A Survey of Sentiment Analysis and Sarcasm Detection: Challenges, Techniques, and Trends

Review Paper

Ahmed Derbala Yacoub
Helwan University Faculty of Computers and Artificial Intelligence
Cairo EGYPT
ahmed_drbala_1073@fci.helwan.edu.eg

Salwa O. Slim
Helwan University Faculty of Computers and Artificial Intelligence
Cairo EGYPT
salwaosama@fci.helwan.edu.eg

Amal Elsayed Aboutabl
Helwan University Faculty of Computers and Artificial Intelligence
Cairo EGYPT
amal.aboutabl@fci.helwan.edu.eg

Abstract *– In recent years, more people have been using the internet and social media to express their opinions on various subjects, such as institutions, services, or specific ideas. This increase highlights the importance of developing automated tools for accurate sentiment analysis. Moreover, addressing sarcasm in text is crucial, as it can significantly impact the efficacy of sentiment analysis models. This paper aims to provide a comprehensive overview of the conducted research on sentiment analysis and sarcasm detection, focusing on the time from 2018 to 2023. It explores the challenges faced and the methods used to address them. It conducts a comparison of these methods. It also aims to identify emerging trends that will likely influence the future of sentiment analysis and sarcasm detection, ensuring their continued effectiveness. This paper enhances the existing knowledge by offering a comprehensive analysis of 40 research works, evaluating performance, addressing multilingual challenges, and highlighting future trends in sarcasm detection and sentiment analysis. It is a valuable resource for researchers and experts interested in the field, facilitating further advancements in sentiment analysis techniques and applications. It categorizes sentiment analysis methods into ML, lexical, and hybrid approaches, highlighting deep learning, especially Recurrent Neural Networks (RNNs), for effective textual classification with labeled or unlabeled data.*

Keywords: *Sarcasm detection, sentiment analysis, natural language processing, Deep learning, machine learning*

1. INTRODUCTION

The paper's objective is to thoroughly study all the related research to understanding people's feelings and identifying sarcasm, whether accomplished using a single language or multiple languages. We'll also compare these studies based on what they aim to solve, what methods they use, what the dataset used, how they test their results, who benefits from their work, what the outcomes are, and what's good and not so good about their approaches. We'll also investigate the problems in sentiment analysis and sarcasm detection, and we'll explore the ways researchers have come up with to tackle these problems. We'll attempt to identify which solutions work best and what's coming next in sentiment analysis and sarcasm detection. Sentiment analysis [1, 2], crucial for understanding public opinions on products and events, has become vital with the internet's growth, but analyzing vast online data poses challenges. The goal [3-11] is to assess sentiment, often on a positive-negative spectrum for written reviews. Sentiment analysis, situated at the intersections of computational linguistics, NLP, and data mining, utilizes techniques from these fields to extract sentiments from text. It explores research areas such as data mining and machine learning within the context of NLP [12,13]. Sarcasm [14-18], a statement conveying the opposite of its intended meaning and often used humorously or critically, poses a challenge for sentiment analysis as it can invert the true sentiment of a statement. Despite being a popular research topic, sarcasm detection [19-26] is crucial as sentiment analysis can misinterpret sarcastic sentences, leading to inaccurate sentiment classifications. The difficulty lies in the nuanced nature of human emotions and expressions conveyed through text, making automatic sarcasm detection a challenging task within natural language processing (NLP) [27, 28]. In sentiment analysis, written opinions often mix several languages, making it difficult to fully capture the text messages, resulting in a more complex classification of the text's polarity. Sentiment analysis struggles with multilingual comments that mix different languages [29, 30]. In tackling this issue, a sentiment analysis tool is essential for sorting reviews in the intended language and discarding others. However, this might make data smaller and influence the accuracy of linear classifiers. Identifying the source language is essential for successful cross-lingual sentiment analysis, as it helps to use suitable resources for sentiment analysis in that language [31].

Some challenges in sentiment analysis and sarcasm detection are:

- Determining sentiment in a sentence becomes difficult with the presence of sarcasm, making it challenging to discern the intended meaning. Detecting sarcasm in text poses a challenge due to context absence, user personality, the lack of expressions and body language, diverse writing styles, noise in text, and limitations in accurately classifying input data for sentiment analysis.
- Classifying people's sentiments is hard when they use different languages, making it complex to decide if the message is positive or negative. Most sentiment analysis focuses on English, but as more people use their languages on social media like Facebook and Twitter, there's a need to handle reviews in different languages.
- Sentiment analysis becomes complicated in cases where contextual information and explicit sentiment words are absent in an implicit text and when individuals express their emotions in an obscure and implicit manner.
- While current word vectors primarily focus on word meaning in sentiment analysis, it's crucial to incorporate sentiment aspects in embeddings because words with different sentiments, like "I'm sad" and "I'm happy," can end up with similar vectors due to shared contexts.

For Methods, the study of sentiment analysis and sarcasm detection categorizes methods into three types: machine learning (ML), lexical-based, and hybrid approaches. Deep learning, especially recurrent neural networks (RNNs), is valuable for classifying textual information due to its flexibility and complex structure. Machine learning involves training a computer with algorithms to detect patterns in large datasets, with deep learning (DL) being a component that uses algorithms for advanced data abstraction. In ML, supervised learning relies on labeled training data to predict outcomes accurately, while unsupervised learning discovers patterns without labeled data. Popular ML techniques include Naive Bayesian, Support Vector Machine, Artificial Neural Networks, and the K-means algorithm.

