Informatics Bulletin, Faculty of Computers and Artificial Intelligence, Helwan University
Published Online *Vol 6 Issue 2, July 2024*
(https://fcihib.journals.ekb.eg)

A Survey of Challenges, Methods, and Trends in Sentiment Analysis and Sarcasm Detection

Ahmed Derbala Yacoub[1], and Amal Elsayed Aboutabl[1], Salwa O. Slim[1],

[1]Computer Science Department, Faculty of Computers and Artificial Intelligence Helwan University, Cairo, Egypt. ahmed_drbala_1073@fci.helwan.edu.eg , amal.aboutabl@fci.helwan.edu.eg , salwaosama@fci.helwan.edu.eg

***Abstract*—This survey paper explores the intersection of Natural Language Processing (NLP), particularly sentiment analysis (SA) and sarcasm detection (SD), within machine learning. It emphasizes the challenges inherent in detecting sarcasm due to its implicit nature and contextual dependencies. The paper comprehensively examines the field, conducting a comparative analysis of various studies ranging from traditional machine learning to advanced deep learning methods. Identified challenges include sarcasm's implicit nature, contextual impact, difficulties with short-text datasets, and the challenge of detecting both sarcasm and sentiment within a multilingual context, especially in low-resource languages. The survey discusses innovative solutions proposed by researchers, spanning classical machine learning to advanced deep learning approaches, such as aspect-based sentiment analysis and advanced transformer models. The methodology, which utilized bidirectional encoder representation transformers (BERT) for aspect-based sentiment analysis, achieves exceptional F1 scores on Twitter and Reddit datasets. Artificial neural networks demonstrate flawless accuracy in detecting sarcasm within emoticons and text. The usage of a modified switch transformer attains an impressive accuracy on the ArSarcasm dataset, highlighting the efficacy of its dynamic switching mechanism. Additionally, incorporating bidirectional LSTM results in significant performance improvements in sentiment and sarcasm classification tasks, evidenced by notable F1 scores.**

***Index Terms*—sarcasm detection, sentiment analysis, natural language processing, machine learning, deep learning.**

I. Introduction

This paper aims to conduct a comprehensive examination of current research on understanding human emotions and detecting sarcasm in both single and multiple languages. It explores the objectives, methodologies, data, testing procedures, outcomes, strengths, and weaknesses of existing research efforts, outlining challenges in sentiment classification and detection of sarcasm while delving into innovative solutions proposed by researchers. Additionally, it seeks to identify effective solutions and anticipate future developments in the fields of sentiment classification and the detection of sarcasm. Understanding and interpreting sentiments as well as detecting sarcasm in textual data are pivotal tasks in the field of natural language processing, particularly given the exponential growth of digital content. Natural language processing is a sub-field of machine learning that aims to enable computers or machines to understand and analyze human language [1]. Sentiment Analysis is used for many applications such as product review analysis, customer feedback, digital marketing, social media monitoring, etc.[2]. Sentiment analysis, known as opinion extraction or opinion mining, is a sub-field of natural language processing responsible for determining whether a given text is positive, negative, or neutral [3]. The existence of sarcasm in a text is a big challenge in the sentiment analysis process since it can ultimately change the meaning of a given text, which leads to errors in the sentiment analysis results if it is not considered. Sarcasm can change the conveyed sentiment from positive to negative [4]. People usually use sarcastic words to express their negative sentiments, so sentiment analysis and sarcasm detection are very correlated [5]. It is a form of language where the expressed meaning is different from the intended meaning, and this can lead to difficulty in detecting sarcasm [6,7]. Sarcasm is a type of negation with the absence of an explicit negation marker or it is known as a positive sentiment polarity with a negative intention. In sentiment analysis and sarcasm detection, it is a major challenge to correctly categorize sarcastic text as negative [8]. Sarcasm detection is a text classification sub-field that is responsible for identifying sarcasm instances in text. It is responsible for categorizing whether a given sentence contains sarcastic text. The implicit nature of sarcasm and sentimental expressions leads to difficulties in identifying them by traditional NLP functions. The absence of context is a big challenge in sarcasm detection and sentiment analysis since it is significant to understand the context in which the text was written [9]. In written language or text communication, the nonexistence of information such as body gestures, facial expressions, and the speaker's emphasis on words makes it difficult for machines to detect sarcasm [10]. Machine learning models can easily detect explicit sarcasm in text in a good manner. However, the implicit nature, context dependency of sarcasm, and word dependency make it difficult to detect sarcasm through traditional machine learning models [11]. More work has been done in the English language, a high-resourced language, that has more resources available on the Internet. However, With low-resource languages, there is a problem while training and testing text classification models [12]. Another challenge facing sentiment analysis tools is how to analyze and detect sentiment and sarcasm in multilingual textual data. Users use their mother language to express sentiments and emotions which leads to the need for a tool that is responsible for handling multiple language contexts. In sentiment analysis, the accuracy has dropped when excluding non-English text and not considering multilingual texts, where very valuable data are not considered. Also, code-mixed text must be addressed and solved [13].

The challenges in sentiment analysis and sarcasm detection are:

- Detecting sarcasm holds significant importance in sentiment analysis as it can notably influence the perceived sentiment of a text.

