2024 27th International Conference on Computer and Information Technology (ICCIT)
20-22 December 2024, Cox's Bazar, Bangladesh

Enhancing Bilingual Sarcasm Detection with an Advanced Ensemble Model: A Dual-Language Deep Learning Framework

Syed Nakibul Islam
Dept. of Electronics and Telecommunication Engineering
Chittagong University of Engineering and Technology
Chattogram, Bangladesh
s.nakib30@gmail.com

Fahad Siddique Faisal
Dept. of Electronics and Telecommunication Engineering
Chittagong University of Engineering and Technology
Chattogram, Bangladesh
fahadsid1770@gmail.com

Md. Farhad Hossain
Dept. of Electronics and Telecommunication Engineering
Chittagong University of Engineering and Technology
Chattogram, Bangladesh
farhad.hossain@cuet.ac.bd

Md. Azad Hossain
Dept. of Electronics and Telecommunication Engineering
Chittagong University of Engineering and Technology
Chattogram, Bangladesh
azad@cuet.ac.bd

***Abstract*—Sarcasm is a type of verbal irony in which the literal expression of a statement contradicts its intended meaning. Sarcasm detection is crucial for interpreting web portal comments, customer feedback, and social media interactions, as misinterpretation can distort sentiment analysis. It improves the ability of automated systems to process opinions, reviews, and dialogues in web series or ordinary communication, thereby ensuring a more nuanced comprehension of user intent. This study has proposed a technique for Bilingual Sarcasm Detection that has used an ensemble model, integrating Both BiLSTM and GRU to detect sarcasm from both English and Bengali text. This study provides valuable insights into the effectiveness of Bi-LSTM and GRU in detecting subtle and context-dependent sarcastic expressions in both Bengali and English Languages. The proposed ensemble model utilizing word embeddings demonstrated superior performance compared to the other deep learning models achieving approximately 97% accuracy on the news headline dataset and 89% on the BanglaSARC dataset.**

***Index Terms*—Sarcasm Detection, Bi-LSTM, GRU, GloVe, Embedding**

I. Introduction

In the ever-evolving landscape of natural language processing, the detection of sarcasm has emerged as a critical challenge due to its nuanced and often context-dependent nature. Detecting sarcasm is a challenging endeavor for machines. As each participant in a live conversation between two individuals can immediately identify the context and facial expressions conveying sarcasm, it gets easy for humans to detect sarcasm. However, machine learning algorithms face significant challenges in discerning sarcasm in text due to the absence of internal context and in text expressions. This work delves into the realm of sarcasm detection for both English and Bengali language using Bidirectional long short-term memory (Bi-LSTM) networks, a class of recurrent neural networks known for their sequential data processing capabilities and GRU, another class of recurrent neural networks, which has competence in capturing sequential dependencies efficiently in data while mitigating the vanishing gradient problem. In this research, GloVe pretrained word embedding is used for English text and Bangla Word2vec is used for bangla text. Here, the proposed model is implemented into two languages, Bengali and English. The performance score of the proposed model for both language outperforms the existing model, and overcome the limitation of rule based AI and other machine learning model. As, conventional machine learning algorithm often not suitable for large categorical dataset and unable to detect contextual information as well as complex linguistic pattern, this ensemble model can overcome this problem and easily detect contextual information.

II. Related Works

A. Machine Learning based Approach :

For sarcasm detection,in past years many approaches have been introduced by using machine learning. A research has been conducted which concluded that SVM(support vector machine) is the most common method which prove the high accuracy in detecting the sarcasm [1]. A novel approach for sarcasm detection is pattern based approach where four sets of feature is proposed and detect sarcasm from the text. This methodology uses several classifier as SVM, KNN and random forest [2], [3].