The main contributions of this paper are:

- The survey paper compares 40 relevant studies on sentiment classification and detection of sarcasm, revealing significant variations in model performance and identifying gaps in the literature. It is a valuable resource, offering insights to guide future research and development in these areas.
- The paper evaluates the performance of diverse methods and algorithms in sarcasm detection and sentiment classification, comparing their accuracy and highlighting strengths and limitations to assist researchers in choosing suitable techniques for specific applications.
- The paper addresses the challenges of multilingual comments in sentiment analysis. It emphasizes the need for a sentiment analysis tool capable of processing multilingual and mixed-code datasets.
- The paper identifies future trends and expectations in sarcasm detection and sentiment classification. It discusses the potential of advanced deep learning techniques, such as transformers and attention mechanisms.

2. LITERATURE REVIEW

In the world of sentiment analysis and sarcasm detection for written communications, a growing area of study driven by the increasing need for a better understanding of textual data is appearing. This review thoroughly examines previous research in sentiment analysis and sarcasm detection. It covers studies focused on specific tasks, including Sentiment Analysis, Sarcasm Detection, Joint Sarcasm Detection and Sentiment Analysis, Multilingual Sentiment Analysis, and Multilingual Sarcasm Detection.

2.1. SENTIMENT ANALYSIS

Sharmin et al. [1] proposed an attention-based convolutional neural network (CNN) approach for sentiment analysis in the Bengali language. The authors use a dataset of Bengali movie reviews to train and evaluate their model. The literature supports the use of deep learning models, especially CNNs, for sentiment analysis tasks, and adding attention mechanisms to CNNs has been shown to improve their performance.

De Kok et al. [2] proposed a novel model for aspect-based sentiment analysis (ABSA) that analyzes sentiments or feelings related to attributes or characteristics of a product or service. The model utilizes ontology features to enhance the accuracy of ABSA models.

The authors use a dataset of hotel reviews to train and evaluate their model.

Elfaik et al. [3] proposed a deep learning approach for sentiment analysis of Arabic text using a deep bidirectional LSTM network with pre-trained word embedding. The authors utilize a dataset of Arabic tweets to train, assess, and compare their model with other state-of-the-art models.

Sarkar et al. [4] proposed a deep learning-based approach for sentiment analysis of Bengali tweets using the SAIL 2015 Bengali dataset, which contains 9,108 Bengali tweets labeled with positive, negative, and neutral sentiments. They employ a Bidirectional Long Short-Term Memory (BILSTM) network with an attention mechanism and incorporate external knowledge from SentiWordNet for improved performance. The proposed approach achieves an accuracy of 55%.

Chen et al. [5] proposed a new pre-processing method for implicit sentiment text classification that converts text features into picture features using FER-Net and TTP. The proposed method achieved an accuracy of 85% on the SMP2019 dataset. The study also introduces a new research direction in implicit sentiment analysis work by combining text classification and image analysis. Nonetheless, the proposed method has limitations in dealing with more detailed implicit sentiment text categorization tasks.

Zhao et al. [6] presented a new approach to sentiment analysis on social media platforms, which involves capturing the implicit text features hidden in user comments and the explicit sentiment network formed by the topological relationships of users to entities. The proposed approach comprises two primary components: a text representation module based on word graphs and a network embedding module for a heterogeneous social graph. Low-dimensional potential representations of network nodes are obtained by operating network embedding techniques, and non-consecutive and long-distance semantics are captured by adding feature learning technology. The original social network is divided into a user entity sentiment network and a user relationship structure network and acquires their node embedding representations individually using two separate autoencoders.

Huang et al. [7] proposed a novel model for sentiment analysis in Vietnamese by leveraging sentiment word embedding alongside transfer learning. The model utilizes a BiLSTM with an attention mechanism and generates sentiment vectors using Word2Vec. The model uses the Vietnamese Students' Feedback Corpus as the training dataset. The authors also demonstrate the effectiveness of using English corpus for transfer learning to improve the sentiment classification of Vietnamese. The proposed sentiment word vector can represent both the context and the sentiment information of words, which enhances the model performance. The accuracy achieved by the proposed model is 86%, which shows its potential in a successful sentiment analysis for the Vietnamese language.

Hussein et al. [8] proposed a method for Arabic sentiment analysis of multi-dialect text that utilizes machine learning techniques. The authors collected a dataset of tweets from different Arabic dialects and pre-processed it by removing stop words, non-alphanumeric characters, and URLs. They then used a feature extraction technique called Term Frequency-Inverse Document Frequency (TF-IDF) to convert the tweets into numerical vectors. The proposed model utilized three machine learning classifiers for sentiment analysis: Naive Bayes, Decision Trees, and Support Vector Machines (SVM). The experiments were conducted using a 10-fold cross-validation technique, and the results show that SVM achieved the highest accuracy of 87.2%, followed by Logistic Regression with 0.85 and Stochastic Gradient Descent with 85%. The authors also conducted experiments with different feature selection techniques and found that the Chi-square feature selection method provided the best results.

Ma et al. [9] proposed a novel model called Attentive LSTM with Common Sense Knowledge (ALCSK), which utilizes a pre-trained commonsense knowledge graph to enrich the aspect embedding layer and to enhance the accuracy of Targeted Aspect-Based Sentiment Analysis (TABSA) methods. The proposed model outperforms existing state-of-the-art methods on benchmark datasets, demonstrating its effectiveness in handling domain-specific knowledge and incorporating commonsense data.

Londhe et al. [10] propose a system that combines Long Short-Term Memory (LSTM), fuzzy logic, and incremental learning for sentiment analysis in restaurant reviews. The system effectively handles multi-category classification and captures cross-category correlations by incorporating binary classifiers and an ensemble approach. Adding fuzzy-encoded LSTM enhances model performance, achieving an accuracy of 81.04% and outperforming previous methods.

Li et al. [11] proposed an enhanced label propagation algorithm (LPA) referred to as the label attenuation propagation model (LAPM) for the automated assessment of emoji sentiment. The paper addresses the need for quantifying emoji sentiment and introduces the use of LPA in the emoji sentiment analysis field. LAPM is employed to compute sentiment uncertainty and emojis sentiment based on the Emoji Link Network (ELN) built for effectively managing and categorizing many social media emojis. The accuracy achieved by the proposed model is 85%, which is higher than the 74% accuracy achieved by LPA.

