

# Sarcasm detection methods based on machine learning and deep learning

**ZhiYu Zhang**

College of Computer Science and Technology, Shanghai University of Electric Power,  
Shanghai, 200070, China

1849053168@mail.shiep.edu.cn

**Abstract.** Sarcasm is an expression of emotion, which often expresses criticism and disgust in a positive tone, and we see it everywhere on social networking platforms. Because of this nature, sarcasm can be challenging to process in Natural Language Processing (NLP) Systems. Failing to detect sarcasm correctly can negatively affect the results of natural language processing tasks such as text correct and sentiment analysis. In recent years, there are more and more studies on sarcasm detection, and the research methods are diverse. According to the survey, the methods based on machine learning (ML) and deep learning (DL) account for the majority. Therefore, this paper mainly summarizes the literature on sarcasm detection based on ML and DL methods in recent years. This paper first introduces the application of sarcasm, then expounds the general architecture of sarcasm detection and makes a distinction between machine learning and deep learning. Then the author investigates the relevant literature and compares the results. Finally, the challenges and research directions in this field are summarized. The purpose of this paper is to facilitate the follow-up research of sarcasm detection.

**Keywords:** sarcasm detection, machine learning, deep learning.

## 1. Introduction

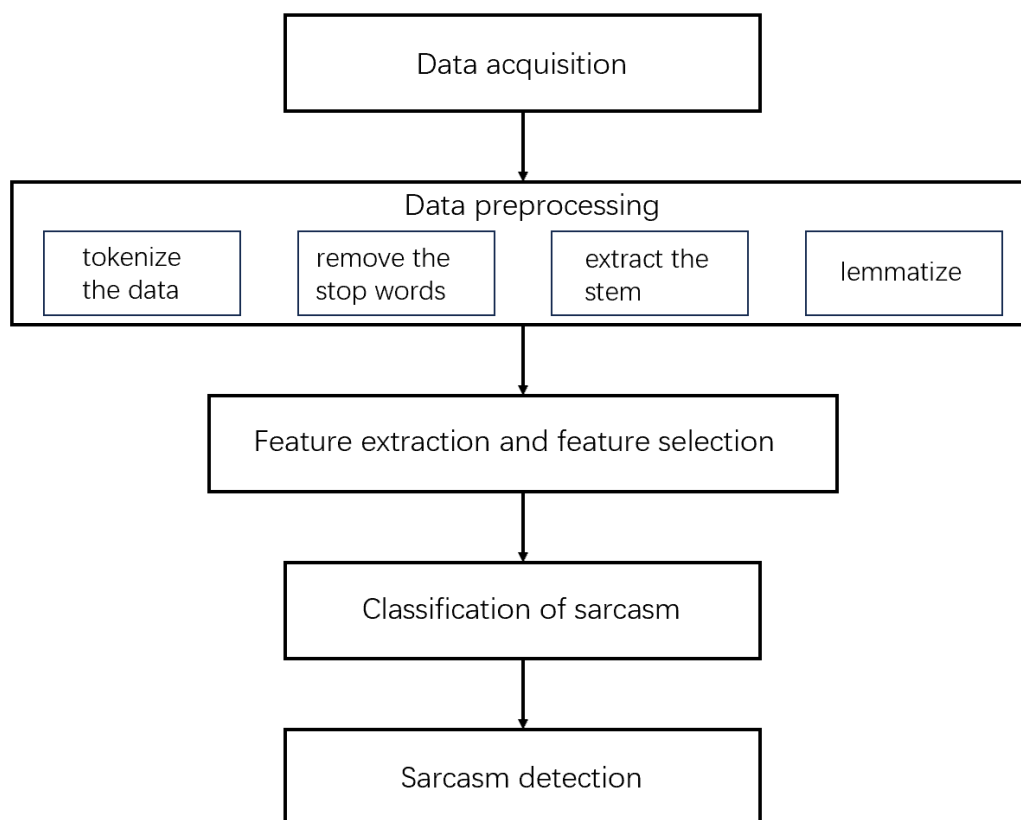
What is sarcasm? Emotion can be expressed through this method. People always use seemingly positive language to express the opposite meaning, in order to cause harm or disapproval to someone. For example, "I love being ignored!" "How I love working on Sundays!". It can be seen that although people use positive words such as "like" and "love" to express it, the emotions displayed are actually sarcastic and critical. In recent years, the network society has begun to develop rapidly. Twitter, Facebook, Instagram and other web social platforms, or Amazon, eBay, Wish and other e-commerce platforms have been highly recognized by netizens. They can express their opinions on any event and share their happiness and sorrow on these social platforms. They can also review products on e-commerce platforms, give praise to products they like, and may post sarcastic comments about products they don't like. Just like that, more and more people are getting involved, and the amount of data on the web is growing exponentially. Therefore, many studies on sarcasm detection use the data of social platforms such as Twitter as the data set for research, so as to obtain the opinions and views of people in different fields.

