Hindawi
IET Software
Volume 2024, Article ID 6819714, 11 pages
https://doi.org/10.1049/2024/6819714

Research Article

Design and Efficacy of a Data Lake Architecture for Multimodal Emotion Feature Extraction in Social Media

Yuanyuan Fan and Xifeng Mi

Jiaozuo Normal College, Jiaozuo 454000, China

Correspondence should be addressed to Yuanyuan Fan; jzfanyy@jzsz.edu.cn

Received 11 December 2023; Revised 12 February 2024; Accepted 23 February 2024; Published 8 March 2024

Academic Editor: Manuel Angel Serrano

In the rapidly evolving landscape of social media, the demand for precise sentiment analysis (SA) on multimodal data has become increasingly pivotal. This paper introduces a sophisticated data lake architecture tailored for efficient multimodal emotion feature extraction, addressing the challenges posed by diverse data types. The proposed framework encompasses a robust storage solution and an innovative SA model, multilevel spatial attention fusion (MLSAF), adept at handling text and visual data concurrently. The data lake architecture comprises five layers, facilitating real-time and offline data collection, storage, processing, standardized interface services, and data mining analysis. The MLSAF model, integrated into the data lake architecture, utilizes a novel approach to SA. It employs a text-guided spatial attention mechanism, fusing textual and visual features to discern subtle emotional interplays. The model's end-to-end learning approach and attention modules contribute to its efficacy in capturing nuanced sentiment expressions. Empirical evaluations on established multimodal sentiment datasets, MVSA-Single and MVSA-Multi, validate the proposed methodology's effectiveness. Comparative analyses with state-of-the-art models showcase the superior performance of our approach, with an accuracy improvement of 6% on MVSA-Single and 1.6% on MVSA-Multi. This research significantly contributes to optimizing SA in social media data by offering a versatile and potent framework for data management and analysis. The integration of MLSAF with a scalable data lake architecture presents a strategic innovation poised to navigate the evolving complexities of social media data analytics.

1. Introduction

With the advent and ubiquity of the Internet, coupled with the exponential advancement of mobile technologies, social networking has become a crucial aspect of modern human interaction. Platforms like Facebook, Twitter, Instagram, and LinkedIn attract billions of users, generating vast amounts of data daily. These networks have transformed the paradigms of social interaction, facilitating the exchange of various content forms and the cultivation of virtual communities. The ascendancy of social networking has significantly streamlined communication, catalyzing the transition from unimodal to multimodal content representation, enriching the texture of digital communication [1, 2]. The data created by these interactions holds latent analytical value. Extracting and dissecting this unstructured data have immense potential, particularly in the domain of sentiment analysis (SA)—a critical tool in various fields, from financial forecasting to public opinion monitoring and crisis management [3–5]. The limitations of unimodal data in conveying nuanced emotions necessitate the exploration of multimodal data for clearer affective communication, which carries profound social utility and significance [6, 7].

Despite advancements, navigating the landscape of multimodal SA presents formidable challenges. The storage, synthesis, and collaborative exploration of intricate datasets pose significant hurdles. The heterogeneity and vast volume of social network data create barriers to smooth cross-domain integration and analysis, resulting in discrepancies in data standards that, in turn, compromise the effectiveness of higher order analytical processes. The intricate task of effectively employing advanced computational methods, including machine learning, graph computation, and deep learning, further complicates comprehensive data utilization.

Table 1: Analysis of social network data types.

Data type	Data format	Raw storage format	Data manipulation features
Text data	Structured	Database, text files	Supports full-text search, natural language processing
Image data	Unstructured	Image files	Requires image processing and analysis technology
Video data	Unstructured	Video files	Requires video processing technology
Audio data	Unstructured	Audio files	Requires audio processing technology
Behavioral data	Structured	Databases	User activity tracking and analysis
Emotional data	Structured	Databases	Sentiment analysis and association mining

Consequently, there is a pressing need for an adept framework capable of addressing these intricate challenges in multimodal data management.

The concept of a data lake, a nascent yet rapidly emerging centralized repository, has drawn significant scholarly attention [8–10]. It enables the storage of extensive raw data arrays, supported by metadata cataloging and governance mechanisms to facilitate a comprehensive suite of big data applications.

In response to these scholarly discourses, this treatise delineates a multitiered research investigation into the architectural design of scalable data lakes and the mining of sentiment associations therein. This paper posits a novel data lake architecture tailored for social multimodal sentiment association mining, demonstrating, through empirical application, the robustness and superiority of the proposed system. The manuscript contributes by presenting a multitiered research investigation into the architectural design of scalable data lakes and the mining of sentiment associations within social networks. The primary contribution lies in the conceptualization and development of a sophisticated data lake architectural framework, specifically calibrated for the mining of multimodal sentiment associations within the dynamic context of social networking environments. Furthermore, the introduction should emphasize the novel integration of the multilevel spatial attention fusion (MLSAF) model with the data lake infrastructure, providing a layered and scalable solution for the comprehensive examination of social network data. These contributions collectively optimize SA within social media data, offering a versatile and potent framework for data management and analysis within the social sphere.

The structure of this paper is as follows: the introduction section (Section 1) sets the stage by highlighting the significance of SA in the context of the rapidly evolving landscape of social media. Section 2 delves into the state of the art, providing a comprehensive overview of the data lake concept and existing research on multimodal SA. Section 3 then presents the methodology in detail, with a focus on the design of the data lake architecture and the construction of the MLSAF model. Section 4 presents the results and discussion, showcasing empirical evidence from experiments conducted on prevalent multimodal sentiment datasets. The conclusion (Section 5) summarizes the key findings and contributions of this research, emphasizing the strategic innovation achieved by integrating MLSAF with the scalable data lake architecture.

2. State of the Art

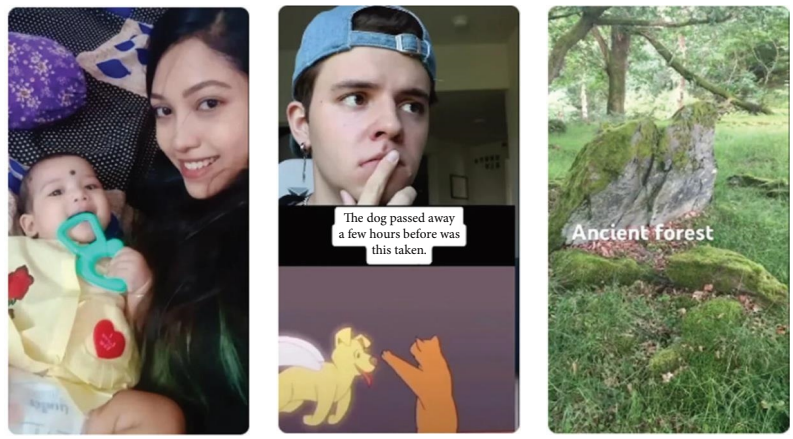
2.1. Data Lake Concept and Social Network Data Analytics. A data lake is a data management and analytics architecture that emphasizes flexibility, real-time capabilities, and data diversity, enabling organizations to store and process large-scale unstructured and structured data [11]. It advocates the philosophy of "any data, any time, any way," allowing users to store and access data in its raw form, thereby promoting data diversity and integration. The availability of a data lake is manifested in its flexibility, real-time processing, and scalability [12]. It can accommodate various types and formats of data, support real-time storage and analysis, and scale horizontally to handle large volumes of data at a lower cost. However, the implementation of a data lake faces challenges, including data quality, security, management complexity, and query performance [13]. By comprehensively addressing these challenges, organizations can fully leverage the advantages of a data lake, enhancing data availability and insights for more effective decision-making and innovation.

