YenCS@DravidianLangTech 2025: Integrating Hybrid Architectures for Fake News Detection in Low-Resource Dravidian Languages

Anusha M D
Department of Computer Science,
Yenepoya Institute of Arts,
Science, Commerce and Management,
Yenepoya (Deemed to be University),
Balmata, Mangalore
anugowda251@gmail.com

Parameshwar R Hegde
Department of Computer Science,
Yenepoya Institute of Arts,
Science, Commerce and Management,
Yenepoya (Deemed to be University),
Balmata, Mangalore
param1000@yahoo.com

Abstract

Detecting fake news in under-resourced Dravidian languages is a rigorous task due to the scarcity of annotated datasets and the intricate nature of code-mixed text. This study tackles these issues by employing advanced machine learning techniques for two key classification tasks, the first task involves binary classification achieving a macro-average F1-score of 0.792 using a hybrid fusion model that integrates Bidirectional Recurrent Neural Network (Bi-RNN) and Long Short-Term Memory (LSTM)-Recurrent Neural Network (RNN) with weighted averaging. The second task focuses on fine-grained classification, categorizing news where an LSTM-GRU hybrid model attained a macro-average F1-score of 0.26. These findings highlight the effectiveness of hybrid models in improving fake news detection for under-resourced languages. Additionally, this study provides a foundational framework that can be adapted to address similar challenges in other under-resourced languages, emphasizing the need for further research in this area.

Keywords: Dravidian Languages, Fake News Detection, Hybrid Models, Multi-Class Classification

1 Introduction

Fake news consists of misleading or false information that imitates the structure and style of authentic news (Devika et al., 2024). Its spread can cause substantial societal misperceptions, sometimes resulting in severe consequences. Hence, distinguishing genuine news from fake news is essential.

If news is inaccurate, it can mislead individuals and contribute to the dissemination of false information (Subramanian et al., 2025). In some cases, fake news is deliberately used to generate rumors or damage the reputation of political figures (Subramanian et al., 2023). To address this challenge, a system has been proposed for detecting fake news. However, given the vast volume of data available on the internet and social media, manually verifying the authenticity of news content remains a significant challenge(Yigezu et al., 2023).

This widespread phenomenon spreads rapidly, affecting a vast number of people on a daily basis. The far-reaching influence of fake news presents substantial risks to national security, economic stability, and public welfare. Regrettably, many individuals remain unaware of the profound consequences fake news can have on crucial issues and often lack the necessary skills to identify and mitigate such challenges(Yigezu et al., 2023). The study was conducted by the organizers of a shared task, which involves two distinct tasks: classifying social media text as either original or fake, and identifying multiple labels in Malayalam news (Subramanian et al., 2024). This research aims to investigate the effectiveness of different machine learning, deep learning, and hybrid models in tackling critical challenges in text classification tasks. Through the use of advanced techniques and optimization of model architectures, the study seeks to contribute to the development of robust solutions for processing complex datasets, with a particular focus on under-resourced and code-mixed languages.

2 Literature Review

In terms of feature extraction, contextual understanding, and enhancing classification accuracy, RNNs and LSTMs have demonstrated remarkable efficacy Waqas and Humphries (2024). With an emphasis on using deep learning frameworks to process complex textual data, this section gives a summary of recent developments in binary and multi-class classification tasks.

Numerous studies have emphasized the potential of deep learning methods for addressing diverse classification tasks. Yigezu et al. (2024) employed an RNN-LSTM model, with hyperparameters opti-

Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 223–227
May 3, 2025 ©2025 Association for Computational Linguistics

mized via grid search. The model demonstrated notable effectiveness in binary classification, achieving an accuracy of 0.82. However, its performance on multi-class tasks was compromised due to the issue of imbalanced data, resulting in a lower score of 0.32. Similarly,Chauhan and Palivela (2021) applied an LSTM-based approach for fake news detection, utilizing GloVe word embeddings to represent text as vectors, tokenization for feature extraction, and N-grams to enhance feature representation. When compared to Alghamdi et al. (2022) fake news detection methods, their model achieved an outstanding accuracy of 99.88%, highlighting the strength of LSTM networks in processing complex textual data and distinguishing between false and genuine news.

Convolutional neural networks (CNN) and recurrent neural networks with long short-term memory (RNN-LSTM) were combined in Goonathilake and Kumara (2020) to create a hybrid model for text classification. Convolution and max-pooling were used by the CNN to extract features, and the RNN-LSTM to record long-term dependencies. Overfitting was lessened by dropout regularization and dense layers, and the Adam optimizer with binary cross-entropy loss attained 92% accuracy.