- The implicit nature of sarcasm poses complexities for traditional natural language processing (NLP) functions, introducing challenges in accurate detection.
- The challenge of accurately categorizing sarcastic text, especially in identifying it as negative, is a notable difficulty due to the intricate interaction between sarcasm and sentiment.
- The absence of context presents a challenge in sarcasm detection, underscoring the importance of understanding the circumstances surrounding the text. Challenges in written communication arise due to the absence of cues such as gestures and facial expressions available in verbal communication, hindering sarcasm detection.
- Traditional machine learning models encounter difficulties with sarcasm due to its dependency on specific words and context, limiting detection accuracy.
- Resource scarcity in low-resource languages impedes the development of effective sentiment analysis and sarcasm detection models.
- Analyzing sentiment and sarcasm across multiple languages proves challenging, necessitating adaptability to diverse linguistic expressions.
- The exclusion of non-English text impacts accuracy, neglecting valuable data and potentially introducing bias.
- the challenge of analyzing sentiment and sarcasm in short texts demands specialized attention due to the concise nature of the datasets.

This review delves into the forefront of research, surveying recent studies in sentiment Analysis and sarcasm detection models. The exploration encompasses an array of methods and approaches employed by researchers, ranging from classical machine learning techniques to cutting-edge deep learning architectures. The approach to sentiment classification or sarcasm detection involves data preparation and preprocessing, followed by feature extraction to convert text into a numerical format suitable for machine learning. Classifier algorithms are then trained on these features to classify text sentiments or detect sarcasm, and their performance is evaluated using appropriate metrics. Figure 1 describes the general methodology for sarcasm detection and sentiment classification tasks. Researchers such as Taha et al. [4] proposed a method leveraging aspect-based sentiment analysis and bidirectional encoder representation transformers to detect sarcasm. This involves a pre-processing phase with hashtag segmentation and the utilization of Glove and FastText embeddings for feature representation, achieving notable F1 scores on Twitter and Reddit datasets. Shah et al. [1] introduced a novel approach with a modified switch transformer for sarcasm detection, showcasing competitive performance on the ArSarcasm dataset. Patrick et al. [6] presented a baseline model integrating sentiment classification and detection of sarcasm using deep learning (CNN) and linear model approaches (SVM), revealing insights into the impact of sentiment analysis on sarcasm detection. Abdelkader et al. [2] proposed a multi-task learning model incorporating a MARBERT encoder, achieving top rankings in sentiment classification and detection of sarcasm tasks. These studies represent a snapshot of the diverse methodologies explored, ranging from transfer learning and multi-task learning to innovative deep learning architectures. The subsequent sections of this review will delve into each of these studies, providing detailed insights into their methodologies, performances, and contributions to the evolving landscape of sentiment classification and detection of sarcasm.

The main contributions of this paper are the following:

- it conducts a comprehensive review of current research, encompassing studies in both single and multiple languages. The analysis examines differences in objectives, methodologies, data utilization, testing procedures, beneficiaries, and outcomes across these studies.
- it identifies the strengths and weaknesses of the approaches used, providing valuable insights into their effectiveness by evaluating and analyzing their results.
- it tackles challenges in sentiment classification and detection of sarcasm, showcasing novel solutions put forth by researchers. It serves as a practical resource for both researchers and practitioners by identifying the most effective solutions.
- it adopts a forward-looking perspective, anticipating future developments in sentiment classification and detection of sarcasm. It offers a roadmap for potential research directions, providing valuable insights for future studies in these fields.

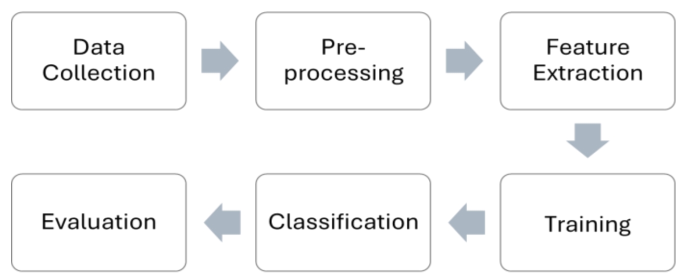

Figure 1: General Method for sentiment classification and Detection of sarcasm

II. Literature Review

The literature review explores recent advancements in sentiment classification and detection of sarcasm Models, crucial components in understanding human communication within the burgeoning digital landscape. With the surge in textual data from social media and online platforms, the need for robust models become paramount. This review covers a range of methodologies, from traditional machine learning to state-of-the-art deep learning approaches. The synthesis of diverse studies aims to provide a comprehensive analysis of the current landscape, offering insights for researchers and practitioners navigating the complexities of sentiment classification and the detection of sarcasm.

II.1 Traditional machine learning

M. V. Rao et al. [3] proposed a model for detecting sarcasm in Amazon product reviews using machine learning algorithms for sentiment analysis. The research begins by collecting a dataset from Amazon, treating each review as a separate document. The data undergoes preprocessing steps, including tokenization, stop word removal, stemming, and lemmatization. Feature extraction follows, employing techniques like term frequency, inverse document frequency(TFIDF), and n-grams. In the classification phase, Support Vector Machines, K Nearest Neighbors, and Random Forest algorithms are utilized. The results, measured in terms of accuracy, reveal that Support Vector Machines achieve the highest accuracy at 67.58%, followed by Random Forest (62.34%) and K Nearest Neighbors (61.08%).

