

Article

Multi-Level Online Public Opinion Sentiment Analysis Method Based on Text Features

Jian Zhao ¹, Yi Sun ¹, Dawei Xu ^{1,*}, Zhejun Kuang ^{1,*}, Lijuan Shi ², Zubin Zhang ¹ and Yong Zheng ¹

¹ School of Computer Science and Technology, Changchun University, Changchun 130022, China; zhaojian@ccu.edu.cn (J.Z.); sunyia2026@163.com (Y.S.); ywanecz14@163.com (Z.Z.); 241501491@mails.ccu.edu.cn (Y.Z.)

² School of Electronic and Information Engineering, Changchun University, Changchun 130022, China; shilj@ccu.edu.cn

* Correspondence: xudw@ccu.edu.cn (D.X.); kuangzhejun@ccu.edu.cn (Z.K.)

Abstract

With the rapid development of social media and online interactive platforms, online public opinion has become a vital information source for public emotional expression, social risk perception, and decision support. However, public opinion texts are typically characterized by short length, obscure semantics, complex emotional expressions, and strong context dependence, making it difficult for traditional lexicon-based or shallow neural network methods to achieve stable and robust performance in sentiment discrimination tasks. To address these issues, this paper proposes BERT-BiLSTM-MHSA-Capsule (BBMC), hereafter referred to as BBMC, an online public opinion sentiment analysis model based on multi-level semantic feature fusion. The model first utilizes the pretrained language model BERT to extract dynamic semantic representations with context-aware capabilities; subsequently, a Bidirectional Long Short-Term Memory (BiLSTM) network is employed to model the bidirectional temporal dependencies within the texts, while a Multi-Head Self-Attention (MHSA) mechanism is introduced to achieve adaptive focusing on key emotional information. Building upon this, a three-layer cascaded capsule network is constructed to achieve structured modeling of high-order emotional attributes through vector neurons and dynamic routing mechanisms, effectively mitigating the loss of spatial feature information caused by traditional pooling and fully connected structures. Experimental results on a manually annotated online public opinion dataset show that BBMC achieves better performance than the evaluated baseline models in terms of accuracy, recall, and F1-score. These results indicate the empirical effectiveness of the proposed task-oriented feature-integration strategy and capsule-based classification head for online public opinion sentiment analysis.

Keywords: online public opinion analysis; sentiment analysis; BERT; BiLSTM; capsule network



Received: 10 June 2026

Revised: 29 June 2026

Accepted: 2 July 2026

Published: 6 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

With the maturity of Web 2.0 technology and the widespread popularization of social media platforms, cyberspace has become an essential venue for the public to express opinions, emotions, and stances. Online public opinion texts not only reflect individual subjective attitudes toward social events but also influence public decision-making, market behavior, and social stability to a certain extent. Therefore, accurately mining latent sentiment tendencies from massive, unstructured online public opinion texts has become

a research hotspot in the fields of natural language processing and social computing. However, unlike traditional long texts or normalized corpora, online public opinion texts generally exhibit the following challenging characteristics:

- Semantics are dense, and emotions are often expressed through metaphors, irony, or contextual implications;
- Emotional expressions have strong context dependence, where the same word may present completely opposite emotional polarities under different event backgrounds;
- There are implicit temporal logics and emotional transmission relationships within the content of public opinion event comments.

In recent years, pretrained language models such as BERT have significantly improved the quality of text representation; however, their outputs are typically input into downstream classifiers as static features, which remains insufficient in depicting the temporal evolution and high-order attribute relationships of public opinion sentiments. Although models based on BERT and Transformers have achieved breakthroughs in capturing contextual semantics, their classification heads usually adopt max-pooling or average-pooling strategies. This scalar compression process inevitably leads to the loss of spatial relationships and emotional intensity attributes between features.

The study focuses on how established representation-learning components can be organized into a task-oriented multi-level framework for online public opinion sentiment analysis. In practical public opinion monitoring, sentiment classification is still often conducted through semi-manual procedures, including event-keyword screening, manual text inspection, and subjective sentiment judgment. Although such procedures are useful for small-scale analysis, they are labor-intensive, difficult to scale to massive volumes of online comments, and susceptible to inconsistency among analysts. Therefore, the scientific and engineering value of this study lies in constructing an automated sentiment classification framework that adapts established NLP components to the practical requirements of online public opinion analysis, including contextual ambiguity, implicit sentiment expression, sentiment-keyword focusing, and category-level sentiment judgment.

The main contributions of this paper are summarized as follows:

- A task-oriented multi-level sentiment analysis framework, BBMC, is developed for online public opinion texts. The framework sequentially integrates contextual semantic representation, bidirectional temporal modeling, attention-based key-feature selection, and capsule-based classification, so that different types of sentiment-related information can be represented at different levels.
- In BBMC, a three-layer cascaded capsule classification head is introduced after the BERT-BiLSTM-MHSA feature extractor. Rather than using conventional pooling or a simple fully connected classifier, the proposed classification head progressively aggregates local semantic features into category-level capsule representations. Its contribution is further examined through module-level and capsule-depth ablation experiments.
- A manually annotated online public opinion dataset is constructed, and the proposed framework is evaluated against traditional machine learning, convolutional, pretrained language model, and hybrid neural network baselines. The experimental results provide empirical evidence for the effectiveness of the proposed integration strategy on the evaluated dataset.

2. Related Work

2.1. Online Public Opinion Sentiment Analysis

The paradigm of online public opinion analysis has transitioned from early feature engineering to deep semantic representation centered on pretrained language models. Representing foundational work, Gueddes et al. [1] enhanced BERT's semantic modeling in social media contexts by integrating sentiment ontologies, while Bahari et al. [2] provided a systematic review of hyperparameter optimization for BERT-based sentiment analysis. Addressing the complexity of social media content, Amirthasaravanan et al. [3] utilized attention mechanisms to achieve effective multimodal feature fusion.

To further mitigate feature sparsity and long-range dependencies, hybrid architectures have gained prominence. Kirti et al. [4] demonstrated the synergy between LSTM and BERT in capturing contextual nuances, Iyer et al. [5] incorporated multi-level attention into a hybrid CNN-BiLSTM model, whereas Yun et al. [6] tailored a BERT-BiLSTM-CNN architecture for Chinese Weibo. To tackle data incompleteness, Yang [7], Afroze [8], and Ji [9] proposed robust solutions involving token-level reconstruction, cross-attention, and feature relevance fusion; concurrently, Shi et al. [10] focused on text-guided reconstruction networks to handle missing modalities. Furthermore, Ping et al. [11] explored the integration of causal reasoning, Wang et al. [12] proposed a knowledge-enhanced semantic composition model, and Tarik et al. [13] optimized multimodal analysis efficiency through innovative feature selection algorithms.

