



ARTICLE

External Knowledge-Enhanced Cross-Attention Fusion Model for Tobacco Sentiment Analysis

Lihua Xie¹, Ni Tang¹, Qing Chen^{1,*} and Jun Li^{2,*}

¹Division of Information Center, Tobacco Sichuan Industry Co., Ltd., Chengdu, 610017, China

²Computer Network Information Center, Chinese Academy of Sciences, Beijing, 100190, China

*Corresponding Authors: Qing Chen. Email: chenqing@sctobacco.com; Jun Li. Email: lijun@cnic.cn

Received: 24 September 2024 Accepted: 03 December 2024 Published: 17 February 2025

ABSTRACT

In the age of information explosion and artificial intelligence, sentiment analysis tailored for the tobacco industry has emerged as a pivotal avenue for cigarette manufacturers to enhance their tobacco products. Existing solutions have primarily focused on intrinsic features within consumer reviews and achieved significant progress through deep feature extraction models. However, they still face these two key limitations: (1) neglecting the influence of fundamental tobacco information on analyzing the sentiment inclination of consumer reviews, resulting in a lack of consistent sentiment assessment criteria across thousands of tobacco brands; (2) overlooking the syntactic dependencies between Chinese word phrases and the underlying impact of sentiment scores between word phrases on sentiment inclination determination. To tackle these challenges, we propose the External Knowledge-enhanced Cross-Attention Fusion model, CITSA. Specifically, in the Cross Infusion Layer, we fuse consumer comment information and tobacco fundamental information through interactive attention mechanisms. In the Textual Attention Enhancement Layer, we introduce an emotion-oriented syntactic dependency graph and incorporate sentiment-syntactic relationships into consumer comments through a graph convolution network module. Subsequently, the Textual Attention Layer is introduced to combine these two feature representations. Additionally, we compile a Chinese-oriented tobacco sentiment analysis dataset, comprising 55,096 consumer reviews and 2074 tobacco fundamental information entries. Experimental results on our self-constructed datasets consistently demonstrate that our proposed model outperforms state-of-the-art methods in terms of accuracy, precision, recall, and F1-score.

KEYWORDS

Tobacco sentiment analysis; natural language processing; cross-attention fusion; external knowledge

1 Introduction

Sentiment analysis, also known as user opinion mining, involves the assessment and exploration of users' expressed viewpoints on various subjects [1]. In a highly competitive market, understanding consumer sentiment can provide a competitive edge. A report by Allied Market Research suggests that the global sentiment analysis market was valued at \$2.6 billion in 2021 and is projected to reach



\$10.5 billion by 2031, growing at a Compound Annual Growth Rate (CAGR) of 15.3%¹. This growth highlights the increasing reliance on sentiment analysis across various sectors to gain a competitive advantage.

Beyond marketing, sentiment analysis has other values. For instance, applying sentiment analysis to the tobacco industry is beneficial for public health and policy making. The World Health Organization (WHO) reports that tobacco kills more than 8 million people each year, with over 7 million of those resulting from direct tobacco use². Effective sentiment analysis can help policymakers understand public perception and the impact of tobacco control policies, potentially leading to more effective public health initiatives.

Overall, this research on sentiment analysis on the tobacco market aims to understand the attitudes of user groups towards specific events or matters, which holds great significance for policy-making, public sentiment monitoring, and product upgrading. Mainstream sentiment analysis tasks [2–4] predominantly rely on publicly available English datasets, such as the SemEval-2014 series and the social media platform Twitter. Due to the severe restrictions on publicity in the tobacco industry, conventional datasets fall short of meeting the specific needs of the tobacco industry, and further analysis based on specific datasets and advanced technical are needed to promote the beneficial development of the tobacco industry and related public policy-making.

In recent years, sentiment analysis has been applied in the tobacco industry [5–7]. Yang et al. [8] constructed a cigarette sentiment analysis dataset using two cigarette products. They employed a conventional online comment dictionary to compute sentiment scores and tendencies, overlooking the profound exploration of user comment information. Building upon a significantly expanded cigarette sentiment analysis dataset, Rui et al. [9] utilized an Attention Layer and a Bidirectional Long Short-Term Memory (BiLSTM) layer to capture contextual features of the text. Hao et al. [10] employed Bidirectional Encoder Representation from Transformers (BERT), Convolutional Neural Network (CNN), and BiLSTM models to comprehensively capture interactive feature representations of target sentences.