Social network data has various sources, complex types, and structures. In this study, we analyze the data sources selected for social network data lake architecture, which mainly include six data sources. These data sources are summarized as shown in Table 1.

2.2. Research on Multimodal Sentiment Analysis. Multimodal SA is a computational study of viewpoints, emotional states, etc., derived from data comprising text, images, audio, or even video, building upon unimodal SA [6, 14]. Extracting emotional features from a single modality is often subject to ambiguity and may require the integration of other modal information to more accurately convey emotional tendencies. For example, Figure 1 illustrates various graphic message contents on YouTube.

Figure 1 shows the image data obtained from YouTube, which can be analyzed as follows, positive and negative emotions can be easily observed in Figures 1(a) and 1(b). However, by solely observing the image in Figure 1(c), one might mistakenly perceive it as depicting a beautiful forest. It is only when we consider the emotional term like "abandoned" from the accompanying text that we can comprehend the negative emotions conveyed by the creator.

Mining the correlation information between images and text is one of the main directions of current multimodal SA research. Ghorbanali et al. [15] proposed a convolutional neural network (CNN)-based framework for multimodal



Our happy moment #love #memories #moments...

You cry you lose□□

Abandoned quarry in the Scottish highlands...

(a) (b) (c)

FIGURE 1: Example of YouTube graphic data. (a) Happiness and satisfaction, (b) depression and disappointment, and (c) an emotionally neutral state.

content sentiment prediction (including tweets of text and images). Yang et al. [16] proposed a CNN-based integrated visual–text SA model. This model utilizes a CNN network to amalgamate emotional features from images and text, enabling sentiment polarity prediction for Chinese Weibo data. Zhang et al. [17] introduced a multimodal feature learning model based on continuous bag of words (CBOW) and denoising autoencoder (DA) for SA of Twitter data. This model can also be applied to other social media datasets. Xu et al. [18] introduced the cross-modal consistent regression (CCR) method for combining visual and textual SA. This approach utilizes deep visual and textual features to construct a regression model. Wang et al. [19] employed the AlexNet network to extract visual features from images and utilized the GloVe model for obtaining textual vector representations. Subsequently, they proposed a method called supervised collective matrix factorization (SCMF), which considers label information during the matrix factorization process to acquire unified representations for sentiment prediction. Additionally, Sevastjanova et al. [20] considered the prevalent phenomenon of visual rhetoric in web advertising images. They used an adaptive encoder to comprehend visual rhetoric in images and incorporated topic analysis within a multitask framework to enhance SA. Yadav and Vishwakarma [21], focusing on cross-modal entity consistency between text and images, introduced a Bidirectional multilevel attention model (BDMLA). Building upon this model, they further proposed entity-level SA for social media posts. Addressing the unequal relationship between text and images in online comments, Huang et al. [22] presented VistaNet, which treats images as supplementary features to text rather than independent information. It utilizes images as attention anchors to emphasize key sentences in the text.

The M-SENA platform stands as a significant contribution to advanced multimodal sentiment analysis, offering not only an open-sourced framework with flexible toolkits and reliable benchmarks but also a modular video SA architecture, contributing valuable resources to the research community [23]. In [24], a pioneering framework known as MWRCMH fusion is presented, addressing the challenges inherent in real-world multimodal sentiment analysis. Through the incorporation of multimodal word correction and cross-modal hierarchical fusion, this work surpasses existing methodologies, showcasing its potential for superior sentiment recognition in practical scenarios. A comprehensive exploration into sentiment analysis (SA) and emotion detection (ED) unfolds in this scholarly review, emphasizing the crucial role of recurrent neural networks and their architectural variants, in handling textual, visual, and multimodal inputs on social networking platforms [25]. The comprehensive surveys navigate the evolution of SA from text-based models to the realm of multimodality, exploring diverse approaches, applications, challenges, and the transformative impact of deep neural architectures, and reflecting the dynamic shift in SA trends [26, 27]. Presenting an innovative multimodal SA system, the study [28] meticulously integrates text- and image-based components, utilizing deep learning for feature learning and classification. In the literature [29], researchers introduced a novel approach, utilizing a cross-modal modeling multitensor fusion network for effective emotion fusion in multimodal SA. Regression and classification experiments were conducted on the CMU-MOSI and CMU-MOSEI datasets. In [30, 31], the establishment of baselines and dataset splits for multimodal SA is discussed, addressing key issues often overlooked in research. Three progressive deep learning architectures are introduced, evaluated across multiple datasets, serving as a benchmark for future multimodal SA research. The growing importance of SA and ED in social media is explored in [32, 33], emphasizing the effectiveness of CNN across textual, visual, and multimodal inputs, accompanied by detailed discussions and optimized algorithmic approaches.

3. Methodology

To address the mining needs of multimodal social media data, this paper proposes an innovative approach that integrates a hierarchical scalable data lake with sentiment association

