



OPEN ACCESS

EDITED BY

Xianzhi Wang,
University of Technology Sydney, Australia

REVIEWED BY

Faraz Masood,
Aligarh Muslim University, India
Khaled Alahmadi,
University of Technology Sydney, Australia

*CORRESPONDENCE

Zhenkai Qin,
✉ qinzhenkai@gxjcx.edu.cn

RECEIVED 28 July 2025

ACCEPTED 24 September 2025

PUBLISHED 08 October 2025

CITATION

Lin Z, Wu D, Yang X, Li L and Qin Z (2025)
FEAM: a dynamic prompting framework for
sentiment analysis with hierarchical
convolutional attention.
Front. Phys. 13:1674949.
doi: 10.3389/fphy.2025.1674949

COPYRIGHT

© 2025 Lin, Wu, Yang, Li and Qin. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with
these terms.

FEAM: a dynamic prompting framework for sentiment analysis with hierarchical convolutional attention

Ziqian Lin¹, Dongze Wu ², Xiangbeng Yang², Lingying Li² and
Zhenkai Qin ^{2*}

¹School of Public Administration, Guangxi Police College, Nanning, China, ²School of Information Technology, Guangxi Police College, Nanning, China

Introduction: This paper proposes FEAM (Fused Emotion-aware Attention Model), a dynamic prompting framework for sentiment analysis. Unlike static or handcrafted templates, FEAM dynamically selects input-specific prompts, aiming to address the challenges of nuanced sentiment expressions, lexical ambiguity, and domain variability.

Methods: The framework integrates a query-aware prompt controller with a BERT encoder to generate contextualized representations. Emotion-aware modulation amplifies sentiment-bearing features, multi-scale convolution captures linguistic patterns at different granularities, and topic-aware attention aligns local cues with global semantics. Experiments are conducted on four benchmark datasets: Rest16, Laptop, Twitter, and FinancialPhraseBank.

Results: FEAM achieves F1-scores of 91.55% on Rest16, 92.83% on Laptop, 93.10% on Twitter, and 90.11% on FinancialPhraseBank, outperforming strong baselines such as transformer-based, graph-enhanced, and prompt-tuned models. Ablation studies verify the contribution of each module, and robustness tests demonstrate resilience to adversarial perturbations and domain shifts.

Discussion: The results show that FEAM effectively improves sentiment classification across diverse and noisy textual domains. By combining dynamic prompting with emotion-aware modeling and hierarchical convolutional attention, FEAM provides a scalable and robust framework for real-world sentiment analysis, with potential extensions in domain adaptation, multimodal integration, and automated prompt discovery.

KEYWORDS

dynamic prompting, soft prompts, sentiment classification, prompt selection, emotion-aware modeling, domain generalization

1 Introduction

Sentiment analysis is a fundamental task in natural language processing (NLP), enabling the extraction of subjective opinions from unstructured text across a wide range of applications, such as social media monitoring, customer feedback analytics, and financial opinion mining [1–3]. Despite its broad utility, sentiment classification remains intrinsically challenging due to its reliance on subtle contextual cues, lexical ambiguity, and domain-sensitive semantics. For example, the phrase “*not bad*” implies a positive sentiment despite a

superficially negative construction, while the word “hot” may express excitement in fashion reviews but denote overheating in electronics. Such variations demand models that are both semantically precise and contextually adaptable.

Prompt-based learning has recently emerged as a parameter-efficient alternative to traditional fine-tuning for large pretrained language models (PLMs) such as BERT and RoBERTa [4]. By prepending task-relevant instructions or continuous prompt embeddings to the input, this paradigm enables downstream adaptation with minimal updates to model weights [5]. However, existing soft prompt approaches typically employ a static prompt vector for all instances, disregarding the semantic diversity and emotional variability inherent in sentiment-laden text. This one-size-fits-all strategy often fails to capture fine-grained polarity, especially when applied across domains with different linguistic expressions or stylistic patterns.

To alleviate these limitations, dynamic prompting methods have been introduced, wherein prompt representations are adapted based on the input context [6–8]. While offering improved flexibility, many such methods require either generating prompts via large decoders or modifying templates at inference time, leading to high computational overhead and reduced interpretability. More importantly, most of these methods are designed for general NLP tasks, with limited emphasis on the unique requirements of sentiment analysis—such as emotional salience, domain sensitivity, and polarity contrast.

To this end, we propose FEAM (Fused Emotion-aware Attention Model), a dynamic prompting framework tailored specifically for sentiment classification [9–11]. FEAM employs a lightweight *query-aware soft prompt selection* strategy, wherein a controller network dynamically retrieves soft prompt vectors from a learnable prompt pool according to the semantic embedding of each input. This enables instance-specific supervision without relying on handcrafted templates or computationally expensive generative mechanisms.

In addition, FEAM enhances sentiment-aware representation through a unified architecture that integrates three key components [1]: a token-wise emotion modulation layer that amplifies sentiment-bearing features [2], a multi-scale convolutional module that captures patterns at varied linguistic granularities, and [3] a topic-aware attention mechanism that aligns local features with global semantic context. These components collectively improve the model’s ability to detect nuanced sentiment expressions across domains, even under noisy or adversarial conditions. An overview of the FEAM framework is illustrated in Figure 1, which serves as the research overview for this work, highlighting the motivation, architectural design, and key findings.

We conduct extensive evaluations on four representative benchmarks—Rest16, Laptop, Twitter, and FinancialPhraseBank—as well as a large-scale single-domain corpus (Combined Dataset). FEAM consistently outperforms strong baselines, including transformer-based, graph-enhanced, and instruction-tuned models, showing a significant improvement in F1-score. Ablation and robustness analyses further validate the contributions of each module and demonstrate the model’s resilience under domain shifts and prompt perturbations [12, 13].

Our main contributions are summarized as follows.

1. We propose a lightweight and template-free dynamic prompting mechanism that enables interpretable, input-conditioned prompt selection without relying on handcrafted templates or generative modules.
2. We introduce FEAM, a framework that integrates dynamic prompting with emotion-aware token modulation, multi-scale convolutional encoding, and topic-aware attention to enhance sentiment representation and generalization.
3. We perform comprehensive evaluations on multi-domain datasets, showing that FEAM achieves superior accuracy, robustness, and domain adaptability compared to existing state-of-the-art models.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 formulates the problem. Section 4 presents the FEAM architecture. Section 5 describes the experimental setup and evaluation results. Section 6 discusses robustness, design implications, and limitations. Section 7 concludes the paper and outlines future directions.

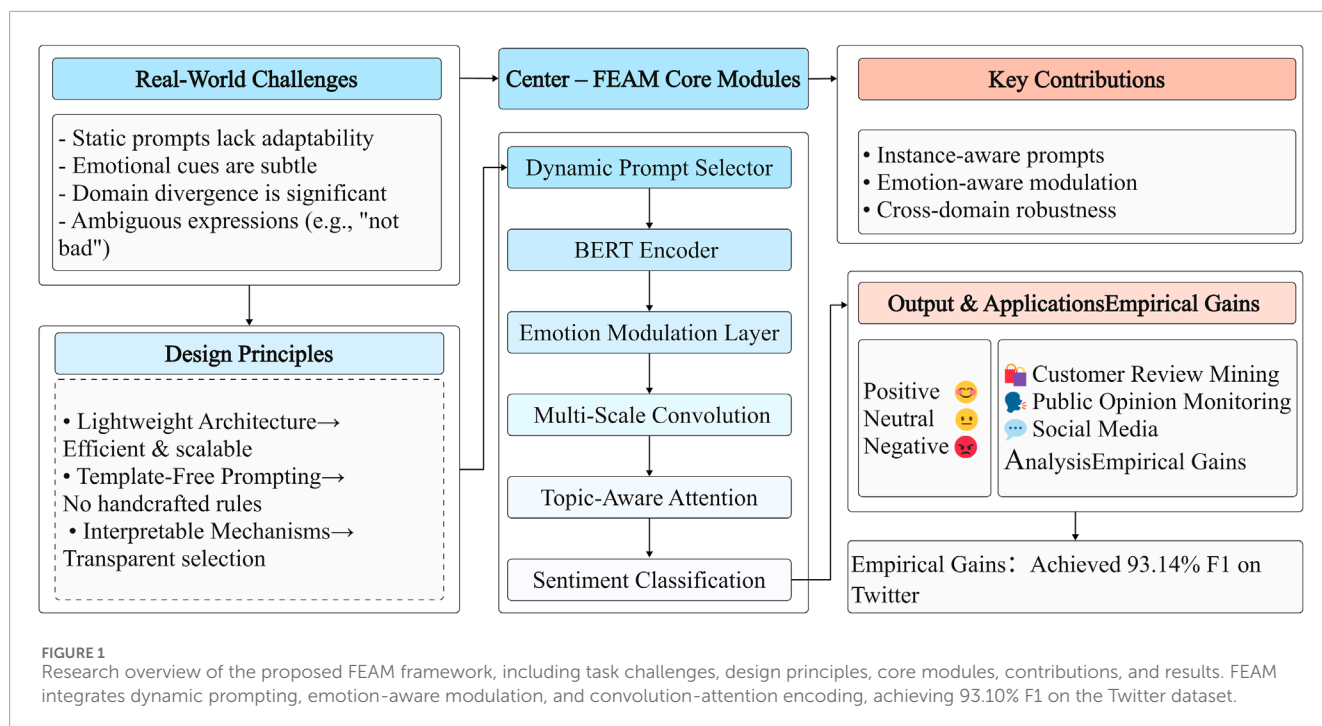
2 Related work

Sentiment analysis has long played a critical role in natural language processing (NLP), serving applications ranging from social media monitoring and public opinion mining to customer experience analysis and market forecasting [14–16]. Early approaches primarily relied on classical machine learning techniques with handcrafted features—such as n-grams, sentiment lexicons, and part-of-speech tags—combined with classifiers like Support Vector Machines and Naive Bayes [17–19]. Despite their simplicity, these methods lacked contextual understanding and exhibited poor generalizability across domains, particularly when dealing with implicit sentiment, sarcasm, or domain-specific expressions [20–23].

