Systematic Literature Review: Computational Approaches for Humour Style Classification

Mary Ogbuka Kenneth[1], Foaad Khosmood[2], and Abbas Edalat[3]

[1,3]Algorithmic Human Development, Department of Computing, Imperial College London
[2]Computer Engineering Department, California Polytechnic State University

February 6, 2024

Abstract

Understanding humour styles is crucial for grasping humour's diverse nature and impact on psychology and artificial intelligence. Humour can have both therapeutic and harmful effects, depending on the style. Though computational humour style analysis studies are limited, extensive research exists, particularly in binary humour and sarcasm recognition. This systematic literature review explores computational techniques in these related tasks, revealing their relevance to humour style analysis. It uncovers common approaches, datasets, and metrics, addressing research gaps. The review identifies features and models that can transition smoothly from binary humour and sarcasm recognition to humour style identification. These include incongruity, sentiment analysis, and various models, like neural networks and transformer-based models. Additionally, it provides access to humour-related datasets, aiding future research.

Keywords— Humour style classification, Sarcasm detection, Binary Humour detection

1 Introduction

Mental health disorders are an increasing public health concern since poor mental health relates to significant emotional distress and contributes significantly to the global illness burden [14]. Duchenne (Genuine) laughter is a common reaction to amusing stimuli or pleasant feelings; it is frequently regarded as having beneficial psychological and physical consequences for humans. As a result, it is considered a possible mental health intervention [67]. Laughter therapy, which is a form of cognitive-behavioural therapy, can enhance a person's quality of life overall as well as their physical, psychological, and social connections [98]. The two main types of laughter therapy are humour-based and self-stimulated laughter yoga therapy [12, 33]. The use of humour as a therapeutic technique to advance physical, emotional, and psychological well-being is a common theme throughout both therapies.

Humour, as a complex and essential aspect of human communication, plays a vital role in social interactions, emotional expression, and cognitive processing [15, 19]. Understanding how individuals perceive and express humour and how it relates to our mental health and relationships has garnered significant interest among researchers, leading to the emergence of the field of humour style classification [85, 56, 57, 39, 58, 62]. Humour style recognition focuses on categorising individuals based on their approaches to humour, shedding light on the diversity and intricacies of humour-related behaviours [15].

Humour style refers to an individual's preferred or characteristic way of expressing humour and engaging in humourous activities. It is a unique and consistent pattern of using humour to interact with others, cope

1

with stress, and express one's personality [62, 92]. The Humour Styles Questionnaire (HSQ), created by Martin et al. [62], has identified four fundamental humour styles, each representing different approaches to humour engagement. First, there is Affiliative Humour, which is characterised by using humour to foster social bonds and bring people together, with a focus on making others laugh. On the other hand, Self-Enhancing Humour highlights using humour as a coping mechanism, finding amusement in life's challenges, and maintaining a positive outlook. The Aggressive Humour style involves using humour as a form of criticism, sometimes at the expense of others, employing sarcasm or teasing. Lastly, there is Self-Deprecating Humour, where individuals humourously acknowledge their own imperfections, using self-mockery as a way to connect with others. These four humour styles, as delineated by the HSQ, provide insights into the multifaceted role of humour in social interactions, emotional expression, and coping mechanisms.

The HSQ is a widely used psychological diagnostic tool designed to determine a person's preferred humour style. It typically consists of 32 items, with respondents rating their responses on a Likert scale, indicating their level of agreement or disagreement with each statement [62].

While early studies on humour styles heavily relied on self-report questionnaires, the evolution of computational methods has ushered in automated techniques for humour style identification [28, 49, 2]. The application of these methodologies in the analysis of extensive textual and multimedia datasets has enhanced our comprehension of humour styles within various contexts.

Although the number of studies specifically focusing on humour style classification using computational models may be limited, there exists a substantial body of research on related tasks, such as binary humour recognition (i.e., determining whether a sentence or scene is humourous or not) and sarcasm recognition. These related tasks share commonalities in humour-related analysis, including feature extraction, representation modelling, and classification algorithms. Understanding the relationship between humour style classification and related tasks is crucial for advancing the field. Lessons learned from humour recognition and sarcasm recognition can be leveraged to improve the accuracy and generalisability of humour style classification models, even with the limited number of specific studies in this domain.

There are few review paper [81] and systematic literature review papers that focused on analysis of computational approaches for Humour [46] and sarcasm recognition [8]. However, there is no review or systematic literature review paper of computational approaches for humour style recognition.

This systematic literature review aims to provide a comprehensive analysis of the utilisation of computational approaches for humour style classification and its related tasks. By synthesising the existing literature on humour recognition and sarcasm recognition, the review aims to identify common methodologies, datasets, evaluation metrics, and insights that can inform and enhance humour style classification.

The specific objectives of this review are as follows:

1. To analyse the computational techniques employed in binary humour recognition and sarcasm recognition.

2. To explore the transferability of features and computational models from related tasks to humour style classification.

3. To identify the strengths and limitations of existing computational-based approaches for humour-related tasks.

4. To highlight potential research gaps and opportunities for future investigations in humour style classification using machine learning.

By addressing these objectives, this review intends to contribute to the advancement of humour style classification research and its practical applications in various domains, including natural language processing, human-computer interaction, and psychological studies.

The rest of this paper is organised as follows: Section 2 describes the review process that was employed to carry out the study. Section 3 displays and explains the outcomes of the review by answering the research questions. Section 4 discusses unresolved challenges and potential future paths in the subject of humour styles and associated tasks. Conclusions were drawn in Section 5.

2 Methodology

This section presents the methodology for conducting the SLR on humour style classification and related tasks, such as binary and multi-class humour recognition and sarcasm recognition using computational models. The SLR aims to provide a comprehensive analysis of the existing literature in this domain [89], exploring the computational techniques utilised and the transferability of methodologies from related tasks to humour style classification.

The SLR methodology encompasses the following key components: inclusion and exclusion criteria, search strategy and databases used, selection process and study identification, data extraction and analysis, and data synthesis. Through a clear and systematic approach, this methodology enables the review to identify and address the fundamental research questions in this domain[41].

2.1 Research Questions

Over the years, a very small number of scholars have used computational methods to accomplish humour style classification. As a result, this study intends to answer the following SLR research questions:

1. Are there open-source annotated humour styles, sarcasm, and humour datasets in the field of humour recognition?

2. What features are commonly extracted for humour style classification and related tasks, such as humour and sarcasm recognition?

3. What are the computational techniques commonly employed in humour style classification and its related tasks?

4. What are the strengths and limitations of existing computational-based approaches for humour-related tasks, including humour style classification?

2.2 Search Strategy and Databases

A systematic and comprehensive literature search was conducted to identify relevant studies. The search strategy involved using appropriate keywords and search terms [66] related to humour style classification, machine learning, computational models, humour recognition, and sarcasm recognition.

Examples of the search string used are as follows:

1. *("humour style" OR "humor style" OR "sarcasm" OR "humour" OR "humor") AND ("detection" OR "recognition" OR "prediction" OR "classification" OR "identification")*

2. *("machine learning" OR "deep learning" OR "artificial intelligence") AND ("humour style" OR "humor style" OR "sarcasm" OR "humour" OR "humor") AND ("detection" OR "recognition" OR "prediction" OR "classification" OR "identification").*

The academic databases that were used in the search are: IEEE Xplore (https://ieeexplore.ieee.org/), ScienceDirect (https://www.sciencedirect.com/), ACM Digital Library (https://dl.acm.org/), Scopus (https://www.scopus.com/), arXiv (https://arxiv.org/), Semantic Scholar (https://www.semanticscholar.org/), Web of Science (https://www.webofscience.com/), Google Scholar (https://scholar.google.com/), Springer Link (https://link.springer.com/), and Database systems and logic programming (DBLP) (https://dblp.uni-trier.de/) The search was carried out without any time restrictions to ensure the inclusion of relevant historical studies while focusing on the specified timeframe for inclusion.

2.3 Inclusion and Exclusion Criteria

To ensure the relevance and consistency of the selected studies, specific inclusion and exclusion criteria were established. The criteria aimed to capture studies that focused on humour style and related task recognition using computational models, while excluding studies that did not align with the research objectives. The inclusion and exclusion criteria used are depicted in Table 1

Table 1: Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Studies that addressed humour style classification using computational techniques	Studies that focused solely on non-computational-based humour style classification approaches
Studies that explored related tasks such as humour recognition and sarcasm recognition	Studies that were not directly related to humour style classification or related tasks
Studies published in peer-reviewed journals and conference proceedings	Studies published in languages other than English
Studies written in English	Studies published before the specified time-frame
Studies published within the last 10 years	Short papers
Full-text article	

2.4 Selection Process and Study Identification

The initial search yielded a substantial number of papers related to humour style classification, binary and multi-class humour recognition, and sarcasm recognition. All search results were imported into reference management software to facilitate the removal of duplicates. Subsequently, a title and abstract screening was conducted to identify papers that met the pre-defined inclusion and exclusion criteria. After the initial screening, the selected papers were obtained in full-text format for further examination.

2.5 Quality Assessment

The quality assessment process is used to ensure the rigour and credibility of the primary papers reviewed in this study[69]. The quality assessment aimed to evaluate the selected primary studies based on specific criteria to determine their methodological soundness and relevance to the research objectives [53, 91].

Quality assessment checklist for the reviewed primary papers:

1. **Research Objectives and Scope:** This checks if the selected studies clearly stated their research objectives and scope in relation to humour style or related task identification. Then we check if the research questions were well-defined and if they were aligned with the overall research objectives of the SLR.

2. **Methodology:** Checks if the studies provided a clear description of their approach to humour style or related task recognition. Similarly, the appropriateness of the computational techniques used is checked. Lastly, the suitability of the datasets for the evaluation of the computational techniques is checked.

3. **Evaluation Metrics:** Did the studies employ appropriate evaluation metrics to assess the performance of their machine learning models, and were the evaluation metrics relevant and aligned with the research objectives?

4. **Results and Discussion:** Did the studies present their results and findings in a clear and concise manner, and were insights into the effectiveness of the computational models provided?

5. **Limitations and Future Directions:** This checks if the studies acknowledged any limitations in their approach or data and discussed their potential impact on the results. Furthermore, we check if the studies propose future research directions in relation to their work.

6. **Citation and References:** Lastly, we check if the studies appropriately cite and reference prior literature relevant to humour and its related tasks.

Table 2: Data Extraction Form

Extracted Data	Description	Data Type
Bibliography	Publication year, author, title and the publisher	General
Study type	Journal, conference, workshop, or lecture paper	General
Dataset	Description of dataset used	Research question
Extracted Features	Description of humorous or sarcastic features extracted and used	Research question
Computational technique	Employed computational method used like rule-based or machine learning	Research question
Performance Measure	Performance measures used to evaluate the computational models	Research question
Findings and contributions	Results obtained and contributions	General

By conducting a thorough quality assessment, the SLR aimed to present a reliable and comprehensive synthesis of the selected studies and their contributions to the field of humour style classification using computational models and related tasks.

2.6 Data extraction and analysis

For the selected studies, a systematic data extraction process was employed to capture relevant information. The data extraction included details such as authors, publication year, study objectives, computational techniques used, datasets employed, evaluation metrics, and key findings.

Following data extraction, a qualitative analysis was performed to identify common methodologies, approaches, and trends across the selected studies. The analysis aimed to provide insights into the computational techniques used in humour and sarcasm classification, highlight their performance, and explore potential transferability to humour style classification.

The analysis and conclusions drawn from the selected studies were used to inform potential research gaps and future directions in humour style classification using computational models. The data extraction form that was used to perform an in-depth analysis for all selected primary studies is presented in Table 2.

2.7 Data Synthesis

Data synthesis for the SLR involves analysing and summarising the information extracted from the selected studies on humour style and related task (humour and sarcasm) recognition using computational models.

The initial search of the research databases based on the predefined search strings yielded about 4290 studies. Out of these initial results, 2518 were deemed irrelevant, 612 were duplicate articles, 951 were excluded based on irrelevant titles and abstracts, and 209 studies were selected for possible inclusion. Lastly, 150 papers were excluded based on the research question. After following the procedures outlined above, 59 papers were chosen as primary papers. Figure 1 shows a visual representation of the results obtained after performing data synthesis.

