

MSD: Multilingual Sarcasm Detection using Deep Learning-Based Model

Ranjeet Kumar¹, Abhinav Kumar^{1,*†}

¹Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, Uttar Pradesh, 211004

Abstract

Sarcasm detection is a crucial task in natural language processing (NLP), where the intended meaning of a statement diverges from its surface-level interpretation. Sarcasm detection plays a vital role in sentiment analysis and opinion mining. Historically, research in this domain has been limited to single-language input text. However, with the rise of social media, there has been a surge in multilingual data, where users express themselves in a mix of languages such as Hinglish, Tamil, and Malayalam. This paper addresses the need for multilingual sarcasm detection by presenting deep learning-based architectures such as BERT and Xlm-RoBERTa. The multilingual Tamil-English and Malayalam-English texts are first translated into their corresponding English text, and then BERT and Xlm-RoBERTa are used to fine-tune their weight for sarcasm identification. The proposed approach achieves promising performance with macro F_1 -score of 0.72 for both BERT and Xlm-RoBERTa models in the case of Tamil-English posts. In contrast, it achieves macro F_1 -scores of 0.71 for BERT and 0.73 for Xlm-RoBERTa in the case of Malayalam-English posts.

Keywords

Sarcasm, Multilingual, Xlm-RoBERTa, BERT, NLP

1. Introduction

Sarcasm is familiar on social media sites, where a message's deeper significance frequently differs from its apparent reading. Sarcasm detection is a significant difficulty in many applications, mainly where it's essential to discern the speaker's true viewpoint, like in discussion forums, customer evaluations, and sentiment analysis tools [1, 2, 3]. Traditionally, lexical clues and certain language components have been used to approach sarcasm detection as a text categorization problem. However, sarcasm detection has become a far more difficult process with the emergence of social media data from platforms such as Instagram and Twitter [4, 5, 6]. Comprehending sarcasm takes more than just recognizing linguistic patterns; it frequently calls for prior knowledge, contextual awareness, or even a particular degree of maturity or intelligence. Sarcasmic content abounds on social media, especially on microblogging platforms, so it is impracticable to detect sarcastic content manually. Because of this, complex algorithms that can automatically identify and understand sarcasm in user-generated content are now required. Since sarcasm is based on intense emotions that come from one's situation, attitude, or relationships with other people, it frequently crosses over into emotion. People's physical and psychological reactions to emotions like love, hate, or fear are mirrored in online communication. Thus, emotion identification in reviews or comments on social media has become an important part of several study fields [7, 8, 9, 10], including sarcasm detection. Sarcastically worded content's emotions can have a big impact on how users understand user opinions, whether for businesses evaluating customer feedback or for online platforms keeping an eye on public discourse [11, 12].

Dealing with content that lacks obvious context, like headlines, is one of the trickiest parts of sarcasm identification. Single-line news headlines are especially challenging to understand, in contrast to tweets or longer social media posts that offer surrounding information for sarcasm identification. It is possible to significantly change the intended meaning of a sarcastic headline by mistaking it for a real statement. This might result in poorly informed decisions in commercial applications or incorrect interpretations of

Forum for Information Retrieval Evaluation, December 12-15, 2024, India

✉ ranjeet.2022is20@mnnit.ac.in (R. Kumar); abhik@mnnit.ac.in (A. Kumar)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

important information. This problem also highlights worries about how sarcasm could spread dangerous content, including disparaging particular races or ethnic groups in ostensibly innocent yet satirical headlines. Tools that can reliably identify sarcasm in this kind of information are essential to reducing these dangers [13].

Automatic sarcasm identification has become more important and complex due to the increased occurrence of multimodal and multilingual data, particularly as users combine languages like Hinglish, Tamil, and Malayalam more frequently. Furthermore, one of the biggest challenges facing conventional sarcasm detection techniques is the absence of context in headlines and short social media posts. To tackle these challenges, this work explores the usability of two different deep learning models: (i) BERT and (ii) Xlm-RoBERTa. To fine-tune these models, Tamil-English and Malayalam-English datasets are first translated into their corresponding English text.

