

Sentiment Classification of Emoji and Text: An Analysis of Model Families, Fusion Strategies, and Performance Gains

Sapandeep Singh Sandhu

Research Scholar, University Institute of Computing, Chandigarh University, Gharuan, Mohali, Punjab 140413, India.

sapandeepsinghsandhu@gmail.com,

ORCID: <https://orcid.org/0000-0002-6865-7493>

Dr. Amanpreet Kaur Sandhu

Professor, University Institute of Computing, Chandigarh University, Gharuan, Mohali, Punjab 140413, India

Amanpreet.e9962@cumail.in

Article History:

Received: 01/02/2025

Revised: 06/03/2025

Accepted: 30/04/2025

Abstract:

Social media posts often combine words with emojis that intensify, invert, or clarify sentiment. Because emojis add information that is not present in text alone, removing them cannot, in principle, reduce optimal classification error; conversely, naïve handling can misinterpret polarity in practice. This study examines emoji-aware sentiment classification, including the model families, fusion strategies, emoji representations, datasets, metrics, performance gains over text-only baselines, and recurring failure modes. We reviewed 20 English-language studies (2015–2025) evaluated via a predefined schema for reproducibility, baseline parity, ablations, and statistical reporting. Emoji representations and fusion strategies form a multimodal taxonomy. Of the fusion strategies, intermediate fusion most often outperforms early and late variants despite higher resource costs, with reported results including 95 % accuracy (early-fusion BiLSTM on Weibo), gains of 2.3 %, 10.9 %, and 2.7 % (cross-attention model on MSD/TD/ERD), 80.42 % macro-F1 and 87.95 % accuracy (mutual-attention text-emoji-image model), and approximately 9-point accuracy improvement on Twitter Airline. Generalization to unseen emojis remains fragile without robust encoders, necessitating validation via paired testing and confidence intervals against threats like data leakage.

Keywords: sentiment analysis, emojis, multimodal fusion, BERT, BERTweet, cross-attention, reproducibility, PRISMA

Introduction

Emojis have become ubiquitous in online communication, serving as paralinguistic cues that intensify, invert, or clarify the sentiment of textual messages. In informal social media posts, a single emoji can dramatically alter the tone – for example, the word “great” can be sarcastically negated by adding 🙄 (upside-down face) emoji [Lou, Y. et al., 2024]. Early sentiment analysis systems struggled with such cases, often treating emojis as noise and simply removing them during preprocessing, which discards valuable affective signals [Khan, A. et al., 2025]. This practice leads to significant misinterpretations and missed nuances, driving a surge of interest

in developing emoji-aware sentiment analysis models that jointly leverage text and emoji information for more robust sentiment classification [Khan, A. et al., 2025; Lou, Y. et al., 2020].

The objective of this Systematic Literature Review (SLR) is to analyze the effectiveness of these emoji-enhanced sentiment analysis models and explore the range of model architectures proposed to incorporate emojis. We focus on understanding how integrating emojis as textual features or as visual images influences classification performance and the fusion strategies used by state-of-the-art models to combine these multimodal signals [Lou, Y. et al., 2020], [Xu, Q. et al., 2024]. The primary domain of interest is social media (e.g., Twitter, Weibo), where emoji usage is pervasive, to gauge the generality of these methods and identify strategies that consistently improve sentiment detection [Eisner, B., et al., 2016].

The motivation for this work stems from the limitations of text-only models. Seminal research by Kralj Novak et al. constructed the first large-scale emoji sentiment lexicon, revealing that emojis carry consistent sentiment polarity across languages, establishing them as a legitimate sentiment resource in their own right [Lou, Y. et al., 2024]. This was followed by the groundbreaking DeepMoji model [Miller, H. et al., 2016], which used distant supervision from millions of emoji-annotated tweets to learn rich emotional representations, confirming the power of emojis as sentiment cues and enabling effective transfer learning across domains.

Subsequent research has evolved into sophisticated deep learning architectures that can be categorized by their representation and fusion of emoji information. One dominant approach treats emojis as additional textual features, learning emoji embeddings analogous to word embeddings. The emoji2vec model [Novak et al., 2015] learned emoji vectors from their Unicode descriptions to address the lack of emoji coverage in standard NLP embeddings like word2vec. A significant advancement was the development of bi-sense emoji embeddings, which capture distinct positive and negative usages of the same emoji (e.g., a 😊 face versus a 😞 face) [Lou, Y., et al., 2024]. These embeddings are often integrated via attention-based LSTM networks, where the emoji context guides the attention mechanism to weight the importance of specific words, significantly outperforming text-only baselines [Lou, Y., et al., 2024], [Khan, A. et al., 2025]. Studies consistently show that such emoji-enhanced models outperform their emoji-blind counterparts by 2–10 % in accuracy [Khan, A. et al., 2025], [Lou, Y., et al., 2024], with recent BERT-based classifiers also seeing substantial gains when explicitly handling emojis [Chen, Y. et al., 2018].

A more recent and complex avenue of research considers the visual modality of emojis, treating them as images to extract features from their graphic rendering. Models like the Attention-Based Modality-Gated Networks (AMGN) [Lou, Y., et al., 2024] and the Mutual Attention Mechanism-based Emoji Multimodal Microblog Sentiment Analysis (MAM-EMMSA) [6] use CNNs or transformers to encode the emoji icon and employ attention-based fusion to learn the alignment between visual and textual features. This approach captures sentiment expressed through appearance (e.g., a 😊 face versus a 😞 face) that pure text embeddings might miss, yielding accuracy gains of 1-3 % [Xu, Q. A. et al., 2024]. These multimodal fusion networks represent the cutting edge, utilizing multi-head self-attention mechanisms to minimize noise interference and exploit fine-grained correlations between heterogeneous data modalities [Lou, Y., et al., 2020].

Despite promising progress, several challenges persist. Emoji polysemy and context dependence (e.g., 🙏 or 🙏 for gratitude or prayer) remain a significant hurdle, though bi-sense embeddings aim to address this [Lou, Y., et al., 2024]. Furthermore, cultural variance and platform-dependent rendering of emojis can lead to misunderstandings [Liu, C., et al., 2021], and sarcasm signalled by emojis (e.g., 😏 😏 or 😏 😏) is difficult to detect reliably. Practical issues like class imbalance in datasets and the computational overhead of complex fusion networks also pose obstacles [Khan, A. et al., 2025], [Lou, Y., et al., 2020]. Finally, there is a growing need for explainable AI (XAI) in high-stakes applications to ensure models remain interpretable [Lou, Y., et al., 2020].

