



International Journal of
Applied Data Science in Engineering and Health
<https://ijadseh.com>



State-of-the-Art Machine Learning Techniques in Sentiment Analysis for Social Media

Mohsen Mohammadagha*, Israel Tshitenge, Ifetilayo Adebambo

Department of Civil Engineering, University of Texas at Arlington, Arlington, Texas 76019, USA

(Email: mxm4340@mavs.uta.edu)

Received date: February 22, 2026; Accepted date: August 9, 2026

Abstract

This review examines the evolution of machine learning and natural language processing (NLP) techniques for social media sentiment analysis, tracing the shift from traditional methods such as SVM and Naïve Bayes to modern deep learning, transformer-based, modern large language models (LLMs), and hybrid architectures. Recent advances-including contextual embeddings, recurrent and attention-based networks, and multimodal approaches that incorporate emojis-have substantially improved classification performance in noisy social media environments. State-of-the-art hybrid models fine-tuned through optimization algorithms now achieve accuracy and F1-scores approaching 97% on large Twitter datasets, while multimodal and multilingual models report F1-scores ranging from 0.57 to 0.83 on more challenging benchmarks. These findings underscore the importance of integrating multiple data types, applying robust preprocessing, and leveraging adaptive neural architectures to analyze global social media content effectively. By synthesizing these developments, this review highlights key trends, methodological challenges, and opportunities for advancing sentiment analysis research.

Keywords: Machine Learning, Sentiment Analysis, Social Media.

Introduction

The rapid expansion of social media platforms has transformed the way individuals and organizations communicate, share opinions, and interact globally. Social media generates vast amounts of user-generated content daily, encapsulating sentiments, opinions, and trends that shape public discourse. Sentiment analysis (SA), a branch of Natural Language Processing (NLP) [1], has emerged as a powerful tool to automatically extract and interpret these sentiments from textual data. Machine learning techniques play a pivotal role in advancing sentiment analysis by enabling more accurate classification and prediction [2] of affective states embedded within noisy and unstructured social media text. This research explores state-of-the-art machine learning methodologies applied to sentiment analysis, emphasizing their significance in understanding public opinion [3] dynamics across diverse domains such as business, politics, and social trends.

The application of machine learning to sentiment analysis [4] has evolved significantly, transitioning from traditional lexicon-based approaches to advanced deep learning architectures. Modern techniques incorporate contextual embeddings, neural networks, and hybrid models [5] that enhance the granularity of sentiment classification. These advancements address challenges

such as noisy text normalization [6], multilingual processing [7], and real-time trend prediction [8]. For example, ontology-driven methods [9] have gained traction for their ability to extract domain-specific information and improve sentiment analysis accuracy by leveraging [10] labeled datasets tailored to specific contexts. This research reviews these cutting-edge methodologies while highlighting their implications for social media monitoring.

Despite the progress in sentiment analysis research, challenges [11] persist in processing noisy text generated on social media platforms. Users often employ informal language, abbreviations, and inconsistent grammar, complicating the extraction of meaningful insights. Moreover, while English dominates sentiment analysis studies, other languages such as Malay [12], remain underexplored due to limited datasets and pre-processing techniques tailored for non-English texts. Addressing these issues is essential for broadening the applicability of sentiment analysis across diverse linguistic [13] and cultural contexts. This research examines existing solutions for noisy text normalization and investigates their effectiveness in multilingual environments [14].

By combining time-series analysis with social network theory [15], researchers can identify emergent topics and forecast shifts in public opinion with remarkable precision. These predictive capabilities are invaluable for businesses seeking [16] to understand consumer behavior, policymakers aiming to gauge public sentiment on policy issues, and researchers exploring societal trends. This research delves into comparing trend prediction methodologies that complement sentiment analysis frameworks [17], demonstrating their potential for these applications.

This research, with Table 1, Overview of State-of-the-Art, aims to provide a comprehensive overview of state-of-the-art machine-learning techniques for sentiment analysis in social media data. By synthesizing insights from recent studies and addressing gaps in existing research, it contributes to the academic discourse on NLP applications in social media monitoring. The findings underscore the importance of frameworks that combine advanced machine learning algorithms and hyperparameter optimization strategies with domain-specific ontologies and preprocessing strategies to achieve higher accuracy in sentiment classification and trend prediction. This research intends to provide insights, findings, or frameworks that can guide researchers in identifying gaps, opportunities, or areas for further investigation within the field. To achieve these aims, the following section provides a systematic synthesis of the evolution of machine learning techniques in social media sentiment analysis, highlighting key methodological advances and persistent gaps that this review seeks to address.

Literature Review

To frame the methodological developments discussed later, this section surveys the progression of sentiment analysis techniques across the past decade. The past decade has witnessed a dramatic evolution in sentiment analysis [18] for social media, fueled primarily by advances in machine learning (ML) and natural language processing (NLP). Traditional sentiment classification methods, which largely depended on lexicon-based sentiment scoring and shallow supervised models such as Support Vector Machines (SVM) and Naïve Bayes, provided meaningful early results, especially where data was relatively clean and structured [19]. However, these methods struggled with the informal, dynamic, and contextually rich nature of content generated on platforms such as Twitter and Facebook. Recent comprehensive reviews have noted a paradigm shift toward deep learning and transformer-based architectures that can better capture the intricacies of social-media text [20]. Transformer models like BERT and GPT leverage self-attention mechanisms to model long-range dependencies and understand semantic nuances, outperforming traditional models across a range of benchmarks. Albladi et al. (2025) detail how transformer-based approaches now dominate Twitter sentiment analysis, adapting efficiently to diverse topics, languages, and noisy environments [20]. Systematic reviews, such as Alsiaity & Orji (2022) [19], confirm that while classical models remain relevant for their simplicity and

speed, deep learning models are increasingly favored for their superior accuracy and ability to generalize across domains. Notably, most studies continue to prioritize English datasets and use accuracy as a benchmark, though there is growing recognition of the need for more robust evaluation metrics and expanded language coverage. Ongoing research focuses on addressing challenges tied to language diversity, domain adaptation, and preprocessing of noisy, user-generated content.

Beyond traditional text-only approaches, a particularly transformative trend has been the integration of multimodal cues, most notably emojis, which provide critical emotional context in short, informal social media posts. Emojis, as visual expressions of emotion and intent, offer a potent and often necessary complement to textual sentiment, especially in the fast-paced, shorthand environment of social media. Al-Azani and El-Alfy (2021) provide compelling evidence that combining emoji features with text-using early and late fusion models-significantly boosts sentiment polarity detection, particularly for morphologically complex and low-resource languages like Arabic. Their results indicate that fusion at the score level enhances overall classification performance, validating the surge of interest in multimodal sentiment frameworks [21]. To operationalize this integration, resources such as EmojiNet (Wijeratne et al., 2017) have emerged, building comprehensive mappings between emoji icons, their English senses, and context-dependent meanings. These resources help address challenges of emoji ambiguity and platform-specific rendering, enabling models to capture the full spectrum of sentiment embedded in social media posts. Collectively, such advances underline the importance of combining neural architectures, robust preprocessing, and multimodal features to achieve state-of-the-art performance-while also exposing enduring difficulties in language representation, normalization, and result generalization across platforms [22].

Building on these multimodal foundations, increasingly sophisticated neural architectures have been developed that combine recurrent, attention-based, and transformer components to capture even richer contextual information. Expanding on these developments, researchers have engineered increasingly complex yet powerful architectures that integrate advanced neural network components and feature fusion techniques. Li et al. (2021) introduced the emoji-text integrated bidirectional long short-term memory (ET-BiLSTM) framework for microblog sentiment classification, combining novel emoji vectorization with BiLSTM-based text representation and attention mechanisms. Their model demonstrated marked improvements over single-modality baselines, achieving higher macro-Precision, macro-Recall, and macro-F1 scores on microblog datasets replete with noisy, expressive content [23]. Similarly, Khalil et al. (2021) utilized multilabel BiLSTM models-leveraging pre-trained word embeddings such as AraVec and ARLSTM-showing these deep learning approaches significantly outperformed older techniques like SVM, Random Forest, and traditional neural networks, particularly in managing colloquial, informal, and code-mixed data. Such findings reinforce the consensus that cutting-edge sentiment analysis hinges on deep and recurrent architectures, with an emphasis on embeddings and attention mechanisms for nuanced emotion and opinion extraction. This trend not only advances accuracy and robustness for standard sentiment polarity tasks but also facilitates multilingual and domain-specific analysis, further expanding the utility of sentiment analysis in global social media applications [24].

