Vol. 8, No. 1, February 2026

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Multimodal Implicit Sentiment Analysis for Tourism Development: A Systematic Literature Review

Yoannes Romando Sipayung[1], Mochamad Agung Wibowo[2], Ridwan Sanjaya[3]

[1]Doctoral Program of Information System, Postgraduate School, Diponegoro University, Semarang, Indonesia

[2]Postgraduate School, Diponegoro University, Semarang, Indonesia

[3]Department of Information System, Soegijapranata Catholic University, Semarang, Indonesia

Received:
December 4, 2025
Revised:
January 21, 2026
Accepted:
February 2, 2026
Published:
March 1, 2026

Corresponding Author:
Author Name*:
Yoannes Romando Sipayung
Email*:
orinandariska8@gmail.com

DOI:
10.63158/journalisi.v8i1.1436

Abstract. This study aims to examine the application of multimodal approaches in implicit sentiment detection within the tourism sector to support data-driven digital development strategies. This review identifies prevailing trends, methodologies, datasets, and scientific novelties in multimodal sentiment analysis capable of capturing hidden emotions, such as sarcasm and ambiguity, in tourist reviews. Using a systematic literature review approach, ten core studies published between 2020 and 2025 were analyzed to identify prevailing research trends, dominant methodological frameworks, commonly used datasets, and emerging scientific contributions. Results demonstrate that multimodal deep learning models—particularly those employing attention-based fusion and contrastive learning—consistently outperform unimodal approaches in recognizing nuanced tourist emotions that are not explicitly stated in text. Despite these advances, the review reveals a significant gap in tourism-specific and Indonesian-context studies, as well as an overreliance on general-purpose social media datasets. This review provides a conceptual and methodological foundation for implementing multimodal implicit sentiment analysis in tourism decision-making systems, enabling destination managers and policymakers to develop early warning mechanisms for tourist dissatisfaction, enhance destination quality assessment, and support more targeted and sustainable tourism development strategies.

Keywords: Multimodal Sentiment Analysis, Implicit Sentiment, Tourism Development, Systematic Literature Review

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

The development of the tourism sector has significant potential to stimulate regional economic growth through increased revenue generation, employment creation, investment attraction, and infrastructure development [1], [2]. These impacts are particularly pronounced in rural areas, where tourism activities can enhance local community income and activate supporting economic sectors [3]. As tourism expands, understanding tourist behavior becomes increasingly critical for ensuring sustainable and competitive destination development. Alongside this growth, tourist behavior has evolved rapidly due to technological advancements and increased digital connectivity. Travelers now rely heavily on online platforms and digital information sources to explore destinations, compare alternatives, and share experiences [4]. This transformation has reshaped how tourism stakeholders monitor destination performance and respond to tourists' needs.

The widespread use of social media and online review platforms has resulted in a substantial volume of user-generated content documenting tourist experiences [5]. These online reviews provide emotionally rich information that is valuable for tourism practitioners and policymakers seeking to understand tourists' perceptions and satisfaction levels [6]. Platforms such as TripAdvisor have become central sources for evaluating destination performance, as they allow users to share opinions, ratings, and visual impressions that collectively reflect tourist preferences and experiences [7], [8], [9], [10], [11], [12].

Sentiment analysis has therefore become an important analytical tool for interpreting tourists' opinions expressed in online reviews. Most existing tourism studies rely on text-based sentiment analysis to classify opinions into positive, negative, or neutral categories [13], [14], [15]. While effective for capturing explicitly stated emotions, text-only approaches face inherent limitations when tourists' express opinions indirectly or ambiguously. Tourist evaluations often rely on nuanced language, contextual cues, or subjective expressions whose meanings cannot be reliably inferred from textual content alone. Moreover, tourist sentiment is not conveyed exclusively through text. Images uploaded by tourists—such as photographs of destinations, hotel ambience, food quality, or surrounding environments—play a critical role in expressing emotional experiences

[16], [17]. When analyzed in isolation, textual or visual data may provide incomplete or misleading interpretations of tourist sentiment. This limitation has motivated the integration of multiple data modalities to better reflect how tourists communicate their experiences in real-world digital environments.

These limitations have motivated the integration of multiple data modalities in sentiment analysis. Multimodal sentiment analysis combines textual and visual information to better reflect how tourists communicate their experiences in real-world digital environments. This integration is particularly important for understanding complex emotional expressions that emerge from the interaction between what tourists write and what they visually depict. Implicit sentiment analysis further complicates this task, as it aims to identify emotions that are not explicitly stated and are often embedded in sarcasm, irony, or contextual incongruity [18], [19], [20], [21], [22]. With the increasing availability of multimodal data on social media and review platforms, integrating textual and visual information has become a promising approach to uncovering these hidden emotional cues [23], [24], [25], [26].

Nevertheless, despite the growing availability of multimodal data, research in tourism sentiment analysis remains largely focused on explicit sentiment and unimodal approaches [27], [28]. Studies addressing implicit sentiment detection have primarily concentrated on product reviews or general social media content rather than tourism-specific contexts [29], [30], [31]. As a result, the potential of multimodal implicit sentiment analysis for understanding complex tourist experiences and supporting tourism development strategies has not been fully explored.

With the increasing use of social media and online review platforms, data comprising both textual and visual content have become increasingly abundant [32][33], reinforcing the need to integrate multiple modalities in sentiment analysis. Accordingly, this study adopts a Systematic Literature Review (SLR) approach aimed at examining trends, methodologies, datasets, and scientific novelties in multimodal sentiment analysis capable of detecting implicit sentiment, in order to support more targeted and data-driven tourism development strategies.

Therefore, this SLR seeks to address the following research questions:

1) RQ1: How can Artificial Intelligence–based models be developed to detect tourists' implicit sentiment by integrating textual and visual modalities?
2) RQ2: Which methods are most commonly employed to analyze implicit multimodal sentiment?
3) RQ3: What types of datasets are predominantly used in multimodal sentiment analysis?
4) RQ4: What forms of scientific novelty and contribution are offered by implicit multimodal sentiment analysis approaches in understanding and optimizing tourist experiences within the tourism sector?

Despite the rapid growth of sentiment analysis research in tourism, several critical gaps remain. First, the existing literature predominantly focuses on explicit sentiment analysis, which is insufficient for capturing implicit emotional expressions such as sarcasm, ambiguity, and cross-modal inconsistency that frequently appear in tourist reviews. Second, although multimodal data combining text and images are increasingly available on tourism platforms, their systematic use for implicit sentiment detection in tourism contexts remains limited, with most studies concentrating on general social media or product-review domains. Third, prior research shows a strong overreliance on general-purpose datasets (e.g., Twitter, Flickr, Weibo), resulting in limited contextual relevance for tourism-specific decision-making, particularly in developing countries such as Indonesia. To address these gaps, this study makes several key contributions. First, it provides a systematic synthesis of multimodal implicit sentiment analysis research in the tourism domain, highlighting dominant architectures, fusion strategies, and emerging methodological trends. Second, it explicitly positions implicit sentiment detection as a critical analytical dimension for understanding nuanced tourist experiences that cannot be captured through text-only or explicit sentiment approaches. Third, this review identifies strategic research opportunities for tourism-specific and Indonesian-context applications, offering a conceptual and methodological foundation for future model development, dataset construction, and the implementation of multimodal sentiment analysis in tourism decision-making systems.