De Kok et al. [5] designed an innovative approach for aspect-based sentiment analysis (ABSA). Their method uses ontology features to improve the accuracy of ABSA models. They utilized a dataset of hotel reviews to train and assess their model, also conducting comparisons with other advanced models. ABSA is a specialized area within sentiment analysis that seeks to determine sentiments expressed toward specific aspects or attributes of a subject. The sentiment classification model utilizes linear Support Vector Machines (SVMs) in two algorithms: a review-based approach and a sentence aggregation method. Leveraging a restaurant ontology with Entity, Property, and Sentiment classes, the model integrates domain-specific knowledge for sentiment classification of specific aspects. Feature generators, adaptors, and parameter optimization enhance the model's performance, with evaluation based on F1 scores and feature selection using tenfold cross-validation.

Goel et al. [12] developed a language-independent system for sentiment analysis on Twitter dataset by translating non-English tweets into English through Google API translator. They chose English because it is considered a high-resourced language. They used Stanford NLP which is the modeling of the English language and has very good knowledge and has all possible combinations for training. Tweets are in an unstructured format, So They process tweets by converting them into meaningful data. The processed tweets are converted to numerical representation utilizing dictionary modeling and feature extraction. They utilized Na¨ıve Bayes as a classifier and for performance evaluation, a confusion matrix is used. Also, they conducted a comparative analysis with the same experimental dataset size for some previous work and obtained that the proposed model has the highest precision value.

Arun et al. [13] presented a sentiment analysis algorithm to classify multilingual tweets into positive, negative, and neutral categories. The proposed algorithm uses natural language processing techniques to detect and translate non-English tweets into English and reduce data sparsity. The algorithm achieved up to 95% accuracy using support vector machines(SVM). The study collected 200 tweets from Twitter API with 65 – 80% in English and the rest in local languages such as Telugu and Hindi. The data set was pre-processed by removing noise data and using regular expressions to correct misspelled words. The ML algorithm included Multinomial Na¨ıve Bayes, Logistic Regression, SVM, Decision Tree, k- Nearest Neighbor, and Random Forest. The results showed that the proposed algorithm outperformed pre-processed tweets in terms of precision, recall, and F1 scores.

Zahedi et al. [14] developed a multilingual sentiment mining system to predict governance. The study focused on Twitter data and collected a dataset of tweets about the government, which was composed of English, Urdu, and Roman Urdu languages. The dataset was cleaned and pre-processed, and five different classification models were applied to determine sentiments. The Logistic Regression model demonstrated the highest accuracy, reaching 75%, closely followed by the Linear Support Vector classifier and the Stochastic Gradient Descent model, both achieving an accuracy of 74%. In the final comparison, both the Multinomial Na¨ıve Bayes and Complement Na¨ıve Bayes models achieved an accuracy of 73%. The study concludes that the ensemble technique could be used to move forward with the acquired results.

Keith et al. [15] presented a baseline model for sentiment classification and detection of sarcasm for the Maltese, English, and Maltese-English languages. Maltese is considered a low-resourced language, English is a high-resourced language, and Maltese-English is a code-switched language. For the dataset, they used the yearly government budget of Malta, which is a multilingual social opinions dataset. In the preprocessing phase, First, they performed data cleaning by Removing any numbers, HTML/XML tags, special characters, and whites-paces. Second, they performed tokenization and stemming. Third, they extracted features by term frequency-inverse document frequency (TF-IDF to convert words into a numerical representation. They utilized traditional supervised machine learning models including, Support Vector Machines, Logistic Regression, Na¨ıve Bayes, Random Forest, and Decision Trees, to build a baseline for sentiment classification and detection of sarcasm tasks. Their experimental results stated that the Complement Na¨ıve Bayes (CNB) obtained the best F1 score for the sentiment analysis task, whereas Nu-Support Vector Classification (Nu-SVC) produced the best results for the sarcasm detection task.

II.2 Deep learning

Shah et al. [1] proposed a novel method for detecting sarcasm by utilizing an adapted switch transformer, incorporating stochastic projections (Embedding utilizing Variational Spatial Gated Units) and Variational Enmesh Experts routing. A switch transformer represents a type of neural network type that uses dynamic switching to decide what is the appropriate number of attention heads in each layer which can help in implementing the computational efficiency and overall performance of the model. The Switch Transformer routing mechanism is responsible for dynamically determining the number of attention heads to use at each layer in the network. The spatial gated unit is a mechanism that is responsible for controlling the flow of information on the network and deciding which features to use. For performance and evaluation, they evaluated their proposed model on the ArSarcasm dataset, and they obtained an accuracy of .83 for the sarcasm detection task and 0.52 for the sentiment analysis task. Additionally, they compared their model with the state-of-the-art models on the same dataset.

Abdelkader, et al. [2] proposed a multi-task learning model that uses a MARBERT encoder, which is a pre-trained BERT-based model, for generating word embedding. The model combines a Multi-task Attention Module for interaction that involves task-specific attention layers to generate contextual- ized word embedding, and to allow interaction between tasks and sharing knowledge for both sentiment classification and detection of sarcasm tasks. The model also includes a classifier task for each task of sentiment classification and detection of sarcasm, which comprises one hidden layer and one output layer. These classifiers receive input by concatenating the embedding of the pooled output and the task-specific output from the multi-task attention module. Performance evaluation is conducted on the ArSarcasm dataset, showing that the proposed model ranks top in sentiment analysis and fourth in sarcasm detection among submitted systems.

