

# Sentiment and Emotion Analysis in Textual Data: A Recent Systematic Literature Review Method, Model and Application

Wan Azzura Wan Ramli<sup>1</sup>, Rabiah Abdul Kadir<sup>2</sup>, Amalia Amalia<sup>3</sup>, Ang Mei Choo<sup>4</sup>

Institute of Visual Informatic, Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia<sup>1, 2, 4</sup>

Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, 20155 Sumatera Utara, Indonesia<sup>3</sup>

**Abstract**—The analysis of sentiment and emotion has become an important research topic in Natural Language Processing (NLP) due to the rapid growth of textual data generated on digital platforms. Still, despite significant progress, the existing literature remains fragmented across methods, modalities, and application domains, making it difficult to obtain a comprehensive understanding of current research trends. This study presents a structured literature review that synthesizes recent advances in sentiment and emotion analysis of textual data. The review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol and systematically examines studies retrieved from the Web of Science (WoS) and Scopus databases. After screening, eligibility evaluation, and Quality Assessment (QA), 50 primary studies published between 2023 and 2025 were selected for analysis. As such, the findings reveal a clear methodological transition from traditional Machine Learning (ML) techniques toward transformer-based architectures and Large Language Models (LLMs). In addition, recent studies increasingly explore multimodal approaches and context-aware emotion modeling to improve sentiment and emotion detection. Despite these advancements, several challenges remain, including the detection of implicit emotions, dataset imbalance, and domain adaptability. Overall, this review provides a structured synthesis of recent developments in textual sentiment and emotion analysis, identifies key research challenges, and outlines potential directions for future studies.

**Keywords**—Sentiment; emotion analysis; textual data; transformer; large language models

## I. INTRODUCTION

Sentiment and emotion analysis of textual data is a fast-growing area of Natural Language Processing (NLP) that focuses on the recognition and classification of emotions and opinions conveyed in the form of text. This field has become one of the most strongly studied areas owing to the rise in user-generated content on social media and other online platforms. Notably, sentiment analysis or opinion mining is a computational process for extracting subjective information like opinions, attitudes and feelings from text [1], [2], [3]. The main aim is to determine whether the feeling conveyed in a text or passage is good, bad or indifferent. Nevertheless, sentiment analysis may further explore more aspects of how to detect certain feelings like joy, sadness, anger, and surprise [4], [5], [6].

Sentiment and emotion analysis has become complex due to the intricacy of the human language. Context is crucial for correctly understanding sentiments. As an example, the term

“Nothing!” can have various meanings based on the context in which it is applied [7]. In addition, sarcasm and irony are also rather problematic, as they usually imply the expression of an opposite feeling to the literal meaning of the words used [7], [8]. Furthermore, to overcome these problems, scholars use different Machine Learning (ML) and deep learning methods, such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Recurrent Neural Networks (RNNs) [7], [9], [10]. With the help of large datasets, these models are also trained to identify patterns and classify sentiments and emotions with high accuracy.

The analysis of sentiment and emotion in application has a broad spectrum. It is applied in the business industry to measure customer satisfaction and customer feedback, allowing businesses to make valuable decisions and enhance their products and services [9], [10]. Accordingly, sentiment analysis in politics can be used to monitor social behavior of policies and political personalities, which offers a helpful input for the preparation of campaigns [11], [12]. It also helps the healthcare industry, especially through comprehending patient experience and mental health conditions by analyzing social media posts and other texts [13], [14], [15]. Furthermore, sentiment analysis finds application in many areas like marketing, finance, and human-computer interaction, which is why it is versatile and essential [16], [17], [18].

Sentiment and emotion analysis has several challenges despite its developments. One of them is that it requires large, domain-specific datasets to be trained successfully [7], [19]. In line with this, the variety of languages and cultural variations in the expression of emotions means that algorithms must be flexible and precise in various settings [12], [19]. At the same time, ethical factors, such as training data bias and privacy issues [6], [20], are key to the development and use of the sentiment analysis tools. Hence, to address the challenges, researchers continue developing new tools and techniques, including the use of multimodal information (text, facial expressions, body language) and the development of more sophisticated models [11], [13], [21].

In summary, textual sentiment and emotion analysis is a promising and powerful field that employs the most appropriate NLP techniques in the identification of meaningful information from large quantities of unstructured data. It is applied across various industries, and it provides the necessary information to make decisions and learn more about human emotions and

opinions. As the field continues to evolve, the problems of context dependence, language diversity, and ethical concerns must be addressed to develop more precise and robust systems of sentiment analysis. Notably, more developments and applications of sentiment and emotion analysis are sure to emerge in the future, given the current studies and progress in this area. This study differs from previous reviews that often examine sentiment analysis and emotion detection separately. It integrates both perspectives and provides a comprehensive synthesis of recent developments in textual affect analysis. Specifically, this review organizes the literature into four main research streams, namely core detection methods, multimodal sentiment and emotion analysis, transformer-based and Large Language Model (LLM) architectures, and domain-specific applications. In addition, this review focuses on recent studies published between 2023 and 2025, encompassing the rapid shift towards transformer-based approaches driven by LLMs in affective computing. By systematically synthesizing these new developments, this study aims to provide a clearer understanding of the current research directions and methodological trends in sentiment and emotion analysis.

The remainder of this study is organized as follows. Section II presents the Research Questions (RQs) guiding the systematic review. Section III describes the research methodology, including the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol, search strategy, screening process, and Quality Assessment (QA) procedure. Section IV discusses the findings of the systematic review across four major thematic areas. Finally, Section V concludes the study by summarizing key findings, highlighting research gaps, and proposing directions for future research.

## II. RESEARCH QUESTION

A Systematic Literature Review (SLR) is a form of research that relies on RQs to give intellectual structure for the review, and is the conceptual center of the review and the methodological focus. It describes the scope and content of the review, which makes the judgment aware of the choice of publications and eliminates irrelevant or subjectively unrelated evidence. Concurrently, clear RQs will enable a comprehensive and systematic search strategy, which will rule out potential omissions and address selection bias, thus contributing to a good and balanced representation of knowledge applicable in the field. Moreover, RQs provide a logical structure for analysis within which to model, compare, and synthesize results of studies to identify significant patterns, trends, and gaps. In addition, the clarity of RQs assists the researcher in avoiding ambiguity, enhances interpretability and also renders the findings of the review relevant and actionable by making sure that the concept is clear and has a focus on analysis. Notably, they also enhance the transparency and reproducibility of the review process, allowing other researchers to reproduce, confirm or generalize the results. Together, RQs form the basis of the scientific rigor of an SLR to guarantee its consistency with its overall goals, be it mapping the state of the art, methodological assessment, or the unresolved problems in an area of research.

The development of RQs is the most important component of the planning stage of an SLR, as it directly influences the

protocol of the review, the search strategy, and the methodology of synthesis [22]. Since the aim of the present SLR is to organize the search for the state of the art in sentiment and emotion analysis of textual information, the PICO model was selected to inform the formation of the RQ. PICO, also known as Population, Interest, and Context, is a mnemonic model frequently utilized in either qualitative or exploratory reviews to guarantee conceptual accuracy and analytical consistency [23]. Building on this, the PICO framework helps formulate clear, focused, and methodologically effective RQs by breaking the research focus into three main constituent elements, thereby enhancing the efficiency of the literature retrieval and synthesis process. According to this framework, the present study formulated two RQs, which are as follows:

RQ1: How have textual datasets (Population) been analyzed using traditional ML and deep learning techniques for emotion and sentiment detection (Interest) within general-purpose textual analysis contexts such as social media, reviews, and online discourse (Context)?

RQ2: How do multimodal data sources that include text combined with visual or audio elements (Population) enhance emotion and sentiment analysis performance (Interest) in contexts where affective meaning is distributed across multiple modalities, such as social media platforms and human-computer interaction environments (Context)?

RQ3: How do transformer-based models and LLMs applied to large-scale textual corpora (Population) improve fine-grained emotion and sentiment detection (Interest) in advanced natural language understanding contexts, including implicit, contextual, and low-resource language scenarios (Context)?

RQ4: How are domain-specific textual datasets, such as those related to healthcare, education, or online harassment (Population), analyzed for sentiment and emotion patterns (Interest) within applied and sensitive real-world contexts that demand ethical, contextual, and interpretive robustness (Context)?

## III. MATERIAL AND METHODS

The SLR covered in the study is conducted within the framework of the PRISMA system, which is a global standard, offering transparency, methodological rigor and reproducibility to the review process [24]. The PRISMA guidelines permit the usage of a comprehensive and unbiased approach to identifying, screening, and filtering studies, which underlie the validity and reliability of the review findings. The emphasis on systematic processes and the integration of high-quality evidence, notably rigorously designed studies, as PRISMA is concerned with, contributes to the minimization of the selection bias and the enhancement of the overall analytical rigor. Correspondingly, the main databases selected for this study are the Web of Science (WoS) and Scopus due to their comprehensive coverage, high-quality indexing and dependability in academic research.

The PRISMA system offers a framework for the review process and divides it into four successive steps: identification, screening, eligibility, and data abstraction. During the identification phase, a search of databases is conducted on a large scale to encompass all the possibly relevant studies. This is followed by the screening step, where the records are

subjected to set inclusion and exclusion criteria to identify irrelevant studies and fit or poor-quality studies. Following this, the eligibility phase would involve an intensive review of the remaining articles in order to ensure that they satisfy the review criteria for the content. Finally, the data abstraction phase speculates on the systematic identification and combination of useful data from the sampled literature to support a sound analysis and interpretation. Collectively, this methodological illumination may facilitate transparency of method as well as analytic rigor that produce plausible outcomes that may inform future research and professional contexts in a significant manner.

### A. Identification

Following the PRISMA model, the identification stage is the most basic phase of an SLR. It establishes the extent and validity of the evidence against which the following analysis operations will be carried out. In this research, Scopus and WoS are the two authoritative bibliographic databases employed in the identification process, and both are known to be of high quality in terms of indexing, coverage of different journals, and a large extent of relevance to interdisciplinary research. At the same time, a set of keywords developed with consideration of sentiment analysis, emotion analysis, and textual data was used to develop the search strategy. These ensure the highest possible recall and conceptual correspondence with the study objectives. As a result, 1,915 records were acquired during this step, comprising 1,320 Scopus publications and 595 WoS publications, as summarized in Table I. This high introductory yield indicates a growth in scholarly interest in the research of affective computing and text analysis, particularly in socially sensitive and complex studies.

TABLE I. THE SEARCH STRING

<b>Scopus</b>	TITLE-ABS-KEY ("sentiment analysis" OR "sentiment" OR "opinion mining" OR "sentiment classification") AND ("emotion analysis" OR "emotion detection" OR "affect analysis" OR "emotional classification") AND ("textual data" OR "text data" OR "text" OR "documents" OR "written content") AND ( LIMIT-TO ( DOCTYPE , "ar")) AND (LIMIT-TO (PUBSTAGE , "final" ) ) AND (LIMIT-TO ( LANGUAGE , "English" ) ) AND ( LIMIT-TO ( SRCTYPE , "j")) AND (LIMIT-TO ( PUBYEAR , 2023) OR LIMIT-TO (PUBYEAR , 2024) OR LIMIT-TO (PUBYEAR , 2025)) <b>Date of Access: Jan 2026</b>
<b>WoS</b>	TS=("sentiment analysis" OR "sentiment" OR "opinion mining" OR "sentiment classification") AND ("emotion analysis" OR "emotion detection" OR "affect analysis" OR "emotional classification") AND ("textual data" OR "text data" OR "text" OR "documents" OR "written content") and 2025 or 2024 or 2023 (Publication Years) and Article (Document Types) and English (Languages) and Early Access or Retracted Publication (Exclude-Document Types) and Article (Document Types) <b>Date of Access: Jan 2026</b>

The spread of records between the two databases also highlights the complementary advantages of these databases and the rationale for using them together in this review. The added records from Scopus represented a larger percentage, attributable to its broad collection of journals and substantial coverage of conference proceedings, especially in the areas of computer science and data-driven research. Instead, WoS has

fewer entries, though it is marked by extremely narrow parameters for the selection of journals and is interested in high-impact publications, thereby enhancing the overall quality of the obtained literature. As such, this study will influence the bias of databases by combining the two databases, which will provide a better and more balanced picture of available research. Notably, inclusiveness was a careful consideration in the identification phase, as it was acknowledged that studies focusing on sensitivity and emotional issues tend to use various terminologies and approaches to the methods. Such strategic breadth at this point of identification enhances the methodology rigor of the review, and a strong basis for the rest of the screening and eligibility analyses will be conducted, and hence, larger and stronger findings will be drawn.