Here are the applications of sarcasm detection. Sarcasm detection is a type of sentiment analysis. It often requires non-literal indicators such as context and tone, so it is difficult for Natural Language Processing (NLP) Systems to make correct categorical judgments, which makes sarcasm detection challenging. But with the rapid increase in tweets and comments on social media in recent years, there is an urgent need for Natural Language Processing (NLP) Systems to correctly identify sarcasm. Sarcasm and satirical metaphors occupy a large part of the content of online social communication. For example, according to [1], the Internet corpus argument obtained from 4 forums.com is composed of 12% satirical statements. It can be seen that many people actually express their dissatisfaction through sarcasm, which makes sentiment analysis more difficult. Many studies have shown that the advancement of sarcasm detection research and the advancement of technology are very useful. According to [2], finding new avenues for business ventures can be achieved by detecting sentiments in posts on social media platforms and e-commerce portals. According to [3], sarcasm detection can facilitate emotion classification. The use of sarcasm in tweets and its influence on sentiment analysis were investigated by D Maynard and M Greenwood [4], according to the study, identifying sarcasm accurately can increase sentiment analysis ability by almost 50%.

## 2. Related knowledge

### 2.1. The general structure for sarcasm detection

According to [5], the general architecture of sarcasm detection can be divided into five steps: (a) data acquisition or data collection (b) data preprocessing, (c) feature extraction and feature selection, (d) sarcasm classification, and (e) sarcasm detection. The sequence is shown in Figure 1.



**Figure 1.** The framework of sarcasm detection.

Common sarcasm detection datasets include Twitter datasets, Amazon product datasets, and Facebook datasets. There are also manually labeled datasets that can be used for sarcasm detection. However, there is no best data set, and many data sets are imbalanced, which is one of the challenges of sarcasm detection. The obtained data sets usually have a lot of non-essential elements and interference elements, so it is usually necessary to tokenize the data, remove the stop word, extract the stem and lemmatize the word. The process of feature extraction involves converting the original data that the algorithm cannot recognize into features that can be recognized. The process of selecting features involves removing irrelevant features and maintaining relevant features. Common selection methods include: term-frequency (TF), Term frequency-inverse Document Frequency (TF-IDF), Part of Speech (POS) Tagging, N-Gram. There are many sarcastic classification methods, and the more popular methods in recent years can be mainly divided into: machine learning, deep learning, dictionary-based methods, etc. Most of the research is mainly based on machine learning and deep learning. Machine learning also includes Support Vector Machine (SVM), Naive Bayes, Random Forest and other methods. Deep learning is mainly based on neural networks such as CNN and RNN.

## 2.2. The difference between the machine learning and the deep learning

The author mainly investigates and discusses sarcasm detection methods based on machine learning and deep learning, so let's first discuss the differences between them. The comparison can be seen in Table 1. The following are mainly discussed from five aspects: Data dependency, Hardware dependency, The method to solve the problem, Execution time and Creating feature projects. In terms of data dependence, machine learning is more suitable for a small amount of data, while deep learning can adapt to a large amount of data, and the greater the amount of data used for training, the better the conclusions drawn by deep learning algorithms. On the hardware side, deep learning algorithms require a lot of matrix multiplication operations, so only GPU can effectively optimize these operations. Machine learning, on the other hand, is more traditional and can be run on low-end machines. In terms of problem solving, machine learning requires dividing the problem into different parts, solving them separately and then combining them to get the final result. Deep learning is about solving problems from start to finish. In terms of execution time, machine learning requires short training time, but if the amount of data processed is large, the execution time will be long. With deep learning, training is a time-consuming thing, but once the model is trained, it only takes a short time to process large amounts of data. The final and most important difference lies in the creation of feature engineering. To reduce data complexity and make features more visible to learning algorithms, feature engineering is the process of creating domain knowledge in feature extractors. It is challenging and costly in terms of time and expertise. Features in machine learning often require expert recognition and manual labeling and coding. Deep learning algorithms try to learn advanced features from data, which goes well beyond traditional machine learning and makes up for shortcomings.