2.2. Application of Capsule Networks in Text Sentiment Analysis

Pooling operations in traditional convolutional architectures often precipitate the irreversible loss of spatial hierarchical information. In contrast, capsule networks preserve spatial constraints between local features and global structures via vector neurons. Patrick et al. [14] provided a comprehensive survey of these advantages, while the potential for extracting fine-grained structural features was further validated in architectures integrating capsule and residual networks [15].

In the NLP domain, capsule networks have demonstrated robust hierarchical feature modeling. Gauraha et al. [16] proposed a capsule-enhanced RoBERTa model, and Pamulaparthi et al. [17] developed a cohesive Transformer-Capsule model based on contextual embeddings. To optimize the structured modeling of dynamic temporal features, the BiERU-CapsNet [18] introduced an integrated scheme combining Bidirectional Emotional Recurrent Units with capsules. For specific applications, the dynamic routing mechanism of capsule networks has shown significant advantages in detecting hate speech on Twitter [19] and sentiment classification for depression identification [20]. These developments provide the theoretical impetus for the multi-level cascaded capsule network proposed in this study to achieve refined emotional feature extraction.

3. Model Construction

3.1. Architecture

This study constructs a multi-level online public opinion sentiment analysis framework, realizing the crawling, processing, and sentiment analysis of public opinion comments to improve the accuracy of sentiment analysis for Weibo comments related to specific public opinion events and provide a scientific basis for guiding and governing negative public opinion. The model framework and specific experimental steps are shown in Figure 1.

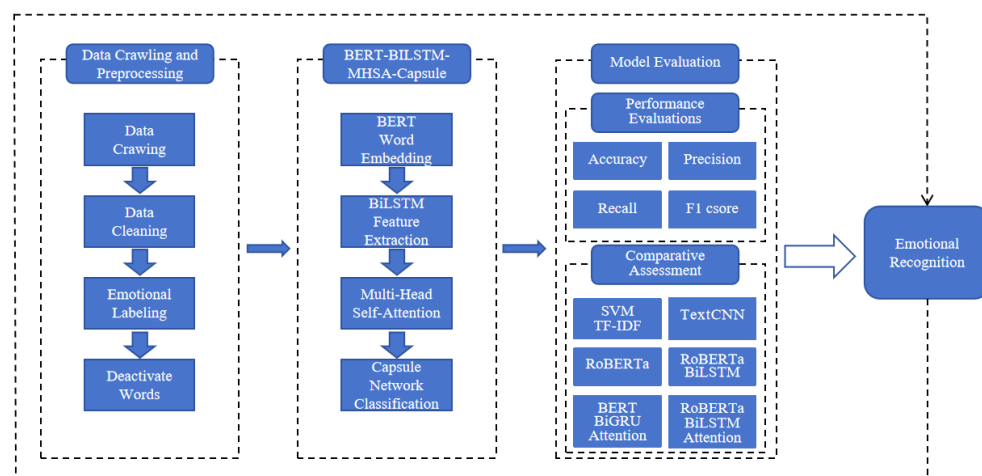


Figure 1. The systematic architecture and experimental flowchart of online public opinion sentiment analysis.

3.2. Data Collection and Preprocessing

The architecture of this model primarily requires data from two aspects, one being Weibo—a public platform for emotional expression and the world’s largest Chinese social networking platform. With its high level of data openness, obtaining relevant data from it for sentiment analysis holds guiding significance for the governance of online public opinion and the promulgation of measures. The specific collection process is shown in Figure 2.

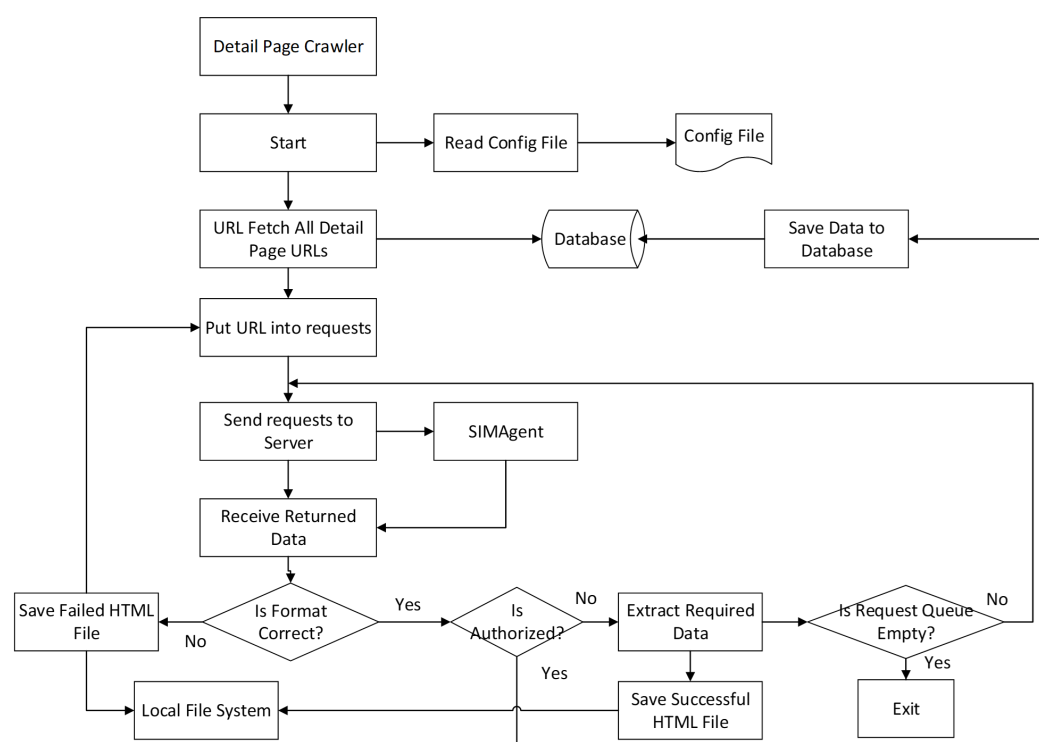


Figure 2. Flowchart of data collection and preprocessing.

3.3. BBMC Model

In the context of online public opinion, the objective of the sentiment analysis task is to extract latent stances and emotional tendencies from unstructured text. Let the

dataset $\mathcal{D} = \{(X_i, y_i)\}_{i=1}^n$ contain n public opinion samples, where X_i denotes the i -th text sample and $y_i \in \{1, \dots, K\}$ denotes its sentiment label. For an input token sequence $X = [w_1, w_2, \dots, w_L]$, the goal of the model is to learn a mapping function $\mathcal{F} : X \rightarrow y$, where K is the number of sentiment categories. This subsection presents BBMC, a multi-level public opinion sentiment analysis framework designed to address obscure semantics and the loss of spatial feature position information in public opinion texts.