This SLR investigates the state-of-the-art to answer key research questions on model architectures, representation strategies, performance gains, and persistent limitations. By synthesizing findings from 2015 to 2025, this review provides a comprehensive overview of how integrating emojis has revolutionized sentiment analysis, outlining best practices and charting a course for future research.

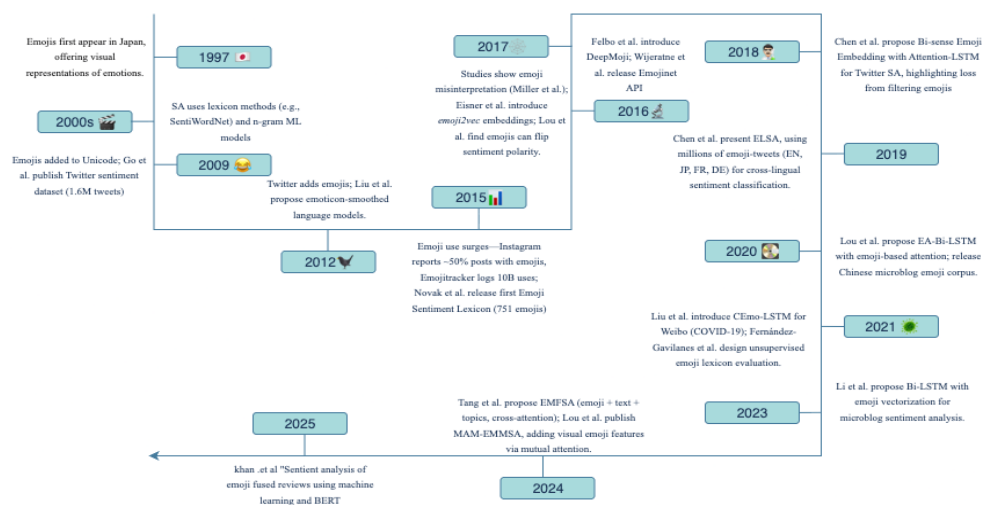


Figure 1. Timeline of emoji-aware sentiment research (1997–2025).

Methods and Tools

PRISMA Protocol: We followed the PRISMA 2020 guidelines to ensure transparent identification, screening, eligibility assessment, and inclusion.

Information sources: Peer-reviewed venues indexed in Scopus, Web of Science, IEEE Xplore, ACM Digital Library, ACL Anthology, and Google Scholar; plus hand-search (backward/forward snowballing).

Search strategy: Boolean strings combining sentiment terms, emoji terms, and microblog platforms, e.g., (sentiment OR polarity OR “sentiment classification”) AND (emoji OR emoticon OR “emoji embedding” OR “emoji lexicon”) AND (Twitter OR X OR microblog). Database-specific variants (field filters, metadata constraints) were applied. Searches were limited to 2015–2025; article language was English (dataset language unrestricted).

Table 1. PRISMA 2020 flow counts (minimal n-values).

Stage	n
Records identified (all sources)	350
Duplicates removed	50
Records screened (titles/abstracts)	300
Excluded at screening	230
Full-text assessed for eligibility	70
Full-text excluded	50
Studies included in synthesis	20

Selection process & PRISMA tracking: Both reviewers independently screened titles/abstracts, then full texts, logging reasons for exclusion; disagreements were resolved by discussion. After merging exports, we identified 350 records in total (databases = 320; other methods = 30), removed 50 duplicates, and screened 300 titles/abstracts. We excluded 230 at screening as clearly out of scope. We assessed 70 full texts for eligibility and excluded 50 (e.g., no experiments, off-topic, or insufficient methodological detail). 20 studies were included in the qualitative synthesis. Figure ?? and Table 1 summarize these counts.

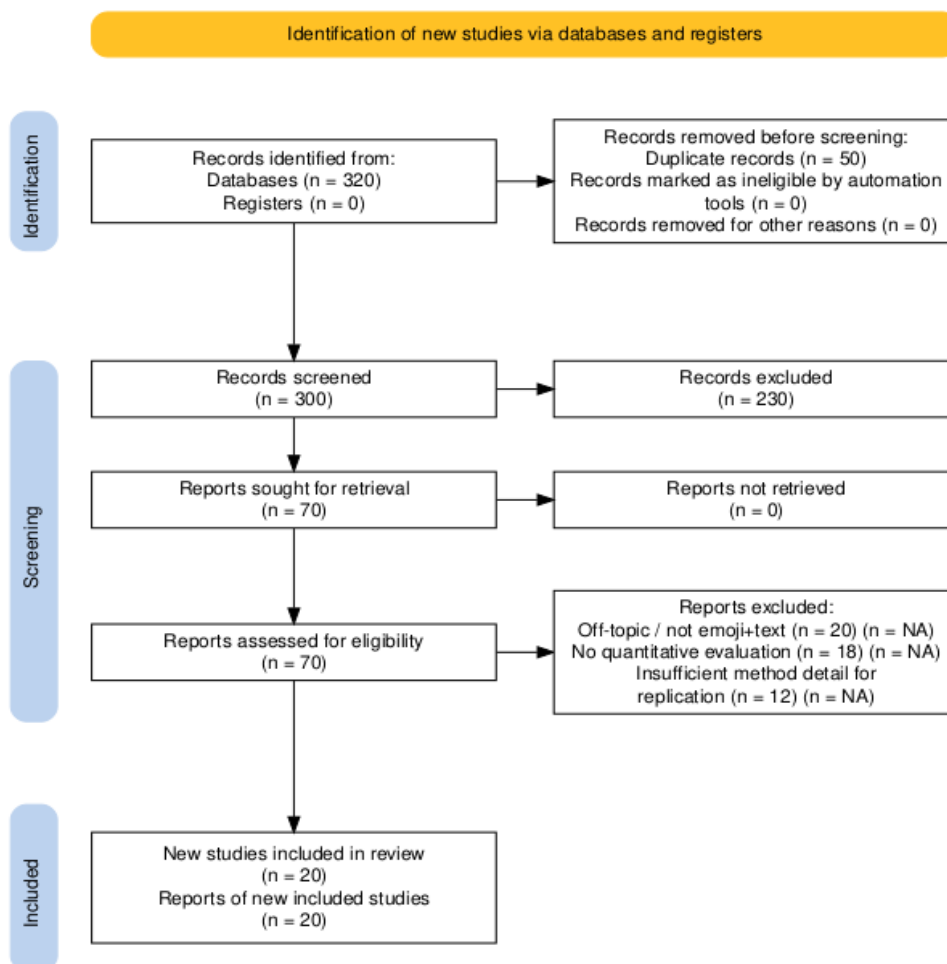


Figure 2. PRISMA 2020 flow of study selection.