However, existing methods have two primary limitations: 1) Neglect of fundamental tobacco information. Current sentiment analysis models often overlook the influence of fundamental tobacco information, such as brand specifics and chemical compositions, which are crucial for consistent sentiment assessment across diverse tobacco products. 2) Overlooking syntactic dependencies. Most models fail to consider the syntactic dependencies between Chinese word phrases and the impact of sentiment scores between word phrases on overall sentiment inclination determination.

To address these challenges, we propose a Cross-Attention Fusion model enhanced with External Knowledge, termed CITSA, for the task of tobacco sentiment analysis. The model comprises three layers: the Initial Embedding Layer (IEL), the Cross Infusion Layer (CIL), and the Textual Attention Enhancement Layer (TAEL). Specifically, the IEL is utilized for initializing consumer comment information and specific external tobacco data. The CIL introduces a novel cross-attention fusion strategy to interactively merge these two data sources. Furthermore, the TAEL incorporates word phrase sentiment score information into the “text-tobacco info” fusion features through the integration of a graph convolution network (GCN) module and a textual attention module. The aim of this proposed model is to enhance sentiment analysis performance in the tobacco domain by leveraging external knowledge and intricate fusion strategies.

¹Emotion Analytics Market

²Tobacco Facts

Furthermore, due to the sensitivity of tobacco-related information, acquiring evaluative data specific to tobacco products proves to be challenging. Yanyue.com (accessed on 30 October 2024) is one of China's most professional cigarette review websites and has been operating for a decade. Compared to other platforms, this site contains approximately 90,000 reviews with fundamental information encompassing all cigarette products in China, including cigarette brand, price, type, tar content, and user ratings associated with each cigarette product, which provide a rich stream of raw data for sentiment analysis. In this study, we overcome this challenge by extracting citizen comment data from Yanyue.com. After subjecting the data to cleaning procedures, we manually annotated the evaluation data into three sentiment categories: positive, neutral, and negative. The main contributions are summarized as follows:

- A novel cross-fusion approach has been introduced to capture and enhance both coarse and fine-grained textual feature representations, aiming to combine consumer comment information and tobacco fundamental information effectively.
- Building upon the text-tobacco feature representation, we have further devised an emotion-oriented syntactic dependency graph. Through the incorporation of a GCN module, the sentiment-syntactic relationships are integrated into consumer comments. Additionally, we introduce a Textual Attention Layer to merge these two feature representations.
- We constructed a tobacco sentiment analysis dataset tailored for the Chinese language and gathered comprehensive basic information related to all mentioned cigarette products.
- Experimental outcomes demonstrate that the proposed model surpasses the performance of state-of-the-art models on our datasets. Specifically, the proposed CITSA model achieved an accuracy of 92.5%, which is a significant improvement over the next best-performing model, the hybrid of Robustly optimized BERT approach and Long Short-Term Memory (RoBERTa-LSTM), at 89.8%. In terms of precision, CITSA scored 92.8%, outperforming RoBERTa-LSTM, which had a precision of 90.1%. For recall, CITSA achieved 92.0%, surpassing RoBERTa-LSTM's 89.3%. CITSA's F1-score was 92.4%, which is higher than RoBERTa-LSTM's 89.7%.

The subsequent sections of this manuscript are structured as follows: [Section 2](#) provides an overview of pertinent literature. In [Section 3](#), an emotion analysis model incorporating external tobacco domain knowledge is introduced. [Section 4](#) delves into the experimental findings and associated validation experiments, while [Section 5](#) offers concluding remarks, along with a prospective outlook on future endeavors.

2 Related Work

In this section, we will analyze the current state of research in text sentiment analysis and attention mechanisms from both task and technical perspectives. We introduce them specifically as follows.

2.1 Sentiment Analysis

The objective of text sentiment analysis is to recognize the emotional polarity conveyed within a given text, making it a crucial research focus in the field of natural language processing. Numerous efforts revolve around three primary approaches: lexical-based rules, machine learning, and deep learning. Specifically, traditional text sentiment analysis models based on lexicons and rules [11–13] require the construction of specialized dictionaries, matching rules, and specific languages. This limitation restricts the model's applicability to small-scale scenarios, necessitating the creation of new strategies for novel issues and thereby demanding substantial human and financial resources.