(a) Number of Studies per Systematic Procedure

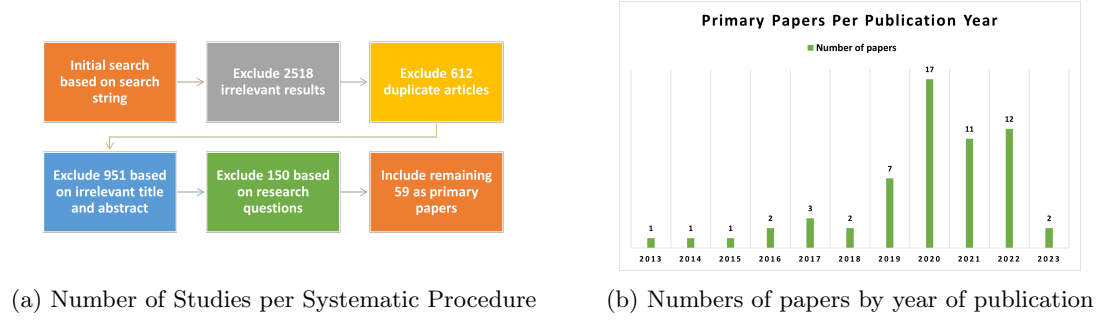(b) Numbers of papers by year of publication

Figure 1: Data Synthesis Results

Figure 1a depicts the number of studies conducted for each systematic procedure.

The 59 primary papers included were published between 2013 and 2023, spanning a period of ten years. Figure 1b displays the number of primary studies by published year. From Figure

1b, it can be seen that 2020 has the highest selected papers with 17 articles, followed by 2022 with 12 articles, and then 2021 with 11 articles. The years with the fewest primary papers are 2013, 2014, and 2015.

3 Results and Discussion

This section presents the results and discussion from the SLR that was conducted to address the stated research questions. Each subsection in this section is dedicated to answering one of the research questions. Additionally, the SLR responses that were gathered from the included primary research paper are reviewed.

A summary of the datasets, extracted features, computational models, and performance metrics used in the 59 primary papers is presented in the appendix in Table 7.

3.1 Availability of Annotated Datasets

An essential foundation for advancing research in humour style, sarcasm, and humour recognition lies in the availability of annotated datasets. These datasets serve as crucial resources that enable the development, training, and evaluation of computational models for understanding and identifying humour-related behaviours.

Through a systematic search and selection process, a total of 39 datasets curated for humour style classification, sarcasm recognition, and humour recognition tasks were identified. The datasets are discussed in the context of the annotation techniques used and the dataset modality.

3.1.1 Dataset Annotation

The two primary methods utilised for building the datasets in the area of humour styles and their related tasks are manual annotation and the distant supervision technique.

1. **Manual Annotation:** In this method, human annotators review each piece of text (such as jokes, tweets, or comments) and manually assign labels indicating whether it contains humour or sarcasm. This approach allows for detailed analysis and accurate labelling, but it can be time-consuming and subject to annotator bias. Nonetheless, this bias can be reduced by allowing multiple annotators to independently label each instance. The final label is then determined by a majority vote among annotators. This majority vote approach helps mitigate errors and subjectivity. The manual annotation method was greatly adapted for labelling multimodal datasets [74, 42, 17, 28, 50, 45]. Riloff et al. [84] used manual annotation to label the text dataset.

2. **Distant Supervision (DS):** In DS, labels are assigned to data instances indirectly based on the presence of some external information or signal. This approach leverages existing knowledge or sources of information to automatically generate labels for a large amount of data. In the context of humour and sarcasm detection, distant supervision involves using emoticons, hashtags, or metadata (such as the context of the text) to infer the label of a text snippet. Most data extracted from Twitter and Reddit made use of this annotation method, as these social media platforms make use of hashtags and emoticons to label text context [49, 2, 93, 71, 1, 37]. DS is advantageous in generating very large labelled data for training machine learning models as it is not resource-intensive. However, distant supervision comes with certain challenges, such as limited nuance, noisy labels, or inadequate generalisation.

Out of the 39 identified datasets, 22 were manually labelled, 15 were labelled using distant supervision, and 2 were labelled with a combination of both methods.

Single Modality Dataset

In the context of humour-related tasks, a single modality dataset typically consists of data from a single mode, such as text, audio, or images, for humour analysis. This means that the dataset captures humour-related information from only one source, limiting the analysis to a specific modality. Examples of single modality datasets are text-only, image-only, or audio-only datasets.

The text modality is the most commonly used single modality for humour and its related tasks. Out of 29 single modality datasets, 28 are text modality [2, 49, 25, 6, 93, 77, 70] and one dataset is image-only modality [30]. Table 3 gives details of single modality datasets used in the literature.

Table 3: Single Modality Datasets

Paper	Task	Name	Annota-tion	Labels	Language	Source	Size	Availability
[49]	Humour	E-tweets	DS	2	English	Twitter	1000	Tweets IDs are available
[2]	Sarcasm	E-tweets	DS	2	English	Twitter	1000	Tweets IDs are available
[6]	Humour	ColBert	DS	2	English	Reddit	200,000	https://rb.gy/tif35
[93, 43]	Humour	rJokes	DS	10	English	Reddit	550,000	https://rb.gy/54fcp
[25]	Humour	Short text		6	English	Different sources	32,980	https://rb.gy/22ksm
[77, 83]	Sarcasm	Tweets	DS, Manual	2	Czech and English	Twitter	7,000	https://rb.gy/22ksm
[70]	Humour	Portuguese jokes	DS	2	Portuguese	Twitter, Newspaper	2,800	https://rb.gy/51cxx
[1]	Sarcasm, Senti-ment	Arsarcasm	Manual	6	Arabic	Twitter	10,547	https://rb.gy/r7fev
[71]	Sarcasm	iSarcasm	Manual	8	English	Twitter	4,484	https://rb.gy/ghp2k
[31]	Sarcasm	IACv2	Manual	2	English	Internet Argument Corpus	4,692	https://nlds.soe.ucsc.edu/sarcasm1

Table 3: Single Modality Datasets

Paper	Task	Name	Annotation	Labels	Language	Source	Size	Availability
[55, 3, 59]	Sarcasm	DS	SARC	2	English	Reddit	175,058	http://nlp.cs.princeton.edu/SARC/
[65, 44]	Sarcasm	News headlines	DS	2	English	The Onion, HuffPost	26,709	https://rishabhmisra.github.io/publications/
[48]	Sarcasm	Twitter-280	DS	2	English	Twitter	42,622	https://rb.gy/io2xa
[11]	Humour	Stierlitz and FUN	DS	2	Russian	VK, Twitter, anekdot	312,877	https://rb.gy/ctxj8
[64, 35, 36, 99, 82]	Humour	1600 One-liner	DS	2	English	Berro, Muted Faith, Proverbs, British National Corpus, Reuters	16,000	Contact Author
[97, 35, 36]	Humour	Pun of the Day	DS	2	English	Pun of the Day, News, Proverbs	4,826	https://rb.gy/06lgn
[100]	Humour	Product Question Answering	Manual	2	English	Product Question Answering Systems	28,716	https://rb.gy/t9bd2/
[63, 94]	Humour	SemEval-2021 Task 7	Manual	2	English	Twitter, Kaggle	10,000	https://semeval.github.io/SemEval2021/
[78, 43]	Humour	English Plaintext Jokes		2	English	Reddit, stupidstuff, wocka	208,000	https://rb.gy/9y453
[68]	Humour	Yelp Dataset	DS	3	English	Yelp	34,000	https://rb.gy/q2y1y
[42] : [16, 4]	Humour	Humicro-edit	Manual	4	English	Reddit	5,000	https://rb.gy/u0nu3
[17]	Humour	Spanish Corpus	Manual	2	Spanish	Twitter	27,000	https://t.ly/WmKRF
[88, 73]	Sarcasm	English-Hindi code-mixed	DS, Manual	2	English, Hindi	Twitter	5,250	https://t.ly/b1acR
[90]	Humour	Naughty-former	DS	4	English	Reddit, Reuters	92,153	https://rb.gy/99tvq
[72, 83]	Sarcasm	IAC	Manual	2	English	Internet Argument Corpus	9,386	https://rb.gy/99tvq

Table 3: Single Modality Datasets

Paper	Task	Name	Annotation	Labels	Language	Source	Size	Availability
[37, 26, 75, 34]	Sarcasm	Sarcasm-Detection	DS	2	English	Twitter	51,189	https://rb.gy/ynvzw
[84, 3, 75, 34]	Sarcasm	Tweet-Sarcasm	Manual	2	English	Twitter	3,200	https://rb.gy/jftmz
[30]	Sarcasm	Yahoo Flickr Sarcasm	DS	2	Image	Flickr	1,846	https://rb.gy/na3jx

Overall, single modality datasets are valuable resources for specific types of humour analysis tasks, especially those that primarily rely on one type of data. Nonetheless, it is essential to acknowledge that humour is a multi-faceted and context-rich phenomenon, and in some cases, incorporating multiple modalities may enhance the accuracy and completeness of the humour analysis.

Multimodal Dataset

A multimodal dataset is a type of dataset that contains multiple modes or types of data, each providing different types of information about a particular subject, event, or phenomenon. The goal of creating multimodal datasets is to enable the study of relationships and patterns that can be discovered by analysing multiple types of data simultaneously. Examples of this modality are text-image, text-audio, and text-audio-video datasets. In this SLR, ten multimodal datasets were identified. Most of the multimodal datasets consist of video-text-audio modality and they were manually annotated. Table 4 gives details of the multimodal datasets used in the primary papers. Table 4 includes the name of the multimodal dataset, the dataset's creator and papers that used it, the task the dataset can be used for, the annotation method used, the number of categories in the data, the language of the data, the sources from which the raw data was extracted, the modality contained in the dataset, and the URL to the dataset, if it is publicly available.

Table 4: Multimodal Datasets

Paper	Task	Name	Annotation	Labels	Language	Source	Type	Availability
[74, 61]	Humour	MHD	Manual	2	English	TV show Big Bang Theory	Video, Text	https://tinyurl.com/4a2d3t3x
[18, 40, 61]	Sarcasm	MUStARD	Manual	2	English	TV show Friends, Golden Girls, Big Bang Theory, Sarcasmaholics Anonymous.	Video, Text, Audio	https://rb.gy/iwm5p
[23]	Sarcasm	SEEmoji MUStARD	Manual	2	English	TV show Friends, Golden Girls, Big Bang Theory, Sarcasmaholics Anonymous.	Video, Emoji, Text, Audio	https://rb.gy/x8q6l

[28, 95, 51, 29]	Humour Styles	Passau-SFCH	Manual	4	German	Football press conference	Video, Text, Audio	Available on request
[50, 27, 40, 61]	Humour	UR-FUNNY	Manual	2	English	Ted Talks	Video, Text, Audio	https://rb.gy/r6xdp
[60]	Humour	Memeplate	Manual	2	Chinese	Weibo, Tieba, Baidu, Yandex	Image, Text	https://rb.gy/rnvqe
[87, 21]	Humour, Sarcasm, offensive, motivation	Memotion	Manual	4	English	Various sources	Image, Text	https://rb.gy/2mm2e
[9]	Humour, sarcasm	MaSaC	Manual	4	Hindi, English	Tv show (Sarabhai versus Sarabhai)	Audio, Text	https://rb.gy/x27c3
[45, 61]	sarcasm	Friends	Manual	2	English	Tv show	Video, Audio, Text	Available on request
[24]	Humour	M2H2	Manual	2	Hindi	Shrimaan Shrimati Phir Se TV show	Video, Audio, Text	https://rb.gy/0jfrv

3.2 Commonly Extracted Features

In the context of humour style classification and related tasks, various features are extracted from the single or multimodal data to capture different aspects of humour or sarcasm. These features can be broadly categorised into three main groups: language-based features, sentiment and tone features, and text representation features.

3.2.1 Language-Based Features

These features involve the linguistic elements of humour that contribute to its comedic effect. They play with the structure, context, and style of language to create humorous situations. In this SLR, eleven language-based features such as punctuation, semantic similarity, n-grams and more were identified.

1. **Contextual Information:** This feature plays a vital role in humour comprehension and interpretation. Humour is often dependent on cultural, social, and linguistic factors, which shape how individuals perceive and respond to humorous content. Key aspects of contextual information include cultural differences, wordplay and language, shared knowledge, social situations, stereotypes, and incongruity. Zhang et al. [99] extracted the latent contextual knowledge at the word level by assigning a contextual similarity score to each instance. Huang et al. [43] extracted contextual information in the form of commonsense features.