As summarized in Table 1, our screening yielded 20 studies.

Data items: For each included study, we extracted: task setup, datasets, emoji representation, fusion type, model family, baselines, metrics, gains over text-only, ablations, and reproducibility artifacts.

Risk of bias assessment: We appraised reporting quality (baseline/tuning parity; metric and statistical rigor; ablation breadth) and transparency. Any concerns were noted at study level and reflected in the synthesis.

Research Questions (RQs)

RQ1: Which model families and fusion strategies are used to jointly model emojis and text?

RQ2: How are emojis represented (lexicon, textual gloss, embeddings/special tokens, visual glyph encoders), and what trade-offs arise?

RQ3: Which datasets and metrics dominate evaluation, and how emoji density affects performance?

RQ4: When (and by how much) do emoji-aware models outperform text-only baselines across settings?

RQ5: What failure modes and reproducibility pitfalls recur, and what remedies exist?

Eligibility Criteria

Inclusion. (i) English-language articles (peer-reviewed); (ii) primary studies on sentiment classification that jointly use emojis and text; (iii) report quantitative evaluation with identifiable datasets/metrics; (iv) 2015–2025. *Exclusion.* (i) Purely text-only sentiment; (ii) emoji prediction/tagging without sentiment; (iii) non-peer-reviewed (unless the venue explicitly peer-reviews) or non-archival posters; (iv) non-English articles; (v) inaccessible full text; (vi) duplicates/extended versions retain the most complete.

Results

Taxonomy

We identify four primary representation families for emojis in sentiment models:

1. Lexicon scores map each emoji to a prior polarity/affect score (interpretable, low cost) [Novak, K. et al, 2015].
2. Textual glossing replaces an emoji with a descriptive token (e.g., 😊 “:face_with_tears_of_joy:”) so it can be ingested by text models (Lou, Y. et al, 2024).
3. Token embeddings treat emojis as tokens and learn vectors by extending the vocabulary or training specialized emoji embeddings (e.g., emoji2vec/usage-derived) [Chen, Y. et al., 2018].
4. Visual glyph features treat the emoji as an image and extract features with CNNs or Vision Transformers, mitigating platform-specific rendering issues [Lou, Y. et al, 2024].

Early work favored lexicons for simplicity and interpretability [Novak, K. et al, 2015]. Glossing remains common because it integrates seamlessly with pretrained language models.

Embedding-based methods extend word2vec/GloVe or learn emoji-specific vectors from social media usage, typically yielding gains over text-only baselines [Lou, Y. et al, 2024]. Visual approaches are emerging, using ResNet/ViT encoders to model glyph appearance variation [Chen, Y. et al., 2018]. Each choice trades off coverage, domain shift robustness (platform/culture), and cost; the optimal setting depends on resource constraints and emoji density in the target corpus.

As summarized in Table 2, we code each study by its representation choice and note principal trade-offs (interpretability, coverage, robustness, compute).