Machine learning-based text-level sentiment analysis builds upon feature engineering and employs machine learning algorithms to establish mathematical models for classifying samples. For instance, in [14], a sentiment analysis algorithm based on Support Vector Machine (SVM) is proposed, and grid search techniques are utilized to optimize its classification performance. In [15,16], four classical machine learning algorithms, namely SVM, K-Nearest-Neighbor, Decision Trees, and Naive Bayes, are employed for sentiment classification on target tweets. In [17], the universal language model fine-tuning strategy is combined with SVM, which ensures the reliability of SVM to a great extent and significantly enhances sentiment classification accuracy. However, although machine learning-based algorithms to some extent address the limitations of traditional lexicon and rule-based methods, these approaches often rely on extensive data, leading to inefficiency and struggling to meet the demands of the big data era. Moreover, these methods struggle to acquire sufficient contextual information from target texts, greatly impacting the accuracy of classification.

With the rapid advancement of deep learning techniques, an increasing number of researchers have focused on utilizing deep learning algorithms to tackle text sentiment analysis problems, and they have achieved significant progress. Reference [18] introduced a BiLSTM model with self-attention mechanism and multi-channel features. This model creates distinct feature channels from existing linguistic knowledge and sentiment resources, enhancing sentiment information through attention mechanisms. Reference [19] developed an approach based on attention mechanisms and BiLSTM to address sentiment analysis challenges in large-scale microblog texts from real-world social networks. Reference [20] introduced a convolutional neural network model incorporating self-attention based on BiLSTM for subjective classification of user comments. The combination of CNN and BiLSTM facilitates the optimal extraction of essential semantics and contextual information. Reference [21] harnessed the strengths of sequential models and Transformers, presenting a hybrid deep learning approach that captures deep-level semantic features, thus enhancing sentiment analysis accuracy. Reference [22] proposed a graph neural network-based approach, integrating syntactic dependency relationships into target texts to reinforce intrinsic dependencies between words, thereby improving sentiment analysis precision.

In our paper, we propose an External Knowledge Enhanced Cross-Attention Fusion Model. Building upon foundational deep learning algorithms, we incorporate a diverse range of external knowledge sources, including tobacco basic information, sentiment lexicons, and Chinese syntactic dependencies. By leveraging cross-attention mechanisms and GCN modules, we integrate this external knowledge with the target text, enhancing the model's stability and specificity.

2.2 *Pre-trained Language Models*

Currently, most state-of-the-art text analysis methods leverage Pre-trained Language Models (PLMs). In the field of sentiment analysis, existing PLM strategies encompass traditional word embedding models like Global Vectors for Word Representation (GloVe) [23], as well as fine-tuning models like BERT [24]. Specifically, BERT employs a Transformer architecture, enabling it to capture bidirectional contextual information of words during pre-training. This characteristic empowers BERT-based sentiment analysis models [20,21,25] to more accurately capture word associations and semantic meanings within sentences. On the other hand, GloVe relies on word co-occurrence statistics, making models of this type [22,26–28] less effective at capturing bidirectional relationships between words. In our paper, we utilize the fine-tuning model BERT as the embedding representation for textual information.

3 Proposed Model

In this section, we provide a detailed introduction to the CITSA model. Specifically, the model's input consists of a sentence of n words represented as $S = [w_1, w_2, \dots, w_n]$, and m pieces of basic information about tobacco denoted as $L = [l_1, l_2, \dots, l_m]$. Our objective is to determine the user's emotional inclination y through the user's expressed opinions, corresponding sentiment scores, and associated tobacco-related information, where $y \in \{0, 1, 2\}$. As shown in Fig. 1, the CITSA model consists of three components: 1) Initial embedding layer. It utilizes a pre-trained model to initialize word embedding representations of user comments and representations of tobacco-related information. 2) Cross-infusion layer. It employs a cross-infusion strategy to fuse and align the text and basic information interactively. This allows the model to consider both the review text and external knowledge about the product, such as brand, price, and other attributes, which can have positive or negative sentiments associated with them. 3) Textual attention enhancement layer. It constructs user comment representations based on sentiment lexicons and enhances text-basic information features through an attention mechanism, followed by a fully connected layer for final sentiment prediction. This enables the model to understand the relationships between different phrases in the review and how they contribute to the overall sentiment, including mixed sentiments.