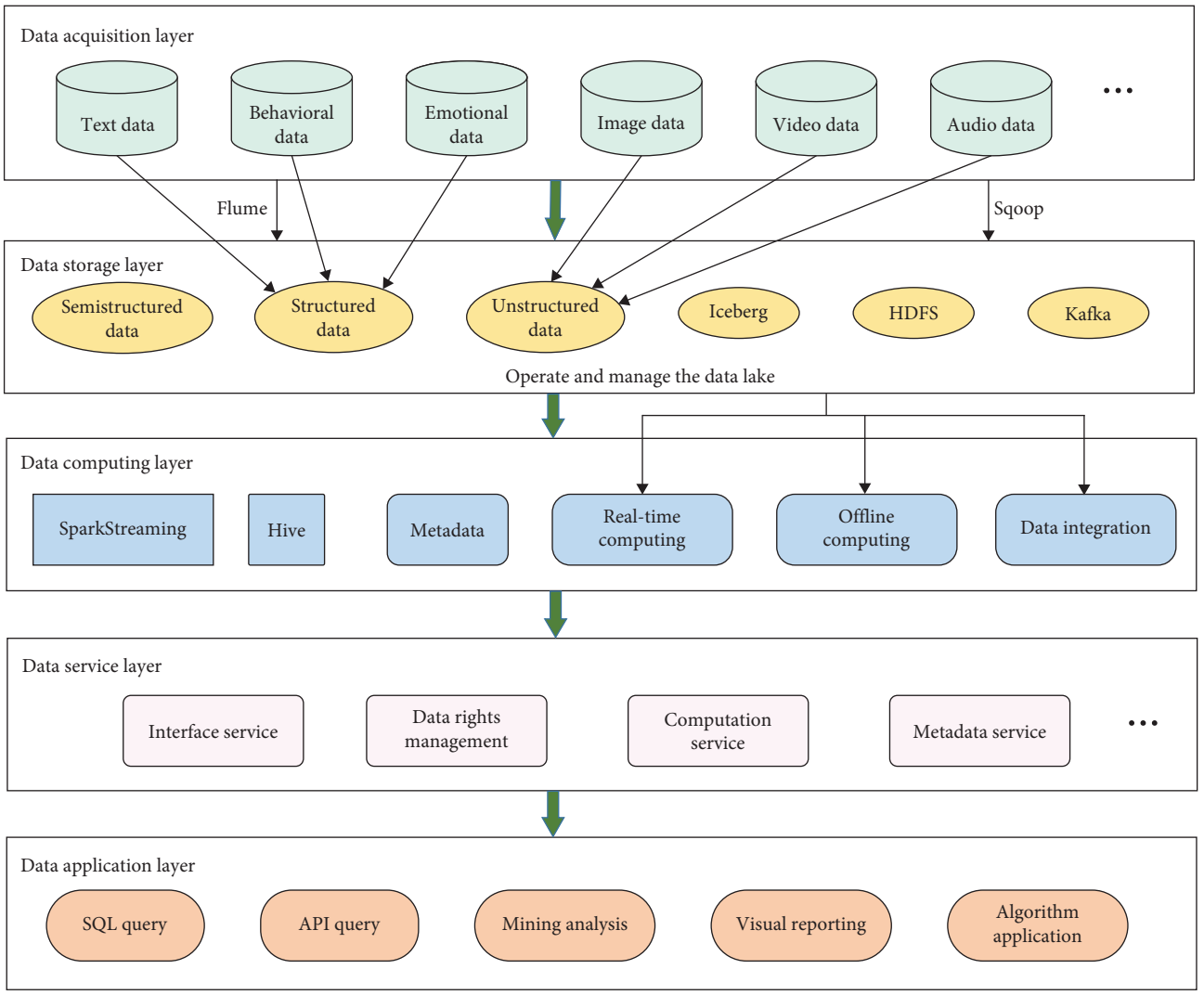


Figure 2: Data lake architecture for social multimodal sentiment association mining.

mining. This approach involves the integration of multimodal social media data into a layered scalable data lake. In conjunction with the MLSAF SA model, it accomplishes correlated sentiment mining in social networks.

3.1. Data Lake Architecture Design. The proposed data lake architecture for social multimodal sentiment association mining consists of five layers, as shown in Figure 2.

Data acquisition layer: This layer combines real-time and offline data collection modes to systematically collect raw data from social networks into the data lake.

Data storage layer: Enhancing storage flexibility without the need for full data reloading, this layer operates in real-time and stores social multimodal data with flexible field modification in tabular format.

Data computing layer: Following a Lambda architecture, this layer combines batch processing for offline data warehousing and real-time data warehousing for critical business operations. Further subdivisions include real-time computing, offline computing, and data integration. Kafka is used for real-time data collection and storage, while Spark Streaming handles real-time data processing. Hive is employed for batch processing of offline data.

Data service layer: Externally provide standardized data-related interface services and computation services through master data and service bus technology.

Data application layer: Provide modular data mining analysis algorithm packages to support network operation managers in data mining and visualization analysis.

3.2. MLSAF Model Construction. The MLSAF model adopts an end-to-end learning approach. In contrast to methods that extract features solely from the highest level convolutional outputs of the image, the MLSAF model incorporates a five-branch text-guided spatial attention module. This module applies spatial attention weighting to CNN's output layers. The weighted feature matrices are used as inputs in subsequent convolutional layers, eventually leading to the final convolutional output layer. The overall architecture of MLSAF is illustrated in Figure 3.