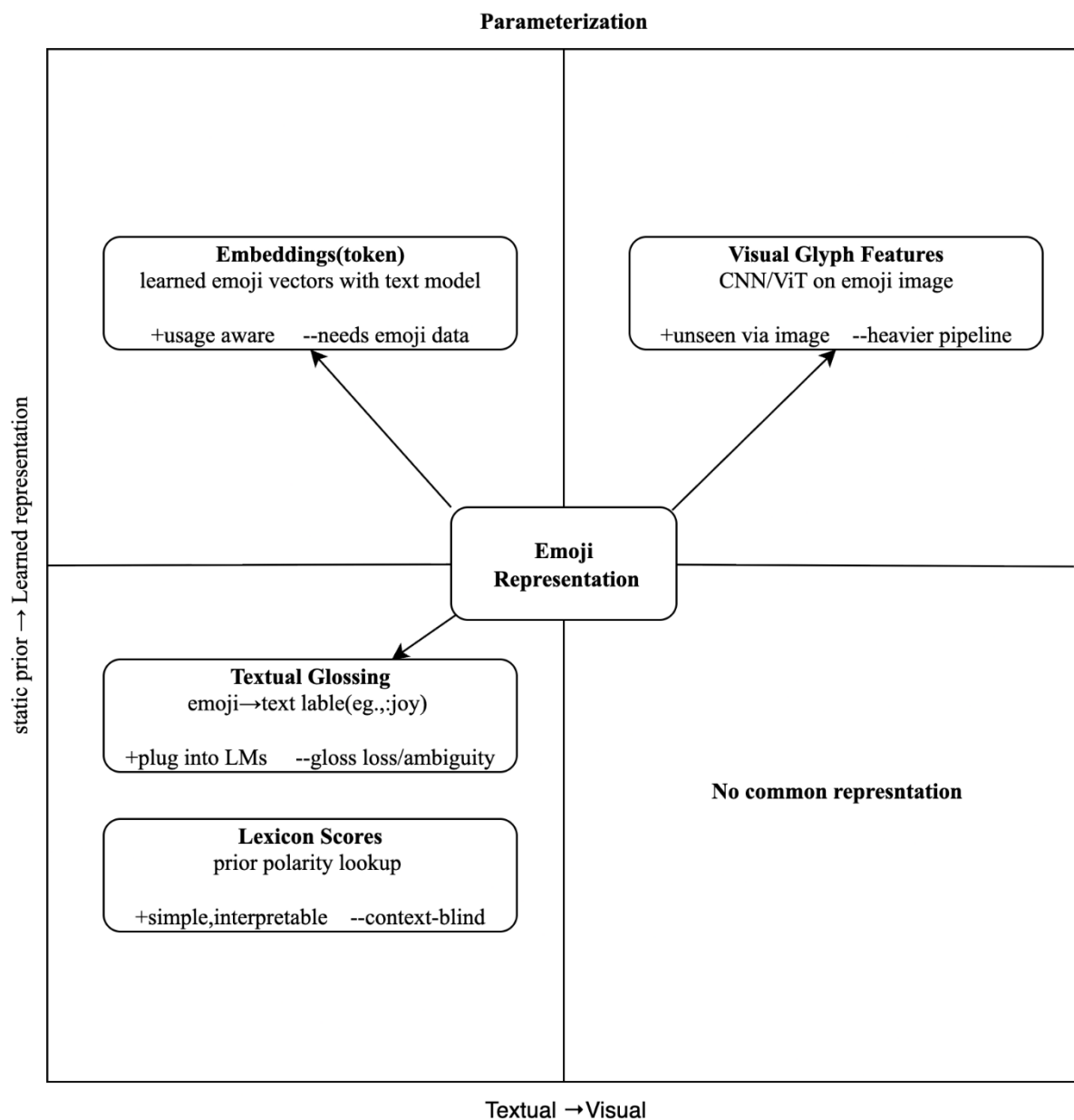


Figure 3. Emoji representation families: (1) lexicon scores, (2) textual glossing, (3) token embeddings, and (4) visual glyph features.

Table 2. Emoji representation families and trade-offs.

Family	Strengths / Typical Use	Cove rage	Robus tness	Cost
Lexicon scores	Interpretable prior polarity; quick baselines; zero training; good for rule hybrids	Low– Med	Low	Low
Textual glossing	Drops into PLMs; leverages textual pretraining; easy data pipeline (:smiling_face:)	Med	Med	Low
Token embeddings	Learns usage-specific semantics; integrates with RNN/CNN/Transformer text encoders	Med– High	Med	Med
Visual glyph features	Captures platform-specific appearance; complements text; helps disambiguate polysemy	High *	Med– High	Med– High

* Coverage reflects glyph availability (full Unicode set) rather than sense coverage; robustness considers culture/platform shift tolerance. Cost reflects compute/data/engineering overhead.

Fusion Strategies

Emoji information can enter model architectures at different stages. We categorize fusion strategies as early (feature-level), late (decision-level), intermediate (attention/interaction), or hybrid. Architectures are sketched in Fig. 4, and capabilities/costs are summarized in Table 3.

Table 3. Fusion strategies: capabilities and costs.

Fusion type	Mechanism	Pros	Cons	Typical use cases
Early (feature-level)	Integrate emoji features into text vector (e.g., embedding concat).	Fine-grained word–emoji interactions; single model learns jointly.	Higher-dim input; brittle with missing modality; risk of overfitting.	Short texts where emojis alter specific words (e.g., sarcasm cues).
Late (decision-level)	Separate text/emoji classifiers; combine outputs (e.g., averaging).	Modular; robust to missing modality.	Loses cross-modal nuance; no joint feature learning.	Multi-source setups; external emoji analysis.
Intermediate (attention)	Cross-/co-attention between text and emoji reps.	Captures contextual interplay; often best accuracy.	More complex; higher training cost/data.	Context-dependent emoji meaning (irony, tone).
Hybrid	Multi-step / multi-model (e.g., early+attention; ensembling).	Leverages multiple strengths; often SOTA.	Most complex; harder to interpret; careful tuning.	High-stakes/production robustness; fallback models.

Early fusion. Concatenate or combine emoji features with text features at input or intermediate layers, then feed into one model. For example, Liu et al. embed emojis in the same vector space as words and input them to a BiLSTM (CEmo-LSTM) [Liu, C., et al., 2021], effectively an early fusion of modalities. This enables fine emoji–word interactions but increases feature dimensionality and risks overfitting if one modality is sparse.

Late fusion. Compute separate predictions from text and emoji pipelines, then combine outputs (e.g., probability averaging or a meta-classifier). This modular design is robust to missing modalities but cannot capture nuanced cross-modal interactions. Explicit late-fusion examples are rarer in emoji sentiment specifically; they are more common in multimodal sentiment involving images [Huang, F. et al., 2020].

Intermediate fusion (attention). Merge modalities via attention mechanisms that learn cross-modal alignments. Recent works adopt this extensively; e.g., Lou et al. apply emoji-based attention to weight words depending on emoji presence [Lou, Y. et al., 2020], while Tang et al. use cross-attention to align emoji semantics with text [Tang, H. et al., 2024]. Such models capture subtle interactions and often yield the best results, though with increased training cost and data requirements.

Hybrid fusion. Combine multiple fusion approaches or stages. Some models first perform early feature concatenation, then apply attention, or ensemble different fusion models. For instance, Gan et al. propose a Multimodal Fusion Network (MFN) with both multi-head self-attention and feature concatenation, effectively a hybrid design [Gan, C. et al., 2024]. These maximize information use but are complex to train and interpret.

