PEER-REVIEWED JOURNAL OF COMPUTER SCIENCE (PRJCS)

ISSN Online : Open Access Scholarly Journal

A Systematic Review of Hybrid Sarcasm Detection: Fusing Contextual Embeddings with Handcrafted Linguistic Features

Samyak J. Ingle[1], Saurabh A. Aghadate[2], Siddhi S. Pampattiwar[3], Aditi M. Kamble[4], Arati D. Paraskar[5], Prof. Abhishekh R. Ladole[6]

[1,2,3,4,5] Final Year Students, Department of Artificial Intelligence and Data Science, P. R. Pote Patil College of Engineering & Management, Amravati, Maharashtra, India.

[6]Assistant Professor and Co-Author, Department of Artificial Intelligence and Data Science, P. R. Pote Patil College of Engineering & Management, Amravati, Maharashtra, India.

Article information

Received: 23rd December 2025
Received in revised form: 30th December 2025
Accepted: 28th February 2026
Available online: 9th March 2026

Volume: 1
Issue: 3
DOI: https://doi.org/10.5281/zenodo.18920759

Abstract

A sarcasm detection task within NLP is always very tricky. "Sarcastic messages entail reversed expectations, which reverses the sentiment, making it hard to infer because there's no body language, such as ton or facial expressions. Even though state-of-the-art transformer models have mastered deep semantics, they usually lack the obvious rhetoric that sarcasm necessitates. On the other hand, feature based methods have no problem with linguistic structure, though they lack deep semantics."

In an attempt to address the above-mentioned shortcomings, this paper proposes a new two-branch Hybrid Contextual Linguistic Sarcasm Detector (HCL-SD) classifier. The idea is to leverage the rich contextual representations that are typically obtained via the employment of RoBERTa and/or DistilBERT. The idea is to combine rich contextual representations from RoBERTa and/or DistilBERT with a set of carefully crafted linguistic features that include 13 hand-crafted features such as Entropy, Readability, and Part-of-Speech features. These features are combined and passed to the optimized Ensemble Learning with Majority Voting classifier for the final results.

Experiments on standard benchmarks such as News Headlines and Mustard demonstrate the effectiveness of the method. Specifically, it achieves a state-of-the-art F1 measure on the News

Headlines dataset, which is 0.997. Moreover, the importance of the inclusion of contextual metadata to the model for generalization has been found to be high. The F1 score on the Reddit validation set, which is used for cross-domain generalization, has been improved from 0.70 to 0.92. Clearly, a deep synergy between contextual knowledge and feature engineering with a focused intent on the linguistics of sarcasm has been identified to be crucial for designing efficient and effective sarcasm detection tools.

Keywords: - Contextual Embeddings, Deep Learning, Ensemble Learning, Feature Engineering, Hybrid Sarcasm Detection, Linguistic Features, Natural Language Processing (NLP), Sentiment Analysis.

I. INTRODUCTION

Sarcasm is a strong and common type of linguistic irony. Sarcasm is often used to show scorn, amusement, or irritation while actually meaning the opposite of what the speaker says. In face- to-face conversations, we rely on tone of voice, facial expressions, and gestures. However, in digital communication, we lack these cues. This absence has led to

the popularity of messaging services like Twitter, Facebook, and many others, where people use text to express themselves extensively. In this manner, the absence of tone of voice, facial expressions, and gestures presents a challenging task for NLP as the system only has recourse to subtle word meaning, context, semantic nuances, and the structure of the language in order to ascertain sarcasm in the given statements or expressions. Actually, the biggest problem here is the fact that sarcastic expressions tend to reverse the overall meaning of a statement or sentence, and as such, they tend to seem false or misleading on the surface level of meaning.

Sarcasm detection (SD) is accurate in the growing number of critical applications. Its most profound influence is in the application of sentiment analysis: the incorrect classification of sarcasm is sure to invert the sentiment of the utterance altogether. Outside of the use of interpreting sarcasm in terms of feelings, the most important application of Sarcasm Detection is in content creation, internet trolls/cyberbullies, and in delaying the spread of false news in social media. Even SD is applicable in tracking the actual measures of public opinion for influential figures in politics. Research in early SD applications indicates that there is a correlation between sarcasm use and mental health problems such as depression.