Taha et al. [4] proposed a method for sarcasm detection that utilized aspect-based sentiment analysis and bidirectional encoder representation transformer approaches to detect the relation or connection between context input and output and determine whether the output is sarcastic or not. In the preprocessing phase, they applied hashtag segmentation to save context metadata for sentences and removed all insignificant data like usernames and URLs. For Feature representation, they employed Glove embedding and fastest embedding. For performance and evaluation, they obtained an F1-score of 0.734 on the Reddit dataset and 0.73 on the Twitter dataset and compared their model with state-of-the-art on both datasets.

Patrick et al. [6] presented a baseline model for sarcasm detection that combines the task of sentiment analysis through transfer learning to address the effects of sentiment analysis on sarcasm detection. The study utilized both deep learning (CNN) and linear model approaches (SVM). For the CNN model, the standalone CNN without transfer learning of sentiment analysis performed better in terms of F1-score compared to the CNN model integrating sentiment analysis. In the case of the SVM model, there was a slight improvement in F1-score, precision, and recall when sentiment analysis was integrated. However, the differences seem relatively small.

Cignarella et al. [7] investigate how dependency-based syntactic features impact the accuracy of irony detection across multiple languages, including Spanish, French, Italian, and English. In this study, the pre-processing and feature extraction phase focuses on preparing data for irony detection through steps such as URL stripping and character normalization. The study introduces novel features, particularly emphasizing the impact of the Universal Dependencies (UD) format on syntactic information. Various features, including n-grams, char-grams, and dependency relations, are extracted, with attention to features that capture negation based on morpho-syntactic cues. The paper then explores a diverse set of models, including traditional classifiers like SVM and LR, neural networks like GRU, and advanced models such as Multilingual BERT. Thorough hyperparameter tuning and the exploration of different embedding, including fastText and dependency-based word2vec, are conducted. The study comprises three main sets of experiments: feature selection, evaluation of dependency-based word embedding, and integration of syntactic features with BERT. Notably, the selection of the best features demonstrated the prominence of syntactic elements, with Italian exhibiting the highest F1 score and emphasizing the impact of resource quality on performance. Dependency-based word embedding, particularly fastText, generally outperformed other approaches, showcasing their efficacy in capturing irony-related nuances. Incorporating syntactic features into the BERT model consistently improved performance across languages, underlining the importance of syntax in irony detection.

Tan et al. [8] introduce a multi-task learning framework that combines sentiment classification and the detection of sarcasm using a deep artificial neural network. By leveraging a Bidirectional LSTM network, the framework achieves significant improvements in sentiment and sarcasm categorization compared to standalone models. The sentiment analysis dataset, extracted from Twitter, and the sarcasm dataset, obtained from the same platform, provide the train- ing data. The results demonstrate an F1-score of 94% for sentiment classification and 93% for sarcasm classification, showcasing the effectiveness of the proposed approach. However, the performance on unbiased datasets, such as those from Reddit and Twitter, remains a challenge due to the difficulty in accurately classifying neutral sentiment.

Gupta Shaina, et al. [9] addressed sentiment analysis challenges in social media by focusing on the detection of sarcasm in both emoticons and textual content. The authors proposed a system employing artificial neural networks (ANN) to classify positive and negative polarities related to emoticons and sarcasm, achieving an impressive 100% accuracy. Their method involves preprocessing steps like stop word removal and filtration on raw test data. Additionally, the generation of values for the filtered text document likely involves assigning numerical values or representations to the words in the text, followed by training the neural network for polarity classification and sarcasm detection. Notably, the comparison with existing work, particularly a lexicon-enhanced rule-based approach, highlights the superior accuracy achieved by the proposed artificial neural network (ANN) method.

Han et al. [10] developed Cross-lingual learning for sarcasm detection in both Arabic and English languages, which uses pre-trained language models like ERNIE-M and DeBERTa to detect sarcasm in Arabic and English texts. The system ranked well in the SemEval-2022 Task 6 dataset, and the paper describes the approach used for the different sub-tasks. The paper employs a multilingual learning method and k-fold train- ing with ensemble techniques. This SemEval task necessitates identifying sarcasm in either a single sentence or pairs of sentences across various language settings, encompassing three distinct sub-tasks. For Task A (sarcasm detection), ERNIE- M (monolingual) performs well in both English and Arabic, while DeBERTa excels in English. For Task B (ironic speech category classification in English), ERNIE-M (monolingual) achieved the best performance. For Task C (determining the sarcastic text among rephrased pairs in English and Arabic), ERNIE-M (multilingual) showed competitive results in both languages, and DeBERTa performed well in English.

Murali et al. [11] employed a deep learning model for detecting sarcasm in the sentiment classification process using bi-directional long short-term memory(Bi-LSTM) which is a type of recurrent neural network (RNN). They utilized the Bi-LSTM model since it can retain numerical representations from previously processed information in the training process, and captures dependencies between words. The experiment is performed on manually and automatically annotated datasets. The results of the experiments indicated that the proposed model achieved better results than the compared algorithms for the vast majority of the selected hashtags.