**Table 1.** Difference between machine learning and deep learning.

	Machine learning	Deep learning
<b>Data dependency</b>	Less data is more advantageous	Requires a lot of data
<b>Hardware dependency</b>	Run on low-end machines	Relies on high-end machines
<b>The method to solve the problem</b>	Solve problems separately, recombine to get results	Solve problems from start to finish
<b>Execution time</b>	Short training time, long problem solving time	Training time is long, problem solving time is short
<b>Creating feature projects</b>	An expert is required to manually annotate the features	Learn advanced features from the data

### 3. Literature review

#### 3.1. Machine learning

S K Bharti et al. proposed a data sarcasm detector based on hyperbolic features. They believe that the most critical tasks for sarcasm detection in text using machine learning are database collection and feature extraction. After they preprocessed the Twitter dataset, they labeled the part-of-speech. In the feature extraction module, interjections and intensifiers build a feature set. Finally, five classifiers, NB, DT, SVM, RF and AdaBoost, are used to get the experimental results [6].

In order to detect sarcasm, Neha Pawar et al. proposed a pattern method based on Twitter dataset. After the data is preprocessed, four features of the data are extracted: The Feature Related to Sentiment, The Feature Related to Punctuation, The Feature Related to Syntactic and Semantic, The Feature is Related to Pattern. After training the data on the model, SVM, KNN and Random Forest algorithm were used to classify the data. Finally, they compared the results of the three models and found that the Random Forest classifier has higher accuracy than other classifiers [7].

On the basis of machine learning, Vinoth et al. proposed another sarcasm language detection and classification technique based on intelligent machine learning (IMLB-SDC). The model includes the following modules: Preprocessing, Feature engineering based on TF-IDF, FS, SVM-based classification and PSO-based parameter adjustment. At the initial stage, different subprocesses are required to pre-process the actual input data into a compatible format. In addition, using PSO algorithm to optimize SVM parameters can improve the performance of sarcasm detection and classification. The author thinks that IMLB - SDC technique can also be improved through deep learning methods [8].

**Table 2.** Summary of the results of the ML method.

Author	Proposed Approach	Dataset	Results
S K Bharti et al. [6]	Machine Learning--A hyperbolic feature-based sarcasm detector	Twitter data	The accuracy of: NB:75.12% DT:80.27% SVM:80.67% RF:80.79% AdaBoost:80.07%
Neha Pawar et al. [7]	Machine Learning--Pattern approach based on Twitter dataset	Twitter data	Random Forest Accuracy:81% Precision:82% Recall:63.34% F1-Score:79%
Vinoth et al. [8]	Machine Learning--IMLB-SDC	—	Precision:94.7% Recall:95.2% F1-Score:94.9%

#### 3.2. Deep learning

Le Hoang Son et al. proposed a deep learning model called sAtt-BLSTM convNet. It is a hybrid of soft attention-based bidirectional long short-term memory and convolutional neural networks, in which punctuation based auxiliary features are also incorporated. The proposed automatic detection framework includes data set acquisition, preprocessing, implementation of Att-BLSTM convolutional network model, and results and analysis. They conducted experiments on both balanced and imbalanced datasets and compared them with baseline deep learning models [9]. A Kumar et al. considered various manual features and then constructed a SVM model that could perform sarcasm detection. A bidirectional long short-term memory (MHA-BiLSTM) network based on multi-head self-attention is proposed. The network uses a variety of manual features. With the help of manual features, the performance of MHA-BiLSTM is much improved compared with BiLSTM [10]. S Sangwan et al.

investigated and found that most existing methods categorize sarcasm detection as "text classification". But they believe that simple text detection is not good enough to distinguish sarcasm from non-sarcasm, and pictures and contextual information are also important factors affecting sarcasm detection. Therefore, they propose a deep learning framework based on RNN, which can accept and process multi-modal input information. The experiment was conducted using tweets from the Instagram platform as a dataset and compared with the performance of existing technologies [11].