Figure 3 shows the comprehensive architecture of BBMC. The framework consists of four primary components: (1) a pretrained BERT layer for extracting dynamic, context-aware semantic representations; (2) a BiLSTM layer to capture long-range temporal dependencies and emotional transmission logic; (3) an MHSA mechanism designed to adaptively focus on salient emotional indicators; and (4) a three-layer cascaded capsule network that utilizes vector neurons and dynamic routing to preserve spatial hierarchical relationships and structural information of high-order sentiment attributes. The rationale of this architecture is not to treat each component as an independent methodological novelty, but to assign a specific representational role to each level of the model. BERT provides contextual token representations, BiLSTM further models sequential dependencies in public opinion comments, MHSA highlights sentiment-relevant tokens or phrases, and the cascaded capsule network transforms the resulting features into class-level vector representations.

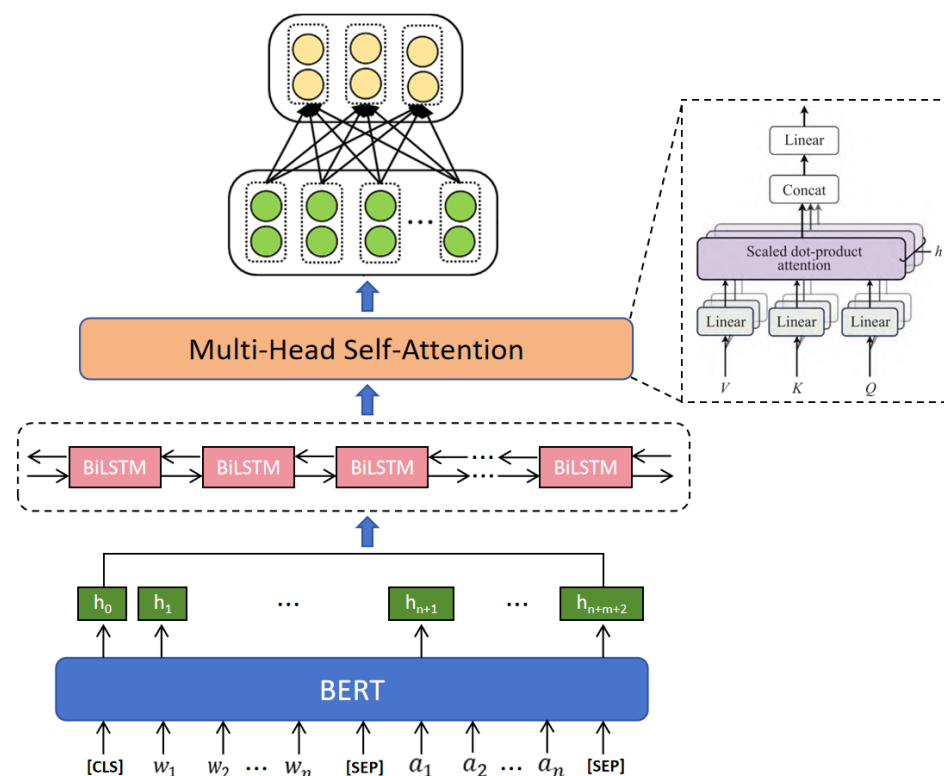


Figure 3. Structural diagram of BBMC.

3.3.1. Dynamic Contextual Semantic Representation Layer

Traditional embedding methods cannot handle polysemy and sentiment reversals in specific contexts within public opinion texts. The first layer of the model employs the Bidirectional Encoder Representations from Transformers. By encoding the input sequence through multiple Transformer blocks, the generated initial semantic features H_{bert} can be expressed as:

$$H_{\text{bert}} = \text{BERT}(x_1, x_2, \dots, x_L) \in \mathbb{R}^{L \times d_{\text{model}}} \quad (1)$$

where $H_{\text{bert}} = [h_1^{\text{bert}}, h_2^{\text{bert}}, \dots, h_L^{\text{bert}}]$, h_t^{bert} denotes the contextual representation of the t -th token, and d_{model} represents the hidden layer dimension. In the BBMC implementation, the BERT encoder was initialized from the pretrained checkpoint bert-base-chinese. This checkpoint contains 12 Transformer encoder layers, a hidden dimension of 768, and 12 self-attention heads. Tokenization was performed using the tokenizer and vocabulary distributed with the same checkpoint. Input sequences were truncated or padded to a maximum length of 128 tokens. Through the Masked Language Model task, BERT learns rich lexical and grammatical patterns, providing a high-dimensional continuous semantic space for subsequent sentiment inference.

3.3.2. Public Opinion Bidirectional Temporal Memory Network Layer

Data comments in online public opinion exhibit significant contextual logical correlations. To capture the bidirectional long-distance dependencies within public opinion texts, a BiLSTM network is introduced.

$$i_t = \sigma(W_i[h_{t-1}, h_t^{\text{bert}}] + b_i) \quad (2)$$

$$f_t = \sigma(W_f[h_{t-1}, h_t^{\text{bert}}] + b_f) \quad (3)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_C[h_{t-1}, h_t^{\text{bert}}] + b_C) \quad (4)$$

$$o_t = \sigma(W_o[h_{t-1}, h_t^{\text{bert}}] + b_o) \quad (5)$$

$$h_t = o_t \odot \tanh(C_t) \quad (6)$$

Finally, by concatenating the forward and backward hidden states at each time step, the temporal-enhanced features are obtained as $H_{\text{lstm}} = [\vec{h}_t \oplus \overleftarrow{h}_t]_{t=1}^L \in \mathbb{R}^{L \times 2d_{\text{hidden}}}$, where $\vec{h}_t$ and $\overleftarrow{h}_t$ denote the forward and backward hidden states, respectively. This layer effectively filters redundant background information in public opinion texts through the forget gate, preserving the core logic of emotional evolution.

3.3.3. Multi-Head Attention Mechanism

In public opinion texts, sentiment tendencies are often determined by a few key terms such as “anger” or “agreement.” To automatically identify these salient signals, the MHSA mechanism is introduced. H_{lstm} is mapped into the query space, key space, and value space:

$$Q = H_{\text{lstm}} W^Q \quad (7)$$

$$K = H_{\text{lstm}} W^K \quad (8)$$

$$V = H_{\text{lstm}} W^V \quad (9)$$

The weight distribution scores are calculated using Scaled Dot Product Attention:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \quad (10)$$

$$Z = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (11)$$

where d_k serves as the scaling factor to prevent the vanishing gradient problem caused by excessively large dot product results. W^O denotes the output projection matrix, which is responsible for integrating the multi-dimensional public opinion sentiment distribution features extracted by h attention heads.