### B. Screening

The screening phase was conducted to narrow the original list of studies to those eligible for the review according to the scope and quality criteria. Among the 1,915 records that were found in Scopus and WoS, titles and abstracts were carefully filtered after a set of inclusion and exclusion criteria was established. An outcome of this process was 425 records considered of interest for subsequent evaluation, including 242 articles in Scopus and 183 in WoS. In particular, the large drop at this point is due to stringent filtering processes that will keep the studies specifically related to sentiment analysis, emotion analysis, and textual data within a reasonable empirical and methodological framework. In essence, by reducing the range of data at the screening stage, the review increases the degree of analytical focus and maintains a sufficiently wide range of literature that can be synthesized meaningfully.

TABLE II. THE SELECTION CRITERION IS SEARCHING

Criterion	Inclusion	Exclusion
Language	English	Non-English
Time line	2023 – 2025	< 2023
Literature type	Journal (Article)	Conference, Book, Review
Publication Stage	Final	In Press
Subject	Computer Science	Besides Computer Science

Accordingly, 1,490 records were eliminated during the screening process as a result of failure to conform to the set standards. Specifically, articles in languages other than English, publications earlier than 2023, and non-primary research products, including conference papers, books, review articles, and in-press articles, were eliminated, as depicted in Table II. These omissions have been made with a strategic purpose in order to have consistency of the methodology, relevance in time and inclusion of peer-reviewed empirical studies that are reflective of the latest developments in the study of sentiment and emotion analysis. Moreover, 147 duplicate records were identified and removed to avoid duplication and possible bias in the review procedure. Taken together, these screening choices contribute to the internal validity and credibility of systematic reviews by emphasizing quality and current journal articles and reducing the presence of noise from outdated, redundant, and heterogeneous sources. Consequently, the filtered dataset offers a strong and narrow base on which the next stage of eligibility

will be done to guarantee the final selection of studies is theoretically applicable and empirically sound.

### C. Eligibility

The eligibility phase is a highly crucial quality-control phase where both methodological rigor and relevance are carefully examined. During the stage of eligibility, deliberation over a publication year cut-off was done between 2023 and 2025 to guarantee time and currency of relevance and methodology. The rationale behind this choice was the accelerated, disruptive change in automated and AI-driven systems over the last several years, especially after the introduction of transformer-based and generative AI into text analysis and educational tools, making them ubiquitous. Most studies published before this date represent older groups of automated tools and have significant differences in modeling capacity, contextual sensitivity, and analytical sophistication compared to modern methods. In addition, restricting the review to literature published during this period guarantees that the synthesized evidence is sufficiently representative of the affordances of technologies at the moment, current methods of practice, and current trends in human interaction with text. This, in turn, increases the relevance, validity, and applicability of the review results. Based on the screening process, 278 full-text articles obtained from Scopus and WoS underwent thorough eligibility evaluation. At this point, each study was thoroughly reviewed to assess its substantive value, empirical basis, and adherence to the fundamental research focus on sentiment analysis, emotion analysis, and textual data. This higher form of examination goes more profoundly than surface relevance and allows only studies that provide useful theoretical, methodological, or analytical insight to be maintained. The eligibility process, therefore, serves to protect the internal validity of the review by ensuring that the studies included are thematically aligned and sufficiently robust to undergo qualitative synthesis.

As a result of this high-fidelity screening, 228 articles were not further considered. The exclusions were mostly made based on studies falling out of the pertinent sphere of discipline, titles of studies which were conceptually and analytically non-relevant, lacked an abstract which explicitly covered the research objectives and could not locate full-text copies, or lacked empirical support for their claims. Remarkably, these were exclusion criteria designed to provide adequate conceptual dilution and to prevent the impact of the interpretive power of the review by including speculative or methodologically weak studies. Following the systematic filtering, 50 studies were found eligible and included in the qualitative analysis. The latter corpus is a high-quality, well-edited set of studies that covers current tendencies in the sphere of methodology and theoretical approaches. Concurrently, the review reduces the evidence base at the eligibility phase, thereby enhancing analytical rigor, consistency, and credibility, which makes the results of the study more reliable and applicable in future research and practice.

The PRISMA flow diagram illustrates the systematic screening process applied in this study. Duplicate records were removed using reference management tools, followed by manual verification. Titles and abstracts were screened to assess relevance to sentiment and emotion analysis in textual data.

Subsequently, full-text articles were evaluated based on the predefined inclusion and exclusion criteria to ensure methodological relevance and research quality.

### D. Data Abstraction and Analysis

An integrative analysis approach was a fundamental analytical strategy in this study, as it helps examine and synthesize evidence from various qualitative research designs in a systematic way. Notably, the main aim of this strategy was to distinguish salient themes and subtopics that defined the body of literature in existence. Following this, structured data collection served as the starting point of the analytical process, providing a basis for the later development of the themes. To ensure transparency and reproducibility, metadata from the selected studies were systematically extracted and organized in a structured matrix. The extracted information includes author names, publication year, dataset used, model architecture, evaluation metrics, and key findings. This structured dataset served as the basis for thematic synthesis and comparative analysis. The authors performed a careful analysis of 50 selected sources to provide statements, findings and conceptual insights directly related to the scope of the current study, as displayed in Fig. 1. Special focus was placed on the critical evaluation of influential studies on sentiment and emotion analysis in textual data, both in terms of the methods adopted and the results.

After this first analytical step, the methodologies utilized in the studies reviewed and their findings were compared and interpreted in a systematic way. Emergent themes were developed by working together with the authors to bring conceptual sense and situational affiliation, and the use of empirical evidence was well-built into their development of emergent themes. During the process of the analysis, an audit trail was maintained to document analytical decisions, interpretive reflections, emerging questions and methodological considerations, which enhanced transparency and reliability. Additionally, to boost the rigor of the analysis process, the authors have cross-validated the identified themes to determine and remove potential inconsistencies in the thematic framework. In line with this, the thematic synthesis was plausible and robust, as deviations or disjointed readings could be resolved through the process of group discussion and agreement.

### E. Quality of Appraisal

The quality of the selected primary studies was rated according to the methodological framework of Kitchenham and Charters [22], which made it possible to conduct a rigorous and quantitative comparison of the included research. To operationalize it, the contemporary study takes into account the QA framework suggested by [25], comprising six predefined QA criteria specifically created to be applied in SLRs. The criteria have been rated using a score scale of three levels: "Yes" (Y), a score of 1 when the criteria was met fully; "Partly" (P), a score of 0.5 when the criteria was met partially and with certain limitations; and "No" (N), a score of 0 when the criteria was not addressed. Interestingly, this systematic review will provide consistency, transparency and comparability in the review process of assessing the methodological rigor of the selected studies. The QA criteria (QA1-QA6) to assess the included studies are outlined in Table III.

TABLE III. QUALITY ASSESSMENT CRITERIA USED IN THE SYSTEMATIC LITERATURE REVIEW

QA1	Is the purpose of the study clearly stated?
QA2	Is the interest and the usefulness of the work clearly presented?
QA3	Is the study methodology clearly established?
QA4	Are the concepts of the approach clearly defined?
QA5	Is the work compared and measured with other similar work?
QA6	Are the limitations of the work clearly mentioned?

These criteria were adapted from the QA framework proposed by [22] for SLRs. The selected criteria focus on evaluating the clarity of research objectives, methodological transparency, conceptual soundness, comparative evaluation, and the reporting of limitations. This framework has been widely adopted in SLR studies to ensure methodological

consistency and comparability across the included literature. Although the criteria are intentionally generic to allow consistent evaluation across heterogeneous studies, aspects such as datasets, model architectures, and evaluation metrics were further examined during the data extraction and synthesis stages of the review.

The determination of the quality was performed by various professionals independently. Each study was rated according to the stated QA criteria, and the scores were then aggregated to provide the total quality score. A study had to score more than 3.0 cumulatively based on the evaluation across all three experts to qualify to be included in the next phase of the review. This was a predetermined threshold to ensure that only studies that demonstrate a desirable level of methodological rigor and quality of reporting would be further analyzed.

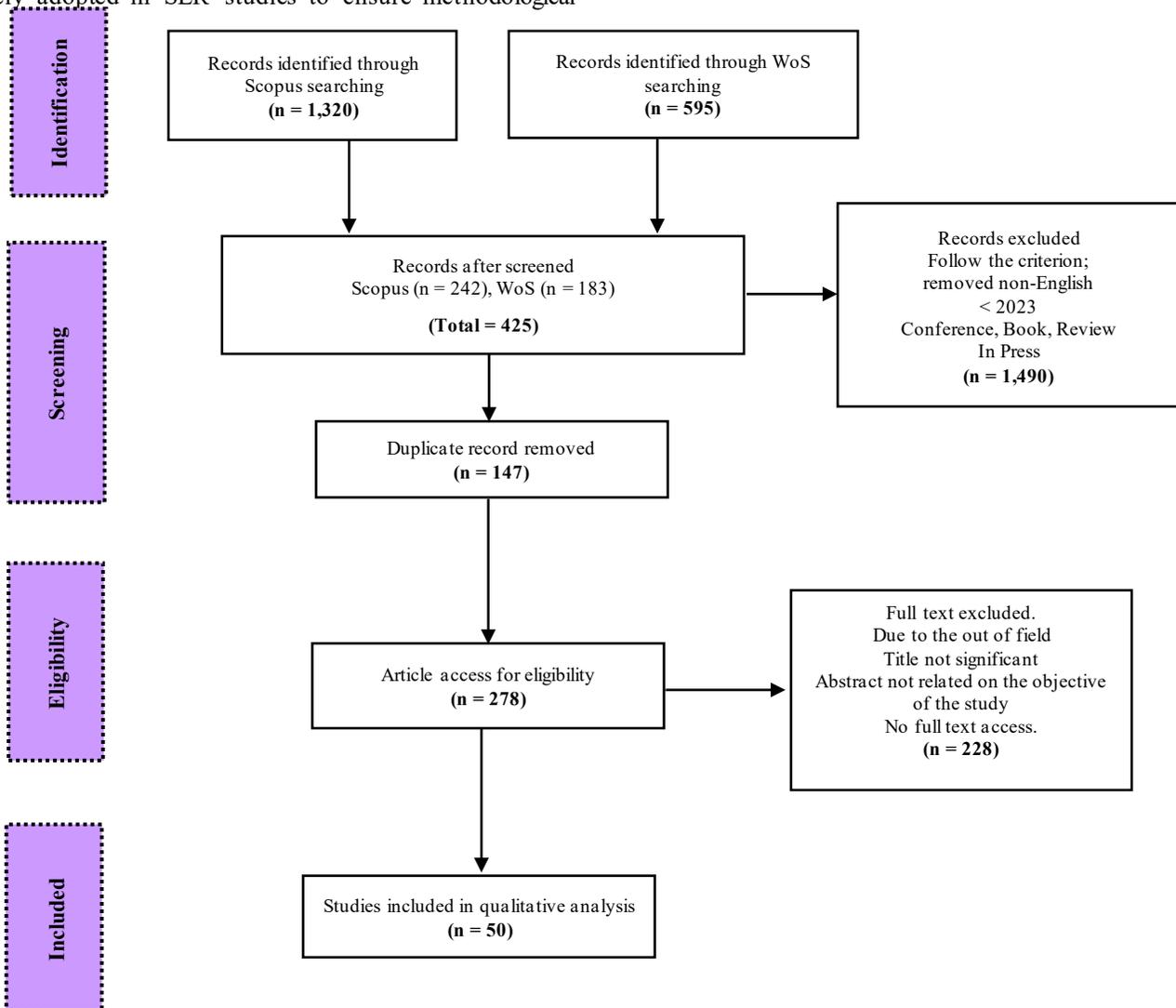


Fig. 1. Flow diagram of the proposed searching study [26].