In parallel with advancements in model architectures, the application of transfer learning and ensemble techniques has yielded promising results, particularly in complex classification and risk assessment tasks. Howard et al. (2020) evaluated a spectrum of models-including rule-based lexicon tools, DeepMoji, Universal Sentence Encoder, and fine-tuned GPT-1 transformers-on the CLPsych 2017 dataset to identify suicide risk from social media posts. Their best-performing approach, which integrated fine-tuned GPT-1 features, achieved a macro-averaged F1 score of 0.572, setting a new standard for risk prediction without extensive user metadata [25]. In public health monitoring, Bengesi et al. (2023) processed over half a million tweets regarding the monkeypox outbreak, employing 56 different models. Their findings identified a pipeline

comprising TextBlob annotation, lemmatization, CountVectorizer, and SVM as the top performer, achieving a notable accuracy of 0.9348. These results illustrate that, in addition to deep learning, ensembles and hybrid pipelines-combining neural models, traditional classifiers, and rigorous preprocessing-play a pivotal role in generating robust sentiment and risk predictions amid the noisy, heterogeneous, and often multilingual data typical of social media. The growing sophistication of these integrated methods underlines the need for both algorithmic innovation and attention to domain-specific challenges [26].

Finally, the most recent literature spotlights the emergence of hybrid models and optimization-driven approaches that further push the boundaries of sentiment analysis in social media environments. Vadakketthil Somanathan Pillai et al. (2024) presented a hybrid framework that amalgamates Bi-GRU and LSTM classifiers with the Hosted Cuckoo Optimization Algorithm, achieving an exceptional 99% accuracy for classifying three emotion classes in Twitter data. This optimization-centric strategy demonstrates that customized model tuning can unlock new levels of performance in handling the complexity and noise inherent in real-world user-generated text [27]. Similarly, Althobaiti (2022) benchmarked a BERT-based model against traditional machine learning methods for offensive language and hate speech detection in Arabic tweets, reporting F1-scores of 84.3% for offensive content and 81.6% for hate speech-substantially superior to SVM and logistic regression. The study also noted that incorporating sentiment and emoji features can further refine results, though the impact depends on dataset characteristics [28]. Together, these studies highlight a converging trend: advanced sentiment analysis for social media increasingly depends on the deployment of context-aware deep neural architectures [29], hybrid models, multimodal feature engineering, and algorithmic optimization [30]. Persistent challenges remain-such as handling noisy, code-mixed, and multilingual texts [31]; expanding the availability of labeled datasets in underrepresented languages [32]; and improving model transparency [33] and explainability. Nevertheless, the literature demonstrates that the fusion of state-of-the-art machine learning models, robust preprocessing, multimodal integration, and domain-aware adaptation represents the most promising pathway for research and practical deployment in sentiment analysis across social media contexts.

Methodology

The insights drawn from the reviewed studies inform the comparative framework adopted in this work. Building on these findings, the methodology outlined below integrates key analytical dimensions-model type, multimodality, and linguistic diversity-to systematically evaluate recent approaches. This study adopts a comprehensive and comparative methodology as shown in Figure 1 to investigate advanced machine learning techniques for sentiment analysis on social media data. Recognizing the complexity and diversity of user-generated content, the research is structured to systematically evaluate the effectiveness of various machine learning paradigms, spanning from traditional algorithms by using hybridization and optimization sorting techniques and analysis to contemporary deep learning and hybrid models. The rationale for this approach stems from the evolving landscape of sentiment analysis [34], where the shift from lexicon-based and shallow models to context-aware neural architectures [35] has been pivotal in addressing the unique challenges posed by social media platforms. By synthesizing and interpreting recent advancements, this methodology aims to bridge gaps in the literature, particularly regarding the integration of multimodal features and the adaptation of models to different linguistic and cultural contexts. The comparative nature of the research ensures a nuanced understanding of the strengths and limitations inherent in each methodological approach.

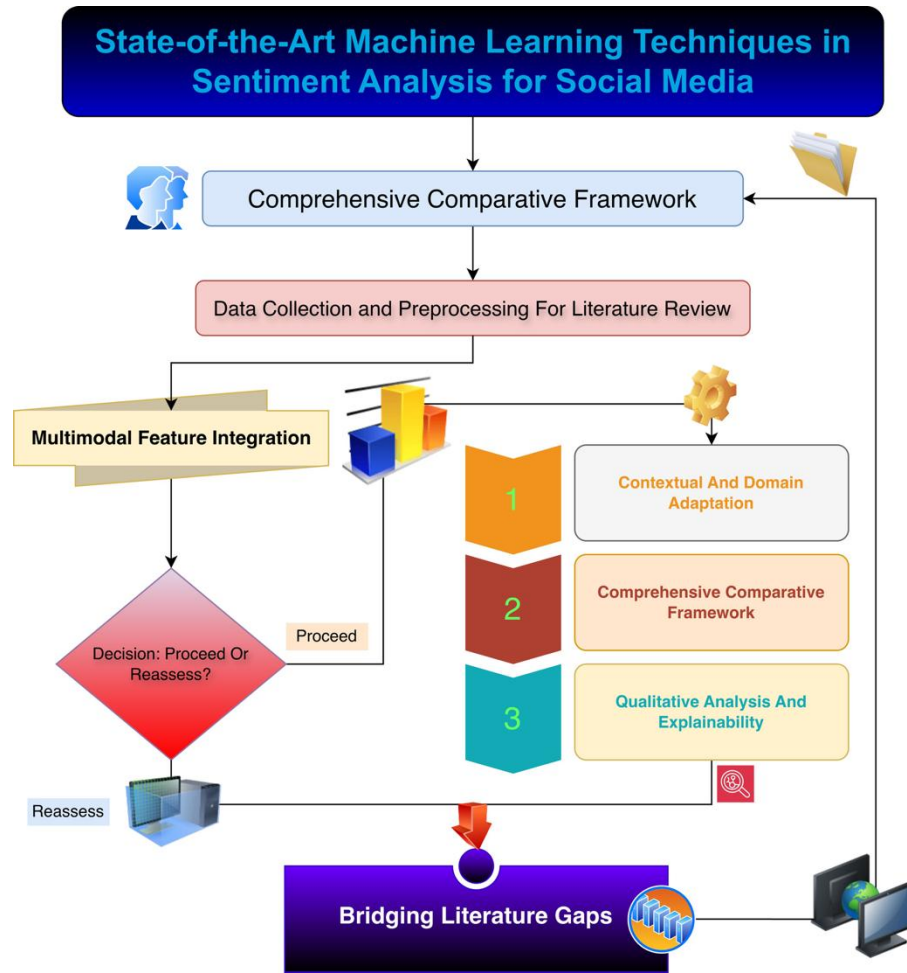


Figure 1. Flowchart of Methodology

For Data Collection and Preprocessing, the methodology emphasizes the importance of data collection and preprocessing as foundational steps in sentiment analysis [36]. Social media platforms, especially Twitter, are selected as primary data sources due to their accessibility and the prevalence of sentiment-rich content. Data is gathered using official APIs [37], ensuring authenticity and relevance to the research objectives. Preprocessing involves several stages: noise removal (such as eliminating links, usernames, and special characters), normalization of informal language, and the handling of multilingual content. Special attention is given to the treatment of emojis and other non-textual elements, as these play a significant role in conveying sentiment on social media. The preprocessing is designed to enhance data quality and ensure that subsequent machine learning models can effectively extract meaningful patterns [38] from the inherently noisy and heterogeneous social media text. This step is crucial for improving model robustness and generalizability across diverse datasets. With a clean and well-structured dataset, the next step involves selecting and evaluating machine learning models capable of capturing social-media-specific linguistic patterns.

Regarding Machine Learning and Deep Learning Approaches [39], this section discusses the machine learning and deep learning approaches used in the reviewed research, focusing on the deployment and evaluation of a range of machine learning techniques. Traditional models, such as Support Vector Machines and Naïve Bayes, are implemented as baselines due to their historical significance and computational efficiency. However, the primarily focuses on advanced deep

learning architectures, including recurrent neural networks (RNNs) [40], long short-term memory networks (LSTMs), bidirectional LSTMs (BiLSTMs) [41], and transformer-based models like BERT and GPT. These models are selected for their demonstrated ability to capture complex linguistic patterns and contextual dependencies within social media text. Additionally, hybrid models that combine neural architectures with optimization algorithms or ensemble strategies are explored to further enhance predictive performance. The comparative analysis of these approaches provides insights into their relative effectiveness in handling the nuances of sentiment expression in dynamic online environments [42].