This study provides an in-depth description of the literature search process, along with the criteria employed to assess the relevance of the reviewed documents. Subsequently, the study presents a synthesis of the most relevant studies in relation to the proposed

research questions. In the concluding section, a comprehensive discussion and the resulting conclusions derived from the analysis are presented.

2. METHODS

2.1. Data Sources and Query

The article search process began by querying online databases commonly used in Systematic Literature Reviews (SLR). Three databases were selected, namely Scopus, ScienceDirect, and ProQuest. These databases were selected because they provide broad coverage of high-impact journals in information systems, computer science, and tourism studies, ensuring access to peer-reviewed and methodologically rigorous publications relevant to multimodal sentiment analysis. The authors constructed specific search strings for each database. The search string applied in this process was as follows: ("Sentiment Analysis" OR "Multimodal Sentiment Analysis") AND ("tourism" OR "travel reviews" OR "tourist behavior") AND ("text and image" OR "text-image fusion" OR "text-image integration")

To ensure consistency, quality, and reproducibility, the search was restricted to articles published between 2020 and 2025, reflecting recent methodological developments in deep learning–based multimodal sentiment analysis. Only English-language publications were included, as English remains the dominant language for international scientific communication in artificial intelligence and tourism research. This restriction was applied to minimize the risk of misinterpretation and to ensure comparability across studies. In addition, the review focused on open-access and institutionally accessible full-text articles. This access criterion was adopted to ensure transparency, verifiability, and replicability of the review process, allowing all included studies to be independently examined and methodologically assessed. Articles without full-text availability were excluded due to limitations in evaluating experimental design, datasets, and technical contributions.

The identification and selection of literature followed a systematic workflow based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, as illustrated in Figure 1. The initial stage involved identifying articles from the three primary databases: Scopus, ScienceDirect, and ProQuest. In total, 227 articles

were retrieved, comprising 4 articles from Scopus, 120 from ScienceDirect, and 103 from ProQuest.

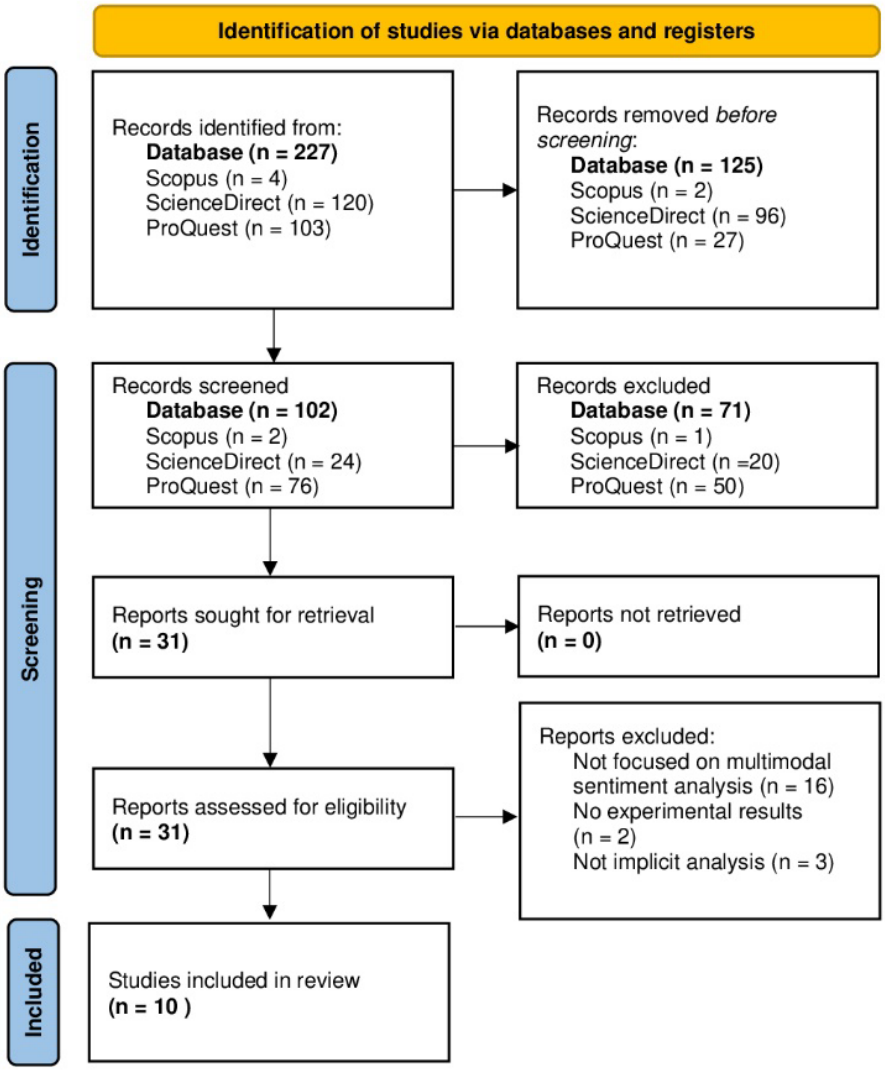

Figure 1. PRISMA flow diagram illustrating the systematic stages of the literature review process

Prior to the screening stage, 125 articles were removed due to irrelevance, including 2 from Scopus, 96 from ScienceDirect, and 27 from ProQuest. Consequently, 102 documents remained and were further screened to assess preliminary eligibility based on titles and abstracts. At this stage, 124 documents were excluded, consisting of 1 from Scopus, 20 from ScienceDirect, and 50 from ProQuest.

Subsequently, 31 articles were deemed to meet the initial eligibility criteria and were advanced to the full-text review stage. No documents were inaccessible or unavailable at this stage. However, following an in-depth assessment, 21 articles were excluded for failing to meet the inclusion criteria, namely: 16 articles did not specifically address multimodal sentiment analysis, 2 articles did not present experimental results or technical evaluations; and 3 articles did not focus on implicit sentiment detection. As a result, 10 articles satisfied all inclusion criteria and were further analyzed as part of the systematic literature review.

3. RESULTS AND DISCUSSION

Based on the systematic selection procedure using the PRISMA protocol outlined in the methodology section, this study identified ten core articles that met the inclusion criteria and were subjected to in-depth analysis. These articles comprise experimental studies published between 2020 and 2025, with a primary focus on the integration of textual and visual modalities in sentiment analysis. A summary of the selected literature is presented in the Table 1.