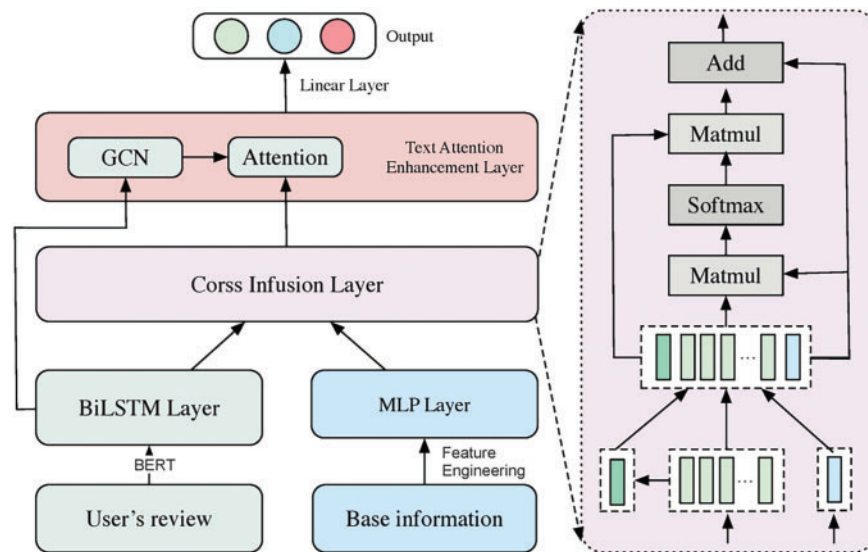


Figure 1: The overview of our model. On the left side, the model's comprehensive architectural diagram is depicted. On the right side, we delve into the internal structure of the Cross Infusion Layer, aiming to achieve a profound fusion between user review features and fundamental tobacco-related characteristics

In the subsequent sections, we will elaborate on these components' internal compositions.

3.1 Initial Embedding Layer

In this layer, we initialize the word embedding representations of user comments and tobacco-related information. Unlike English text, which uses spaces as separators, Chinese text lacks obvious delimiters. Therefore, prior to generating word embedding representations, we tokenize the target text using the “jieba” library, resulting in a sequence of k -word phrases denoted as $Q = [a_1, a_2, \dots, a_k]$. Based on this, we utilize the pre-trained BERT model to capture the initial representations of each

word phrase. Subsequently, the sentence embedding is denoted as $E_s = [b_1, b_2, \dots, b_k] \in \mathbb{R}^{k \times \text{dm}}$, where dm represents the dimension of word phrase embeddings.

Due to the potential disruption of the original contextual relationships by Chinese word segmentation, we introduce a bidirectional LSTM module after the pre-trained BERT model. This module fine-tunes the initial word phrase embeddings to mitigate the impact of segmentation on text context. For the i -th word phrase, the computation of the bidirectional hidden layer representation is as follows:

$$h_{lstm}^{\rightarrow} = LSTM(x_i, h_{i-1}^{\rightarrow}) \quad (1)$$

$$h_{lstm}^{\leftarrow} = LSTM(x_i, h_{i-1}^{\leftarrow}) \quad (2)$$

$$E_f = h_{lstm}^{\rightarrow} \oplus h_{lstm}^{\leftarrow} \quad (3)$$

where $h_{i-1}^{\rightarrow} \in \mathbb{R}^{N \times d}$ and $h_{i-1}^{\leftarrow} \in \mathbb{R}^{N \times d}$ represent the input of the forward and backward hidden layers obtained at the previous time step, respectively.

Furthermore, the comprehensive nature of tobacco's fundamental information, including aspects like cigarette brand, price, type, and user ratings, effectively captures the foundational characteristics of cigarettes. This plays a crucial role in assessing the sentiment of tobacco comments. To maximize the utilization of this information, we initialize each aspect as a discrete feature and align the dimensions of text and fundamental information through a multi-layer perceptron. This aligned input is denoted as $B = [p_1, p_2, \dots, p_k] \in \mathbb{R}^{k \times d}$, where k represents the multiple aspects of tobacco's fundamental information.

3.2 Cross Infusion Layer

As illustrated in Fig. 1 (the right segment), we have devised a word phrase-level cross-attention fusion strategy to delve into and fuse the underlying correlations between user comment information and tobacco-related basics. Specifically, for the user comment data, we begin by employing a pooling function to capture its coarse-grained discourse-level feature representation, denoted as E_c . This representation acts as a complement to the word phrase-level feature representation. Subsequently, we concatenate these six features together, which then serve as the input to our cross-fusion layer.

$$E_g = \text{concat}(E_f, E_e, E_c) \in \mathbb{R}^{(k+v+1) \times d} \quad (4)$$

where $\text{concat}(\hat{A} \cdot)$ denotes the concatenation function.