Hiyama et al. [12] proposed a neural network-based sentiment analysis model that uses an attention mechanism. The model consists of four layers: word-embedding, bidirectional LSTM, attention, and classification layers. The embedding layer converts the input into a fixed-size continuous vector. The bidirectional

LSTM layer produces contextual word vectors, while the attention layer assesses the significance of these contextual word vectors and computes a sentence vector. Finally, the classification layer utilizes the sentence vector to predict the sentiment polarity. The authors provide details on each layer and how they are used to enhance sentiment analysis results.

Liu et al. [13] proposed an algorithmic framework for cross-domain aspect-level sentiment analysis using the Amazon product review dataset, which comprises reviews for products across five domains: Books, DVD disks, Electronics, Kitchen appliances, and Videos. The proposed algorithmic framework utilizes the convolutional, adversarial, and BERT models. The authors Altered the CNN architecture by introducing a gated activation unit to improve its performance. The proposed model achieved an accuracy of 88.9%.

2.2. SARCASM DETECTION

Bagate et al. [14] proposed a model for detecting sarcasm in tweets without relying on specific clues such as hashtags. The dataset used for this research was collected from Kaggle and consisted of 51,188 tweets. The proposed model uses machine learning classification methodology and deep learning embedding techniques, which include data pre-processing steps such as tokenization, lower casing, stop word removal, and stemming. The model utilizes a stacking way that combines the outcomes of logistic regression and a Long Short-Term Memory (LSTM) recurrent neural network, feeding them into a light gradient boosting technique. The evaluation metrics used for the model are F1-score and confusion matrix. The proposed model achieves better results in sarcasm detection without the presence of hashtag 'sarcasm' compared to existing methods.

Zhang et al. [15] developed a method to enhance the performance of sarcasm detection with limited labeled data by applying adversarial training. The authors employed different neural network architectures, like Convolutional Neural Networks (CNN) and Hierarchical Recurrent Neural Networks (HRNN) and incorporated self-attention mechanisms into these models. For evaluation, the authors conducted evaluations using six different models on three datasets from the Internet Argument Corpus (IAC), Rhetorical Questions, and Hyperbole. The findings demonstrated that their adversarial training technique significantly improved sarcasm detection performance.

Zhang et al. [16] proposed a novel deep learning-based model for detecting sarcasm in Chinese text using contextual information from social networks. The model includes a Pair of concurrent modules, sketchy reading, and Intensive Reading, which use an attention mechanism and a hierarchical method to extract features from the input. The output from the two modules is then concatenated and passed through a fully connected layer to make the final prediction.

Tanya et al. [17] proposed a hybrid approach combining deep learning and traditional machine learning methods for detecting sarcasm in social media, using the News Headlines Dataset. The authors pre-process the dataset by removing usernames, stop words, and non-alphanumeric characters. To classify sarcasm patterns, authors used supervised machine learning techniques, hybrid neural network architectures, ensemble hybrid approaches, and standard models implementing word embedding. The results show that the deep learning model, specifically the RNN with LSTM, demonstrates the highest level of accuracy of 96%, followed by The RNN utilizing GRU with an accuracy of 95%, and the CNN-RNN hybrid model with an accuracy of 90%. These accuracies are much higher than those achieved by other supervised models.

Sharma et al. [18] proposed a hybrid method for sarcasm detection in social media. The model incorporates sentence embedding generation from Autoencoder, bidirectional encoder representation transformer (BERT), and the universal sentence encoder (USE), aiming to accommodate a wide range of content types. It undergoes verification using three real-world social media datasets, achieving higher accuracy than previous state-of-the-art frameworks. The proposed model employs pre-processing, pre-training, and classification phases, utilizing three sentence-based techniques to provide the final categorization of sarcastic or non-sarcastic comments.

Kavitha et al. [19] developed a novel deep learning-based sarcasm detection classification model (DLE-SDC) that involves a pre-processing stage, Glove-based word vector representation, a combination of convolutional neural network and recurrent neural network (CNN-RNN) based classification, and an algorithm for teaching and learning based optimization (TLBO) for hyperparameter tuning. The model is validated using a benchmark dataset and evaluated based on precision, recall, accuracy, and F1 score.

Sengar et al. [20] introduce a novel method for sarcasm detection in text. It employs feature engineering techniques based on contrasts within sarcastic sentences, focusing on positive phrases followed by negative situations or vice versa in tweets. The dataset, sourced from Twitter, includes 151,900 sarcastic and 330,692 non-sarcastic tweets. The approach involves pre-processing, contrast sentiment feature extraction, pattern recognition, and N-grams feature extraction. A neural network with ReLU activation outperforms traditional machine learning models in accuracy and f1-score for sarcasm detection.

Rao et al. [21] developed a sarcasm detection model using Twitter data. The dataset used for the experiment is the Twitter headlines dataset.It contains around 30k tweets. The data is processed using various Python libraries like Word2Vec, NLTK, SciPy, TextBlob, Plotly, Matplotlib, and NumPy. The pre-processing steps involve data cleaning, tokenization, stemming, lemmatization, and feature extraction using word embeddings. The sarcasm detector utilizes a long short-term memory (LSTM) network that ensures careful monitoring of the input data dependencies. The model can achieve an accuracy rate as high as 92% depending on the dataset and the number of epochs employed. The model classifies input data based on a prediction variable, represented as a two-dimensional array. A front-end engine was created for the RNN model, capable of detecting sarcasm in grammatically correct input and notifying users of incorrect input and spelling errors.

Bhardwaj et al. [22] proposed a deep learning model that uses BERT to pre-process the input sentence for sarcasm detection. The authors report an accuracy of 99% on the news headline dataset, which is a valuable improvement over the previous approach.