The rest of the paper is organized as follows: Section 2 lists related literature for identifying sarcastic social media posts. Section 3 discusses the proposed methodology, results of the proposed model are listed in Section 4, and Section 5 concludes the paper.

2. Literature Review

The identification of sarcasm is a challenging topic in natural language processing (NLP) that has attracted a lot of study interest since it relies on subtle linguistic clues such as context, exaggeration, and incongruity. Recognizing sarcasm is essential in several applications, including sentiment analysis, opinion mining, and social media content regulation. A variety of techniques have been investigated to increase the accuracy of sarcasm recognition models [14, 15, 16]. This is especially correct when working with multimodal data; here, the user can express himself by a combination of speech, text, and images.

2.1. Conventional Methods

Sarcasm Detection through Sentiment and Incongruity: Detection of sarcasm using sentiment analysis is a well-known research on sarcasm detection that uses a model that includes sentiment and data to identify sarcasm. This architecture covered representations that are sentiment and incongruity-oriented to this configuration. By using the most used dataset on sarcasm detection, This study shows some of the main shortcomings of other used models in the literature, especially in compiling big texts, and emphasizes the need to include sentiment and sarcasm detection. [17]

Multimodal Sarcasm Detection: Detection of sarcasm that can be handled with multimodal data, photos, text, and audio is more important as the quantity of data generated from social media platforms such as Instagram and Twitter keeps growing. The Multimodal Learning framework is one of the most used methods in this area. By aligning text with visual and oral modalities, this framework enhances the extraction of contextual and emotional elements using a cross-modal target attention mechanism. In the case of the previous modal that prioritizes shared properties, the above framework indicates and shows both the shared and different elements of sentiment analysis and detection of sarcasm. This modal uses the junction between sentiment analysis and sarcasm detection to improve the model's capability to grasp task-specific distinctions. The effectiveness of this multimodal data in improving the detection of sarcasm was explained by extensive testing of the MIL model on the MUSTARD datasets, which is a significant gain over what is already in place [18].

Deep Learning Approaches with Contextual Features: In the case of social media datasets on sarcasm identification, especially tweets, the use of contextual feature extraction in different conjunctions using different deep-learning methods was reported. The writer wrote a system that collects manually created contextual properties based on linguistics with a Convolutional Neural Network(CNN) to extract features. Word embeddings were completed using FastText elements, which combined verbs, pronouns, and event words to enhance contextual awareness. Using different machine learning classifiers to compare the accuracy of the model, the author found that this model most outperformed other models

Table 1

Overall data statistics for Tamil-English and Malayalam-English social media posts

SNo	Language	Class	Train	Validation	Test
1	Malayalam	Not-Sarcastic	10689	2305	2314
		Sarcastic	2499	521	512
		Total	13188	2826	2826
2	Tamil	Not-Sarcastic	21740	4630	4621
		Sarcastic	7830	1706	1717
		Total	29570	6336	6338

on the same datasets in terms of F1-score. To detect sarcasm, deep learning combined with contextual awareness is the most used method, providing feature complexity and model accuracy [19].

Sarcasm Detection in Code-Mixed Conversations: Bedi et al. [20] completed a noticeable investigation focusing on the complexities associated with sarcasm identification and classification in code-mixed languages, especially those that contain Hindi and English. The researcher demonstrated an MSH-COMICS model and presented MaSaC, a unique multimodal dataset for the classification of sarcasm identification. This architecture can analyze the complexities of code-mixed datasets.[6] This uses a tree structural approach that concentrates on the basis of sentences and different dialog. On the basis of this experiment, MSH-COMICS accuracy is improved than previous models by obtaining a higher F1-score value of classification and categorization of sarcasm identification. [20]

3. Proposed Methodology

The overall flow diagram of the proposed methodology can be seen in Figure 1.