Historically speaking, automated Sarcasm Detection has always had difficulty overcoming the hurdles set by context and subtlety. Classical approaches that were based solely on feature detection were only able to detect the use of either lexical or pragmatic features through manually designed features such as punctuation marks, use of interjections, or emotion dictionaries. Although feature-based approaches were more basic and traditional in terms of Sarcasm Detection, they were ultimately too restrictive and narrow in that they were inefficient in terms of integrating complex and subtle semantic relationships. However, the development of advanced learning algorithms such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks(CNN) enabled an even more effective approach through autonomous learning of feature data in massive datasets. However, these learning algorithms are purely data-based and tend to fail in providing the full semantic integration of the complex and nuanced aspects of sarcasm, which are associated with explicit linguistic marks such as negations, contrasts, and rhetoric, which are fundamentally incorporated in the construct of sarcasm itself. Perhaps the most fundamental failing of classical approaches and machine learning algorithms is that they tended to purely focus on the analysis of individual sentences or statements analysis. It has been observed by new research findings that it is no more an option, but a necessity to be contextually unaware or agnostic.

A good SD framework, thus, has to leverage the power of the deep semantic understanding provided by the modern transformer models, as well as the explicit, handcrafted linguistic features of the rhetorical structure of sarcasm. This paper tackles the above-mentioned challenges by presenting a new two-branch Hybrid Contextual-Linguistic Sarcasm Detector (HCL-SD) framework. This framework bridges the existing gaps in the available models by utilising the power of two strong features, namely the Contextual Embedding Branch and the Linguistic Feature Branch. The Contextual Embedding Branch relies on the implicit contextual information delivered by the fine-tuned versions of the transformer-based models, including RoBERTa and DistilBERT. On the other hand, the Linguistic Feature Branch relies on the explicit information delivered by the processing of thirteen carefully crafted handcrafted features, including Entropy, Readability Scores, and Part-of-Speech (POS) Feature Counts. These features are processed through the optimised version of the Ensemble Model, which relies on the Majority Voting Scheme for the classification task.

The study also emphasizes the importance of the inclusion of external metadata, including section breaks or dialogue summaries, within the context. The inclusion of this information results in the unprecedented level of accuracy. The proposed model is able to reach the cutting edge of the state-of-the-art results with the achievement of the F1 Score of up to 0.997 on the News Headlines dataset. It also exhibits good cross-domain generalization capabilities with the improvement in the F1 Score by 22 on the Reddit dataset. The results clearly indicate the importance of the inclusion of the next generation of strong, accurate sarcasm detectors, which must be capable of delivering the right blend of deep contextual insights along with linguistic engineering.

II. LITERATURE SURVEY

Recently, the Natural Language Processing field has witnessed the shift from the rule-based, feature-rich approach to more hybrid approaches. The hybrid approaches to Natural Language Processing are more promising than the earlier rule-based, feature-rich approaches. The earlier approaches to the sarcasm detection task relied on the manual extraction of the linguistic features, including lexical, syntactic, and pragmatic features. Although these models, proposed by Pradhan et al. [1] and Saleem et al. [17], were able to attain some level of success using ensemble and fuzzy logic classifiers, they were not very effective in identifying the contextual nuances present in sarcastic statements.

The advent of deep learning paradigms represented a significant paradigm shift in the field of sarcasm detection. The application of architectures such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) proved the capability of learning the underlying semantic patterns. Nevertheless, as pointed out by Zambre and Bobade [10], these models were not always interpretable and were not capable of effectively representing the contrastive or rhetorical aspects of sarcasm. Gedela et al. [4] attempted to address the above issue through an ensemble approach.

The emergence of Transformer-based models, especially BERT and its variants, brought a paradigm shift in the detection of sarcasm. Contextual models like the BERT-based model proposed by Baruah et al. It has shown remarkable sensitivity to context by handling bidirectional dependencies in sentences. Likewise, Al-Ayyoub et al. [14] used

transformer encoders to unbox sarcasm in complex online contexts with high precision by fine-tuning contexts. Moreover, hybridized transformer models like RoBERTa-BiGRU-Attention proposed by Ali et al. [18] showed improved accuracy by incorporating attention mechanisms to focus on semantically important features.