3 Methodology

The methodology describes the structured process followed to identify fake news in Dravidian languages. This process encompasses text preprocessing, tokenization, and padding, which are crucial steps for preparing the data for analysis and efficient model training.

3.1 Dataset

The dataset employed in this research originates from the Shared Task on Fake News Detection in Dravidian Languages, organized at DravidianLangTech@NAACL 2025.

Table 1 and Table 2 provide the class-wise distribution of the dataset both the task respectively. This dataset is crucial for advancing fake news detection in Dravidian languages, a less explored area in computational linguistics.

Classes	Train	Test	Dev
Original	1658	512	409
Fake	1599	507	406
Total	**3257**	**1019**	**815**

Table 1: Class-wise Distribution of Dataset for Task A

Classes	Train Set	Test Set
Half True	145	24
False	1,251	149
Partly False	44	14
Mostly False	242	63
Total	**1,682**	**250**

Table 2: Class-wise Distribution of Dataset for Task B

3.2 Pre-processing

The pre-processing pipeline involves tokenization using the Keras Tokenizer[1] after the data has been cleaned up by eliminating stopwords, mentions, punctuation, and numbers. In order to maintain consistent input dimensions, sequences are then padded to 100 words. Lastly, for multi-class classification, labels are encoded using one-hot encoding method.

3.3 Feature Extraction

Semantic representations of words in the dataset are derived using FastText embeddings, as detailed in FastText. FastText, an extension of Word2Vec (Church, 2017), represents words as bags of character n-grams, enabling it to generate meaningful embeddings even for out-of-vocabulary words. Each word in the Malayalam text is mapped to a dense vector using pre-trained FastText embeddings. This approach enhances the performance of subsequent machine learning and deep learning models by improving their understanding of language patterns in code-mixed text(Umer et al., 2023).

3.4 Model Building

1. Task A: Binary Classification

   In this task, a hybrid model approach combined with an ensemble strategy using weighted averaging, known as the fusion model (Alyahyan, 2025), is employed to classify social media posts, particularly YouTube comments, into fake or original categories.

   - Bi-RNN: uses tokenized and padded sequences, with a pre-trained embedding layer and a Bi-RNN layer (128 units) to capture contextual dependencies in both forward and backward directions. The output is passed through a dense layer with ReLU activation before the final classification layer. This structure allows the model to effectively learn from

[1] https://keras.io/keras_hub/api/tokenizers/tokenizer/

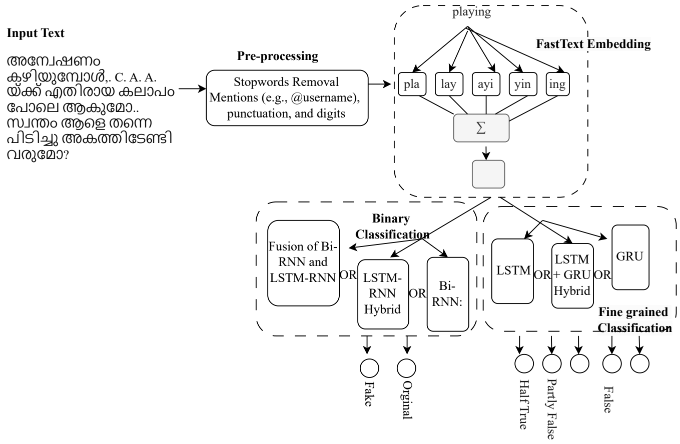

Figure 1: Proposed Methodology for Binary and Fine-Grained Classification

sequential data with complex dependencies(Yang et al., 2022).

- LSTM-RNN Hybrid: combines an embedding layer, an LSTM layer (128 units), and a Simple RNN layer (64 units) to extract sequential features(Telmem et al., 2024). The output is processed by a dense layer for classification.
- Fusion of Bi-RNN and LSTM-RNN (Ensemble Method): The predictions from the Bi-RNN and LSTM-RNN models are integrated through weighted averaging, utilizing the distinct advantages of each model to enhance performance. (Telmem et al., 2024). Figure 2 illustrates the fusion model.

  Experiments with ensembles of DNN, LSTM, RNN, and GRU models were also conducted. The predictions from each model were weighted and aggregated, enhancing robustness and performance in classification tasks.