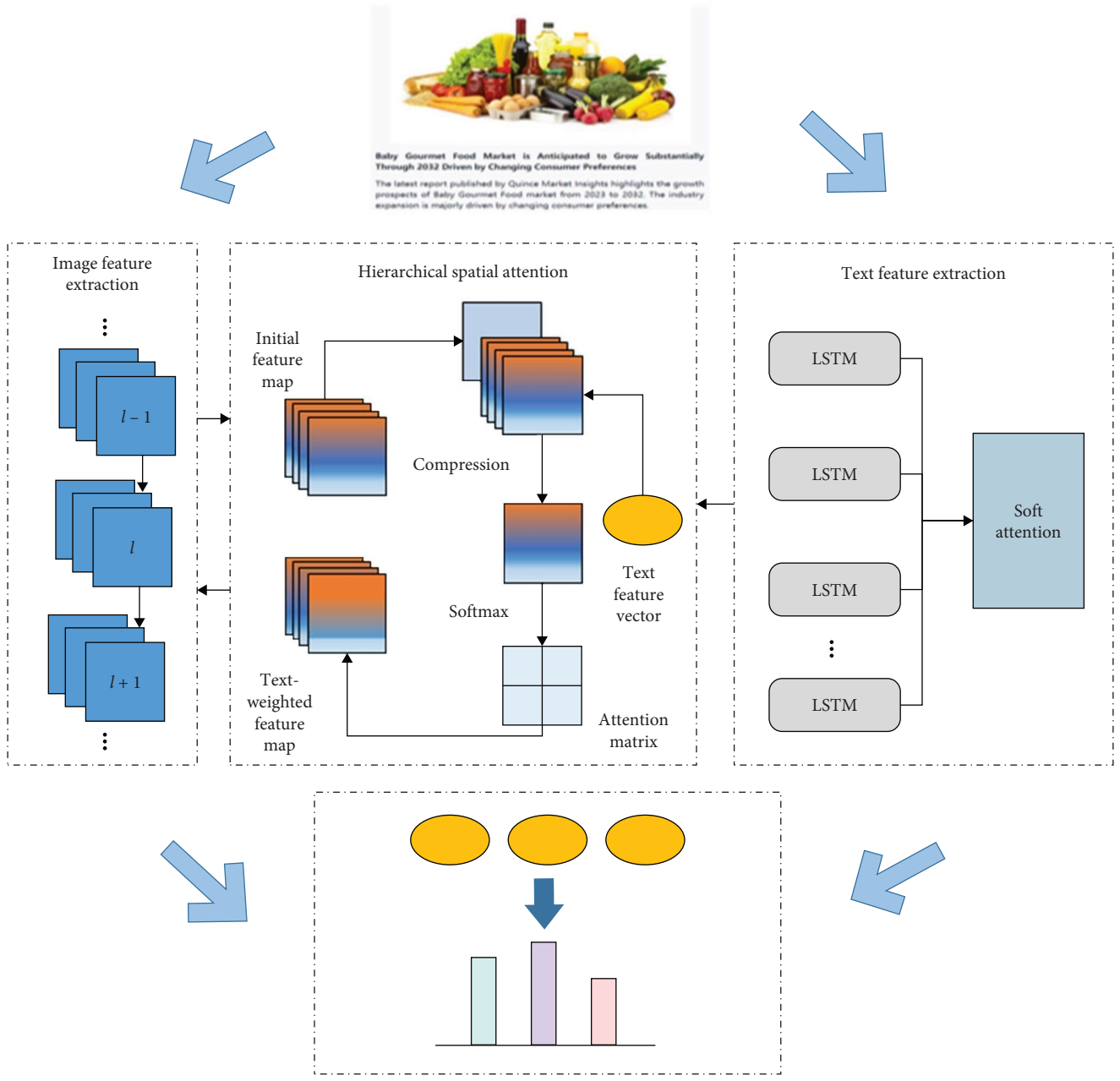


FIGURE 3: Basic structure of MLSAF.

3.2.1. Text Feature Extraction. The preprocessed tweeted text is used as input to the feature extraction network. The given text sequence $M = [m_1, \mathrm{m}, \cdots, m_T]$. Where m_t is a one-hot vector representation of the word at position t, and the subscript T denotes the total length of the text sequence. The words are first embedded into the vector space using the embedding matrix W_{glove}:

$$i_t = W_{\text{glove}} m_t \in R^E. \tag{1}$$

For each word embedding vector, the model uses a long short-term memory (LSTM) network for further encoding. The LSTM accepts the word embedding i_t as input and outputs a new hidden state vector b_t.

$$b_t = \text{LSTM}(i_t). \tag{2}$$

Since the textual sentiment semantics is affected by the contextual content, the MLSAF model introduces a bidirectional LSTM mechanism, which connects the hidden state vectors generated by the forward LSTM and the backward LSTM to get the final vector representation $b_t = [\vec{b}_t, \overleftarrow{b}_{nt}]$ for each word. Each word in a sentence is different, and some words provide more valid information about the sentiment. In order to calculate and assign the weight of each word in sentiment categorization, the model incorporates a soft attention mechanism.

$$p_t = \tanh\left(M_b b_t + h_b\right), \tag{3}$$

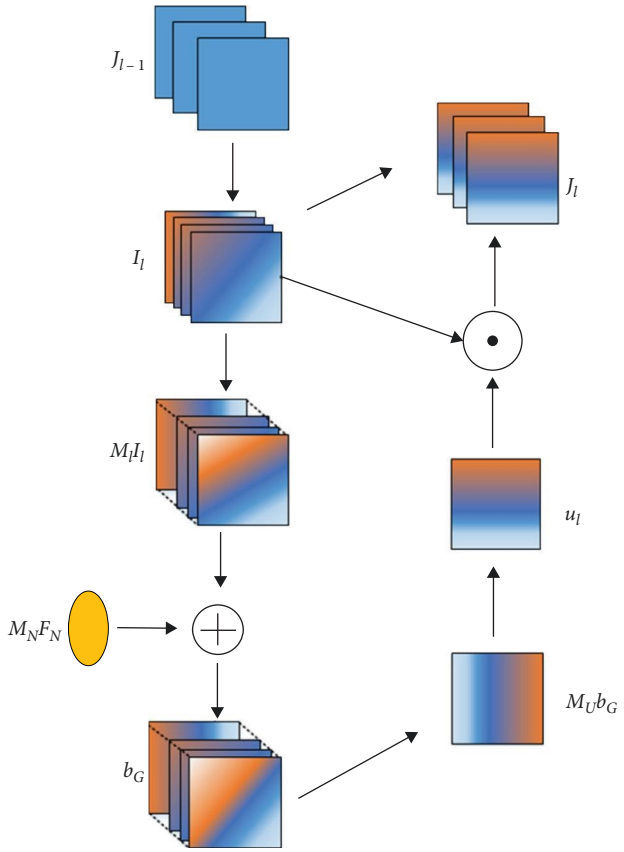


FIGURE 4: Text-guided MSAM.

$$\alpha_t = \exp(p_t) / \sum_t \exp(p_t), \tag{4}$$

where p_t is a nonnormalized attention score that measures the relationship between word b_t and text sentiment. The text semantic vector F_N of the text as a whole can be computed by weighted average of the word features as in Equation (5).

$$F_N = \sum_t \alpha_t b_t. \tag{5}$$

3.2.2. Image Feature Extraction. The MLSAF model expands spatial attention beyond the highest convolution layer by computing text-guided visual attention weights on feature maps at every convolutional layer. To further demonstrate the overall process of image feature extraction, the structural details of the text-guided multilevel spatial attention module (MSAM) are further demonstrated using Figure 4.

Formally, it is assumed that the model will generate layer l features of the image. At layer l, the text vector F_N will determine the spatial attention weights u_l and the image feature map J_l modulated by the attention weights.

$$u_l = \Phi(F_N, I_l), \tag{6}$$

$$J_l = f(I_l, u_l), \tag{7}$$

where $\Phi(\,)$ is the spatial attention function, see Equation (9)–(10) for details. $f(\,)$ is the modulo function that linearly combines the image features with the attention weights, Equation (11). I_l is denoted as the graphic feature matrix output from the feature mapping of the $l-1$ convolutional layer.

$$I_l = \mathrm{CTT}(J_{l-1}). \tag{8}$$

Given the text vector F_N and the image feature matrix $I_l \in R^{C\times W}$ of layer l, where W is the number of regions of the image matrix of the layer and C is the number of channels of the layer. The inputs are first projected into the same dimensions by a single-layer neural network. Then, the attention probability u_l of the text corresponding to each image region is generated by the softmax function guided by the text vector F_N.