The advantage of the multi-head attention mechanism lies in its ability to capture dependencies between different positions in the input sequence, and it significantly improves learning efficiency through parallel computation. Through multiple attention heads, the model can independently learn and capture different features in multiple representation subspaces, thereby enhancing the model's expressiveness and robustness.

3.3.4. Three-Layer Cascaded Capsule Network Clustering Layer

By introducing vector neuron capsules to replace scalar neurons in traditional CNNs, the capsule network redesigns the connection structure between layers, thereby better preserving the spatial and hierarchical structural information of the text. In each layer, the features represented by capsules are expanded from one-dimensional vectors to multi-dimensional vectors and are optimized via the dynamic routing algorithm. During the model training process, the dynamic routing mechanism continuously adjusts the coupling coefficients between the child capsule layer and the parent capsule layer, optimizing their inter-dependencies. Furthermore, a squash function is employed to jointly determine which textual information should be transmitted from the child capsule layer to the parent capsule layer, thereby minimizing information loss and enhancing feature extraction capabilities. Ultimately, the high-level feature representation of the text is obtained by merging and concatenating all vectors in the parent capsule layer.

To address the loss of structural information caused by traditional fully connected layers when processing complex feature space relationships, this paper constructs a three-layer cascaded capsule network. Unlike traditional scalar neurons, cascaded capsules utilize vector neurons to model sentiment attributes, where the length of the vector represents the probability of a specific sentiment category, and the direction of the vector encodes the attribute features of the sentiment. Vector mapping from features to attributes: The sentiment prediction vector mapped from low-level capsules to high-level capsules is denoted as $u_{j|i}$. To achieve precise polarity clustering, the model employs a dynamic routing algorithm to iteratively update the coupling coefficients c_{ij} . The coupling coefficients c_{ij} are iteratively updated by the dynamic routing algorithm based on prediction agreement:

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \quad (12)$$

Weighted Combination: A weighted combination of prediction vectors from low-level capsules is performed to learn the mapping relationship from local components to the global representation.

$$s_j = \sum_i c_{ij} \hat{u}_{j|i} \quad (13)$$

Non-linear Squashing Function: To maintain the orientation of sentiment features while normalizing the vector magnitude, the squash function is utilized:

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \cdot \frac{s_j}{\|s_j\|} \quad (14)$$

Coupling Coefficient Update: In each iteration, the coupling coefficients are updated by calculating the directional agreement between the prediction vectors of the low-level capsules and the target capsules. When the prediction vector aligns with the direction of the target capsule, the connection strength is enhanced.

$$b_{ij} \leftarrow b_{ij} + \langle \hat{u}_{j|i}, v_j \rangle \quad (15)$$

The classification capsule layer represents the probability of each category through the length of the final output capsule vectors. Each category is represented by a specific capsule, where the probability of the category is proportional to the magnitude of the corresponding squashed capsule output. The calculation formula for classification is as follows:

$$P(y = i|x) = \frac{\exp(\|v_i\|)}{\sum_{k=1}^K \exp(\|v_k\|)} \quad (16)$$

Ultimately, a three-layer cascaded capsule network is employed to achieve step-by-step clustering and determination of public opinion text, transitioning from local semantic fragments to global sentiment entities.

4. Results and Discussion

4.1. Dataset Sources and Experimental Setup

The experimental dataset comprises 20,000 real-world public opinion entries collected from mainstream social media platforms via a self-developed distributed crawler. After removing duplicated posts, advertisements, irrelevant content, and entries with excessive missing or unreadable text, 17,224 valid samples were retained, categorized into negative (7454), positive (7119), and neutral (2651) classes. To ensure high-quality annotation, we implemented a double-blind labeling process involving three linguistics graduate students. Before annotation, all annotators were provided with unified labeling guidelines. Comments expressing support, approval, satisfaction, or positive emotional tendency toward the public opinion event were labeled as positive; comments expressing criticism, dissatisfaction, anger, anxiety, or other negative emotional tendency were labeled as negative; and comments mainly describing facts, forwarding information, or expressing no clear emotional polarity were labeled as neutral. Two primary annotators independently labeled the data according to these criteria, while a third expert served as an arbitrator to resolve any discrepancies. The inter-rater reliability was validated using Cohen's Kappa coefficient, achieving a score of 0.82, which indicates substantial agreement and ensures the reliability of the empirical results. Finally, the dataset was partitioned into training, validation, and testing sets with a ratio of 8:1:1, as detailed in Table 1. The dataset shows a certain degree of class imbalance, with the neutral class accounting for 15.40% of the corpus. Public benchmark datasets are valuable for evaluating model generalization; however, many existing sentiment benchmarks differ from online public opinion monitoring scenarios in terms of data source, event context, annotation granularity, and discourse characteristics. Since this study focuses on event-oriented online public opinion sentiment classification, the main experiments were conducted on the manually annotated public opinion dataset constructed for this application scenario. To preserve the original distribution of real-world online public opinion data, we did not apply oversampling, undersampling, or synthetic resampling. Instead, the train/validation/test split was performed while maintaining the original class distribution, and macro-F1 was used as the validation metric to reduce the dominance of majority classes during model selection.

Table 1. Statistical distribution of the experimental dataset.

Sentiment Category	Sample Size	Proportion (%)	Average Length
Negative	7454	43.27	128
Positive	7119	41.33	115
Neutral	2651	15.40	86
Total	17,224	100	112

The experiments were implemented using the PyTorch 2.0 framework, with the hardware computations conducted on an NVIDIA Tesla V100 GPU (32 GB) (NVIDIA, Santa Clara, CA, USA). To improve reproducibility, the preprocessing procedure, annotation criteria, data split ratio, pretrained checkpoints, optimization strategy, early-stopping rule, and random seeds are explicitly reported in this section and Table 2. For reproducibility, the exact pretrained BBMC and the BERT-based ablation models used `bert-base-chinese`, whereas the RoBERTa, RoBERTa-BiLSTM, and RoBERTa-BiLSTM-Attention baselines used `hfl/chinese-roberta-wwm-ext`. The pretrained tokenizer corresponding to each checkpoint was used without vocabulary modification. Each model was trained for a maximum of 20 epochs. Model performance on the validation set was evaluated after each epoch, and the checkpoint with the highest validation macro-F1-score was retained for test-set evaluation. No class-weighted loss was used in the reported experiments, so that all compared models were evaluated under the same data distribution and optimization protocol. Early stopping was applied with a patience of three epochs if the validation macro-F1-score did not improve. The BERT parameters and the newly initialized downstream layers were optimized using learning rates of 2×10^{-5} and 1×10^{-3} , respectively. Weight decay was set to 0.01, and dropout with a probability of 0.3 was applied to the BiLSTM output and downstream classification layers. To account for stochastic variation, each model was independently trained and evaluated using five random seeds: 42, 52, 62, 72, and 82. For each seed, the same train/validation/test split and experimental protocol were used across all models. Accuracy, precision, recall, and F1-score are reported as the mean and standard deviation over the five runs. To further assess whether the observed improvements are statistically reliable, paired two-tailed *t*-tests were conducted based on the seed-level results. In addition, 95% confidence intervals were calculated for the paired mean differences between BBMC and the strongest comparison models.