Integration of Multimodal and Contextual Features can be a distinguishing aspect of this methodology is the incorporation of multimodal features, particularly the integration of emojis alongside textual data. Emojis are systematically mapped to their contextual meanings using resources such as EmojiNet, allowing models to interpret the full spectrum of sentiment embedded in social media posts. Early and late fusion techniques are employed to combine emoji and text features, enabling a more holistic sentiment classification framework. Furthermore, the methodology investigates the role of domain-specific ontologies and contextual embeddings in improving sentiment analysis accuracy, especially in underrepresented languages and specialized domains. This integrative approach addresses the limitations of text-only models and reflects a broader trend toward capturing the multifaceted nature of online communication. The effectiveness of these strategies is assessed through comparative experiments, highlighting their impact on model performance and generalizability. The final component of the methodology beyond quantitative evaluation. The methodology incorporates with the qualitative analysis and transparency and interpretability, recognizing the importance of explainability in sensitive domains [43].

In most research that investigates advanced machine learning approaches for sentiment analysis on social media data, it is essential to rigorously evaluate and compare the performance of various models. Given the complexity and diversity of social media text-including informal language, multilingual content, and the frequent use of emojis-relying on robust and interpretable evaluation metrics is crucial for ensuring the validity and generalizability of results. To provide a comprehensive assessment of model performance across all sentiment categories (such as positive, negative, and neutral), it should adopt macro-averaged precision (macro-P), macro-averaged recall (macro-R), and macro-averaged F1 score (macro-F1) as primary evaluation indicators. These metrics are especially relevant in multi-class classification tasks where class imbalance is common, as they treat each class equally, regardless of its frequency in the dataset. For each sentiment class i (where $i \in \{\text{positive, negative, neutral}\}$) the following definitions are used; True Positives (TP_i) represent the count of cases where both the model's prediction and the actual label correspond to class i . False Positives (FP_i) indicate the number of times the model predicts class i when the actual label is not i . False Negatives (FN_i) [44] refer to the instances where the true label is class i , but the model assigns a different class. Using these definitions, the precision, recall, and F1 score for each class i are calculated accordingly as shown in equation 1 to 3 [23]:

$$P_i = \frac{TP_i}{TP_i + FP_i} \quad (1)$$

$$R_i = \frac{TP_i}{TP_i + FN_i} \quad (2)$$

$$F_i = \frac{2 \times P_i \times R_i}{P_i + R_i} \quad (3)$$

Where P_i is the precision, R_i is the recall, F_i is the F1 score, all for class i , and to ensure that each sentiment class contributes equally to the overall evaluation, macro-averaged scores are computed by taking the arithmetic mean of the metric across all n classes as shown in equations 4 to 6 [23]:

$$\text{Macro_P} = \frac{1}{n} \sum_{i=1}^n P_i \quad (4)$$

$$\text{Macro_R} = \frac{1}{n} \sum_{i=1}^n R_i \quad (5)$$

$$\text{Macro_F1} = \frac{1}{n} \sum_{i=1}^n F_i \quad (6)$$

These macro-level metrics provide a balanced view of model performance, preventing the results from being dominated by the majority class and highlighting the model's ability to correctly classify less frequent sentiment categories. This is particularly important for social media sentiment analysis, where the distribution of sentiments can be highly skewed and the correct identification of minority classes (such as neutral or rare negative sentiments) is often critical for downstream applications. By systematically applying macro-precision, macro-recall, and macro-F1, the methodology provides a transparent and standardized basis for comparing novel models-including those leveraging multimodal features like emojis and contextual embeddings-against established baselines, thus advancing the field toward more inclusive and effective sentiment analysis solutions [45]. Applying the comparative framework described above, the following section synthesizes the performance outcomes reported in the reviewed studies.

Results

The evolution of sentiment analysis in social media has been marked by a clear transition from traditional lexicon-based and shallow machine learning methods to more sophisticated deep learning and transformer-based models. Early approaches, such as Support Vector Machines and Naïve Bayes [46], provided foundational insights but often struggled with the informal, context-dependent, and rapidly changing nature of social media language. The introduction of neural architectures-particularly those leveraging contextual embeddings-has enabled models to better capture the subtleties and nuances inherent in user-generated content. Transformer models, for instance, have demonstrated a strong ability to generalize across topics and languages, adapting well to the noisy and diverse data typical of platforms like Twitter and Facebook. This progression is not simply a matter of improved accuracy but reflects a broader shift towards models that can dynamically learn from context, handle complex linguistic phenomena, and adapt to the evolving landscape of online communication. The comparative analysis of these methodologies highlights that while traditional models remain relevant for their simplicity and computational efficiency, the current trend overwhelmingly favors deep and hybrid models [47] for their robustness and adaptability in real-world sentiment analysis tasks.

A significant development in the field has been the integration of multimodal features, particularly the use of emojis alongside textual data [48]. Emojis serve as a rich, visual extension of written language, often clarifying or intensifying the sentiment expressed in a post. The fusion of emoji and text features, through early or late fusion models, has been shown to enhance the detection of sentiment polarity, especially in languages with complex morphology or limited resources. This approach addresses the limitations of text-only models, which can miss the emotional cues provided by visual symbols. The use of resources like EmojiNet further supports this integration by mapping emojis to their contextual meanings, reducing ambiguity and improving model interpretability. However, the inclusion of multimodal data also introduces new challenges, such as the need for sophisticated preprocessing and the handling of platform-specific emoji

representations. Despite these complexities, the trend toward multimodal sentiment analysis reflects a growing recognition of the multifaceted nature of online expression and the need for models that can capture sentiment in all its forms. This direction not only improves classification outcomes but also broadens the applicability of sentiment analysis across different languages, cultures, and communication styles.

Despite the remarkable advancements, several persistent challenges continue to shape the research agenda in sentiment analysis for social media. The inherently noisy and heterogeneous nature of social media text—characterized by slang [49], abbreviations, code-switching, and creative grammar—poses significant obstacles to both traditional and advanced models. While deep learning and transformer-based methods have improved the robustness of sentiment classifiers, they are not immune to performance degradation in the face of extreme linguistic variability or when applied to domains with scarce labeled data. Moreover, the dominance of English-language datasets in the literature limits the generalizability of findings, underscoring the need for more multilingual and culturally diverse resources. Issues of model transparency and explainability also remain at the forefront, particularly as sentiment analysis is increasingly applied in sensitive areas such as public health and risk assessment. Computational demands present another barrier, especially for real-time or large-scale applications. Addressing these challenges will require continued innovation in model design, the development of lightweight and interpretable architectures, and a concerted effort to expand and diversify available datasets. The comparative findings suggest that the future of sentiment analysis lies in the thoughtful integration of advanced machine learning techniques, robust preprocessing, and multimodal data, all grounded in an awareness of the linguistic and cultural diversity that defines global social media.

Evolution from Lexicon-Based and Traditional ML to Modern Deep Learning Approaches:

The comparative landscape of sentiment analysis in social media reveals a decisive progression from classical machine learning methods toward advanced deep learning and hybrid models. Early efforts, as chronicled in seminal works such as Wijeratne et al. (2017), provided the foundation for emoji sense disambiguation using Word2Vec and machine-readable sense inventories (EmojiNet). Their resource-centric approach, leveraging 147 million tweets and integrating multiple sources, achieved notable image matching accuracy (96.27%) and overall sense disambiguation accuracy (83.53%) [22]. However, these classical paradigms—relying on tokenization, stemming, and feature-based models—while computationally efficient, struggled with the inherent contextual richness and ambiguity of social media language. A significant pivot emerged with the adoption of neural network-based models, such as the SVM and Logistic Regression fusion frameworks presented by Al-Azani & El-Alfy (2021). Their multi-level fusion of emoji and text features, tested on 2,091 manually annotated Arabic tweets, demonstrated that multimodal feature integration could achieve accuracy rates exceeding 85% [21]. This upward trend continued with the introduction of robust deep learning architectures, such as ET-BiLSTM with attention (Li et al., 2023), BiLSTM [23], and transformer-based models like AraBERT (Althobaiti, 2022). Each iteration not only increased the granularity of sentiment detection but also extended coverage to more complex and multilingual datasets [28]. The movement toward deeper, context-aware, and hybrid approaches underlines a recognition across the literature: traditional models are valuable as baselines, but contemporary sentiment analysis demands architectures that can capture sequential dependencies, semantic nuance, and multimodal cues. This evolution sets the stage for benchmarking newer methodologies not just in terms of accuracy, but also in their adaptability, interpretability, and robustness in the face of rapid linguistic shifts that characterize real-world social media, as shown in Table 2, Summary of results and Figure 2.