$$b_G = \tanh(M_lI_l \oplus (M_NF_N + h_G)), \tag{9}$$

$$u_l = \mathrm{softmax}\,(M_Ub_G + h_U). \tag{10}$$

where $M_l \in R^{z\times C}$ and $M_N \in R^{z\times d}$ are the transformation matrices that map the image visual features and text vectors into the same vector space. $M_u \in R^{1\times z}$ provides the compression rules in the channel direction. $h_G \in R^z$ and $h_u \in R^1$ are the bias terms of the linear transformations. Addition between matrices and vectors is accomplished by adding each column of the matrix to the vector. Calculate the product of pixel regions and corresponding region weights in the feature map based on attention distribution. Then, in the process of image feature generation, encode visual information related to the text.

$$J_l^w = u_l^w I_l^w. \tag{11}$$

where $J_l^w \in R^{1\times c}$ and $I_l^w \in R^{1\times C}$ refers to the visual feature vectors of the regions indexed by m in J_l and I_l. u_l^w refers to the text-guided attention weight on the w-th pixel region. In the convolution process of CNN, the convolutional layers learn higher level visual features as the receptive field expands. In order to obtain visual features containing multilevel associations, the output of the last convolutional layer is extracted as the final image sentiment feature F_X.

$$F_X = \sum_w u_L^w I_L^w. \tag{12}$$

where L is the total number of layers in the CNN convolutional layers.

3.2.3. Sentiment Classification. A fusion layer is first used to aggregate the visual features F_X and textual features F_N of the existing graphic reviews into a final multimodal representation. Then, a softmax classifier is added on top for sentiment classification.

$$F_{mul} = \tanh(M_F[F_X, F_N] + h_F), \tag{13}$$

$$\rho = \mathrm{softmax}\,(M_{mul}F_{mul} + h_{mul}). \tag{14}$$

The cross-entropy loss is used as the objective function of softmax to train the model in a supervised manner.

$$\text{loss} = -\frac{1}{D}\sum_{d}\widehat{\rho}_{\text{d}}\log(\rho_{\text{d}}). \quad (15)$$

In the proposed model, feature fusion for sentiment classification is achieved through a dedicated fusion layer. This layer serves to aggregate both visual and textual features extracted from the graphic reviews, creating a final multimodal representation. Specifically, the fusion layer combines the information encapsulated in Equations (13) and (14). Following the fusion process, a softmax classifier is introduced to facilitate sentiment classification. The training of the model is conducted in a supervised manner, utilizing the cross-entropy loss as the objective function for optimizing the softmax classifier (equation not provided here). This approach ensures effective integration of both visual and textual features in the SA task, enhancing the model's capacity to capture multimodal aspects of emotion expression in social media content.

4. Result Analysis and Discussion

4.1. Dataset Selection. The experiments are based on the MVSA dataset, which is a commonly used dataset in graphical SA tasks [34]. It contains two independent datasets, MVSA-Single and MVSA-Multi, whose samples are graphic comments collected from Twitter. Among them, MVSA-Single contains 5,129 image–text pairs, each with one sentiment label, and MVSA-Multi consists of 19,600 image–text pairs, each with three sentiment labels. Due to different annotations, inconsistent data in the MVSA-Multi dataset were processed according to these rules in this experiment: when two or more of the three sentiment labels matched, they were retained as the sentiment labels; when data had both positive and negative sentiment labels, they were removed. The dataset was also cleaned, resulting in a final count of 4,511 image–text pairs for the MVSA-Single dataset and 17,024 image–text pairs for the MVSA-Multi dataset.

4.2. Experimental Setup. According to the source of the dataset, "glove. twitter.27B.200d" is selected as the text embedding; the five convolutional modules of the MLSAF model are initialized using the pretrained VGG-T4SA FT-A. The initial learning rate of the network was 0.001, the batch size was 64, and the epoch was 100. The optimal parameters were trained by back propagation using the RMSProp update rule. To avoid overfitting, Dropout (value set to 0.1) is used after the fully connected layer. Early stopping technique was used, and the patience was set to 10, i.e., the training was stopped when the loss value of the validation set did not decrease for 10 consecutive times. The MVSA dataset is divided according to the ratio of 8 : 1 : 1, which is used as the training set, validation set, and test set, respectively.

4.3. Analysis of Experimental Results of MLSAF Model

4.3.1. Performance Analysis. Table 2 shows the results of comparing the accuracy and F1 value of the MLSAF model proposed in this paper with four other models (from Literature [35–39]) on the MVSA dataset.

Table 2: Comparison results of different models on two datasets.

Model	MVSA-Single		MVSA-Multi	
	Accuracy	F1 value	Accuracy	F1 value
Literature [35]	62.6	63.3	62.4	61.5
Literature [36]	70.1	69.9	70.2	69.5
Literature [37]	70.9	69.7	70.8	69.6
Literature [38]	72.6	72.3	72.5	72.4
Literature [39]	73.2	73.6	74.1	75.8
MLSAF	77.6	76.4	75.3	77.1

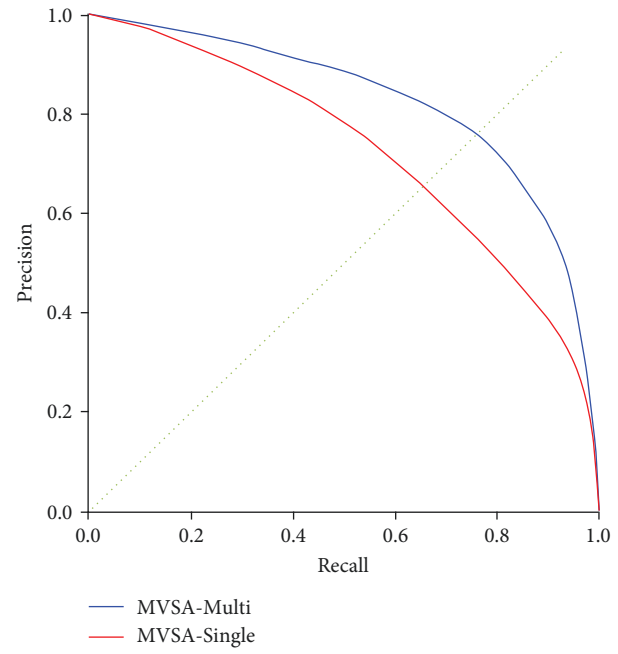


Figure 5: MLSAF model P–R curve.