B. Deep Learning based Approach :

This section provides information about the work and research previously done about sarcasm detection from the text . The overview of that research and the methodology in different works are discussed in this section. Different

type of methodology is used in this field. Machine learning and Deep learning approaches have been used for different cases. However deep learning is the most common approach that is used in this field. In the deep learning, The LSTM has been used in research to categorize sarcasm using a context-based method. LSTM uses paragraph2vec to extract features that make sarcasm detection simple [4]. A hybrid model is proposed that integrates both LSTM and CNN techniques for sarcasm detection [5]. Genetic optimization is employed in the LSTM to optimize the hyper parameters derived from the LSTM. Besides, A more sophisticated and enhanced form of LSTM has been presented in the suggested methodology for sarcasm identification. This has improved LSTM which combines the global max pooling layer with LSTM to achieve better results [6]. For the detection of sarcasm in deep learning, a hybrid method is constructed multiple times using LSTM, CNN, and RNN models [7]–[9]. A2Text-Net which is a novel approach in the classification system for sarcasm that introduces three distinct layers and yields superior results in comparison to conventional approaches [10]. Another model integrating a bidirectional LSTM variational auto-encoder that demonstrates utility when applied to an unbalanced dataset. Some of the research in this field take place by comparing the outcomes of different methodology in the deep learning approach [11]. There has been a considerable impact on the research field as a result of the comparison of outcomes about numerous different deep learning hybrid approaches [12], [13].
The proposed methodology has used two distinct networks to make predictions, and eventually the final prediction is obtained through exhaustive analysis by integrating the two separate predictions made by the networks. This is a key difference between the proposed methodology and the current approaches, which use one network and two networks in sequentially.

III. Proposed Methodology

In this work, a comprehensive ensemble model has been proposed to detect sarcasm from the textual data. This ensemble framework combines Bidirectional Long Short-Term Memory (BiLSTM) and Gated Recurrent Unit (GRU) models. This model commences with segregation and extensive preprocessing of two distinct sections of dataset which ensures that the data undergoes a rigorous cleansing and tokenization process. Raw input has been efficiently converted into vector token through this process.

The proposed model is designed for two different languages, Bengali and English. For two different languages, two distinct word embeddings are used in order to enhance the performance of the proposed model. A GloVe(Global Vector) of high dimension is employed which introduces extensive statistical relationship of words with large corpora for News Headline Dataset. Furthermore, for BanglaSarc dataset, bangla word2vec embedding is used which is capable of handling complex pattern of Bengali language. The performance of the model is optimized by these dual word embeddings, which provides superior result in evaluation metrics.

After the data are preprocessed and tokenized, the proposed model recognizes word dependencies along with contextual relationship from the textual data by utilizing the Bi-LSTM and GRU layer separately. The outputs of two separate layers are incorporated and transferred to the dense layer. Dense layer utilizes the integrated output of the previous layer and provide the final output. The integration of the learned featured from both models, the dense layer enables the optimal decision making process to identify the sarcasm from two different languages.

In this proposed methodology, textual features can be extracted using the suggested method since GRU uses the short-term dependencies to reduce computational complexity while BI-LSTM extracts the long-term dependencies from the text. This integration of two different network enhances the extraction of diverse textual feature which eventually improves the result of whole feature representation. The architectural structure of the model is depicted in figure 1.

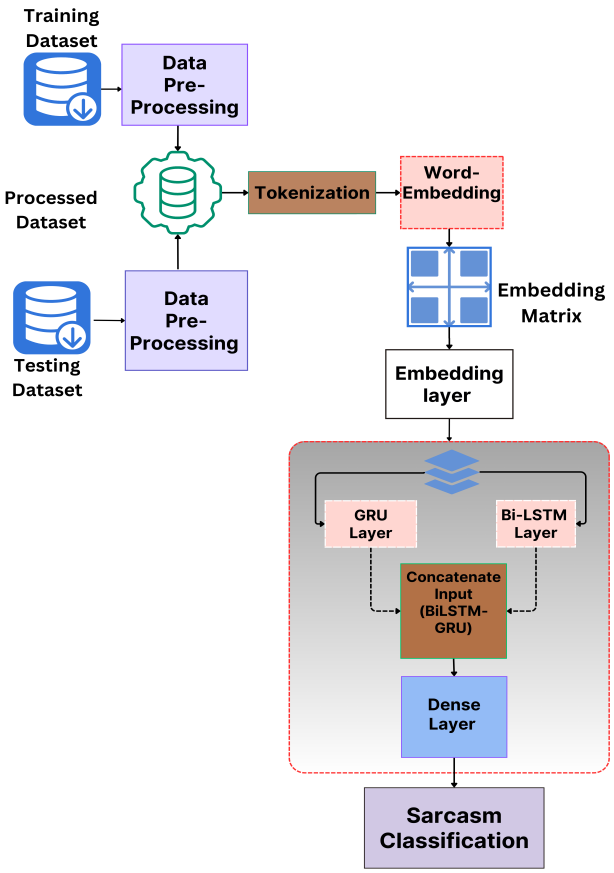

Fig. 1. Proposed Model Architecture

A. Data

1) News Headline Dataset: The dataset used for English language in this investigation comprises online headlines obtained from two websites: The Onion and The HuffPost. It is partitioned into two sections, which are subsequently combined to create a single dataset that is applied for both training and validation. There are 55,328 samples in the dataset, with 25,358 being classified as sardonic (denoted as 1) and 29,970 as non-sarcastic (denoted as 0).