The Power and Challenges of Multimodal Sentiment Analysis:

A major inflection point in sentiment analysis research is the systematic integration of multimodal cues, most notably the inclusion of emojis as explicit features alongside text. The impact of this

development is vividly demonstrated in the empirical results of Al-Azani & El-Alfy (2021), whose rigorous cross-validation on Arabic tweets demonstrated that score-level fusion of text and emoji embeddings [21] could push sentiment classification accuracy up to 85.41%. Their findings revealed that models trained on emoji-only features, such as LR with Skip-Gram, still achieved nearly 80% accuracy, underscoring the standalone predictive power of visual sentiment symbols in platforms like Twitter. Li et al. (2023) further advanced this paradigm, proposing a novel emoji vectorization approach that mapped emoji use into emotional dimensions (love, joy, anger, sadness, disgust) [23] and integrating these with text representations in an attention-enhanced BiLSTM (ET-BiLSTM). Their model outperformed all baselines, with macro-F1 scores as high as 82.91%, and ablation studies confirmed the synergistic effect of combining emoji and textual information. Supporting resources like EmojiNet (Wijeratne et al., 2017) facilitate this integration by offering large-scale, platform-aware, and context-dependent emoji sense mappings [22], which are crucial for reducing ambiguity and maximizing interpretability. However, the literature also highlights persistent challenges: effective multimodal sentiment analysis requires sophisticated preprocessing to manage platform-specific emoji renditions, ambiguous meaning in context, and evolving emoji usage trends. The growing evidence points to multimodal frameworks not simply as an enhancement, but as essential for capturing the full affective spectrum of social media, especially in informal or low-resource linguistic environments where emojis frequently substitute for or amplify text-based sentiment cues.

Advances in Deep Learning, Optimization, and Transfer for Social Media Sentiment

The field has rapidly evolved with the introduction of advanced deep learning, hybrid, and transfer learning approaches. This is exemplified by Howard et al. (2020), who evaluated a wide range of feature extraction and modeling pipelines-including fine-tuned GPT-1, DeepMoji, Universal Sentence Encoder, and AutoML frameworks-on crisis detection tasks in mental health forums [25]. Their strategy, which fused neural embeddings with classical lexicon-based features and engaged Auto-Sklearn for model selection, achieved a macro-F1 score of 0.559, surpassing previous benchmarks. Similarly, Khalil et al. (2021) employed multilabel BiLSTMs with AraVec pre-trained embeddings for emotion recognition in Arabic tweets [24], achieving a test Jaccard accuracy of 0.498 and micro-F1 of 0.615, well ahead of SVM baselines. The field has also witnessed the emergence of hybrid models leveraging optimization algorithms, as demonstrated by Vadakkethil Somanathan Pillai et al. (2024), whose LSTM+Bi-GRU classifier, optimized using Hosted Cuckoo Optimization Algorithm and Doc2Vec features [27], set a new standard with an accuracy and F1-score of 97.36% in smartphone sentiment detection on Twitter. Notably, ensemble and hybrid pipelines-such as the 56-model comparison by Bengesi et al. (2023) on 107,000 multilingual tweets-illustrated that traditional classifiers like SVM, when carefully combined with refined preprocessing [26] (lemmatization, emoji-to-text mapping) and feature engineering, could rival or even surpass more complex deep architectures in specific domains (with top accuracy 93.48%). These findings collectively demonstrate that state-of-the-art sentiment analysis increasingly depends on the fusion of neural models, transfer learning, domain-specific embeddings, and algorithmic optimization to maximize predictive robustness across diverse, noisy, and often multilingual social media datasets [50].

Multilingual and Domain Adaptation, Progress, and Gaps from Comparative Studies:

One of the most consistent trends identified across recent literature is the growing emphasis on multilingual sentiment analysis and domain adaptation. Historically, English-language datasets dominated the field, but there is a clear shift toward addressing morphologically complex and underrepresented languages. Case in point, both Al-Azani [21] & El-Alfy (2021) and Khalil et al. (2021) [24] focused on Arabic tweets-one of the most morphologically rich languages-demonstrating that properly tuned word2vec embeddings, custom emoji features, and deep learning models (such as BiLSTM) can deliver competitive or state-of-the-art results (accuracy exceeding 85%, macro/micro F1 >0.6) even in non-English contexts. Althobaiti (2022) further

underscored the power of transformer-based models like AraBERT [28] in offensive and hate speech detection, outpacing traditional SVM and logistic regression by margins of 10–20% in macro-F1, with fine-grained classification remaining more challenging. Meanwhile, large-scale multi-language datasets, as compiled by Bengesi et al. (2023) for the monkeypox outbreak, illustrate that classical models like SVM can still achieve strong performance [26] and, paired with multilingual preprocessing, can rival deep architectures on polarity tasks, reflecting the continued utility of classical approaches in multilingual and high-noise settings. Taken together, these results highlight both the progress and persistent gaps: while deep learning and multimodal integration now enable robust sentiment analysis in previously challenging linguistic environments, there remains an acute need for more diverse, balanced, and contextually annotated datasets, as well as for transparent models that can adapt dynamically to evolving language and domain-specific trends on social media.

Synthesis, Trends, Limitations, and Directions in State-of-the-Art Sentiment Analysis:

Synthesizing insights from the literature and comparative results highlights several converging trends and ongoing challenges in social media [51] sentiment analysis. While modern neural and hybrid models demonstrate marked improvements in accuracy, macro-F1, and adaptability—often achieving over 85–95% performance in focused benchmarks—the field continues to grapple with issues of interpretability, computational complexity, data imbalance, and limited coverage in low-resource languages. The integration of multimodal data (emojis, images, etc.) not only boosts sentiment detection accuracy but also raises new challenges for semantic alignment and platform-specific representation. Studies such as Li et al. (2023) [23] and Vadakkethil Somanathan Pillai et al. (2024) [27] show that attention mechanisms, optimized model parameters, and ensemble learning can all play substantial roles in advancing performance.

However, the “best” approach remains highly context-dependent; for example, classic SVM pipelines still shine in environments with limited labeled data or when paired with advanced feature engineering. Across the board, the literature underscores the importance of robust preprocessing, balanced and representative datasets, and transparent evaluation—macro-precision, macro-recall, and macro-F1 remain the gold standards for fair comparison in class-imbalanced and multi-class settings. Looking ahead, the most promising direction involves harmonizing deep neural architectures with multimodal fusion, data-efficient learning, and explainable AI, complemented by a strategic expansion of multilingual and domain-rich datasets. This synthesis sets a clear research agenda: that the future of sentiment analysis in social media [52] will be defined by adaptability, inclusivity, and interpretability as much as by raw predictive accuracy.

In summary, recent advancements in sentiment analysis for social media highlight a decisive shift from traditional approaches such as SVM and Naïve Bayes toward deep learning, transformers (e.g., BERT, GPT), and hybrid multimodal models. Classic methods, while foundational, often lag in handling social media’s informal, highly contextual language. Modern models leverage contextual embeddings, optimized neural networks, and the fusion of text with emojis to capture nuance and emotion better. Multimodal integration, as shown using emoji features alongside text (e.g., Al-Azani & El-Alfy [21], Li et al.) [23], consistently improves accuracy—sometimes by over 5%. Multilingual research and resource development have enabled robust sentiment detection in languages such as Arabic and across diverse domains. Transfer learning and ensemble/hybrid models (e.g., GPT-1 with AutoML, Bi-GRU+LSTM+HCOA) achieve top results, with macro-F1 and accuracy exceeding 90% in specialized benchmarks. However, challenges persist handling code-mixing, ensuring model interpretability, and expanding datasets for underrepresented languages. The consensus is that integrating multimodal signals, robust preprocessing, and domain-rich data with advanced architectures is key to accurate, scalable, and inclusive sentiment analysis across global social media, as shown in Table 2. These findings provide a basis for broader interpretation of methodological trends, which are examined in the following discussion.