3.1. Dataset Description

Malayalam-English and Tamil-English text data are used for this experiment. This dataset has social media posts and their corresponding labels, sarcastic and non-sarcastic [21]. The characteristics of datasets include casual conversations, specifically in Tamil and Malayalam, film industries, and cinemas in south India. For example, “Rajapapanod pokan para ... njangalkku njangade stephen chettan undallo ” and “Ee pattinum dislike adicha aalkkarod anik puncham mathram ” these types of sentences are present in this dataset that shows richness in covering sarcasm in different forms, making it a more worthy resource for developing and filtering sarcasm identification algorithms in low resource languages. languages [16]. The overall statistics of the datasets can be seen in Table 1. This dataset is separated into training, testing and validation sets for the Malayalam and Tamil languages. The Validation data for Malayalam has 2826 samples with 2305 non-sarcastic labels and 521 sarcastic labels. There are 13188 samples in the training data, with 2499 sarcastic and 10689 non-sarcastic labels. In testing datasets there are no labels presents in this testing dataset there are 2826 rows in it [5]. In the validation dataset of Tamil there are 636 samples which have 4630 non-sarcastic labels and 1706 sarcastic labels. In training datasets there 29570 samples which have 7830 sarcastic labels and 21740 non-sarcastic labels. In case of testing datasets there are 6338 rows are presents without any labels. This section indicates that the model are fairly trained and evaluated for both the languages.

The GoogleTranslator¹ module from the deep translator package was used to translate the dataset. The translator was configured to convert the text into English automatically after identifying the source language, which may be either Tamil or Malayalam code-mixed text.

3.2. Translation Process

The dataset used in this research consists of Tamil-English and Malayalam-English code-mixed text, which needed to be translated into English for further analysis. To achieve this, we employed the

¹<https://translate.google.co.in/>

GoogleTranslator module from the deep translator package. This tool was selected for its ability to automatically detect the source language, whether Tamil or Malayalam, and convert the text into English. Given the nature of the dataset, which contains a wide range of casual conversations and social media posts, accurately capturing linguistic nuances, especially sarcasm, was crucial. The automatic translation process offered a fast and efficient way to process the large dataset, ensuring consistency in handling the initial translations. However, automatic translation tools are not without limitations, particularly when dealing with sarcasm, informal language, and code-mixed text. To address this, we implemented a secondary step in the translation process, where human validation was carried out by a team of five language experts. Each team member is proficient in both the source languages (Tamil and Malayalam) and English, allowing them to critically assess the quality and accuracy of the machine-generated translations. During this validation phase, the team thoroughly reviewed the translated text, comparing it to the original to identify any mistranslations, contextual inaccuracies, or issues where the automatic translator failed to capture the correct meaning. If any discrepancies were found, the team members manually corrected the translations to ensure they reflected the original intent, especially for detecting sarcasm, which often involves subtle or indirect cues. By combining machine translation with expert human validation, we ensured that the translations were both accurate and contextually appropriate, making the dataset highly reliable for developing and testing sarcasm identification algorithms in low-resource languages.