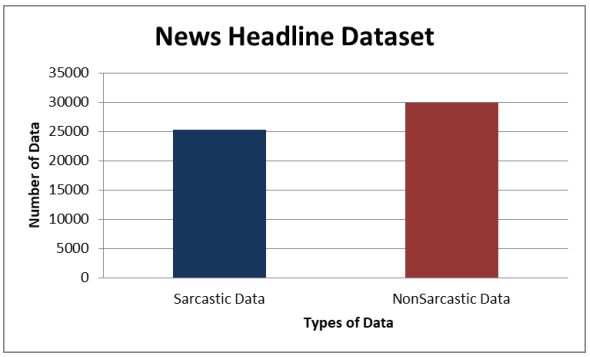

Fig. 2. Sarcastic headline and non sarcastic headline in News Headline Dataset

2) BanglaSarc: The dataset consists of 10,224 comments, of which 3,906 are classed as sarcastic (denoted as 1) and 6,318 as non-sarcastic (denoted as 0). The comments of the dataset were gathered from a variety of internet social media sites, including Facebook, YouTube, and a few websites that host blogs [14]. This dataset will assist in the study of identifying sarcasm, recognizing people's emotions, detecting various forms of Bengali expressions, and other fields due to the limited amount of data collection of categorized comments in Bengali.

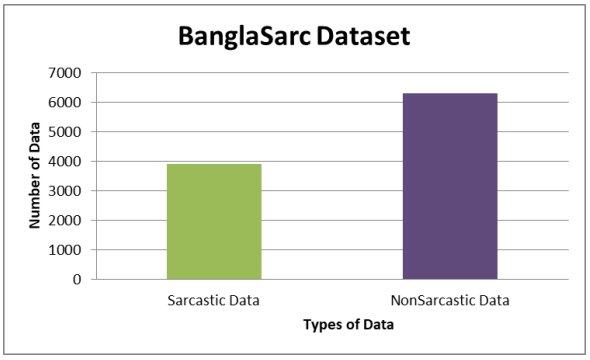

Fig. 3. Sarcastic headline and non sarcastic headline in BanglaSarc

B. Data Pre-Processing :

Here, data pre-processing is subdivided into two components, data cleaning and tokenization. Data cleaning process involves removing noise, such as punctuation and special characters, correcting misspellings, and standardizing text by converting it to a uniform case. In addition, stop words are removed, and text normalization techniques, such as stemming or lemmatization, are also applied to reduce words to their base forms.

Tokenization is a significant step in natural language processing (NLP) that involves dividing text into smaller units, such as words or subwords, known as tokens that facilitates the conversion of text into a structured format.

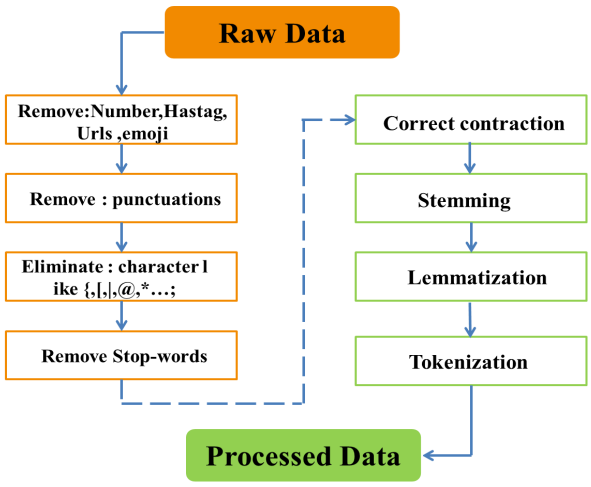

Fig. 4. Steps in Data Pre-Processing

C. Word Embedding :

1) GloVe : Global Vectors for Word Representation (GloVe) is an unsupervised learning system that produces vector representations of words by collecting detailed co-occurrence information from a given text corpus. This algorithm identifies and extracts semantic similarities and connections among words. This research employed glove word embeddings which were trained on a dataset of 42 billion words, with a dimension of 300.