Table 1. Summary of Implicit Multimodal Sentiment Analysis SLR (2020-2025)

Author & Year	Method	Research Focus	Dataset	Result	Limitations
Yin & Chen (2023) [34]	TF-MMATI: RoBERTa-BiLSTM, ResNet50, Transformer Encoder, DAIMA & ADIMA multimodal attention.	Multimodal sentiment (text + image)	Twitter-15 & Twitter-17.	Accuracy reached 78.66% and 72.67%.	Highly dependent on complex architecture; only uses Twitter datasets.
Hu et al. (2025) [35]	QwenLM multimodal model (image–text), RoBERTa untuk sentiment, VR-based emotional visualization,	Multimodal sentiment (text + image)	UGC: 9,207 texts + 29,742 images from Dianping (Qianmen Street, Beijing).	The model is capable of detecting multidimensional emotions, generating VR visualizations that increase user attention and engagement	Focus on the cultural domain; do not use standard benchmarking datasets.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Author & Year	Method	Research Focus	Dataset	Result	Limitations
	eye-tracking analysis.			according to eye-tracking results.	
Boumhidi, Benlahbib & Nfaoui (2023) [36]	Sentiment XLNet, deteksi sarkasme (text + image + emoji), ResNet50, popularity score, demand score, news-influence score.	Multimodal data (text, images, emojis, interactions, news).	Twitter	The model approximates the user's ground truth rating. Sentiment accuracy improves after incorporating multimodal sarcasm detection.	Small dataset (250 tweets per asset); highly specialized domain (cryptocurrency); sarcasm detection is still ineffective in text alone.
Yang & Chen (2023) [37]	Improved PageRank (text sentiment), ECA-ResNeXt50 (image sentiment), cross-modal fusion (KL divergence + dynamic weighting).	Multimodal sentiment (text + image)	Flickr & Twitter	The accuracy reached 88.20%.	Relatively small dataset.
Yuehua Han & Zhifen Xu, (2024) [38]	Multi-Granular View Dynamic Fusion Model (MVDFM): BERT (teks), ResNet50 (gambar), Dynamic Gated Self-Attention, Triple-View Factorized High-Order Pooling	Multimodal sentiment (text + image)	Twitter-2015 & Twitter-2017	Accuracy: 78.78% (Twitter-2015), 73.89% (Twitter-2017). F1: 74.48% & 72.47%.	Does not specifically test implicit sentiment; dataset is limited to two Twitter corpora.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Author & Year	Method	Research Focus	Dataset	Result	Limitations
Silva et al., (2024) [39]	Transfer learning pada image classifier (VGG16, VGG19, ResNet50, EfficientNetB0); CNN	Developing a multimodal sentiment classification framework based on scene context (background, objects, and facial expressions).	Flickr; Instagram; Sentiment140 (text); FI (Face Images); B-T4SA; Emotion6.	Multimodal accuracy reached 78.84%, higher than unimodal.	Just focus on face-related visual data.
Zhizhong Liu (2023) [40]	CNN, BiLSTM/LSTM, DenseNet/VGG.	sentiment classification by combining visual and text features.	General domain multimodal dataset (image–text combination).	The combination of DL models can strengthen multimodal representation.	Complex architecture but does not show strong generalization; small-scale dataset.
Mu, Chen, Li, Dai & Dai (2025) [41]	SECIF Model BERT (teks), ResNet101, GMHA, ICN (Improved Capsule Network), Cross-Modal Interaction, KL Divergence + Cross-Entropy Loss	Sentimen multimodal (teks + gambar)	11.165 post Sina Weibo (teks + gambar); MVSA-Single dan MVSA-Multiple (benchmark).	Akurasi pada dataset Weibo: 90.86%; pada MVSA-Single: akurasi 79.10% (+6.42% dari baseline).	• Gagal memahami sarkasme & implicit sentiment secara mendalam.
Ling Jixian et al (2022) [42]	Embedding, BiLSTM, Attention, k-max pooling, CNN, XGBoost.	Multimodal information classification for disaster management.	Weibo, WeChat, Twitter, Facebook (text + images).	Accuracy: >85% (CN), >95% (EN).	Does not use Transformers; multimodal fusion is still simple; not geared towards implicit sentiment.

Author & Year	Method	Research Focus	Dataset	Result	Limitations
Shujun Wei & Song Song (2022) [13]	CNN, MTCNN, VGG16	Sentimen multimodal (teks + gambar)	C-Tourism (China), T-Tourism (Twitter).	The combination of facial expressions and text gives the best performance.	Does not handle implicit sentiment & sarcasm; face detection fails on tilted poses/occlusion.

3.1. Text- and Image-Based Implicit Sentiment Detection Models for Tourists (RQ1)

From an architectural perspective, the development of models for detecting implicit sentiment follows an integrated framework that involves parallel feature extraction from textual and visual modalities. Textual information is typically processed using Transformer-based language models such as BERT or RoBERTa to capture contextual semantics, while visual content is analyzed using Convolutional Neural Networks (CNNs) or Vision Transformers (ViT) to extract high-level visual representations. The extracted features are subsequently combined within a multimodal fusion layer—often implemented through attention mechanisms or cross-modal interaction modules—to model the relationships and potential inconsistencies between text and images. Yin and Chen demonstrate that attention-driven interaction modules significantly improve performance by modeling fine-grained dependencies between textual and visual features, particularly in cases of semantic incongruity that signal implicit sentiment [22].

In tourism-related applications, Wei and Song show that combining facial expressions in images with textual reviews improves sentiment classification accuracy compared to unimodal approaches, indicating that visual affective cues play a complementary role in interpreting tourist emotions [13]. Although their study primarily targets explicit sentiment, the architectural pattern aligns with more recent implicit sentiment frameworks that emphasize contradiction and alignment between modalities as a core signal of hidden emotion [31].

The workflow begins with the use of Transformer-based models (such as BERT or RoBERTa) to capture complex linguistic contexts, while visual features are extracted using Convolutional Neural Networks (CNNs) such as ResNet or Vision Transformers (ViT).

The core component of this architecture lies in the fusion layer, which is responsible for aligning and modeling the interactions between the two modalities. This integration process enables the model to identify discrepancies between what is written and what is visually depicted, which are key indicators of implicit sentiment, including sarcasm and ambiguity in tourist reviews.

Conceptually, this pipeline enables the detection of implicit sentiment through semantic mismatch. For instance, a neutral or mildly positive textual statement accompanied by a negative visual scene (e.g., poor room conditions or overcrowded attractions) can be interpreted as sarcastic or implicitly negative. This mechanism is consistent with findings by Liu et al., who demonstrate that multimodal sarcasm detection benefits substantially from modeling cross-modal inconsistencies rather than relying on textual polarity alone [31]. Therefore, the reviewed literature supports the argument that implicit tourist sentiment is best inferred through attention-based or interaction-driven fusion mechanisms that explicitly model contradiction and alignment between text and images.

The core function of the fusion layer in such models is therefore not merely feature aggregation, but cross-modal alignment and discrepancy detection, which are essential for identifying implicit sentiment such as sarcasm, irony, or understated dissatisfaction in tourism reviews. Attention-based fusion mechanisms, in particular, allow the model to assign greater importance to the modality that provides stronger emotional signals, depending on the context of the review.

It is important to clarify that the model illustrated in Figure 2 does not represent a newly proposed or empirically trained model. Instead, it is a conceptual framework synthesized from the systematic analysis of the reviewed studies. The architecture reflects recurring design patterns identified across the ten core articles, including parallel modality encoding, attention-driven fusion, and implicit sentiment inference based on cross-modal inconsistencies. As such, the model serves as an integrative abstraction that consolidates dominant methodological approaches rather than introducing a novel algorithmic contribution.

By synthesizing these recurring architectural components, this review highlights how existing multimodal sentiment analysis techniques can be systematically adapted to

tourism-specific applications, particularly for detecting implicit emotional expressions embedded in real-world tourist reviews.