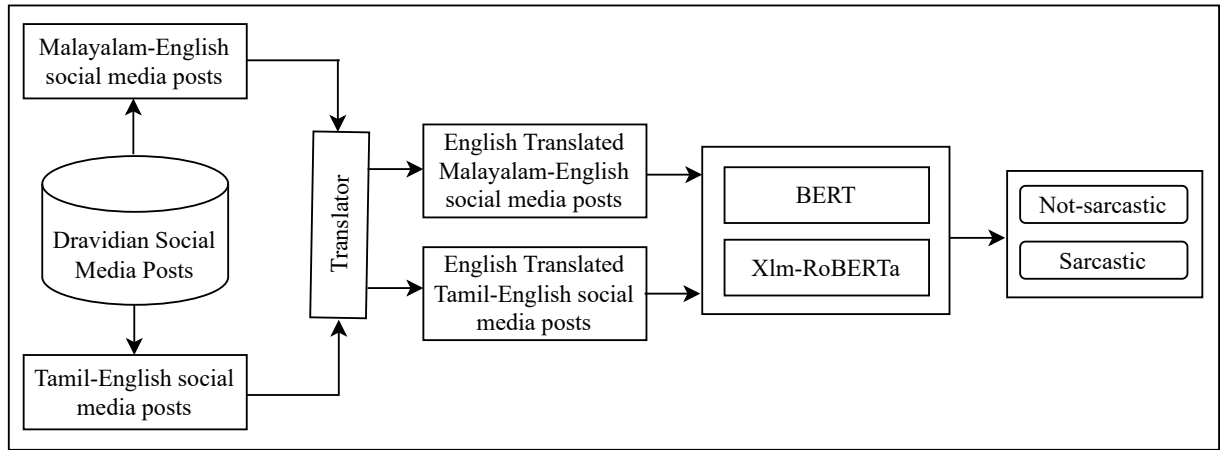


Figure 1: Overall flow diagram of the proposed model

3.3. Model Selection and Training

Two different deep learning-based models such as BERT and Xlm-RoBERTa were fine-tuned on the English translated Tamil-English and Malayalam-English to classify it into sarcastic and not-sarcastic classes.

- **BERT:** BERT (Bidirectional Encoder Representations from Transformers) [22] is a deep learning model designed by Google that has achieved state-of-the-art performance on a wide variety of Natural Language Processing (NLP) tasks, including text classification. BERT has several advantages: (i) Pre-trained contextualized embeddings: BERT can handle long-range dependencies and understands words in the context of the entire sentence; (ii) Transfer learning: Pre-training on a massive dataset means that fine-tuning on specific tasks requires significantly less data, (iii) State-of-the-art performance: BERT has achieved top results in many text classification benchmarks. Therefore, this work uses this model to fine-tuned with the translated Tamil-English and Malayalam-English social media posts.

- **XLM-RoBERTa:** XLM-RoBERTa is an extension of the original BERT (Bidirectional Encoder Representations from Transformers) and RoBERTa (Robustly Optimized BERT Pretraining Approach) models. It leverages the strengths of RoBERTa while focusing on multilingual understanding by being pre-trained on text in 100 languages. Unlike previous multilingual models like mBERT (Multilingual BERT), XLM-RoBERTa doesn't rely on language-specific tokens, making it a language-agnostic model. There are several advantages of using Xlm-RoBERTa model for text classification: (i) Multilingual support: XLM-R is specifically designed for multilingual tasks, making it highly suitable for text classification in multiple languages. (ii) Cross-lingual transfer: The model can be fine-tuned on one language and perform well on other languages, even with little or no training data for the target language. (iii) Handling of low-resource languages: XLM-R performs well even in languages with limited training data because of its extensive multilingual pre-training. (iv) Contextual understanding: Like BERT and RoBERTa, XLM-R understands words in the context of the entire sentence, providing rich, contextualized word embeddings for classification tasks. Due to the robustness of Xlm-RoBERTa model, this paper utilized it and fine-tuned it on the translated English social media posts.

We begin by preprocessing the data, where we initialize the Ktrain text transformer tailored to the chosen model. Each input text is capped at 30 tokens to ensure consistency. Both the text inputs and corresponding labels are transformed into the model's expected format for training. The selected pre-trained Transformer model is fine-tuned using the *Adam* optimizer, configured for classification tasks. The model is trained for 50 epochs with a learning rate of 5×10^{-5} and a batch size of 32.

4. Results

Table 2
Results for Tamil-English and Malayalam-English posts

Models	Class	Tamil-English			Malayalam-English		
		Precision	Recall	F1-score	Precision	Recall	F1-score
Bert-Base-Multilingual	Non-sarcastic	0.84	0.89	0.86	0.89	0.92	0.90
	Sarcastic	0.63	0.53	0.58	0.57	0.47	0.52
	Macro avg	0.73	0.71	0.72	0.73	0.70	0.71
	Weighted avg	0.78	0.79	0.78	0.83	0.84	0.83
Xlm-Roberta-Base	Non-sarcastic	0.83	0.88	0.86	0.90	0.92	0.91
	Sarcastic	0.63	0.53	0.57	0.60	0.51	0.55
	Macro avg	0.73	0.71	0.72	0.75	0.72	0.73
	Weighted avg	0.78	0.79	0.78	0.84	0.85	0.84