2) Word2vec : Bangla word2vec is a proficient word embedding method which is competent in understanding the complex pattern of Bengali language. It transforms text into vector representation using the concealed meaning of a particular word according to the use of that word in text. Bangla word2vec is adept at capturing the compound nuances of languages and helps to recognize similarities and distinctions throughout various dialects in a precise way.

IV. Experimental Result :

The proposed ensemble model exhibits a significant improvement in accuracy and other result of evaluation metrics including precision and recall. The performance of the model outperforms current methodologies on News Headline Dataset and utilizes a large number of parameters. The accuracy of the proposed approaches for News Headline Dataset is 0.967. Again, For BanglaSarc dataset, the model has demonstrated the accuracy of 0.886 which indicates the strong ability of the model to detect the sarcasm from the Bengali text. Furthermore, other

performance indicators of the model including precision and recall are also measured. The result of the evaluation metrics of the model indicates the superior performance of the proposed model for detecting sarcasm in both Bengali and English text. The high precision score guarantees the reliability of sarcastic classifications, while the robust recall demonstrates the model's capacity to identify most sarcastic occurrences.

In addition, the accuracy versus loss plot of the two distinct datasets indicates that the model aligns appropriately with the data, without exhibiting overfitting or underfitting. The model generalizes effectively with unseen data without the need for customization or failure to comprehend significant patterns. The model fits both Bengali and English datasets properly without overfitting or underfitting and improve the robustness of the model.

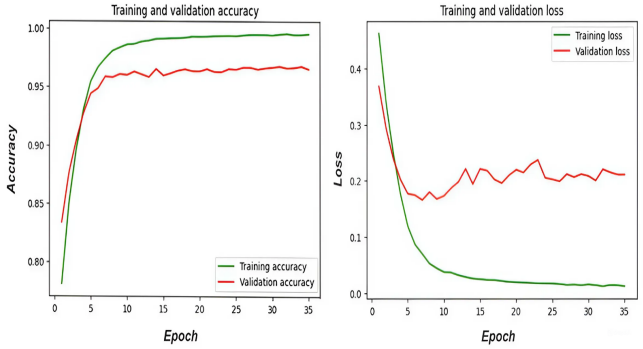

Fig. 5. Accuracy vs Loss plot of News Headline Dataset

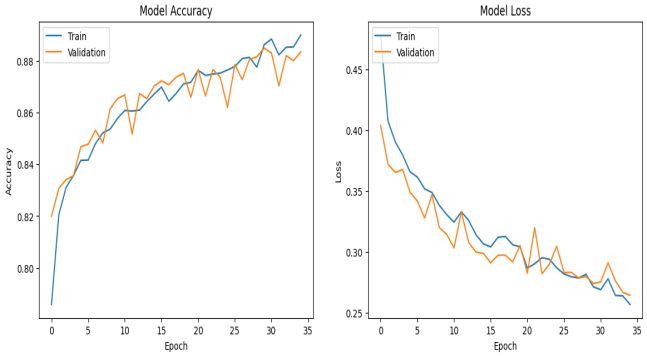

Fig. 6. Accuracy vs Loss plot of BanglaSarc Dataset

The qualitative assessment for the proposed model is illustrated in the confusion matrix. TP = 5788, TN = 4900, FP = 138, FN = 239 for the News Headline Dataset. Furthermore, the values for the BanglaSARC dataset are as follows: TP=610, TN=1199, FP=66, and FN=169.

Figure 7 and 8 represent the confusion matrix For News Headline Dataset and BanglaSarc dataset respectively. Table I displays the summary of the proposed model's results for two distinct datasets.