Ghosh et al. [23] focused on the class imbalance problem in sarcasm detection on Twitter. The authors extracted Indian tweets and manually classified them as sarcastic or non-sarcastic. They used N-gram features for tokenization and applied the Synthetic Minority Oversampling Technique (SMOTE) to augment the minority class data points. They used two classifiers, BNB and LRC, to Assess the performance of each and make a comparison before and after SMOTE. The results showed that SMOTE was effective when dealing with the imbalance in class issues and enhancing the classifier performance. BNB reached maximum accuracy with a low-level oversampling rate, while LRC achieved the same precision with a 50% oversampling rate.

Mohan et al. [24] address the challenge of detecting sarcasm in textual data, particularly in the context of customer feedback and social media. The paper proposes a hybrid model that combines Bidirectional Encoder Representations from Transformers (BERT) and Graph Convolutional Networks (GCN) for accurately detecting sarcasm. The model utilizes affective and adjacency graphs, BERT embeddings, and GCN to capture advanced structural and semantic patterns within sarcastic content. Experimental results showcase the outstanding performance of the proposed approach compared to baseline approaches, achieving high accuracy and f1 scores on publicly available sarcasm datasets.

Sharma et al. [25] introduce a hybrid ensemble approach designed for detecting sarcasm on social networking platforms. The model combines word-based and sentence-based models such as GloVe, Word2Vec, and BERT embeddings with a fuzzy logic module to overcome the limitations of individual methods and improve sarcasm detection accuracy. The study evaluates the model using real-world datasets from Twitter and Reddit, demonstrating its applicability and potential in practical and business contexts.

Kumar et al. [26] propose a hybrid deep-learning approach for detecting sarcasm in Hindi text and addressing challenges in languages with limited resources. The model utilized word-emojis embedding and demonstrated high accuracy and F-score in detecting sarcasm in Hindi tweets. The paper discusses tweet pre-processing, word embeddings for semantic relationships, and the representation of emojis as vectors (Emoji2vec). The hybrid model combines convolutional neural networks (CNNs) and long short-term memory (LSTM) networks to capture local and long-term features in the text.

Hasnat et al. [27] propose an automated system for detecting sarcasm in social media, specifically on Twitter. The study utilizes techniques such as Bag of Words, TF-IDF, and word embeddings to extract features from the text and employs a Long Short-Term Memory (LSTM) network for identifying sarcasm. The results demonstrate the effectiveness of the LSTM model with word embeddings, achieving a high accuracy score of 99.01%.

Ali et al. [28] introduce a novel deep-learning model called GMP-LSTM (Global Max Pooling with Long-Short Term Memory) for sarcasm identification in news headlines. The model replaces the LSTM layer with a GlobalMaxPool1D layer, incorporates dense layers with ReLU activation functions, and achieves promising accuracy in detecting sarcasm.

2.3. MULTILINGUAL SENTIMENT ANALYSIS

Khanvilkar et al. [29] proposed a system to determine the polarity of product reviews using ordinal classification and provide recommendations to users based on the achieved polarity. The methodology involves language standardization, preprocessing, applying machine learning algorithms such as SVM and Random Forest, determining polarity, and using sentiment analysis and user profile information for product recommendation.

Londhe et al. [30] proposed a hybridized model that combines the Social Eagle Algorithm and deep bidirectional long short-term memory (BiLSTM) to predict sentiment effectively. The model uses Transliteration to standardize the different languages and extract features from the data. The simulation results show that the proposed model outperforms current state-of-the-art methods with an accuracy of 91%, precision of 89%, recall of 91%, and F1 measure of 89%.

Manias et al. [31] performed a comparative study of two multilingual sentiment analysis methods, three multilingual deep learning classifiers, and a zero-shot classification method. The evaluation utilizes English dataset from Kaggle and another multilingual Twitter dataset. The multilingual deep learning classifiers demonstrated high performance, especially those pre-trained and evaluated exclusively on monolingual data, the process of transferring inference when using these models with multilingual data. However, the zero-shot classification method fails to achieve high accuracies in monolingual data compared to multilingual data.

Kanfoud et al. [32] introduced SentiCode, a novel method for sentiment analysis across multiple languages and domains. SentiCode employs a unified model using a singular level of abstraction and pseudocode, avoiding the need for separate models for each language. The approach generates a language-independent code, incorporating shared linguistic attributes like part-of-speech tags to extract sentiment-bearing words. The SentiCode vocabulary includes ADJ, ADV, NOUN, VERB, NOT, POS, and NEG. Experimental evaluations, encompassing English, Arabic, German, and Russian, demonstrated SentiCode's 70% accuracy, striking a balance between efficiency and computational time (around 67 seconds), making it well-suited for real-time applications.

Goel et al. [33] presented a study on developing a language-independent system for sentiment analysis of tweets using NLP and classification algorithms. The study utilizes a Twitter dataset comprising 2000 entries and applies preprocessing steps to filter the data. Additionally, the study employs the Google Translate API to develop a language-independent system. The Data modeling and training using Stanford NLP, then classification using the Naïve Bayes machine learning algorithm.

2.4. MULTILINGUAL SARCASM DETECTION

Rao et al. [34] proposed a sarcasm detection model for Amazon product reviews using machine learning. They collected a dataset, treated each review as a separate document, and applied preprocessing steps like tokenization and stemming. Feature extraction involved techniques such as TFIDF and n-grams. Classification employed Support Vector Machines, K Nearest Neighbors, and Random Forest algorithms. Results showed that Support Vector Machines achieved the highest accuracy at 67.58%, followed by Random Forest (62.34%) and K Nearest Neighbors (61.08%).