#### IV. RESULT AND DISCUSSION

The results of the QA of the 50 primary studies (PS1 -PS50) suggest that the overall methodological quality of the selected literature is, in general, high enough and corresponds to the

predetermined inclusion criterion. According to the scoring scheme adopted, all studies obtained a percentage score that was far beyond the rejection point of 50, set as the absolute minimum, and a range of 66.7% to 91.7% is displayed in Table IV. In comparison, studies with the highest scores demonstrated

well-stated research objectives, clear research methodology, and a clear description of the proposed methodology. Meanwhile, moderate changes in scores could be explained by the fact that the partial reporting in comparative evaluations (QA5) and insufficiently discussed limitations of the research (QA6) were frequently omitted in abstracts. However, none of the studies

had a score below 50% quality. Hence, none of the evaluated articles was filtered out during this phase. The result of this outcome proves that all the selected primary studies have an adequate degree of rigor and relevance to be retained in further stages of the SLR.

TABLE IV. PERFORMANCE OF QUALITY ASSESSMENT

RQ	Primary Studies (PS)	Author	QA1	QA2	QA3	QA4	QA5	QA6	Total	%
RQ3	PS1	[27]	Y	Y	Y	Y	P	N	4.5	75.00
RQ3	PS2	[28]	Y	Y	Y	Y	Y	P	5.5	91.70
RQ2	PS3	[29]	Y	Y	Y	Y	Y	N	5.0	83.30
RQ1	PS4	[30]	Y	P	Y	Y	P	N	4.0	66.70
RQ4	PS5	[31]	Y	Y	Y	Y	Y	N	5.0	83.30
RQ1	PS6	[32]	Y	P	Y	Y	P	N	4.0	66.70
RQ4	PS7	[33]	Y	Y	Y	Y	P	N	4.5	75.00
RQ1	PS8	[34]	Y	Y	Y	Y	Y	P	5.5	91.70
RQ4	PS9	[35]	Y	Y	Y	Y	P	N	4.5	75.00
RQ1	PS10	[36]	Y	Y	Y	Y	Y	N	5.0	83.30
RQ2	PS11	[37]	Y	Y	Y	Y	Y	P	5.5	91.70
RQ2	PS12	[38]	Y	Y	Y	Y	P	P	5.0	83.30
RQ3	PS13	[39]	Y	Y	Y	Y	Y	N	5.0	83.30
RQ4	PS14	[40]	Y	P	Y	Y	P	N	4.0	66.70
RQ4	PS15	[41]	Y	Y	Y	Y	P	N	4.5	75.00
RQ2	PS16	[42]	Y	Y	Y	Y	P	N	4.5	75.00
RQ1	PS17	[43]	Y	Y	Y	Y	Y	N	5.0	83.30
RQ2	PS18	[44]	Y	Y	Y	Y	Y	P	5.5	91.70
RQ3	PS19	[45]	Y	Y	Y	Y	Y	P	5.5	91.70
RQ1	PS20	[46]	Y	Y	Y	Y	Y	N	5.0	83.30
RQ2	PS21	[47]	Y	Y	Y	Y	Y	N	5.0	83.30
RQ3	PS22	[48]	Y	Y	Y	Y	Y	N	5.0	83.30
RQ3	PS23	[49]	Y	Y	Y	Y	P	N	4.5	75.00
RQ4	PS24	[50]	Y	Y	Y	Y	Y	P	5.5	91.70
RQ4	PS25	[51]	Y	Y	Y	Y	P	N	4.5	75.00
RQ4	PS26	[52]	Y	Y	Y	Y	Y	P	5.5	91.70
RQ1	PS27	[53]	Y	Y	Y	Y	Y	P	5.5	91.70
RQ4	PS28	[54]	Y	Y	Y	Y	P	N	4.5	75.00
RQ4	PS29	[55]	Y	Y	Y	Y	Y	N	5.0	83.30
RQ3	PS30	[56]	Y	Y	Y	Y	P	N	4.5	75.00
RQ1	PS31	[57]	Y	Y	Y	Y	P	P	5.0	83.30
RQ3	PS32	[58]	Y	Y	Y	Y	Y	N	5.0	83.30
RQ1	PS33	[59]	Y	Y	Y	Y	P	N	4.5	75.00
RQ4	PS34	[60]	Y	Y	Y	Y	P	N	4.5	75.00
RQ2	PS35	[61]	Y	Y	Y	Y	P	N	4.5	75.00
RQ3	PS36	[62]	Y	Y	Y	Y	Y	N	5.0	83.30
RQ4	PS37	[63]	Y	Y	Y	Y	P	N	4.5	75.00
RQ1	PS38	[64]	Y	Y	Y	Y	P	N	4.5	75.00
RQ4	PS39	[65]	Y	Y	Y	Y	P	N	4.5	75.00
RQ4	PS40	[66]	Y	Y	Y	Y	P	P	5.0	83.30
RQ1	PS41	[67]	Y	Y	Y	Y	P	N	4.5	75.00
RQ1	PS42	[68]	Y	Y	Y	Y	P	N	4.5	75.00
RQ1	PS43	[69]	Y	Y	Y	Y	Y	N	5.0	83.30
RQ2	PS44	[70]	Y	Y	Y	Y	Y	N	5.0	83.30
RQ1	PS45	[71]	Y	Y	Y	Y	P	N	4.5	75.00
RQ1	PS46	[72]	Y	Y	Y	Y	Y	P	5.5	91.70
RQ3	PS47	[73]	Y	Y	Y	Y	Y	N	5.0	83.30
RQ3	PS48	[74]	Y	Y	Y	Y	Y	N	5.0	83.30
RQ4	PS49	[75]	Y	Y	Y	Y	P	N	4.5	75.00
RQ3	PS50	[76]	Y	Y	Y	Y	P	N	4.5	75.00

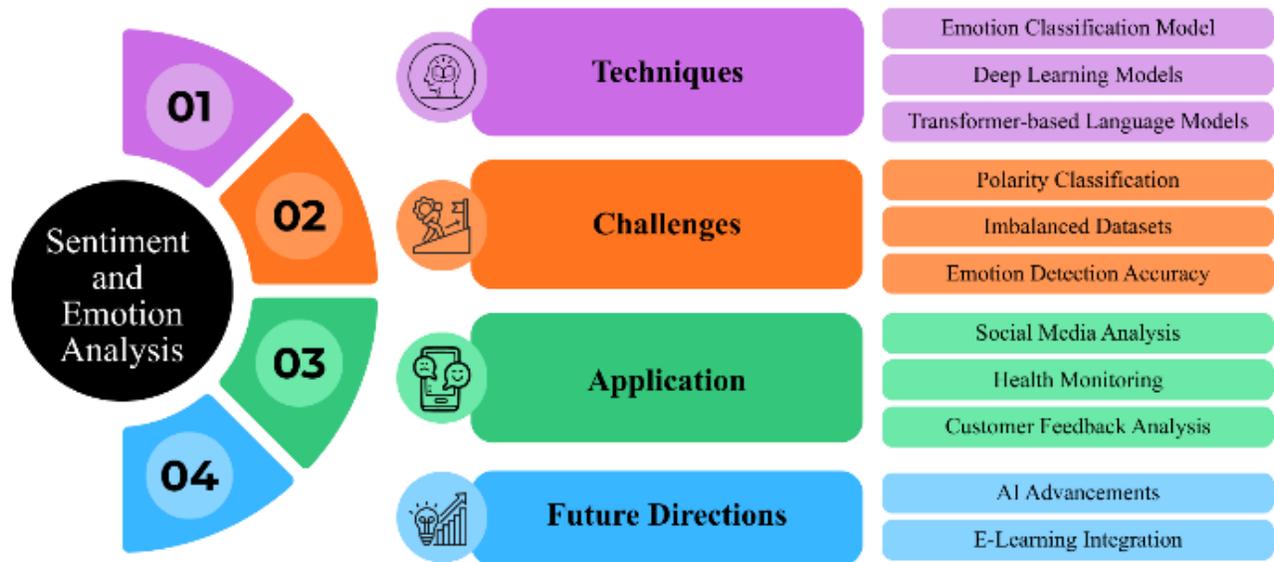


Fig. 2. Conceptual maps of sentiment and emotion analysis research landscape.

### A. Overview of Research Landscape in Sentiment and Emotion Analysis

Fig. 2 presents a conceptual overview of the research landscape in sentiment and emotion analysis derived from the synthesis of the reviewed studies. The figure illustrates the interconnected relationship between four major dimensions identified in the literature, namely analytical techniques, methodological challenges, application domains, and future research directions. In terms of techniques, recent studies increasingly employ deep learning and transformer-based language models for emotion classification and sentiment detection [36], [48], [62]. However, several challenges remain, including limitations in polarity-based sentiment interpretation, imbalanced datasets, and difficulties in accurately detecting complex emotional expressions [57], [59]. The reviewed studies also demonstrate a wide range of application domains, particularly in social media analysis, health monitoring, and customer feedback interpretation [34], [53], [60]. At the same time, emerging research directions highlight the integration of artificial intelligence into adaptive systems such as e-learning platforms and intelligent affect-aware applications [38], [55], [65]. Overall, the conceptual map highlights the dynamic interaction between methodological development and real-world applications, while also revealing existing research gaps that motivate further advancement in sentiment and emotion analysis.

Another observation derived from the reviewed studies concerns the types of datasets commonly used in sentiment and emotion analysis research. Among the selected studies, social media datasets dominate the research landscape, particularly datasets derived from platforms such as Twitter, Reddit, and other online discussion forums [41], [53], [69]. These datasets are widely used because they contain large volumes of user-generated textual data that reflect spontaneous emotional expressions and public opinions. In addition, benchmark datasets such as GoEmotions are frequently utilized for fine-grained emotion classification tasks [48], [64]. Several studies

also rely on domain-specific datasets, including healthcare feedback, e-learning discussions, customer reviews, and political discourse [34], [38], [60]. From a linguistic perspective, the majority of datasets are primarily English-based, although a growing number of studies explore multilingual and low-resource language datasets to support cross-lingual sentiment and emotion analysis [27], [62], [74]. These findings indicate that dataset diversity plays a crucial role in shaping the development and evaluation of sentiment and emotion analysis models.