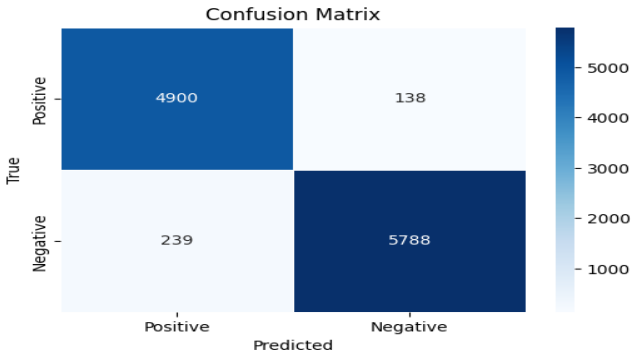

Fig. 7. Confusion Matrix for News Headline Dataset

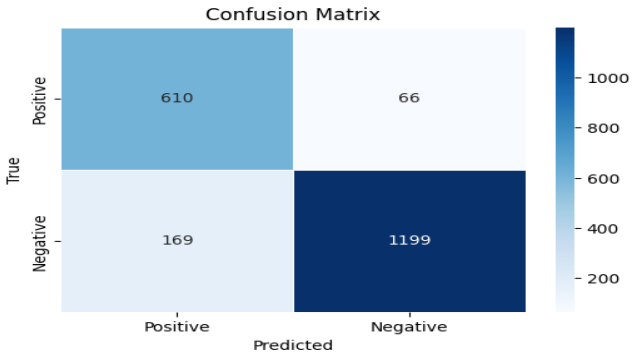

Fig. 8. Confusion Matrix for BanglaSarc Dataset

TABLE I
RESULT OF THE PROPOSED MODEL

Word Embedding	Dataset	Evaluation Metrics		
		Accuracy	Precision	Recall
GloVe(Global Vector)	News Headline Dataset	**0.966**	**0.976**	**0.960**
Bangla Word2vec	BanglaSarc Dataset	**0.885**	**0.902**	**0.783**

A visual representation of the comparison between different evaluation metrics result has been shown in the figure 9. The accuracy, precision, and recall of the News Headline Dataset are represented by A1, P1, and R1, respectively. Similarly, the accuracy, precision, and recall of the BanglaSarc Dataset are represented by A2, P2, and R2.

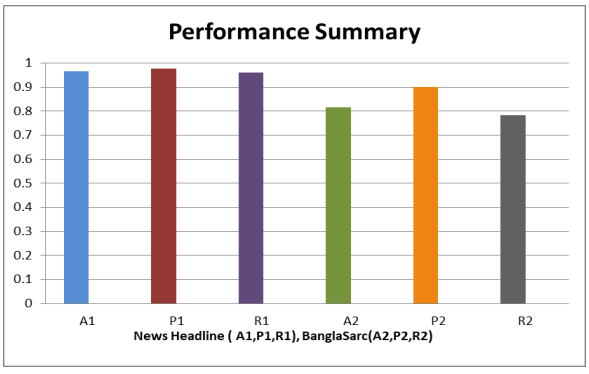

Fig. 9. Performance Summary of Proposed Model

TABLE II
COMPARISON WITH EXISTING WORK

Existing Model	Dataset	Accuracy
Attention Based BiLSTM [9]	News Headline Dataset	0.897
LSTM [15]		0.951
Hybrid Ensemble model based on Fuzzy Logic [16]		0.908
Multi-Head self Attention Based GRU [17]		0.916
LSTM-RNN [18]		0.838
Ensemble LSTM,word2vec,GloVe [19]		0.889
Proposed Model	News Headline Dataset	**0.967**
	Bangla Sarc Dataset	**0.886**

Table II presents a comparative analysis with prior research studies. The model is trained in an efficient way to detect sarcasm from the textual data by extracting information about context, linguistics pattern related a particular text. As two distinct layers generate two unique predictions independently, the probability of a false prediction is diminished by utilizing the combination of the predictions from separate layers. This makes the model enable to determine appropriately whether a text is sarcastic or not. The proposed model indicates the ability to diminish the complexity of multilingual text and amend complex natural language processing(NLP) applications such as opinion mining, text interpretation across various languages.

V. CONCLUSION

The ensemble model presented in this research is a high-dimensional GloVe embedding that incorporates both Bi-LSTM and GRU models. The proposed model extracts salient features, including semantic meanings, contextual information, and sequential dependencies. The evaluation results for precision, recall, and accuracy serve as evidence that the proposed methodology surpasses all other deep learning models and offers a distinctive, efficient solution for the detection of sarcasm in both Bangla and English languages. In comparison to other models, the model also exhibited superior accuracy within a restricted number of iterations. Furthermore, the model is capable of extracting a significantly greater number of parameters than other existing models. In summary, the ensemble model that has been introduced in this research exhibits exceptional performance in the detection of sarcasm in both Bengali and English text.