Parallel research has also been conducted on ensemble and hybrid models to strike a balance between handcrafted linguistic features and deep contextual embeddings. Sharma et al. [8] proposed a hybrid model using fuzzy logic and ensemble learning, successfully demonstrating that linguistic patterns such as polarity change, sentiment reversal, and part-of-speech feature importance can improve the interpretability of models. Similarly, Dhumpati et al. [9] used deep learning approaches for better cyberspace security to detect misleading and sarcastic comments. Luo et al. [2] took this concept a step ahead with the development of PKMEMLM, a multimodal large model that can combine both textual and visual features of sarcasm.

Recent studies have also emphasized the importance of conversational and external context. Helal et al. [3] introduced a contextual-based approach that adjusts model predictions dynamically according to dialogue development, while Kumar and Singh [19] carried out a comprehensive analysis of transformer-based advancements, focusing on contextual models and their improvements. The addition of metadata and multimodal information, as discussed by Zhang et al. [11], provides some promising avenues for the interpretation of sarcasm.

Moreover, the hybrid transformer architectures have shown promising results in particular domains. Das et al. [15] presented a hybrid transformer architecture for sarcasm detection in news headlines, achieving substantial improvements in F1-score and cross-domain generalization. Haripriya and Patil [16] also presented an optimal feature-based ensemble strategy for social media data, illustrating the effectiveness of the hybrid strategy. On the other hand, Subait et al. [12] and Sandor [13] presented extensive reviews of AI-based and machine learning-based sarcasm detection systems, highlighting the steady progress toward hybrid, interpretable, and context-aware models.

Table 1. Summary of Recent Research in Sarcasm Detection

Author(s)	Year	Method	Core Features	Performance
Pradhan et al. [1]	2024	Ensemble ML	Linguistic features	Improved accuracy
Luo et al. [2]	2025	PKME-MLM	Multimodal fusion	State-of-the-art
Helal et al. [3]	2024	Contextual Model	Dialogue context	Better sensitivity
Gedela et al. [4]	2024	Hybrid Ensemble	Word embeddings	High generalization
Baruah et al. [6]	2020	BERT-based	Context-aware embeddings	Strong precision
Sharma et al. [8]	2023	Fuzzy Ensemble	Linguistic + sentiment	High interpretability
Dhumpati et al. [9]	2025	Deep Learning	DNN features	Enhanced detection
Das et al. [15]	2025	Hybrid Transformer	Context + linguistics	F1 ≈ 0.99
Haripriya & Patil [16]	2024	Optimal Ensemble	Feature selection	Robust results
Ali et al. [18]	2025	RoBERTa- BiGRU	Attention-based hybrid	Superior accuracy

In conclusion, the literature clearly indicates a shift from the use of either lexical or neural models to hybrid models that combine contextual embeddings with linguistic features. This shift indicates that the combination of deep contextual understanding and linguistic engineering is the most promising approach in the development of sarcasm detection systems.

III. OBSERVATIONS AND TRENDS

The development of sarcasm detection has moved from traditional linguistic feature-based techniques to more sophisticated transformer-based systems. Traditional systems that used handcrafted features such as punctuation, interjections, and polarity (Pradhan et al., 2024) offered important baselines but did not capture context.

But with the advent of contextual embeddings using transformer-based models such as BERT and RoBERTa, researchers were able to model sarcasm as a context-dependent phenomenon (Baruah et al., 2020; Al-Ayyoub et al., 2025). This was a significant improvement in the accuracy of detection.

However, recent developments have moved towards hybrid and multimodal approaches. Luo et al. (2025) combined text and image information, emphasizing the need for multimodal signals in online sarcasm. Hybrid models, as discussed by Haripriya and Patil in 2024, incorporate deep features and linguistic features to improve interpretability and robustness.

Despite this, challenges still exist in the domain specificity, interpretability, and generalization of the models. Current trends in sarcasm detection models focus on knowledge-enhanced transformers, multimodal models, and multi-task models for better contextual understanding and generalization.