### B. Core Methods for Emotion and Sentiment Detection in Textual Information

Recent literature on fundamental techniques for emotion and sentiment detection all repeat the claim that conventional sentiment analysis based on polarity is inadequate to reflect delicate emotional expressions in textual information, particularly in informal, domain-specific, or linguistically challenging settings. Results from several studies indicate that linguistic phenomena, including negation, emotional dependency, and feature representation, have a great impact on the performance of classification. According to [30], they proved that negation is mishandled to create systematic misclassification of emotions, and their findings indicate an improvement of measurable accuracy with the explicit transformation of negation terms via antonym replacement. These same issues in shallow linguistic processing are replicated by [59], who discovered that solely statistical measures like Term Frequency-Inverse Document Frequency (TF-IDF) do not sufficiently capture emotional signals in short Indonesian texts, but that semantic representations do considerably improve the accuracy of detection. To add to this point of view, [67] transcended single-emotion forecasting and discloses that emotions tend to be combined in systematic structures. Its results illustrate elevated associative correlations across classes of emotions, which are in Ekmanian posts on social media. Together, these results indicate that emotion detection can be improved with techniques that have an explicit structure that

captures linguistic structure, contextual dependencies, and inter-emotional relations instead of being based extensively on surface-level lexical statistics. These findings indicate that traditional sentiment analysis methods that rely solely on surface-level lexical features often struggle to capture subtle emotional relationships in text, especially when linguistic structures such as denial statements and emotions co-occur.

The other significant literature finding is the increasing efficiency of deep learning and transformer-based models in analyzing emotions and sentiments, especially with task-specific adaptations. Various research works have noted that the hybrid or optimized deep neural models always perform better than traditional ML methods. [34] demonstrated that hybrid systems that integrate deep neural networks with optimization algorithms provide almost state-of-the-art accuracy in patient sentiment analysis with lower computational requirements than large transformer models. Similar improvements in performance are viewed in language-specific settings, [46] discovered that a transfer-based Bidirectional Encoder Representations from Transformers-Bidirectional Long Short-Term Memory (BERT-BiLSTM) model enhances the classification of Persian emotions by directing model attention to features with salient emotional values. During the analysis of literary texts, [36] cited that an actively tuned BERT model is significantly more effective at detecting subtle, rhetorically embedded emotions than lexicon-based and conventional neural approaches. These results are supported by [64], who revealed that transformer variants, especially Robustly Optimized BERT Pre-training Approach (RoBERTa), have a higher ability to classify emotion in a fine-grained manner, even in complex emotional conditions such as gratitude. As a whole, the discourse of these works suggests that deep contextual representations combined with domain adaptation or architecture optimization are at the heart of the further development of emotion detection accuracy in textual genres. This development reflects a broader methodological shift in the field, where research is increasingly prioritizing context-aware neural architectures over traditional feature-based models to better capture nuanced emotional signals.

In addition to the field of architecture, many studies also focus on the significance of emotion-aware representations and data strategies for overcoming the drawbacks associated with data scarcity, emotional flow, and cross-domain generalization. Furthermore, it was demonstrated in [43] that the ability to incorporate emotional distributions as sequential structures allows models to learn the temporal dynamics of emotions in documents, which subsequently improves emotion-detecting and similar emotion-related tasks. This includes irony classification and hate speech classification. The same

representational advantages are evident in [71], who established that emotion-enriched word embeddings restructure semantic spaces better than control embeddings, especially in morphologically rich languages such as Turkish. To overcome the problem of data availability, [69] revealed that distant supervision with emojis can yield large-scale emotion labeling with competitive classification results, while eliminating reliance on expensive manual annotation. The compatible information is given by [57], who stated that simulations of emotion detection by fine-tuned pre-trained models in general beat prompt-only LLMs unless the prompts are carefully designed and the emotions are classified into categories. Lastly, [72] also presented a different view, as it demonstrated that unsupervised optimization-based methods are capable of robust sentiment and emotion classification without massive labeled datasets. Taken together, these arguments describe a shift towards emotionally informed methods of representation, scalable labeling methods, and hybrid learning methods, leading to the maturation of basic methods for analyzing textual emotion and sentiment. Overall, the reviewed studies highlight a clear methodological development in textual emotion and sentiment detection. These early approaches relied heavily on lexical statistics or hand-crafted linguistic features that often faced limitations in capturing contextual emotional cues and their relationships [30], [59]. In contrast, deep learning and transformer-based architectures demonstrated stronger capabilities for modeling dependencies and subtle emotional cues in complex text environments [32], [36], [64]. At the same time, emerging strategies such as emotion-aware embeddings, remote supervision via emoji, and hybrid optimization frameworks can address ongoing challenges related to data scarcity and domain variability [34], [43], [69]. However, despite the progress of this methodology, several challenges remain, such as context-dependent implicit emotion detection and the development of models that can be effectively generalized across domains and languages [57]. The synthesis of the reviewed studies suggests that effective emotion detection relies heavily on integrating contextual modeling, linguistically informed representations, and scalable data strategies. To gain a systemic account of the dynamic research landscape in textual emotion and sentiment detection, a concise overview of the primary methodological approaches, their areas of interest, and key contributions is presented in Table V. Table V indicates that transformer-based architectures such as BERT and RoBERTa consistently demonstrate higher performance across multiple datasets, often achieving accuracy levels above 85%. In comparison, hybrid deep learning models typically report performance between 80% and 90%.

TABLE V. CORE METHODS IN EMOTION AND SENTIMENT DETECTION

Authors	Core Method/ Model	Key Focus	Methodological Contribution	Key Findings
[30]	Negation-assisted ML model	Linguistic structure (negation handling)	Explicit antonym replacement for negation terms	Significant improvement in emotion classification accuracy, achieving 88.75% accuracy, demonstrating the importance of negation handling in sentiment polarity detection.
[59]	Statistical vs Semantic Feature Extraction	Feature representation in short texts	Comparative evaluation of TF-IDF vs semantic embeddings	Semantic features significantly outperform purely statistical features for short-text emotion detection.
[67]	Emotion Association Mining	Inter-emotional relationships	Association rule mining across Ekman's emotion classes	Strong co-occurrence patterns among emotions in social media posts.

[34]	Hybrid Deep Learning	Patient sentiment analysis	Optimized DNN architecture with computational efficiency	Achieved 96% to 98% accuracy, outperforming several transformer baselines while maintaining lower computational cost.
[46]	BERT BiLSTM model	Language-specific adaptation (Persian)	Transfer learning with emotion-focused attention	Achieved 86% accuracy for emotion classification and 88% for binary sentiment classification, outperforming several baseline deep learning models.
[36]	Tuned BERT Model	Literary text sentiment analysis	Dynamic hyperparameter tuning of BERT	Superior detection of subtle, rhetorically embedded emotions.
[64]	Transformer-based Fine-Grained Classification	Fine-grained emotion categories	Advanced neural architectures on the GoEmotions dataset	RoBERTa achieved superior performance in fine-grained emotion classification with F1-score up to 0.8458 on the GoEmotions dataset.
[43]	Distributional Emotion Embeddings	Emotion-aware representation	Embedding emotional distributions into semantic space	Improved emotion-sensitive sentiment analysis.
[71]	Emotion-enriched Word Embeddings	Morphologically complex language (Turkish)	Integration of emotional features into embedding training	Reorganized semantic space improves classification performance.
[69]	Distant Supervision + Deep Learning	Large-scale emotion labeling	Emoji-based automatic labeling strategy	Competitive performance with reduced manual annotation cost.
[57]	Fine-tuned vs Prompt-based LLMs	Pre-trained model comparison	Systematic evaluation of prompting vs fine-tuning	Fine-tuned models consistently outperform prompt-only LLMs unless carefully structured prompts are applied.
[53]	Metaheuristic-optimized GNN	Large-scale Twitter emotion analysis	A Hadoop-based distributed framework with optimization	Scalable and improved performance in social media emotion analysis.
[68]	Pretrained Deep Learning Models	Comparative architecture evaluation	Evaluation of multiple pretrained DL models	Pretrained transformer models such as RoBERTa achieved 75% accuracy, outperforming several traditional ML baselines.
[72]	Unsupervised Optimization (Game Theory + MCDM)	Label scarce environments	Optimization-based unsupervised sentiment classification	Robust performance without large labeled datasets.
[32]	Emotion Vocabulary Recognition	Emotion lexicon recognition	B-BLLC-CL algorithm for emotion vocabulary extraction	Enhanced detection of emotion-bearing lexical units.

### C. Multimodal Emotion and Sentiment Analysis

The results of recent multimodal emotion and sentiment analysis research studies all emphasize the fact that single-modality models are insufficient to describe the depth of human affect in the context of real-world digital content. Studies of social media videos and multimedia on the internet indicate that emotional messages can be expressed through a combination of textual, visual and sound signals, and breaking down the text, photo, or sound can easily lead to an interpretation that is either incomplete or inaccurate. For example, when using videos on social media such as TikTok or Instagram reels, emotions are often conveyed simultaneously through facial expressions, tone of voice, background music, and caption text. Therefore, models that integrate these multimodal signals can interpret deeper meanings more accurately than text-only sentiment analysis systems. In addition, the results of [29] revealed that the emotional reactions of the audience to short videos on social media are highly interconnected with the emotions of the audience, but the model accuracy is undermined by the heterogeneous resolution of the videos, noisy audio, and irregular text. Their findings indicated that adaptive preprocessing and multimodal fusion are significant ways of enhancing weighted accuracy and F1 scores. The same can be inferred from [61], whose study on Urdu video reviews indicated that audio, text, and visual frame co-integration results in more consistent sentiment classification than unimodal pipelines, specifically in a low-resource language context. In line with this, the argument in [61] brought to light that the cross-modal

interactions are used to counter the weaknesses that are presented by a single modality, such as an unintelligent tone of voice or insufficient text. Notably, these findings are aligned with those of [38], who reported that multimodal aggregation of facial expressions, fatigue indicators, and textual feedback enables more reliable identification of learner disengagement than visual or textual analysis alone. Taken together, these papers imply that multimodal systems are more inclusive of affective states as they represent complementary emotional information that cannot be deduced based on one source of data.

The other common theme in the findings relates to the role of high-order fusion strategies and mechanisms of attention to resolve the imbalance of modality and noise. According to a number of studies, naive fusion methods tend to favor text and underuse visual and acoustic information, resulting in poor performance. According to [47], it has been proven that sentiment analysis models that rely excessively on textual cues do not take into account spatial and temporal information in other modalities. According to their findings, sentiment prediction with multi-scale spatial memory and text-guided interaction mechanisms can be substantially enhanced when using benchmark data. On the same note, [44] underscored that negative information, including modality-level inconsistency and irrelevant features, is a significant contributor to performance drop in multimodal emotion analysis. Their contrastive-based feature-cleansing methodology is demonstrated to improve robustness and accuracy through the removal of the non-contributory signals during the fusion step.

In agreement with this view, [70] presented that fusion of attention based on CNN-retrieved speech features and BERT-based textual features enables models to progressively balance between modality effects. This leads to better emotion recognition across several datasets. Similarly, [37] discovered that the addition of sentiment and the intensity of emotion as auxiliary tasks within a multi-task multimodal structure enhances the detection of offensive content by more than 7%. As such, these discussions provide stress on the fact that the increase in performance in multimodal emotion analysis is due to the addition of modalities and to properly created fusion strategies that address noise, dominance of one modality over another, and alignment of context.