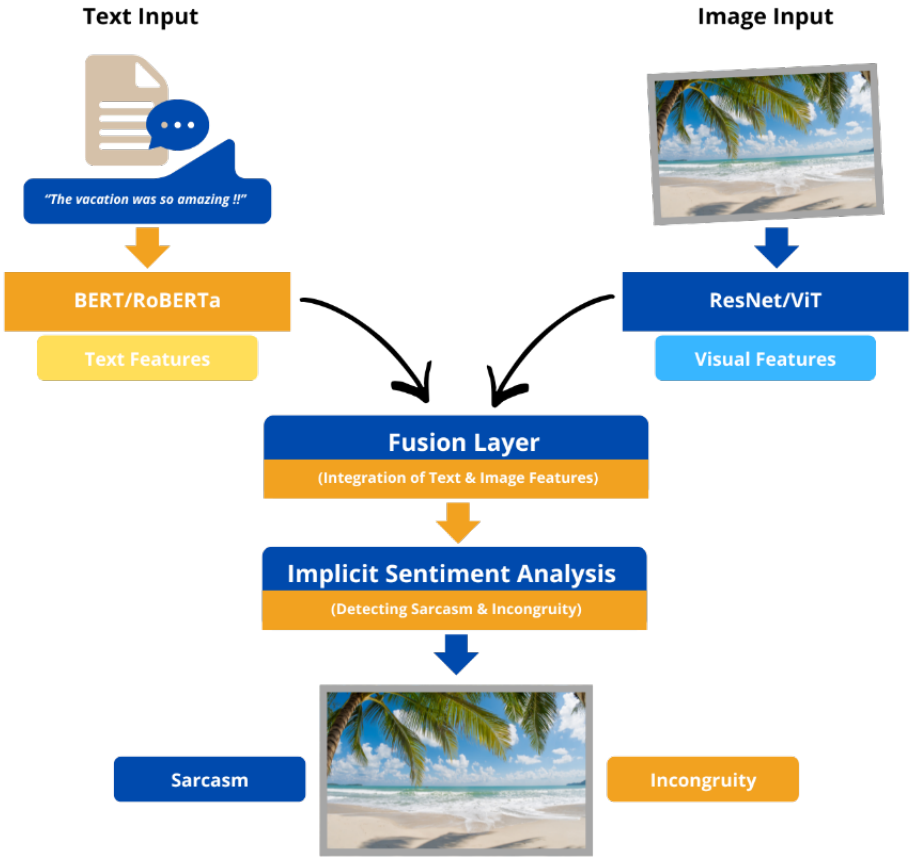

Figure 2. Text & Image-Based Implicit Tourist Sentiment Detection Model

Figure 2 illustrates a Text- and Image-Based Implicit Sentiment Detection Model for tourists, which integrates textual and visual analysis to identify implicit sentiment in tourist reviews. The process begins with textual input analyzed using BERT/RoBERTa models and visual input processed through ResNet/ViT to extract textual and visual features, respectively. These features are subsequently combined within a fusion layer to detect inconsistencies between textual and visual content, such as sarcasm or ambiguity. The final output is an implicit sentiment analysis that determines whether a review conveys sarcasm or cross-modal inconsistency.

3.2. Most Commonly Used Methods for Implicit Multimodal Sentiment Analysis (RQ2)

Based on the review of the selected literature, a clear trend emerges in the choice of algorithms for multimodal sentiment analysis, as illustrated in Figure 2. Attention-based

fusion methods constitute the most dominant approach (60%), owing to their effectiveness in assigning appropriate weights to the most relevant features from both textual and visual modalities. Furthermore, the emergence of Contrastive Learning techniques (20%) represents a recent innovation aimed at enhancing feature representations to become more discriminative in distinguishing hidden emotions. Meanwhile, CNN/RNN/Hybrid approaches (20%) continue to be employed as foundational methods for conventional feature extraction. These findings indicate that current research has shifted from simple fusion strategies toward more sophisticated attention mechanisms, which significantly improve detection accuracy when dealing with ambiguous datasets.

The systematic review indicates that attention-based fusion mechanisms dominate current multimodal sentiment research. Studies such as Yin and Chen report consistent performance gains over early-fusion (feature concatenation) and late-fusion (decision-level integration) approaches. These models dynamically weight modality-specific features, allowing the system to prioritize the modality that carries stronger emotional signals in a given context [22].

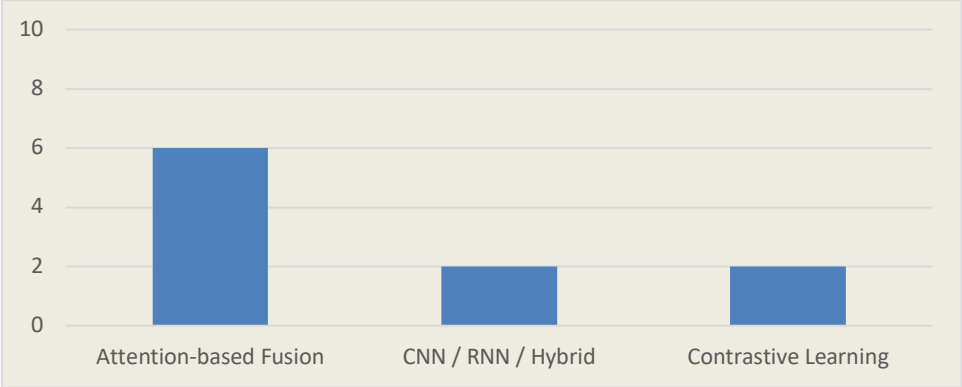

Figure 3. Distribution of Method Categories in Implicit Multimodal Sentiment Analysis

Based on Figure 3 the systematic review of the selected literature, attention-based fusion mechanisms emerge as the most dominant and effective methodological approach for implicit multimodal sentiment analysis. Compared to simpler fusion strategies—such as early fusion through feature concatenation or late fusion via independent modality predictions—attention mechanisms provide a more nuanced and context-aware integration of textual and visual information.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

The primary reason attention-based fusion outperforms simpler fusion lies in its ability to dynamically weight modality-specific features according to their contextual relevance. In tourism reviews, emotional cues are often unevenly distributed across modalities. For example, a review may contain neutral or ambiguous textual expressions accompanied by highly expressive visual content, or vice versa. Simple fusion methods treat all features equally, which can dilute critical emotional signals. In contrast, attention mechanisms selectively emphasize the modality—or specific features within a modality—that contributes more strongly to sentiment interpretation, thereby improving the detection of implicit emotions such as sarcasm or understatement.

Furthermore, attention-based models enable fine-grained cross-modal interaction, allowing the system to capture semantic alignment or contradiction between textual and visual representations. This capability is particularly important for implicit sentiment detection, where emotional meaning often emerges from cross-modal inconsistency rather than from explicit sentiment markers. Studies reviewed in this SLR consistently report higher classification accuracy when attention-driven fusion is employed, especially in datasets containing ambiguous or implicitly expressed emotions.

Despite these advantages, attention mechanisms introduce additional computational complexity compared to simpler fusion strategies. Attention-based fusion typically involves matrix operations whose computational cost increases with feature dimensionality and sequence length, leading to higher memory usage and longer training times. This complexity becomes more pronounced when Transformer-based encoders are used for both text and images, particularly in large-scale tourism datasets with high-resolution visual content.