2. **Incongruity:** Incongruity is one of the primary theories of humour [86, 52]. Incongruity simply refers to a cognitive phenomenon where there is a disconnect or misalignment between what is expected and what actually occurs. Incongruity theory suggests that people find things funny when there is an unexpected twist or discrepancy, challenging their initial assumptions or mental schemas [52, 86, 43]. In order to identify incongruity for humour Yang et al. [97] measured the semantic disconnection (disconnection and repetition) in a sentence using the Word2Vec model. Ziser et al. [100] identified this feature difference between the question and its corresponding product. Other literature that utilises this feature for humour are [43, 68, 16].

3. **Ambiguity:** Ambiguity in the context of humour refers to the presence of multiple possible interpretations within a joke. This uncertainty in meaning can lead to humorous outcomes as individuals navigate through different interpretations, often resulting in an unexpected or amusing conclusion [97, 36, 35, 10]. Yang et al. [97] captured the ambiguity of the sentence using WordNet. Fan et al. [36], Morales and Zhai [68], Hasan et al. [40], Xie et al. [94], Barbieri et al.[7] and Fan et al. [35] also identified and extracted ambiguity in sentences.

4. **Phonetic Style:** The term "phonetic style" in humour refers to using phonetic features to produce comic effects, such as sounds, rhythms, wordplay, and speech patterns [97]. Humour is produced by changing the way words are pronounced and how they sound. An essential component of verbal humour that draws on the auditory and phonological features of language is phonetic style. The alliteration and rhyme features were used by Yang *e al*[97] to identify the phonetic style. Alliteration simply refers to the repetition of the same initial consonant sound in a sequence of words to create a playful and rhythmic effect [97]. Alliteration was also used by [68, 94] for humour identification.

5. **Exaggeration:** Exaggeration is a humour technique that includes dramatically emphasising or overstating something in order to produce a humourous effect [49]. It frequently entails exaggerating, intensifying, or enhancing an event, description, or ingredient beyond what would be realistic. Kamal and Abulaish [49] used the exaggeration feature to identify self-deprecating humour style. The identified exaggeration consists of an intensifier in the form of an adverb, adjective, or interjection, along with its frequency count.

6. **Part-of-Speech (POS) :** POS features are linguistic elements that provide information about the grammatical category of words in a sentence. These features play a significant role in humour analysis as they help capture the syntactic structure and grammatical patterns that contribute to the comedic effect [77]. This feature was used by Ptacek et al., [77], Ghosh and Veale [37] and Jariwala [44] for sarcasm detection. Kamal and Abulaish [49] used personal pronouns like "I", "me", "my", and "we" to identify self-deprecating humour style (which involves intentionally highlighting one's own flaws or vulnerabilities in a lighthearted and often relatable way to evoke laughter from others). Kamal and Abulaish [2] also made use of the POS features to identify self-deprecating sarcasm.

7. **N-gram-based features:** N-grams can be used to capture linguistic patterns and context in text analysis [54]. N-grams help identify recurring word sequences, phrases, and linguistic patterns that contribute to the comedic effect of humour. N-gram (frequency) was used by Barbieri et al.[7], Ptacek et al. [77], Joshi et al. [45], Jariwala [44], Riloff et al.[84] and Swami et al. [88] for sarcasm detection.

8. **Punctuation Features:** In the context of humour, this refers to the utilisation and manipulation of punctuation marks within the text to enhance the comedic effect or convey a specific tone. Punctuation marks, such as commas, periods, exclamation points, question marks, and ellipses, can contribute to the overall humorous intent of a text by influencing the reader's perception of timing, emphasis, and tone. Kumar et al. [59] identified that sarcasm employs behavioural aspects like low tones, facial gestures, or exaggeration. These low tones, facial gestures, or exaggerations are expressed using punctuation in sentences. Hence, to detect sarcasm Kumar et al. [59], Ptacek et al. [77], Jariwala [44] and Barbieri et al. [7] extracted punctuation features.

9. **Term Frequency-Inverse Document Frequency (TF-IDF):** This is a technique used to quantify the importance of words within a collection of documents. The meaning of the word increases proportionally to the number of times in the text the word appears [80]. Words with high TF-IDF scores in a sentence are likely to be indicative of whether it is humorous or not. Pandey and Singh [73] utilised this feature for sarcasm detection. Alexandru and Gîfu [4] and Ahuja and Sharma [3] applied it to humour detection.

10. **Semantic similarity:** This refers to numerical values or scores that quantify how similar or related two pieces of text, words, or concepts are in terms of their meaning [20]. This feature aims to capture the degree of likeness in meaning between words. This feature was used by Xie et al.[94] for humour recognition.

11. **Out-of-vocabulary (OOV) words :** OOV are words that are not typically found in a natural language processing environment's vocabulary. For humour identification in Portuguese, Oliveira et al. [70] utilised this feature to count the number of terms that a Portuguese word2vec model's pre-trained vocabulary did not include.

3.2.2 Sentiment and Tone

Sentiment and tone refer to the emotional aspects of language that contribute to the overall mood of a statement. In humour analysis, understanding the sentiment and tone helps determine whether a statement is intended to be funny, sarcastic, or even offensive.

1. **Polarity:** This is the inherent positive, negative, or neutral nature of a sentiment or opinion expressed in a piece of text, speech, or communication. It involves assessing whether the content conveys a positive or negative sentiment towards a particular subject, topic, or situation. The appearance of a positive meaning word in a negative context or vice versa has greatly been used to identify sarcasm [44]. Zhang et al. [99] extracted the affective polarity which consists of emotional polarity, and intensity. Jariwala [44] used polarity as part of the features for sarcasm detection.

2. **Subjectivity:** This refers to the fact that the perception and interpretation of what is humorous can vary from person to person based on their individual preferences, experiences, cultural background, and personal context. Humour is inherently subjective, and what one person finds funny, another might not. This subjectivity adds depth and complexity to the study and analysis of humour. In the paper by Zhang et al. [99] subjectivity was computed in the range of 0 to 1, where 0 represents very objective and 1 represents very subjective. Ziser et al. [100] identified the subjectivity feature based on the sentiment polarity.

3. **Interpersonal Effect:** Interpersonal effects in terms of humour refer to the impact that humour has on social interactions, relationships, and communication between individuals. One important concept of humour is its social or hostility focus, especially regarding its effect on receivers. The interpersonal effect essentially shows that humour is associated with sentiment and subjectivity. The interpersonal effect feature was extracted by Yang *e al*[97] based on the words polarity and subjectivity.

4. **Sentiment-based features:** Sentiment refers to the emotional tone or attitude expressed in a piece of text [76, 5, 96]. It encompasses a range of emotions beyond just positive and negative, including joy, anger, surprise, and more [5]. This feature was used by Bouazizi and Otsuki [13], Kumar et al. [59], Jariwala [44], Hasan et al. [40], Ghosh and Veale [37], Joshi et al.[45], Riloff et al.[84] and Du et al. [32] for sarcasm detection.

5. **Emoticons:** These are textual symbols created by combining characters and symbols to represent facial expressions, objects, or concepts. It is used to convey feelings or ideas. This feature was used by Swami et al. [88] for sarcasm detection.

6. **Acoustic Features:** This refers to various attributes of sounds or speech patterns. These features are extracted from audio data and can provide valuable insights into the comedic effect and emotional tone of the content. This feature is used in multi-modal humour identification. Acoustic features encompass aspects such as pitch, intonation, rhythm, tempo, and prosody, which contribute to the overall delivery and interpretation of humour. In addition to textual cues, acoustic features add an extra layer of complexity to the understanding of humour. They help convey nuances that may not be fully captured through text alone. The following articles made use of the acoustic features for humour classification: [40, 95, 27, 50, 18, 9, 24, 61].

7. **Visual Features:** This refers to the visual cues and elements present in humorous content, particularly in images, videos, or other visual media. These features play a significant role in enhancing the comedic effect and conveying humour to the audience. Visual humour often relies on the juxtaposition of images, unexpected visuals, or exaggerated expressions to create laughter and amusement. Integrating visual cues with text and audio-based humour analysis provides a more comprehensive understanding of humour across different modalities. The literature that utilised the visual features for humour includes: [40, 95, 27, 50, 18, 60, 21, 24, 61]

3.2.3 Text Representation

Text representation involves transforming textual content into numerical forms that can be processed by machine learning algorithms. In humour analysis, text representation methods help capture the underlying patterns and features that contribute to humour.

1. **Word Embedding:** This entails the representation of words as dense vectors of real numbers in a continuous vector space. Word embedding captures the semantic relationship between words in a sentence. The papers that made use of word embedding for sarcasm detection include [65, 38, 23, 9, 83, 37, 34, 87] while the papers that used this feature for humour identification are [49, 35, 27, 16, 24, 3, 50, 87].

2. **Deep Features:** Deep features refer to complex and abstract representations of textual, visual, or auditory content that have been learned by deep learning models. These features capture high-level patterns [53], relationships, and representations that are not directly interpretable by humans but are useful for humour and sarcasm identification tasks. For sarcasm detection, the papers that made use of deep features are [59, 1, 30, 37, 26, 75, 87]. While for humour detection, the literature that used deep features are [95, 50, 18, 60, 21, 90, 61, 3, 87]

3. **Bag of Word (BOW):** is a text representation technique used to represent text data as numerical vectors [79]. In BOW, each sentence or phrase is represented as a vector, where each dimension corresponds to a word in the vocabulary, and the value in the vector represents the frequency of that word in the document [79]. This feature was used by Ghosh and Veale [37] for sarcasm detection.

Table 5 provides a summary of the features used in the primary literature to detect either humour styles, humour, sarcasm, or related-tasks. Table 5 includes the major categories of the extracted features, including the features themselves, the tasks they were used for, the papers they were used in, and the overall number of times they were used.

Table 5: Summary of Extracted Features

Categories	Features	Humour	Sarcasm	Humour Style	Number of studies
Language-based	Contextual Information	[99]			1
	Incongruity	[16, 43, 68, 97, 100]			5
	Ambiguity	[35, 36, 40, 68, 94, 97]	[7]		7
	Phonetic Style	[68, 94, 97]			3
	Exaggeration			[49]	1
	Part-of-Speech		[77, 44, 37]	[49, 2]	5
	N-gram		[7, 77, 88, 44, 45, 84]		6
	Punctuation Features		[7, 59, 77, 44]		4
	TF-IDF	[4]	[73]		2
	Semantic Similarity	[94]			1
	OOV words	[70]			1
Sentiment and Tone	Polarity	[99]	[44]		2
	Subjectivity	[99, 100]			2
	Interpersonal Effect	[97]			1
	Sentiment-based	[40]	[13, 59, 32, 44, 45, 37, 84]		8
	Emoticons		[88]		1

13

	Acoustic Features	[27, 40, 50, 95, 29, 51, 9, 24, 61]	[18]	[28]	11
	Visual Features	[27, 40, 50, 95, 22, 29, 51, 74, 24, 61]	[18, 22, 59]	[28]	14
Text Representation	Word Embedding	[16, 27, 35, 49, 9, 24, 50, 27, 87]	[65, 38, 83, 34, 87]		14
	Deep Features	[18, 21, 50, 60, 95, 29, 51, 74, 90, 61, 87]	[1, 30, 59, 37, 26, 75, 87]	[28]	19
	Bag of Words		[37]		1

3.3 Commonly Employed Computational Techniques

In this section, we delve into the rich landscape of computational techniques that have been explored in existing works within the realm of humour style classification and related tasks such as humour and sarcasm recognition. These techniques encompass a diverse range of methodologies, from traditional machine learning (ML) approaches to state-of-the-art neural network models, transformer models, and specialised models tailored specifically to humour and sarcasm tasks.