The rise of deep learning introduced automatic hierarchical representation learning into sentiment classification. Convolutional neural networks (CNNs) were effective for capturing localized polarity cues, while recurrent models such as LSTMs and GRUs could model sequential dependencies [24]. Nonetheless, these architectures typically require large-scale labeled data and often degrade in performance when deployed in domain-shifted or noisy scenarios.

More recently, prompt-based learning has emerged as a parameter-efficient paradigm for adapting pretrained language models (PLMs) to downstream tasks. Approaches such as Prefix-Tuning, P-Tuning v2, and Prompt-Tuning [26] reformulate tasks as conditional language modeling by prepending task-specific tokens or embeddings to the input sequence. While soft prompt tuning offers improved flexibility and efficiency by avoiding full model fine-tuning [27], it typically employs a static prompt across all inputs. This limits its ability to capture the diverse sentiment cues found in real-world text, especially in cases involving emotional nuance, sarcasm, or domain-dependent semantics.

To address this, dynamic prompting strategies have been proposed to enable instance-conditioned prompt generation or selection [28, 29]. These include routing-based controllers, attention-based prompt retrieval, and embedding-conditioned



generation. While these methods improve adaptability, they are often computationally expensive and offer limited interpretability. Moreover, most are designed for generic NLP tasks, and their effectiveness in sentiment analysis—where polarity is often subtle and context-dependent—remains underexplored.

Another complementary direction is domain adaptation, which aims to mitigate performance degradation across heterogeneous sentiment domains. Techniques such as adversarial training, contrastive learning, and domain-aware alignment have been developed to bridge distributional gaps [30, 31]. Prompt-based domain adaptation, which adjusts input representations via prompts to emphasize domain-relevant patterns, provides a lightweight alternative. However, these methods rarely integrate sentiment-specific refinements such as emotion-guided modulation or affective reasoning.

Hybrid neural architectures combining CNNs and transformers have also demonstrated competitive performance for sentiment tasks [32–34]. CNN modules effectively extract local features such as short phrases or negations, while transformer-based encoders model global dependencies. Yet, these hybrid models generally lack dynamic adaptability and seldom incorporate explicit emotion modeling mechanisms.

In parallel, emotion-aware representation learning has gained momentum for enhancing sentiment understanding. Approaches such as gated attention [35], lexicon-guided enhancement [36], and emotion-channel modeling [37] have been introduced to highlight affective content. However, such methods remain largely disconnected from the prompt learning paradigm. The synergy between prompt selection and emotion-aware encoding is seldom explored in current literature.

Although notable progress has been made across prompt-based adaptation, domain generalization, and emotion-sensitive modeling, existing efforts remain fragmented, rarely unifying

dynamic prompt selection, emotion-aware modulation, and multi-scale hierarchical modeling in a coherent architecture. To address this gap, we propose FEAM, a framework that integrates instance-specific prompting with emotion-guided feature enhancement, multi-scale convolution (Conv), and topic-aware attention (Attn), offering a lightweight yet sentiment-focused alternative that effectively bridges these gaps for robust sentiment classification across diverse and noisy textual domains.

As shown in Table 1, existing methods such as P-Tuning v2, SynPrompt, and Adaptive Prompt Retrieval each have limitations: P-Tuning v2 uses static prompts that fail to adapt to diverse input contexts, SynPrompt relies on template-based prompts that overlook sentiment signals, and Adaptive Prompt Retrieval, while input-dependent, neglects sentiment-aware representation. In contrast, FEAM integrates dynamic instance-specific prompting with emotion-guided feature enhancement, multi-scale convolution (Conv), and topic-aware attention (Attn), offering a lightweight yet sentiment-focused alternative that effectively bridges these gaps for robust sentiment classification across diverse and noisy textual domains.

3 Problem formulation

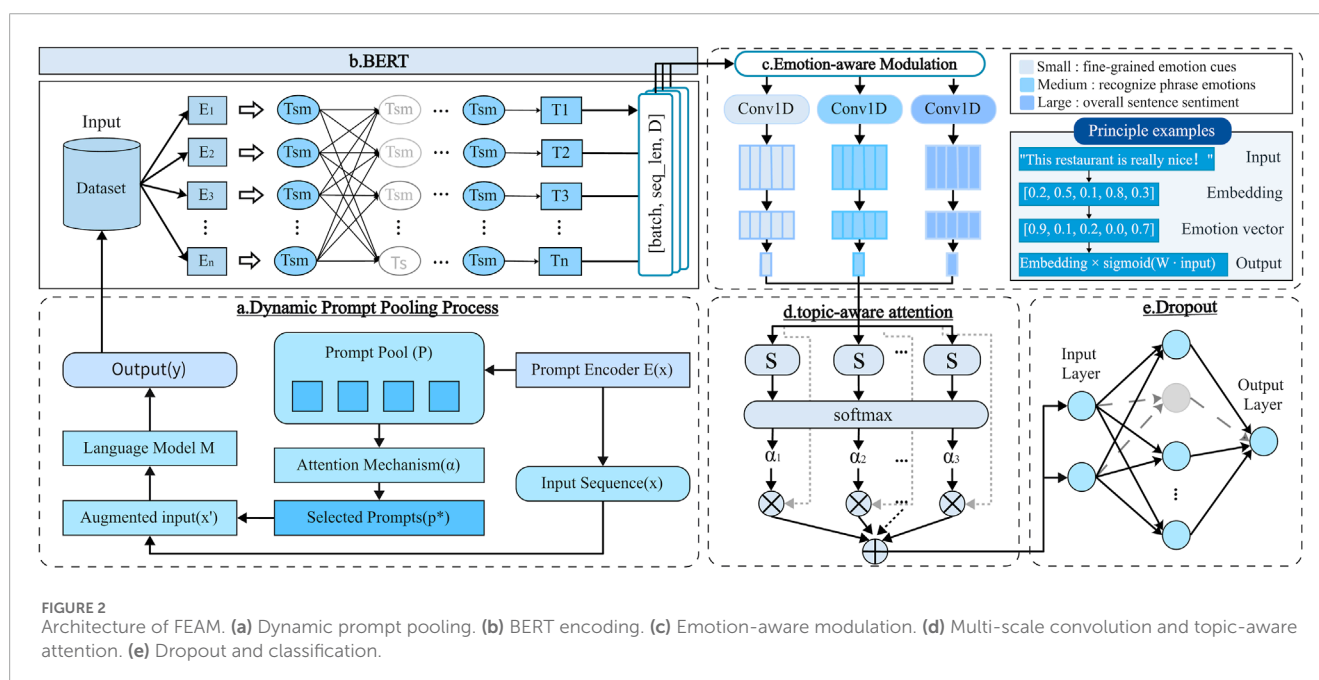
We formalize sentiment classification with dynamic prompting as a supervised learning task. Given a dataset $\mathcal{D} = \{(q_i, c_i)\}_{i=1}^N$, where each input q_i is a tokenized sentence and $c_i \in C = \{\text{Positive}, \text{Neutral}, \text{Negative}\}$ is the corresponding sentiment label, the objective is to learn a prediction function f that maps inputs to labels based on instance-aware prompt representations, as shown in Equation 1:

$$\hat{c}_i = f(q_i) = \arg \max_{c \in C} P(c | q_i, P_{q_i}) \quad (1)$$

TABLE 1 Comparison between FEAM and representative prompt-based approaches.

Method	Prompt type	Adaptivity	Sentiment-aware design
P-Tuning v2 [38]	Continuous soft prompts	Static across instances	✗
SynPrompt [39]	Syntax-enhanced prompts	Template-dependent, partial adaptivity	✗
Adaptive Prompt Retrieval [40]	Retrieved prompts from memory	Input-dependent retrieval	✗
FEAM (ours)	Dynamic soft prompt pool	Fully input-conditioned, query-aware selection	✓ (EM + Conv + Attn)

EM, emotion modulation; Conv, Convolution; Attn, Attention.



Here, P_{q_i} denotes a dynamically selected soft prompt tailored to q_i , computed by attending over a learnable prompt pool. Unlike static prompting, this formulation enables input-specific alignment between prompts and semantic-emotional features. The task poses challenges such as handling diverse sentiment expressions, adapting to cross-domain variation, and capturing both local polarity cues and global contextual semantics.

4 Methodology

To address the limitations of static, manually designed prompts—which often result in performance instability across diverse tasks and input distributions—we propose the FEAM framework. FEAM integrates instance-aware dynamic prompting with hierarchical semantic modeling to enhance both adaptability and robustness in sentiment classification. As illustrated in Figure 2, the framework consists of two main stages: dynamic prompt selection and hierarchical feature refinement.

In the first stage, *Dynamic Prompt Selection and Input Encoding*, the input is reformulated by prepending a dynamically generated soft prompt to the token sequence of the query. Specifically,

FEAM maintains a learnable prompt pool composed of multiple soft prompt vectors. A query-aware controller network processes the semantic embedding of the input sentence and computes a set of attention weights, which are used to aggregate relevant prompt vectors into a composite prompt $P_q \in \mathbb{R}^{m \times d}$, where m denotes the prompt length and d is the hidden dimension. This instance-specific prompt is then concatenated with the token embeddings of the input sentence and passed through a pretrained BERT encoder, producing contextualized representations that incorporate both task-specific supervision and domain-sensitive semantics.

In the second stage, *Hierarchical Feature Modulation and Aggregation*, the contextual representations are refined to emphasize sentiment-relevant and structurally significant information. A token-level sentiment modulation layer (see Figure 2b) employs a gating mechanism to selectively adjust embedding dimensions based on the emotional salience of each token. The modulated features are subsequently passed through a multi-scale convolutional module (Figure 2a) consisting of parallel convolutional filters with varying kernel sizes, enabling the extraction of sentiment cues at different levels of linguistic granularity. To further improve semantic coherence, a topic-aware attention mechanism (Figure 2c)

aggregates the multi-scale features using global sentence-level contextual information.

Finally, the aggregated representation is fed into a softmax classification layer to predict the sentiment polarity of the input (i.e., positive, neutral, or negative). By jointly leveraging dynamic prompt selection, emotion-aware feature modulation, and multi-scale hierarchical encoding, FEAM achieves robust and generalizable performance across diverse sentiment classification scenarios.