Another lesson the literature is helping learn is connected to the task-specific issue of sarcasm, offensiveness, and contextual ambiguity, which are especially challenging to mitigate with text-only models. For example, He [42] established that the detection of sarcasm improved significantly when emojis, patterns of punctuations, and other visual features, which are normally removed during preprocessing, were retained. According to their results, multimodal representations allow more credible polarity interpretations, even though the textual sentiment and the intent behind the sentence are in opposition. This observation is consistent with [37], who revealed that information of emotional intensity when represented by images and text is more effective in identifying harmful and offensive material, particularly when it concerns sensitive situations like women's harassment. Furthermore, [29] noted that emotional manifestations in short videos tend to intensify popular mood, which supports the necessity of multimodal systems with the ability to record nuanced affective indicators through media. In human-centered and educational tasks, [38] claimed that multimodal emotion detection can enable timely interventions,

as disengagement patterns are observed, though invisible to unimodal systems. In these studies, the conclusion is always that multimodal emotion and sentiment analysis will aid in the contextual accuracy, decrease the misclassification due to uncertain cues, and extend the generalization across areas. The literature is thus indicative of a unification of multimodal, attention-based, and task-based architectures as a guiding line in future sentiment and emotion analysis studies. The literature, therefore, reflects a convergence toward a multimodal, attention-driven, and task-aware architecture as a foundational direction for future sentiment and emotion analysis research.

Overall, the reviewed studies have provided important comparative insights into multimodal sentiment and emotion analysis approaches. However, early multimodal systems still relied on simple fusion techniques that combined textual, visual and acoustic features. At the same time, more recent studies have increasingly adopted attention-based and cross-modal interaction architectures to better capture the interdependencies between modalities. This suggests a clear methodological trend towards more sophisticated multimodal fusion strategies that can address modality imbalance, contextual ambiguity, and noisy real-world data. Despite these advances, several challenges remain to be resolved. In particular, multimodal models still struggle with data heterogeneity, limited availability of large-scale multimodal datasets, and the difficulty of aligning emotional signals across modalities. These limitations suggest that future research should focus on developing more robust, stable multimodal representation learning techniques, using scalable datasets that support emotion detection in real-world environments. Table VI systematically summarizes the multimodal emotion and sentiment comparative matrix of multimodal emotion and sentiment analysis approaches.

TABLE VI. MULTIMODAL EMOTION AND SENTIMENT ANALYSIS APPROACHES

Authors	Primary Modality Combination	Fusion Strategy Type	Methodological Focus	Key Contribution / Empirical Findings
[29]	Text + Audio + Video	Adaptive preprocessing + Hybrid fusion	Multi-scale resolution adaptation, RoBERTa text modeling, automatic audio segmentation	Improved weighted accuracy (+4.17%) and F1 (+7.29%), demonstrated correlation between short video emotions and audience sentiment
[37]	Text + Image	Multi-task multimodal fusion	Emotion and sentiment intensity as auxiliary tasks in offensive detection	>7% performance improvement, introduced WCSE off dataset, improved harassment-related content detection
[38]	Text + Facial Video + Behavioral Cues	Deep fusion (CNN-LSTM + EmoNet + BERT)	Learner disengagement detection integrating fatigue, emotion, and sentiment	88.5% fatigue precision, 91.5% sentiment accuracy, multimodal aggregation improves early disengagement detection
[42]	Text + Emoji + Image	Feature-level fusion with structured preprocessing	Sarcasm detection preserving emojis, punctuation, and repeated characters	Stable 76% accuracy demonstrated the importance of scenario-specific preprocessing
[44]	Text + Audio + Image	Contrastive feature cleansing + multi-head attention	Removal of modality-level noise and inconsistent emotional signals	Outperformed SOTA on MOSI & MOSEI, enhanced robustness against negative information
[47]	Text + Audio + Visual	Text-guided cross-attention + Spatial-memory fusion	Multi-scale representation integration, reduction of text dominance	Significant gains over benchmark models, improved inter-modal alignment
[61]	Text + Audio + Frames	Early fusion + Model ensembling	Urdu low-resource multimodal sentiment framework (UMSA)	>80% accuracy, demonstrated cross-modal interaction benefits in a low-resource setting
[70]	Speech + Text	Attention-based late fusion	CNN (speech spectrogram) + BERT (text) integration	88.4% accuracy (MOSEI), attention improves modality weighting & emotion recognition

#### D. Transformer, Large Language Model (LLM), and Advanced Architectures

The recent discoveries in the transformer and LLM literature suggest a marked change from task-specific and heavily supervised pipelines to more flexible architectures, able to adapt to multilingual, low-resource, and multifaceted affective-signals environments. Many studies highlight that traditional sentiment and emotion models fail when there is a lack of laboratory data or more linguistic diversity. Most recent results reported by [27] suggest that zero-shot learning with transformer models facilitates successful multilingual sentiment and emotion classification without task-conditioned training and reveal dynamic trends in perception driven by geopolitical events in Twitter data. These issues are also mentioned in [62], the results of which indicate that fine-tuned cross-lingual transformers, especially Cross-lingual Language Model-RoBERTa (XLM-R), are much better objects in comparison to traditional classifiers and multilingual baselines when faced with language of code-switching and informal social media. These findings suggest that cross-lingual embeddings and transfer learning methods have very significant enhancements in linguistic strength. In line with this view, [76] proved that even in politically sensitive and pragmatically rich fields, fine-tuned transformer models achieve predictable sentiment and emotion classification, with pragmatic nuances remaining a source of error. Combined, the results indicate that transformer-based designs offer scalable multilingual and domain-specific affect and affect analysis, particularly when zero-shot or cross-lingual learning processes are in place.

The other outstanding stream is concerned with integration of emotion and sentiment analysis, as well as auxiliary or related tasks, by use of multi-task learning and sophisticated architectural development. The results of [39] demonstrated that sentiment and emotion detection can be included as auxiliary tasks within a multi-task transformer framework, significantly improving the performance of propaganda detection in Arabic news. According to the recorded Macro-F1 advances, affective cues are complementary to detecting manipulative materials. Similar results are noted by [74], in which on-code-mixed Hinglish data, the detection of multi-tasking sentiment and emotion recognition is boosted with no use of ensemble methods. [45] continued this discussion by indicating that LLMs used in zero-shot, few-shot, and fine-tuning scenarios can be useful to detect cyberbullying, harmfulness, sarcasm, sentiment, and emotion in captions of memes, and that accuracy always improves with the task. All this indicates that shared representations of affective and semantic information in transformer-based multi-task architectures are advantageous for generalization and can also reduce the dependency on large task-specific datasets.

In addition to the accuracy of classification, recent work also indicates the methodological innovations in the field of LLM-based emotion analysis connected to representation learning, prompt engineering, and context generation. [48] asserted that ensemble strategies together with attention mechanisms enhance the multilabel emotion classification of ambiguous social media

texts significantly, resolve the problems of class imbalance, and sparsity in short texts. On a similar note, [56] demonstrated that contextual sentences generated by LLC can optimize the emotion lexicon vectorization, especially in the case of Turkish. This suggests that generative models can enhance the availability of affective resources in low-resourced languages. Meanwhile, [58] presented additional evidence that the performance of LLMs is extremely sensitive to timely design, and that no single prompt provides the best outcome in the context of sentiment, emotion, and other similar tasks in the Arabic NLP. Their results highlight the significance of systematic prompt-evaluation systems. [28] also discussed culturally and theologically complicated literature and highlighted that LLMs can grasp emotional semantics in Quran translations to a reasonable level. However, it is not always compatible with human interpretation. Lastly, [49] demonstrated that conversational LLMs such as ChatGPT are capable of modeling emotional intelligence across all dialogue datasets, and that advanced architectures continue to play an increasing role in empathetic human-computer interaction. The overall results of these studies suggest that the advancements in the area of sentiment and emotion analysis in future will rely on the size of the model, task-sensitive adaptation, prompt maximization, and context-sensitive learning of representations.

Overall, a comparative study of the examined research demonstrates some key methodological improvements in transformer-based and LLM-driven mood and emotion analysis. Earlier methodologies predominantly depended on supervised transformer models trained on domain-specific datasets. However, contemporary research increasingly investigates zero-shot, few-shot, and cross-lingual learning methods to mitigate the constraints imposed by the scarcity of labeled data. This transition signifies a broader research trend towards scalable architectures capable of managing multilingual and low-resource language contexts. Nonetheless, despite the encouraging efficacy of transformer-based models and LLMs, numerous problems persist. Particularly, the identification of subtle or culturally specific emotional expressions, the responsiveness of models to prompt formulation, and the constrained interpretability of expansive architectures persist as notable research obstacles. In addition, recent studies highlight several methodological risks associated with LLM-based emotion analysis, including sensitivity to prompt design, hallucinated outputs, and potential cultural or dataset bias. These issues may affect the reliability of emotion interpretation, particularly in socially sensitive contexts where misclassification of emotions could lead to misleading conclusions. These insights underscore the need for forthcoming investigations to focus on well-defined transformer architectures, resilient prompt engineering methodologies, and integrative approaches that combine linguistic expertise with comprehensive contextual representations to enhance the reliability and clarity of emotion-sensitive NLP systems. Table VII outlines a systematic synthesis of these transformer-based and LLM-driven methods, summarizing their architectural designs, learning approaches and contributions to empirical data in multilingual and multi-task contexts.

TABLE VII. TRANSFORMER, LLM AND ADVANCED ARCHITECTURES

Authors	Core Architecture / Model	Learning Strategy	Primary Research Focus	Key Empirical Contribution
[27]	Transformer (BERT/GPT-Based)	Zero-Shot Multilingual Learning	Multilingual Perception Analysis (Twitter)	Demonstrated Scalable Zero-Shot Sentiment and Emotion Classification Without Task-Specific Training, Revealed Dynamic Geopolitical Perception Trends
[62]	XLNet (Fine-Tuned)	Cross-Lingual Transfer Learning	Multilingual and Code-Mixed Sentiment/Emotion	Achieved F1 = 90.3%, Handling Code-Switching Improved Accuracy By 8.9%
[73]	Arabic BERT (Further Pre-Trained)	Within-Task and Cross-Task Adaptation	Arabic Sentiment and Emotion	Further Pre-Training Improved Performance Up To 4.71%, Addressed Low-Resource Adaptation Gap
[76]	Fine-Tuned Hungarian BERT	Supervised Fine-Tuning	Political Sentiment and Emotion	Demonstrated an Effective Fine-Tuned Transformer for Domain-Specific Political Texts, Identified Pragmatic Nuance Limitations
[39]	Transformer-Based MTL	Multi-Task Learning	Arabic Propaganda Detection	Macro-F1 = 0.778, Integrating Sentiment & Emotion as Auxiliary Tasks Improved Propaganda Detection
[74]	XLNet (MTL)	Multi-task Transfer Learning	Code-Mixed Hinglish Sentiment and Emotion	MTL Outperformed Single-Task Baselines Without Ensemble Techniques
[45]	LLMs (Zero-Shot, Few-Shot, Fine-Tuning)	Multi-Task LLM Adaptation	Harmful Meme and Cyberbullying Detection	Improved Cyberbullying Accuracy by 7.94%, Demonstrated Cross-Task LLM Capability
[48]	EmoBERTA-X (Transformer + Attention + Ensemble)	Multilabel Classification with Attention	Fine-Grained Emotion Detection	Accuracy Improved By 4.32%, Addressed Class Imbalance & Short-Text Ambiguity
[56]	BERT + LLM-Generated Context	Representation Enhancement	Emotion Lexicon Vectorization (Turkish & English)	LLM-Generated Context Significantly Improved Vector Quality, Especially in Low-Resource Turkish
[58]	Multiple LLMs (GPT-4o, Llama, Claude, ALLAM)	Prompt Evaluation Framework	Arabic NLP Prompt Optimization	Proposed Multi-Dimensional Prompt Scoring; Showed No Universal Optimal Prompt
[28]	LLM-Based Emotion Analysis	Human-Validated LLM Evaluation	Emotion Preservation in Quran Translations	LLMs Captured Emotional Semantics at Fair Level (K = 0.29-0.35), Revealed Cultural-Pragmatic Limits
[49]	ChatGPT	Conversational Emotion Modeling	Emotional Intelligence in Dialogue	Demonstrated Conversational LLM Capability in Modeling Empathetic Emotional Responses