3.3.1 Traditional Machine Learning (TML) Models

These techniques encompass classical machine learning algorithms. These traditional ML models are well established and interpretable. However, they require manual feature engineering to capture humour-related patterns effectively. The different traditional ML models used in the reviewed papers are Support Vector Machine (SVM), K-nearest neighbour (KNN), Random Forest (RF), Decision Tree (DT), Naive Bayes (NB), Logistic Regression (LR) and ensemble models like Bagging and Boosting. The most commonly used TML method for humour detection is the SVM, with 6 papers using it for binary humour classification, followed by NB with 5 papers, RF with 3 papers, Ensemble with 2 papers, and LR, DT, and KNN with 1 paper each. Also for sarcasm identification, the SVM model was mostly adapted with 11 papers utilising it for sarcasm detection, followed by NB with 3 papers. The KNN, RF, and LR were adapted equally in two papers, each using them for sarcasm identification. The least used TML model was the DT and ensemble, with only 1 paper utilising them.

Furthermore, it was noticed that only three papers [28, 45, 18] utilised the TML models precisely SVM for multi-modal humour/sarcasm identification. For multimodal datasets, the best F1-score obtained by the TML (SVM) model was 79.8% [45].

In the identification of a single humour style (self-deprecating), the DT, RF, and Bagging showed to be effective with F1-scores of 94.8% [2], 87% [49], and 93% [2], respectively. It can be inferred that all these models have DT in common, either as a single model or as a base model. This shows that DT is well-suited for the detection of self-deprecating and self-enhancing humour styles. Similarly, Christ et al. [28] used SVM to identify the direction and sentiment of a joke, given that jokes are assigned humour style labels based on their direction and sentiment. For example, the self-enhancing humour style has a positive sentiment and is self-directed, while the aggressive humour style has a negative sentiment and is others-directed.

Overall, the SVM performed well across several papers, with an F1-score range of 93% [77] to 51% [84] for binary sarcasm identification and up to 88% [6] F1-score for binary humour. The least performing model was KNN with an F1-score of 31% [3] for sarcasm detection.

In conclusion, a total of 19 reviewed papers made use of TML models for humour detection, while 22 reviewed papers used them for sarcasm analysis. In addition, the SVM model was the most commonly used model for both sarcasm and humour identification in text datasets because it is effective in high-dimensional spaces, which makes it well-suited for text classification, and it is also robust to overfitting as it can generalise well to unseen data when trained with a small dataset.

3.3.2 Neural Network (NN) Models

NN models have gained prominence for their ability to automatically learn complex patterns from data. These models excel at capturing sequential and contextual information, making them suitable for humour or sarcasm analysis tasks. The NN models adapted in the 59 reviewed papers are: Long short-term memory (LSTM), Convolutional Neural Network (CNN), Gated Recurrent Unit (GRU), Recurrent Neural Networks (RNN), Artificial Neural Network (ANN) and Bidirectional-LSTM. The most used NN model for humour task was LSTM, with 8 reviewed papers using it, followed by CNN with 4 reviewed papers, whereas CNN was the most used for the sarcasm task, with 7 reviewed papers utilising it as compared to LSTM with only 4 papers using it for the sarcasm task.

Out of the 59 reviewed papers, 17 made use of multimodal datasets, and out of these 17 papers, 11 made use of neural network (NN) models for humour-related task. This shows that the NN is the most commonly used computational model for multimodal humour or sarcasm detection. Looking at the NN performance for multimodal datasets, the LSTM model had the best performance with an F1-score of 81.4% [9] for humour and 68.6% [9] for sarcasm detection. The next best performing model is the RNN model, with an F1-score of up to 76.7% [23].

Considering the single modality models, 11 papers made use of NN modals for sarcasm or humour detection on text, while one paper used it on an image dataset [30]. Kamal and Abulaish [47] used LSTM for detection of self-deprecating sarcasm in text. Whereas, Das and Anthony [30] used the CNN model to identify sarcasm in images. The CNN model obtained an accuracy of 84% for this task [30]. The LSTM and CNN models showed promising performance for text modality, with both achieving an accuracy of 84% [100] for the humour classification task and an F1-score and accuracy ranging from 81% to 96% sarcasm-related task [73, 3].

3.3.3 Transformer-Based (TB) Models

These models are built upon a unique architecture that employs self-attention mechanisms and feed-forward neural networks, allowing them to capture complex dependencies and relationships in data. The TB models used in the included literature are as follows: Bidirectional Encoder Representations from Transformers (BERT) [74, 73, 3, 90], Robustly Optimised BERT (RoBERTa) [60, 3, 90], Decoding-enhanced BERT with disentangled attention (DeBERTa) [90], XLNet [40, 3], GPT-2 [94]and A Lite BERT (ALBERT) [40]. Looking at the multimodal datasets, the transformer-based models (TBM) were adapted by only 3 papers [74, 60, 40] and these papers mostly used them as the baseline models for comparison and not as the main classification models. Nonetheless, the TB models produced good classification results with BERT obtaining an accuracy of 67.98% [74] and 69.62% [40] and XLNet having an accuracy of up to 72.43% [40]. Li et al. [60] used a combination of two transformer models for classification. The combinations include RoBERTa + cross-covariance image transformer (XCiT) and RoBERTa + bidirectional encoder representation from image transformers (BEiT).

Out of the 40 included papers that performed humour or sarcasm detection on text datasets 9 papers utilised only a single TB model without a combination with other model types [3, 90, 73, 16, 6, 43, 93, 94, 63]. Two papers [16, 93] used TB models BERT and RoBERTa as Regressors to predict the level of funniness of a humour. Weller and Kelvin [93] predicted the level of funniness of a joke on a scale of 0 to 10 and then got a Root Mean Square Error (RMSE) of 1.619, 1.614 and 1.739 for BERT, RoBERTa, and XLNet, respectively. On the other hand, Cao [16] predicted the funniness of an edited news headline on a scale of 0 to 3 and got RMSE scores of 0.528 and 0.516 for BERT and RoBERTa, respectively. The TB models performed significantly well across the remaining 7 papers that used them as classifiers. For instance, in the work by Annamoradnejad and Zoghi [6] BERT got an accuracy and F1-score as high as 98.2% and XLNet got an accuracy of 91.6% and F1-score of 92%. Meaney et al. [63] BERT's model obtained a 91% accuracy and 92.8% f1-score, while Pandey and Singh [73] got an F1-score of 97% using BERT. Given the recent popularity of GPT models, only Xie et al. [94] used GPT-2 for humour detection. Their GPT-2 model got an accuracy of 83.7% and an F1-score of 83.6%.

Overall, given the performance of TB models on humour-related tasks, it can be concluded that they are well-suited for text-based humour or sarcasm recognition tasks.

Table 6 summarises the computational models that were applied in the initial research on humour and its related task recognition. The table lists the computational models that were employed, the tasks for which they were utilised, and the number of studies that did so. The # in the table means number.

Table 6: Summary of Studies and Computational Model Types

Traditional Machine Learning Models		
Model	Sarcasm # of studies	Humour # of studies	Humour Styles # of studies
SVM	11	6	1
NB	3	5	1
RF	2	3	1
LR	2	1	-
DT	1	1	1
KNN	2	1	-
Others	1	2	1
Neural Networks Models		
Model	Sarcasm # of studies	Humour # of studies	Humour Styles # of studies
LSTM	5	8	1
CNN	7	4	-
GRU	2	1	-
RNN	1	2	-
ANN	1	2	-
Others	5	2	-
Transformer-Based Models		
Model	Sarcasm # of studies	Humour # of studies	Humour Styles # of studies
BERT	2	8	-
RoBERTa	1	3	-
XLNet	1	3	-
GPT-2	-	1	-
Others	-	2	-
Specialised Models		
Model	Sarcasm # of studies	Humour # of studies	Humour Styles # of studies
Specialised models	6	9	-

3.3.4 Specialised Models

Some research efforts on humour-related tasks have focused on developing specialised models tailored specifically to humour or sarcasm tasks. These models go beyond general-purpose natural language processing techniques and instead incorporate domain-specific knowledge and linguistic features that are particularly relevant to humour analysis. These specialised models include: Multi-level Memory Network-based on

Sentiment Semantics (MMNSS) [83], Multimodel Self Attention Model (MSAM) [74], Humour Knowledge Enriched Transformer (HKT) [40],Muti-dimension attention-based BiLSTM (MDA-BiLSTM) [31], Context-Aware Hierarchical Network (CAHN) [27], Modality-Invariant and Specific Representations for Multimodal Sentiment Analysis (MISRMSA) [24], BERT-LSTM [73], Semantic Incongruity and Commonsense Knowledge Network (SICKNet) [43], Multi-dimension Intra-Attention Network (MIARN) [71], Multi-Head Attention (MHA) [82, 16], Contextual Extension of memory fusion Network (C-MFN) [50] and Phonetics and Ambiguity Comprehension Gated Attention Network (PACGA) [36].

These specialised models have widely been used in both multimodal and single modality humour or sarcasm detection. The collective findings from 13 included studies employing these specialised models consistently underline their superiority over general-purpose natural language processing techniques. For instance, SICKNet, with its focus on semantic incongruity detection and commonsense knowledge extraction, demonstrated a significant 3% F1-score improvement over BERT and a remarkable 6% improvement over RoBERTa on the same humour task, as reported by Huang et al. [43]. Similarly, Hasan et al. [40]'s HKT model recorded impressive performance increments of 5% and 8% when compared to XLNet and BERT, respectively, for humour detection in multimodal datasets.

These findings collectively underscore the importance of specialised models in advancing the state of the art in humour and sarcasm detection. The field benefits from these models' ability to capture nuanced linguistic patterns, leverage domain-specific knowledge, and effectively address the unique challenges posed by humour analysis.

3.4   Strengths and Limitations of Existing Computational Approaches

Evaluating the strengths and limitations of existing computational-based approaches, encompassing both features and models, within the domain of humour style classification and related tasks yields the following insights:

3.4.1   Strengths

1. **Feature Diversity:** The extensive array of features, including contextual information, incongruity, and sentiment-based metrics, offers flexibility in capturing the multifaceted aspects of humour styles. This feature diversity allows for a nuanced analysis of humour expressions.

2. **Interpretability:** Traditional ML models such as SVM, NB, and DT provide transparent decision-making processes, aiding in understanding the role of individual features in classifying humour styles. This interpretability can be valuable for model validation and debugging.

3. **Contextual Awareness:** Based on the reviewed papers, the transformer-based models, notably BERT, RoBERTa, and GPT-2, excel at grasping contextual nuances, making them well-suited for humour styles that heavily rely on context.

4. **Multimodal Capability:** Specialised models like MSAM and MISRMSA cater to multimodal sentiment analysis, enabling the incorporation of acoustic and visual cues into humour style classification. This is particularly beneficial for humour expressed across various modalities.

5. **Transferability:** The adaptability of existing models from humour and sarcasm detection to humour style classification demonstrates their robustness and potential for re-purposing. This can accelerate research efforts by building upon established architectures.

3.4.2   Limitations and Challenges

1. **Subjectivity and Ambiguity:** Humour, by its nature, is subjective and context-dependent. Existing models may struggle to capture the diverse interpretations of humour and sarcasm, leading to potential misclassifications or limited generalisations.

2. **Data Biases:** Computational models are highly dependent on training data. Biases present in the training data can perpetuate stereotypes or under-represent certain humour styles, sarcasm, affecting model fairness and performance.

3. **Generalisation:** Models may perform well on specific humour types seen in the training data but struggle with humour styles that exhibit unique linguistic or contextual characteristics. Achieving robust generalisation remains a challenge.

4. **Multimodal Integration:** While specialised models cater to multimodal data, integrating features from different modalities seamlessly remains a complex task. Ensuring a balanced contribution from textual, acoustic, and visual features is challenging.

5. **Model Overhead:** Some advanced models, particularly transformer-based models, may require substantial computational resources and longer training times, limiting their accessibility for researchers with limited computational capabilities.

3.4.3 Areas for Improvement

1. **Diverse Datasets:** Developing diverse and balanced datasets that encompass various humour styles is essential for training models that generalise well across a wide range of comedic expressions.

2. **Bias Mitigation:** Implementing techniques to mitigate biases in training data and models is critical for ensuring fairness in humour style and related-task classification.

3. **Interpretable Models:** Enhancing the interpretability of complex models like transformers can aid in understanding their decision-making processes and improving model trustworthiness.

4. **Multimodal Fusion:** Developing effective strategies for integrating textual, acoustic, and visual features to create a holistic understanding of humour styles is an area ripe for innovation.

5. **Transfer Learning:** Investigating transfer learning approaches to leverage pre-trained models for humour style and related-task classification can optimise resource utilisation and enhance performance.