4.1 Prompt-based supervision encoding

To handle semantic variability and domain-specific sentiment divergence, the first stage of FEAM introduces an *instance-aware dynamic prompt encoding* strategy. This component enables the model to adaptively select soft prompt representations that align with the semantic content of each input, thereby providing query-specific supervision without relying on handcrafted templates or static patterns.

4.1.1 Instance-aware dynamic prompt selection

To enable input-specific adaptation, we introduce a dynamic prompt selection mechanism in which a controller network attends over a learnable prompt pool to generate soft prompts tailored to each query instance. Unlike conventional prompting paradigms that rely on fixed templates or demonstration-based in-context learning, the proposed approach facilitates continuous, differentiable, and semantically aligned prompt construction conditioned on the input.

The prompt pool is defined as a learnable tensor $\mathcal{P} \in \mathbb{R}^{N \times m \times d}$, where N denotes the number of candidate prompts, m is the length of each prompt in tokens, and d is the hidden embedding dimension. Given an input instance x_q , a query embedding vector $\mathbf{q} \in \mathbb{R}^d$ is computed by aggregating the [CLS] token representation with the mean of all token embeddings from the input sequence. This embedding is then passed through a controller network to produce an attention distribution over the prompt pool, as shown in Equation 2:

$$\alpha_i = \frac{\exp(\mathbf{q}^\top \mathbf{k}_i)}{\sum_{j=1}^N \exp(\mathbf{q}^\top \mathbf{k}_j)} \quad \text{for } i = 1, \dots, N \quad (2)$$

In this formulation, \mathbf{k}_i denotes the learned key vector corresponding to the i -th prompt. The final instance-specific prompt $P_s \in \mathbb{R}^{m \times d}$ is obtained via attention-weighted aggregation over the candidate prompts, as shown in Equation 3:

$$P_s = \sum_{i=1}^N \alpha_i \cdot \mathcal{P}_i \quad (3)$$

Let $E_q \in \mathbb{R}^{l_q \times d}$ represent the token embeddings of the query input. The prompt-augmented encoder input is then constructed by concatenating P_s with E_q , as shown in Equation 4:

$$X' = \text{Concat}(P_s, E_q) \in \mathbb{R}^{(m+l_q) \times d} \quad (4)$$

The resulting sequence X' is subsequently processed by a pretrained BERT encoder to obtain contextualized representations. Both the prompt pool and the controller network are trained jointly with the end-task objective, enabling the model to learn instance-aware prompt selection strategies that improve semantic alignment and task performance.

4.1.2 Contextual representation via BERT

To obtain contextually enriched representations from the prompt-augmented input, we utilize a pretrained BERT encoder. Owing to its bidirectional self-attention mechanism, BERT is well-suited for modeling long-range dependencies and capturing subtle sentiment expressions—both of which are essential for accurate sentiment inference under diverse linguistic and domain-specific conditions.

Let $X = \{x_1, x_2, \dots, x_n\}$ denote the input sequence formed by concatenating the dynamically selected prompt with the tokenized query. Each token x_i is embedded by summing its token, positional, and segment embeddings, as shown in Equation 5:

$$H_i^0 = E_{x_i} + E_{pos_i} + E_{seg_i} \quad (5)$$

These initial embeddings H^0 are subsequently propagated through a stack of L transformer layers, each comprising a multi-head self-attention mechanism followed by a position-wise feed-forward network. The attention operation is computed as shown in Equation 6:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \quad (6)$$

The output of the multi-head attention module is produced by concatenating the individual attention heads and projecting them through a linear transformation, as shown in Equation 7:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W^O \quad (7)$$

To facilitate gradient flow and training stability, each transformer layer includes residual connections and layer normalization, as defined in Equation 8:

$$H^{l+1} = \text{LayerNorm}\left(\text{FFN}(H^l) + H^l\right) \quad (8)$$

After processing through all L layers, the output $H_{\text{BERT}} \in \mathbb{R}^{n \times d}$ encodes the contextual semantics of the entire input, integrating both the content of the query and the semantics induced by the instance-specific prompt. These representations are then forwarded to the emotion-aware modulation module for sentiment-focused refinement.

4.2 Hierarchical Feature Modulation and Aggregation

While the *instance-aware dynamic prompting* mechanism enables contextual alignment by injecting adaptive supervision, it does not explicitly enhance sentiment-relevant features or handle syntactic and semantic variability. To overcome these limitations, the second phase of the proposed FEAM framework introduces a hierarchical feature refinement pipeline. This phase consists of three key components [1]: an emotion-aware gating layer that selectively amplifies sentiment-bearing tokens [2], a multi-scale convolutional module that captures features across diverse linguistic granularities, and [3] a topic-aware attention mechanism that integrates local features with global sentence-level semantics to guide final representation aggregation.

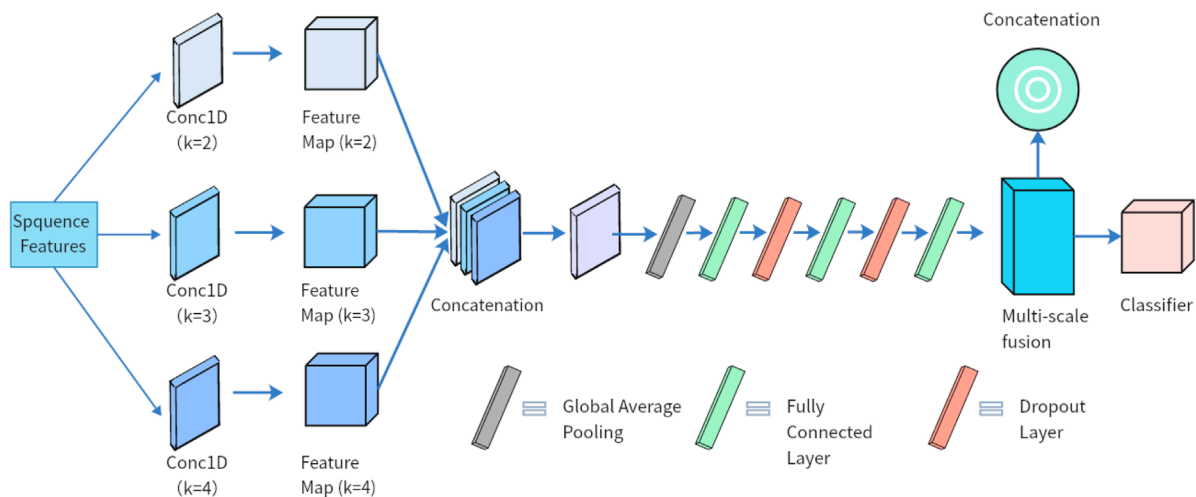


FIGURE 3 Architecture of the multi-scale convolutional module. Parallel 1D convolutions with kernel sizes of 2, 3, and four extract features at different linguistic granularities. Outputs are concatenated and globally pooled to form a unified representation.

4.2.1 Emotion-aware modulation

One major challenge in sentiment classification lies in the inconsistent encoding of emotionally salient tokens (e.g., “not,” “great,” “terrible”) by general-purpose language encoders such as BERT. These models often struggle to distinguish sentiment-critical expressions from neutral context, particularly in domain-divergent or low-signal scenarios.

To alleviate this limitation, we introduce an emotion-aware modulation layer that operates immediately after the BERT encoder. This layer applies a token-wise gating mechanism to selectively amplify or suppress token representations based on their emotional salience. Specifically, for each token embedding $h_i \in \mathbb{R}^d$, a sentiment gate $g_i \in [0, 1]^d$ is computed using a sigmoid-activated linear transformation. The gated representation \tilde{h}_i is obtained as shown in Equation 9:

$$\tilde{h}_i = g_i \odot h_i \quad (9)$$

where \odot denotes element-wise multiplication. The resulting modulated sequence $\tilde{H} = \{\tilde{h}_1, \dots, \tilde{h}_n\}$ preserves the syntactic structure of the input while selectively enhancing sentiment-relevant components, thereby improving downstream interpretability and discrimination.

4.2.2 Multi-scale convolutional feature extraction

Sentiment indicators can manifest at multiple linguistic granularities, ranging from short phrases (e.g., “too bad”) to full clauses (e.g., “did not meet expectations”). To capture such hierarchical patterns, we incorporate a multi-scale convolutional module that applies parallel convolutions with varying kernel sizes, as illustrated in Figure 3.

The input to this module is the modulated sequence $\tilde{H} \in \mathbb{R}^{n \times d}$ obtained from the previous layer. After transposing to $H^T \in \mathbb{R}^{d \times n}$,

three parallel 1D convolutional filters with kernel sizes $k \in \{2, 3, 4\}$ are applied to extract local features, as shown in Equation 10:

$$F_k = \text{ReLU}(\text{Conv}_k(H^T)), \quad k \in \{2, 3, 4\} \quad (10)$$

To ensure alignment among branches with differing temporal resolutions, each output is truncated to a common length n' , as defined in Equation 11:

$$\tilde{F}_k = F_k[:, :n'], \quad n' = \min(n_2, n_3, n_4) \quad (11)$$

The truncated feature maps are then concatenated along the channel dimension to yield a unified multi-scale representation F_{multi} , as shown in Equation 12:

$$F_{\text{multi}} = \text{Concat}(\tilde{F}_2, \tilde{F}_3, \tilde{F}_4) \quad (12)$$