Khandagale et al. [35] proposed a system for detecting sarcasm in Hindi-English code-mixed tweets using machine learning and deep learning-based models. The authors compare the performance of various classification models and observe that the Random Forest and Logistic Regression classifiers yielded the highest F-score of 96%. The proposed system includes data pre-processing, feature extraction, feature selection, and classification models. The authors report that the Bi-LSTM encoder with an attention framework yields the best results among deep learning models with an F-score of 95%.

2.5. JOINT SARCASM DETECTION AND SENTIMENT ANALYSIS

Suhaimin et al. [36] proposed a sentiment analysis framework that integrates sarcasm detection. The proposed model comprises six modules: pre-processing, feature extraction, feature selection, initial sentiment classification, sarcasm detection, and final sentiment classification. The evaluation of this framework utilized Malay social media data, with results demonstrating that sarcasm detection leads to better sentiment classification performance. They use non-linear SVM for the initial sentiment classification module.

Modak et al. [37] extensively explore the correlation between sarcasm and sentiment analysis, utilizing Twitter data. They employ an innovative approach involving the fusion of K-means clustering, Principal Component Analysis (PCA), and Support Vector Machine (SVM) classifiers. The research methodology involves several phases: dataset collection and processing, feature extraction and reduction, clustering by K-means, and classification using SVM, Random Forest (RF), and K-nearest neighbor (KNN). The results show that the integration of PCA, K-means, and SVM classifier achieves a precision of 91.67%, recall of 92.67%, and accuracy of 91.90% for sarcasm detection. The study emphasizes the significance of considering textual information and sentiment diffusion patterns in analyzing sentiment in Twitter data.

M. Pal et al. [38] present a comprehensive study on sarcasm detection and sentiment analysis in Bengali. By collecting news headlines and comments from Twitter, the authors create datasets for both tasks. The sarcasm detection phase utilizes an LSTM model with GloVe embeddings, achieving high accuracy and performance. In the sentiment analysis phase, various classifiers undergo training using TF-IDF vectorization and n-gram features, with the best results obtained using the trigram feature and linear SVM.

Sait et al. [39] propose a novel approach called deep learning with natural language processing enabled SA (DLNLP-SA) for detecting sarcasm in textual data and its impact on sentiment analysis. The paper introduces the DLNLP-SA technique, which involves pre-processing, feature vector conversion using the N-gram feature extraction technique, and sarcasm classification using the MHSA-GRU (multi-head self-attention-based gated recurrent unit) model. The study also incorporates MFO (mayfly optimization) for hyperparameter optimization. The proposed DLNLP-SA model demonstrates superior performance compared to existing approaches, achieving an accuracy of 97.61% on a Twitter dataset.

Yin et al. [40] proposed a multi-task deep neural network for joint sarcasm detection and sentiment analysis (MT-SS). The model uses bidirectional gated recurrent units incorporating an attention network module to capture task-specific local feature representations and employing convolutional neural networks to capture global feature representations. The experiments utilized two datasets, and the proposed model outperforms existing approaches on both datasets.

3. COMPARATIVE ANALYSIS

This analysis provides insights into the strengths and weaknesses of different models, showcasing the impact of techniques, embeddings, and datasets on the performance of sentiment analysis and sarcasm detection models across various languages.

Table 1. Comparative Analysis of Models in Sarcasm Detection and Sentiment Analysis

Ref	lang	Dataset	Algorithm	Features	Acc.
[1]	Bangla	social media dataset	CNN-AM	Word embedding	72.06
[2]	English	restaurant reviews	SVM	ontology features	81.19
[3]	Arabic	Arabic Health Services Dataset	BiLSTM	Sent2seq-pre-trained word embedding	92.6
[4]	Bengali	SAIL 2015 Bengali dataset	BILSTM	SentiWordNet	55.73
[5]	Chinese	SMP2019 dataset	FER-Net	TTP (text to picture)	85.64
[6]	English	Weibo towards Movies dataset	non-linear layers	Graph-based	78.3
[7]	VN	Vietnamese Students Feedback	BiLSTM-AM	Word2Vec	86.64
[8]	Arabic	Multi-Dialect restaurant reviews	SVM	TF-IDF	87.7
[9]	English	SentiHood dataset	LSTM-AM	Commonsense Knowledge	89.32
[10]	English	restaurant reviews	LSTM, fuzzy logic	N-Grams	81.04
[11]	English	six microblogs' datasets	LAMP	Word embedding	85.2
[12]	Japanese - English	hotel reviews and movie reviews	BiLSTM-AM	word-embedding	87.2
[13]	English	Amazon product review dataset	CNN and Adversarial model	BERT	88.9
[14]	English	Twitter Dataset	LR -LSTM	XGBoost	73.1
[15]	English	generic sarcasm corpus	HRNN attention	Glove	75.3
[16]	Chinese	Chinese sarcasm dataset	AM and (MLP)	Encoder	69.53
[17]	English	News Headline Dataset	LSTM	TFIDF	96.11
[18]	English	news headlines dataset	Ensemble Model with Fuzzy Logic	Word2Vec, GloVe, and BERT	90.81
[19]	English	News Headlines Dataset	CNN-RNN	Glove	94.05
[20]	English	Twitter dataset	ILR and NN	Contrast Sentiment -N-grams	59
[21]	English	Twitter headlines dataset	LSTM	Word2Vec	92
[22]	English	News Headline Dataset	CNN-LSTM	BERT	99.63
[23]	Indian	786 Indian tweets	LRC	N-Grams	88
[24]	English	News Headlines Dataset	BERT and GCN	BERT	90.7
[25]	English	news headlines dataset	Ensemble Model with Fuzzy Logic	Word2Vec, GloVe, and BERT	90.81
[26]	Hindi	Twitter sarcasm dataset	CNN-LSTM	(CBOW) with Emoji Embedding	97.35
[27]	English	Twitter sarcasm dataset	LSTM	TFIDF and word embedding	99.01
[28]	English	News Headlines dataset	GMP-LSTM	word-embedding	92.5
[29]	Multi	user reviews	RF	Word embedding	95.03
[30]	multi	Twitter sentiment analysis dataset	BiLSTM	TF-IDF	91.57
[31]	Multi	Twitter dataset	MSC-VNN	Word embedding	84.31
[32]	multi	Amazon reviews	MLP	SentiCode-TF-IDF	70
[33]	Multi	Twitter dataset	NB	dictionary modeling	79.4
[34]	English	Amazon product reviews	SVM	TF-IDF, N-grams	67.58
[35]	Hindi-English	Code -mixed dataset	RF or LR	Sarcasm tokens, emoticons, n-grams	96
[36]	Malay-English	social media dataset	non-li SVM	(syntactic pragmatic- and prosodic)	90.5
[37]	English	tweets from Twitter	PCA-K-mean-SVM	component analysis model	91.9
[38]	Bengali	News headlines and News comments	LSTM	GloVe	91.94
[39]	English	Twitter dataset	MHSA-GRU	N-Grams	97.61
[40]	English	Sentiment and sarcasm dataset	Bi-GRU+AMCNN	Glove	92.01