However, the reviewed studies suggest that the performance gains achieved through attention-based fusion often justify the increased computational cost, especially in analytical scenarios where accuracy in detecting nuanced tourist sentiment is prioritized over real-time processing. For practical deployment in tourism decision-support systems, several studies recommend employing optimization strategies such as dimensionality reduction, lightweight attention variants, or pre-trained encoders with frozen layers to balance performance and efficiency.

The findings indicate that attention-based fusion represents a methodological trade-off between interpretive accuracy and computational efficiency. While simpler fusion methods remain suitable for resource-constrained environments, attention mechanisms are better suited for applications requiring robust interpretation of implicit tourist sentiment in complex multimodal data.

3.3. Most Commonly Used Datasets for Multimodal Sentiment Analysis (RQ3)

Most prior studies employ datasets that integrate both textual and visual modalities, such as CMU-MOSEI, the Twitter Multimodal Dataset, TripAdvisor, and Ctrip. These datasets are commonly selected because they reflect user expression patterns on social media and review platforms, where opinions are accompanied by destination images and informal language containing implicit meanings. The datasets utilized by Yin and Chen (n.d.) demonstrate particularly high relevance, as they encompass tourist reviews that convey emotions through both visual content and experiential narratives [23].

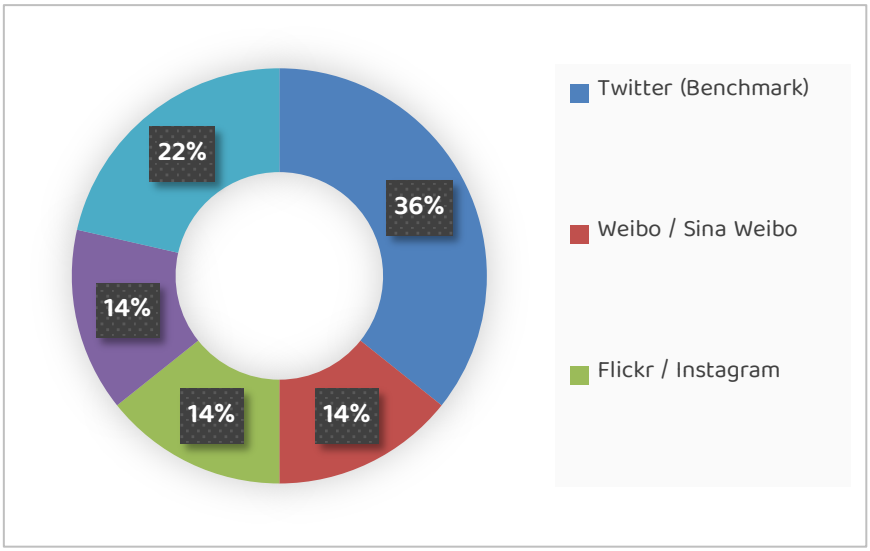

Figure 4. Frequency of Use of Dataset Types

Figure 4 indicates that data sources in multimodal sentiment analysis are still predominantly drawn from general social media platforms (such as Twitter and Weibo) as well as benchmark datasets like MVSA, primarily due to their abundant data availability. Although tourism-specific platforms such as TripAdvisor and Ctrip have begun to be utilized, their representation remains relatively limited compared to general-purpose datasets. The dominance of generic social media datasets highlights a significant

research gap, emphasizing the need to develop tourism-domain-specific datasets that are more contextualized and capable of capturing the unique characteristics and nuanced emotional expressions of tourists with greater accuracy.

Although tourism-specific multimodal datasets are increasingly recognized as essential for accurately capturing tourist sentiment, their collection and use present several methodological and practical challenges. One major difficulty lies in data heterogeneity, as tourism platforms host diverse forms of user-generated content, including informal text, multilingual expressions, emojis, and images with varying visual quality and semantic relevance. This heterogeneity complicates annotation processes and reduces consistency across datasets.

Another challenge concerns domain specificity and contextual richness. Unlike general social media datasets, tourism reviews often contain location-dependent references, culturally embedded expressions, and experiential narratives that require domain knowledge for accurate sentiment interpretation. Annotating implicit sentiment in such data is particularly demanding, as sarcasm or dissatisfaction may be expressed subtly through cross-modal inconsistencies rather than explicit sentiment markers. Consequently, tourism-specific multimodal datasets are costly and time-consuming to curate, leading many studies to rely on readily available general-purpose datasets such as Twitter or Weibo.

In addition to technical challenges, the development of tourism-specific multimodal datasets raises important ethical and privacy considerations. User-generated tourism data frequently include identifiable information, such as faces in images, geolocation metadata, or personal travel details. The collection and reuse of such data must therefore comply with data protection regulations and platform usage policies. Failure to anonymize visual content or remove sensitive metadata can expose users to privacy risks, particularly when datasets are redistributed for research purposes.

Ethical concerns also extend to consent and representational bias. Tourists rarely provide explicit consent for their reviews and images to be used in machine learning research, which raises questions about fair data usage, especially when datasets are collected at scale. Moreover, tourism datasets often overrepresent certain destinations,

demographics, or travel behaviors, potentially biasing sentiment models and limiting their generalizability across regions and cultural contexts.

Several authors highlight the methodological challenges of constructing tourism-domain multimodal datasets. Chen et al emphasize that location-specific references, cultural expressions, and heterogeneous visual content complicate annotation consistency [34]. Nip & Berthelier Ethical and privacy concerns also arise, as tourism images frequently contain identifiable individuals or geolocation metadata, requiring careful anonymization and compliance with platform usage policies [32].

Addressing these challenges requires a more responsible and transparent approach to dataset construction. Several studies reviewed in this SLR emphasize the importance of ethical data sourcing, including the use of publicly available data in accordance with platform policies, anonymization of personal identifiers, and clear documentation of dataset limitations. Future research is encouraged to develop tourism-domain-specific multimodal datasets that balance methodological rigor with ethical accountability, particularly for applications in policy-making and destination management.

3.4. Scientific Novelty of Implicit Multimodal Sentiment Analysis (RQ4)

Implicit multimodal sentiment analysis introduces a new paradigm in the analysis of tourist opinions by integrating textual and visual information. Unlike traditional text-based approaches that focus solely on explicit expressions, multimodal approaches are capable of interpreting implicit emotions by exploiting the relationships between text and images. Several studies, such as those by Liu et al., demonstrate that the integration of Cross-Modal Attention mechanisms and Contrastive Learning enables models to capture deeper emotional contexts, including sarcasm, ambiguity, and visual meanings that are not explicitly conveyed through textual content [25].

The scientific novelties introduced by implicit multimodal sentiment analysis can be summarized as follows:

1) Expansion of traditional sentiment analysis boundaries, implicit multimodal approaches extend conventional sentiment analysis by combining textual and visual data to capture hidden emotional meanings that cannot be detected through text-only analysis.

2) Deep integration of NLP and Computer Vision, the adoption of Transformer-based models, such as BERT for textual data and Vision Transformers (ViT) for visual data, facilitates a deeper integration of Natural Language Processing (NLP) and Computer Vision within the tourism domain.
3) Innovative learning mechanisms for implicit emotion detection, the application of Cross-Modal Attention and Contrastive Learning represents a major innovation, enabling models to understand semantic and emotional alignment between textual and visual modalities, particularly in cases of implicit sentiment such as sarcasm or contextual ambiguity.