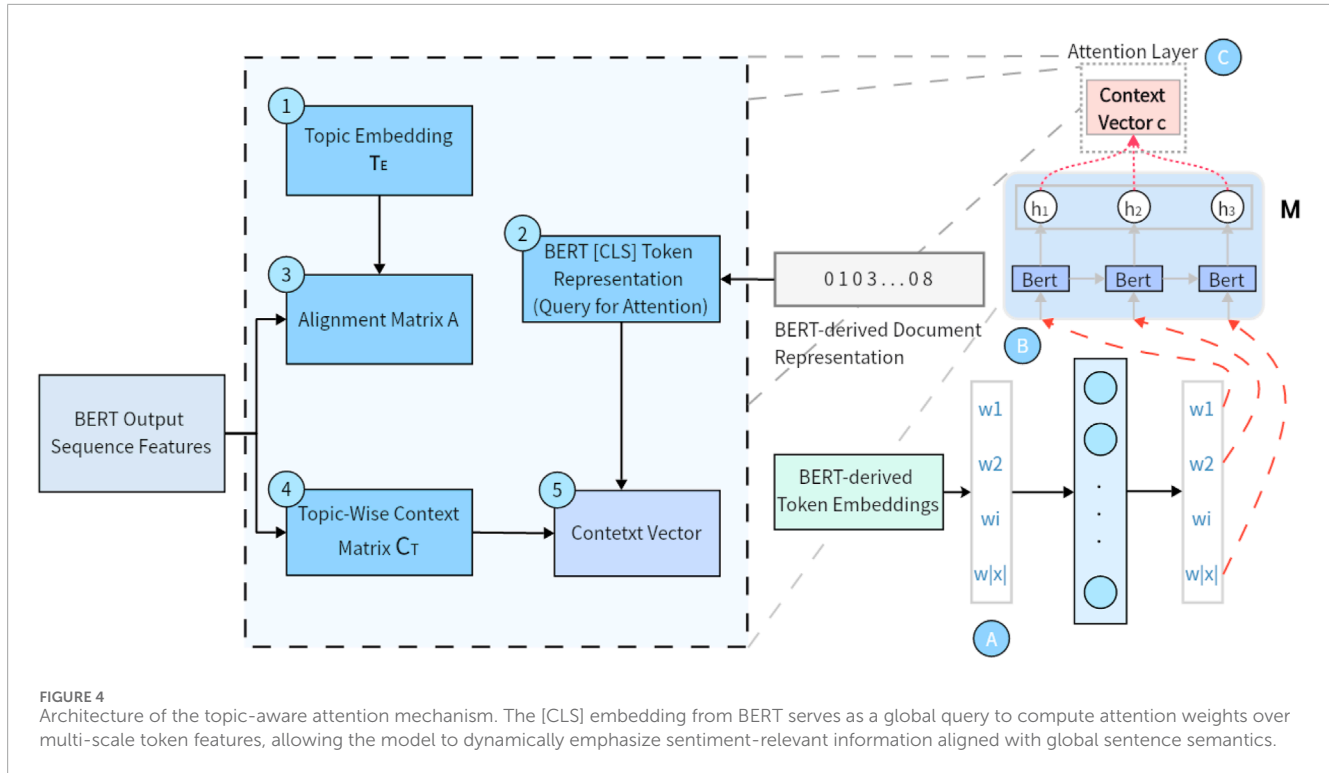
A global average pooling operation is subsequently applied along the temporal axis to obtain a fixed-dimensional representation vector f_{pool} , as shown in Equation 13:

$$f_{\text{pool}} = \frac{1}{n'} \sum_{i=1}^{n'} F_{\text{multi}}[:, i] \quad (13)$$

This pooled feature vector aggregates sentiment-relevant information across different granularities and serves as a compact, expressive representation for downstream attention and classification modules.

4.2.3 Topic-aware attention and classification

While the multi-scale convolutional module is effective in capturing localized sentiment features, it lacks the capacity to align such features with the global semantic structure of the input. To address this limitation, we introduce a topic-aware attention mechanism that modulates token-level importance based on sentence-level context, as illustrated in Figure 4.



Let $h_{[\text{CLS}]} \in \mathbb{R}^d$ denote the global sentence representation obtained from BERT. A query vector \mathbf{q} is computed by projecting $h_{[\text{CLS}]}$ into the attention space, as shown in Equation 14:

$$\mathbf{q} = W_q h_{[\text{CLS}]} \quad (14)$$

Each multi-scale token feature $f_i \in \mathbb{R}^d$ is compared with \mathbf{q} to compute attention weights via softmax normalization, as shown in Equation 15:

$$\alpha_i = \frac{\exp(f_i^T \mathbf{q})}{\sum_{j=1}^n \exp(f_j^T \mathbf{q})} \quad (15)$$

The attention-weighted sentence representation F_{att} is then obtained by computing the weighted sum over token features, as defined in Equation 16:

$$F_{\text{att}} = \sum_{i=1}^n \alpha_i f_i \quad (16)$$

This global representation is subsequently passed through a linear classification layer to predict sentiment polarity, as shown in Equation 17:

$$\hat{y} = \text{softmax}(W_{\text{cls}} F_{\text{att}} + b_{\text{cls}}) \quad (17)$$

Model parameters are optimized using cross-entropy loss between the predicted distribution \hat{y} and the ground-truth label distribution y , as defined in Equation 18:

$$\mathcal{L}_{\text{CE}} = -\sum_{i=1}^3 y_i \log(\hat{y}_i) \quad (18)$$

Taken together, the integration of topic-aware attention enables the model to align sentiment-expressive local features with global sentence-level semantics. In conjunction with the emotion-aware modulation and multi-scale convolution modules, this component enhances FEAM's ability to construct sentiment-focused, domain-robust representations suitable for nuanced classification tasks.

5 Experiments

5.1 Datasets

To assess the generalization capability and robustness of the proposed FEAM framework under diverse linguistic conditions, we conduct experiments on four sentiment classification datasets: Rest16, Laptop, Twitter, and FinancialPhraseBank. These datasets span restaurant reviews, consumer electronics, social media, and financial news, offering diverse lexical and stylistic patterns for comprehensive evaluation.

The Rest16 and Laptop datasets originate from the SemEval-2016 Task 5 benchmark, which was designed for aspect-based sentiment classification. Rest16 contains hospitality reviews, while Laptop includes product reviews from the electronics domain.

Sentence-level conversion. To adapt these corpora into sentence-level classification, aspect annotations were aggregated: if all aspects shared the same polarity, it was assigned as the sentence label; in case of disagreement, the majority label was kept, and sentences with irreconcilable conflicts were discarded. This ensured clean sentence-level labels without severe imbalance.

TABLE 2 Dataset statistics.

Dataset	Training set	Validation set	Test set	Total
Rest16	1,257	292	283	2,332
Laptop	4,142	530	617	5,789
Twitter	4,131	552	621	5,307
Financial	4,010	529	585	5,115

The Twitter dataset consists of user-generated posts, characterized by informal grammar, slang, and noisy expressions, providing a challenging real-world benchmark.

The FinancialPhraseBank dataset contains financial news headlines and company-related statements, annotated by experts with sentence-level polarity. Its domain-specific vocabulary (e.g., “dividend”, “merger”) makes it valuable for testing model generalization in professional contexts.

All four datasets were standardized into three sentiment classes (*positive*, *neutral*, *negative*). Table 2 summarizes their train/validation/test splits. A unified preprocessing pipeline (lowercasing, tokenization, sequence truncation) was applied to ensure comparability across domains.

To illustrate domain divergence, Figure 5 shows word clouds: Rest16 emphasizes hospitality terms, Laptop highlights electronics vocabulary, Twitter reflects informal open-domain language, and FinancialPhraseBank contains finance-specific expressions.

5.2 Experimental environment and hyperparameter settings

All experiments were conducted on a dedicated single-node computational server equipped with an NVIDIA A10 GPU (24 GB VRAM), 64 GiB of system memory, and running the Ubuntu 20.04 LTS operating system. The model implementation was based on the PyTorch deep learning framework (version 2.1.2), with CUDA 12.1 utilized to enable hardware-accelerated training and inference.

Model training was carried out over 20 full epochs using a mini-batch size of 32. Optimization was performed using the AdamW optimizer, with an initial learning rate set to 5×10^{-5} and a weight decay coefficient of 1×10^{-2} to prevent overfitting. To further stabilize the optimization process and promote convergence, a cosine annealing scheduler was employed to dynamically adjust the learning rate across epochs. Regularization was incorporated via a dropout mechanism with a rate of 0.3 applied to selected layers. In addition, gradient clipping with a maximum norm of 1.0 was enforced to address the risk of gradient explosion.

The FEAM architecture is composed of several inductive bias modules designed to enhance the representation of sentiment-relevant features. First, a multi-scale convolutional encoding block is employed, consisting of three parallel 1D convolutional layers with kernel sizes of 2, 3, and 4, respectively. Each layer comprises

128 filters, enabling the model to extract sentiment cues at various linguistic granularities—ranging from bi-grams to short clauses. In addition, the dynamic prompt pool is configured with $N = 5$ prompts of length $m = 10$, and a lightweight controller network projects input embeddings ($1536 \rightarrow 768 \rightarrow 768$) to compute prompt selection scores. A topic-aware attention mechanism then aligns local convolutional features with global semantic context for final classification.

A concise overview of the system specifications and hyperparameter configurations is provided in Table 3. These settings were held constant across all experiments to ensure comparability and reproducibility of results.

5.3 Comparison experiment

To rigorously evaluate the effectiveness of the proposed FEAM framework, we benchmark it against a set of state-of-the-art baselines across three sentiment classification tasks: Rest16, Laptop, and Twitter. The comparison includes standard transformer models (e.g., BERT, KDGN), graph-based architectures (e.g., EK-GCN, DGGCN), hybrid convolution-attention models (e.g., MambaForGCN + BERT), and a customized large language model Prompt-Qwen-2.5B. Specifically, Prompt-Qwen-2.5B is fine-tuned using LoRA (Low-Rank Adaptation) and prompt tuning in a fully supervised manner. The fine-tuning process uses cross-entropy loss with a learning rate of 2×10^{-5} , batch size of 8, and a maximum sequence length of 128.

As reported in Tables 4, 5, and others, FEAM consistently outperforms all competitive baselines in both accuracy and F1-score across the four datasets. The results demonstrate statistically significant improvements over both conventional neural architectures and large-scale instruction-tuned models.

On the Rest16 benchmark, FEAM achieves an F1-score of 91.55%, surpassing Prompt-Qwen-2.5B (90.09%) by 1.46 points and RoBERTa-large (89.09%) by 2.46 points. It also outperforms classical transformer variants such as BERT and KDGN + BERT by more than 11 points, highlighting FEAM’s superior ability to model sentiment-rich content in domain-specific, low-resource settings.

In the Laptop domain, which presents technical and fragmented sentiment expressions, FEAM attains 92.83% F1, exceeding Prompt-Qwen-2.5B by 1.61 points and RoBERTa-large (85.45%) by 7.38 points. Compared to graph-based approaches like EK-GCN + BERT (77.59%), the performance gap exceeds 15 points, underscoring the importance of adaptive prompt selection in complex domain-specific scenarios.