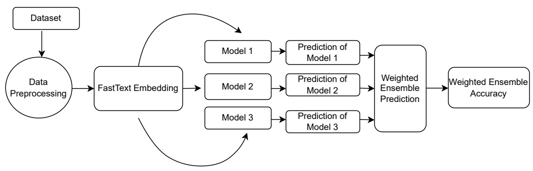

Figure 2: Architecture of the Fusion Model

2. Task B: Fine-grained classification

   The methodology focuses on detecting and classifying fake news in Malayalam-language news articles. This is achieved by employing a range of advanced model architectures to categorize the articles into five predefined categories.

   - GRU: employs an embedding layer initialized with a pre-trained embedding matrix, followed by a Conv1D layer (128 filters) with ReLU activation, and a MaxPooling1D layer for downsampling. The model includes a GRU layer (128 units) to capture sequential dependencies, followed by a Dropout layer (0.2) for regularization(Xu, 2024). The output layer uses softmax activation to handle the multi-class classification task with 5 output classes.
   - LSTM + GRU Hybrid: begins with an embedding layer, followed by an LSTM layer (128 units) with return_sequences=True to capture long-term dependencies. A GRU layer (64 units) processes the LSTM output to extract further sequential features Mousa et al. (2024). After applying a Dropout layer (0.2), the model proceeds through a Dense layer (64 units) with ReLU activation. The final output layer uses softmax activation, suitable for multi-class classification.
   - LSTM: starts with an embedding layer, followed by an LSTM layer (128 units) to capture sequential dependencies in the data. After a Dropout layer (0.2), the output is flattened and passed through a Dense layer (64 units) with ReLU ac-

tivation (Telmem et al., 2024). The final output layer uses softmax activation, classifying the data into one of the 5 output classes.

Following the completion of all experiments for binary classification, the Bi-RNN, Fusion Model, and LSTM+GRU Hybrid exhibited promising accuracy. These models were then submitted to the task organizer, where the Fusion Model attained the highest performance on the test set. In Task B, the three models outlined in the methodology were submitted, with the LSTM+GRU Hybrid emerging as the top performer among them. The results of the submitted models on the development set are presented in Tables 3 and 4. The proposed methodology is illustrated in Figure 1. The implementation code is available on GitHub.

4 Results

The results from the test set reveal differences in performance between Task A and Task B. In Task A, i.e., Fake vs Original news classification, the Fusion Model achieved a macro-average score of 0.792, demonstrating strong performance in distinguishing between fake and original news. However, in Task B (fine-grained fake news classification), the LSTM + GRU Hybrid model scored a much lower macro-average of 0.26, highlighting the difficulty of classifying nuanced categories like Half True, False, and Mostly False.

Model	**Precision**	**Recall**	**F1-score**
Bi-RNN	0.75	0.75	0.75
Fusion Model	0.83	0.82	0.82
LSTM-RNN Hybrid	0.81	0.79	0.79

Table 3: Performance of Task A on Development Set.

Table 3 summarizes the metrics, including accuracy and F1-scores, for Task A, while Table 4 outlines the development set performance for Task B. Test set results and Ranking for both tasks can be accessed on the Fake News Detection in Dravidian Languages DravidianLangTech@NAACL 2025 task page.

Model	**Precision**	**Recall**	**F1-Score**
GRU-Model	0.63	0.67	0.63
LSTM Model	0.58	0.63	0.60
LSTM + GRU Hybrid	0.94	0.93	0.93

Table 4: Performance of Task B on Development Set

In this study, multi-class classification (F1-score 0.26) is hindered by class imbalance and the inability to discern subtle categories. Because sequential models don't have a deep understanding of context, it's more difficult to spot sarcasm and implicit misinformation. In environments with limited resources, real-time deployment is limited by the high computational cost. Reliability is decreased when fact-checking procedures are absent. Interpretability and trust are impacted by deep learning models' black-box nature. Explainability strategies and attention-based models should be investigated in future research.

The low score in Task B may be due to class imbalance and semantic overlap between categories, making it harder for the model to distinguish subtle differences. While the hybrid model captures sequential patterns, it may lack the ability to encode deeper contextual cues, especially without attention mechanisms or transformer-based embeddings. Additionally, the macro-average F1-score penalizes poor performance in labels with minimal classes, further lowering the overall score. These results suggest the need for class balancing, more advanced embeddings like BERT or XLM-R, and attention mechanisms to improve performance.