During the cross-fusion stage, a self-attention mechanism learning strategy is introduced to explore and understand the internal correlations of the fused features. Specifically, we initially learn the query/key/value feature mappings: $\text{query} = W_q E_g \in \mathbb{R}^{(k+v+1) \times d}$, $\text{key} = W_k E_g \in \mathbb{R}^{(k+v+1) \times d}$, and $\text{value} = W_v E_g \in \mathbb{R}^{(k+v+1) \times d}$, where W_q , W_k , and W_v represent three trainable weights. Subsequently, we compute the attention scores for the fused features and enhance the stability of the fusion features through masking:

$$\beta = \text{Softmax}(\text{query} * \text{ket}^T) \in \mathbb{R}^{(k+v+1) \times (k+v+1)} \quad (5)$$

$$E_t = \beta E_g \quad (6)$$

where E_t represents novel tobacco feature representation acquired through the self-attention mechanism.

Moreover, we combine the original features with the attention-based features using a residual strategy to obtain the fused features of text and tobacco fundamental information. Formally, the residual fusion features E_f use are calculated as follows:

$$E_{fuse} = E_f + E_g \quad (7)$$

3.3 Attention Layer

GCN Module: The ultimate objective of tobacco sentiment analysis is to determine the emotional inclination of comment text. Consequently, text plays a pivotal role in sentiment assessment, with syntactic relationships influencing the inherent emotional connections between text word phrases. This factor is equally vital for tobacco sentiment analysis. Drawing inspiration from the principles presented in the syntactic paper, we construct an emotion-aware syntactic dependency graph before text augmentation:

$$A[i, j] = \begin{cases} \text{SSW}(w_i, w_j), & \text{if node } i \text{ is connected to } j \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where $\text{SSW}(\cdot)$ represents the Sentiment Score of Words in the sentiment lexicon. It's worth noting that due to the fact that sentiment scores between words related by syntactic dependencies better express the emotional inclination of the text, we focus on computing sentiment scores only for word phrases that have syntactic relationships, for ease of calculation.

Moreover, we devised a GCN Layer that leverages a graph convolutional network to integrate syntactic dependencies into the fused feature representations. This helps in capturing the nuanced relationships between different aspects of the product mentioned in the review, such as how the flavor and price might be related in the context of the overall sentiment. Formally, the computation of the text representation h_l with syntactic dependency is as follows:

$$h_i^l = \sum_{j=1}^n \xi_i A_{ij}^c W^l \bar{e}_j^{l-1} + b^l \quad (9)$$

where $\xi_i = 1/(d_i + 1)$ is employed for matrix normalization, where d_i signifies the degree of the node. W^l and b^l are the trainable weight and bias of the l -th GCN layer. \bar{e}_j^{l-1} signifies the hidden layer representation of the previous GCN layer, while e_0 represents the initial feature representation of the user comments.

Attention Module: As the cross-fusion approach did not account for the impact of sentiment scores and syntactic relationships, we formulated a textual attention module to amend the fusion of text and tobacco information by incorporating comment text information encompassing sentiment scores and syntactic relationships. By focusing on computing sentiment scores for word phrases that have syntactic relationships, the model can better understand the context in which positive and negative sentiments are expressed. Concretely, the recalibrated fusion of text and tobacco information, denoted as E^z , is computed as follows:

$$\beta_v = \sum_{i=1}^n (e_v^f)^T e_i^f \quad (10)$$

$$\gamma_v = \frac{\exp(\beta_v)}{\sum_{j=1}^n \exp(\beta_j)} \quad (11)$$

$$E^z = \sum_{i=1}^N \gamma_i (E_{fuse})_i \quad (12)$$

where v is the initialized multimodal feature of text and image, $h_i^{G'}$ is the output of the GCN layer, and γ represents the attention score.

Finally, we employ the ultimate representation z to forecast the conclusive sentiment orientation $P(y|z)$ by means of the fully connected layer:

$$P(y|z) = \sigma(W_f z + b_f) \quad (13)$$

where W_f and b_f are the learnable weight and bias parameters, σ denotes an activation function such as Softmax, and y represents the predicted label.