REFERENCES

[1] S. M. Sarsam, H. Al-Samarraie, A. I. Alzahrani, and B. Wright, "Sarcasm detection using machine learning algorithms in twitter: A systematic review," *International Journal of Market Research*, vol. 62, no. 5, pp. 578–598, 2020.

[2] N. Pawar and S. Bhingarkar, "Machine learning based sarcasm detection on twitter data," in *2020 5th international conference on communication and electronics systems (ICCES)*, pp. 957–961, IEEE, 2020.

[3] D. Šandor and M. B. Babac, "Sarcasm detection in online comments using machine learning," *Information Discovery and Delivery*, no. ahead-of-print, 2023.

[4] S. Khotijah, J. Tirtawangsa, and A. A. Suryani, "Using lstm for context based approach of sarcasm detection in twitter," in *Proceedings of the 11th international conference on advances in information technology*, pp. 1–7, 2020.

[5] D. M. Ashok, A. N. Ghanshyam, S. S. Salim, D. B. Mazahir, and B. S. Thakare, "Sarcasm detection using genetic optimization on lstm with cnn," in *2020 International Conference for Emerging Technology (INCET)*, pp. 1–4, IEEE, 2020.

[6] R. Ali, T. Farhat, S. Abdullah, S. Akram, M. Alhajlah, A. Mahmood, and M. A. Iqbal, "Deep learning for sarcasm identification in news headlines," *Applied Sciences*, vol. 13, no. 9, p. 5586, 2023.

[7] S. S. Salim, A. N. Ghanshyam, D. M. Ashok, D. B. Mazahir, and B. S. Thakare, "Deep lstm-rnn with word embedding for sarcasm detection on twitter," *2020 International Conference for Emerging Technology (INCET)*, pp. 1–4, 2020.

[8] R. Misra and P. Arora, "Sarcasm detection using hybrid neural network," *arXiv preprint arXiv:1908.07414*, 2019.

[9] R. Misra and P. Arora, "Sarcasm detection using news headlines dataset," *AI Open*, vol. 4, pp. 13–18, 2023.

[10] L. Liu, J. L. Priestley, Y. Zhou, H. E. Ray, and M. Han, "A2text-net: A novel deep neural network for sarcasm detection," in *2019 IEEE first international conference on cognitive machine intelligence (CogMI)*, pp. 118–126, IEEE, 2019.

[11] S. Chatterjee, S. Bhattacharjee, K. Ghosh, A. K. Das, and S. Banerjee, "Class-biased sarcasm detection using bilstm variational autoencoder-based synthetic oversampling," *Soft Computing*, vol. 27, no. 9, pp. 5603–5620, 2023.

[12] S. Kaya and B. Alatas, "Sarcasm detection with a new cnn+ bilstm hybrid neural network and bert classification model," *International Journal of Advanced Networking and Applications*, vol. 14, no. 3, pp. 5436–5443, 2022.

[13] S. Sinha and M. Choudhary, "Sarcasm detection using deep learning approaches: A review," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 11, no. 6, pp. 50–58, 2023.

[14] T. S. Apon, R. Anan, E. A. Modhu, A. Suter, I. J. Sneha, and M. G. R. Alam, "Banglasarc: A dataset for sarcasm detection," in *2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, pp. 1–5, IEEE, 2022.

[15] A. Barhoom, B. S. Abu-Nasser, and S. S. Abu-Naser, "Sarcasm detection in headline news using machine and deep learning algorithms," 2022.

[16] D. K. Sharma, B. Singh, S. Agarwal, N. Pachauri, A. A. Alhussan, and H. A. Abdallah, "Sarcasm detection over social media platforms using hybrid ensemble model with fuzzy logic," *Electronics*, vol. 12, no. 4, p. 937, 2023.

[17] R. Akula and I. Garibay, "Interpretable multi-head self-attention architecture for sarcasm detection in social media," *Entropy*, vol. 23, no. 4, p. 394, 2021.

[18] K. Bari, "Sarcasm detection of newspaper headlines using lstm-rnn," in *2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, pp. 604–608, IEEE, 2023.

[19] D. K. Sharma, B. Singh, and A. Garg, "An ensemble model for detecting sarcasm on social media," in *2022 9th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 743–748, IEEE, 2022.