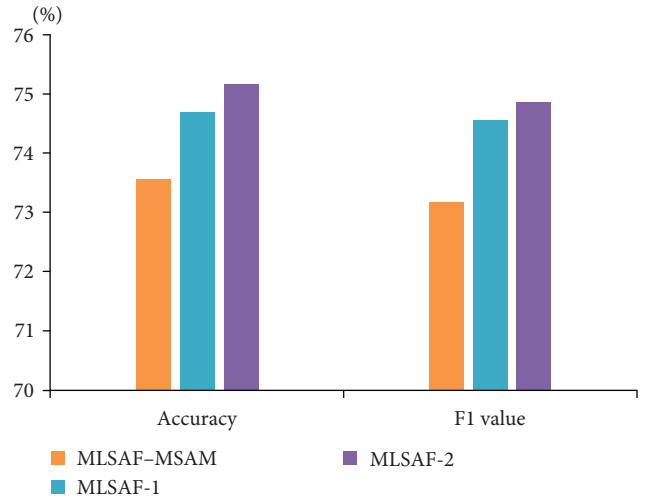
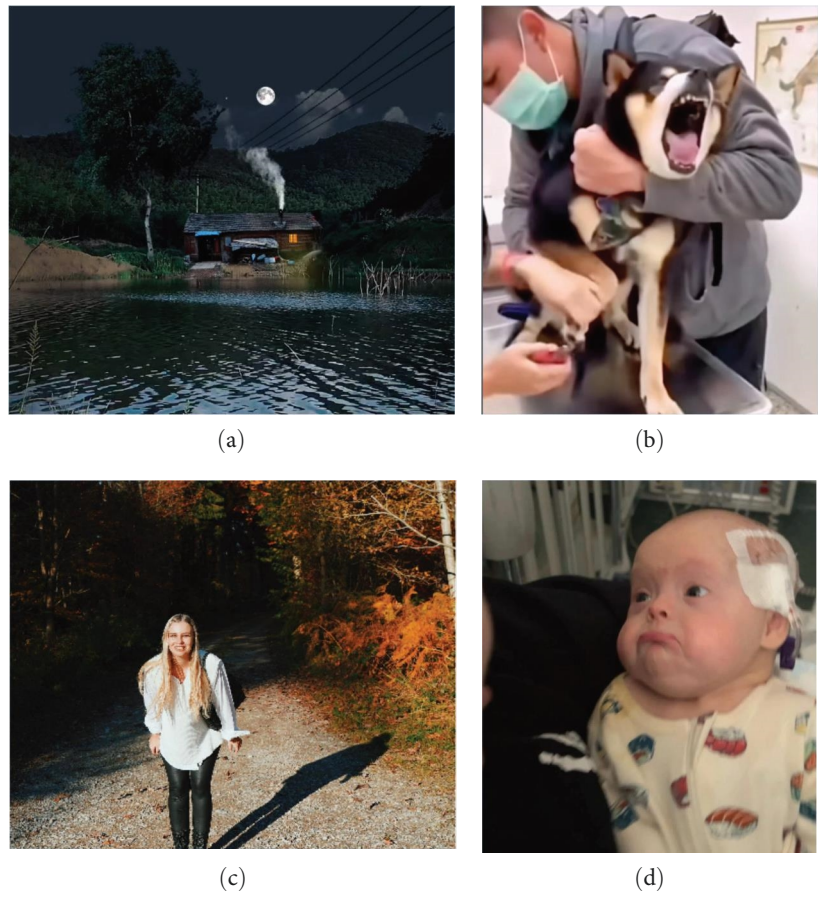


Figure 6: Ablation experiment results.

As can be seen from Table 2, the [35] has the worst performance because it only performs sentiment classification by extracting features for fusion. In this research, it is



(a) (b) (c) (d)

Figure 7: Legend of test samples. (a) I can give you a description of a peaceful village night. (b) The vet scared my dog so much he clipped his nails. (c) The nice weather became the reason for someone's smile today. (d) This girl has started social smiling, and every once in a while, she throws in a big dramatic frown!

used as a benchmark model against which the performance of all other models is measured. Al-Tameemi et al. [36] employs a deep learning neural network to acquire feature representations from various modalities, demonstrating a substantial performance enhancement in comparison to Ramasubramanian [35]. The average accuracy and F1 value of [36] on the two datasets surpass those of [35] by 12% and 10.4%, respectively. The [37] adds to this the effect of image information on text information. This model's performance improves the average accuracy and F1 by 13.4% and 11.6%, respectively, compared to [35]. However, a drawback of [37] is its disregard for the interaction between visual and textual information. This is improved in the [38], which not only learns multimodal features but also considers the interrelationships among multimodal features. So, its average accuracy and F1 value are 2.4% and 3.9% higher than the literature [37], respectively. However, its disadvantage is that the use of coarse-grained attention mechanism tends to lead to information loss. The [39] utilizes text, vision, and audio to predict emotions. In comparison to [38], its average accuracy and F1 values are increased by 0.8% and 1.8%, respectively, compared to [37]. Compared to the [39], the MLSAF model we propose demonstrates superior accuracy and F1 values, showcasing improvements of 6% and 3.8% on the MVSA-Single dataset and 1.6% and 1.7% on the MVSA-Multi dataset, respectively. It shows that the proposed model is the best among all the compared models. This is because the proposed model considers both the contextual information of the text and emphasizes the importance of channel information in visual attention, which enriches feature information. Additionally, it reintegrates the interaction between multimodal information into feature vectors through an attention mechanism and highlights the importance of each modality. The experimental results prove that the proposed model can effectively improve classification performance.

Since most of the data in the MVSA dataset is unbalanced, the P–R plot (precision–recall curve) of the MLSAF model on the two datasets is plotted as shown in Figure 5. This graph helps to understand the actual effect and role of the model, and can also be used to improve the model performance. From Figure 5, it can be seen that the proposed MLSAF model effectively achieves results in learning the multimodal sentiment classification task.

4.3.2. Ablation Study. In order to analyze the model effect and explore the influence of each module of the model on the overall performance at a deeper level, this section makes the following changes to the original model and retrains it:

Table 3: Comparison model test results with MLSAF–MSAM.

Images	Text	Real emotion	MLSAF–MSAM	MLSAF
(a)	I can give you a description of a peaceful village night.	Neutral	Neutral	Neutral
(b)	The vet scared my dog so much he clipped his nails.	Negative	Negative	Negative
(c)	The nice weather became the reason for someone's smile today.	Positive	Positive	Positive
(d)	This girl has started social smiling, and every once in a while, she throws in a big dramatic frown!	Positive	Negative	Positive

(i) MLSAF–MSAM: eliminate the channel and spatial attention module in the visual feature extraction, in this way, to study the influence of the MSAM on the model performance. (ii) MLSAF-1: only the unidirectional effect of image information on text information is considered. (iii) MLSAF-2: directly cascade two features after bidirectional feature fusion to form the final output and input the fully connected layer for final classification. The experiments are carried out on the MVSA-Multi dataset, and the results are shown in Figure 6.