In conclusion, the strengths and limitations of existing computational-based approaches in humour style and related-task classification highlight the multifaceted nature of this task. Addressing these challenges while capitalising on their strengths will advance the state of the art in computational humour analysis and its applications across various domains.

4 Transferable Features and Models for Humour Style Recognition

Drawing from the established features and computational models within the domains of humour styles, binary humour recognition, and sarcasm detection, several features and models can be inferred to be potentially transferable for humour style recognition. These identified features and models provide a solid groundwork upon which to construct and assess computational methodologies tailored for the specific challenge of humour style classification.

4.1 Transferable Features

1. **Contextual Information:** Context plays a vital role in humour understanding. Contextual features that capture the surrounding text can be highly transferable for humour style classification, as different humour styles often rely on context for interpretation. This feature is especially useful for identifying affiliative humour style, where understanding the context of social interactions is key to recognising humour and camaraderie, and also for identifying aggressive humour styles that rely on the context of criticism or ridicule. Hence, recognising the contextual cues, like a confrontational tone or sharp comments, is essential to discerning the aggressive humour style.

2. **Incongruity:** Incongruity, which represents the element of surprise or unexpectedness in humour, is a fundamental aspect of many humour styles. Models designed for sarcasm detection often consider incongruity. This makes the incongruity feature valuable for recognising aggressive humour styles, as this humour style is greatly based on sarcasm or mockery.

3. **Ambiguity:** Ambiguity comprehension is crucial in humour recognition, especially for humour styles that involve wordplay or double meanings. This feature is relevant for humour styles like self-deprecating humour, where ambiguity or self-criticism can be used to create comedic effect.

4. **Phonetic Style (Alliteration):** Humour entails using phonetic features to produce comic effects, such as sounds, rhythms, wordplay, and speech patterns. This feature is applicable to self-enhancing humour, which may involve playful phonetic elements for self-affirmation.

5. **Part-of-Speech:** Understanding the grammatical structure of text is important for many humour styles. Part-of-speech features can help identify linguistic patterns associated with different humour styles. This feature is useful for identifying self-deprecating or self-enhancing humour, as it often involves specific linguistic patterns and self-referential language like "me", "myself", "I", "we", and "us".

6. **N-gram:** N-gram features capture short sequences of words, which can be relevant for identifying recurring linguistic patterns characteristic of certain humour styles. This feature is valuable for recognising positive sentiment humour styles like self-enhancing and affiliative styles that may involve recurring phrases or inside jokes among a group of friends.

7. **Semantic Similarity:** Semantic similarity measures can help determine the relatedness of words or phrases, which is essential for humour styles involving semantic nuances and word associations. This feature is transferable to the humour styles classification, as it is important for understanding the underlying sentiment in self-enhancing humour and identifying affiliative humour styles that rely on shared meanings.

8. **Sentiment-based and Polarity:** The four humour styles are greatly reliant on sentiment and polarity, given that each humour style can be classified as either having a positive or negative sentiment. For instance, this feature is relevant for negative humour styles (self-deprecating and aggressive), where sentiment reversal and criticism are part of the humour.

9. **Acoustic Features:** The acoustic features that may be transferable for the effective identification of different humour styles are pitch and tone, timing and rhythm, and voice modulation. The acoustic features related to pitch and tone can convey emotional nuances in speech. These features can be relevant for identifying self-enhancing humour, which often involves cheerful and upbeat delivery to enhance one's mood. The timing and rhythm of speech or delivery can influence the perception of humour. These features can be applied to humour styles like self-deprecating and affiliative humour, where timing and rhythm play a role in comedic timing or creating a friendly atmosphere. Voice modulation, including changes in volume and intensity, can be relevant for aggressive humour styles as it may reflect sarcasm or mockery.

10. **Visual Features:** Visual features have been shown to be effective in emotion recognition. Given that the four humour styles are associated with emotions, this makes visual cues valuable for the identification of the various humour styles. The visual features that are transferable from the reviewed papers are facial expressions, gestures, body language, and visual context. Visual features related to facial expressions, such as smiles and laughter, can be indicative of self-enhancing and affiliative humour styles, where humour is often expressed through positive facial cues. Visual cues like sad or sarcastic facial expressions can be indicative of self-deprecating and aggressive humour styles, where humour is often expressed through negative facial cues. Gestures and body language can provide context for humour styles, especially in self-deprecating humour, where individuals may use physical gestures to complement their humour. The visual context of a scene or image can influence the perception of humour. Visual context can be relevant for identifying negative humour styles (self-deprecating and aggressive) in visual content, where the context may contain elements of mockery or sarcasm.

4.2 Transferable Computational Models

Based on the analysis of the included research papers, it is evident that a broad spectrum of existing models, ranging from traditional machine learning models to specialised task-specific models, can be suitably adapted

for the task of humour style classification. This adaptability is supported by several compelling reasons:

1. **Traditional ML Models:** Traditional ML models like SVM, NB, RF, DT, LR, and KNN are highly adaptable to different feature sets. They can utilise a wide range of linguistic and contextual features, which are also relevant for humour style classification. Furthermore, these models are often more interpretable, making it easier to analyse which features contribute to the classification decision. This interpretability can help in understanding how different features relate to humour styles.

2. **Neural Network Models:** Neural network models, particularly recurrent models like LSTM, RNN, and GRU, are adept at capturing sequential and contextual information, which is vital for humour style understanding. They can capture dependencies between words and phrases that contribute to the identification of humour styles. Similarly, neural networks are highly flexible and can be adapted to different input data and tasks. This adaptability makes them suitable for the diverse characteristics of humour styles.

3. **Transformer Models:** Transformer models have been identified as excelling at capturing contextual information and semantic relationships within text. They can comprehend the subtleties and nuances often associated with humour styles. In addition, pre-trained transformer models, such as BERT and RoBERTa, have learned rich linguistic representations from vast amounts of text data. Fine-tuning these models for humour style classification can leverage this pre-trained knowledge.

4. **Specialised Models:** Specialised models are designed with humour, sarcasm, or sentiment analysis tasks in mind. They often incorporate domain-specific knowledge and architectural components that can be directly relevant to humour style classification. Moreover, some specialised models, like MSAM or MISRMSA, are designed for multimodal sentiment analysis. Given that humour styles can be expressed through text, audio, and visuals, these models can handle the multimodal nature of humour style recognition.

In summary, these existing models are transferable for humour style classification because they possess essential qualities such as adaptability, contextual understanding, and task-specific expertise. By fine-tuning these models with relevant datasets and features specific to humour styles, researchers and practitioners can leverage their strengths to recognise and classify humour styles effectively. However, it is crucial to carefully design experiments, pre-process data, and select appropriate features to maximise their transferability.

5 Research Gaps and Opportunities

Within the domain of computational-based approaches for humour style classification and related tasks, this SLR has revealed a spectrum of research gaps and promising opportunities for future investigations. Below, we present a consolidated overview of these identified research gaps and potential directions for future research.

1. **Diverse and Balanced Datasets:** Existing research heavily relies on social media, in particular X (formerly Twitter), and Reddit and English-language datasets, which may not capture the full spectrum of humour across languages and cultures. Future research should focus on constructing diverse and balanced datasets that encompass various humour types and languages, reducing dataset bias. Currently, there is a scarcity of comprehensive humour style datasets that cover all four primary humour styles: affiliative, self-deprecating, aggressive, and self-enhancing. The absence of such datasets hinders the development and evaluation of computational models for humour style classification.

2. **Contextual Analysis of Humour Styles:** Existing studies have not explored humour style classification extensively in the context of mental health and relationships. Future research should investigate how humour styles, especially affiliative and self-enhancing styles, can impact individuals' well-being, self-esteem, intimacy, and mood in relationships and mental health.

3. **Subjectivity and Labeling Bias:** Humour is inherently subjective and context-dependent, posing challenges to accurately labelling datasets. Future research should address the subjectivity of labelling humour styles and their related tasks and develop techniques to mitigate labelling bias, ensuring the quality of annotated datasets.

4. **Exploration of Understudied Humour Styles:** While self-deprecating humour has received significant attention, there is a lack of research into the less explored humour styles, such as aggressive, affiliative, and self-enhancing. Investigating these understudied styles is vital for a comprehensive understanding of humour.

5. **Multiclass Classification Models:** Current research predominantly relies on binary classification for humour. Future investigations should focus on the development of multiclass classification models that differentiate between affiliative, self-deprecating, aggressive, and self-enhancing humour styles.

6. **Multimodal Data Integration:** Humour often involves multiple modalities, including text, audio, video, and images. Future investigations should explore effective strategies for integrating these modalities seamlessly to achieve a holistic understanding of humour.

7. **Specific Feature Identification:** To improve classification accuracy, it is essential to identify features specific to different forms of humour, such as sarcasm, irony, and wordplay. Tailoring features to each form of humour can enhance the precision of classification models.

8. **Context Detection:** Developing models capable of identifying appropriate contexts for humour and sarcasm usage is crucial for natural human-machine communication. Understanding when and where humour is suitable enhances the user experience.

9. **Cross-Cultural and Multilingual Analysis:** Investigating the cross-cultural and multilingual aspects of humour styles and its related-task can lead to better translation and interpretation of humour across languages and cultures. This research direction involves creating language-specific datasets and exploring how humour styles vary globally.

10. **Enhanced Feature Engineering:** Future research should explore advanced feature engineering techniques that consider not only linguistic features but also common sense, background knowledge, and emerging text modalities like emoji to improve humour style classification.

11. **Domain-Aware Datasets:** Researchers can work on creating domain-specific humour datasets to investigate the adaptation of humour styles in different contexts, such as humour in legal documents, medical reports, or technical literature.

12. **Humour Style-Based Interventions:** Future research can explore the potential of using humour style-based interventions in mental health and relationship counselling. Developing AI systems that adapt humour styles to individuals' needs and preferences can be beneficial.

In conclusion, the field of humour style classification and related tasks offers numerous research gaps and exciting opportunities for future investigations. Developing comprehensive datasets, exploring understudied humour styles, mitigating labelling bias, and improving contextual understanding are crucial steps towards a more comprehensive and nuanced analysis of humour styles and its related-tasks and their implications in various domains, including mental health and relationships.

6 Conclusion

This SLR has illuminated the dynamic landscape of computational-based approaches to humour style classification and related tasks. It encapsulated not only the richness of features and model advancements but also the enduring challenges of subjectivity, bias, and multimodal integration. Crucially, it underscored research gaps calling for comprehensive humour style datasets, exploration of contextual dimensions, multiclass classification, seamless multimodal integration, and cross-cultural analysis. The SLR serves as a clarion call to further explore this evolving field, advancing nuanced computational models and applications with profound implications. Moreover, it identified features and computational models for related tasks (binary humour and sarcasm detection) that are transferable for humour style classification tasks. Features include incongruity, sentiment and polarity, ambiguity, acoustic, visual, contextual information, and more, while computational models encompass traditional ML models, neural network models, transformer-based models, and specialised models. Lastly, it also identified existing humour and sarcasm datasets, providing easy access to resources for future research endeavours.

References

[1] Ibrahim Abu-Farha and Walid Magdy. "From Arabic Sentiment Analysis to Sarcasm Detection: The ArSarcasm Dataset". In: *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools.* 2020, pp. 11–16. URL: https://github.com/iabufarha/ArSarcasm.

[2] Muhammad Abulaish and Ashraf Kamal. "Self-Deprecating Sarcasm Detection: An Amalgamation of Rule-Based and Machine Learning Approach". In: *Proceedings - 2018 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2018.* Institute of Electrical and Electronics Engineers Inc., Jan. 2019, pp. 574–579. ISBN: 9781538673256. DOI: 10.1109/WI.2018.00-35.

[3] Ravinder Ahuja and S. C. Sharma. "Transformer-Based Word Embedding With CNN Model to Detect Sarcasm and Irony". In: *Arabian Journal for Science and Engineering* 47.8 (Aug. 2022), pp. 9379–9392. ISSN: 21914281. DOI: 10.1007/s13369-021-06193-3.

[4] Dan Alexandru and Daniela Gîfu. "Tracing humor in edited news headlines". In: *Smart Innovation, Systems and Technologies.* Vol. 197. Springer Science and Business Media Deutschland GmbH, 2021, pp. 187–196. ISBN: 9789811573828. DOI: 10.1007/978-981-15-7383-5{_}16.