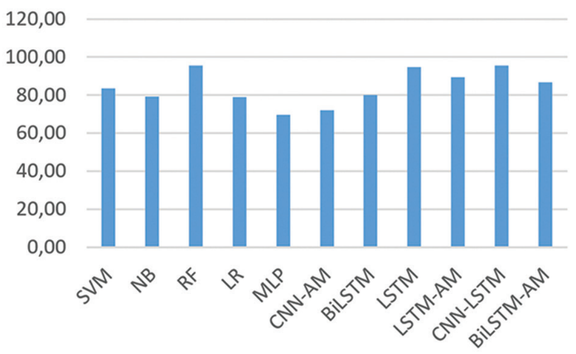

Fig. 1. Average Accuracy Comparison of Machine Learning Algorithms

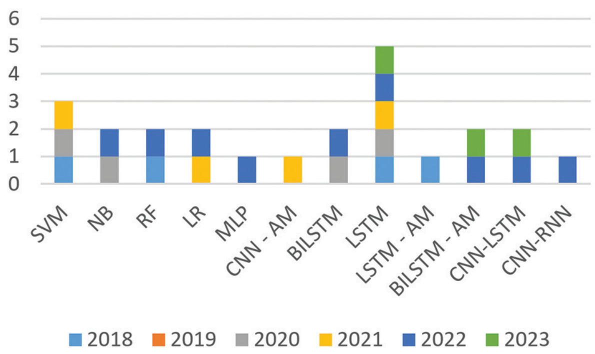

Fig. 2. Evolution of Machine and Deep Learning Algorithms Over the years

The performance of sentiment analysis and sarcasm detection models varies significantly due to factors like language, datasets, and algorithms. Integrating BERT word embedding techniques with LSTM or BiLSTM architectures consistently achieves high accuracy. However, accuracy depends on factors like feature engineering, word embedding algorithm, and model architecture. BERT-based models show promise in sentiment analysis and sarcasm detection. Future trends include the continued improvement of accuracy in sarcasm detection using advanced deep-learning models like BERT and LSTM. There is a focus on developing models capable of detecting sarcasm without explicit indicators. In sentiment analysis, we expect that progress in transformers and attention mechanisms will enhance performance. Multilingual sentiment analysis poses challenges, but we anticipate models that handle code-mixed and dialect-specific sentiment analysis. Integrating external knowledge sources and the increased use of transfer learning with pre-trained language models like BERT and GPT are expected to become more prevalent in sentiment analysis tasks. For sarcasm detection, notable models achieving the best accuracies include the one in [27], which employs LSTM for English Twitter sarcasm detection, achieving an impressive accuracy of 99.01%. This suggests that the model has a good capability to understand and classify sarcastic expressions in English tweets, and the model in [22] using CNN-LSTM with BERT for English News Headline sarcasm detection achieved a very high accuracy of 99.63%, showcasing the power of BERT embeddings in sentiment analysis tasks. For notable performance, models like the Hindi Twitter sarcasm detection model in [26] using CNN-LSTM with (CBOW) and Emoji Embedding achieved an accuracy of 97.35%, indicating its effectiveness in handling sarcasm detection in Hindi tweets, and the sarcasm detection model in [25] for English news headlines, using an ensemble model with fuzzy logic and combining Word2Vec, GloVe, and BERT embeddings, achieved an accuracy of 90.81%. For Multilingual Performance: The model in [35] for the Hindi-English code-mixed dataset, using Random Forest (RF) or Logistic Regression (LR) with features like sarcasm tokens, emoticons, and n-grams, achieved an accuracy of 96%, indicating effectiveness in handling multilingual sentiment analysis. For Dataset Influence: Performance varies across datasets, with some models achieving high accuracy on specific datasets, such as the model in [22] on the News Headline dataset. In sentiment analysis, a comprehensive evaluation of various models reveals intriguing insights. Notably, Random Forest with Word embedding excels with an impressive 95.03% accuracy in multi-user reviews [29], demonstrating its robustness in capturing nuanced sentiments. English Tweets from Twitter achieve high accuracy, notably with PCA-K-mean-SVM [9] and [30] BiLSTM with TF-IDF reaching 91.9% and 91.57% accuracy, respectively. These models showcase the effectiveness of diverse techniques in social media sentiment analysis. In multilingual analysis, the Non-linear SVM with syntactic pragmatic- and prosodic-based features achieves a commendable 90.5% accuracy in Malay-English social media [36]. A prominent trend highlights the consistent success of models leveraging advanced embeddings like Word2Vec, TF-IDF, and BERT. The versatility of techniques such as SVM, BiLSTM, and Random Forest, coupled with nuanced linguistic understanding, proves crucial for high accuracy in sentiment analysis across diverse datasets and languages. Tailoring models to specific linguistic nuances emerges as a good factor for optimal performance [9, 29, 30, 36].