From Figure 6, it can be seen that after changing the model structure, the performance of the models decreased to a certain extent. Compared with the MLSAF–MSAM model after eliminating the multilayer spatial attention mechanism, the accuracy and F1 value of the original model decreased by 3.9% and 3.3%, respectively. This underscores the enhanced effect of the attention mechanism, which is based on the channel domain and spatial domain, on image feature information. This enhancement assists the model in accurately localizing key components. In the MLSAF-1 model, only the unidirectional effect of image information on text information is retained, resulting in a decrease of 0.8% and 0.6% in accuracy and F1 value on the MVSA-Multi dataset, respectively. This reflects the importance of attentional interaction and shows that a close relationship between words and images can improve the model's performance. In MLSAF-2, the removal of the modality fusion module resulted in a decrease of 0.33% and 1.28% in the model on the MVSA-Multi dataset, respectively. This highlights the necessity of focusing on informative features across multiple modalities, as not all features contribute equally to the task.

4.3.3. Case Study. In order to further evaluate the MLSAF model performance, four samples from the dataset were randomly selected for experiments on MLSAF–MSAM and MLSAF. Among them, Figure 7 shows an example of the pictures in the selected samples, and Table 3 shows the results for each sample on both models, with model prediction errors highlighted in bold. Table 3 also describes the information conveyed in Figure 7. The four illustrations in Figure 7 correspond to each other in Table 3.

In the case of Figure 7(d), MLSAF–MSAM made an incorrect prediction. It is hypothesized that the model may have excessively amplified less relevant image details, leading to an inaccurate emotional interpretation of the text. The classification results indicate that, among the four selected samples, MLSAF–MSAM correctly predicted three, while the MLSAF model accurately predicted all four, achieving a 100% accuracy rate. These results validate the contribution of the MSAM module in enhancing image features, thereby effectively improving sentiment classification performance.

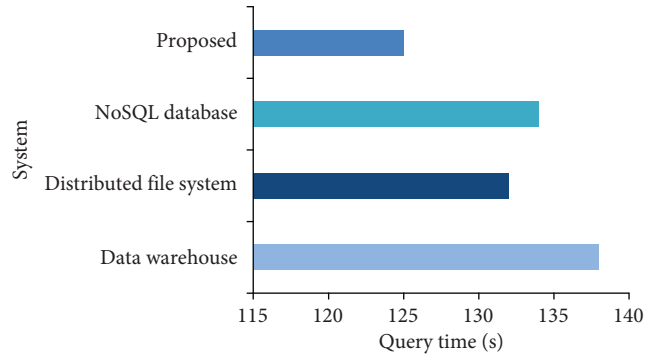


Figure 8: Data query time test.

4.4. Data Lake Platform Application Results and Analysis. In order to evaluate the performance of the designed data lake architecture platform, access query experiments are conducted on the data analysis layer of the data lake platform using the proposed MLSAF model. On the basis of the above experiments, structured text data and unstructured image data from social networks are selected and input to the data lake for scanning test of query speed. Meanwhile, a comparative study is conducted with several other systems with similar functions (data warehouse, distributed file system, and NoSQL database). The test results are presented in Figure 8.

5. Conclusion

In the midst of rapid advancements in network technologies, the volume and complexity of data generated by social media platforms have witnessed unprecedented growth. This expansion presents challenges to conventional data warehousing solutions, often constrained by scalability and performance bottlenecks. The imperative for nuanced cross-domain analyses of this vast dataset necessitates innovative architectural solutions. This scholarly investigation focuses on the conceptualization and development of a data lake architectural framework, intricately designed to facilitate the storage, management, and retrieval of multimodal data from social media networks. The study commences by constructing a hierarchical, scalable framework for a data lake, meticulously tailored to handle the storage, management, and retrieval of multimodal data from social media networks. Building upon this foundational architecture, the research introduces an advanced

approach for multimodal sentiment association mining. This methodology is anchored by the MLSAF model, which adeptly integrates high-level semantic features with low-level visual cues from images, thereby enriching the emotional context of associated textual data. The incorporation of the MLSAF model into the data lake infrastructure enables a comprehensive exploration of sentiment associations, enhancing analytical capabilities within the social network milieu. Empirical trials utilizing the MVSA multimodal sentiment dataset substantiate the model's validity and unveil its robust performance in SA tasks. Furthermore, comparative analyses through systematic data querying across multiple platforms validate the technical and functional ascendancy of the proposed data lake system. In summary, this scholarly endeavor offers substantial theoretical and practical advancements for addressing the intricacies of social network analytics amidst the ever-expanding volumes of social media data. The synergistic deployment of a scalable data lake architecture, in conjunction with the MLSAF model, represents a strategic innovation poised to effectively navigate the evolving complexities of the social media data landscape.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request. The images in this article are from the open source dataset (MVSA), which can be downloaded via a Google search, or via the URL https://mcrlab.net/research/mvsa-sentiment-analysis-on-multi-view-social-data.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] B. Singh and D. K. Sharma, "Predicting image credibility in fake news over social media using multi-modal approach," *Neural Computing and Applications*, vol. 34, no. 24, pp. 21503–21517, 2022.

[2] S. Bayrakdar, I. Yucedag, M. Simsek, and I. A. Dogru, "Semantic analysis on social networks: a survey," *International Journal of Communication Systems*, vol. 33, no. 11, Article ID e4424, 2020.

[3] C. N. Dang, M. N. Moreno-García, and F. D. L. Prieta, "An approach to integrating sentiment analysis into recommender systems," *Sensors*, vol. 21, no. 16, Article ID 5666, 2021.

[4] A. L. Karn, R. K. Karna, B. R. Kondamudi et al., "Customer centric hybrid recommendation system for E-Commerce applications by integrating hybrid sentiment analysis," *Electronic Commerce Research*, vol. 23, no. 1, pp. 279–314, 2023.

[5] J. Zhao, F. Xiong, and P. Jin, "Enhancing short-term sales prediction with microblogs: a case study of the movie box office," *Future Internet*, vol. 14, no. 5, Article ID 141, 2022.

[6] G. Chandrasekaran, T. N. Nguyen, and D. J. Hemanth, "Multimodal sentimental analysis for social media applications: a comprehensive review," *WIREs Data Mining and Knowledge Discovery*, vol. 11, no. 5, 2021.

[7] N. Xu, W. Mao, P. Wei, and D. Zeng, "MDA: multimodal data augmentation framework for boosting performance on sentiment/emotion classification tasks," *IEEE Intelligent Systems*, vol. 36, no. 6, pp. 3–12, 2020.

[8] M. Derakhshannia, C. Gervet, H. Hajj-Hassan, A. Laurent, and A. Martin, "Data lake governance: towards a systemic and natural ecosystem analog," *Future Internet*, vol. 12, no. 8, Article ID 126, 2020.

[9] J. Kachaoui and A. Belangour, "From single architectural design to a reference conceptual meta-model: an intelligent data lake for new data insights," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 4, pp. 1460–1465, 2020.