Table 2. Core hyperparameter settings of the model.

Parameter Name	Parameter Value
Batch Size	32
BERT Learning Rate	2×10^{-5}
Other Layer Learning Rate	1×10^{-3}
BiLSTM Hidden Size	256
Attention Heads (<i>h</i>)	8
Routing Iterations (<i>r</i>)	3
Primary Capsule	8 caps $\times$ 64 dim
Intermediate Capsule	4 caps $\times$ 32 dim
Class Capsule	3 caps $\times$ 16 dim
Optimizer	AdamW
Maximum Epochs	20
Validation Metric	Macro-F1
Early-Stopping Criterion	No improvement in validation macro-F1
Early-Stopping Patience	3 epochs
Model-Selection Criterion	Highest validation macro-F1
Dropout Rate	0.3
Weight Decay	0.01
Random Seeds	42, 52, 62, 72, 82
Statistical Test	Paired two-tailed <i>t</i> -test over five seeds
Confidence Interval	95% CI of paired mean differences
BERT Checkpoint	<code>bert-base-chinese</code>
RoBERTa Checkpoint	<code>hfl/chinese-roberta-wwm-ext</code>
Maximum Sequence Length	128

4.2. Comparison with Baseline Models

To evaluate the performance of the proposed BBMC model, we selected representative baselines from different modeling levels, including traditional machine learning, convolutional neural networks, pretrained Transformer encoders, and Transformer-based hybrid architectures. Specifically, SVM + TF-IDF was used as a traditional feature-engineering baseline, TextCNN was used as a neural convolutional baseline, RoBERTa was used as a strong pretrained Transformer-based baseline, and RoBERTa-BiLSTM and RoBERTa-BiLSTM-Attention were included as enhanced Transformer-based hybrid baselines. This baseline design was intended not only to compare BBMC with standard sentiment classifiers, but also to examine whether the proposed capsule-based classification head provided additional benefit beyond pretrained Transformer representation, sequential modeling, and attention-enhanced feature extraction. For a fair comparison, all models were implemented using the same preprocessing pipeline, data split, experimental protocol, and hardware environment described previously. The comparative experimental results are presented in Table 3.

Table 3. Performance comparison of different models on the public opinion dataset.

Model Name	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM + TF-IDF	77.42 ± 0.17	77.68 ± 0.21	77.03 ± 0.13	77.33 ± 0.18
TextCNN	82.35 ± 0.17	82.64 ± 0.23	81.95 ± 0.13	82.24 ± 0.16
RoBERTa	86.24 ± 0.19	86.55 ± 0.25	85.83 ± 0.18	86.13 ± 0.18
RoBERTa-BiLSTM	89.07 ± 0.15	89.38 ± 0.21	88.66 ± 0.14	88.95 ± 0.15
RoBERTa-BiLSTM-Attention	90.99 ± 0.19	91.31 ± 0.26	90.56 ± 0.19	90.86 ± 0.18
BBMC (Ours)	94.23 ± 0.17	93.85 ± 0.16	93.29 ± 0.16	93.56 ± 0.16

According to the repeated-run results presented in Table 3 and Figure 4, BBMC achieves the highest average performance among all evaluated models across the four core evaluation metrics. Traditional models like SVM are limited by manual feature engineering, struggling to capture complex semantic correlations, while TextCNN shows inherent deficiencies in modeling long-distance dependencies despite its local feature extraction capabilities. Although RoBERTa-based models show clear improvements over conventional neural architectures due to their contextual semantic representations, BBMC achieves further gains over the selected Transformer-based and hybrid baselines under the same experimental setting. Compared with the strongest baseline, RoBERTa-BiLSTM-Attention, the paired *t*-test over five random seeds shows that BBMC improves accuracy by 3.24 percentage points, with a 95% confidence interval of 2.97–3.52 percentage points ($p < 0.001$), and improves F1-score by 2.70 percentage points, with a 95% confidence interval of 2.44–2.97 percentage points ($p < 0.001$). Specifically, by replacing traditional pooling layers with a cascaded capsule network, BBMC utilizes vector neurons and a dynamic routing mechanism to precisely capture spatial hierarchical relationships. This approach effectively mitigates feature information loss and proves that BBMC demonstrates enhanced robustness and significant performance gains when navigating complex public opinion landscapes, including those involving metaphor and irony.

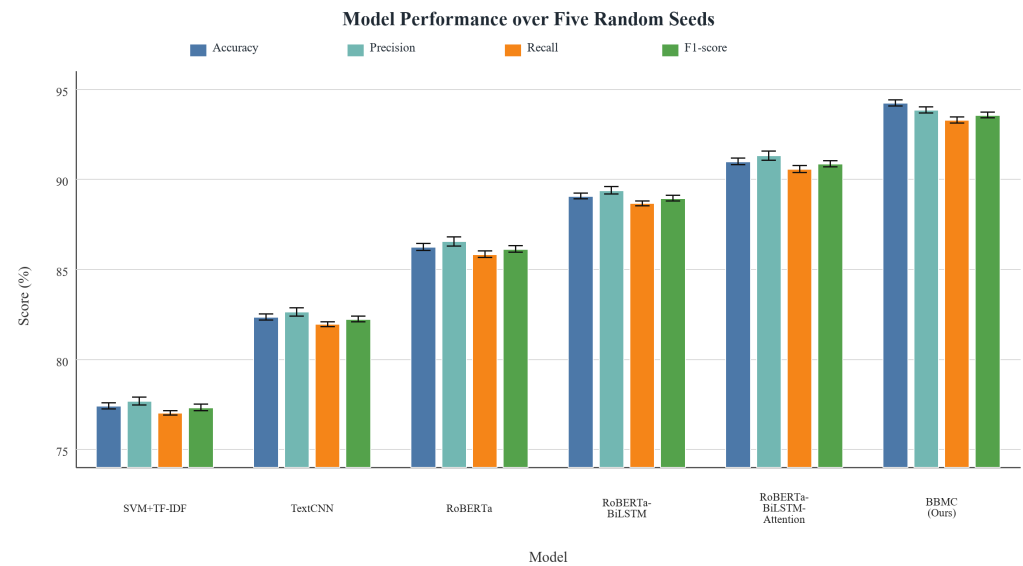


Figure 4. Performance comparison of different sentiment analysis models.