D M Ashok et al. investigated various approaches to sarcasm detection, such as rule-based approaches, supervised or semi-supervised learning, and neural networks. After that, they propose a deep learning framework with BERT as the embedding layer, and the LSTM layer's hyperparameters, window size and number, and hidden units have been optimized using genetic algorithms and convolutional neural networks. The experiment was conducted after collecting and preprocessing Twitter data. The final accuracy is about 93-95% [12]. S S Salim et al. aimed at making it easier for companies to carry out sarcasm detection, they studied algorithms that could improve sarcasm detection. After investigating sarcasm detection, they proposed RNN-LSTM model. The model consists of four modules: Data set, Preprocessing, Word embedding and RNN-LSTM. They mined satirical tweets created by Twitter's public accounts as a dataset to experiment with. In the future, RNN-LSTM can be used to detect non-literal emotions [13]. D Sahoo et al. proposed a model based on CNN-GRU to detect the nature of comments, that is, sarcasm or non-sarcasm. They chose GRU over LSTM because GRU has fewer gates than LSTM, which means that training GRU requires fewer parameters and shorter runtime, but the performance is comparable to LSTM. The experiment uses the data set provided by the SARC corpus and is classified according to sarcasm and non-sarcasm. The data set they used is balanced. They compared the results with existing methods and found that the model performed better [14].

**Table 3.** Summary of the results of the DL method (a).

Author	Proposed Approach	Dataset	Results
Le Hoang Son et al. [9]	Deep Learning--sAtt-BLSTM convNet	Twitter data	Accuracy:93.71% Recall:92.67% Precision:90.49% F1-Score:88.29% Recall of Balanced:83.03% Precision of Balanced:72.63% F1-Score of Balanced:77.48%
A Kumar et al. [10]	Deep Learning--MHA-BiLSTM	Reddit data	Recall of Imbalanced:53.71% Precision of Imbalanced:60.26% F1-Score of Imbalanced:56.79%
S Sangwan et al. [11]	Deep Learning--A deep learning framework based on RNN	Instagram data	Accuracy of Text:66.17% Accuracy of Text+Image:70.0% Accuracy of Text+Image+Transcript:71.5%
D M Ashok et al. [12]	Deep Learning--MHA-BiLSTM	Twitter data	Accuracy is about 93-95%
S S Salim et al. [13]	Deep Learning--RNN-LSTM	Twitter data	Accuracy:85.23%
D Sahoo et al. [14]	Deep Learning--CNN-GRU	Data set provided by SARC corpus(Balanced)	Recall:83.06% Precision:79.59% F1-Score:77.47%

D M Kumar and A Patidar preprocessed the data with machine learning algorithm firstly, and then decoded the meaning of words in the form of real-value vectors so that the model could perform better. They then proposed a stack-based bidirectional Long Short-term memory (Stk-BiLSTM) model that

improves overall performance primarily by identifying important words in set of words or a sentence. They thought the Twitter dataset was messy because it used too many hashtags, so they collected real headlines from two news sites with all the facts, and then categorized them [15]. The task of automatically identifying sarcastic instances in short articles was the focus of M S Razali et al. The study involves extracting the best features by combining a deep learning architecture with contextual features that are carefully hand-crafted. Their proposed approach focuses on using deep learning architectures to create deep features, which are then combined with reasonably hand-crafted features. The dataset is pulled from news content from Onion and CNN. By comparing the results of SVM, KNN, LR, DT and DISCR, it is found that logistic regression (LR) is the highest in this work, and the performance compared with the existing works is also the best. This study demonstrates that the task of NLP detection can be further optimized by combining supervised learning with manual feature extraction [16]. M Jeyakarthic and J Senthilkumar [17] proposed a new OBiLSTM-SDC algorithm designed to identify sarcastic tweets and comments on social network platforms. The model includes preprocessing, TFIDF vectorizer, BiLSTM classifier and CSO hyperparameter optimizer. The TF-IDF technique removes feature engineering and calculates the largest appropriate items in both true and false news from the data set [18]. Therefore, it helps to get the best efficiency. Hyperparameter fine-tuning of the BiLSTM model was performed using CSO techniques to improve the classifier results. The model is tested with Twitter data set, and the results of comparison with other methods are given.