Ma et al. [16] introduced a model named Attentive LSTM with CommonSense Knowledge (ALCSK). This model innovatively integrates a pre-trained commonsense knowledge graph into the aspect embedding layer. The results revealed that the proposed ALCSK model surpasses existing state-of-the-art techniques on benchmark datasets, showcasing its efficacy in managing domain-specific knowledge and integrating common-sense knowledge. The primary focus of this model is to augment the accuracy of Targeted Aspect-Based Sentiment Analysis (TABSA) methods by incorporating common-sense knowledge. The study specifically demonstrated notable success on the English SentiHood dataset, achieving an accuracy of 89.32% with the integration of LSTM-AM and common-sense Knowledge Embedding.

Souza et al. [17] presented a comprehensive series of experiments regarding diverse methods for consolidating the produced features in the BERT output layer, emphasizing the task of sentiment analysis in the context of the Brazilian Portuguese language. The study includes BERT models fine-tuned on corpora in Brazilian Portuguese and the multilingual version, considering various aggregation approaches and publicly available datasets partitioned into predefined training, validation, and test sets. Two foundational models, namely m-BERT and BERTimbau, were employed in conjunction with two design approaches: pre-trained and fine-tuned. The results of the experiments indicated that BERT consistently outperformed TF-IDF in terms of ROC-AUC values across most scenarios. Nevertheless, it's worth noting that TF-IDF strikes a favorable balance between predictive performance and computational overhead. The best result of ROC-AUC (%) is 98.4, achieved by the fine-tuned BERTimbau model on the UTLC-Apps dataset.

Savini et al. [18] explored the use of BERT pre-trained language models for sarcasm detection, an important task in natural language processing. The authors establish robust benchmark models for sarcasm detection and propose a framework for transfer learning that fine-tunes BERT models on relevant intermediate tasks, specifically sentiment categorization and emotion recognition. The authors experiment on three datasets and show that their Models based on BERT consistently outshine numerous earlier models. The pre-trained IMDB dataset model attains a leading F1-Score of 80.85%, underscoring the efficacy of BERT through intermediate task transfer learning, revealing a correlation between sarcasm and sentiment.

Parveen et al. [19] introduce a CNN-based classification model for detecting sarcastic text. By integrating implicit and clear representations of brief or short text, the model improves sarcasm classification accuracy. The proposed model employs feature engineering, feature selection, and classification processes, utilizing CNNs and sub-networks to filter word concepts and extract character features. Conceptualization of brief or short text is achieved through a knowledge base, with predefined word embedding and learned character embedding used for representation. Evaluation of Amazon and Twitter datasets demonstrates the CNN model's superior performance, outperforming other classifiers in terms of precision and recall.

Moholkar et al. [20] proposed a sentiment classification model that uses a recurrent neural network (RNN) with Long Short-Term Memory (LSTM) units and Word2Vec embedding. The system consists of three main stages: pre-processing, word embedding, and classification. In the pre-processing stage, the input text is pre-processed to remove stop words, and punctuation, and to convert the text to lowercase. In the word embedding stage, the pre-processed text is transformed into a fixed-size vector representation using Word2Vec embedding. Finally, the classification stage uses an RNN-LSTM model to categorize the sentiment of the input text as positive or negative. The proposed system achieves high accuracy on various benchmark datasets and outperforms traditional machine learning models for sentiment classification.

Omran et al. [21] tackle the scarcity of Arabic dialect datasets, particularly Bahraini dialects, by employing a translation approach. The study focuses on sentiment analysis using an Amazon product review dataset in Bahraini dialects, translated into English and Standard Arabic. The dataset consists of balanced positive and negative reviews (2500 each). The sentiment analysis procedure employs a deep learning model based on stacked LSTM, achieving an AUC value of 98.46% for the Bahraini dialect dataset.

Sharmin et al. [22] proposed a method to analyze sentiments in Bengali utilizing a deep learning model, particularly a convolutional neural network (CNN). They trained and tested their model using a set of Bengali movie reviews. Incorporating attention mechanisms into CNNs has proven to enhance their performance.

Elfaik et al. [23] introduced a novel deep learning method for sentiment analysis in Arabic text by employing a deep bidirectional Long Short-Term Memory (LSTM) network with pre-trained word embedding. The study utilized a dataset consisting of Arabic tweets for both training and evaluating their model, while also conducting comparisons with other leading models in the field. The incorporation of pre-trained word embedding stands out as a noteworthy contribution. Results demonstrated that their proposed model achieved superior performance, boasting an accuracy of 92.61%, surpassing other state-of-the-art models.

Huang et al. [24] presented a novel approach to sentiment analysis in Vietnamese. Their methodology involved the utilization of sentiment embedding for word and transfer learning. The model was built around a Bi-directional Long Short-Term Memory with an attention mechanism and sentiment vectors were generated using Word2Vec. The training dataset, comments of students in Vietnamese, was employed, and the authors showcased the effectiveness of incorporating an English corpus for transfer learning to improve sentiment classification in Vietnamese. The proposed sentiment word vector successfully captured both contextual and sentiment information, leading to enhanced model performance. With an achieved accuracy of 86%, the model demonstrated significant potential for effective sentiment classification in Vietnamese.