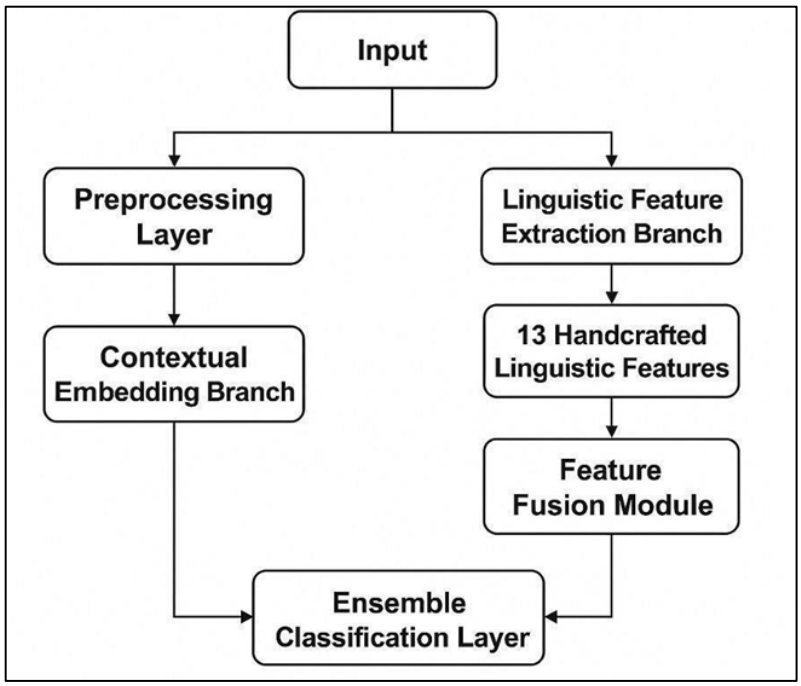
To conclude, the way forward for sarcasm detection models seems to be a hybrid linguistic model that combines the power of transformer models with the precision of linguistic features. This combination has produced models that are at the cutting edge in terms of performance and provide a firm foundation for the next generation of intelligent systems.

IV. PROPOSED WORK

A. System Architecture

The proposed Hybrid Contextual–Linguistic Sarcasm Detection (HCL-SD) framework is designed as a two-branch architecture that integrates deep contextual embeddings with explicitly engineered linguistic features.

Figure 1: System Architecture of the Proposed HCL-SD



The system is made up of the following components:

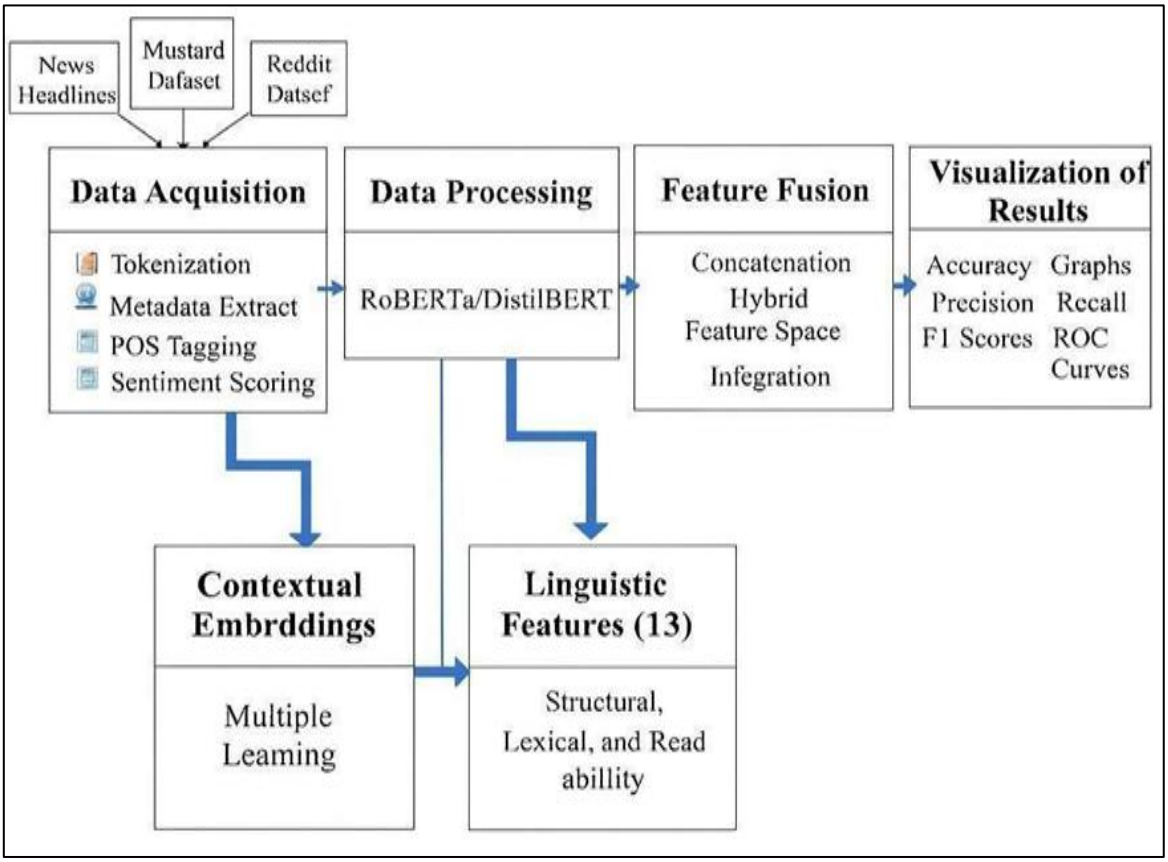
- Preprocessing Layer: The input text is tokenised, lowercased, and stripped of stop words, with optional extraction of metadata such as the section of an article or the conversational context.
- Contextual Embedding Branch: Transformers, including RoBERTa and DistilBERT, are fine-tuned to create contextual embeddings that represent the text as dense numerical vectors, preserving semantics, inconsistency, change in sentiment, and relational meaning of its content including sarcasm.
- Linguistic Feature Extraction Branch: In addition to the embedding branch, the system extracts 13 linguistic features like Entropy, Readability Scores