4.3. Ablation Study Analysis

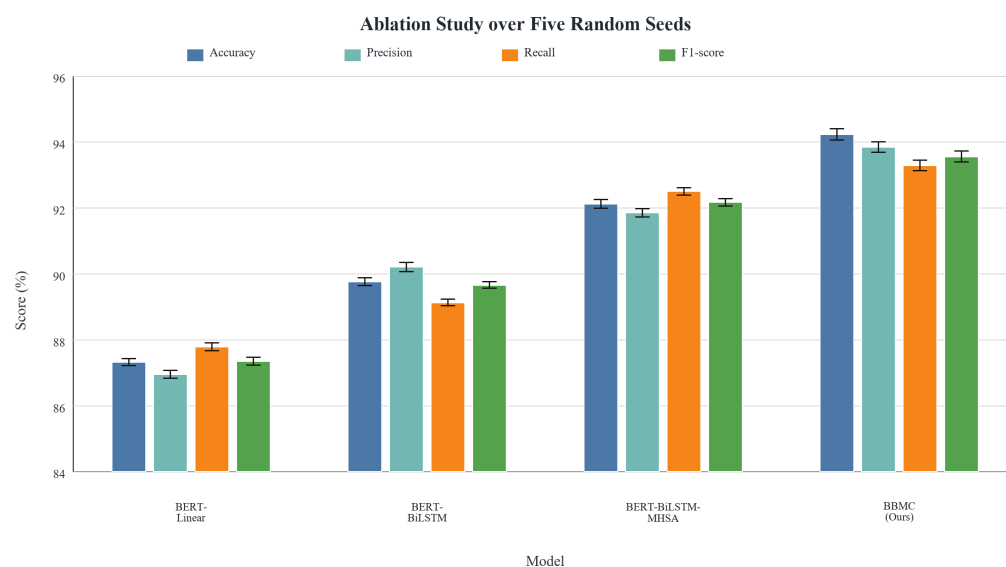
To investigate the specific contribution of each module to the overall performance of the model, four groups of comparative models were designed for the ablation study.

As shown in Table 4 and Figure 5, all ablation results are reported as mean $\pm$ standard deviation over five independent runs. Compared with BERT-Linear, BERT-BiLSTM improves the average F1-score from 87.35% to 89.66%, indicating that bidirectional temporal modeling contributes to capturing contextual dependencies in online public opinion texts. This indicates that although online public opinion texts are relatively short, they still possess significant internal logical connections and emotional transmission relationships. Through its bidirectional structure, BiLSTM effectively captures these long-distance dependencies, thereby enhancing the depth of feature representation. Following the introduction of the MHSA mechanism, all performance metrics of the model are further improved. This verifies that within complex public opinion texts, the contributions of different words to sentiment polarity are not uniform. MHSA adaptively assigns higher weights to key emotional signals, enabling the precise extraction of core features amidst noise interference.

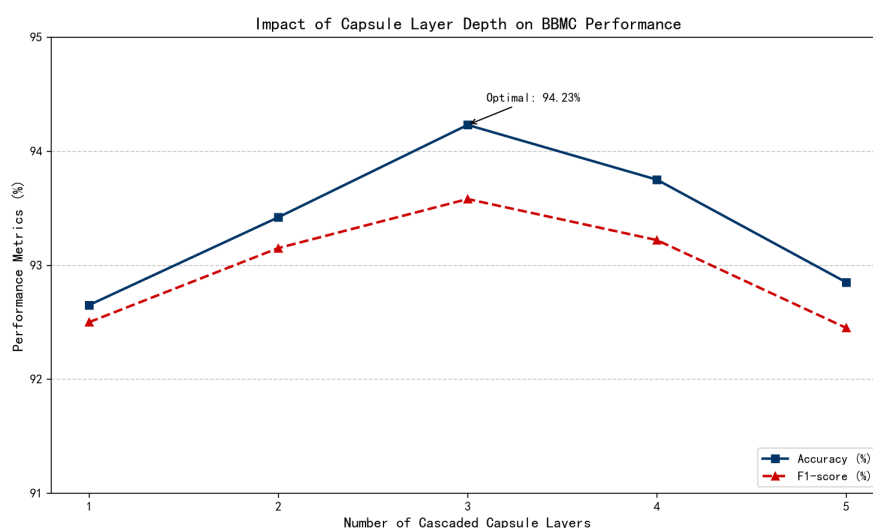
Compared with BERT-BiLSTM-MHSA, BBMC increases the average accuracy from 92.12% to 94.23% and the average F1-score from 92.17% to 93.56% after introducing the three-layer cascaded capsule classification head. The paired *t*-test further indicates that the improvement introduced by the capsule classification head is statistically significant, with an accuracy gain of 2.11 percentage points, a 95% confidence interval of 1.98–2.24 percentage points ($p < 0.001$), and an F1-score gain of 1.39 percentage points, along with a 95% confidence interval of 1.28–1.50 percentage points ($p < 0.001$). This significant improvement demonstrates the clear advantages of capsule networks over traditional scalar neurons and fully connected structures; traditional classification heads typically employ pooling operations, which lead to the loss of relative positional and spatial hierarchical information between words. In contrast, the BBMC model utilizes vector neurons and a dynamic routing mechanism to successfully achieve step-by-step clustering, transitioning from local semantic fragments to global sentiment entities. This structured modeling approach effectively mitigates sentiment determination errors caused by implicit semantics and strong contextual dependency in public opinion analysis, fully demonstrating the superiority of BBMC in handling complex sentiment attribute mappings.

Table 4. Ablation results of different BBMC variants on the public opinion dataset.

Model Name	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
BERT-Linear	87.32 ± 0.11	86.95 ± 0.12	87.78 ± 0.12	87.35 ± 0.12
BERT-BiLSTM	89.76 ± 0.12	90.21 ± 0.14	89.13 ± 0.10	89.66 ± 0.10
BERT-BiLSTM-MHSA	92.12 ± 0.13	91.85 ± 0.13	92.50 ± 0.11	92.17 ± 0.11
BBMC (Ours)	94.23 ± 0.17	93.85 ± 0.16	93.29 ± 0.16	93.56 ± 0.16

**Figure 5.** Results of the ablation study.

To further verify the rationality of the proposed three-layer cascaded capsule network in processing the spatial relationships of public opinion features, comparative experiments were conducted by varying the number of cascaded capsule layers while keeping the parameters of BERT, BiLSTM, and MHSA constant. The experimental results are illustrated in Figure 6.