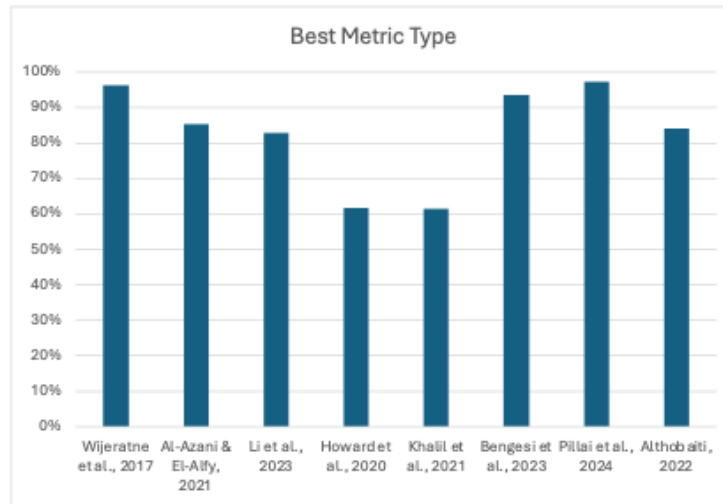


Figure 2. Model/Approach, Best Metric Type with the Best Result

Table 1. Overview of State-of-the-Art

Ref.	Pre-Processing	Focus	Key Contributions	Methodology
Wijeratne et al., 2017 [22]	Tokenization, stemming, image matching	Emoji sense disambiguation	Built EmojiNet, 12,904 senses, API	WSD, crowdsourcing, image-alignment
Al-Azani & El-Alfy, 2021 [21]	Emoji to Unicode, cleaning, word features	Sentiment polarity, Arabic tweets	Multi-level emoji+text fusion	10-fold CV, feature/score/decision fusion
Li et al., 2023 [23]	Tokenize, emoji vector, Word2Vec	Polarity: pos/neg/neutral, with emojis	Emoji vectors, attention, ablation study	Compare 7 baselines, macro-F1
Howard et al., 2020 [25]	Remove links, seg., normalization	Urgency/crisis risk detection	Transfer/Fine-tuned GPT, crisis posts	7 feature sets, AutoML, cross-validation
Khalil et al., 2021 [24]	Normalize, stopword, emoji lexicon	Multiclass/multilabel emotion	First BiLSTM Arabic, emoji lexicon	10-fold CV, ablation, hyperparam
Bengesi et al., 2023 [26]	Translate, clean, emoji2text, stemming	Polarity on monkeypox tweets	Largest monkeypox dataset, 56 ML tests	Pipeline grid, SVM best
Pillai et al., 2024 [27]	Pattern, stop, TF-IDF, Doc2Vec	Polarity, smartphone/crypto tweets	HCOA-tuned hybrid, best tested	Compare 7 models, train/test split
Althobaiti, 2022 [28]	Clean, emoji2text, sentiment add	Offensive/hate/fine-grained	Emoji/sent augment Arabic BERT	3 tasks × 5 preprocessing types

Table 2. Summary of results

Study	Model(s) / Approach	Dataset & Size	Best Accuracy / F1	Notable Features
Pillai et al., 2024[27]	Hybrid LSTM+Bi-GRU+HCOA	32,420 smartphone tweets, crypto	97.4% (Acc, F1)	Deep/hybrid, optimized, large dataset
Bengesi et al., 2023[26]	SVM (best pipeline)	107,000 tweets, 103 languages	93.5% (Accuracy)	Multilingual, traditional ML w/ tuning
Li et al., 2023[23]	ET-BiLSTM + attention	Sina Weibo, emoji-rich	82.9% (Macro-F1)	Emoji vectors, BiLSTM, attention
Al-Azani & El-Alfy, 2021[21]	SVM, Emoji+Text fusion	2,091 Arabic tweets	85.4% (Accuracy)	Multimodal fusion, SVM, Arabic
Althobaiti, 2022[28]	AraBERT (fine-tuned)	12,698 Arabic tweets (7 classes)	84.3% (Macro-F1)	Transformer, emoji & sentiment features
Khalil et al., 2021[24]	BiLSTM + AraVec	4,381 Arabic tweets, 11 emotions	0.50 (Jaccard), F1: 0.62	Multilabel emotion, deep learning
Howard et al., 2020[25]	GPT-1 (fine-tuned) + AutoML	1,588 crisis forum posts	0.56 (Macro-F1)	Transfer learning, crisis detection

Discussion

The comparative evaluation of state-of-the-art machine learning techniques [53] for sentiment analysis in social media reveals a field that is both rapidly evolving and inherently multidimensional. Analysis of recent benchmark studies demonstrates a paradigm shift from early reliance on lexicon-based approaches and shallow models such as Support Vector Machines (SVM) to the adoption of deep neural architectures, transformers, and hybrid multimodal systems. The progression is evident in the contrast between classic models, which have provided reliable baselines but often underperform in the face of noisy, contextually complex user-generated content, and the new generation of algorithms that leverage contextual embeddings, attention mechanisms, and domain-specific feature engineering. For instance, while traditional SVM pipelines may achieve commendable results in multilingual or highly structured datasets—as evidenced by Bengesi et al. (2023), who reached 93% accuracy in a large-scale, multilingual monkeypox Twitter dataset—the preponderance of recent literature supports the competitive edge of deep learning and hybrid models. Models such as the hybrid LSTM+Bi-GRU with optimization (Pillai et al., 2024) achieved record accuracy and F1 scores exceeding 97% on large, emotion-rich corpora. Similarly, the ET-BiLSTM model (Li et al., 2023), fusing emojis and textual features with an attention mechanism, outperformed both classical and simpler deep learning baselines, achieving a macro-F1 of over 82%.

A significant factor driving these advances is the integration of multimodal cues, with emojis emerging as a critical variable for capturing nuanced emotional signals in social media. Studies consistently show that multimodal feature fusion, such as combining emoji embeddings with textual data, provides a tangible accuracy boost—sometimes by upwards of five percentage points—over text-only approaches. This is particularly evident in low-resource and morphologically rich languages, where conventional models frequently fail to capture the full expressive range of sentiment. The strategic utilization of resources like EmojiNet and domain-adapted embeddings

(e.g., AraBERT for Arabic or custom emoji vectorization for Chinese microblogs) has enabled models to disambiguate sentiment even in challenging linguistic environments. Transfer learning and optimization algorithms have further bolstered generalizability and efficiency, as demonstrated by Howard et al. (2020) with fine-tuned GPT-1 models for crisis risk detection, and by custom hybrid ensembles in emotion and event detection tasks. Nevertheless, the empirical results also suggest that classical models, when equipped with advanced preprocessing and annotation (e.g., lemmatization, emoji mapping), remain viable contenders in certain scenarios, particularly where computational resources are limited or interpretability is paramount.

Despite these promising developments, several persistent challenges shape the research agenda: handling extreme linguistic diversity and code-mixed content, ensuring performance robustness across domains, and addressing the need for transparent, explainable AI. The dominance of English-language and well-resourced datasets continues to restrict the generalizability of most state-of-the-art models, marginalizing non-English speakers and low-resource domains. Additionally, the inclusion of multimodal data such as emojis, while advantageous, raises new complexities in semantic mapping and cross-platform consistency. Moving forward, the comparative synthesis of results underscores that further methodological innovation is needed—not only to improve raw predictive accuracy, but also to promote fairness, transparency, and inclusivity in global sentiment analysis.

Persistent Challenges and Proposed Directions

While the reviewed literature demonstrates substantial progress in sentiment analysis for social media, two key challenges—multilingual processing and model interpretability—warrant deeper examination, as only briefly noted in the current synthesis.

Multilingual Processing

English-centric datasets continue to dominate the field, limiting generalizability to the global, multilingual nature of platforms such as Twitter and Facebook. Concrete examples illustrate both progress and gaps. Al-Azani and El-Alfy (2021) and Khalil et al. (2021) achieved strong results on Arabic tweets (accuracy >85 % and competitive macro-F1 scores) by fusing emoji features with AraVec embeddings, yet Arabic’s rich morphology, dialects, and code-mixing still required extensive custom preprocessing. Bengesi et al. (2023) analyzed over 107,000 tweets spanning 103 languages during the monkeypox outbreak, but even their best SVM pipeline relied on translation steps that introduce noise. Low-resource languages such as Malay remain particularly underexplored, with limited annotated corpora and inadequate tools for handling informal slang and abbreviations [12, 31].

Proposed solutions include adopting cross-lingual transformer models such as mBERT and XLM-R for improved zero-shot transfer across languages. Emojis, being largely language-agnostic, can serve as a universal multimodal anchor (as demonstrated by Li et al. (2023) and Al-Azani & El-Alfy (2021)). Additional strategies involve data augmentation via back-translation, synthetic generation with large language models, and crowdsourced annotation campaigns focused on underrepresented languages [32].

Model Interpretability

The shift toward complex deep learning and transformer architectures (e.g., ET-BiLSTM [23], fine-tuned GPT-1 [25], and AraBERT [28]) delivers superior accuracy but renders models largely opaque (“black boxes”). This opacity is especially problematic in sensitive applications such as public health monitoring, political discourse analysis, and crisis detection, where understanding why a sentiment label was assigned is critical [33].