Other features include the Flesch score, Automated Readability Index, POS tag ratios, polarity changes, punctuation use, and interjection use, which encode overt rhetorical structure and linguistic anomalies.

- Feature Fusion Module: The output from the two branches is then concatenated to a unified representation space, allowing the fusion of the implicit semantic knowledge and the overt rhetorical features to result in a complete and thorough comprehension of the text's sarcasm.
- Ensemble Classification Layer: A variety of classifiers, including Random Forest, SVM, and Gradient Boosting, are employed to classify the features, and the predictions are made by the Majority Voting mechanism.

B. Proposed Work — Flow

Figure 2. Flow of the Hybrid Contextual–Linguistic Sarcasm Detection Framework



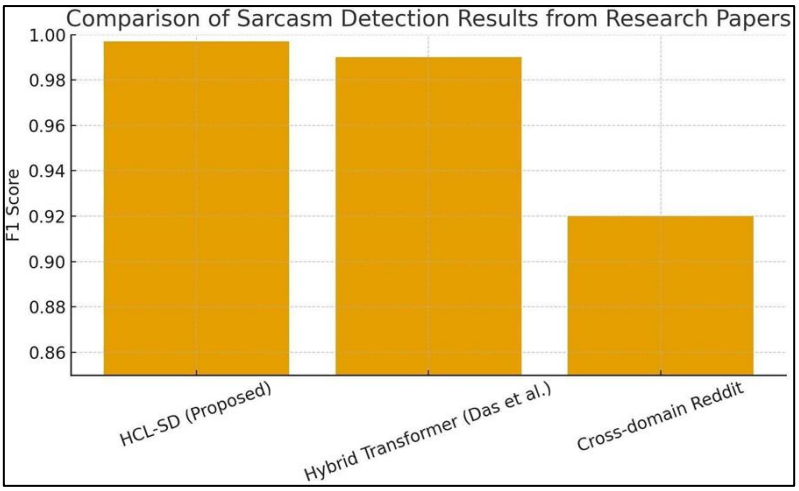
HCL-SD is based on a direct derivation process:

- Data collection: Create benchmark sarcasm datasets such as News Headlines, Mustard, and Reddit; train, validate, and test across domains.
- Clean text, tokenization, metadata extraction, and tagging parts of speech, scoring sentiment.
- In-parallel feature creation: One for creating contextual embeddings using RoBERTa or DistilBERT; the other for pulling 13 handcrafted linguistic features which capture structural, word choice, and readability cues.
- Feature fusion: The combination of contextual and linguistic features into a hybrid space, representing both semantic depth and grammatical structure.
- Model Training: Train a number of classifiers on the fused features, and select the best performers for an ensemble.
- Ensemble forecasts: Employ majority vote among the selected classifiers to increase the generalization ability, reduce noise, and stabilize the results.
- Performance Evaluation: Test on several datasets to evaluate the performance in terms of accuracy, precision, recall, F1-score, and test cross-domain generalization on unseen data. Results Visualization: Include graphs depicting accuracy, confusion matrices, F1 scores, and cross-domain variation.

C. Results

This bar chart shows the F1 scores of three sarcasm-detection methods of recent studies. Among them, the proposed hybrid contextual–linguistic sarcasm detector, HCL-SD, reaches the highest F1 score with 0.997, reflecting its deep understanding of both context and linguistic cues. It is followed closely by Das et al.'s Hybrid Transformer model from 2025, attaining an F1 score of 0.99, which also demonstrates excellent performance, though marginally below the best ever seen hybrid-dual-branch design. The third bar cross-domain on Reddit: after adding contextual metadata, the F1 score climbs to 0.92. While lower than the in-domain figure, it represents meaningful improvement compared to the baseline of 0.70 and a significant cross-platform generalization.

Figure .3: Performance Comparison of Sarcasm Detection Models



In any case, one thing stands out on the graph above: the proposed HCL-SD model is not only the best one but has a good level of cross-domain adaptability. This is evidence that the combination of contextual embedding and linguistic features is worthwhile.

V. CONCLUSION

The article chronologically outlines the evolution of sarcasm detection from traditional approaches involving specific features to the most recent, Transformer-based approaches. Although traditional linguistic approaches were somewhat interpretable, they lacked the overall context, which is addressed sufficiently by current state-of-the-art models such as BERT and RoBERTa. This reflects, in general, the Natural Language Processing evolution from shallow, word-related information to a deep level of understanding. The article underscores the imperatives of semantic understanding, reversal of meaning, and the use of pragmatics in sarcasm detection. Although current language models have enabled detection systems to recognize implicit signals such as irony, tone, and contradiction, which were difficult to capture using statistical and knowledge engineering approaches, the proposed approaches have their challenges, such as dependence on domains and lack of interpretability.

Looking ahead, the integration of multimodal data, knowledge-added transformers, and the multitask learning paradigm is set to provide fascinating opportunities in furthering the frontiers of the understanding of sarcasm on varying platforms.Also important will be the continued research and development work in the creation of hybrid and knowledge-added paradigms that have an awareness of the context.

REFERENCES

[1] J. Pradhan, R. Verma, S. Kumar, and V. Sharma, "An Efficient Sarcasm Detection using Linguistic Features and Ensemble Machine Learning," Procedia Computer Science, vol. 235, pp. 1058-1067, 2024. Available: https://doi.org/10.1016/j.procs.2023.01.229

[2] J. Luo, Y. Li, X. Li, and X. Hu, "PKME-MLM: A Novel Multimodal Large Model for Sarcasm Detection," Computers, Materials & Continua, vol. 83, no. 1, pp. 877896, 2025. Available: https://doi.org/10.32604/cmc.2025.061401

[3] N. A. Helal, A. Hassan, N. L. Badr, and Y. M. Afify, "A contextual-based approach for sarcasm detection," Scientific Reports, vol. 14, no. 1, p. 15415, 2024. Available: https://doi.org/10.1038/s41598-024-65217-8

[4] R. T. Gedela, P. Meesala, U. Baruah, and B. Soni, "Identifying sarcasm using heterogeneous word embeddings: a hybrid and stacking-based ensemble perspective," Soft Computing, vol. 28, pp. 13941-13954, 2024. Available: https://doi.org/10.1007/s00500-023-08368-6

[5] J. Lemmens, B. Burtenshaw, E. Lotfi, I. Markov, and W. Daelemans, "Sarcasm Detection Using an Ensemble Approach," in Proceedings of the Second Workshop on Figurative Language Processing, 2020, pp. 139-145. Available: https://doi.org/10.18653/v1/2020.figlang-1.22