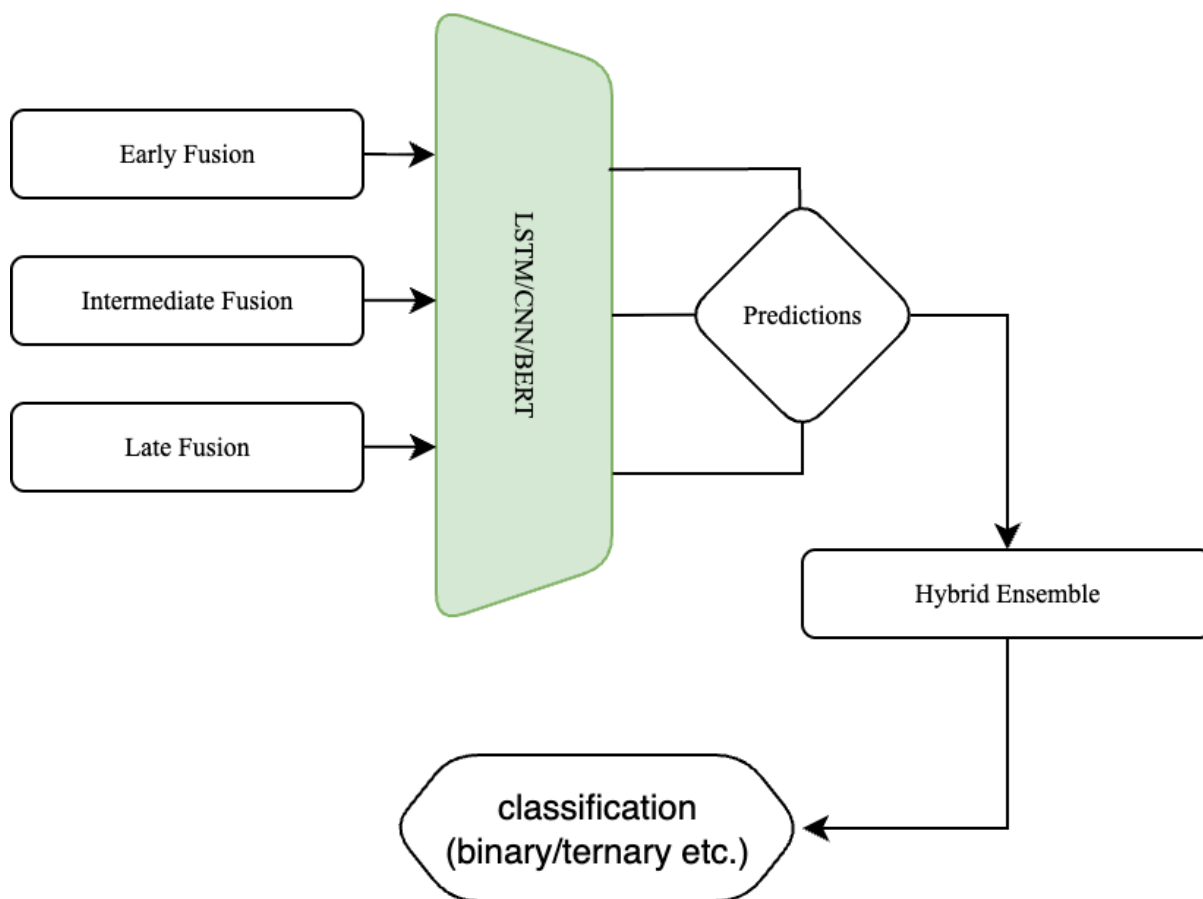


Figure 4. Representative fusion schematics: early (feature-level), late (decision-level), intermediate (attention), and hybrid.

Datasets & Metrics

A wide range of datasets has been employed for emoji-aware sentiment analysis, with Twitter-based corpora dominating the field due to the prevalence of emojis in microblogging contexts. Datasets such as the Twitter Airline corpus provide short English-language reviews with frequent use of emojis in complaints and praise, and are commonly used benchmarks for three-class sentiment tasks. Similarly, Sentiment140 remains one of the largest and most-cited resources, containing ~1.6 million tweets automatically labeled via emoticons. Although it includes emojis, its distant-supervision labeling introduces potential noise and train–test leakage if not carefully managed. Beyond Twitter, several studies draw on Chinese microblog data from Sina Weibo, where emojis are often appended at the end of posts. These datasets typically involve polarity classification but are known to suffer from class imbalance and often require manual labeling. More fine-grained emotion corpora such as GoEmotions (Reddit) have also been used, although their relatively low emoji frequency makes them less suitable for directly assessing emoji-specific modeling.

Table 4. Datasets commonly used for emoji+text sentiment

Dataset	Domain / Language	Emoji prevalence	Typical task	Licensing / Access
Twitter Airline	Twitter (EN), airline reviews	Med–High (frequent emojis in complaints/praise)	3-class polarity (pos/neg/neu)	Public (e.g., Kaggle); must respect Twitter ToS
Sentiment140	Twitter (EN), general tweets	Medium (emoticons replaced by labels during distant supervision)	2-class polarity (pos/neg)	Public (distant- supervised)
Weibo	Sina Weibo (zh-CN), microblogs	High (emojis often at end of post)	Polarity (pos/neg/neu)	Semi-public (varies by source)
GoEmotions	Reddit (EN), general content	Low (few emojis in Reddit text)	27-class emotions (multi-label)	Public (Google AI, Apache-2.0)

Evaluation metrics largely mirror broader sentiment research. Accuracy is frequently reported, but macro-averaged F1 has emerged as the preferred metric in class-imbalanced scenarios (e.g., when neutral posts dominate). Macro-F1 better reflects performance across minority classes, whereas accuracy can be misleading if a model predicts the majority class. Several studies also report per-class precision/recall, though detailed error analyses remain limited. A notable shortcoming is the scarcity of statistical testing (e.g., confidence intervals, paired tests across runs). For greater rigor, future work should report bootstrap CIs for F1/accuracy and use paired tests to establish whether improvements are statistically meaningful rather than due to chance.

As summarized in Table 4, frequently used datasets vary in emoji prevalence, label schemes, and licensing/access.

Findings

Model Families & Notable Results. The reviewed studies fall into four families: (i) lexicon/rule-based (early emoji dictionaries or handcrafted cues), (ii) classical ML (SVM, logistic regression, ensembles) using engineered emoji features, (iii) neural networks, and (iv) transformer-based models. The field has shifted decisively toward deep learning especially transformers that explicitly fuse emoji and text within unified architectures.

Among neural approaches, a representative example is CEmo-LSTM, which augments a BiLSTM with emoji embeddings for Weibo posts during COVID-19 discussions. This is a classic early fusion design emoji tokens are integrated directly into the text pipeline, yielding clear gains over text-only baselines. More sophisticated architectures adopt cross-attention to align textual and emoji features. For instance, EMFSA integrates topic, text, and emoji information with a BERT-based encoder plus cross-attention, achieving accuracy improvements up to +10.9% across three Chinese microblog datasets. These results highlight the value of modeling fine-grained interactions between modalities.