On the Twitter dataset—characterized by high lexical variability, informal expressions, and semantic noise—FEAM achieves a new state-of-the-art with an F1-score of 93.10%, outperforming Prompt-Qwen-2.5B (91.13%) and RoBERTa-large (87.68%) by 5.42 points. It also surpasses strong hybrid models like MambaForGCN + BERT (76.88%), demonstrating its robustness in real-world, user-generated text environments.

On the FinancialPhraseBank dataset, FEAM achieves an F1-score of 90.11%, exceeding Prompt-Qwen-2.5B (89.00%) and RoBERTa-large (84.14%) by 1.11 and 5.97 points, respectively. This indicates that FEAM’s adaptive prompting and sentiment modulation mechanisms generalize effectively to domain-specific

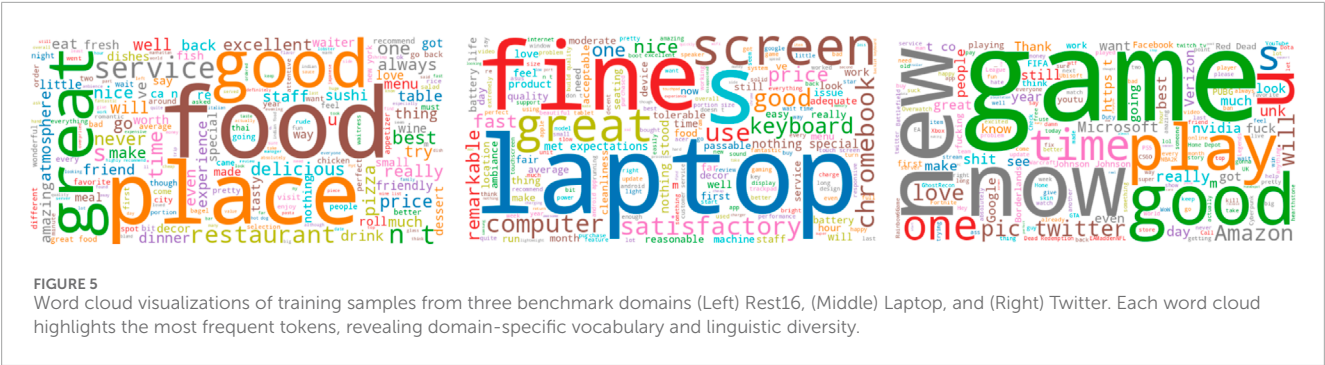


TABLE 3 Experimental environment and hyperparameter configurations used in FEAM training.

Setting	Value
GPU	NVIDIA A10 (24 GB)
System Memory	64 GiB
Operating System	Ubuntu 20.04
Python Version	3.10
CUDA Version	12.1
PyTorch Version	2.1.2
Batch Size	32
Number of Epochs	20
Learning Rate	5×10^{-5}
Weight Decay	1×10^{-2}
Dropout Rate	0.3
Gradient Clipping	1.0 (maximum norm)
Convolution Kernels	2, 3, 4 (128 filters each)
Optimizer	AdamW
Learning Rate Scheduler	CosineAnnealingLR
Prompt Pool Size (N)	5 prompts
Prompt Length (m)	red10 tokens per prompt
Controller Network	Two linear layers: 1536→768→768 (ReLU, dropout 0.1)

financial sentiment tasks, which are often challenging due to specialized vocabulary and subtle sentiment cues.

These consistent gains can be attributed to FEAM’s architectural design. Its *instance-aware dynamic prompting* module adapts supervision to the semantic properties of each input. The *sentiment modulation layer* enhances emotionally salient features, while the

multi-scale convolution and topic-aware attention mechanisms jointly capture both local sentiment expressions and global contextual cues.

In summary, FEAM establishes a new performance benchmark across all four tasks under evaluation, validating its effectiveness in cross-domain sentiment classification under the few-shot prompting paradigm. Its adaptability, lightweight supervision, and robustness to linguistic noise make it highly suitable for deployment in sentiment-driven applications across sectors such as e-commerce, financial analysis, customer experience management, and social media monitoring.

5.4 Ablation studies

To evaluate the individual contributions of each architectural component in the proposed FEAM framework, we conduct a series of ablation experiments by selectively disabling key modules. Specifically, we examine the impact of removing the following components [1]: the few-shot dynamic prompting mechanism (DP) [2], the sentiment modulation layer (SM) [3], the multi-scale convolutional encoder (Conv), and [4] the topic-aware attention mechanism (TA). Results are averaged over multiple runs and presented in Table 6.

The results clearly show that each module makes an indispensable contribution to FEAM’s performance. Removing the dynamic prompting mechanism (w/o DP) causes the most severe degradation, with F1 dropping from 93.14% to 61.18% and recall plunging from 93.28% to 50.93%. This emphasizes that instance-aware prompt selection is the cornerstone for maintaining generalization and robustness under few-shot or cross-domain conditions.

Excluding the sentiment modulation layer (w/o SM) also results in a sharp performance decline, with F1 decreasing by 30.62 points and recall falling from 93.28% to 51.20%. This confirms the gating layer’s crucial role in amplifying emotionally salient features while filtering out neutral or noisy tokens, especially in informal and short-text environments like Twitter.

When the multi-scale convolution module is removed (w/o Conv), F1 drops by 28.40 points (to 64.74%), despite precision remaining relatively higher (77.80%). This highlights the necessity of capturing sentiment cues at multiple granularities, particularly phrases and local syntactic structures that attention-only encoders tend to overlook.

TABLE 4 Accuracy, precision, recall, and F1-score (%) on Rest16 and Laptop (averaged over three runs with random seeds 42, 43, and 44, with standard deviation).

Model	Rest16				Laptop			
	Acc	P	R	F1	Acc	P	R	F1
BERT	85.79 ± 0.80	85.10 ± 0.70	84.90 ± 0.75	80.09 ± 0.85	79.91 ± 0.82	79.50 ± 0.78	79.20 ± 0.80	76.00 ± 0.83
KDGN + BERT	86.10 ± 0.82	85.60 ± 0.75	85.40 ± 0.78	80.29 ± 0.84	80.80 ± 0.83	80.40 ± 0.80	80.10 ± 0.81	78.55 ± 0.85
EK-GCN + BERT	87.01 ± 0.85	86.50 ± 0.80	86.30 ± 0.82	81.94 ± 0.86	81.32 ± 0.84	81.00 ± 0.81	80.70 ± 0.82	77.59 ± 0.85
DGGCN + BERT	87.65 ± 0.84	87.20 ± 0.81	87.00 ± 0.82	82.55 ± 0.85	81.30 ± 0.85	81.10 ± 0.82	80.90 ± 0.83	79.19 ± 0.86
DCN-CA + BERT	86.89 ± 0.83	86.40 ± 0.80	86.10 ± 0.81	80.32 ± 0.85	81.50 ± 0.86	81.20 ± 0.82	81.00 ± 0.84	78.51 ± 0.87
IA-HiNET + BERT	87.72 ± 0.85	87.30 ± 0.81	87.00 ± 0.83	82.65 ± 0.86	81.53 ± 0.87	81.20 ± 0.83	80.90 ± 0.84	77.97 ± 0.88
APSCL + BERT	86.79 ± 0.84	86.30 ± 0.80	86.00 ± 0.82	81.84 ± 0.85	79.45 ± 0.85	79.00 ± 0.82	78.80 ± 0.83	76.56 ± 0.86
MambaForGCN + BERT	86.68 ± 0.85	86.20 ± 0.82	85.90 ± 0.83	80.86 ± 0.86	81.80 ± 0.87	81.50 ± 0.83	81.20 ± 0.85	78.59 ± 0.88
RoBERTa-large	88.76 ± 0.85	89.94 ± 0.83	88.76 ± 0.82	89.09 ± 0.84	92.63 ± 0.87	84.62 ± 0.82	86.36 ± 0.83	85.45 ± 0.85
Prompt-Qwen-2.5B	90.19 ± 0.88	90.00 ± 0.84	89.80 ± 0.86	90.09 ± 0.87	91.22 ± 0.89	91.00 ± 0.85	90.90 ± 0.87	91.22 ± 0.88
FEAM	91.52 ± 0.90	91.55 ± 1.68	91.79 ± 0.28	91.55 ± 0.56	92.84 ± 0.87	92.83 ± 1.21	92.86 ± 1.87	92.83 ± 1.05

TABLE 5 Accuracy, precision, recall, and F1-score (%) on Twitter and FinancialPhraseBank (averaged over three runs with random seeds 42, 43, and 44, with standard deviation).

Model	Twitter				Financial			
	Acc	P	R	F1	Acc	P	R	F1
BERT	75.92 ± 0.85	75.50 ± 0.80	75.40 ± 0.82	75.18 ± 0.84	76.50 ± 0.88	76.20 ± 0.85	76.00 ± 0.86	75.90 ± 0.87
KDGN + BERT	75.00 ± 0.85	74.60 ± 0.83	74.40 ± 0.84	76.10 ± 0.86	77.00 ± 0.89	76.70 ± 0.86	76.50 ± 0.87	76.40 ± 0.88
EK-GCN + BERT	77.64 ± 0.87	77.20 ± 0.83	77.00 ± 0.84	75.55 ± 0.86	78.00 ± 0.90	77.80 ± 0.87	77.50 ± 0.88	77.30 ± 0.89
DGGCN + BERT	75.89 ± 0.85	75.50 ± 0.82	74.90 ± 0.84	75.16 ± 0.86	78.20 ± 0.90	78.00 ± 0.87	77.80 ± 0.88	77.60 ± 0.89
DCN-CA + BERT	76.94 ± 0.86	76.50 ± 0.83	76.30 ± 0.85	75.07 ± 0.87	77.80 ± 0.89	77.50 ± 0.86	77.30 ± 0.87	77.10 ± 0.88
IA-HiNET + BERT	77.59 ± 0.88	77.10 ± 0.85	76.90 ± 0.86	76.85 ± 0.87	78.50 ± 0.91	78.20 ± 0.87	78.00 ± 0.88	77.80 ± 0.89
APSCL + BERT	75.88 ± 0.86	75.50 ± 0.83	75.20 ± 0.85	75.36 ± 0.87	77.20 ± 0.89	76.90 ± 0.86	76.70 ± 0.87	76.50 ± 0.88
MambaForGCN + BERT	77.67 ± 0.87	77.20 ± 0.84	76.50 ± 0.86	76.88 ± 0.88	77.90 ± 0.90	77.60 ± 0.87	77.40 ± 0.88	77.20 ± 0.89
RoBERTa-large	87.70 ± 0.86	88.67 ± 0.84	87.69 ± 0.83	87.68 ± 0.85	85.08 ± 0.87	87.38 ± 0.85	85.08 ± 0.86	84.14 ± 0.88
Prompt-Qwen-2.5B	91.20 ± 0.90	91.00 ± 0.86	90.80 ± 0.88	91.13 ± 0.89	89.50 ± 0.92	89.30 ± 0.88	89.10 ± 0.90	89.00 ± 0.91
FEAM	93.14 ± 2.70	93.10 ± 2.23	93.28 ± 2.04	93.10 ± 2.25	90.12 ± 0.42	91.05 ± 0.29	90.12 ± 0.42	90.11 ± 0.40

Ablating the topic-aware attention mechanism (w/o TA) leads to a 19.78-point F1 reduction, with precision decreasing from 93.10% to 80.77%. This indicates that aligning local sentiment-bearing features with the global sentence-level context is essential for coherent and accurate sentiment classification.