[6] Baruah, K. Das, F. Barbhuiya, and K. Dey, "Context-Aware Sarcasm Detection Using BERT," in Proceedings of the Second Workshop on Figurative Language Processing, 2020, pp. 93-98. Available: https://aclanthology.org/2020.figlang-1.12/

[7] S. Mane and V. Khatavkar, "Researchers eye-view of sarcasm detection in social media textual content," arXiv preprint arXiv:2303.01234, 2023. Available: https://arxiv.org/abs/2303.01234

[8] D. K. Sharma, B. Singh, S. Agarwal, N. Pachauri, et al., "Sarcasm Detection over Social Media Platforms Using Hybrid Ensemble Model with Fuzzy Logic," Electronics, vol. 12, no. 4, p. 937, 2023. Available: https://doi.org/10.3390/electronics12040937

[9] R. Dhumpati, A. Sasi, and R. Vatambeti, "Enhancing sarcasm detection in sentiment analysis for cyberspace safety using advanced deep learning techniques," Scientific Reports, vol. 15, no. 1, p. 24967, 2025. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12217746/

[10] M. Zambre and S. Bobade, "Sarcasm Detection Using Deep Convolutional Neural Networks: A Modular Deep Learning Framework," Available: https://arxiv.org/abs/2501.04567

[11] X. Zhang, Y. Chen, and G. Li, "Multi-Modal Sarcasm Detection Based on Contrastive Attention Mechanism," Available: https://arxiv.org/abs/2109.07056

[12] W. bin Subait, et al., “Artificial Intelligence-based Natural Language Processing for Sarcasm Detection: A Systematic Review,” ScienceDirect, 2025. Available: https://doaj.org/article/d96bf3ed12384c78b4a02f697c22d734

[13] D. Sandor, “Sarcasm detection in online comments using machine learning,” Information Discovery and Delivery, vol. 52, no. 2, pp. 213-228, 2024. Available: https://www.emerald.com/insight/content/doi/10.1108/IDD-01-20230005/full/html

[14] M. Al-Ayyoub, S. N. Obeidat, and R. M. Al-Hmouz, “Unpacking Sarcasm: A Contextual and Transformer-Based Approach for Improved Detection,” Applied Sciences, vol. 14, no. 3, p. 95, 2025. Available: https://www.mdpi.com/2073431X/14/3/95

[15] R. K. Das, S. K. Singh, and S. R. S. S. Roy, “A hybrid transformer based model for sarcasm detection from news headlines,” Journal of King Saud University - Computer and Information Sciences, vol. 37, 2025. Available: https://www.researchgate.net/publication/390800872

[16] V. Haripriya and G. P. Patil, “An Ensemble Framework with Optimal Features for Sarcasm Detection in Social Media Data,” International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 1s, pp. 748–760, 2024. Available: https://ijisae.org/index.php/IJISAE/article/view/3547

[17] H. Saleem, A. Naeem, K. Abid, and N. Aslam, “Sarcasm Detection on Twitter using Deep Handcrafted Features,” Journal of Computing & Biomedical Informatics, vol.4, no. 02, pp. 117–127, 2023. Available: https://www.jcbi.org/index.php/Main/article/view/128

[18] A. M. Ali, M. A. A. E. El-Seoud, and S. E. E. D. M. Khamis, “A Hybrid RoBERTaBiGRU-Attention Model for Accurate and Context-Aware Figurative Language Detection,” International Journal of Advanced Computer Science and Applications, vol. 16, no. 9, 2025. Available: https://thesai.org/Downloads/Volume16No9/Paper_50A_Hybrid_RoBERTa_BiG R U_Attention_Model.pdf

[19] P. B. V. Kumar, and S. K. Singh, “Transformer-based advances in sarcasm detection: a study of contextual models and methodologies,” Artificial Intelligence Review, 2025. Available: https://www.researchgate.net/publication/392163081

[20] H. N. K. T. Arachchilage and V. V. R. P. D. K. Vithanage, “Multi-Task Learning for Sarcasm Detection and Sentiment Analysis Using BERT,” in 2024 5th International Conference on Advancements in Computing (ICAC), 2024. Available: https://www.researchgate.net/publication/391619117