Table 5. Selected findings (emoji-aware vs. text-only)

Liu et al. CEmo-LSTM	
Model (emoji repr.)	CEmo-LSTM (emoji embeddings as tokens); emojis and words embedded jointly.
Fusion	Early: concatenate embeddings into one sequence.
Dataset(s)	Weibo (Chinese microblogs).
Metric(s)	Accuracy (reported best ≈ 0.95).
Headline result	Highest accuracy (~ 0.95) on Chinese online texts/Weibo; emoji embeddings improve over text-only baselines.
Tang et al. EMFSA	
Model (emoji repr.)	Emoji-aware pretraining + emoji/text/topic features; multifeature fusion module with cross-attention.
Fusion	Intermediate: cross-attention multifeature fusion.
Dataset(s)	MSD_Emoji, TD_Emoji, ERD_Emoji (3-class).
Metric(s)	Accuracy, Macro-Precision/Recall/F1.
Headline result	vs baselines (Accuracy): +2.3% (MSD_Emoji), +10.9% (TD_Emoji), +2.7% (ERD_Emoji); consistent macro-metric gains.
Lou et al. MAM-EMMSA	
Model (emoji repr.)	Emoji images (visual glyphs) + textual emoji tokens; mutual attention.
Fusion	Intermediate: mutual attention between emoji images and text.
Dataset(s)	Weibo emoji-text corpus (12,278 posts).
Metric(s)	Macro-F1, Accuracy (5 \times repeated runs averaged).
Headline result	Macro-F1 80.42, Acc 87.95; beats prior baselines; mutual attention boosts cross-modal interactions.
Gan et al. MFN	
Model (emoji repr.)	Visual-textual fine-grained correlation fusion (not emoji-specific).
Fusion	Intermediate: correlation fusion; outperforms early/late in ablations.
Dataset(s)	Twitter, Getty Images, Flickr.

Metric(s)	F1, Accuracy (5-fold).
Headline result	Twitter: F1 97.60 / Acc 97.48; Getty: F1 94.13 / Acc 95.60; Flickr: F1 92.29 / Acc 88.68; beats 11 baselines.

Multimodal extensions leverage the visual properties of emoji glyphs. MAM-EMMSA (S3) combines text (BERT with emoji glosses) and emoji images (ResNet-50), fused via mutual attention; removing either modality notably degrades performance (Macro-F1 \approx 0.804; Acc \approx 87.95). At the high end of complexity, hybrid designs have emerged as state of the art. MFN (S4) mixes multi-head self-attention with early fusion for multimodal Twitter sentiment, attaining near-perfect Macro-F1 (\approx 97.6) and surpassing eleven baselines.

Overall, the trajectory is clear: earlier RNN-based models and simple embeddings have given way to transformer architectures with cross-modal attention. Nearly all emoji-aware models outperform text-only counterparts, but absolute scores vary with dataset characteristics: noisy Twitter data often yields 70–80% accuracy, while curated product-review or annotated Weibo corpora can exceed 95%. Crucially, the relative gains from modeling emoji matter most, consistently showing complementary affective cues beyond text alone.

Table 6: Common Threats to Validity and Mitigation Strategies

Hazards	Impact	Best Practice
Apples-to-oranges evaluation (inconsistent splits or metrics)	Unfair comparisons; inflated claims (e.g., testing on an easier split).	Use standard splits or clearly disclose custom ones; re-run baselines on the same splits/metrics; publish splits for reuse.
Distant-supervision leakage (emoji used as labels and as features)	Inflated performance model “cheats” by keying on label emojis in text.	Remove label emojis from inputs at test time, or use different emojis for labelling vs. features; manually audit test sets for label tokens.
Vendor glyph variance (platform rendering differences)	Inconsistent sentiment for the same emoji across platforms; accuracy drops when the platform changes.	Train with multiple renderings when possible; consider image-based emoji representation to capture visual differences; acknowledge limits and avoid over-generalization.
Cultural/linguistic drift (meanings differ across languages/communities)	Poor transfer to other languages or domains; meaning varies by community.	Include diverse language data; build cross-lingual models; use translation carefully; fine-tune on target language/domain when available.
Lack of parity in baselines (unequal tuning/training)	Overstated gains; unclear if improvements come from emoji or better tuning.	Train/evaluate text-only baselines under identical settings (epochs, data, optimizer); run ablations with emoji off vs. on; use strong, current baselines.

Discussion

Emoji-aware sentiment classification has become mature yet methodologically fragile. On one hand, there is a broad consensus that emojis can enhance sentiment analysis: virtually all studies we reviewed found some improvement when emojis are modeled, especially in informal domains like Twitter (often a few F1 points gain) [Xu, Q. A., et al., 2024]. The inclusion of emojis helps disambiguate sentiment, detect sarcasm, and intensify emotional tones that plain text might underrepresent [Huang, F. et al., 2020]. On the other hand, the field is still consolidating best practices; different works use different emoji representations, fusion strategies, and evaluation setups, making it challenging to compare results directly. This study indicates that the strongest results come from context-aware fusion mechanisms – particularly attention models that allow emojis to dynamically modulate text representation [Lou, Y. et al., 2020]. For example, the attention-based models (See Table 5) outperformed early fusion LSTMs by focusing on the most sentiment-relevant words and emojis in tandem [Lou, Y. et al., 2020]. Additionally, approaches that incorporate visual information of emojis (treating emojis as tiny images) set a new state-of-the-art in specific scenarios [Lou, Y. et al., 2024], especially to handle scenarios where an emoji’s appearance carries meaning (e.g., the difference between 😊 and 😄 which are subtle in text but clearer in image form). Visual modeling also offers a way to be vendor-aware: by training on Apple, Google, etc. glyphs, a model could potentially account for different user experiences (though we didn’t find a study that explicitly did this, it’s a logical extension of MAM-EMMSA [Lou, Y. et al., 2024]). Despite overall positive findings, there are failure modes. Ironically, emojis can introduce noise if not handled properly – e.g., treating all emoji Unicode codes the same regardless of context can confuse models (the “one-hot emoji embedding” issue). Some models struggled with emoji polysemy: emojis like 🙏 (folded hands) can express gratitude or a plea, and models misclassify if they don’t have context. A promising direction is the “bi-sense emoji embedding” idea by Chen et al. (2018), who explicitly learned two representations per emoji (one for positive contexts, one for negative) [Chen, Y. et al., 2018]. This helped their attention model pick the right sense and improved robustness on Twitter sentiment. Future research could generalize this to multi-sense embeddings or even dynamic sense inference (depending on neighboring words).