Most prior sentiment analysis studies in tourism have focused on explicit sentiment detection, which relies on overt emotional markers in textual data, such as positive or negative adjectives, sentiment lexicons, or polarity scores. These approaches assume a direct correspondence between linguistic expressions and emotional states, making them effective for clearly stated opinions but limited in handling ambiguity, understatement, or sarcasm. As a result, explicit sentiment models often misclassify tourist reviews in which dissatisfaction is implied rather than directly expressed, particularly when textual cues are neutral but visual content conveys negative experiential signals.

In contrast, implicit multimodal sentiment analysis introduces a fundamental methodological shift. Rather than relying solely on explicit textual polarity, implicit approaches focus on uncovering hidden emotional meanings that emerge from cross-modal interactions between text and images. The novelty of this paradigm lies in its ability to detect sentiment through semantic incongruity, contextual mismatch, and latent emotional cues, such as a positive textual description accompanied by an image depicting poor service quality or degraded destination conditions. This capability directly addresses a critical limitation of explicit sentiment analysis, which typically treats text as the primary or sole carrier of emotional meaning.

From a theoretical perspective, implicit multimodal sentiment analysis extends the conceptual boundaries of sentiment research by treating emotion as an emergent property of multimodal interaction rather than a direct function of lexical polarity. This aligns with earlier work on implicit sentiment and commonsense reasoning in text [18],

[19] and expands it into the visual domain, where emotional meaning is often conveyed through aesthetic and contextual cues.

Another key novelty is the integration of advanced fusion mechanisms, particularly attention-based and contrastive learning approaches, which enable models to dynamically prioritize the most informative modality. Unlike explicit sentiment models that aggregate features in a static or linear manner, implicit multimodal models capture fine-grained alignment and contradiction between modalities. This allows for a more realistic representation of how tourists communicate experiences in digital environments, where emotions are often conveyed indirectly and visually.

3.5. Conceptual Framework of Contributions

Based on the synthesis of the reviewed studies, this SLR proposes a conceptual framework summarizing the contributions of implicit multimodal sentiment analysis to tourism research. The framework consists of four interrelated layers:

1) Data Layer

   This layer comprises multimodal tourism data sourced from online platforms, including textual reviews and accompanying images. Unlike traditional approaches that prioritize text-only data, this layer emphasizes the complementary role of visual content in expressing tourist emotions.

2) Analytical Layer

   At this stage, textual and visual data are processed through deep learning encoders (e.g., Transformer-based language models and CNN/ViT-based vision models). Cross-modal fusion mechanisms—particularly attention-based and contrastive learning methods—enable interaction modeling between modalities to detect implicit sentiment cues.

3) Interpretative Layer

   This layer focuses on identifying implicit emotional signals, such as sarcasm, irony, and cross-modal inconsistency. By interpreting discrepancies between textual intent and visual evidence, the framework extends beyond polarity classification to capture nuanced tourist experiences.

4) Application Layer

   The final layer translates analytical outputs into tourism decision-support applications, including early warning systems for tourist dissatisfaction,

destination quality assessment, and evidence-based policy formulation. This layer highlights the practical relevance of implicit multimodal sentiment analysis for sustainable tourism development.

By positioning implicit multimodal sentiment analysis within this conceptual framework, this study clearly distinguishes its contribution from prior explicit sentiment approaches. Rather than replacing traditional sentiment analysis, implicit multimodal methods extend its analytical boundaries, offering a more comprehensive and context-sensitive understanding of tourist emotions. This distinction underscores the scientific novelty of the reviewed approaches and their strategic importance for future tourism analytics research.

The findings of this Systematic Literature Review indicate that implicit multimodal sentiment analysis represents an emerging direction for comprehensively understanding tourist behavior and experiences. An analysis of the ten core studies reveals that most recent research emphasizes the integration of text and image modalities to detect emotions that are not explicitly expressed.

The majority of prior studies have focused on explicit sentiment, namely opinions that are clearly and directly stated in text. While some research has addressed implicit sentiment, it has largely been confined to product reviews or e-commerce domains rather than tourism. This study emphasizes cross-modal feature extraction to detect implicit sentiment, thereby capturing the nuanced emotional expressions of tourists that are often not fully represented by textual descriptions alone. Moreover, this research is among the first to apply an implicit multimodal sentiment analysis framework within the context of Indonesian-language tourist reviews and Indonesia's digital social data ecosystem.

Despite the significant potential of implicit multimodal sentiment analysis, the review of the ten core articles identifies several technical challenges that require further investigation:

1) Cultural and contextual sensitivity limitations, although attention-based fusion models demonstrate high accuracy, many existing architectures still face substantial challenges in detecting sarcasm and culturally dependent implicit

meanings. Tourist sentiment is often embedded in linguistic ambiguity that can only be interpreted through a deep semantic understanding of simultaneous text–image interactions.

2) Overreliance on general-purpose datasets, most existing studies continue to rely heavily on general datasets such as Twitter, Flickr, or Weibo. This reliance creates a research gap, as linguistic and expressive patterns on tourism-specific platforms (e.g., TripAdvisor or destination-focused reviews) differ significantly from those found on general social media platforms.
3) Limited exploitation of visual ambience and aesthetics, current studies tend to utilize image data primarily for object classification or facial expression recognition. However, these models remain insufficient in extracting "ambience" or aesthetic attributes of destinations, which often serve as key implicit indicators of tourist satisfaction.

The findings of this Systematic Literature Review provide a strategic foundation for data-driven tourism development through the integration of textual and visual information. Such integration enables the mapping of destination quality based on tourists' visual perceptions beyond explicit textual evaluations. Furthermore, the detection of sarcastic and implicit sentiment functions as an early warning system for destination managers, allowing them to respond proactive

3.6. Discussion

This Systematic Literature Review synthesizes evidence from ten core studies to demonstrate that implicit multimodal sentiment analysis represents a critical methodological advancement for understanding nuanced tourist emotions. Across the reviewed literature, a consistent pattern emerges in which Transformer-based textual encoders and CNN/ViT-based visual models are integrated through attention-driven or contrastive fusion mechanisms to capture cross-modal alignment and semantic incongruity. This synthesis indicates that sentiment in tourism contexts should be conceptualized not as a direct function of lexical polarity, but as an emergent property of interaction between what tourists express verbally and what they visually document. By consolidating these findings, this study positions implicit multimodal analysis as a bridge between traditional sentiment analytics and experiential interpretation, enabling a more holistic representation of tourist perceptions and destination quality.

Despite these methodological advances, the synthesis reveals structural limitations that constrain theoretical generalization and practical deployment. First, the dominance of general-purpose datasets (e.g., Twitter, Weibo, MVSA) limits the contextual validity of models for tourism-specific applications, where emotions are shaped by cultural norms, place-based references, and experiential narratives. Second, current models prioritize object-level or facial-expression features, while underexploring higher-order visual attributes such as ambience, spatial aesthetics, and environmental context, which are often central to implicit tourist satisfaction or dissatisfaction. Third, although attention-based fusion improves performance, it also increases computational complexity, creating a gap between experimental success and real-time implementation in tourism decision-support systems.

Future research should therefore move in three strategic directions. Methodologically, there is a need to develop lightweight and interpretable fusion architectures that retain the representational power of attention mechanisms while reducing computational overhead. From a data perspective, constructing tourism-domain-specific multimodal datasets—particularly in underrepresented regions such as Indonesia—should be prioritized, with annotation schemes explicitly designed to capture sarcasm, cultural nuance, and cross-modal inconsistency. Ethically grounded data governance frameworks must also be integrated to address privacy, consent, and representational bias in the use of user-generated visual content.