Hiyama et al. [25] introduced a sentiment classification model based on neural networks with an embedded attention mechanism. The model comprises four key layers: word embedding, bidirectional LSTM, attention, and classification. The word-embedding layer is responsible for transforming input into a fixed-size continuous vector, followed by the bidirectional LSTM layer generating contextual word vectors. The attention layer then evaluates the importance of these vectors and computes a sentence vector. Finally, the classification layer employs the sentence vector to predict sentiment polarity. Notably, the model obtained an accuracy of 87.2% on the TSUKUBA corpus, which includes Japanese

TABLE 1 : Comparative Analysis of Models in Sarcasm Detection and Sentiment Analysis

Ref	Approach	Datasets	Results
Shah et al. , 2023 [1]	Modified Switch Transformer	ArSarcasm	Accuracy (SD: 0.83, SA: 0.52)
Abdelkader,et al. ,2021[2]	Multi-task Learning (MAR-ERT)	ArSarcasm	Recall(SA:0.7183,SD :0.7122)
M.V. Rao et al. , 2021 [3]	ML Algorithms (SVM, KNN, RF)	Amazon product reviews	Accuracy (SVM: 67.58%, RF: 62.34%, KNN: 61.08%)
Taha et al. , 2020 [4]	Aspect-Based SA + BERT	Twitter, Reddit	F1Score(Twitter:0.73, Reddit:0.734)
De Kok et al. , 2018 [5]	Ontology Features with SVM	Hotel reviews	F1-score sentence level: 0.792 review level: 0.8119
Patrick et al. , 2022 [6]	Transfer Learning (CNN-SVM)	I-Sarcasm Eval dataset	F1Score CNN+SA: 0.1679, CNN: 0.2024,SVM+SA: 0.2738,SVM: 0.2721
Cignarella et al., 2020[7]	Various Models (SVM, LR, GRU, BERT)	English, Spanish, French, and Italian Datasets	highest F1Score (Best Features : 0.702, for Bert + Syntax :.785)
Tan et al., 2023 [8]	Bidirectional LSTM	Twitter	F1-score (SA: 94%, SD: 93%)
Gupta, et al., 2020 [9]	Artificial Neural Networks	Social media data	Accuracy (100%)
Han et al., 2022 [10]	Cross-lingual learning (ERNIE-M, DeBERTa)	SemEval- 2022 Task 6	F1-score: Task (A): 56.91, Task (B): 7.99, Task (C): 90.50,
Murali et al. , 2023 [11]	Bi-LSTM	datasets annotated manually and automatically	Highest Accuracy(automatically: 98.76% , manually :78.42%)
Goel et al. , 2020 [12]	Na¨ıve Bayes	Twitter dataset	Precision :0.78
Arun et al. , 2020 [13]	Multinomial Na¨ıve Bayes, SVM, Decision Tree, RF, etc.	Multilingual tweets	Accuracy(SVM:95% ,RF : 93%)
Zahedi et al. , 2022 [14]	LR, LSV, NB, CNB	Twitter data in multiple languages	Accuracy LR:75% LSV:74% N :73% CNB:73%
Keith et al., 2022 [15]	traditional machine learning (LR, SVC, RF)	Malta’s Annual Government Budget	SD: 0.931 (LinearSVC) SA: 0.739,
Ma et al., 2018 [16]	Attentive LSTM With CommonSense Knowledge	English Sen- hood	Accuracy (89.32%)
Souza et al. , 2022 [17]	BERT Mod-els	Brazilian Portuguese	ROC-AUC: 98.4%
Savini et al. , 2022 [18]	BERT	TransferIMDB	F1-Score : 80.85
Parveen et al. , 2023 [19]	CNN-based Model	Amazon, Twitter datasets	precision:0.92 recall:0.89 F1-score :0.90
Omran et al. , 2023 [21]	Stacked LSTM	Bahraini dialect dataset	AUC: 98.46%
Elfaik et al. , 2020 [23]	Bidirectional LSTM with Pre-trained embedding	Arabic tweets	Accuracy:92.61%
Huang et al., 2022 [24]	BiLSTM with Attention	Vietnamese dataset	Accuracy: 86%
Hiyama et al. , 2018 [25]	Neural Network with Attention	TSUKUBA corpus	Accuracy (87.2%)

hotel reviews and English movie reviews.

III. Comparative Analysis

The comparative analysis reveals a rich landscape of methodologies, each offering unique insights and contributions to the challenging domains of sarcasm detection and sentiment analysis. The nuanced exploration of linguistic features, cross-lingual adaptability, and the transformative impact of pre-trained models collectively contribute to advancing the state-of-the-art in understanding and interpreting complex language nuances in digital communication.

The varying results in sentiment analysis studies come from several factors. Firstly, the choice of methodology and algorithm is crucial, with deep learning models often achieving high accuracy due to their ability to capture complex patterns. Conversely, simpler models may struggle with nuanced data. Secondly, dataset quality and diversity greatly influence performance, with manually annotated or diverse datasets leading to more robust models. Additionally, differences in evaluation metrics also impact perceived performance. Overall, the combination of algorithmic sophistication, dataset quality, and evaluation methodology collectively determines observed results, emphasizing the need for comprehensive analysis when interpreting performance outcomes.