### E. Domain Specific and Applied Emotion, Sentiment Analysis

In recent domain-specific research, it is evident that emotion and sentiment analysis is no longer generic opinion mining. For example, in healthcare monitoring, sentiment and emotion analysis of patient comments on online forums or hospital feedback platforms can help detect early signs of stress, dissatisfaction, or anxiety. Rather, it is context-sensitive systems specific to particular social, commercial, educational, legal and health-related contexts. In applications to the consumer, results have always indicated that use of emotional signals results in quantifiable results over behavior-only or content-only methods. According to [31], incorporating multimodal emotional information into consumer targeting systems significantly enhances individualization and interaction, where emotion-sensitive architectures are more accurate and adaptable than marketing analytics. Related performance incentives are observed in [63], where joint semantic and emotional optimization can be used to classify sentiment in fresh agricultural product reviews, which is helpful when the consumer feedback is sparse and ambiguous. [40] also highlighted that sentiment polarity is inverted in user-generated reviews with sarcasm. The results revealed that hybrid deep learning architectures with BERT and convolutional models are much more effective at detecting sarcasm. Together, these works highlight that applied sentiment and emotion analysis in a commercial and consumer context is most useful when models explicitly take into consideration domain-specific language phenomena. This includes sarcasm, emphasis on emotion, and

situation-specific interpretation, as opposed to a situation-independent sentiment polarity.

The results of studies conducted in the areas of education and mental health demonstrate a consistent pattern: emotion-aware analytics can be used to obtain the insights that are not represented by traditional assessment tools. There are examples on online learning platforms where analyzing discussion posts or student feedback can reveal emotional states such as frustration, confusion, or low motivation, allowing instructors to intervene more effectively. [35] indicated that models that integrate textual and numerical aspects during counseling conversations enhance the quality of emotion classification prediction, which indicates the possibility of AI-assisted mental health tools that can identify stress, depression, and positive affect in real-time. In professional development contexts, [50] demonstrated that automatically calculated sentiment and emotion ratings of teacher assignments are strongly related to performance outcomes across demographic categories, indicating patterns unnoticed using post-program surveys. Equivalent inferences are made by [65], whose findings suggested that fusion of several attention processes into a neural framework improves students' emotional processing by identifying latent semantic content in learning documents. Moreover, multi-task term-weighting strategies that combine sentiment and emotion information are revealed to be better at predicting user satisfaction in e-learning systems than more traditional embedding fusion strategies [55]. Essentially, these results indicate that emotion and sentiment analysis is an

evaluative instrument in the educational and psychological fields and a complementary indicator that helps a person understand learning behaviors, engagement, and well-being beyond superficial measures.

Another application research demonstrates the applicability of emotion sentiment analysis in the context of societal, legal and popular discourse, where interpretability and contextual responsiveness are crucial. [41] demonstrated that the analysis of sentiment and emotion in discussions on Reddit about cultural appropriation presents a complex combination of trust, fear, anger, and sadness. This suggests that the discourse of the population cannot be satisfactorily described using a dominant sentiment. Recent studies in health communication [60] identified significant disparities between the affective states expressed by health institutions and those of the population on Instagram. This is specifically true in the context of fear and anger in discussions about vaccines, which highlights the need to synchronize emotional framing with the response of the audience. Explainability represents one of the core issues in the field of social media analytics, as [52] observed that when explainable AI methods are included with deep learning models, transparency and trust in emotional prediction findings are enhanced. Both [54] and [75] offered evidence in the field of forensic and legal science that emotional and sentiment cues in textual evidence can substantially improve the detection of deception and predict the accuracy of legal judgment in particular sentencing tasks. Notably, emotional scores alone do not have as strong an impact as sentiment signals. In these

practical contexts, the results are consistent in the point that domain-specific emotion and sentiment analysis needs to strike a balance between predictive performance, interpretability and contextual awareness. This ensures it should be used as a decision-support process rather than a technical classification task.

Overall, the reviewed studies demonstrate that domain-specific sentiment and emotion analysis requires methodological adaptations that reflect contextual characteristics of each application domain. While commercial and consumer-oriented applications emphasize predictive performance and personalization, educational as well as mental health applications prioritize early intervention and behavioral understanding. By contrast, legal and societal discourse analysis places stronger emphasis on interpretability and ethical considerations. These observations indicate a growing trend toward context-aware emotion analysis systems tailored to specific decision-making environments. Still, challenges remain in balancing predictive accuracy with transparency and ethical responsibility, particularly in sensitive domains such as healthcare, law, and public discourse. In response, future research should focus on explainable and ethically grounded emotion-aware systems capable of supporting responsible decision-making in real-world applications. Table VIII organized synthesis of these specific applications of the domain, methodological adaptations and empirical contributions is outlined in Table VIII to shed light on the operationalization of emotion-sentiment analysis across various applied settings.

TABLE VIII. DOMAIN SPECIFIC AND APPLIED EMOTION, SENTIMENT ANALYSIS

Authors	Application Domain	Methodological Approach	Emotion, Sentiment Integration Strategy	Key Empirical Contribution
[31]	Consumer targeting/Digital marketing	CNN + BiLSTM + attention (multimodal)	Facial, vocal, and textual emotion fusion	21.6% improvement over behavior-only models; 92.4% emotion recognition rate
[63]	Agricultural product reviews	Electra + XLNet + BiGRU + multi-head attention	Semantic-emotional optimization with keyword-based emotion dictionary	Superior classification over baseline models via collaborative semantic-emotion modeling
[40]	Sarcasm in user reviews	CNN + BERT hybrid framework	Polarity inversion handling via sarcasm modeling	Achieved F-score of 95%; improved sarcasm-aware sentiment detection
[35]	Mental health counseling	Hybrid textual + numerical ML models	Structural + lexical emotion features integration	Random Forest achieved 90.05% accuracy, supporting AI-assisted counseling tools
[50]	Teacher professional development	Machine-assisted sentiment/emotion scoring	Performance correlation modeling	Significant association between affective scores and teacher performance
[65]	Student emotion analysis	Multi-attention neural model	Global-local attention for latent emotion detection	F1 improvement up to 6.91%; enhanced hidden semantic extraction
[55]	E-Learning satisfaction	SEMAR (multi-task term weighting + CNN + BiLSTM)	Joint sentiment-emotion multi-task weighting	Avg. F1 = 91.82%; outperformed embedding fusion methods
[41]	Cultural appropriation discourse (Reddit)	Topic modeling + emotion analysis	Temporal sentiment and emoji-based emotion analysis	Identified complex emotion mix (trust, fear, anger, sadness) in public discourse
[60]	Health communication (Instagram)	BERTopic + sentiment-emotion analysis	Topic-emotion alignment analysis	Identified fear gaps between health organizations and public responses
[52]	Social media emotion decoding	BiLSTM + XAI (LIME, SHAP)	Explainable sentiment-emotion modeling	>90% accuracy; enhanced transparency via XAI
[54]	Forensic deception analysis	Psycholinguistic NLP + LDA + emotion analysis	Deception + emotion temporal correlation	Improved suspect identification via emotional and deception cues
[75]	Legal judgment prediction	BERT + ML models + NRC emotion lexicon	Sentiment and emotion score integration in sentencing prediction	Sentiment signals had a stronger predictive impact than emotion alone

## V. CONCLUSION

The purpose of the SLR was to synthesize the current body of knowledge on sentiment and emotion analysis in textual data in a comprehensive, up-to-date manner, to bring together methodological developments, trends in applications and current challenges in the area. The review systematically analyzed peer-reviewed journal articles concerning the topic published within the range of 2023-2025 and was only retrieved using the WoS and Scopus databases, guided by the PRISMA protocol. Accordingly, a total of 50 primary studies were identified through an intensive identification, screening, eligibility and quality appraisal process to be used in the qualitative synthesis. The review aimed to respond to an important set of RQs with regard to the computational approaches used, multimodal data integration, the use of transformer-based and LLMs, and domain-specific sentiment and emotion analysis. In organizing the analysis, this study focuses on providing a structured synthesis of recent developments in sentiment and emotion analysis, highlighting methodological trends, new architectures, and domain-specific applications. This integrative perspective helps bridge the literature's findings and offers a clearer understanding of the evolving research landscape in textual affective computing.

Furthermore, the synthesis of results demonstrates a number of common and strong patterns across the reviewed articles. To begin with, conventional sentiment analysis methods that rely exclusively on polarity alone are becoming known to be too simplistic to represent subtle and implicit forms of emotion in textual data. Due to this, there has been an informal shift in the methodology to deep learning and transformer-based models, which provide a more detailed contextual representation and better performance in the detection of fine-grained emotions. Second, the review determines an increased interest in multimodal emotion and sentiment analysis, in which text-based information is complemented by visual and audio information to increase the interpretative accuracy, especially in socially complex and noisy contexts such as social media and e-learning platforms. Third, transformer models and LLMs have become the leading architectures, which makes it possible to facilitate multilingual processing, low-resource language analysis, and multi-task learning frameworks that simultaneously model sentiment, emotion, and similar phenomena. Lastly, the analysis identifies an increase in domain-specific applications, such as education, healthcare, mental health, legal analysis, and public discourse, where emotion-aware systems are now being appreciated as being able to assist in decision-making, monitoring, and intervention. All these items indicate that the field has matured in the sense of being context-aware, ethically sensitive and application-driven.

The strengths of this review are that it integrates high-quality, more recent research and is organized around thematic topics, presenting four consistent streams of research, within which future research should be framed. Nonetheless, the review is beneficial due to a combination of methodological trends and application areas that will give practical ideas to system designers, educators, policymakers, and practitioners willing to implement sentiment and emotion analysis tools. Meanwhile, several limitations should be acknowledged when interpreting

the findings of this review. First, the literature search was limited to two major databases, namely WoS and Scopus. Although these databases provide high-quality, indexed publications, relevant studies available in other databases or digital libraries may not have been captured, potentially introducing database bias. Second, the review included only English-language journal articles, potentially introducing language bias by excluding relevant research published in other languages. Third, the study focused on publications from 2023 to 2025 to capture the most recent developments in transformer-based and LLM-driven approaches. While this time frame ensures the relevance of the reviewed technologies, it may omit earlier influential studies that contributed to the development of sentiment and emotion analysis methods. Thus, it is recommended to further conduct future studies on cross-linguistic and cross-cultural aspects, develop explainable, ethically based models, and examine longitudinal and real-time emotion processing in dynamic textual conditions. Altogether, this review can contribute to the necessity of systematic evidence synthesis to contribute to the development of theoretical knowledge and methodological integrity in sentiment and emotion analysis and to serve as a solid basis for future empirical research and conscientious use in more emotion-sensitive digital ecosystems.

## ACKNOWLEDGMENT

The authors acknowledge the use of AI-assisted tools in supporting the preparation of this manuscript. Scopus AI supported the development of preliminary conceptual maps, ChatGPT (version 5.2) assisted in constructing search strings, matrix tables, and organizing data analysis, and QuillBot was used for precise grammatical refinement and language editing. All outputs were critically reviewed and revised to ensure accuracy, rigor, and originality, and the authors take full responsibility for the final content. The author expresses gratitude for the financial support received from the Minister of Higher Education Malaysia under Universiti Kebangsaan Malaysia Research Grant with a Project Code: WARISAN-2025-024, which enabled this research to be completed.