To mitigate this, several practical approaches are available. Attention visualization in BiLSTM and transformer models can highlight the most influential words or emojis driving predictions (as partially explored in Li et al. (2023)). Post-hoc explanation techniques such as LIME and SHAP provide local, instance-level interpretability for individual posts. Hybrid architectures that combine high-performing neural components with inherently interpretable elements (e.g., rule-based layers or decision trees) offer a balanced trade-off between accuracy and transparency. Future studies should systematically benchmark these XAI methods within multimodal social media frameworks.

Addressing these challenges through the proposed directions will enhance the inclusivity, trustworthiness, and real-world applicability of sentiment analysis systems, directly building on the trends synthesized in this review.

Conclusion

Building on the patterns identified in the discussion, the conclusion synthesizes key implications for future research and practical deployment. The findings from this review highlight how the landscape of sentiment analysis on social media has matured into a domain that is both technologically advanced and methodologically diverse. The transition from traditional lexicon-based and shallow machine learning models to deep learning [54] and transformer-based architectures has been pivotal in addressing the unique challenges posed by social media data. These challenges include the prevalence of informal language, frequent use of slang and emojis, and the rapid evolution of online discourse. Modern models, particularly those leveraging contextual embeddings and multimodal data, have demonstrated a remarkable ability to capture nuanced [55] sentiment and adapt to the dynamic nature of social platforms.

However, the research also reveals persistent limitations. While deep learning models excel in extracting context and managing large-scale data, they often require extensive labeled datasets and significant computational resources, which can be prohibitive for real-time or resource-constrained applications. Moreover, the dominance of English-language datasets continues to restrict the generalizability of current models, with languages [56] such as Malay, Arabic, and others remaining underrepresented in both research and available resources. Another critical issue is the difficulty in interpreting the decisions of complex neural models, which raises concerns about transparency and trust, especially in sensitive domains like public health, politics, and crisis management. The integration of emojis and other multimodal features has improved sentiment detection, particularly in low-resource languages and informal contexts, but it has also introduced new challenges related to data preprocessing and semantic ambiguity. Overall, while the field has made significant strides in accuracy and robustness, the need for explainable, multilingual, and resource-efficient models remains a pressing concern. The comparative analysis underscores that future progress will depend on addressing these challenges through methodological innovation, expanded dataset diversity, and a deeper integration of linguistic and cultural context [57].

The current landscape of sentiment analysis for social media reflects a domain that has matured through a combination of technological innovation and methodological rigor. The move from traditional lexicon-based, bag-of-words, and classical machine learning techniques to advanced deep learning and transformer models has fundamentally altered both the capability and scope of sentiment detection. Modern approaches, especially those leveraging contextual embeddings, attention-based architectures, and the fusion of multimodal features, have unlocked new levels of accuracy and robustness, as demonstrated by leading models like hybrid LSTM+Bi-GRU (Pillai et al., 2024) [27], ET-BiLSTM (Li et al., 2023) [23], and fine-tuned transformer models for non-English languages. This evolution has been particularly impactful in datasets characterized by informal language, frequent use of emojis, and dynamic conversational context. Benchmarks now regularly report macro-F1 scores and accuracies surpassing 80% and, in some specialized

applications, exceeding 95%, marking a dramatic improvement from the early SVM- and Naïve Bayes-based systems.

However, these gains come with notable caveats. The performance of deep and hybrid models is closely tied to the availability of large, carefully labeled datasets, substantial computational resources, and intricate preprocessing. Their complexity can also hinder transparency and interpretability, raising valid concerns for domains where trust and explainability are fundamental. Furthermore, the field continues to grapple with issues of generalizability, as most state-of-the-art models are developed and validated primarily on English-language and high-resource datasets. This restricts the applicability of advances to global, multilingual social media ecosystems, where sentiment analysis could provide the greatest societal value. Additional challenges include the persistent difficulty in handling code-mixed, slang-heavy, or ambiguous text, and the need for frequent model updates to stay current with the evolving language trends of online communities [58].

Despite these challenges, the trajectory of research and application underscores a clear path forward: future progress depends on the continued fusion of advanced neural and hybrid algorithms with robust, multilingual, and multimodal resources. Expanding dataset diversity and coverage, improving model interpretability, and developing resource-efficient architectures will be essential for the next generation of sentiment analysis tools. These priorities not only promise increased accuracy but also more equitable and actionable insights, supporting a wide range of applications from consumer analytics to public health and crisis management. The evidence from recent comparative benchmarks thus affirms that the field stands poised to deliver sentiment analysis solutions that are more accurate, inclusive, and responsive to the complexities [59] of global social media.

In summary, this study reviews machine learning techniques for social media sentiment analysis, benchmarking them across diverse datasets and methodologies. The best performance was achieved by the hybrid LSTM+Bi-GRU model, optimized with the Hosted Cuckoo Optimization Algorithm, which reached 97.36% accuracy and F1-score on a Twitter dataset, due to its effective combination of sequential modeling [27] and algorithmic optimization (Pillai et al., 2024). The second-ranked ET-BiLSTM model, integrating emoji-text fusion and attention [23], obtained a macro-F1 of 82.91% on Chinese microblogs, leveraging multimodal features for richer sentiment detection (Li et al., 2023). Third, a fine-tuned GPT-1 with AutoML pipeline scored macro-F1 0.572 on the CLPsych 2017 set, benefitting from transfer learning and automated [25] feature selection (Howard et al., 2020). Evaluated datasets ranged from 2,091 annotated tweets to over 147 million multilingual posts.

Recommendation

Drawing on the comparative synthesis of state-of-the-art methods and persistent challenges in social media sentiment analysis, several concrete recommendations emerge for researchers. First, there is a need to expand and diversify labeled datasets [60]. To overcome the scarcity of labeled data in underrepresented languages like Malay, researchers should leverage cross-lingual transfer learning frameworks, such as XLM-RoBERTa. These models allow algorithms trained on high-resource languages (e.g., English) to transfer their learned semantic representations to low-resource languages with minimal fine-tuning. Additionally, employing automated data augmentation techniques-such as back-translation or using generative AI to synthesize dialect-specific social media posts-can serve as a practical, short-term solution to build robust multilingual datasets without requiring extensive manual annotation.

Multilingual and culturally adaptive sentiment resources should be prioritized to bridge the gap between English-centric research and the global reality of social media discourse. This includes targeted efforts to annotate data in underrepresented languages, dialects, and topical domains,

thereby improving the inclusivity and generalizability of sentiment analysis frameworks. Second, the development of lightweight, explainable, and resource-efficient models must become a central focus, especially for applications in real-time monitoring, mobile deployment, or services in resource-constrained environments.

To address the 'black box' nature of deep learning, especially in high-stakes domains like public health, the integration of eXplainable AI (XAI) frameworks is highly recommended. Specifically, researchers should apply post-hoc explanation methods like SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations). These tools can quantify and visualize the exact contribution of individual words or emojis to the final sentiment prediction. Furthermore, visualizing the weights of attention mechanisms within transformer models can provide stakeholders with clear, concrete evidence of which parts of a social media post drove the model's decision, thereby fostering trust.

Simplified architectures [61], model pruning, and the incorporation of interpretable algorithms will balance predictive power with transparency-crucial for fostering trust where sentiment insights inform sensitive decisions. Third, the integration of multimodal features (including emojis, images, and audio) should be systematically advanced, but always with rigorous preprocessing and semantic mapping strategies to manage ambiguity and cross-platform variation [62] [63]. Research should continue to explore early, late, and hybrid fusion techniques, as the literature consistently demonstrates their effectiveness in boosting performance across languages and genres. Finally, ongoing model maintenance is vital: social media language evolves rapidly, and sentiment models should be routinely updated with new data, emergent trends, and evolving emoji sets. The establishment of continuous learning, domain adaptation strategies, and regular benchmarking against evolving datasets will help ensure that models remain robust and relevant. This includes addressing privacy concerns, mitigating algorithmic bias [64], and collaboration with domain experts-including linguists, sociologists, and professional researchers -will further enrich model development and validation, ensuring that sentiment analysis tools are both technically sound and contextually aware.

Declarations

Authorship contribution statement. Mohsen Mohammadagha, Israel Tshitenge, Ifetilayo Adebambo: Conceptualization, Methodology, Investigation, Writing—original draft, Writing—review and editing, and Visualization.