The comparative analysis across the diverse literature on sarcasm detection and sentiment analysis provides a nuanced understanding of the strengths and weaknesses of various methodologies. Taha et al. [4] integration of aspect-based sentiment analysis and BERT exhibits commendable performance, as evidenced by competitive F1 scores on both Twitter and Reddit datasets. The approach's reliance on contextual information and sophisticated language representations positions it as a formidable contender in social media sentiment analysis. Shah et al.[1] modified switch transformer, while showcasing efficacy in sarcasm detection with dynamic attention mechanisms, reveals a more modest performance in sentiment analysis. This highlights the inherent challenge of crafting models adaptable to both tasks, as intricate attention mechanisms may excel in one aspect but falter in another. Transfer learning emerges as a recurring theme, with Patrick et al. [6] demonstrating the impact of integrating sentiment analysis into sarcasm detection models. The nuanced improvements in performance, particularly with the CNN model, underscore the potential synergy between sentiment and sarcasm classification tasks. Multi-task learning, exemplified by Abdelkader et al. [2] MARBERT model, showcases the benefits of knowledge sharing between sentiment and sarcasm classification. The model's top-ranking performance in sentiment analysis and competitive standing in sarcasm detection positions it as a versatile solution for understanding diverse linguistic nuances. Gupta Shaina et al. [9] exploration of artificial neural networks (ANN) for sentiment analysis and emoticon-based sarcasm detection reveals a striking 100% accuracy, underscoring the robustness of deep learning in handling the complexities of social media language. The emphasis on precision in sentiment analysis by Goel et al. [12] in multiple languages through Na¨ıve Bayes demonstrates the importance of language-specific considerations in model design. The incorporation of syntactic features, as explored by Cignarella et al. [7], proves pivotal in irony detection across multiple languages. The study's comprehensive exploration of traditional classifiers like SVM and LR, alongside advanced models like GRU and BERT, highlights the need for a nuanced understanding of linguistic structures for effective sentiment and sarcasm analysis. Cross-lingual learning, a challenging frontier, is tackled adeptly by Han et al. [10], achieving competitive results in sarcasm detection in both English and Arabic. The study's success suggests the feasibility of adapting models to diverse linguistic contexts, a crucial consideration in the era of globalized digital communication. The analysis further explores the efficacy of pre-trained language models, such as BERT, in sentiment analysis and sarcasm detection. Souza et al. [17] and Savini et al. [18] demonstrate the consistent superiority of BERT over traditional methods like TF-IDF, reaffirming the transformative impact of pre-trained language representations in understanding complex linguistic nuances. CNN-based models, as exemplified by Parveen et al. [19] and Hiyama et al. [25], exhibit prowess in capturing implicit and clear representations for sarcasm detection and sentiment analysis. Those studies showcase the efficiency of CNN in handling the intricate interplay of language elements.

I. Conclusion

This survey paper thoroughly explores the intersection of Natural Language Processing (NLP), sentiment, and sarcasm classification in machine learning. It highlights the challenges posed by sarcasm's implicit nature and contextual dependencies, offering insights into the complexities of accurate detection. The paper not only identifies these challenges but also examines innovative solutions spanning traditional and advanced machine learning methods. Its contribution lies in a comprehensive review of current research in single and multiple languages, analyzing the strengths and weaknesses of various approaches. By addressing challenges in sentiment and sarcasm classification, the paper serves as a practical resource for researchers and practitioners. The comparative analysis across studies showcases notable achievements, such as impressive F1 scores in sentiment analysis and the successful use of advanced models like the modified switch transformer for sarcasm detection. The bidirectional LSTM demonstrates significant improvements in both sentiment and sarcasm-related tasks. As the field progresses, the paper encourages careful consideration of specific use cases when selecting models, laying the foundation for future developments in NLP, sentiment analysis, and sarcasm detection across diverse linguistic contexts. In future research, focusing on sarcasm-related features and their impact on sentiment determination could provide valuable insights into improving both sarcasm detection and sentiment classification. Addressing dataset quality issues, including scarcity of sarcasm samples, is crucial for enhancing the reliability and effectiveness of models. Additionally, handling multilingual datasets requires more than just translation; leveraging features from the source language can lead to more accurate results. Similarly, addressing code-mixed datasets, which contain multiple languages within the same text, presents an opportunity for developing specialized models capable of effectively processing such complex linguistic data. By addressing these areas, future research can advance the field of sarcasm detection and sentiment classifications toward robust and adaptable methodologies.

References

[1] Shah, S and Shah, Syed and Ullah, Asad and Rizwan, Atif and Atteia, Ghada and Alabdulhafith, Maali. (2023). Arabic Sentiment Analysis and Sarcasm Detection Using Probabilistic Projections-Based Variational Switch Transformer. IEEE Access. PP. 1-1. 10.1109/AC- CESS.2023.3289715.

[2] Abdelkader El Mahdaouy, Abdellah El Mekki, Kabil Essefar, Nabil El Mamoun, Ismail Berrada, and Ahmed Khoumsi. 2021. Deep Multi- Task Model for Sarcasm Detection and Sentiment Analysis in Arabic Language. In Proceedings of the Sixth Arabic Natural Language Pro- cessing Workshop, pages 334–339, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

[3] M. V. Rao and S. C., ”Detection of Sarcasm on Amazon Product Reviews using Machine Learning Algorithms under Sentiment Analysis,” 2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 2021, pp. 196-199, doi: 10.1109/WiSPNET51692.2021.9419432.

[4] Taha Shangipour ataei, Soroush Javdan, and Behrouz Minaei-Bidgoli. 2020. Applying Transformers and Aspect-based Sentiment Analysis approaches on Sarcasm Detection. In Proceedings of the Second Workshop on Figurative Language Processing, pages 67–71, Online. Association for Computational Linguistics

[5] de Kok, Sophie , Punt, Linda , Puttelaar, Rosita , Ranta, Karoliina , Schouten, Kim , Frasincar, Flavius. (2018). Review-aggregated aspect- based sentiment analysis with ontology features. Progress in Artificial Intelligence. 7. 10.1007/s13748-018-0163-7.