[5] Nayan Varma Alluri and Neeli Dheeraj Krishna. "Multi Modal Analysis of memes for Sentiment extraction". In: *Proceedings of the IEEE International Conference Image Information Processing.* Vol. 2021-November. Institute of Electrical and Electronics Engineers Inc., 2021, pp. 213–217. ISBN: 9781665433617. DOI: 10.1109/ICIIP53038.2021.9702696.

[6] Issa Annamoradnejad and Gohar Zoghi. "ColBERT: Using BERT Sentence Embedding in Parallel Neural Networks for Computational Humor". In: (Apr. 2020). URL: http://arxiv.org/abs/2004.12765.

[7] Francesco Barbieri, Horacio Saggion, and Francesco Ronzano. "Modelling Sarcasm in Twitter, a Novel Approach". In: *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis.* Association for Computational Linguistics, 2014, pp. 50–58. URL: http://sempub.taln.upf.edu/tw/wassa2014/.

[8] Alexandru Costin Băroiu and Ștefan Trăușan-Matu. *Automatic Sarcasm Detection: Systematic Literature Review.* Aug. 2022. DOI: 10.3390/info13080399.

[9] Manjot Bedi et al. "Multi-modal Sarcasm Detection and Humor Classification in Code-mixed Conversations". In: *IEEE Transactions on Affective Computing* (2021). ISSN: 19493045. DOI: 10.1109/TAFFC.2021.3083522.

[10] Tristan A. Bekinschtein et al. "Why clowns taste funny: The relationship between humor and semantic ambiguity". In: *Journal of Neuroscience* 31.26 (June 2011), pp. 9665–9671. ISSN: 02706474. DOI: 10.1523/JNEUROSCI.5058-10.2011.

[11] Vladislav Blinov, Valeriia Bolotova-Baranova, and Pavel Braslavski. "Large Dataset and Language Model Fun-Tuning for Humor Recognition". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.* 2019, pp. 4027–4032. URL: https://www.e1.ru/talk/forum/.

[12] Yevgen Bogodistov and Florian Dost. "Proximity begins with a smile, but which one? Associating non-Duchenne smiles with higher psychological distance". In: *Frontiers in Psychology* 8.AUG (Aug. 2017). ISSN: 16641078. DOI: 10.3389/fpsyg.2017.01374.

[13] Mondher Bouazizi and Tomoaki Otsuki. "A Pattern-Based Approach for Sarcasm Detection on Twitter". In: *IEEE Access* 4 (2016), pp. 5477–5488. ISSN: 21693536. DOI: 10.1109/ACCESS.2016.2594194.

[14] Daniel Bressington et al. *The effects of group-based Laughter Yoga interventions on mental health in adults: A systematic review*. Oct. 2018. DOI: 10.1111/jpm.12491.

[15] Arnie Cann and Chelsea Matson. "Sense of humor and social desirability: Understanding how humor styles are perceived". In: *Personality and Individual Differences* 66 (2014), pp. 176–180. ISSN: 01918869. DOI: 10.1016/j.paid.2014.03.029.

[16] Di Cao. "Self-Attention on Sentence Snippets Incongruity for Humor Assessment". In: *Journal of Physics: Conference Series*. Vol. 1827. 1. IOP Publishing Ltd, Mar. 2021. DOI: 10.1088/1742-6596/1827/1/012072.

[17] Santiago Castro et al. "A Crowd-Annotated Spanish Corpus for Humor Analysis". In: *Proceedings ofthe Sixth International Workshop on Natural Language Processing for Social Media*. Association for Computational Linguistics, 2018, pp. 7–11. URL: https://pln-fing-udelar.github.io/.

[18] Santiago Castro et al. "Towards Multimodal Sarcasm Detection (An Obviously Perfect Paper)". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019, pp. 4619–4629. URL: https://github..

[19] Arjun Chandrasekaran et al. "We are humor beings: Understanding and predicting visual humor". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2016-December. IEEE Computer Society, Dec. 2016, pp. 4603–4612. ISBN: 9781467388504. DOI: 10.1109/CVPR.2016.498.

[20] Dhivya Chandrasekaran and Vijay Mago. "Evolution of Semantic Similarity – A Survey". In: *Association for Computing Machinery*. 1.1 (Apr. 2020), pp. 1–35. DOI: 10.1145/3440755. URL: http://arxiv.org/abs/2004.13820%20http://dx.doi.org/10.1145/3440755.

[21] Dushyant Singh Chauhan, Asif Ekbal, and Pushpak Bhattacharyya. "All-in-One: A Deep Attentive Multi-task Learning Framework for Humour, Sarcasm, Offensive, Motivation, and Sentiment on Memes". In: *Proceedings ofthe 1st Conference ofthe Asia-Pacific Chapter ofthe Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, Dec. 2020, pp. 281–290. URL: http://www..

[22] Dushyant Singh Chauhan, Asif Ekbal, and Pushpak Bhattacharyya. *All-in-One: A Deep Attentive Multi-task Learning Framework for Humour, Sarcasm, Offensive, Motivation, and Sentiment on Memes*. Tech. rep., pp. 281–290. URL: http://www..

[23] Dushyant Singh Chauhan et al. "An emoji-aware multitask framework for multimodal sarcasm detection". In: *Knowledge-Based Systems* 257 (Dec. 2022). ISSN: 09507051. DOI: 10.1016/j.knosys.2022.109924.

[24] Dushyant Singh Chauhan et al. "M2H2: A Multimodal Multiparty Hindi Dataset For Humor Recognition in Conversations". In: *Proceedings of the ACM on Human-Computer Interaction*. Vol. 5. CSCW1. Association for Computing Machinery, Apr. 2021, pp. 1–5. DOI: 10.1145/1122445.1122456.

[25] Dilpreet Chawla et al. *Short Text-Corpus For Humor Detection*. Sept. 2019.

[26] Zheng Lin Chia et al. "Machine Learning and feature engineering-based study into sarcasm and irony classification with application to cyberbullying detection". In: *Information Processing and Management* 58.4 (July 2021). ISSN: 03064573. DOI: 10.1016/j.ipm.2021.102600.

[27] Akshat Choube and Mohammad Soleymani. "Punchline Detection using Context-Aware Hierarchical Multimodal Fusion". In: *ICMI 2020 - Proceedings of the 2020 International Conference on Multimodal Interaction*. Association for Computing Machinery, Inc, Oct. 2020, pp. 675–679. ISBN: 9781450375818. DOI: 10.1145/3382507.3418891.

[28] Lukas Christ et al. "Multimodal Prediction of Spontaneous Humour: A Novel Dataset and First Results". In: *Transactions on Affective Computing* 20.10 (Sept. 2022), pp. 1–15. URL: http://arxiv.org/abs/2209.14272.

[29] Lukas Christ et al. "The MuSe 2022 Multimodal Sentiment Analysis Challenge: Humor, Emotional Reactions, and Stress". In: *MuSe 2022 - Proceedings of the 3rd International Multimodal Sentiment Analysis Workshop and Challenge*. Association for Computing Machinery, Inc, Oct. 2022, pp. 5–14. ISBN: 9781450394840. DOI: 10.1145/3551876.3554817.

[30] Dipto Das and Anthony J. Clark. "Sarcasm detection on Flickr using a CNN". In: *ACM International Conference Proceeding Series*. Association for Computing Machinery, Aug. 2018, pp. 56–61. ISBN: 9781450365406. DOI: 10.1145/3277104.3277118.

[31] Yufeng Diao et al. "A Multi-Dimension Question Answering Network for Sarcasm Detection". In: *IEEE Access* 8 (2020), pp. 135152–135161. ISSN: 21693536. DOI: 10.1109/ACCESS.2020.2967095.

[32] Yu Du et al. "An Effective Sarcasm Detection Approach Based on Sentimental Context and Individual Expression Habits". In: *Cognitive Computation* 14.1 (Jan. 2022), pp. 78–90. ISSN: 18669964. DOI: 10.1007/s12559-021-09832-x.

[33] Abbas Edalat. *Self-initiated humorous protocols: New approach for learning to laugh*. Tech. rep. 2021, pp. 1–3.

[34] Christopher Ifeanyi Eke, Azah Anir Norman, and Liyana Shuib. "Context-Based Feature Technique for Sarcasm Identification in Benchmark Datasets Using Deep Learning and BERT Model". In: *IEEE Access* 9 (2021), pp. 48501–48518. ISSN: 21693536. DOI: 10.1109/ACCESS.2021.3068323.

[35] Xiaochao Fan et al. "Humor detection via an internal and external neural network". In: *Neurocomputing* 394 (June 2020), pp. 105–111. ISSN: 18728286. DOI: 10.1016/j.neucom.2020.02.030.

[36] Xiaochao Fan et al. "Phonetics and Ambiguity Comprehension Gated Attention Network for Humor Recognition". In: *Complexity* 2020 (2020). ISSN: 10990526. DOI: 10.1155/2020/2509018.

[37] Aniruddha Ghosh and Tony Veale. "Fracking Sarcasm using Neural Network". In: *Proceedings of NAACL-HLT*. San Diego: Association for Computational Linguistics, June 2016, pp. 161–169. URL: http://www.maltparser.org/.

[38] Debanjan Ghosh, Weiwei Guo, and Smaranda Muresan. "Sarcastic or Not: Word Embeddings to Predict the Literal or Sarcastic Meaning of Words". In: *Proceedings ofthe 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2015, pp. 17–21. URL: https://github.com/debanjanghosh/sarcasm.

[39] William P Hampes. "The Relation Between Humor Styles and Empathy". In: *Europe's Journal of Psychology* 6.3 (2007), pp. 34–45. URL: www.ejop.org.

[40] Kamrul Hasan et al. *Humor Knowledge Enriched Transformer for Understanding Multimodal Humor*. Tech. rep. 2021. URL: www.aaai.org.

[41] N. F. Hordri et al. "A systematic literature review on features of deep learning in big data analytics". In: *International Journal of Advances in Soft Computing and its Applications* 9.1 (2017), pp. 32–49. ISSN: 20748523.

[42] Nabil Hossain, John Krumm, and Michael Gamon. "President Vows to Cut Taxes Hair: Dataset and Analysis of Creative Text Editing for Humorous Headlines". In: *Proceedings of NAACL-HLT 2019*. Minneapolis: Association for Computational Linguistics, June 2019, pp. 133–142. URL: https://cloud.google.com/bigquery.

[43] Penglong Huang et al. "SICKNet: A Humor Detection Network Integrating Semantic Incongruity and Commonsense Knowledge". In: *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*. Vol. 2022-October. IEEE Computer Society, 2022, pp. 288–296. ISBN: 9798350397444. DOI: 10.1109/ICTAI56018.2022.00049.

[44] Vaishvi Prayag Jariwala. "Optimal Feature Extraction based Machine Learning Approach for Sarcasm Type Detection in News Headlines". In: *International Journal of Computer Applications* 177.46 (2020), pp. 975–8887.

[45] Aditya Joshi et al. "Harnessing Sequence Labeling for Sarcasm Detection in Dialogue from TV Series 'Friends'". In: *Proceedings ofthe 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL)*. Association for Computational Linguistics, 2016, pp. 146–155. URL: http://www.imdb.com/title/tt0108778/.

[46] Antony Kalloniatis and Panagiotis Adamidis. "Computational Humor Recognition: A Systematic Literature Review". In: *Research Square* (2023), pp. 1–31. DOI: 10.21203/rs.3.rs-2552754/v1. URL: https://doi.org/10.21203/rs.3.rs-2552754/v1.

[47] Ashraf Kamal and Muhammad Abulaish. "An LSTM-Based Deep Learning Approach for Detecting Self-Deprecating Sarcasm in Textual Data". In: *Proc. of the 16th Intl. Conference on Natural Language Processing*. 2019, pp. 201–210. URL: https://bit.ly/2WsUkUk.

[48] Ashraf Kamal and Muhammad Abulaish. "CAT-BiGRU: Convolution and Attention with Bi-Directional Gated Recurrent Unit for Self-Deprecating Sarcasm Detection". In: *Cognitive Computation* 14.1 (Jan. 2022), pp. 91–109. ISSN: 18669964. DOI: 10.1007/s12559-021-09821-0.