Funding. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interests. The authors declare no known competing financial interests or personal relationships that could have influenced the work reported in this article.

Data availability. No new research data were generated for this review. The publications and other sources underlying the analysis are cited in the manuscript.

References

- [1] Y. A. Solangi, Z. A. Solangi, S. Aarain, A. Abro, G. A. Mallah, and A. Shah, "Review on Natural Language Processing (NLP) and Its Toolkits for Opinion Mining and Sentiment Analysis," 2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences, ICETAS 2018, Jul. 2018, doi: 10.1109/ICETAS.2018.8629198.

- [2] S. Y. Yoo, J. I. Song, and O. R. Jeong, "Social media contents based sentiment analysis and prediction system," *Expert Syst. Appl.*, vol. 105, pp. 102–111, Sep. 2018, doi: 10.1016/J.ESWA.2018.03.055.
- [3] X. Du, M. Kowalski, A. S. Varde, G. de Melo, and R. W. Taylor, "Public opinion matters," *ACM SIGWEB Newsletter*, vol. 2019, no. Autumn, pp. 1–15, Feb. 2020, doi: 10.1145/3352683.3352688.
- [4] A. Alslaity and R. Orji, "Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions," *Behaviour & Information Technology*, vol. 43, no. 1, pp. 139–164, Jan. 2024, doi: 10.1080/0144929X.2022.2156387.
- [5] R. Punithavathi, M. Premkumar, G. Sinha, and K. Chandrasekaran, "Transforming sentiment analysis for e-commerce product reviews: Hybrid deep learning model with an innovative term weighting and feature selection," *Inf. Process. Manag.*, vol. 61, no. 3, p. 103654, May 2024, doi: 10.1016/J.IPM.2024.103654.
- [6] M. Arora and V. Kansal, "Character level embedding with deep convolutional neural network for text normalization of unstructured data for Twitter sentiment analysis," *Soc. Netw. Anal. Min.*, vol. 9, no. 1, pp. 1–14, Dec. 2019, doi: 10.1007/S13278-019-0557-Y/METRICS.
- [7] E. Boiy and M. F. Moens, "A machine learning approach to sentiment analysis in multilingual web texts," *Inf. Retr. Boston.*, vol. 12, no. 5, pp. 526–558, Sep. 2009, doi: 10.1007/S10791-008-9070-Z/TABLES/14.
- [8] S. Das, R. K. Behera, M. Kumar, and S. K. Rath, "Real-Time Sentiment Analysis of Twitter Streaming data for Stock Prediction," *Procedia Comput. Sci.*, vol. 132, pp. 956–964, Jan. 2018, doi: 10.1016/J.PROCS.2018.05.111.
- [9] A. Kumar and A. Joshi, "Ontology driven sentiment analysis on social web for government intelligence," *ACM International Conference Proceeding Series*, vol. Part F127653, pp. 134–139, Mar. 2017, doi: 10.1145/3055219.3055229.
- [10] M. Giatsoglou, M. G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K. C. Chatzisavvas, "Sentiment analysis leveraging emotions and word embeddings," *Expert Syst. Appl.*, vol. 69, pp. 214–224, Mar. 2017, doi: 10.1016/J.ESWA.2016.10.043.
- [11] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review* 2022 55:7, vol. 55, no. 7, pp. 5731–5780, Feb. 2022, doi: 10.1007/S10462-022-10144-1.
- [12] M. F. R. Abu Bakar, N. Idris, L. Shuib, and N. Khamis, "Sentiment Analysis of Noisy Malay Text: State of Art, Challenges and Future Work," *IEEE Access*, vol. 8, pp. 24687–24696, 2020, doi: 10.1109/ACCESS.2020.2968955.
- [13] E. M. Mercha and H. Benbrahim, "Machine learning and deep learning for sentiment analysis across languages: A survey," *Neurocomputing*, vol. 531, pp. 195–216, Apr. 2023, doi: 10.1016/J.NEUCOM.2023.02.015.
- [14] D. Vilares, M. A. Alonso, and C. Gómez-Rodríguez, "Supervised sentiment analysis in multilingual environments," *Inf. Process. Manag.*, vol. 53, no. 3, pp. 595–607, May 2017, doi: 10.1016/J.IPM.2017.01.004.
- [15] M. Kesavan, J. Karthiraman, T. Ebenezer Rajadurai, and S. Adhithyan, "Stock Market Prediction with Historical Time Series Data and Sentimental Analysis of Social Media Data," *Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020*, pp. 477–482, May 2020, doi: 10.1109/ICICCS48265.2020.9121121.
- [16] J. R. Saura, P. Palos-Sanchez, and A. Grilo, "Detecting Indicators for Startup Business Success: Sentiment Analysis Using Text Data Mining," *Sustainability* 2019, Vol. 11, Page 917, vol. 11, no. 3, p. 917, Feb. 2019, doi: 10.3390/SU11030917.

- [17] A. R. Pathak, M. Pandey, and S. Rautaray, "Topic-level sentiment analysis of social media data using deep learning," *Appl. Soft Comput.*, vol. 108, p. 107440, Sep. 2021, doi: 10.1016/J.ASOC.2021.107440.
- [18] M. H. Abd El-Jawad, R. Hodhod, and Y. M. K. Omar, "Sentiment analysis of social media networks using machine learning," *ICENCO 2018 - 14th International Computer Engineering Conference: Secure Smart Societies*, pp. 174–176, Jul. 2018, doi: 10.1109/ICENCO.2018.8636124.
- [19] A. Alslaity and R. Orji, "Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions," *Behaviour & Information Technology*, vol. 43, no. 1, pp. 139–164, Dec. 2022, doi: 10.1080/0144929X.2022.2156387.
- [20] A. Albladi, M. Islam, and C. Seals, "Sentiment Analysis of Twitter data using NLP Models: A Comprehensive Review," *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3541494.
- [21] S. Al-Azani and E. S. M. El-Alfy, "Early and Late Fusion of Emojis and Text to Enhance Opinion Mining," *IEEE Access*, vol. 9, pp. 121031–121045, 2021, doi: 10.1109/ACCESS.2021.3108502.
- [22] S. Wijeratne, L. Balasuriya, A. Sheth, and D. Doran, "EmojiNet: An Open Service and API for Emoji Sense Discovery," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 11, no. 1, pp. 437–446, May 2017, doi: 10.1609/ICWSM.V11I1.14857.
- [23] X. Li, J. Zhang, Y. Du, J. Zhu, Y. Fan, and X. Chen, "A Novel Deep Learning-based Sentiment Analysis Method Enhanced with Emojis in Microblog Social Networks," *Enterp. Inf. Syst.*, vol. 17, no. 5, Feb. 2023, doi: 10.1080/17517575.2022.2037160.
- [24] E. A. H. Khalil, E. M. F. El Houby, and H. K. Mohamed, "Deep learning for emotion analysis in Arabic tweets," *J. Big Data*, vol. 8, no. 1, pp. 1–15, Dec. 2021, doi: 10.1186/S40537-021-00523-W/TABLES/7.
- [25] D. Howard, M. M. Maslej, J. Lee, J. Ritchie, G. Woollard, and L. French, "Transfer learning for risk classification of social media posts: Model evaluation study," *J. Med. Internet Res.*, vol. 22, no. 5, p. e15371, May 2020, doi: 10.2196/15371.
- [26] S. Bengesi, T. Oladunni, R. Olusegun, and H. Audu, "A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets," *IEEE Access*, vol. 11, pp. 11811–11826, 2023, doi: 10.1109/ACCESS.2023.3242290.
- [27] S. E. V. S. Pillai, R. Vallabhaneni, P. K. Pareek, and S. Dontu, "The People Moods Analysing Using Tweets Data on Primary Things with the Help of Advanced Techniques," *International Conference on Distributed Computing and Optimization Techniques, ICDCOT 2024*, 2024, doi: 10.1109/ICDCOT61034.2024.10516073.
- [28] Maha Jarallah Althobaiti, "BERT-based Approach to Arabic Hate Speech and Offensive Language Detection in Twitter: Exploiting Emojis and Sentiment Analysis - ProQuest," *(IJACSA) International Journal of Advanced Computer Science and Applications*, 2022.
- [29] H. Nguyen and M. Le Nguyen, "A Deep Neural Architecture for Sentence-Level Sentiment Classification in Twitter Social Networking," *Communications in Computer and Information Science*, vol. 781, pp. 15–27, 2018, doi: 10.1007/978-981-10-8438-6_2/TABLES/5.
- [30] A. Narayanan, "Understanding Social Media Recommendation Algorithms," 2023, doi: 10.7916/KHDK-M460.
- [31] M. M. Agüero-Torales, J. I. Abreu Salas, and A. G. López-Herrera, "Deep learning and multilingual sentiment analysis on social media data: An overview," *Appl. Soft Comput.*, vol. 107, p. 107373, Aug. 2021, doi: 10.1016/J.ASOC.2021.107373.