4. CONCLUSION

This paper gives a thorough overview of sentiment classification, focusing on the challenges, especially in detecting sarcasm. Sentiment analysis is crucial for understanding people's opinions. Sarcasm makes Sentiment analysis difficult because it can lead to misclassification. The paper discusses different methods for sarcasm detection and their performance. It also explores the challenges of handling comments in multiple languages and the importance of language identification. The paper compares previous studies on sentiment analysis and sarcasm detection, noting variations in accuracy based on languages, datasets, and algorithms. Deep learning models like BERT and LSTM show promise for high accuracy. The paper identifies trends in sarcasm detection and sentiment analysis, including advanced deep learning, and addressing class imbalances. Transfer learning and pre-trained language models are expected to become more common. The key takeaway is that improving models with advanced techniques and addressing specific challenges will enhance sentiment analysis, aiding decision-making in various fields. Finally, practitioners should prioritize advanced embeddings (Word2Vec, TF-IDF, BERT) and versatile techniques (SVM, BiLSTM, Random Forest) for optimal sentiment analysis. Ethical considerations, including bias mitigation, transparent model decisions, and continuous monitoring, are essential to ensure responsible and effective sentiment analysis practices.

5. REFERENCES

[1] S. Sharmin, D. Chakma, "Attention-based convolutional neural network for Bangla sentiment analysis", AI & Society, Vol. 36, No. 1, 2021, pp. 381-396

[2] S. de Kok, L. Punt, R. van den Puttelaar, K. Ranta, K. Schouten, F. Frasincar, "Review-aggregated aspect-based sentiment analysis with ontology features", Progress in Artificial Intelligence, Vol. 7, No. 4, 2018, pp. 295-306.

[3] H. Elfaik, El H. Nfaoui, "Deep Bidirectional LSTM Network Learning-Based Sentiment Analysis for Arabic Text", Journal of Intelligent Systems, Vol. 30, No. 1, 2021, pp. 395-412.

[4] K. Sarkar, "Sentiment Analysis of Bengali Tweets Using Deep Learning", Proceedings of the International Conference on Computational Intelligence in Data Science, Chennai, India, 20-22 February 2020, pp. 71-84.

[5] M. Chen, K. Ubul, X. Xu, A. Aysa, Muhammat, Mahpirat, "Connecting Text Classification with Image Classification: A New Preprocessing Method for Implicit Sentiment Text Classification", Sensors, Vol. 22, No. 5, 2022.

[6] X. Zhao, Y. Liu, Z. Jin, "Multiplex network embedding for implicit sentiment analysis", Complex and Intelligent Systems, Vol. 7, No. 6, 2021, pp. 3179-3193.

[7] Y. Huang, S. Liu, L. Qu, Y. Li, "Correction to: Effective Vietnamese Sentiment Analysis Model Using Sentiment Word Embedding and Transfer Learning", Proceedings of the International Conference of Pioneering Computer Scientists, Engineers and Educators, Taiyuan, China, 18-21 September 2020, pp. 36-46.

[8] A. Hussein, I. Moawad, R. Badry, "Arabic Sentiment Analysis for Multi-dialect Text using Machine Learning Techniques", International Journal of Advanced Computer Science and Applications, Vol. 12, No. 12, 2021.

[9] Y. Ma, H. Peng, E. Cambria, "Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM", Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, No. 1, 2018.

[10] A. Londhe, P. V. R. D. P. Rao, "Dynamic Classification of Sentiments from Restaurant Reviews Using Novel Fuzzy-Encoded LSTM", International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 10, No. 9, 2022, pp. 112-124.

[11] D. Li, X. Luo, X. Wei, R. Xue, "Emojis Sentiment Analysis Based on Big Social Media Data", Proceedings of the International Conference on Applications and Techniques in Cyber Security and Intelligence, Shanghai, China, 11-13 July 2018, pp. 56-63.

[12] Y. Hiyama, H. Yanagimoto, "Word polarity attention in sentiment analysis", Artificial Life and Robotics, Vol. 23, No. 3, 2018, pp. 311-315.

[13] N. Liu, J. Zhao, "A BERT-Based Aspect-Level Sentiment Analysis Algorithm for Cross-Domain Text", Computational Intelligence and Neuroscience, Vol. 2022, 2022, p. 8726621.

[14] R. Bagate, S. Ramadass, "Sarcasm detection of tweets without sarcasm: Data science approach", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 33, 2021, pp. 993-1001.

[15] Q. Zhang, G. Chen, D. Chen, "Adversarial Training for Sarcasm Detection", Proceedings of the International Conference on Cognitive Computing, Seattle, WA, USA, 25-30 June 2018, pp. 43-54.

[16] L. Zhang, X. Zhao, X. Song, Y. Fang, D. Li, H. Wang, "A Novel Chinese Sarcasm Detection Model Based on Retrospective Reader", Proceedings of the International Conference on Multimedia Modeling, Phu Quoc, Vietnam, 6-10 June 2022, pp. 267-278.

[17] T. Sharma, N. Rani, A. Mittal, N. Rathee, "Sarcasm Detection in Social Media Using Hybrid Deep Learning and Machine Learning Approaches", Proceedings of the International Conference on Artificial Intelligence and Speech Technology, Delhi, India, 12-13 November 2022, pp. 455-464.

[18] D K. Sharma, B. Singh, S. Agarwal, H. Kim, R. Sharma, "Sarcasm Detection over Social Media Platforms Using Hybrid Auto-Encoder-Based Model", Electronics, Vol. 11, No. 18, 2022.