Theoretically, future studies should extend implicit sentiment models toward experiential intelligence, incorporating commonsense reasoning, spatial semantics, and affective computing to better model how tourists perceive and emotionally evaluate destinations. Practically, longitudinal and real-time deployments of multimodal systems could be explored to support early warning mechanisms for destination management, adaptive marketing strategies, and sustainable tourism policy formulation. By advancing along these directions, future research can transform implicit multimodal sentiment analysis from a predominantly technical innovation into a robust analytical foundation for evidence-based and context-sensitive tourism development.

4. CONCLUSION

This study synthesizes recent advances in implicit multimodal sentiment analysis and demonstrates their value for capturing nuanced tourist emotions that extend beyond explicit textual polarity. By integrating textual and visual modalities through attention-based and contrastive fusion approaches, the reviewed literature highlights how cross-modal alignment enhances the interpretability and predictive accuracy of sentiment models in tourism contexts. However, the findings also reveal persistent gaps, particularly the limited availability of tourism-specific multimodal datasets and the high computational complexity of state-of-the-art architectures, which constrain real-world adoption. To translate these methodological insights into practice, policymakers and destination managers should support the development of ethically governed, region-specific multimodal data infrastructures and invest in scalable analytical systems that can inform evidence-based tourism planning and sustainable destination management.

The primary contribution of this study lies in identifying the potential application of multimodal sentiment analysis within the Indonesian tourism context, which has remained largely underexplored. The review demonstrates that leveraging multimodal digital data provides a strategic foundation for stakeholders to assess destination quality based on authentic visual perceptions, develop early warning systems for tourist complaints, and optimize marketing strategies through content personalization.

Nevertheless, this study also identifies several challenges, including the heavy reliance on general-purpose social media datasets and the limited capability of existing models to deeply extract aesthetic or “ambience-related” aspects of destinations. Therefore, future research is expected to focus on the development of tourism-domain-specific datasets for Indonesia, as well as the enhancement of model architectures capable of understanding cross-modal semantic interactions in a more contextualized manner, in order to support the sustainability of national tourism development strategies.

ACKNOWLEDGMENT

The authors gratefully acknowledge the academic support, insightful guidance, and constructive feedback provided throughout the preparation of this study. Appreciation

is also extended to Universitas Diponegoro for its institutional support, which contributed to strengthening the conceptual development, methodological rigor, and overall quality of this article.

REFERENCES

[1] K. K. Anele and C. C. Sam-Otuonye, "Sustainable Tourism: Evidence from Lake Toba in North Sumatra, Indonesia," *ASEAN J. Hosp. Tour.*, vol. 19, no. 1, pp. 52–62, 2021, doi: 10.5614/ajht.2021.19.1.05.

[2] I. W. R. Junaedi *et al*, "Investment Opportunities and Tourism Business Development in The Village of Siallagan Village, Batak Adat Village," *Int. Bus. Account. Res. J.*, vol. 7, no. 2, pp. 253–268, 2023.

[3] Y. Wang and H. Bai, "The impact and regional heterogeneity analysis of tourism development on urban-rural income gap," *Econ. Anal. Policy*, vol. 80, no. October, pp. 1539–1548, 2023, doi: 10.1016/j.eap.2023.10.031.

[4] D. P. Ramadhani, A. Alamsyah, M. Y. Febrianta, and L. Z. A. Damayanti, "Exploring Tourists' Behavioral Patterns in Bali's Top-Rated Destinations: Perception and Mobility," *J. Theor. Appl. Electron. Commer. Res.*, vol. 19, no. 2, pp. 743–773, 2024, doi: 10.3390/jtaer19020040.

[5] H. Tao, D. Yang, H. Zhou, and Y. Jian, "Travel experiences documented in online reviews influence travelers' travel intentions," *Sci. Rep.*, vol. 15, no. 1, pp. 1–13, 2025, doi: 10.1038/s41598-025-25971-9.

[6] E. A. Mensah, D. N. A. Odame, I. Ankrah, T. Obuobisa-Darko, and R. E. Hinson, "From reviews to reflections: Understanding tourist sentiments and satisfaction in African destinations through user-generated content," *Ann. Tour. Res. Empir. Insights*, vol. 6, no. 1, 2025, doi: 10.1016/j.annale.2025.100174.

[7] J. G. Martínez-Navalón, V. Gelashvili, and A. Gómez-Ortega, "Evaluation of User Satisfaction and Trust of Review Platforms: Analysis of the Impact of Privacy and E-WOM in the Case of TripAdvisor," *Front. Psychol.*, vol. 12, no. September, pp. 1–12, 2021, doi: 10.3389/fpsyg.2021.750527.

[8] W. Kim, S. B. Kim, and E. Park, "Mapping tourists' destination (Dis)satisfaction attributes with user-generated content," *Sustain.*, vol. 13, no. 22, pp. 1–16, 2021, doi: 10.3390/su132212650.

[9] I. Nawawi, K. F. Ilmawan, M. R. Maarif, and M. Syafrudin, "Exploring Tourist Experience through Online Reviews Using Aspect-Based Sentiment Analysis with Zero-Shot Learning for Hospitality Service Enhancement," *Inf.*, vol. 15, no. 8, 2024, doi: 10.3390/info15080499.

[10] E. A. Mensah, D. N. A. Odame, I. Ankrah, T. Obuobisa-Darko, and R. E. Hinson, "From reviews to reflections: Understanding tourist sentiments and satisfaction in African destinations through user-generated content," *Ann. Tour. Res. Empir. Insights*, vol. 6, no. 1, p. 100174, 2025, doi: 10.1016/j.annale.2025.100174.

[11] A. Budhi and I. G. A. G. Witarsana, "Pengaruh Tripadvisor Electronic Word Of Mouth Terhadap Online Booking Decision Tamu Domestik Di Bali," *J. Kepariwisataan Destin. Hosp. dan Perjalanan*, vol. 6, no. 2, pp. 203–218, 2022, doi: 10.34013/jk.v6i2.414.

[12] A. N. Candrea *et al.*, "How Do Visitors to Mountain Museums Think? A Cross-Country Perspective on the Sentiments Decoded from TripAdvisor Reviews," *Electron.*, vol. 14, no. 8, 2025, doi: 10.3390/electronics14081637.

[13] S. Wei and S. Song, "Sentiment Classification of Tourism Reviews Based on Visual and Textual Multifeature Fusion," *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022, doi: 10.1155/2022/9940817.

[14] D. Erdoğan *et al.*, "Developing a Deep Learning-Based Sentiment Analysis System of Hotel Customer Reviews for Sustainable Tourism," *Sustain.*, vol. 17, no. 13, 2025, doi: 10.3390/su17135756.

[15] D. Ariyus, D. Manongga, and I. Sembiring, "Enhancing Sentiment Analysis of Indonesian Tourism Video Content Commentary on TikTok: A FastText and Bi-LSTM Approach," *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 6, pp. 18020–18028, 2024, doi: 10.48084/etasr.8859.

[16] S. Wei and S. Song, "Sentiment Classification of Tourism Reviews Based on Visual and Textual Multifeature Fusion," *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022, doi: 10.1155/2022/9940817.