5 Conclusion

This study examined hybrid models that focus on binary and multi-class classification for the detection of fake news in Dravidian languages. In binary classification, the Fusion Model obtained a robust macro-average F1-score of 0.792, whereas the LSTM + GRU hybrid model had trouble in multi-class classification, achieving an F1-score of 0.26. While highlighting the usefulness of hybrid models for binary tasks, these results also point to the need for more sophisticated techniques in multi-class classification. Attention mechanisms, transformers, context-aware embeddings, and language-specific preprocessing methods for Dravidian languages could all be included in future research.

References

Jawaher Alghamdi, Yuqing Lin, and Suhuai Luo. 2022. A comparative study of machine learning and deep learning techniques for fake news detection. *Information*, 13(12):576.

Saleh Alyahyan. 2025. Fusionnet remote a hybrid deep learning ensemble model for remote image classification in multispectral images. *Discover Computing*, 28(1):3.

Tavishee Chauhan and Hemant Palivela. 2021. Optimization and improvement of fake news detection using deep learning approaches for societal benefit. *International Journal of Information Management Data Insights*, 1(2):100051.

Kenneth Ward Church. 2017. Word2vec. *Natural Language Engineering*, 23(1):155–162.

K Devika, B Haripriya, E Vigneshwar, B Premjith, Bharathi Raja Chakravarthi, et al. 2024. From dataset to detection: A comprehensive approach to combating malayalam fake news. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 16–23.

MDP P Goonathilake and PPN V Kumara. 2020. Cnn, rnn-lstm based hybrid approach to detect state-of-the-art stance-based fake news on social media. In *2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer)*, pages 23–28. IEEE.

Ramin Mousa, Mitra Khezli, Mohamadreza Azadi, Vahid Nikoofard, and Saba Hesaraki. 2024. Classifying objects in 3d point clouds using recurrent neural network: A gru lstm hybrid approach. *arXiv preprint arXiv:2403.05950*.

Malliga Subramanian, , B Premjith, Kogilavani Shanmugavadivel, Santhia Pandiyan, Balasubramanian Palani, and Bharathi Raja Chakravarthi. 2025. Overview of the Shared Task on Fake News Detection in Dravidian Languages: DravidianLangTech@NAACL 2025. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Malliga Subramanian, Bharathi Raja Chakravarthi, Kogilavani Shanmugavadivel, Santhiya Pandiyan, Prasanna Kumar Kumaresan, Balasubramanian Palani, B Premjith, K Vanaja, S Mithunja, K Devika, et al. 2024. Overview of the second shared task on fake news detection in dravidian languages: Dravidianlangtech@ eacl 2024. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 71–78.

Malliga Subramanian, Bharathi Raja Chakravarthi, Kogilavani Shanmugavadivel, Santhiya Pandiyan, Prasanna Kumar Kumaresan, Balasubramanian Palani, Muskaan Singh, Sandhiya Raja, Vanaja, and Mithunajha S. 2023. Overview of the shared task on fake news detection from social media text. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Meryam Telmem, Naouar Laaidi, Youssef Ghanou, Sanae Hamiane, and Hassan Satori. 2024. Comparative study of cnn, lstm and hybrid cnn-lstm model in amazigh speech recognition using spectrogram feature extraction and different gender and age dataset. *International Journal of Speech Technology*, 27(4):1121–1133.

Muhammad Umer, Zainab Imtiaz, Muhammad Ahmad, Michele Nappi, Carlo Medaglia, Gyu Sang Choi, and Arif Mehmood. 2023. Impact of convolutional neural network and fasttext embedding on text classification. *Multimedia Tools and Applications*, 82(4):5569–5585.

Muhammad Waqas and Usa Wannasingha Humphries. 2024. A critical review of rnn and lstm variants in hydrological time series predictions. *MethodsX*, page 102946.

Cong Xu. 2024. Cnn-gru model for ecg signal classification using ucr time series data. *Advances in Engineering Innovation*, 12:31–35.

Xinzhi Yang, Zili Fang, Wenbo Zhang, Lixia Xi, Xiaoguang Zhang, and Nan Cui. 2022. Fiber nonlinear compensation using bi-directional recurrent neural network model based on attention mechanism. In *2022 Asia Communications and Photonics Conference (ACP)*, pages 426–428. IEEE.

Mesay Yigezu, Olga Kolesnikova, Grigori Sidorov, and Alexander Gelbukh. 2024. Habesha@ dravidianlangtech 2024: Detecting fake news detection in dravidian languages using deep learning. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 156–161.

Mesay Gemeda Yigezu, Girma Yohannis Bade, Olga Kolesnikova, Grigori Sidorov, and Alexander F Gelbukh. 2023. Multilingual hope speech detection using machine learning. In *IberLEF@ SEPLN*.