In summary, the ablation study validates that all four components—dynamic prompting, sentiment modulation, multi-scale convolutional encoding, and topic-aware attention—are critical and complementary. Their synergistic integration ensures FEAM maintains superior performance across noisy, informal, and domain-diverse input conditions.

TABLE 6 Ablation Study: Average Accuracy, Precision, Recall, and F1-score (%) on the Twitter Dataset (mean \pm std, averaged over three runs with random seeds 42, 43, and 44).

Models	Acc. (%)	P (%)	R (%)	F1 (%)
FEAM (Full Model)	93.28 \pm 2.70	93.10 \pm 2.23	93.28 \pm 2.04	93.14 \pm 2.25
w/o DP (Dynamic Prompting)	84.28 \pm 1.52	71.05 \pm 1.46	50.93 \pm 1.38	61.18 \pm 1.42
w/o SM (Sentiment Modulation)	85.71 \pm 1.63	75.18 \pm 1.55	51.20 \pm 1.44	62.52 \pm 1.49
w/o Conv (Convolution Module)	79.71 \pm 1.44	77.80 \pm 1.39	54.60 \pm 1.30	64.74 \pm 1.36
w/o TA (Topic-Aware Attention)	80.77 \pm 1.48	80.77 \pm 1.41	66.07 \pm 1.33	73.36 \pm 1.37

It is worth noting that, in Table 6, accuracy drops appear marginal compared to the drastic decrease in F1. This discrepancy arises from the class imbalance in the Twitter dataset, which in our setting contains 5,307 samples with neutral instances being the majority (2,844), followed by positive (1,682) and negative (781). Accuracy therefore remains relatively stable when the model still correctly predicts the majority class, even if its performance on minority classes deteriorates. In contrast, F1—being more sensitive to per-class precision and recall—exposes the sharp degradation in handling positive and negative cases. Additional per-class analysis confirmed that variants without DP or SM suffer severe recall loss on minority categories, which explains the sharp F1 drop despite only moderate accuracy decline.

6 Discussion

6.1 Generalization and feature space visualization

To assess the generalization ability of FEAM under low-resource and cross-domain conditions, we construct a Combined Dataset by merging the Rest16, Laptop, and Twitter datasets into a unified multi-domain benchmark. This setting reflects real-world scenarios in which input distributions vary widely across topics, writing styles, and sentiment expressions.

Table 7 presents the average precision, recall, and F1-score on the combined test set. FEAM is evaluated against several strong baselines, including transformer-based models and prompt-enhanced architectures.

As shown in Table 7, FEAM consistently achieves the best results across all metrics. In particular, it surpasses Prompt-Qwen-2.5B by more than 10 F1 points (91.30% vs. 81.14%) and outperforms BERT by over 27 points (91.30% vs. 64.10%). These improvements highlight FEAM's superior capability to generalize across heterogeneous domains and to effectively capture sentiment signals even under limited supervision and diverse linguistic distributions.

To further examine the quality of the learned representations, we visualize the output embeddings of FEAM using t-distributed stochastic neighbor embedding (t-SNE). The resulting 2D projection of the final-layer sentence embeddings is shown in Figure 6.

TABLE 7 Accuracy, Precision, Recall, and F1-score (%) on the Combined Multi-Domain Dataset (mean \pm std, averaged over three runs with random seeds 42, 43, and 44).

Model	Acc. (%)	P (%)	R (%)	F1 (%)
BERT	74.07 \pm 0.92	65.20 \pm 0.88	63.10 \pm 0.90	64.10 \pm 0.91
MSAPT	76.98 \pm 0.95	72.40 \pm 0.90	70.90 \pm 0.93	71.59 \pm 0.92
BERT-DAAT	87.05 \pm 1.10	75.80 \pm 1.03	73.60 \pm 1.07	74.62 \pm 1.05
MoKa-ADA	84.11 \pm 1.05	72.00 \pm 0.98	70.90 \pm 1.02	71.43 \pm 1.01
Prompt-Qwen-2.5B	88.60 \pm 1.15	81.70 \pm 1.08	80.60 \pm 1.11	81.14 \pm 1.10
FEAM	91.39 \pm 1.20	91.50 \pm 1.15	91.10 \pm 1.18	91.30 \pm 1.17

Despite the diverse and noisy nature of the combined input space, the t-SNE visualization reveals clear and interpretable clustering of sentiment categories. Negative samples are tightly grouped, indicating strong polarity separation. Meanwhile, neutral and positive instances exhibit moderate overlap—reflecting the inherent semantic ambiguity of weakly polarized expressions. This latent structure illustrates that FEAM's architecture—integrating instance-aware prompting, emotion-aware feature modulation, and hierarchical convolution-attention encoding—effectively organizes the semantic space for robust sentiment classification.

In conclusion, FEAM not only delivers state-of-the-art performance in terms of quantitative metrics but also produces well-structured, semantically coherent latent spaces across domains. These findings highlight the model's scalability, interpretability, and potential for deployment in practical, real-world sentiment analysis applications [41, 42].

6.2 Impact of adversarial perturbation strength

To evaluate the robustness of FEAM under adversarial input conditions, we perform controlled experiments using the Fast Gradient Method (FGM), which introduces perturbations into the embedding layer during training. We systematically vary the

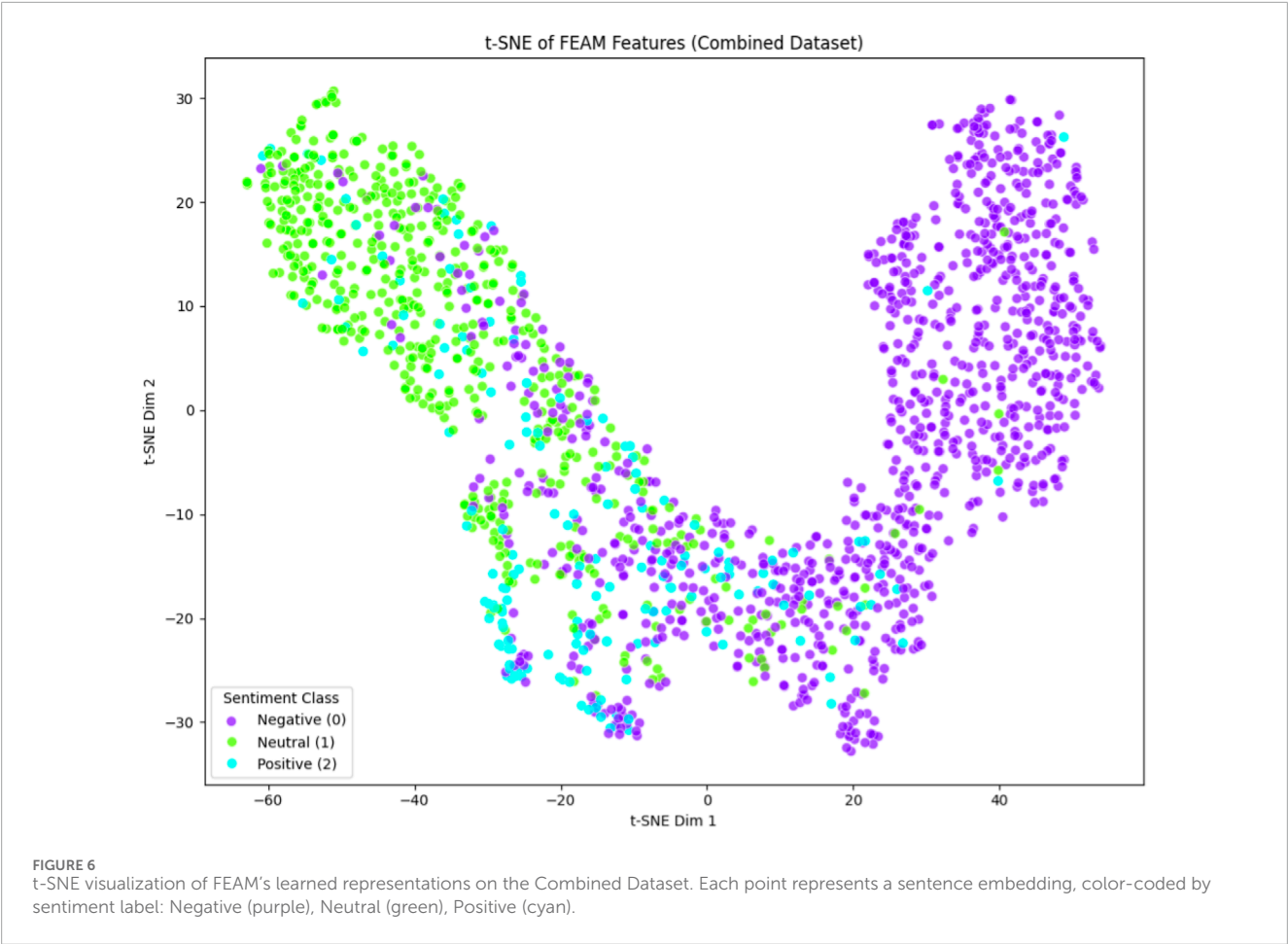


TABLE 8 Accuracy, Precision, Recall, and F1-score (%) under Varying Adversarial Perturbation Strength (ϵ) Using FGM (mean \pm std, averaged over three runs with random seeds 42, 43, and 44).