**Figure 6.** Impact of the number of capsule layers on BBMC performance.

When the number of layers is 1–2, although the model can perform preliminary feature modeling using vector neurons, its capacity for structural modeling of complex sentiment attributes remains insufficient due to the shallow architecture, resulting in an accuracy of 92.10%—lower than that of the three-layer configuration. Model performance

reaches its peak at 94.23% when the depth is set to three layers. This demonstrates that a three-layer cascaded structure best aligns with the semantic hierarchical logic of public opinion texts: (1) the Primary Capsule Layer receives and vectorizes features extracted by the MHSA; (2) the Intermediate Capsule Layer performs preliminary sentiment feature aggregation via the dynamic routing mechanism; and (3) the Sentiment Class Capsule Layer completes the final polarity determination. This architecture effectively mitigates the spatial information loss typically caused by traditional pooling operations. However, model performance declines significantly when the depth increases to four layers or more. This degradation can be attributed to two main factors: first, since capsule networks rely on iterative dynamic routing, an excessively deep structure causes computational drift of coupling coefficients during transmission, thereby introducing unnecessary routing noise. Second, the substantial increase in parameters prone to inducing overfitting on the dataset of 17,224 public opinion samples leads to a weakened generalization capability.

To further investigate the discriminative accuracy of the BBMC model across different sentiment polarities, this study examines the normalized confusion matrix on the test set. This per-class analysis was included because the neutral class is less frequent than the positive and negative classes, and therefore overall accuracy alone may not fully reflect the model's behavior on minority-class samples. As shown in Figure 7, the model exhibits exceptional performance on “Positive” and “Negative” samples—categories representing strong emotional intensity—achieving recognition accuracies of 95.2% and 94.8%, respectively. This success is primarily attributed to BERT's capability in capturing deep semantics and the capsule network's precise clustering of sentiment entity attributes. Although the recognition rate for “Neutral” samples of 92.3% is slightly lower than that of the two polar extremes, a tracking analysis of False Positives reveals that negative samples misclassified as neutral often contain subtle irony or suffer from semantic loss due to internet abbreviations. Nevertheless, the confusion matrix indicates that neutral samples remain more difficult to distinguish than positive and negative samples, which is consistent with the relatively lower proportion and weaker emotional intensity of neutral texts in the dataset. The results suggest that the three-layer cascaded capsule network, facilitated by the dynamic routing algorithm, helps improve the discrimination of sentiment categories with blurred boundaries, although further comparison using baseline confusion matrices would be needed to quantify category-level confusion reduction more precisely.

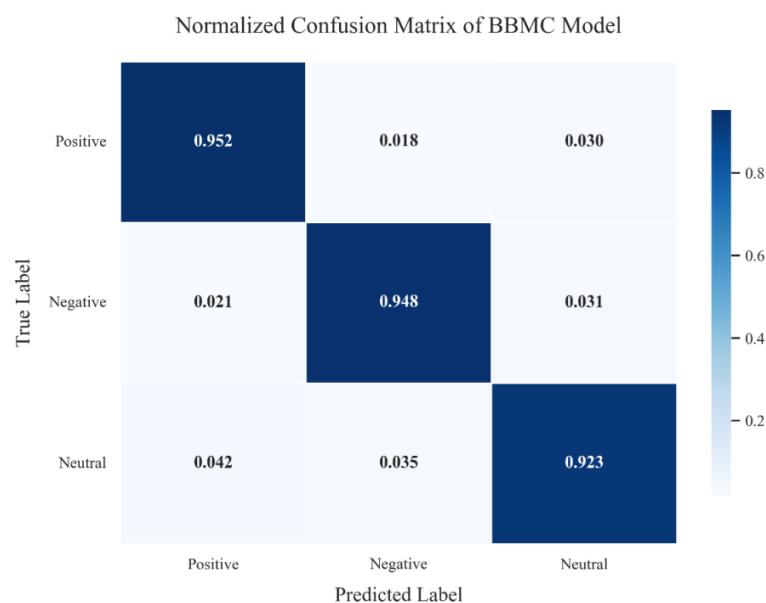


Figure 7. Confusion matrix of BBMC.

5. Conclusions

5.1. Summary of Major Research Findings

To address challenges such as implicit semantics, complex emotional expressions, and context dependency in online public opinion texts, this paper proposes BBMC, a multi-level sentiment analysis framework for online public opinion sentiment classification. Based on experimental validation using 17,224 processed real-world public opinion data samples, the following conclusions are reached:

1. **Effectiveness of Multi-level Feature Fusion:** Experimental results over five independent runs show that the BBMC model achieves an average accuracy of $94.23\% \pm 0.17$ and an average F1-score of $93.56\% \pm 0.16$. Compared to traditional baseline models such as TextCNN and RoBERTa, the proposed model demonstrates superior precision and stability when processing complex semantic logic.
2. **Structural Feature Capture Capability:** Ablation studies confirm that the introduction of a cascaded capsule network effectively overcomes the limitations of traditional fully connected layers regarding the loss of feature space information. Through the dynamic routing mechanism, the model can accurately cluster local semantic capsules into global sentiment entities, significantly enhancing its ability to identify metaphorical scenarios such as irony and double negatives.

5.2. Limitations and Future Work

Although the proposed model has achieved significant performance improvements, certain limitations remain to be addressed in future research:

1. **Computational Efficiency and Optimization:** While the BBMC model performs excellently on the current corpus, the multiple iterations of dynamic routing result in a higher computational overhead compared to linear classifiers. Consequently, the response latency in large-scale real-time stream monitoring scenarios requires further optimization.
2. **Cross-dataset Generalization:** The current evaluation is mainly based on a manually annotated online public opinion dataset. Although this dataset reflects the practical characteristics of event-oriented public opinion monitoring, additional experiments on public benchmark datasets would further strengthen the evidence for model robustness and generalization. Future work will evaluate BBMC on publicly available sentiment analysis benchmarks and cross-domain public opinion datasets.
3. **Multimodal Data Fusion:** Current forms of public opinion expression are increasingly diverse; beyond text, images, short videos, and internet memes carry substantial emotional weight. This study currently focuses solely on textual features. Future work will explore multimodal fusion technologies to achieve a more comprehensive perception of online sentiment.
4. **Integration of External Knowledge Bases:** Internet slang and “memes” evolve rapidly, often involving domain-specific abbreviations. Future research plans to introduce knowledge graphs or external semantic lexicons to enhance the model’s depth of understanding regarding background knowledge in specific vertical domains (e.g., financial or medical public opinion).