The performance of XLM-RoBERTa-Base and BERT-Base-Multilingual models fine-tuned on Tamil-English and Malayalam-English can be seen in Table 2. In terms of identifying non-sarcastic posts, both models performed admirably, with precision and F_1 -scores continuously surpassing 0.80 in both languages. Both XLM-RoBERTa-Base and BERT-Base-Multilingual achieved F_1 -scores of 0.86 for non-sarcastic postings in the Tamil-English dataset. However, both models had trouble distinguishing sarcasm; their F_1 -scores dropped to about 0.57 and 0.58, respectively, indicating how difficult it is to spot sarcastic content in code-mixed data. XLM-RoBERTa-Base fared marginally better than BERT-Base-Multilingual for the Malayalam-English dataset. This was especially true in the sarcastic class, where XLM-Roberta-Base obtained an F_1 -score of 0.55 whereas BERT achieved an F_1 -score of 0.52. This slight discrepancy implies that XLM-Roberta-Base is more appropriate for sarcasm detection in postings written in Malayalam and English. With a weighted average F_1 -score of 0.84 against 0.83 for BERT-Base-Multilingual, XLM-RoBERTa-Base outperformed the other model in terms of overall performance as measured by their weighted F_1 -scores. The confusion matrix and ROC curve of the

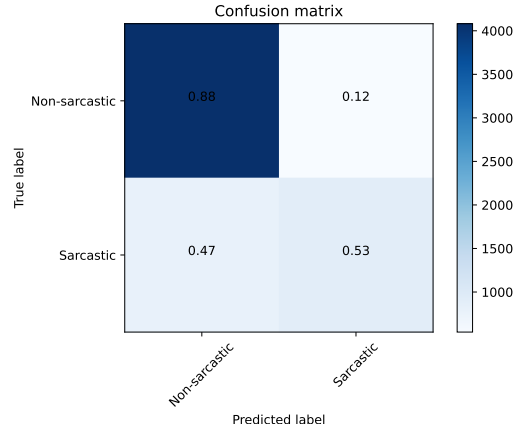


Figure 2: Confusion matrix for Tamil-English (Xml-RoBERTa)

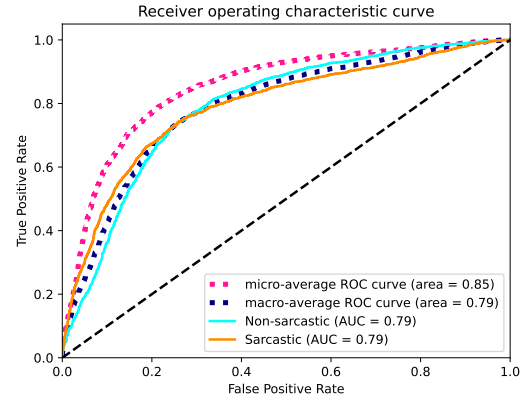


Figure 3: ROC Curve for Tamil-English (Xml-RoBERTa)

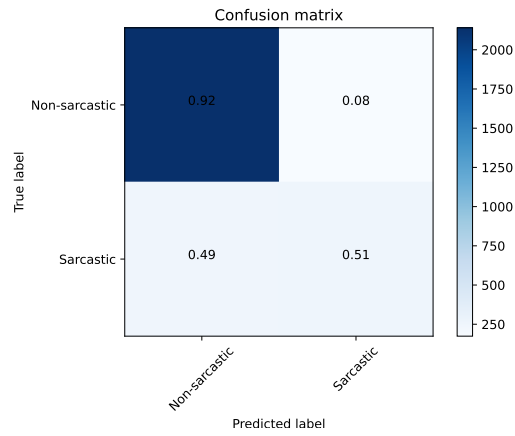


Figure 4: Confusion matrix for Malayalam-English (Xml-RoBERTa)