**Table 3.** Summary of the results of the DL method (b).

Author	Proposed Approach	Dataset	Results
D M Kumar and A Patidar [15]	Deep Learning--Stk-BiLSTM	The real news headlines	Accuracy:86.47% Precision:82.39% Recall:86.73% F1-Score:84.50%
M. S. Razali et al. [16]	Deep Learning--Deep learning architecture combined with reasonable hand-crafted features	News contents pulled from Onion and CNN	Random Forest Precision:95% Recall:94% F1-Score:94%
M Jeyakarthic and J Senthilkumar [17]	Deep Learning--OBiLSTM-SDC	Twitter data	Accuracy:94.6% Precision:94.5% Recall:95.2% F1-Score:94.9%

By observing Table 2 and Table 3, we can find the following conclusions. First, in recent years, the proportion of research based on ML and DL methods is obviously more inclined to deep learning research. Second, comparing the results of deep learning and machine learning, it can be found that the deep learning-based model is better at detecting sarcasm in the data set. This is because by learning different levels of abstraction, a deep learning model can simulate higher-level features and capture more complex relationships between modeling inputs and outputs [19]. Moreover, from the difference between the two mentioned above, with the expansion of the amount of data in the future, deep learning will be more adaptable, so deep learning models will be more favored by researchers.

#### 4. Limitations of methods and challenges of sarcasm detection

##### 4.1. Limitation

Although the deep research approach mentioned above is superior to the machine learning approach, both have limitations. As can be seen in Table 4, for machine learning, features need to be marked manually, which will waste a lot of manual power and time, and feature knowledge is limited. For deep learning, if you want to train the model, you need a lot of data, and it is clear that there is not yet a unified golden data set. The data imbalance is a problem of both models.

**Table 4.** Limitations of both approaches

Machine learning	Manual feature extraction Limited feature knowledge Data imbalance
Deep Learning	Automated feature extraction Black box model Need huge datasets Data imbalance

According to [20], Chatterjee et al. proposed a BVA model and combined five oversampling techniques for research. The results indicate that SVM-SMOTE and ADASYN are the most effective methods for reducing class imbalance in those corpus. However, they mention that the study still has limitations, such as not considering other resampler paradigms, and the work is based only on a binary classification task between satirical and non-satirical texts. This work has already opened the way to solving the problem of data imbalance, and hopefully there will be better progress in the future.

#### 4.2. Challenges

The challenges of sarcasm detection are many, for example:

- ① Most of the time, people will use recognized events or specific languages, objects, etc., to satire, and the current sarcasm detection is only limited to text and other content, how to make the satirize detection system to keep up with the rhythm of the network is a challenge.
- ② Detect sarcasm in informal language.
- ③ People's tweets and comments are often filled with emojis. The recognition and detection of emojis is also important.
- ④ Detect sarcasm in different languages. Language habits vary greatly from language to language or language family, which is also one of the challenges.
- ⑤ Identify sarcasm in images. This requires a combination of image processing techniques, which makes irony detection challenging.
- ⑥ Identify sarcasm in audio. Sarcasm detection of audio clips is also a wide area to be explored [21].

#### 5. Conclusion

This article covers the definition and application scenarios of sarcasm detection, explains the difference between ML and DL in NLP technology, summarizes the research results in the field in recent years, and lists some challenges at the end. Through the comparison of research results, it is found that the method based on deep learning performs better in sarcasm detection. After reaching the conclusion, the author also makes an explanation. The author hopes that this paper will be useful for future research.

#### References

- [1] M. A. Walker, J. E. F. Tree, P. Anand, R. Abbott and J. King, "A corpus for research on deliberation and debate", *Proc. LREC*, vol. 12, pp. 812-817, 2012.
- [2] Adarsh M J;Pushpa Ravikumar.Sarcasm detection in Text Data to bring out genuine sentiments for Sentimental Analysis[A].2019 1st International Conference on Advances in Information Technology (ICAIT)[C],2019
- [3] N.Majumder, S. Poria, H. Peng, N. Chhaya, E. Cambria and A. Gelbukh, "Sentiment and Sarcasm Classification With Multitask Learning," in *IEEE Intelligent Systems*, vol. 34, no. 3, pp. 38-43, 1 May-June 2019, doi: 10.1109/MIS.2019.2904691.
- [4] D. Maynard and M. Greenwood, "Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis", *Language Resources and Evaluation Conference (LREC)*, 2014.