Author Contributions: Y.S.: Conceptualization, methodology, data collection, data curation, original draft preparation, writing—review and editing; D.X. and Z.K.: data curation, original draft preparation, writing—review and editing; J.Z. and L.S.: supervision, project administration, writing—review and editing; Z.Z. and Y.Z.: conceptualization, original draft preparation, writing—review and editing, supervision, project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Jilin Scientific and Technological Development Program (No. 20260203028SF) and the Science and Technology Innovation Project of the Changchun Talent Support Program (No. 2025ccyc20).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw social media data used in this study cannot be publicly released due to platform terms of service and privacy considerations. The processed and anonymized data, annotation guidelines, and experimental details are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gueddes, A.; Mahjoub, M.A. BERT-OntoSent: Combining BERT language model with sentiment ontology for enhanced sentiment analysis on social media. *J. Inf. Telecommun.* **2026**, *10*, 1–23. <https://doi.org/10.1080/24751839.2025.2528363>.
2. Bahari, A.R.S.; Utomo, F.S.; Berililana, B. Systematic Review of Hyperparameter Adjustment and Evaluation Metrics in BERT-Based Sentiment Analysis. *J. La Multiapp* **2026**, *7*, 312–325. <https://doi.org/10.37899/journallamultiapp.v7i2.3046>.
3. Amirthasaravanan, A.; Choudhry, M.D.; Jeevanandham, S.; Ramasamy, S.; Suvalakshmi, K.; Karthikeswaran, D. Enhanced sentiment analysis on social media using BERT and multimodal attention-based fusion. *Automatika* **2025**, *67*, 69–90. <https://doi.org/10.1080/00051144.2025.2598897>.
4. Kirti; Divya; Kaur, S.; Gufran, S.; Singh, D.; Mishra, D.D. Sentiment Analysis Using LSTM and BERT: A Comparative Analysis Using Natural Language Processing Algorithms. In *Human-Smart City Interactions and User-Citizen Experiences*; Springer: Cham, Switzerland, 2026; pp. 131–154. https://doi.org/10.1007/978-3-032-06364-9_7.
5. Iyer, Y.; Gnanaprasanambikai, L. A Hybrid GloVe-BERT Fusion Model with Multi-Level Attention-Based CNN-BiLSTM for Sentiment Analysis. *ECTI Trans. Comput. Inf. Technol.* **2026**, *20*, 15–25. <https://doi.org/10.37936/ecti-cit.2026201.263366>.
6. Yun, C.H.; Rong, H.R. Sentiment Analysis of Chinese Weibo Based on BERT-BiLSTM-CNN. In *2026 6th International Conference on Neural Networks, Information and Communication Engineering (NNICE)*; IEEE: Piscataway, NJ, USA, 2026; pp. 858–861. <https://doi.org/10.1109/NNICE68970.2026.11465567>.
7. Yang, Z.; He, Q.; Yu, M.; Du, N.; Lu, Y. TCTR: Text-Guided Contrastive Learning with Token-Level Reconstruction Network for missing modalities in multimodal sentiment analysis. *Inf. Fusion* **2026**, *126*, 103571. <https://doi.org/10.1016/j.inffus.2025.103571>.
8. Afroze, S.; Hossain, M.R.; Hoque, M.M.; Siddique, N. MulMoSenT: Multimodal Sentiment Analysis for a Low-Resource Language Using Textual-Visual Cross-Attention and Fusion. *Inf. Fusion* **2026**, *131*, 104129. <https://doi.org/10.1016/j.inffus.2026.104129>.
9. Ji, M.; Zhou, J.; Wei, N. AFR-BERT: Attention-based mechanism feature relevance fusion multimodal sentiment analysis model. *PLoS ONE* **2022**, *17*, e0273936. <https://doi.org/10.1371/journal.pone.0273936>.
10. Shi, P.; Hu, M.; Nakagawa, S.; Zheng, X.; Shi, X.; Ren, F. Text-Guided Reconstruction Network for Sentiment Analysis With Uncertain Missing Modalities. *IEEE Trans. Affect. Comput.* **2025**, *16*, 1825–1838. <https://doi.org/10.1109/TAFFC.2025.3541743>.
11. Ping, P.; Huang, C.; Ding, W.; Liu, Y.; Miyajima, C.; Takeda, K. Distracted driving detection based on the fusion of deep learning and causal reasoning. *Inf. Fusion* **2023**, *89*, 121–142. <https://doi.org/10.1016/j.inffus.2022.08.009>.
12. Wang, S.; Wei, P.; Kong, Q.; Mao, W. A knowledge enhanced learning and semantic composition model for multi-claim fact checking. *Knowl.-Based Syst.* **2024**, *304*, 112439. <https://doi.org/10.1016/j.knosys.2024.112439>.
13. Tarik, N.; Jadhav, A. An Efficient Feature Selection Based Novel Deep Learning Models for Multi-Modal Sentimental Analysis in Social Media Platform. *Concurr. Comput. Pract. Exp.* **2026**, *38*, e70568. <https://doi.org/10.1002/cpe.70568>.
14. Kwabena Patrick, M.; Adebayo, F.A.; Ayidzoe, A.M.; Baagyire, E.Y. Capsule Networks: A survey. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 1295–1310. <https://doi.org/10.1016/j.jksuci.2019.09.014>.
15. Zhang, X.; Mao, Y.; Yang, Q.; Zhang, X. A Plant Leaf Disease Image Classification Method Integrating Capsule Network and Residual Network. *IEEE Access* **2024**, *12*, 44573–44585. <https://doi.org/10.1109/ACCESS.2024.3377230>.
16. Gauraha, R.; Agrawal, A.K.; Dubey, P. Capsule-enhanced RoBERTa for hierarchical sentiment analysis on social media texts. *Discov. Artif. Intell.* **2026**, *6*, 147. <https://doi.org/10.1007/s44163-026-00872-6>.
17. Pamulaparthi, L.; Sumalakshmi, C.H. A Cohesive Transformer-Capsule Network Model for Sentiment and Aspect Analysis Using Contextual Embeddings. *Eng. Technol. Appl. Sci. Res.* **2026**, *16*, 33954–33963. <https://doi.org/10.48084/etasr.16650>.
18. Singh, M.; Panwar, S.; Choudhary, S. BiERU-CapsNet: Bidirectional Emotional Recurrent Unit with Capsule Network, a Novel Approach for Textual Sentiment Analysis. *SN Comput. Sci.* **2026**, *7*, 183. <https://doi.org/10.1007/s42979-026-04782-3>.

19. Perera, S.; Perera, I.; Ahangama, S.; Rajasingham, S. Capsule Network-Based Dynamic Routing to Detect Hate Speech on Twitter. In *2026 6th International Conference on Advanced Research in Computing (ICARC)*; IEEE: Piscataway, NJ, USA, 2026; pp. 1–6. <https://doi.org/10.1109/ICARC68737.2026.11453877>.
20. Sagili, S.R.; Shibi, B.; Praveen, R.V.S.; Anjana, P. Sentiment Classification for Depression Detection: Integrating Capsule Networks with CNNs on Review Data. In *2025 Emerging Technologies for Intelligent Systems (ETIS)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–7. <https://doi.org/10.1109/ETIS64005.2025.10961685>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.