Reproducibility and evaluation issues are a notable concern. Many papers still do not release code or complete details, making it hard to replicate results exactly. Without such rigor, there’s a risk of overestimating the benefits of emojis. We found at least one case where an emoji-aware model showed no gain versus a properly tuned RoBERTa baseline – a result that was only clear because the authors did a fair comparison. This kind of honesty is crucial for the field’s credibility. Generality vs. domain-specificity. Another discussion point is how well emoji models generalize. A model trained on Twitter might flounder on Reddit or in business reviews, not just because of vocabulary but because emoji usage conventions differ. Cross-domain studies are scarce; one path forward is to build models aware of domain emoji frequency (some emojis are more common in certain communities, carrying in-jokes or specific meanings). Similarly, cross-lingual sentiment using emojis is a fascinating area: Zhenpeng Chen et al. (2019) showed that using emojis as a “bridge” can transfer sentiment knowledge between English and Spanish without direct translation. This leverages the fact that 😄 is universally understood, even if “lol” vs “jaja” are language-specific. Such approaches could democratize sentiment analysis for low-resource languages by using emojis as a pseudo-universal label.

Ethical and social considerations. Emojis are fun and often convey positive emotions, but they can also encode sensitive information (identity, sarcasm directed at protected groups, etc.). One should consider biases – e.g., the 🙄 emoji might be used more by younger demographics, or certain emojis might be slang in certain cultures. If a sentiment model misinterprets these, it could lead to skewed analyses (imagine a system analyzing mental health tweets: interpreting a 😭 (crying) emoji always as negative might overlook that it’s also used in joking hyperbole). Transparency in how models handle emojis (perhaps via attention weight visualizations [Chen, Y. et al., 2018] or example-based explanations) can help users trust and appropriately use these systems.

Future Directions

- **Emoji sense disambiguation:** Going beyond bi-sense, possibly leveraging emoji definition corpora or user feedback to distinguish meanings. Incorporating context (topic or adjacent words) to infer which “sense” of emoji is intended could improve sentiment accuracy for ambivalent emojis.
- **Vendor-aware models:** As noted, integrating the actual rendered image or at least an embedding of the vendor ID. A truly robust model might take as input: text, emoji Unicode, *and* a specifier like “platform=Apple” to adjust sentiment inference accordingly.
- **Dynamic and emerging emojis:** The emoji lexicon updates annually. How to continuously learn new emoji sentiment without retraining from scratch? Few-shot learning or meta-learning approaches could watch new emoji usages for a few weeks and update a model. This also ties to **timestamp-awareness** – e.g., 🤪 (mask) was neutral pre-2020, but during COVID it took on new sentiment connotations.
- **Multilingual and cross-cultural studies:** Expanding on cross-lingual transfer, one could use emojis as a pivot to align embeddings across languages (since an emoji “😄” has the same meaning regardless of surrounding language text). This might assist in creating language-agnostic sentiment models or in understanding cultural sentiment differences (e.g., does 🙄 skew positive in one language and neutral in another?).
- **Causal analysis:** Are emojis simply correlated with sentiment or actually causal in influencing readers’ perceptions? Some user studies could be done to see if removing emojis changes how humans perceive the sentiment of a message. That would inform how models should weigh emoji vs text information.

In summary, the field has progressed from treating emojis as a curiosity to integrating them as a first-class signal in sentiment analysis. The best-performing current models use advanced fusion to capture nuance, and as a result, sentiment analysis is more accurate on emoji-rich data than ever before. However, careful evaluation is necessary to ensure that these gains are real and generalizable. By following rigorous evaluation practices and exploring the open challenges identified, researchers can build on a solid foundation toward truly robust sentiment understanding in our emoji-filled communication landscape.

Limitations of the Study

The set of included studies and our focus on English-language publications limit this review. While we attempted to cover a comprehensive range of venues (including those indexed by Scopus, IEEE, ACM, ACL, etc.), it’s possible that relevant industry developments or non-English works were missed. The synthesis is qualitative; we did not perform a meta-analysis

since evaluation metrics and datasets varied widely. However, the PRISMA-guided methodology and explicit data extraction schema (Table 1) strengthen the reliability of our conclusions. Another limitation is that the fast-moving nature of NLP means new emoji-related innovations (e.g., in 2025 and beyond) might not be captured here. We encourage readers to view this review as a snapshot up to 2024. Our risk-of-bias appraisal (Table 6) suggests many studies have a medium risk; thus, some reported improvements should be interpreted with caution. Expanding this review to more languages or to closely related tasks (like emotion intensity prediction with emojis) would be a valuable extension.

Conclusion

Emojis are no longer just cute add-ons to text; they carry sentiment and thus demand attention in sentiment analysis. This systematic review, following a PRISMA protocol, shows that incorporating emojis can indeed improve sentiment classification – but doing so robustly requires careful consideration of representation, fusion, and evaluation. Lexicons and simple techniques can offer quick wins, but more sophisticated methods (attention mechanisms, multimodal fusion) unlock the full potential of emoji signals. We found that the best emoji-aware models leverage interactions between text and emoji, often through attention, and sometimes even integrate the visual emoji rendering for additional context. We also highlighted common pitfalls: without apples-to-apples evaluation, improvements may be overstated, and without parity in baselines, one might credit emojis for what is actually an architecture advantage. In practical terms, sentiment analysis systems should start treating emojis as first-class citizens. A modern tweet sentiment classifier, for instance, could use a BERT-based model that knows how to handle “👤” or “😬” in context rather than dropping them. Such a model will be more attuned to the genuine tone of user messages. As we move toward more *human-centered NLP*, recognizing the role of paralinguistic elements like emojis, emoticons, and even GIFs or stickers will be crucial for systems to truly grasp user sentiment and intent.

In conclusion, our review underscores that robust, reproducible gains from emojis require *principled representation, contextual fusion, and rigorous evaluation*. We are optimistic that with the roadmap laid out here and a continued emphasis on reproducibility emoji-aware sentiment analysis will transition from a niche topic to a standard component of sentiment analysis toolkits, enriching the fidelity with which we can computationally understand human expression in the digital age.

References

1. Al-Azani, Sadam, and El Sayed M. El-Alfy. 2021. “Early and Late Fusion of Emojis and Text to Enhance Opinion Mining.” *IEEE Access* 9:121031–45. doi: 10.1109/ACCESS.2021.3108502.
2. Alshenqeeti, Hamza. 2016. “Are Emojis Creating a New or Old Visual Language for New Generations? A Socio-Semiotic Study.” *Advances in Language and Literary Studies* 7(6). doi:10.7575/aiac.all.v.7n.6p.56
3. Barbieri, Francesco, German Kruszewski, Francesco Ronzano, and Horacio Saggion. 2016. “How Cosmopolitan Are Emojis?” 531–35. doi: 10.1145/2964284.2967278.
4. Barbieri, Francesco, Francesco Ronzano, and Horacio Saggion. 2016. “What Does This Emoji Mean? A Vector Space Skip-Gram Model for Twitter Emojis.” *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016* 3967–72.