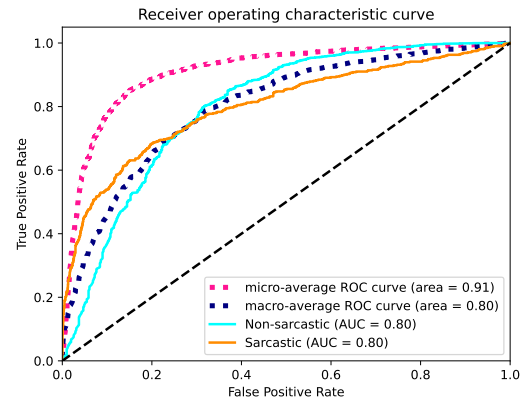


Figure 5: ROC Curve for Malayalam-English (Xml-RoBERTa)

Xlm-RoBERTa model for the Tamil-English language can be seen in Figures 2 and 3, respectively. Similarly, the confusion matrix and ROC curve for the Xlm-RoBERTa model for the Malayalam-English language can be seen in Figures 4 and 5, respectively.

5. Conclusion

In this work, we assessed two cutting-edge multilingual transformer models for sarcasm detection using code-mixed datasets in Tamil-English and Malayalam-English: BERT-Base-Multilingual and XLM-Roberta-Base. With F_1 -scores continuously above 0.80, both models showed good performance in identifying non-sarcastic text in both languages. The models' difficulties with the sarcastic class, on the other hand, were evident in their lower F_1 -scores (0.52 to 0.58), which underscores how difficult it is to detect sarcasm in code-mixed data. For the Tamil-English dataset, both models fared similarly. However, XLM-Roberta-Base did slightly better, especially for the Malayalam-English dataset, where it beat Bert-Base-Multilingual in both the non-sarcastic and sarcastic classes. In comparison to Bert-Base-Multilingual, which obtained a weighted average F_1 -score of 0.83 for Malayalam-English, the XLM-Roberta-Base model demonstrated a stronger overall capacity to handle code-mixed data. These findings imply that although transformer models are helpful in identifying non-sarcastic content, more effort is necessary to enhance the detection of sarcasm, especially in low-resource, code-mixed languages. Future work may investigate better identifying sarcasm in such challenging datasets by integrating more contextual features such as multimodal data or sophisticated fine-tuning methods.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