[6] Patrick Hantsch and Nadav Chkroun. 2022. connotation_clashers at SemEval-2022 Task 6: The effect of sentiment analysis on sarcasm detection. In Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022), pages 945–950, Seattle, United States. Association for Computational Linguistics.

[7] Cignarella, Alessandra & Basile, Valerio & Sanguinetti, Manuela & Bosco, Cristina & Rosso, Paolo & Benamara, Farah. (2020). Multilingual Irony Detection with Dependency Syntax and Neural Models.

[8] Tan, Yik & Chow, Chee Onn & Kanesan, Jeevan & Chuah, Joon Huang & Lim, YongLiang. (2023). Sentiment Analysis and Sarcasm Detection using Deep Multi-Task Learning. Wireless Personal Communications.
129. 10.1007/s11277-023-10235-4.

[9] Gupta, Shaina & Singh, Ravinder & Singla, Varun. (2020). Emoticon and Text Sarcasm Detection in Sentiment Analysis. 10.1007/978-981-15-0029-9_1.

[10] Han, Yaqian & Chai, Yekun & Wang, Shuohuan & Sun, Yu & Huang, Hongyi & Chen, Guanghao & Xu, Yitong & Yang, Yang. (2022). X-PuDu at SemEval-2022 Task 6: Multilingual Learning for English and Arabic Sarcasm Detection.

[11] Murali Krishna, Muddada and Midhunchakkaravarthy, and Vankara, Jayavani. (2023). Detection of Sarcasm Using Bi-Directional RNN Based Deep Learning Model in Sentiment Analysis. Journal of Advanced Research in Applied Sciences and Engineering Technology. 31. 352-362. 10.37934/araset.31.2.352362.

[12] Goel, Pragati & Goel, Vikas & Gupta, Amit. (2020). Multilingual Data Analysis to Classify Sentiment Analysis for Tweets Using NLP and Classification Algorithm. 10.1007/978-981-15-0694-9_26.

[13] Arun, K. & Srinagesh, Ayyagari. (2020). Multi-lingual Twitter sentiment analysis using machine learning. International Journal of Electrical and Computer Engineering (IJECE). 10. 5992. 10.11591/ijece.v10i6.pp5992-6000.

[14] Zahedi, Reza , Alajlan, Naif , Zahedi, Hooman , Rabczuk , Ti- mon. (2022), “Multilingual Sentiment Mining System to Prognos- ticate Governance“, Computers, Materials, Continua, 71. 389-406. 10.32604/cmc.2022.021384.

[15] Keith Cortis and Brian Davis. 2022. Baseline English and Maltese-Maltese-English classification Models for Subjectivity Detection, Sentiment Analysis, Emotion Analysis, Sarcasm Detection, and Irony Detection. In Proceedings of the 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages, pages 161–168, Marseille, France. European Language Resources Association.

[16] Ma, Yukun , Peng, Haiyun , Cambria, Erik. (2018). Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM. Proceedings of the AAAI Conference on Artificial Intelligence. 32. 10.1609/aaai.v32i1.12048.

[17] Souza, Frederico , Souza Filho, Joao. (2022). BERT for Sentiment Analysis: Pre-trained and 15 Fine-Tuned Alternatives. 10.1007/978-3- 030-16 98305-5-20.

[18] Savini, Edoardo , Caragea, Cornelia. (2022). Intermediate-Task Transfer Learning with BERT for Sarcasm Detection. Mathematics. 10. 844. 93 10.3390/math10050844.

[19] Parveen, Summia & Saradha, S. & Krishnaraj, N.. 60 (2023). An Efficient Detection and Classification of Sarcastic by Using CNN Model. 10.1007/978-62 981-19-7447-2 17.

[20] Moholkar, Kavita , Rathod, Krupa , Rathod, Krishna , Tomar, Mritunjay , Rai, Shashwat. (2020), “Sentiment Classification Using Recurrent Neural Network“, 10.1007/978-3-030-28364-3-49.

[21] Omran, Thuraya & Sharef, Baraa & Grosan, Crina & Li, Yongmin. (2023), “Sentiment Analysis of Multilingual Dataset of Bahraini Dialects, Arabic, and English“, Data 8. 68. 10.3390/data8040068.

[22] Sharmin, Sadia , Chakma, Danial. (2021). Attention-based convolutional neural network for Bangla sentiment analysis. AI , SOCIETY. 36. 10.1007/s00146-020-01011-0.

[23] Elfaik, Hanane , Nfaoui, El Habib. (2020). Deep Bidirectional LSTM Network Learning-Based Sentiment Analysis for Arabic Text. Journal of Intelligent Systems. 30. 395-412. 10.1515/jisys-2020-0021.

[24] Huang, Yong , Liu, Siwei , Qu, Liangdong , Li, Yongsheng. (2022). Correction to: Effective Vietnamese Sentiment Analysis Model Using Sentiment Word Embedding and Transfer Learning. 10.1007/978-981-15-7984-4-48.

[25] Hiyama, Yohei, Yanagimoto, Hidekazu. (2018). Word polarity at-attention in sentiment analysis. Artificial Life and Robotics. 23. 1-5. 10.1007/s10015-018-0439-9.