- [5] J. aboobaker and E. Ilavarasan, "A Survey on Sarcasm detection and challenges," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 1234-1240, doi: 10.1109/ICACCS48705.2020.9074163.
- [6] S. K. Bharti, R. Naidu and K. S. Babu, "Hyperbolic Feature-based Sarcasm Detection in Tweets: A Machine Learning Approach," 2017 14th IEEE India Council International Conference (INDICON), Roorkee, India, 2017, pp. 1-6, doi: 10.1109/INDICON.2017.8487712.
- [7] Neha Pawar;Sukhada Bhingarkar.Machine Learning based Sarcasm Detection on Twitter Data[A].2020 5th International Conference on Communication and Electronics Systems (ICCES)[C].2020
- [8] Vinoth, D., Prabhavathy, P. An intelligent machine learning-based sarcasm detection and classification model on social networks. J Supercomput 78, 10575–10594 (2022). <http://doi.org.shiep.vpn358.com/10.1007/s11227-022-04312-x>
- [9] Le Hoang Son;Kumar, A.;Sangwan, S.R.;Arora, A.;Nayyar, A.;Abdel-Basset, M..Sarcasm Detection Using Soft Attention-Based Bidirectional Long Short-Term Memory Model With Convolution Network[J].IEEE Access,2019,Vol.7: 23319-23328
- [10] A. Kumar, V. T. Narapareddy, V. Aditya Srikanth, A. Malapati and L. B. M. Neti, "Sarcasm Detection Using Multi-Head Attention Based Bidirectional LSTM," in IEEE Access, vol. 8, pp. 6388-6397, 2020, doi: 10.1109/ACCESS.2019.2963630.
- [11] S. Sangwan, M. S. Akhtar, P. Behera and A. Ekbal, "I didn't mean what I wrote! Exploring Multimodality for Sarcasm Detection," 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 2020, pp. 1-8, doi: 10.1109/IJCNN48605.2020.9206905.
- [12] D. M. Ashok, A. Nidhi Ghanshyam, S. S. Salim, D. Burhanuddin Mazahir and B. S. Thakare, "Sarcasm Detection using Genetic Optimization on LSTM with CNN," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-4, doi: 10.1109/INCET49848.2020.9154090.
- [13] S. S. Salim, A. Nidhi Ghanshyam, D. M. Ashok, D. Burhanuddin Mazahir and B. S. Thakare, "Deep LSTM-RNN with Word Embedding for Sarcasm Detection on Twitter," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-4, doi: 10.1109/INCET49848.2020.9154162.
- [14] D. Sahoo, N. R. Paul, R. C. Balabantaray and A. U. Khan, "Sarcasm Detection Using Deep Learning," 2021 19th OITS International Conference on Information Technology (OCIT), Bhubaneswar, India, 2021, pp. 331-335, doi: 10.1109/OCIT53463.2021.00072.
- [15] D. M. Kumar and A. Patidar, "Sarcasm Detection Using Stacked Bi-Directional LSTM Model," 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2021, pp. 1-5, doi: 10.1109/ICAC3N53548.2021.9725488.
- [16] M. S. Razali, A. Abdul Halin, Y. -W. Chow, N. Mohd Norowi and S. Doraisamy, "Context-Driven Satire Detection With Deep Learning," in IEEE Access, vol. 10, pp. 78780-78787, 2022, doi: 10.1109/ACCESS.2022.3194119.
- [17] M. Jeyakarthic and J. Senthilkumar, "Optimal Bidirectional Long Short Term Memory based Sentiment Analysis with Sarcasm Detection and Classification on Twitter Data," 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, 2022, pp. 1-6, doi: 10.1109/MysuruCon55714.2022.9972540.
- [18] M.J. Awan, A. Yasin, H. Nobanee, A. A. Ali, S. Z. hahzad, M. Nabeel, et al., "Shahzad Fake News Data Exploration and Analytics", Electronics, vol. 10, no. 19, pp. 2326, 2021.
- [19] B. Singh and D. K. Sharma, "A Survey of Sarcasm Detection Techniques in Natural Language Processing," 2023 6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2023, pp. 1-6, doi: 10.1109/ISCON57294.2023.10112176.
- [20] Chatterjee, S., Bhattacharjee, S., Ghosh, K. et al. Class-biased sarcasm detection using BiLSTM variational autoencoder-based synthetic oversampling. Soft Comput 27, 5603–5620 (2023).



- [21] R. Gupta, J. Kumar, H. Agrawal and Kunal, "A Statistical Approach for Sarcasm Detection Using Twitter Data," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 633-638, doi: 10.1109/ICICCS48265.2020.9120917.