- [1] R. Pandey, A. Kumar, J. P. Singh, S. Tripathi, Hybrid attention-based long short-term memory network for sarcasm identification, *Applied Soft Computing* 106 (2021) 107348.
- [2] R. Pandey, A. Kumar, J. P. Singh, S. Tripathi, A hybrid convolutional neural network for sarcasm detection from multilingual social media posts, *Multimedia Tools and Applications* (2024) 1–29.
- [3] B. R. Chakravarthi, Hope speech detection in youtube comments, *Social Network Analysis and Mining* 12 (2022) 75.
- [4] Y. Wang, X. Xu, Z. Gao, X. Shi, A prior probability of speaker information and emojis embedding approach to sarcasm detection, in: *2021 IEEE International Conference on Engineering, Technology & Education (TALE)*, IEEE, 2021, pp. 1033–1038.
- [5] B. R. Chakravarthi, N. Sripriya, B. Bharathi, K. Nandhini, S. Chinnaudayar Navaneethakrishnan, T. Durairaj, R. Ponnusamy, P. K. Kumaresan, K. K. Ponnusamy, C. Rajkumar, Overview of the shared task on sarcasm identification of Dravidian languages (Malayalam and Tamil) in DravidianCodeMix, in: *Forum of Information Retrieval and Evaluation FIRE - 2023*, 2023.
- [6] B. R. Chakravarthi, N. Sripriya, B. Bharathi, K. Nandhini, S. C. Navaneethakrishnan, T. Durairaj, R. Ponnusamy, P. K. Kumaresan, K. K. Ponnusamy, C. Rajkumar, Overview of the shared task on sarcasm identification of dravidian languages (malayalam and tamil) in dravidiancodemix, in: *Forum of Information Retrieval and Evaluation FIRE-2023*, 2023.
- [7] S. Saumya, A. Kumar, J. P. Singh, Filtering offensive language from multilingual social media contents: A deep learning approach, *Engineering Applications of Artificial Intelligence* 133 (2024) 108159.
- [8] J. Kumari, A. Kumar, Ja-nlp@ It-edi-2023: Empowering mental health assessment: A roberta-based approach for depression detection, in: *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, 2023, pp. 89–96.
- [9] A. Kumar, J. Kumari, J. Pradhan, Explainable deep learning for mental health detection from english and arabic social media posts, *ACM Transactions on Asian and Low-Resource Language Information Processing* (2023).
- [10] A. Kumar, S. Saumya, A. Singh, Detecting dravidian offensive posts in miot: A hybrid deep learning framework, *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* (2023). URL: <https://doi.org/10.1145/3592602>. doi:10.1145/3592602, just Accepted.
- [11] T. S. Apon, R. Anan, E. A. Modhu, A. Suter, I. J. Sneha, M. G. R. Alam, Banglasarc: A dataset for sarcasm detection, in: *2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, IEEE, 2022, pp. 1–5.
- [12] P. K. Roy, An advanced learning approach for detecting sarcasm in social media posts: Theory and solutions, *Social Science Quarterly* 105 (2024) 1857–1874.
- [13] J. A. Josephine, S. K. Maharana, M. A. A. Walid, T. Thulasimani, M. S. Alam, M. Tiwari, Hybrid particle swarm optimization with deep learning driven sarcasm detection on social media, in: *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*, IEEE, 2022, pp. 687–693.
- [14] R. Ratnavel, R. G. Joshua, S. Varsini, M. A. Kumar, Sarcasm detection in tamil code-mixed data using transformers, in: *International Conference on Speech and Language Technologies for Low-resource Languages*, Springer, 2023, pp. 430–442.
- [15] R. Sivakumar, J. M. C, B. M. Jenefer, Identifying the type of sarcasm in dravidian languages using deep-learning models., in: *FIRE (Working Notes)*, 2023, pp. 278–286.
- [16] N. Sripriya, T. Durairaj, K. Nandhini, B. Bharathi, K. K. Ponnusamy, C. Rajkumar, P. K. Kumaresan,

- R. Ponnusamy, C. Subalalitha, B. R. Chakravarthi, Findings of shared task on sarcasm identification in code-mixed dravidian languages, FIRE 2023 16 (2023) 22.
- [17] A. Bhat, A. Chauhan, A deep learning based approach for multimodal sarcasm detection, in: 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), IEEE, 2022, pp. 2523–2528.
- [18] A. Kumar, V. T. Narapareddy, V. A. Srikanth, A. Malapati, L. B. M. Neti, Sarcasm detection using multi-head attention based bidirectional lstm, Ieee Access 8 (2020) 6388–6397.
- [19] M. S. Razali, A. A. Halin, N. M. Norowi, S. C. Doraisamy, The importance of multimodality in sarcasm detection for sentiment analysis, in: 2017 IEEE 15th Student Conference on Research and Development (SCOReD), IEEE, 2017, pp. 56–60.
- [20] M. Bedi, S. Kumar, M. S. Akhtar, T. Chakraborty, Multi-modal sarcasm detection and humor classification in code-mixed conversations, IEEE Transactions on Affective Computing 14 (2021) 1363–1375.
- [21] B. R. Chakravarthi, S. N. B. B. N. K. T. Durairaj, R. Ponnusamy, P. K. Kumaresan, K. K. Ponnusamy, C. Rajkumar, Overview of sarcasm identification of dravidian languages in DravidianCodeMix@FIRE-2024, in: Forum of Information Retrieval and Evaluation FIRE - 2024, DAIICT, Gandhinagar, 2024.
- [22] A. Vaswani, Attention is all you need, Advances in Neural Information Processing Systems (2017).