[10] A. Khatiwada, R. Shraga, W. Gatterbauer, and R. J. Miller, "Integrating data lake tables," *Proceedings of the VLDB Endowment*, vol. 16, no. 4, pp. 932–945, 2022.

[11] A. A. Munshi and Y. A. R. I. Mohamed, "Data lake lambda architecture for smart grids big data analytics," *IEEE Access*, vol. 6, pp. 40463–40471, 2018.

[12] K. Peddireddy, "Kafka-based architecture in building data lakes for real-time data streams," *International Journal of Computer Applications*, vol. 185, no. 9, pp. 1–3, 2023.

[13] R. Hai, C. Koutras, C. Quix, and M. Jarke, "Data lakes: a survey of functions and systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 12, pp. 12571–12590, 2023.

[14] A. Ghorbanali, M. K. Sohrabi, and F. Yaghmaee, "Ensemble transfer learning-based multimodal sentiment analysis using weighted convolutional neural networks," *Information Processing & Management*, vol. 59, no. 3, Article ID 102929, 2022.

[15] A. Ghorbanali, M. K. Sohrabi, and F. Yaghmaee, "Multiple transfer learning-based multimodal sentiment analysis using weighted convolutional neural network ensemble," *Journal of Modeling in Engineering*, vol. 21, no. 72, pp. 83–97, 2023.

[16] W. Yang, T. Yuan, and L. Wang, "Micro-blog sentiment classification method based on the personality and bagging algorithm," *Future Internet*, vol. 12, no. 4, Article ID 75, 2020.

[17] K. Zhang, Y. Geng, J. Zhao, J. Liu, and W. Li, "Sentiment analysis of social media via multimodal feature fusion," *Symmetry*, vol. 12, no. 12, 2020.

[18] J. Xu, Z. Li, F. Huang, C. Li, and S. Y. Philip, "Social image sentiment analysis by exploiting multimodal content and heterogeneous relations," *IEEE Transactions on Industrial*, vol. 17, no. 4, pp. 2974–2982, 2020.

[19] B. Wang, G. S. Fang, and S. Kamei, "Matrix factorization with topic and sentiment analysis for rating prediction," *International Journal of Networking and Computing*, vol. 11, no. 2, pp. 198–214, 2021.

[20] R. Sevastjanova, M. El-Assady, A. Bradley, C. Collins, M. Butt, and D. Keim, "Visinreport: complementing visual discourse analytics through personalized insight reports," *IEEE Transactions on Visualization and Computer*, vol. 28, no. 12, pp. 4757–4769, 2021.

[21] A. Yadav and D. K. Vishwakarma, "A deep multi-level attentive network for multimodal sentiment analysis," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 19, no. 1, pp. 1–19, 2023.

[22] F. Huang, K. Wei, J. Weng, and Z. Li, "Attention-based modality-gated networks for image-text sentiment analysis," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 16, no. 3, pp. 1–19, 2020.

[23] H. Mao, Z. Yuan, H. Xu, W. Yu, Y. Liu, and K. Gao, "M-sena: an integrated platform for multimodal sentiment analysis," arXiv preprint arXiv: 2203.12441, 2022.

[24] J. Huang, P. Lu, S. Sun, and F. Wang, "Multimodal sentiment analysis in realistic environments based on cross-modal

hierarchical fusion network," *Electronics*, vol. 12, no. 16, Article ID 3504, 2023.
[25] J. V. Tembhurne and T. Diwan, "Sentiment analysis in textual, visual and multimodal inputs using recurrent neural networks," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 6871–6910, 2021.
[26] R. Das and T. D. Singh, "Multimodal sentiment analysis: a survey of methods, trends and challenges," *ACM Computing Surveys*, vol. 55, no. 13s, pp. 1–38, Article ID 270, 2023.
[27] S. Lai, X. Hu, H. Xu, Z. Ren, and Z. Liu, "Multimodal sentiment analysis: a survey," *Displays*, vol. 80, Article ID 102563, 2023.
[28] A. Ghosh, B. C. Dhara, C. Pero, and S. Umer, "A multimodal sentiment analysis system for recognizing person aggressiveness in pain based on textual and visual information," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 4, pp. 4489–4501, 2023.
[29] X. Yan, H. Xue, S. Jiang, and Z. Liu, "Multimodal sentiment analysis using multi-tensor fusion network with cross-modal modeling," *Applied Artificial Intelligence*, vol. 36, no. 1, Article ID 2000688, 2022.
[30] N. Majumder, D. Hazarika, A. Gelbukh, E. Cambria, and S. Poria, "Multimodal sentiment analysis using hierarchical fusion with context modeling," *Knowledge-Based Systems*, vol. 161, pp. 124–133, 2018.
[31] S. Poria, N. Majumder, D. Hazarika, E. Cambria, A. Gelbukh, and A. Hussain, "Multimodal sentiment analysis: addressing key issues and setting up the baselines," *IEEE Intelligent Systems*, vol. 33, no. 6, pp. 17–25, 2018.
[32] S. Poria, A. Hussain, and E. Cambria, *Multimodal Sentiment Analysis*, vol. 8, p. 5, Springer International Publishing, Cham, 2018.
[33] T. Diwan and J. V. Tembhurne, "Sentiment analysis: a convolutional neural networks perspective," *Multimedia Tools and Applications*, vol. 81, no. 30, pp. 44405–44429, 2022.
[34] H. Wang, X. Li, Z. Ren, M. Wang, and C. Ma, "Multimodal sentiment analysis representations learning via contrastive learning with condense attention fusion," *Sensors*, vol. 23, no. 5, Article ID 2679, 2023.
[35] P. Ramasubramanian, "Disaster management using deep learning on social media," *International Journal of Applied Science and Engineering*, vol. 18, no. 2, pp. 1–8, 2021.
[36] I. K. S. Al-Tameemi, M. R. Feizi-Derakhshi, S. Pashazadeh, and M. Asadpour, "Interpretable multimodal sentiment classification using deep multi-view attentive network of image and text data," *IEEE Access*, vol. 11, pp. 91060–91081, 2023.
[37] W. Liao, B. Zeng, J. Liu, P. Wei, and J. Fang, "Image-text interaction graph neural network for image-text sentiment analysis," *Applied Intelligence*, vol. 52, no. 10, pp. 11184–11198, 2022.
[38] X. Wang, J. He, Z. Jin, M. Yang, Y. Wang, and H. Qu, "M2lens: visualizing and explaining multimodal models for sentiment analysis," *IEEE Transactions on Visualization and Computer*, vol. 28, no. 1, pp. 802–812, 2021.
[39] K. Kim and S. Park, "AOBERT: All-modalities-in-One BERT for multimodal sentiment analysis," *Information Fusion*, vol. 92, pp. 37–45, 2023.