[49] Ashraf Kamal and Muhammad Abulaish. "Self-deprecating Humor Detection: A Machine Learning Approach". In: *Computer Lingustics*. Ed. by Le-Minh Nguyen et al. Vol. 1215. Communications in Computer and Information Science. Singapore: Springer Singapore, 2020, pp. 483–484. ISBN: 978-981-15-6167-2. DOI: 10.1007/978-981-15-6168-9. URL: http://link.springer.com/10.1007/978-981-15-6168-9.

[50] Md Kamrul Hasan et al. "UR-FUNNY: A Multimodal Language Dataset for Understanding Humor". In: *Proceedings ofthe 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. 2019, pp. 2046–2056. URL: www.ted.com.

[51] Alexander Kathan et al. "A Personalised Approach to Audiovisual Humour Recognition and its Individual-level Fairness". In: *MuSe 2022 - Proceedings of the 3rd International Multimodal Sentiment Analysis Workshop and Challenge*. Association for Computing Machinery, Inc, Oct. 2022, pp. 29–36. ISBN: 9781450394840. DOI: 10.1145/3551876.3554800.

[52] Bruce F. Katz. "A Neural Resolution of the Incongruity-resolution and Incongruity Theories of Humour". In: *Connection Science* 5.1 (Jan. 1993), pp. 59–75. ISSN: 13600494. DOI: 10.1080/09540099308915685.

[53] Mary Ogbuka Kenneth et al. "A Systematic Literature Review on Face Morphing Attack Detection (MAD)". In: *Lecture Notes on Data Engineering and Communications Technologies*. Vol. 109. Springer Science and Business Media Deutschland GmbH, 2022, pp. 139–172. DOI: 10.1007/978-3-030-93453-8{_}7.

[54] Vlado Kešelj et al. "N-gram-based Author Profiles for Authorship Attribution". In: *Pacific Association for Computational Linguistics* (2003), pp. 1–10.

[55] Mikhail Khodak, Nikunj Saunshi, and Kiran Vodrahalli. "A Large Self-Annotated Corpus for Sarcasm". In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki: European Language resources Association (ELRA), May 2018. URL: https://www.reddit.com.

[56] Tracy Kubert and Elena Korshakova. *Identifying IPA humor styles*. Oct. 2020. DOI: 10.1002/pra2.411.

[57] Nicholas Kuiper, Gillian Kirsh, and Nadia Maiolino. "Identity and Intimacy Development, Humor Styles, and Psychological Well-Being". In: *Identity* 16.2 (Apr. 2016), pp. 115–125. ISSN: 1532706X. DOI: 10.1080/15283488.2016.1159964.

[58] Nicholas A. Kuiper and Catherine Leite. "Personality impressions associated with four distinct humor styles". In: *Scandinavian Journal of Psychology* 51.2 (Apr. 2010), pp. 115–122. ISSN: 00365564. DOI: 10.1111/j.1467-9450.2009.00734.x.

[59] Avinash Kumar et al. "Sarcasm Detection Using Multi-Head Attention Based Bidirectional LSTM". In: *IEEE Access* 8 (2020), pp. 6388–6397. ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2963630.

[60] Zefeng Li et al. "Memeplate: A Chinese Multimodal Dataset for Humor Understanding in Meme Templates". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 13551 LNAI. Springer Science and Business Media Deutschland GmbH, 2022, pp. 527–538. ISBN: 9783031171192. DOI: 10.1007/978-3-031-17120-8{_}41.

[61] Zhi-Song Liu, Robin Courant, and Vicky Kalogeiton. "FunnyNet: Audiovisual Learning of Funny Moments in Videos". In: *Computer Vision - ACCV 2022: 16th Asian conference on Computer Vision*. Macao, Dec. 2022, pp. 433–450. URL: http://www.lix.polytechnique.fr/vista/projects/2022_accv_liu.

[62] Rod A Martin et al. "Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire". In: *Journal of Research in Personality* 37 (2003), pp. 48–75. URL: www.elsevier.com/locate/jrp.

[63] J A Meaney et al. "SemEval-2021 Task 7: HaHackathon, Detecting and Rating Humor and Offense". In: *Proceedings ofthe 15th International Workshop on Semantic Evaluation (SemEval-2021)*. Bangkok: Association for Computational Linguistics, Aug. 2021, pp. 105–119. URL: https://www.kaggle.com/.

[64] Rada Mihalcea and Carlo Strapparava. *Making Computers Laugh: Investigations in Automatic Humor Recognition*. Tech. rep. 2005, pp. 531–538. URL: http://www.mutedfaith.com/funny/life.htm.

[65] Rishabh Misra and Prahal Arora. "Sarcasm detection using news headlines dataset". In: *AI Open* 4 (Jan. 2023), pp. 13–18. ISSN: 26666510. DOI: 10.1016/j.aiopen.2023.01.001.

[66] Sanjay Misra. "A Step by Step Guide for Choosing Project Topics and Writing Research Papers in ICT Related Disciplines". In: *Third International Conference, ICTA*. Ed. by Sanjay Misra and B Muhammad-Bello. Minna, Nigeria: Springer Nature, 2020, pp. 727–744. URL: http://www.springer.com/series/7899.

[67] Ramon Mora-Ripoll. "Potential health benefits of simulated laughter: A narrative review of the literature and recommendations for future research". In: *Complementary Therapies in Medicine* 19 (2011), pp. 170–177. DOI: 10.1016/j.ctim.2011.05.003.

[68] Alex Morales and Chengxiang Zhai. "Identifying Humor in Reviews using Background Text Sources". In: *Proceedings ofthe 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2017, pp. 492–501. URL: www.yelp.com.

[69] Chitu Okoli. "A guide to conducting a standalone systematic literature review". In: *Communications of the Association for Information Systems* 37.1 (2015), pp. 879–910. ISSN: 15293181. DOI: 10.17705/1cais.03743.

[70] Hugo Gonçalo Oliveira, André Clemêncio, and Ana Alves. "Corpora and Baselines for Humour Recognition in Portuguese". In: *Proceedings ofthe 12th Conference on Language Resources and Evaluation (LREC 2020)*. Marseille: European Language Resources Association, May 2020, pp. 1278–1285. URL: https://ltpf.files.wordpress.com/2011/01/.

[71] Silviu Oprea and Walid Magdy. "iSarcasm: A Dataset of Intended Sarcasm". In: *Association for Computational Linguistics (ACL)* (Nov. 2019). URL: http://arxiv.org/abs/1911.03123.

[72] Shereen Oraby et al. "Creating and Characterizing a Diverse Corpus of Sarcasm in Dialogue". In: *arXiv preprint* (Sept. 2017), pp. 1–11. URL: http://arxiv.org/abs/1709.05404.

[73] Rajnish Pandey and Jyoti Prakash Singh. "BERT-LSTM model for sarcasm detection in code-mixed social media post". In: *Journal of Intelligent Information Systems* 60.1 (Feb. 2023), pp. 235–254. ISSN: 15737675. DOI: 10.1007/s10844-022-00755-z.

[74] Badri N Patro et al. "Multimodal Humor Dataset: Predicting Laughter tracks for Sitcoms". In: *Computer Vision fundation* (2021), pp. 576–585. URL: https://delta-lab-iitk.github.io/.

[75] Rolandos Alexandros Potamias, Georgios Siolas, and Andreas Georgios Stafylopatis. "A transformer-based approach to irony and sarcasm detection". In: *Neural Computing and Applications* 32.23 (Dec. 2020), pp. 17309–17320. ISSN: 14333058. DOI: 10.1007/s00521-020-05102-3.

[76] Tambe Pravin, Dongare Tejaswini, and Tarate Varsha. *Survey on: A Privacy Preserving of Medical Image using watermarking*. Tech. rep. 2015, pp. 2321–0613. URL: www.ijsrd.com.

[77] Tomáš Ptáček, Ivan Habernal, and Jun Hong. "Sarcasm Detection on Czech and English Twitter". In: *25th International Conference on Computational Linguistics*. Ireland, Aug. 2014, pp. 213–223. URL: http://www.mturk.com.

[78] Taivo Pungas. *A dataset of English plaintext jokes*. 2017.

[79] Wisam A. Qadar, Musa A. Ameen, and Bilal I. Ahmed. "An Overview of Bag of Words; Importance, Implementation, Applications, and Challenges". In: *Proceedings of the 5th International Engineering Conference (IEC2019)*. Erbil, 2019, pp. 200–205. ISBN: 9781728143774.

[80] Shahzad Qaiser and Ramsha Ali. "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents". In: *International Journal of Computer Applications* 181.1 (July 2018), pp. 25–29. DOI: `10.5120/ijca2018917395`.

[81] C. Ramakristanaiah et al. "A Survey on Humor Detection Methods in Communications". In: *Proceedings of the 5th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2021*. Institute of Electrical and Electronics Engineers Inc., 2021, pp. 668–674. ISBN: 9781665426428. DOI: `10.1109/I-SMAC52330.2021.9640751`.

[82] Lu Ren et al. "ABML: attention-based multi-task learning for jointly humor recognition and pun detection". In: *Soft Computing* 25.22 (Nov. 2021), pp. 14109–14118. ISSN: 14337479. DOI: `10.1007/s00500-021-06136-y`.

[83] Lu Ren et al. "Sarcasm Detection with Sentiment Semantics Enhanced Multi-level Memory Network". In: *Neurocomputing* 401 (Aug. 2020), pp. 320–326. ISSN: 18728286. DOI: `10.1016/j.neucom.2020.03.081`.

[84] Ellen Riloff et al. "Sarcasm as Contrast between a Positive Sentiment and Negative Situation". In: *In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. 2013, pp. 1–11.

[85] Vassilis Saroglou, Christelle Lacour, and Marie-Eve Demeure. "Bad Humor, Bad Marriage: Humor Styles in Divorced and Married Couples". In: *Europe's Journal of Psychology* 6.3 (2011), pp. 94–121. URL: `www.ejop.org`.

[86] Tabea Scheel and Christine Gockel. *Humor at Work in Teams, Leadership, Negotiations, Learning and Health*. Vol. 1. 2017, pp. 1–151. URL: `http://www.springer.com/series/10143`.

[87] Chhavi Sharma et al. "SemEval-2020 Task 8: Memotion Analysis-The Visuo-Lingual Metaphor!" In: *Proceedings of the 14th International Workshop on Semantic Evaluation*. Barcelona: Online, Dec. 2020, pp. 759–773. URL: `https://chrome.google.com/webstore/detail/fatkun-batch-download-ima/nnjjahlikiabnchcpehcpkdeckfgnohf?hl=en`.

[88] Sahil Swami et al. "A Corpus of English-Hindi Code-Mixed Tweets for Sarcasm Detection". In: *arXiv* (May 2018), pp. 1–9. URL: `http://arxiv.org/abs/1805.11869`.

[89] Martha Kirk Swartz. "The PRISMA statement: A guideline for systematic reviews and meta-analyses". In: *Journal of Pediatric Health Care* 25.1 (2011), pp. 1–2. ISSN: 08915245. DOI: `10.1016/j.pedhc.2010.09.006`. URL: `http://dx.doi.org/10.1016/j.pedhc.2010.09.006`.

[90] Leonard Tang et al. "The Naughtyformer: A Transformer Understands Offensive Humor". In: *axXiv* (Nov. 2022). URL: `http://arxiv.org/abs/2211.14369`.

[91] Pablo Vicente Torres-Carrion et al. "Methodology for systematic literature review applied to engineering and education". In: *IEEE Global Engineering Education Conference, EDUCON* 2018-April.August (2018), pp. 1364–1373. ISSN: 21659567. DOI: `10.1109/EDUCON.2018.8363388`.

[92] Livia Veselka et al. "Relations between humor styles and the Dark Triad traits of personality". In: *Personality and Individual Differences* 48.6 (Apr. 2010), pp. 772–774. ISSN: 01918869. DOI: `10.1016/j.paid.2010.01.017`.

[93] Orion Weller and Kevin Seppi. *The rJokes Dataset: a Large Scale Humor Collection*. Tech. rep. 2020, pp. 11–16. URL: `https://github.com/orionw/rJokesData.`.

[94] Yubo Xie, Junze Li, and Pearl Pu. "Uncertainty and Surprisal Jointly Deliver the Punchline: Exploiting Incongruity-Based Features for Humor Recognition". In: *Proceedings ofthe 59th Annual Meeting ofthe Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing.* 2021, pp. 33–39. URL: http://www.punoftheday.com/.