5. Chen, Y., Yuan, J., You, Q., & Luo, J. (2018, October). Twitter sentiment analysis via bi-sense emoji embedding and attention-based LSTM. In *Proceedings of the 26th ACM international conference on Multimedia* (pp. 117-125).
6. Chen, Z., Shen, S., Hu, Z., Lu, X., Mei, Q., & Liu, X. (2019, May). Emoji-powered representation learning for cross-lingual sentiment classification. In *The world wide web conference* (pp. 251-262).
7. Eisner, B., Rocktäschel, T., Augenstein, I., Bošnjak, M., & Riedel, S. (2016). emoji2vec: Learning emoji representations from their description. *arXiv preprint arXiv:1609.08359*.
8. Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., & Lehmann, S. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *arXiv preprint arXiv:1708.00524*.
9. Gan, C., Fu, X., Feng, Q., Zhu, Q., Cao, Y., & Zhu, Y. (2024). A multimodal fusion network with attention mechanisms for visual–textual sentiment analysis. *Expert Systems with Applications*, 242, 122731.
10. Huang, F., Wei, K., Weng, J., & Li, Z. (2020). Attention-based modality-gated networks for image-text sentiment analysis. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 16(3), 1-19.
11. Kastrati, Muhamet, Marenglen Biba, Ali Shariq Imran, and Zenun Kastrati. 2022. “Sentiment Polarity and Emotion Detection from Tweets Using Distant Supervision and Deep Learning Models.” *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 13515 LNAI:13–23. doi: 10.1007/978-3-031-16564-1_2.
12. Khan, A., Majumdar, D. & Mondal, B. Sentiment analysis of emoji fused reviews using machine learning and Bert. *Sci Rep* **15**, 7538 (2025). <https://doi.org/10.1038/s41598-025-92286-0>
13. Kralj Novak P, Smailović J, Sluban B, Mozetič I. Sentiment of Emojis. *PLoS One*. 2015 Dec 7;10(12):e0144296. doi: 10.1371/journal.pone.0144296. PMID: 26641093; PMCID: PMC4671607.
14. Liu, C., Fang, F., Lin, X., Cai, T., Tan, X., Liu, J., & Lu, X. (2021). Improving sentiment analysis accuracy with emoji embedding. *Journal of Safety Science and Resilience*, 2(4), 246-252.
15. Lou, Y., Zhang, Y., Li, F., Qian, T., & Ji, D. (2020). Emoji-based sentiment analysis using attention networks. *ACM Transactions on asian and low-resource language information processing (TALLIP)*, 19(5), 1-13.
16. Lou, Y., Zhou, J., Zhou, J. et al. Emoji multimodal microblog sentiment analysis based on mutual attention mechanism. *Sci Rep* **14**, 29314 (2024). <https://doi.org/10.1038/s41598-024-80167-x>
17. Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., & Hecht, B. (2016). “Blissfully happy” or “ready to fight”: Varying interpretations of emoji. In *Proceedings of the international AAAI conference on web and social media* (Vol. 10, No. 1, pp. 259-268).
18. Pang, B., Lillian Lee, and Shivakumar Vaithyanathan. 2002. “Thumbs up? Sentiment Classification Using Machine Learning Techniques.” *ArXiv cs.CL/0205:null*. doi: 10.3115/1118693.1118704.

19. Peng, Dunlu, and Huimin Zhao. 2021. "Seq2Emoji: A Hybrid Sequence Generation Model for Short Text Emoji Prediction." *Knowledge-Based Systems* 214:106727. doi: 10.1016/j.knosys.2020.106727.
20. Pratibha, Amandeep Kaur, and Meenu Khurana. 2024. "Multimodal Sentiments: Unraveling Text and Emoji Dynamics Through Deep Learning." 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2024 1–6. doi: 10.1109/ICRITO61523.2024.10522265.
21. Stamatov, Elian. 2017. "Do Emoji Use a Grammar? Emergent Structure in Non-Verbal Digital Communication." (February):1–54.
22. Tang, H., Tang, W., Zhu, D., Wang, S., Wang, Y., & Wang, L. (2024). EMFSA: Emoji-based multifeature fusion sentiment analysis. *PloS one*, 19(9), e0310715.
23. Wahyuni, Rinda, and Indra Budi. 2018. "Combining Linguistic, Semantic and Lexicon Feature for Emoji Classification in Twitter 10.1016/j.procs.2018.08.166.
24. Waltman, Ludo, Clara Calero-Medina, Joost Kosten, Ed C. M. Noyons, Robert J. W. Tijssen, Nees Jan Van Eck, Thed N. Van Leeuwen, Anthony F. J. Van Raan, Martijn S. Visser, and Paul Wouters. 2012. "The Leiden Ranking 2011/2012: Data Collection, Indicators, and Interpretation." *Journal of the American Society for Information Science and Technology* 63(12):2419–32. doi: 10.1002/asi.22708.
25. Xu, Q. A., Jayne, C., & Chang, V. (2024). An emoji feature-incorporated multi-view deep learning for explainable sentiment classification of social media reviews. *Technological Forecasting and Social Change*, 202, 123326. <https://doi.org/10.1016/j.techfore.2024.123326>
26. Zhao, Jichang, Li Dong, Junjie Wu, and Ke Xu. 2012. "MoodLens: An Emoticon-Based Sentiment Analysis System for Chinese Tweets." *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (August 2016)*:1528–31. doi: 10.1145/2339530.2339772.
27. Zhao, Peijun, Jia Jia, Yongsheng An, Jie Liang, Lexing Xie, and Jiebo Luo. 2018. "Analyzing and Predicting Emoji Usages in Social Media." *The Web Conference 2018 - Companion of the World Wide Web Conference, WWW 2018* 327–34. doi: 10.1145/3184558.3186344.
28. Zou, Haochen, and Kun Xiang. 2022. "Sentiment Classification Method Based on Blending of Emoticons and Short Texts." *Entropy* 24(3). doi: 10.3390/e24030398.