ϵ	Combined dataset				Twitter dataset			
	Acc. (%)	P (%)	R (%)	F1 (%)	Acc. (%)	P (%)	R (%)	F1 (%)
None	91.39 \pm 1.20	91.50 \pm 1.15	91.10 \pm 1.18	91.30 \pm 1.17	93.14 \pm 2.70	93.10 \pm 2.23	93.28 \pm 2.04	93.10 \pm 2.25
0.5	90.62 \pm 0.90	90.40 \pm 0.86	90.10 \pm 0.88	90.23 \pm 0.89	92.89 \pm 0.92	92.70 \pm 0.88	92.60 \pm 0.90	92.68 \pm 0.91
1.0	90.44 \pm 0.91	90.00 \pm 0.87	89.40 \pm 0.89	89.65 \pm 0.90	92.35 \pm 0.93	92.10 \pm 0.89	92.30 \pm 0.91	92.21 \pm 0.92
2.0	89.21 \pm 0.93	89.00 \pm 0.89	89.20 \pm 0.91	89.15 \pm 0.92	91.50 \pm 0.94	91.40 \pm 0.90	91.60 \pm 0.92	91.57 \pm 0.93

perturbation magnitude $\epsilon \in \{0, 0.5, 1.0, 2.0\}$, where $\epsilon = 0$ corresponds to the clean, unperturbed baseline. All other training configurations are held constant to ensure comparability.

We assess model performance on two test sets [1]: the cross-domain Combined Dataset (a union of Rest16, Laptop, and Twitter), and [2] the highly informal, user-generated Twitter Dataset. The results are summarized in Table 8.

As shown in Table 8, FEAM maintains consistently strong performance across all perturbation levels. At the highest perturbation strength ($\epsilon = 2.0$), the model still achieves 89.15% F1 on the Combined Dataset and 91.57% on the Twitter dataset, underscoring its robustness to adversarial noise injected at the embedding layer.

For mild perturbations (e.g., $\epsilon = 0.5$), performance degradation is minimal—an F1-score drop of about 1.07 points on the Combined Dataset (from 91.30% to 90.23%) and only 0.42 points on Twitter (from 93.10% to 92.68%). Interestingly, recall remains relatively stable, suggesting that weak adversarial noise may serve as a form of implicit regularization, enhancing generalization capability. However, as ϵ increases to 1.0 and 2.0, performance begins to decline more noticeably: F1 falls to 89.65% and 89.15% on the Combined Dataset, and to 92.21% and 91.57% on Twitter, reflecting greater disruption to the model’s ability to capture fine-grained sentiment cues.

The observed robustness can be attributed to FEAM’s architectural design. Specifically, the instance-aware dynamic

prompting mechanism provides context-sensitive supervision, the sentiment modulation layer amplifies sentiment-bearing signals, and the multi-scale convolutional attention module jointly captures local and global sentiment patterns. This synergy enables FEAM to maintain high stability even under adversarial perturbations [44–46].

In conclusion, FEAM demonstrates strong resilience under adversarial conditions. Even when $\epsilon = 2.0$, it sustains nearly 90% F1 on the Combined Dataset and over 91% on Twitter, a critical capability for real-world deployment in domains such as social media sentiment monitoring, e-commerce review mining, and opinion tracking—where noisy or corrupted user input is frequent and unavoidable.

6.3 Implications and limitations

The findings of this study offer several key insights into advancing *instance-aware dynamic prompting* for sentiment analysis in heterogeneous domains. First, the results underscore the critical role of prompt adaptability in enhancing model generalization. Unlike static templates or handcrafted demonstrations, dynamically selecting soft prompts based on input semantics enables FEAM to better align with task-relevant and domain-sensitive cues. This underscores the potential of modular prompt selection mechanisms that are both interpretable and context-aware.

Second, FEAM's architectural synergy—integrating dynamic prompt supervision, emotion-aware gating, and multi-scale convolutional encoding—proves highly effective in capturing both localized sentiment expressions and global semantic dependencies. This layered design allows the model to perform robustly across inputs with varying syntactic and lexical characteristics. Consequently, FEAM shows strong promise for deployment in real-world, domain-specific scenarios such as healthcare opinion monitoring, legal sentiment detection, and educational feedback analysis, where affective cues are often subtle, ambiguous, or domain-dependent.

Despite its strengths, the current work has several limitations. First, FEAM relies on general-purpose pretrained encoders (e.g., BERT), which may not fully capture domain-specific terminology, stylistic nuances, or linguistic conventions. Future research may explore domain-adaptive pretraining or hybrid architectures that incorporate symbolic knowledge sources—such as sentiment lexicons or expert-annotated rules—into the prompt construction or representation refinement process.

Second, although FEAM's soft prompts are learned in an input-conditioned manner, they are initialized randomly and trained end-to-end as latent vectors, limiting interpretability and controllability. Future efforts may investigate automated prompt discovery and optimization, potentially leveraging contrastive learning, meta-prompt tuning, or generative prompt synthesis using large language models (LLMs) to improve prompt diversity and semantic richness.

Third, the present framework is restricted to unimodal textual input. However, sentiment in many real-world settings—such as social media, customer service, or online reviews—is often conveyed through multiple modalities, including text, images, audio, and video. Extending FEAM to support multimodal architectures would facilitate more comprehensive sentiment understanding in noisy and dynamic environments.

In conclusion, this work introduces a unified and interpretable framework for sentiment classification based on instance-aware dynamic prompting, demonstrating strong generalization, robustness, and semantic coherence across domains. To enhance its practical applicability, future research should pursue directions such as domain adaptation, prompt interpretability, automated prompt generation, and multimodal extension—laying the foundation for more flexible, scalable, and context-sensitive sentiment analysis systems.

7 Conclusion

This paper presents FEAM, a novel framework for sentiment classification that integrates *instance-aware dynamic prompting* with hierarchical feature modeling. By unifying dynamic soft prompt selection, emotion-aware token modulation, multi-scale convolutional encoding, and topic-aware attention, FEAM effectively addresses key challenges in sentiment analysis, such as domain variability, input-level ambiguity, and limited supervision. In contrast to traditional fine-tuning or static prompting approaches, FEAM dynamically selects prompt representations conditioned on the semantic characteristics of each input, thereby enhancing contextual alignment and domain robustness.

Comprehensive experiments across single-domain and multi-domain benchmarks demonstrate that FEAM consistently outperforms a range of strong baselines, including pretrained transformer models, domain-adaptive architectures, and recent prompt-based methods. The model also exhibits strong resilience to adversarial perturbations and maintains robust performance across various prompt configurations. Ablation studies confirm the critical contributions of each architectural component, with dynamic prompting and emotion-aware modulation proving particularly impactful in boosting sentiment discrimination accuracy.

Looking forward, promising directions for future research include automated prompt synthesis, improved alignment between learned prompt representations and internal semantic features, and integration with large-scale language models to further enhance reasoning capabilities. Moreover, extending FEAM to support multimodal sentiment analysis—by incorporating visual, auditory, or contextual metadata—would further expand its applicability to real-world opinion mining scenarios such as social media monitoring, customer feedback interpretation, and public discourse analysis [47, 48]. In addition, we plan to complement the current quantitative evaluation with qualitative interpretability analysis of dynamic prompts. Future work will explore visualization of prompt selection, token-level attribution, and case studies to provide deeper insights into how instance-aware prompts guide sentiment prediction, thereby improving transparency and trustworthiness.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://github.com/NUSTM/ACOS/tree/main/data> <https://www.kaggle.com/datasets/jp797498e/twitter-entity-sentiment-analysis> <https://alt.qcri.org/semeval2016/task5/>.

Ethics statement

Ethical approval was not required for the study involving human data in accordance with the local legislation and institutional requirements. Written informed consent was not required, for either participation in the study or for the publication of potentially/indirectly identifying information, in accordance with the local legislation and institutional requirements. The social media data was accessed and analyzed in accordance with the platform's terms of use and all relevant institutional/national regulations.

Author contributions

ZL: Conceptualization, Formal Analysis, Investigation, Methodology, Resources, Supervision, Writing – original draft. DW: Data curation, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. XY: Conceptualization, Formal Analysis, Investigation, Project administration, Resources, Supervision, Validation, Writing – review and editing. LL: Conceptualization, Investigation, Methodology, Resources, Validation, Writing – original draft. ZQ: Writing – original draft, Resources, Validation, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was supported by the Guangxi Key R&D Plan Project (Project No:

AB25069130) of Guangxi Science and Technology, and the Guangxi Philosophy and Social Sciences Project (Project No: 24TQF007).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Anand S, Devulapally NK, Bhattacharjee SD, Yuan J. Multi-label emotion analysis in conversation via multimodal knowledge distillation. In: *Proceedings of the conference*. New York, NY, USA: Association for Computing Machinery (2023). doi:10.1145/3581783.3612517
2. Xue P, Li Y, Wang S, Liao J, Zheng J, Fu Y, et al. Sentiment classification method based on multitasking and multimodal interactive learning. In: *Proceedings of the 22nd Chinese national conference on computational linguistics (Harbin, China: Chinese information processing society of China)* (2023). p. 315–27.
3. Ma F, Zhang Y, Sun X. Multimodal sentiment analysis with preferential fusion and distance-aware contrastive learning. In: *Proceedings of the IEEE international conference on multimedia and expo (ICME)*. Brisbane, Australia (2023). p. 1367–72. doi:10.1109/ICME55011.2023.00237
4. Fradet N, Gutowski N, Chhel F, Briot JP. Byte pair encoding for symbolic music. In: *Proceedings of the 2023 conference on empirical methods in natural language processing*. Singapore: Association for Computational Linguistics (2023). p. 2001–20.
5. Wu J, Liu X, Yin X, Zhang T, Zhang Y. Task-adaptive prompted transformer for cross-domain few-shot learning. *Proc AAAI Conf Artif Intelligence* (2024) 38:6012–20. doi:10.1609/aaai.v38i6.28416
6. Hu K, Zhang Y. Prompt-based active learning with ensemble sampling strategy for small data learning in the scientific text classification scenarios. *SSRN Electron J* (2024). doi:10.2139/ssrn.5027128
7. Weng W, Pratama M, Zhang J, Chen C, Yapp Kien Yie E, Savitha R. Cross-domain continual learning via clamp. *Inf Sci* (2024) 676:120813. doi:10.1016/j.ins.2024.120813
8. Wang B, Yu D. A divide-and-conquer strategy for cross-domain few-shot learning. *Electronics* (2025) 14:418. doi:10.3390/electronics14030418
9. Biswas S, Saw R, Nandy A, Naskar AK. Attention-enabled hybrid convolutional neural network for enhancing human-robot collaboration through hand gesture recognition. *Comput Electr Eng* (2025) 123:110020. doi:10.1016/j.compeleceng.2024.110020
10. Gong Z, Chanmean M, Gu W. Multi-scale hybrid attention integrated with vision transformers for enhanced image segmentation. In: *Proceedings of the 2nd international conference on algorithm, image processing and machine vision (AIPMV) (Zhenjiang, China)* (2024). p. 180–4. doi:10.1109/AIPMV62663.2024.10691911
11. Xu B, Zhang X, Zhang X, Sun B, Wang Y. Ddnet: a hybrid network based on deep adaptive multi-head attention and dynamic graph convolution for eeg emotion recognition. *Signal Image Video Process.* (2025) 19:293. doi:10.1007/s11760-025-03876-4
12. Hu W, Wu J, Yuan G. Some convergently three-term trust region conjugate gradient algorithms under gradient function non-lipschitz continuity. *Scientific Rep* (2025) 14:10851. doi:10.1038/s41598-024-60969-9
13. Zhang Z, Yuan G, Qin Z, Luo Q. An improvement by introducing lbfgs idea into the adam optimizer for machine learning. *Expert Syst Appl* (2025) 232:129002. doi:10.1016/j.eswa.2025.129002
14. Rahman Jim J, Talukder MAR, Malakar P, Kabir MM, Nur K, Mridha MF. Recent advancements and challenges of nlp-based sentiment analysis: a state-of-the-art review. *Nat Lang Process J* (2024) 6:100059. doi:10.1016/j.nlp.2024.100059
15. Islam MS, Kabir MN, Ghani NA, Zamli KZ, Zulkifli NSA, Rahman MM, et al. Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach. *Artif Intelligence Rev* (2024) 57:62. doi:10.1007/s10462-023-10651-9
16. Li E, Li T, Liang T, Kang A, Chen K, Luo H. Cross-lingual sentiment analysis empowered by emotional mutual reinforcement through emojis. *Int J Machine Learn Cybernetics* (2025) 16:6031–45. doi:10.1007/s13042-025-02610-3
17. Joshi G, Soni NR. *Sentiment analysis based on machine learning techniques: a comprehensive review*. ResearchGate (2023).
18. Anwar K, Garg A, Ahmad S, Wasid M. Machine learning models for sentiment analysis: state of the art and applications. In: *Proceedings of the 15th international conference on computing, communication and networking technologies (ICCCNT)*. Kamand, India (2024). p. 1–6.

19. Brandão JG, Castro Junior AP, Gomes Pacheco VM, Rodrigues CG, Belo OMO, Coimbra AP, et al. Optimization of machine learning models for sentiment analysis in social media. *Inf Sci* (2025) 694:121704. doi:10.1016/j.ins.2024.121704
20. Farah HA, Kakisim AG. Enhancing lexicon based sentiment analysis using n-gram approach. In: *ICAIAE 2021. Engineering cyber-physical systems and critical infrastructures*. Cham: Springer (2023). p. 213–21. doi:10.1007/978-3-031-09753-9_17
21. Helal NA, Hassan A, Badr NL, Afify YM. A contextual-based approach for sarcasm detection. *Scientific Rep* (2024) 14:15415. doi:10.1038/s41598-024-65217-8
22. Priya M, Vijaya Kumar K, Vennila P, Prasanna MA. Sarcasam analysis in social networks using deep learning algorithm. *Proced Comp Sci* (2025) 252:10. doi:10.1016/j.procs.2025.01.010
23. Liu Z, Xu L, Zhang S. Unified vision-language pretraining for image captioning and vqa. *arXiv* (2025). doi:10.1609/aaai.v34i07.7005
24. Huang M, Xie H, Rao Y, Liu Y, Poon LKM, Wang FL. Lexicon-based sentiment convolutional neural networks for online review analysis. *IEEE Trans Affective Comput* (2022) 13:1337–48. doi:10.1109/TAFFC.2020.2997769
26. Xiang Y, Zhang A, Guo J, Huang Y. Adaptive multimodal prompt-tuning model for few-shot multimodal sentiment analysis. *Int J Machine Learn Cybernetics* (2025) 16:4899–911. doi:10.1007/s13042-025-02549-5
27. Yintao J, Jiajia C, Lingling M, Hongying Z. Dynamic prompt learning and dependency relation based generative structured sentiment analysis model. In: *Proceedings of the 23rd Chinese national conference on computational linguistics (volume 1: main conference) (Taiyuan, China: Chinese information processing society of China)* (2024).
28. Zou H, Wang Y. Large language model augmented syntax-aware domain adaptation method for aspect-based sentiment analysis. *Neurocomputing* (2025) 625:129472. doi:10.1016/j.neucom.2025.129472
29. Ghosal D, Hazarika D, Roy A, Majumder N, Mihalcea R, Poria S. Kingdom: knowledge-guided domain adaptation for sentiment analysis. In: *Proceedings of the 58th annual meeting of the Association for computational linguistics (Online: Association for computational linguistics)* (2020).
30. Huertas-Tato J, Martin A, Huertas-Garcia A, Camacho D. Generating authorship embeddings with transformers. In: *Proceedings of the 2022 international joint conference on neural networks (IJCNN)*. Padua, Italy (2022). p. 1–8. doi:10.1109/IJCNN55064.2022.9892173
31. Yan X, Huang P, Li Z. Self-improving prompting for large language models. *arXiv* (2023). doi:10.48550/arXiv.2210.11610
32. Jahin M, Shovon MSH, Mridha ME, Islam MR, Watanobe Y. A hybrid transformer and attention based recurrent neural network for robust and interpretable sentiment analysis of tweets. *Scientific Rep* (2024) 14:24882. doi:10.1038/s41598-024-76079-5
33. Cheng Y, Sun H, Chen H, Li M, Cai Y, Cai Z, et al. Sentiment analysis using multi-head attention capsules with multi-channel cnn and bidirectional gru. *IEEE Access* (2021) 9:60383–95. doi:10.1109/ACCESS.2021.3073988
34. Kota VR, Munisamy SD. High accuracy offering attention mechanisms based deep learning approach using cnn/bi-lstm for sentiment analysis. *Int J Intell Comput Cybernetics* (2022) 15:61–74. doi:10.1108/IJICC-06-2021-0109
35. Islam SMM, Dong X, de Melo G. Domain-specific sentiment lexicons induced from labeled documents. In: *Proceedings of the 28th international conference on computational linguistics (Barcelona, Spain (Online): International Committee on computational linguistics)* (2020).
36. Dubey G, Kaur K, Chadha A, Raj G, Jain S, Dubey AK. A domain knowledge infused gated network using integrated sentiment prediction framework for aspect-based sentiment analysis. *Evolving Syst* (2025) 16:12. doi:10.1007/s12530-024-09625-1
37. Yekrangi M, Nikolov NS. Domain-specific sentiment analysis: an optimized deep learning approach for the financial markets. *IEEE Access* (2023) 11:70248–62. doi:10.1109/access.2023.3293733
38. Liu X, Ji K, Fu Y, Du Z, Yang Z, Tang J. P-tuning v2: prompt tuning can be comparable to fine-tuning universally across scales and tasks. In: *Proceedings of the 60th annual meeting of the Association for computational linguistics (volume 2: short papers)*. Association for Computational Linguistics (2022). p. 61–8. doi:10.18653/v1/2022.acl-short.8
39. Yin W, Liu C, Xu Y, Wahla AR, He Y, Zheng D. Synprompt: syntax-aware enhanced prompt engineering for aspect-based sentiment analysis. In: *Proceedings of the joint international conference on computational linguistics, Language resources and evaluation (LREC-COLING 2024) (European Language resources Association (ELRA))* (2024). p. 15469–79.
40. Zhang P, Chai T, Xu Y. Adaptive prompt learning-based few-shot sentiment analysis. *arXiv preprint arXiv:2205.07220* (2022). doi:10.48550/arXiv.2205.07220
41. Zhou H, Wan X, Proleev L, Mincu D, Chen J, Heller KA, et al. Batch calibration: rethinking calibration for in-context learning and prompt engineering. *arXiv* (2024). doi:10.48550/arXiv.2309.17249
42. Huang D, Wang Z. Explainable sentiment analysis with deepseek-r1: performance, efficiency, and few-shot learning. *arXiv* (2025). doi:10.1109/MIS.2025.3614967
44. Chun J. Multisentimentarcs: a novel method to measure coherence in multimodal sentiment analysis for long-form narratives in film. *Front Comp Sci* (2024) 6:1444549. doi:10.3389/fcomp.2024.1444549
45. Xian Y, Qin Y, Xiang Y. Auto-verbalizer filtering for prompt-based aspect category detection. *Front Phys* (2024) 12:1361695. doi:10.3389/fphy.2024.1361695
46. Chen Q, Korneder J, Rawashdeh OA, Wang Y, Louie WYG. Improving optimal prompt learning through multilayer fusion and latent dirichlet allocation. *Front Robotics AI* (2025) 12:1579990. doi:10.3389/frobt.2025.1579990
47. Qin Z, Luo Q, Zang Z, Fu H. Multimodal gru with directed pairwise cross-modal attention for sentiment analysis. *Scientific Rep* (2025) 15:10112. doi:10.1038/s41598-025-93023-3
48. Lu W, Wang J, Wang T, Zhang K, Jiang X, Zhao H. Visual style prompt learning using diffusion models for blind face restoration. *Pattern Recognition* (2025) 161:111312. doi:10.1016/j.patcog.2024.111312