitous computing, electronics and mobile communication conference (UEMCON)*, pages 540–546. IEEE.