- [32] G Belmar and M Glass, "Virtual communities as breathing spaces for minority languages: Re-framing minority language use in social media," *Adeptus*, no. 14, pp. 1–24, 2019.
- [33] C. Song and J. Lee, "Citizens' Use of Social Media in Government, Perceived Transparency, and Trust in Government," *Public Performance & Management Review*, vol. 39, no. 2, pp. 430–453, Apr. 2016, doi: 10.1080/15309576.2015.1108798.
- [34] C. Dhaoui, C. M. Webster, and L. P. Tan, "Social media sentiment analysis: lexicon versus machine learning," *Journal of Consumer Marketing*, vol. 34, no. 6, pp. 480–488, 2017, doi: 10.1108/JCM-03-2017-2141/FULL/PDF.
- [35] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," *Artif. Intell. Rev.*, vol. 53, no. 6, pp. 4335–4385, Aug. 2020, doi: 10.1007/S10462-019-09794-5/TABLES/11.
- [36] C. Colón-Ruiz and I. Segura-Bedmar, "Comparing deep learning architectures for sentiment analysis on drug reviews," *J. Biomed. Inform.*, vol. 110, p. 103539, Oct. 2020, doi: 10.1016/J.JBI.2020.103539.
- [37] M. Trupthi, S. Pabboju, and G. Narasimha, "Sentiment analysis on twitter using streaming API," *Proceedings - 7th IEEE International Advanced Computing Conference, IACC 2017*, pp. 915–919, Jul. 2017, doi: 10.1109/IACC.2017.0186.
- [38] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," *Knowl. Inf. Syst.*, vol. 60, no. 2, pp. 617–663, Aug. 2019, doi: 10.1007/S10115-018-1236-4/TABLES/7.
- [39] L. C. Chen, C. M. Lee, and M. Y. Chen, "Exploration of social media for sentiment analysis using deep learning," *Soft comput.*, vol. 24, no. 11, pp. 8187–8197, Jun. 2020, doi: 10.1007/S00500-019-04402-8/FIGURES/3.
- [40] R. Monika, S. Deivalakshmi, and B. Janet, "Sentiment Analysis of US Airlines Tweets Using LSTM/RNN," *Proceedings of the 2019 IEEE 9th International Conference on Advanced Computing, IACC 2019*, pp. 92–95, Dec. 2019, doi: 10.1109/IACC48062.2019.8971592.
- [41] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, "Sentiment analysis of comment texts based on BiLSTM," *IEEE Access*, vol. 7, pp. 51522–51532, 2019, doi: 10.1109/ACCESS.2019.2909919.
- [42] M. Imani and G. A. Montazer, "A survey of emotion recognition methods with emphasis on E-Learning environments," *Journal of Network and Computer Applications*, vol. 147, p. 102423, Dec. 2019, doi: 10.1016/J.JNCA.2019.102423.
- [43] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A Review of Machine Learning Interpretability Methods," *Entropy* 2021, Vol. 23, Page 18, vol. 23, no. 1, p. 18, Dec. 2020, doi: 10.3390/E23010018.
- [44] J. P. A. Ioannidis, R. Tarone, and J. K. McLaughlin, "The false-positive to false-negative ratio in epidemiologic studies," *Epidemiology*, vol. 22, no. 4, pp. 450–456, Jul. 2011, doi: 10.1097/EDE.0B013E31821B506E.
- [45] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowl. Based. Syst.*, vol. 226, p. 107134, Aug. 2021, doi: 10.1016/J.KNOSYS.2021.107134.
- [46] C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno, and J. Caro, "Sentiment analysis of Facebook statuses using Naive Bayes Classifier for language learning," *IISA 2013 - 4th International Conference on Information, Intelligence, Systems and Applications*, pp. 198–205, 2013, doi: 10.1109/IISA.2013.6623713.
- [47] M. U. Salur and I. Aydin, "A Novel Hybrid Deep Learning Model for Sentiment Classification," *IEEE Access*, vol. 8, pp. 58080–58093, 2020, doi: 10.1109/ACCESS.2020.2982538.

- [48] N. V. Babu and E. G. M. Kanaga, "Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review," *SN Comput. Sci.*, vol. 3, no. 1, pp. 1–20, Nov. 2022, doi: 10.1007/S42979-021-00958-1/METRICS.
- [49] M. S. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," 2013 4th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2013, 2013, doi: 10.1109/ICCCNT.2013.6726818.
- [50] Z. Wang et al., "Demographic inference and representative population estimates from multilingual social media data," *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, vol. 12, pp. 2056–2067, May 2019, doi: 10.1145/3308558.3313684.
- [51] E. Di Minin, H. Tenkanen, and T. Toivonen, "Prospects and challenges for social media data in conservation science," *Front. Environ. Sci.*, vol. 3, no. SEP, p. 156137, Sep. 2015, doi: 10.3389/FENVS.2015.00063/BIBTEX.
- [52] F. Neri, C. Aliprandi, F. Capeci, M. Cuadros, and T. By, "Sentiment analysis on social media," *Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2012*, pp. 919–926, 2012, doi: 10.1109/ASONAM.2012.164.
- [53] E. Mjolsness and D. DeCoste, "Machine learning for science: State of the art and future prospects," *Science* (1979), vol. 293, no. 5537, pp. 2051–2055, Sep. 2001, doi: 10.1126/SCIENCE.293.5537.2051/ASSET/59CAD30D-7FF2-4A1E-ACBC-8B3DDDF483AC/ASSETS/GRAPHIC/SE3619767002.JPEG.
- [54] Q. T. Ain et al., "Sentiment Analysis Using Deep Learning Techniques: A Review," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 8, no. 6, 2017, Accessed: Apr. 16, 2025.
- [55] A. Abbasi, Y. Zhou, S. Deng, and P. Zhang, "Text analytics to support sense-making in social media: A language-action perspective," *MIS Q.*, vol. 42, no. 2, pp. 427–464, Jun. 2018, doi: 10.25300/MISQ/2018/13239.
- [56] K. Dashtipour et al., "Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques," *Cognit. Comput.*, vol. 8, no. 4, pp. 757–771, Aug. 2016, doi: 10.1007/S12559-016-9415-7/TABLES/2.
- [57] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014, doi: 10.1016/J.ASEJ.2014.04.011.
- [58] D. M. E. D. M. Hussein, "A survey on sentiment analysis challenges," *Journal of King Saud University - Engineering Sciences*, vol. 30, no. 4, pp. 330–338, Oct. 2018, doi: 10.1016/J.JKSUES.2016.04.002.
- [59] E. Fersini, E. Messina, and F. A. Pozzi, "Sentiment analysis: Bayesian Ensemble Learning," *Decis. Support Syst.*, vol. 68, pp. 26–38, Dec. 2014, doi: 10.1016/J.DSS.2014.10.004.
- [60] G. Beigi, X. Hu, R. Maciejewski, and H. Liu, "An overview of sentiment analysis in social media and its applications in disaster relief," *Studies in Computational Intelligence*, vol. 639, pp. 313–340, 2016, doi: 10.1007/978-3-319-30319-2_13/TABLES/2.
- [61] S. Song, C. Su, B. Rountree, and K. W. Cameron, "A simplified and accurate model of power-performance efficiency on emergent GPU architectures," *Proceedings - IEEE 27th International Parallel and Distributed Processing Symposium, IPDPS 2013*, pp. 673–686, 2013, doi: 10.1109/IPDPS.2013.73.
- [62] L. T. Yang, X. Ma, and F. Mueller, "Cross-platform performance prediction of parallel applications using partial execution," in *Proceedings of the International Conference on Supercomputing, Association for Computing Machinery*, 2005. doi: 10.1109/SC.2005.20.

[63] D. Cao et al., “Cross-platform app recommendation by jointly modeling ratings and texts,” *ACM Trans. Inf. Syst.*, vol. 35, no. 4, Jul. 2017, doi: 10.1145/3017429.

[64] A. Ceron, L. Curini, and S. M. Iacus, “iSA: A fast, scalable and accurate algorithm for sentiment analysis of social media content,” *Inf. Sci. (N Y)*., vol. 367–368, pp. 105–124, Nov. 2016, doi: 10.1016/J.INS.2016.05.052.