[17] Kania Alma Tiara, A. Sanjaya, D. M. Sabilla, and Indriana, "Analisis Sentimen Destinasi Wisata Saung Angklung Udjo," *Altasia J. Pariwisata Indones.*, vol. 6, no. 2, pp. 144–155, 2024, doi: 10.37253/altasia.v6i2.9278.

[18] A. Balahur, J. M. Hermida, and A. Montoyo, "Detecting Implicit Expressions of Sentiment in Text Based on Commonsense Knowledge," *Proc. Annu. Meet. Assoc. Comput. Linguist.*, pp. 53–60, 2011.

[19] D. Zhou, J. Wang, L. Zhang, and Y. He, "Implicit Sentiment Analysis with Event-Centered Text Representation," *EMNLP 2021 - 2021 Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 6884–6893, 2021, doi: 10.18653/v1/2021.emnlp-main.551.

[20] A. Joshi, P. Bhattacharyya, and M. J. Carman, "Research Question Formation Yun," vol. 0, no. 0, 2017.

[21] A. Dataset and T. A. Dataset, "Edinburgh Research Explorer From Arabic Sentiment Analysis to Sarcasm Detection : The From Arabic Sentiment Analysis to Sarcasm Detection :," 2020.

[22] X. Wang, X. Li, Y. Yin, and Y. Li, "Implicit aspect-based generative model for sentiment analysis based on prompt learning," *2024 5th Int. Conf. Big Data Artif. Intell. Softw. Eng. ICBASE 2024*, pp. 94–97, 2024, doi: 10.1109/ICBASE63199.2024.10762313.

[23] M. Chen, K. Ubul, X. Xu, A. Aysa, and M. Muhammat, "Connecting Text Classification with Image Classification: A New Preprocessing Method for Implicit Sentiment Text Classification," *Sensors (Basel)*, vol. 22, no. 5, 2022, doi: 10.3390/s22051899.

[24] M. Devani, D. H. Padheriya, V. Jadeja, D. J. Jani, and A. Patel, "Multimodal Sentiment Analysis on Product Review Text and Image Using Machine Learning," *African J. Biomed. Res.*, vol. 27, no. 4, 2024, doi: 10.53555/ajbr.v27i4s.6119.

[25] K. Zhang, Y. Geng, J. Zhao, J. Liu, and W. Li, "Sentiment analysis of social media via multimodal feature fusion," *Symmetry (Basel)*, vol. 12, no. 12, pp. 1–14, 2020, doi: 10.3390/sym12122010.

[26] F. Amalia, U. G. Mada, and K. Kunci, "Multimodalitas dalam unggahan di Twitter yang dianggap mengandung pelecehan seksual," vol. 6, pp. 781–794, 2023.

[27] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 4, p. 102048, 2024, doi: 10.1016/j.jksuci.2024.102048.

[28] J. Chen, J. Cong, M. Li, Y. Sun, and J. Zhang, "T-ECBM: a deep learning-based text-image multimodal model for tourist attraction recommendation," *Sci. Rep.*, vol. 15, no. 1, pp. 1–16, 2025, doi: 10.1038/s41598-025-25630-z.

[29] P. Chen and L. Fu, "Enhancing Multimodal Tourism Review Sentiment Analysis Through Advanced Feature Association Techniques," *Int. J. Inf. Syst. Serv. Sect.*, vol. 15, no. 1, pp. 1–21, 2024, doi: 10.4018/IJISSS.349564.

[30] Z. Liu, T. Yang, W. Chen, J. Chen, Q. Li, and J. Zhang, "Sentiment analysis of social media comments based on multimodal attention fusion network," *Appl. Soft Comput.*, vol. 164, no. January, p. 112011, 2024, doi: 10.1016/j.asoc.2024.112011.

[31] Y. Liu, Z. Zheng, B. Zhou, J. Ma, L. Sun, and R. Xia, "Multimodal Sarcasm Detection Based on Multimodal Sentiment Co-training," *Proc. - 2022 IEEE SmartWorld, Ubiquitous Intell. Comput. Auton. Trust. Veh. Scalable Comput. Commun. Digit. Twin, Priv. Comput. Metaverse, Autonomous & Trusted Vehicles, 2022*, pp. 508–515, 2022.

[32] J. Y. M. Nip and B. Berthelier, "Social Media Sentiment Analysis," *Encyclopedia*, vol. 4, no. 4, pp. 1590–1598, 2024, doi: 10.3390/encyclopedia4040104.

[33] Y. Chen *et al.*, "Mining Social Media Data to Capture Urban Park Visitors' Perception of Cultural Ecosystem Services and Landscape Factors," *Forests*, vol. 15, no. 1, 2024, doi: 10.3390/f15010213.

[34] X. Xiao *et al.*, "Collaborative fine-grained interaction learning for image–text sentiment analysis," *Knowledge-Based Syst.*, vol. 279, p. 110951, 2023, doi: 10.1016/j.knosys.2023.110951.

[35] H. Hu, Y. Wan, K. Y. Tang, Q. Li, and X. Wang, "Affective-Computing-Driven Personalized Display of Cultural Information for Commercial Heritage Architecture," *Appl. Sci.*, vol. 15, no. 7, pp. 1–20, 2025, doi: 10.3390/app15073459.

[36] A. Boumhidi, A. Benlahbib, and E. H. Nfaoui, "Aggregating Users' Online Opinions Attributes and News Influence for Cryptocurrencies Reputation Generation," *J. Univers. Comput. Sci.*, vol. 29, no. 6, pp. 546–568, 2023, doi: 10.3897/jucs.85610.

[37] H. Yang and J. Chen, "Art appreciation model design based on improved PageRank and ECA-ResNeXt50 algorithm," *PeerJ Comput. Sci.*, vol. 9, pp. 1–17, 2023, doi: 10.7717/PEERJ-CS.1734.

[38] Y. Han and Z. Xu, "Fostering college students' mental well-being: the impact of social networking site utilization on emotion management and regulation," *BMC Psychol.*, vol. 12, no. 1, 2024, doi: 10.1186/s40359-024-02186-7.

[39] N. Silva, P. J. S. Cardoso, and J. M. F. Rodrigues, "Multimodal Sentiment Classifier Framework for Different Scene Contexts," *Appl. Sci.*, vol. 14, no. 16, 2024, doi: 10.3390/app14167065.

[40] Z. Liu, B. Zhou, D. Chu, Y. Sun, and L. Meng, "Modality translation-based multimodal sentiment analysis under uncertain missing modalities," *Inf. Fusion*, vol. 101, no. April 2023, p. 101973, 2024, doi: 10.1016/j.inffus.2023.101973.

[41] G. Mu, Y. Chen, X. Li, L. Dai, and J. Dai, "Semantic enhancement and cross-modal interaction fusion for sentiment analysis in social media," *PLoS One*, vol. 20, no. 4 April, pp. 1–26, 2025, doi: 10.1371/journal.pone.0321011.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[42] L. Jixian, A. Gang, S. Zhihao, and S. Xiaoqiang, "Social Media Multimodal Information Analysis based on the BiLSTM-Attention-CNN-XGBoost Ensemble Neural Network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 12, pp. 104–111, 2022, doi: 10.14569/IJACSA.2022.0131215.