## REFERENCES

- [1] B. H. Kasthuriarachchy, K. De Zoysa, and H. L. Premaratne, "Context-aware sentiment classification," in 15th International Conference on Advances in ICT for Emerging Regions, ICTer 2015 - Conference Proceedings, 2016, p. 276. doi: 10.1109/ICTER.2015.7377709.
- [2] H. Zhu, "Sentiment Analysis of 2021 Canadian Election Tweets," in Proceedings of SPIE - The International Society for Optical Engineering, 2023. doi: 10.1117/12.2667211.
- [3] D. Shanmugapriyaa, M. S. Aravind, S. B. Prakath, S. Deivarani, and K. Rajarajeshwari, "A Comparative Study of Sentiment Analysis on Flipkart Dataset using Naïve Bayes Classifier Algorithm," in 14th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2023, 2023, pp. 1110–1118. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85174312884&partnerID=40&md5=28d23b3e26ba78392315d50a6bf73c73>.
- [4] Y. N. Prajapati, S. Yadav, S. Sharma, U. K. Patel, and S. Tomar, "Analysis of Underlying Emotions in Textual Data Using Sentiment Analysis Which Classifies Text In to Positive, Negative or Neutral Sentiments," in 2023 14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023, 2023. doi: 10.1109/ICCCNT56998.2023.10307310.
- [5] H. A. Sayyed, S. Rushikesh Sugave, S. Paygude, and N. Jazdale, "Study and Analysis of Emotion Classification on Textual Data," in Proceedings

- of the 6th International Conference on Communication and Electronics Systems, ICCES 2021, 2021, pp. 1128–1132. doi: 10.1109/ICCES51350.2021.9489204.
- [6] P. Apte and S. Sri Khetwat, “Text-based emotion analysis: Feature selection techniques and approaches,” in *Advances in Intelligent Systems and Computing*, P. Dutta, A. Abraham, A. Bhattacharya, S. Dutta, and J. K. Mandal, Eds., Springer Verlag, 2019, pp. 837–847. doi: 10.1007/978-981-13-1498-8\_74.
- [7] P. Chinmayi, G. B. Regulwar, N. Anoop, K. G. Reddy, and P. Sunitha, “Sentimental Analysis of Product Reviews with Combined CNN-LSTM Model using Deep Learning,” in *2024 1st International Conference on Cognitive, Green and Ubiquitous Computing, IC-CGU 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/IC-CGU58078.2024.10530690.
- [8] N. V. Babu and E. Grace Mary Kanaga, “Multiclass Text Emotion Recognition in Social Media Data,” in *Lecture Notes in Electrical Engineering*, D. S. Sisodia, L. Garg, R. B. Pachori, and M. Tanveer, Eds., Springer Science and Business Media Deutschland GmbH, 2023, pp. 371–387. doi: 10.1007/978-981-99-0085-5\_30.
- [9] G. I. Ahmad and J. Singla, “Machine learning techniques for sentiment analysis of indian languages,” *International Journal of Recent Technology and Engineering*, vol. 8, no. 2 Special Issue 11, pp. 3630–3636, 2019, doi: 10.35940/ijrte.B1456.0982S1119.
- [10] J. K. Periasamy, M. B. V R, and D. Jothishree, “A Comparative Study of ConvoBiGRU and ConvoBiLSTM on Amazon Dataset,” in *2022 International Conference on Communication, Computing and Internet of Things, IC3IoT 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/IC3IoT53935.2022.9767728.
- [11] S. Deori, T. Chutia, H. J. Saikia, A. Boro, and N. Baruah, “Bridging techniques and applications in sentiment analysis: Approaches, challenges, and emerging trends,” *Eng. Appl. Artif. Intell.*, vol. 164, 2026, doi: 10.1016/j.engappai.2025.113255.
- [12] F. M. Plaza-del-Arco, “Emotion analysis in Spanish for the medical domain; Analisis de emociones en español para el dominio médico,” in *CEUR Workshop Proceedings*, 2017. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85033711231&partnerID=40&md5=2001fda4aa798de4a18de0ae651a1c24>.
- [13] S. Kumar, R. Prabha, and S. Samuel, “Sentiment Analysis and Emotion Detection with Healthcare Perspective,” in *Studies in Computational Intelligence*, vol. 1024, Springer Science and Business Media Deutschland GmbH, 2022, pp. 189–204. doi: 10.1007/978-981-19-1076-0\_11.
- [14] B. Calabrese and M. Cannataro, “Sentiment analysis and affective computing: Methods and applications,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, K. Amunts, T. Lippert, L. Grandinetti, and N. Petkov, Eds., Springer Verlag, 2016, pp. 169–178. doi: 10.1007/978-3-319-50862-7\_13.
- [15] D. N. Dwivedi and G. Mahanty, “Mental health in messages: Unravelling emotional patterns through advanced text analysis,” in *Using Machine Learning to Detect Emotions and Predict Human Psychology*, IGI Global, 2024, pp. 187–208. doi: 10.4018/9798369319109.ch009.
- [16] A. Baqach and A. Battou, “Towards a user-oriented adaptive system based on sentiment analysis from text,” in *E3S Web of Conferences*, EDP Sciences, 2021. doi: 10.1051/e3sconf/202129701010.
- [17] B. A. Reddy, U. A. Nayak, S. Srinivasan, E. Angela Jean Mary, A. Chauhan, and R. Rajagopal, “Sentiment and emotion analysis using machine learning techniques,” in *Exploring Psychology, Social Innovation and Advanced Applications of Machine Learning*, IGI Global, 2025, pp. 173–190. doi: 10.4018/979-8-3693-6910-4.ch009.
- [18] G. Phonsa, S. P. Tiwari, N. Malik, and V. Kumar, “Words to Emotions: A Deep Learning Perspectives on Text Sentiments,” in *Lecture Notes in Networks and Systems*, 2025, pp. 45–57. doi: 10.1007/978-981-96-3352-4\_4.
- [19] M. Xi, Q. Wu, J. He, J. Wen, S. Xiao, and J. Yang, “Comparative Study on Linguistic Representation Models for Enhancing Multimedia Security,” in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, BMSB, IEEE Computer Society, 2025. doi: 10.1109/BMSB65076.2025.11165579.
- [20] D. Pandya, A. Jadeja, M. A. Khan, S. B. Trivedi, M. A. Ramnath, and B. P. Satish, “Significance of Sentiment Analysis with Text-Based Mining Approach,” in *Lecture Notes in Networks and Systems*, V. S. Rathore, V. Piuri, R. Babo, and V. Tiwari, Eds., Springer Science and Business Media Deutschland GmbH, 2024, pp. 315–323. doi: 10.1007/978-981-97-3991-2\_27.
- [21] J. R. Jim, M. A. R. Talukder, P. Malakar, M. M. Kabir, K. Nur, and M. F. Mridha, “Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review,” *Natural Language Processing Journal*, vol. 6, 2024, doi: 10.1016/j.nlp.2024.100059.
- [22] B. Kitchenham, “Guidelines for performing systematic literature reviews in software engineering,” Technical report, Ver. 2.3 EBSE Technical Report. EBSE, 2007.
- [23] C. Lockwood, Z. Munn, and K. Porritt, “Qualitative research synthesis: Methodological guidance for systematic reviewers utilizing meta-aggregation,” *Int. J. Evid. Based. Healthc.*, vol. 13, no. 3, pp. 179–187, 2015, doi: 10.1097/XEB.000000000000062.
- [24] M. J. Page et al., “The prisma 2020 statement: An updated guideline for reporting systematic reviews,” *Medicina Fluminensis*, vol. 57, no. 4, pp. 444–465, 2021, doi: 10.21860/medflum2021\_264903.
- [25] A. Abouzahra, A. Sabraoui, and K. Afdel, “Model composition in Model Driven Engineering: A systematic literature review,” *Inf. Softw. Technol.*, vol. 125, no. May, p. 106316, 2020, doi: 10.1016/j.infsof.2020.106316.
- [26] A. D. Moher D, Liberati A, Tetzlaff J, “PRISMA 2009 Flow Diagram,” 2009.
- [27] S. Shaikh, S. Y. Yildirim, M. Abomhara, and E. Zoto, “Multilingual user perceptions analysis from twitter using zero shot learning for border control technologies,” *Soc. Netw. Anal. Min.*, vol. 15, no. 1, 2025, doi: 10.1007/s13278-025-01434-x.
- [28] A. Almarzoqi and M. Alsuhaibani, “Leveraging large language models for detecting and preserving emotions in quran translations,” *Journal of King Saud University - Computer and Information Sciences*, vol. 37, no. 9, 2025, doi: 10.1007/s44443-025-00269-y.
- [29] Q. Wei, Y. Zhou, S. Xiang, L. Xiao, and Y. Zhang, “MEAS: Multimodal Emotion Analysis System for Short Videos on Social Media Platforms,” *IEEE Trans. Comput. Soc. Syst.*, vol. 12, no. 5, pp. 2398–2410, 2025, doi: 10.1109/TCSS.2024.3490846.
- [30] S. S. J. Soloman and N. Baydetti, “Negation Assisted Emotion Detection from Textual Data Using Machine Learning,” *SN Comput. Sci.*, vol. 6, no. 7, 2025, doi: 10.1007/s42979-025-04388-1.
- [31] T. S. A. Aziz, V. Ramalingam, K. S. Nandini Prasad, D. N. P. David Neels Pon Kumar, and A. Arun Kumar, “Integrating Emotion Aware AI for Hyper Personalized Consumer Targeting in Next Generation Man Machine Computing Environments,” *Journal of Machine and Computing*, vol. 5, no. 4, pp. 2019–2037, 2025, doi: 10.53759/7669/jmc202505158.
- [32] L. Ren and S. A. A. Adi Kasuma, “Emotion Vocabulary Recognition in English Writing Based on the B-BLLC-CL Algorithm,” *KSII Transactions on Internet and Information Systems*, vol. 19, no. 9, pp. 2815–2835, 2025, doi: 10.3837/tiis.2025.09.001.
- [33] K.-C. Hung, A. Amjad, Y.-P. Chao, and H.-T. Chang, “Emotion recognition using deep learning models in Chinese speech,” *Entertain. Comput.*, vol. 55, 2025, doi: 10.1016/j.entcom.2025.101039.
- [34] A. Rastogi, R. Singh, M. Z. Khan, I. Aljubayri, and F. M. Ghabban, “Patient Emotion and Sentiment Analysis Using Deep Learning,” *Journal of Innovative Image Processing*, vol. 7, no. 3, pp. 582–601, 2025, doi: 10.36548/jiip.2025.3.001.
- [35] I. Ramayanti et al., “Sentiment and Emotion Classification Models Using Hybrid Textual and Numerical Features: A Case Study of Mental Health Counseling,” *Journal of Applied Data Sciences*, vol. 6, no. 3, pp. 1482–1494, 2025, doi: 10.47738/jads.v6i3.632.
- [36] J. Feng, “Improving sentiment analysis in literary texts through bidirectional encoder representations: A BERT-based approach,” *Journal of Computational Methods in Sciences and Engineering*, vol. 25, no. 5, pp. 4650–4664, 2025, doi: 10.1177/14727978251338004.
- [37] G. V. Singh, S. Ghosh, M. Firdaus, A. Ekbal, and P. Bhattacharyya, “Unmasking offensive content: a multimodal approach with emotional understanding,” *Multimed. Tools Appl.*, vol. 84, no. 28, pp. 33381–33404, 2025, doi: 10.1007/s11042-025-20603-w.