[95] Haojie Xu et al. "Hybrid Multimodal Fusion for Humor Detection". In: *MuSe 2022 - Proceedings of the 3rd International Multimodal Sentiment Analysis Workshop and Challenge.* Association for Computing Machinery, Inc, Oct. 2022, pp. 15–21. ISBN: 9781450394840. DOI: 10.1145/3551876.3554802.

[96] Zijian Xu and Yunyan Xie. "Attention Method Analysis in Sentiment Analysis for Humor Level Evaluation". In: *2022 IEEE 14th International Conference on Computer Research and Development, ICCRD 2022.* Institute of Electrical and Electronics Engineers Inc., 2022, pp. 178–185. ISBN: 9781728177212. DOI: 10.1109/ICCRD54409.2022.9730184.

[97] Diyi Yang et al. *Humor Recognition and Humor Anchor Extraction.* Tech. rep. 2015, pp. 17–21. URL: http://hosted.ap.org/dynamic/fronts/HOME?SITE=AP.

[98] Jong Eun Yim. *Therapeutic benefits of laughter in mental health: A theoretical review.* July 2016. DOI: 10.1620/TJEM.239.243.

[99] Donghai Zhang et al. "Investigations in automatic humor recognition". In: *Proceedings - 2017 10th International Symposium on Computational Intelligence and Design, ISCID 2017.* Vol. 1 2018-January. Institute of Electrical and Electronics Engineers Inc., July 2017, pp. 272–275. ISBN: 9781538636749. DOI: 10.1109/ISCID.2017.160.

[100] Yftah Ziser, Elad Kravi, and David Carmel. "Humor Detection in Product Question Answering Systems". In: *SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.* Association for Computing Machinery, Inc, July 2020, pp. 519–528. ISBN: 9781450380164. DOI: 10.1145/3397271.3401077.

A Appendices

A.1 Review of Primary Papers

Table 7: Reviewed Papers

Paper	Task	Dataset	Feature	Computational Model	Performance Metrics
[74]	Humour	MHD	Deep, visual	BERT, LSTM, Multimodal Self Attention	F1-Score, Accuracy, ROC
[40]	Humour	UR-FUNNY, MUStARD	Sentiment, Ambiguity, Acoustic, Visual	ALBERT, BERT, XL-Net, Humour Knowledge Enriched Transformer(HKT)	Accuracy
[61]	Humour	MHD, MUStARD, UR-FUNNY, Friends	Deep, Acoustic, Visual	FunnyNet (Cross-Attention Fusion)	F1-Score, Accuracy, Precision, Recall
[18]	Sarcasm	MUStARD	Deep, Acoustic, Visual	Random Majority, SVM	F1-Score, Precision, Recall

Table 7: Reviewed Papers (Continued)

Paper	Task	Dataset	Feature	Computational Model	Performance Metrics
[23]	Sarcasm	SEEmoji MUS-tARD	Word embedding, Acoustic, Visual, emoji Embedding	RNN (Gated Multi-modal Unit)	F1-Score, Precision, Recall
[28]	Humour Styles	Passau-SFCH	Deep, Acoustic, Visual	Gated Recurrent Units (GRUs), SVM	AUC
[95]	Humour	Passau-SFCH	Deep (BERT), Acoustic (Deep-Spectrum), Visual (VGGface2)	Transformer-BiLSTM, BiLSTM	AUC
[51]	Humour	Passau-SFCH	Acoustic (eGeMAPS, Deep-Spectrum), Visual (VGGface2, Facial Action Units)	GRUs	AUC
[29]	Humour	Passau-SFCH	Deep (BERT), Acoustic (Deep-Spectrum), Visual (VGGface2)	LSTM	AUC
[50]	Humour	UR-FUNNY	Word embedding (Glove), Acoustic (COVAREP), Visual (OpenFace)	Contextual Extension of memory fusion Network (C-MFN)	Accuracy
[27]	Humour	UR-FUNNY	Word embedding (Glove), Acoustic (COVAREP), Visual (OpenFace)	Context-Aware Hierarchical Network (CAHN)	Precision, F1-score, Recall, Accuracy
[87]	Humour, Sarcasm, offensive, motivation	Memotion	Word-embedding (Glove), Deep (CNN)	CNN-LSTM	F1-Score
[60]	Humour	Memeplate	Deep	BERT, ResNet-50, BEit,Faster-RCNN, XCit, WenLan	Accuracy, F1-score
[21]	Humour, Sarcasm, offensive, motivation	Memotion	Deep (BERT, ResNet)	Attention deep model	F1-Score, Precision, Recall
[9]	Humour, sarcasm	MaSaC	Word Embedding (FastText), Acoustic (Librosa)	LSTM	F1-Score, Precision, Recall, Accuracy
[45]	sarcasm	Friends	N-gram, Punctuations, Sentiment	Naive Bayes, SVM	Precision, Recall, F1-score

Table 7: Reviewed Papers (Continued)

Paper	Task	Dataset	Feature	Computational Model	Performance Metrics
[24]	Humour	M2H2	Word embedding (fastText), Visual (ResNet), acoustic (openSMILE)	Dialogue-RNN, LSTM, Modality-Invariant and-Specific Representations for Multimodal Sentiment Analysis	Precision, Recall, F1-score
[30]	Sarcasm	Yahoo Flickr Sarcasm	Deep	CNN	Accuracy
[49]	Humour style	E-tweets	Self-deprecating, Exaggeration, Word Embedding, ambiguity, polarity, subjectivity, phonetic style	RF	Precision, Recall, F1-score
[2]	Sarcasm (Humour style)	E-tweets [77]	Punctuation Features, POS	DT, NB, Bagging	Precision, Recall, F1-score
[47]	Sarcasm (Humour style)	E-tweets [77]	N-gram Features, POS	LSTM	Precision, Recall, F1-score
[6]	Humour	ColBert	Incongruity, Deep	DT, SVM, XGBoost, XLNet, BERT	Accuracy, precision, recall, F1-score
[93]	Humour	rJokes	Deep feature	BERT, roBERTa, XLNet	RMSE, Pearson, spearman
[43]	Humour	rJokes, English Plaintext Jokes	Incongruity, Contextual information, Deep	BERT, RoBERTa, Semantic Incongruity and Commonsense Knowledge (SICKNet)	Accuracy, F1-score
[77]	Sarcasm	Tweets	N-gram, punctuation, emoticons, POS	SVM	F1-score, confidence interval
[83]	Sarcasm	Tweets [77], IAC [72]	Sentiment, Deep, word embedding	Multi-level memory network based on sentiment semantics	F1-score, precision, recall
[70]	Humour	Portuguese jokes	TF-IDF, N-gram, polarity, phonetic style (alliteration), ambiguity, Incongruity, Named Entities	SVM, RF	F1-score, recall, precision
[1]	Sarcasm, Sentiment	Arsarcasm	Deep, sentiment, subjectivity	BiLSTM	F1-score, recall, precision
[71]	Sarcasm	iSarcasm	Deep	LSTM, CNN, Multi-dimension Intra-Attention Network (MIARN)	F1-score, precision, recall

Table 7: Reviewed Papers (Continued)

Paper	Task	Dataset	Feature	Computational Model	Performance Metrics
[31]	Sarcasm	IACv2	Word embedding, OOV words	Muti-dimension attention-based BiL-STM	Precision, Recall, F1-measure
[59]	Sarcasm	SARC	word embedding, sentiment, punctuation	SVM, Multi-Head self-Attention based BiL-STM	F1-score, precision, recall
[3]	Sarcasm	SARC [55], Tweet-Sarcasm [84]	Word embedding, TF-IDF	CNN	Accuracy, f1-score, precision, recall, AUC
[44]	Sarcasm	News headlines	N-gram, POS, polarity, sentiment, punctuation,	SVM	Accuracy, f1-score, precision, recall
[65]	Sarcasm	News headlines	Word embedding, N-gram	Attention-based Hybrid Neural Network	Accuracy
[48]	Sarcasm	Twitter-280	Self-deprecating pattern, word embedding, n-gram	Convolution and Attention with BiGRU	Accuracy, f1-score, precision, recall
[11]	Humour	Stierlitz and FUN	TF-IDF, Deep features	SVM, Universal Language Model Fine-tuning (ULMFiT)	F1-score, recall
[35]	Humour	1600 One-liner [64], Pun of the Day [97]	Word embedding, ambiguity	Internal and external attention neural network	Accuracy, precision, f1-score, recall
[36]	Humour	1600 One-liner [64], Pun of the Day [97]	Ambiguity, Phonetic style, Deep, Word embedding	Phonetics and Ambiguity Comprehension Gated Attention network (PACGA)	Precision, Recall, and F1-score
[99]	Humour	1600 One-liner	Contextual knowledge, polarity, subjectivity, phonetic, incongruity, ambiguity	Attention-based BiGRU	Precision, recall, f1-measure
[82]	Humour	1600 One-liner	POS, word embedding	Multi-Head Attention (MHA)	Precision, recall, F1-score
[97]	Humour	Pun of the Day	Incongruity, Ambiguity, Phonetic Style, Polarity, subjectivity, BOW, Word embedding	KNN	Accuracy, f1-score, precision, recall
[100]	Humour	Product Question Answering	Deep, Incongruity, Subjectivity, N-gram	LR, NB, LSTM, CNN	Accuracy, Recall
[63]	Humour	SemEval-2021 Task 7	BOW, TF-IDF, Deep	NB, SVR, BERT	Accuracy, F1-score
[94]	Humour	SemEval-2021 Task 7	Word embedding, Semantic similarity, Phonetic style, ambiguity	SVM, GPT-2	F1-score, precision, recall, accuracy

Table 7: Reviewed Papers (Continued)

Paper	Task	Dataset	Feature	Computational Model	Performance Metrics
[68]	Humour	Yelp Dataset	Incongruity, Ambiguity, Phonetic Style (Alliteration)	NB, AdaBoost, Perceptron	Accuracy
[42]	Humour	Humicro-edit Dataset	N-gram, word embedding (GloVe)	RF, LSTM	Accuracy
[16]	Humour	Humicro-edit Dataset	Word embedding (GloVe), Incongruity, Deep (BERT)	Multi-Head Attention (MHA),RoBERTa, BERT	RMSE
[4]	Humour	Humicro-edit Dataset	TF-IDF	NB, SVM, NN	Accuracy
[88]	Sarcasm	English- Hindi code-mixed	N-gram, Emoticons	SVM, RF	F1-score
[73]	Sarcasm	English- Hindi code-mixed	TF-IDF, word embedding (GloVe), Deep (BERT)	NB, SVM, RF, DT, KNN, LR, DNN, CNN, LSTM, BERT, BERT-LSTM	F1-score, precision, recall
[90]	Humour	Naughty-former	Deep Feature	BERT, RoBERTa, De-BERTa	Accuracy, precision, recall, F1-score
[72]	Sarcasm	IAC	N-gram, Word embedding (Word2Vec)	SVM	Precision, Recall, F1-Score
[83]	Sarcasm	IAC, Tweets[77]	Word Embedding (GloVe), Sentiment, Deep feature	MMNSS	Precision, Recall, F1-Score
[37]	Sarcasm	Sarcasm-Detection	BOW, POS, Sentiment, Deep	Recursive-SVM, CNN, LSTM, CNN+LSTM+ DNN	Precision, Recall, F1-Score
[26]	Sarcasm	Sarcasm-Detection	TF-IDF, word embedding	CNN	f1-score
[75]	Sarcasm	Sarcasm-Detection, Tweet-Sarcasm	Deep feature	Combined RCNN and RoBERTa	Precision, recall, accuracy, AUC, f1-score
[34]	Sarcasm	Sarcasm-Detection, Tweet-Sarcasm	Word embedding (GloVe), Deep (BERT), Sentiment,	Bi-LSTM	Precision, recall, accuracy, AUC, f1-score
[84]	Sarcasm	Tweet-Sarcasm	Sentiment, N-gram	SVM	Precision, recall, f1-score
[3]	Sarcasm	Tweet-Sarcasm, SARC	Word embedding (GloVe), TF-IDF	NB, LR, SVM, RF, KNN,CNN, LSTM, GRU, LSTM-AM, GRU-AM, BERT, ELECTRA, XLNet, RoBERTa, XLM-RoBERTa, ULMFIT	Precision, recall, accuracy, AUC, f1-score