[19] K. Kavitha, S. Chittineni, "An Intelligent Metaheuristic Optimization with Deep Convolutional Recurrent Neural Network Enabled Sarcasm Detection and Classification Model", International Journal of Advanced Computer Science and Applications, Vol. 13, No. 2, 2022.

[20] C. Sengar, J. Nirmala, "Sarcasm Detection in Tweets as Contrast Sentiment in Words Using Machine Learning and Deep Learning Approaches", Proceedings of the International Conference on Machine Learning, Image Processing, Network Security and Data Sciences, Singapore, Silchar, India, 30-31 July 2020, pp. 73-84.

[21] R. P. Rao, S. Dayanand, K. R. Varshitha, K. Kulkarni, "Sarcasm Detection for Sentiment Analysis: A RNN-Based Approach Using Machine Learning", Proceedings of High Performance Computing and Networking, 2022, pp. 47-56.

[22] S. Bhardwaj, M. Prusty, "BERT Pre-processed Deep Learning Model for Sarcasm Detection", National

Academy Science Letters, Vol. 45, No. 2, 2022, pp. 203- 208.

[23] K. Ghosh, A. Banerjee, M. Bhattacharjee, S. Chatterjee, "Improved Twitter Sarcasm Detection by Addressing Imbalanced Class Problem", Vol. 165, 2021, pp. 135-145.

[24] A. Mohan, A. M. Nair, B. Jayakumar, S. Muraleedharan, "Sarcasm Detection Using Bidirectional Encoder Representations from Transformers and Graph Convolutional Networks", Procedia Computer Science, Vol. 218, 2023, pp. 93-102.

[25] D. Sharma, B. Singh, S. Agarwal, N. Pachauri, A. Alhussan, H. Abdallah, "Sarcasm Detection over Social Media Platforms Using Hybrid Ensemble Model with Fuzzy Logic", Electronics, Vol. 12, 2023, p. 937.

[26] A. Kumar, S. R. Sangwan, A. K. Singh, G. Wadhwa, "Hybrid Deep Learning Model for Sarcasm Detection in Indian Indigenous Language Using Word-Emoji Embeddings", ACM Transactions on Asian and Low-Resource Language Information Processing, Vol. 22, No. 5, 2023, p. 25.

[27] H. Saleem, A. Naeem, K. Abid, N. Aslam, "Sarcasm Detection on Twitter using Deep Handcrafted Features", Journal of Computing & Biomedical Informatics, Vol. 4, No. 2, 2023, pp. 117-127.

[28] R. Ali, T. Farhat, S. Abdullah, S. Akram, M. Alhajlah, A. Mahmood, M. Iqbal, "Deep Learning for Sarcasm Identification in News Headlines", Applied Sciences, Vol. 13, No. 9, 2023.

[29] G. Khanvilkar, D. Vora, "Sentiment Analysis for Product Recommendation Using Random Forest", International Journal of Engineering and Technology, Vol. 7, 2018, pp. 87-89.

[30] D. Londhe, A. Kumari, "Multilingual Sentiment Analysis Using the Social Eagle-Based Bidirectional Long Short-Term Memory", International Journal of Intelligent Engineering and Systems, Vol. 15, 2022, pp. 479-493.

[31] G. Manias, A. Kiourtis, A. Mavrogiorgou, D. Kyriazis, "Multilingual Sentiment Analysis on Twitter Data Towards Enhanced Policy Making", Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations, Hersonissos, Greece, 17-20 June 2022, pp. 325-337.

[32] R. Kanfoud, A. Bouramoul, "SentiCode: A new paradigm for one-time training and global prediction in multilingual sentiment analysis", Journal of Intelligent Information Systems, Vol. 59, 2022.

[33] P. Goel, V. Goel, A. K. Gupta, "Multilingual Data Analysis to Classify Sentiment Analysis for Tweets Using NLP and Classification Algorithm", Proceedings of Advances in Data and Information Sciences, 2020, pp. 271-280.

[34] M. V. Rao, Sindhu C., "Detection of Sarcasm on Amazon Product Reviews using Machine Learning Algorithms under Sentiment Analysis", Proceedings of the Sixth International Conference on Wireless Communications, Signal Processing and Networking, Chennai, India, 25-27 March 2021, pp. 196-199.

[35] K. Khandagale, H. Gandhi, "Sarcasm Detection in Hindi-English Code-Mixed Tweets Using Machine Learning Algorithms", Proceedings of the International Conference on Computing in Engineering & Technology: Applied Computational Technologies, Lonere, India, 12-13 February 2022, pp. 221-229.

[36] M. Suhairi, M. S. Md Suhaimin, M. Hanafi, M. Hijazi, R. Alfred, F. Coenen, "Modified framework for sarcasm detection and classification in sentiment analysis", TELKOMNIKA Indonesian Journal of Electrical Engineering, Vol. 13, 2019, pp. 1175-1183.

[37] S. Modak, A. Mondal, "Sentiment Analysis of Twitter Data Using Clustering and Classification", Proceedings of Third International Conference on Computing, Communications, and Cyber-Security, Vol. 421, 2023, pp. 651-664.

[38] M. Pal, R. Prasad, "Sarcasm Detection followed by Sentiment Analysis for Bengali Language: Neural Network & Supervised Approach", Proceedings of the International Conference on Advances in Intelligent Computing and Applications, Kochi, India, 1-3 February 2023, pp. 1-7.

[39] A. R. W. Sait, M. K. Ishak, "Deep Learning with Natural Language Processing Enabled Sentimental Analysis on Sarcasm Classification", Computer Systems Science and Engineering, Vol. 44, 2023, pp. 2553-2567.

[40] C. Yin, Y. Chen, W. Zuo, "Multi-Task Deep Neural Networks for Joint Sarcasm Detection and Sentiment Analysis", Pattern Recognition and Image Analysis, Vol. 31, 2021, pp. 103-108.