- [38] Q. El Maazouzi and A. Retbi, "Multimodal Detection of Emotional and Cognitive States in E-Learning Through Deep Fusion of Visual and Textual Data with NLP," *Computers*, vol. 14, no. 8, 2025, doi: 10.3390/computers14080314.
- [39] L. Al-Henaki, H. Al-Khalifa, and A. al-Salman, "Enhancing Propaganda Detection in Arabic News Context Through Multi-Task Learning," *Applied Sciences (Switzerland)*, vol. 15, no. 15, 2025, doi: 10.3390/app15158160.
- [40] A. Parkar and R. Bhalla, "A Hybrid Framework for Sarcasm Detection Using CB Technique," *International Journal of Basic and Applied Sciences*, vol. 14, no. 3, pp. 116–124, 2025, doi: 10.14419/z470zd97.
- [41] M. M. Mostafa, "Global cultural appropriation discourse on reddit: a topic modeling approach," *Neural Comput. Appl.*, vol. 37, no. 20, pp. 15345–15368, 2025, doi: 10.1007/s00521-025-11277-4.
- [42] E. D. Wahyuni, T. L. M. Suryanto, and H. Arviani, "Deep Learning Multimodal Sarcasm Detection in Social Media Comments: The Role of Memes and Emojis," *Journal of Artificial Intelligence and Technology*, vol. 5, pp. 192–201, 2025, doi: 10.37965/jait.2025.0699.
- [43] C. M. Liapis, A. Karanikola, and S. Kotsiantis, "Enhancing sentiment analysis with distributional emotion embeddings," *Neurocomputing*, vol. 634, 2025, doi: 10.1016/j.neucom.2025.129822.
- [44] R. Wang et al., "Contrastive-Based Removal of Negative Information in Multimodal Emotion Analysis," *Cognit. Comput.*, vol. 17, no. 3, 2025, doi: 10.1007/s12559-025-10463-9.
- [45] A. K. Indira Kumar, G. Sthanusubramoniani, D. Gupta, A. R. Nair, Y. A. Alotaibi, and M. Zakariah, "Multi-task detection of harmful content in code-mixed meme captions using large language models with zero-shot, few-shot, and fine-tuning approaches," *Egyptian Informatics Journal*, vol. 30, 2025, doi: 10.1016/j.eij.2025.100683.
- [46] A. Khodaei, A. Bastanfard, H. Saboohi, and H. Aligholizadeh, "A Transfer-Based Deep Learning Model for Persian Emotion Classification," *Multimed. Tools Appl.*, vol. 84, no. 17, pp. 17435–17463, 2025, doi: 10.1007/s11042-024-19668-w.
- [47] X. He et al., "Text-guided multi-level interaction and multi-scale spatial-memory fusion for multimodal sentiment analysis," *Neurocomputing*, vol. 626, 2025, doi: 10.1016/j.neucom.2025.129532.
- [48] F. H. Labib, M. Elagamy, and S. N. Saleh, "EmoBERTa-X: Advanced Emotion Classifier with Multi-Head Attention and DES for Multilabel Emotion Classification," *Big Data and Cognitive Computing*, vol. 9, no. 2, 2025, doi: 10.3390/bdccc9020048.
- [49] P. Pradeep et al., "Sentiment and Emotion Modeling in Text-based Conversations utilizing ChatGPT," *Engineering, Technology and Applied Science Research*, vol. 15, no. 1, pp. 20042–20048, 2025, doi: 10.48084/etasr.9508.
- [50] Y. Liu, L.-K. Soh, G. Trainin, G. Nugent, and W. M. Smith, "Investigating relationships of sentiments, emotions, and performance in professional development K-12 CS teachers," *Computer Science Education*, vol. 35, no. 1, pp. 28–59, 2025, doi: 10.1080/08993408.2023.2298162.
- [51] S. S. Gorshkov, D. I. Ignatov, A. Y. Chemysheva, V. L. Goiko, and V. V. Kashpur, "Identifying Top-Performing Students via VKontakte Social Media Communities Using Advanced NLP Techniques," *IEEE Access*, vol. 13, pp. 962–979, 2025, doi: 10.1109/ACCESS.2024.3521857.
- [52] D. Muhammad, I. Ahmed, K. Naveed, and M. Bendecheche, "Explainable AI Models for Decoding Emotional Subtexts on Social Media," *Complexity*, vol. 2025, no. 1, 2025, doi: 10.1155/cplx/9258956.
- [53] X. Wang, Y. Li, and F. Chen, "Twitter data emotion analysis using Hadoop and metaheuristic optimized Graphical Neural Network," *Front. Artif. Intell.*, vol. 8, 2025, doi: 10.3389/frai.2025.1672252.
- [54] J. Adkins, A. Al Bataineh, and A. Khanal, "A psycholinguistic NLP framework for forensic text analysis of deception and emotion," *Front. Artif. Intell.*, vol. 8, 2025, doi: 10.3389/frai.2025.1669542.
- [55] S. Sandiwamo, D. I. Indra Sensuse, H. Budi Santoso, D. Hidayat, and A. S. Nyamawe, "SEMAR: A Multi-Task Term Weighting Approach for Sentiment and Emotion-Based E-Learning Users' Satisfaction Analysis," *IEEE Access*, vol. 13, pp. 187584–187601, 2025, doi: 10.1109/ACCESS.2025.3627657.
- [56] S. Kumova Metin and H. Aka Uymaz, "Exploring the Effectiveness of LLM-Generated Context on Emotion Lexicon Word Vectorization: A Comparative Study on Turkish and English," *IT Prof.*, vol. 27, no. 5, pp. 52–58, 2025, doi: 10.1109/MITP.2025.3572550.
- [57] A. Pajon-Sanmartin, F. de Arriba-Pérez, S. García-Méndez, F. Leal, B. Malheiro, and J. Carlos Burguillo-Rial, "Unraveling Emotions With Pre-Trained Models," *IEEE Access*, vol. 13, pp. 182458–182473, 2025, doi: 10.1109/ACCESS.2025.3623877.
- [58] D. Refai, M. S. Al-Shaibani, and I. Ahmad, "Is This the Best Prompt? Scoring Prompts for Arabic NLP Across LLMs," *IEEE Access*, vol. 13, pp. 171468–171492, 2025, doi: 10.1109/ACCESS.2025.3616181.
- [59] A. D. P. Ariyanto, F. K. Fikriah, and A. F. Setyawan, "Impact of Statistical and Semantic Features Extraction for Emotion Detection on Indonesian Short Text Sentences," *CommIT Journal*, vol. 19, no. 1, pp. 1–13, 2025, doi: 10.21512/commit.v19i1.11680.
- [60] A. P. Vit and A. Magid, "Exploring Topics, Emotions, and Sentiments in Health Organization Posts and Public Responses on Instagram: Content Analysis," *JMIR Infodemiology*, vol. 5, 2025, doi: 10.2196/70576.
- [61] S. S. Malik et al., "Multi-Modal Emotion Detection and Sentiment Analysis," *IEEE Access*, vol. 13, pp. 59790–59810, 2025, doi: 10.1109/ACCESS.2025.3552475.
- [62] S. S. Almalki, "Sentiment Analysis and Emotion Detection Using Transformer Models in Multilingual Social Media Data," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 3, pp. 324–333, 2025, doi: 10.14569/IJACSA.2025.0160332.
- [63] Y. Dong, Y. Zhang, G. Wu, X. Chen, L. Tu, and G. Fan, "Sentiment classification method of fresh agricultural product reviews based on semantic and emotional optimisation," *International Journal of Intelligent Information and Database Systems*, vol. 17, no. 1, pp. 32–56, 2025, doi: 10.1504/IJIIIDS.2025.143485.
- [64] R. S. Sitoula, M. Pramanik, and R. Panigrahi, "Fine-Grained Classification for Emotion Detection Using Advanced Neural Models and GoEmotions Dataset," *Journal of Soft Computing and Data Mining*, vol. 5, no. 2, pp. 62–71, 2024, doi: 10.30880/jscdm.2024.05.02.005.
- [65] J. Xiao, "Student Emotion Analysis by Integrating Attention Mechanism Algorithm and Neural Network Algorithm," *IEIE Transactions on Smart Processing and Computing*, vol. 13, no. 6, pp. 642–653, 2024, doi: 10.5573/IEIESPC.2024.13.6.642.
- [66] D. Heaton, J. Clos, E. Nichele, and J. E. Fischer, "The ChatGPT bot is causing panic now – but it'll soon be as mundane a tool as Excel: analysing topics, sentiment and emotions relating to ChatGPT on Twitter," *Pers. Ubiquitous Comput.*, vol. 28, no. 6, pp. 875–894, 2024, doi: 10.1007/s00779-024-01811-x.
- [67] F. Naznin, I. Hazarika, D. Laskar, and A. K. Mahanta, "Mining association between different emotion classes present in users posts of social media," *Soc. Netw. Anal. Min.*, vol. 14, no. 1, 2024, doi: 10.1007/s13278-024-01241-w.
- [68] D. K. Aidam, B.-B. Benuwa, S. O. Oppong, and E. Nwiah, "Sentiment and emotion analysis using pretrained deep learning models," *Journal of Data, Information and Management*, vol. 6, no. 3, pp. 277–295, 2024, doi: 10.1007/s42488-024-00129-w.
- [69] M. Kastrati, Z. Kastrati, A. Imran, and M. Biba, "Leveraging distant supervision and deep learning for twitter sentiment and emotion classification," *J. Intell. Inf. Syst.*, vol. 62, no. 4, pp. 1045–1070, 2024, doi: 10.1007/s10844-024-00845-0.
- [70] F. Makhmudov, A. Kutlimuratov, and Y.-I. Cho, "Enhancing Multimodal Emotion Recognition through Attention Mechanisms in BERT and CNN Architectures," *Applied Sciences (Switzerland)*, vol. 14, no. 10, 2024, doi: 10.3390/app14104199.
- [71] H. Aka Uymaz and S. Kumova Metin, "Emotion-enriched word embeddings for Turkish," *Expert Syst. Appl.*, vol. 225, 2023, doi: 10.1016/j.eswa.2023.120011.
- [72] N. Punetha and G. Jain, "Game theory and MCDM-based unsupervised sentiment analysis of restaurant reviews," *Applied Intelligence*, vol. 53, no. 17, pp. 20152–20173, 2023, doi: 10.1007/s10489-023-04471-1.
- [73] W. Alshehri, N. Al-Twairish, and A. Alothaim, "Affect Analysis in Arabic Text: Further Pre-Training Language Models for Sentiment and Emotion," *Applied Sciences (Switzerland)*, vol. 13, no. 9, 2023, doi: 10.3390/app13095609.
- [74] S. Ghosh, A. Priyankar, A. Ekbal, and P. Bhattacharyya, "Multitasking of sentiment detection and emotion recognition in code-mixed Hinglish

- data.” *Knowl. Based. Syst.*, vol. 260, 2023, doi: 10.1016/j.knosys.2022.110182.
- [75] M. Farhadishad, M. Kazemifard, and Z. Rezaei, “Predicting Court Judgment in Criminal Cases by Text Mining Techniques,” *Journal of Information Technology Management*, vol. 15, no. 2, pp. 204–222, 2023, doi: 10.22059/jitm.2023.350464.3206.
- [76] I. Uveges and O. Ring, “HunEmBERT: A Fine-Tuned BERT-Model for Classifying Sentiment and Emotion in Political Communication,” *IEEE Access*, vol. 11, pp. 60267–60278, 2023, doi: 10.1109/ACCESS.2023.3285536.