3.4 Training

We train our CITSA model by the Adam algorithm [29] to minimize the following loss function:

$$\ell(\theta) = -\sum_{i=1}^S y_i \log \bar{y}_i + \lambda \|\theta\|_2^2 \quad (14)$$

where S is the number of samples, y_i and \bar{y}_i represent the ground truth and the predicted value, respectively. θ is the parameters that need to be trained and λ is the coefficient for L2-regularization.

4 Experiment

4.1 Datasets

In this study, we focused on 2344 tobacco brands available on the Chinese online platform “Yanyue.com” as our research subject. Utilizing Python, we conducted web scraping to gather a total of 84,770 consumer comments and 2344 entries of tobacco fundamental information from the years 2012 to 2023. Furthermore, we employed techniques such as regular expression matching and string manipulation to perform an in-depth data-cleaning process. This process primarily revolved around removing advertisements, eliminating duplicate comments, stripping away webpage tags, and filtering out data containing only numerical or special characters, which were deemed irrelevant. To ensure data consistency and usability, we also employed methods such as capitalization conversion, simplified traditional character conversion, and numerical normalization.

Following these processes, we constructed a tobacco consumer review dataset containing 55,096 consumer comments and 2074 entries of tobacco fundamental information. We manually annotated these review data, categorizing them into three sentiment inclinations: positive, neutral, and negative. To facilitate model training and evaluation, we randomly divided the dataset into training and testing sets in a 4:1 ratio, ensuring the model’s performance was validated across diverse data. Statistics of the datasets are summarized in Table 1, providing a detailed distribution of data for both the training and testing sets. All the web crawling data involved in this study were collected at a controlled frequency. Additionally, the collected data underwent preprocessing (data anonymization) and strictly adhered to the realm of academic research.

Table 1: Statistic of datasets

Sentiment polarity	Sample size	Tobacco	Basic tobacco information	Review
Positive (2)	Train: 18,426	Yunyan (impression brown)	Content: Tar (10 mg), Alkaline (1.0 mg), Carbon Monoxide (10 mg) Packaging: Paper Hard Pack Ratings in Various Dimensions: Flavor (7.8), Appearance (7.6), Value for Money (6.5), Overall (7.3)	I really love this cigarette; it’s a standout product from Yunyan, with its cigar-like design offering a unique appeal, making it my top choice in its price range.
	Test: 4588			

(Continued)

Table 1 (continued)

Sentiment polarity	Sample size	Tobacco	Basic tobacco information	Review
Neutral (1)	Train: 10,374	Liqun (Blue Sky)	Content: Tar (10 mg), Alkaline (1.0 mg), Carbon Monoxide (11 mg) Packaging: Paper Hard Pack Ratings in Various Dimensions: Flavor (7.5), Appearance (7.1), Value for Money (6.7), Overall (7.1)	Today I bought a pack and noticed the warning label had been changed to: "Our company reminds you that smoking is harmful to health. Please do not smoke in non-smoking areas." I want to ask, did I buy a fake one?
Negative (0)	Test: 2862 Train: 15,276	Jia Tianxia (Landscape)	Content: Tar (8 mg), Alkaline (0.5 mg), Carbon Monoxide (11 mg) Packaging: Paper Hard Pack Ratings in Various Dimensions: Flavor (4.4), Appearance (3.8), Value for Money (7.2), Overall (4.9)	This cigarette is an inconsistent product, hastily put together by Guangxi Zhongyan. It's essentially just a rebranded version of a premium product, making it a knockoff of the famous Jiatiantian product.
	Test: 3570			

Note: Among these are three sentiment orientations: Positive, Neutral, and Negative. Furthermore, we present the manifestations of each sentiment category across three attributes, namely, cigarette brand names, fundamental cigarette information, and user comments.

4.2 Experimental Setup

In the experimental configuration, the BERT pre-trained model is harnessed to capture the representation of textual content and aspect-based features. This architecture incorporates a total of 12 encoder layers, each endowed with a dimensionality of 768. The hidden dimension is 300. The batch size is determined as 32, and a dropout rate of 0.2 is introduced within the model architecture to mitigate overfitting. Performance evaluation is based on four key metrics, namely Accuracy, Precision, Recall, and F1-score.

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{PN}} \quad (15)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (16)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (17)$$

$$\text{F1 - score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (18)$$

where TP (True Positives), FP (False Positives), TN (True Negatives), and PN (False Negatives) are defined as presented in Table 2.

Furthermore, we leverage the Stanford CoreNLP toolkit³ for the extraction of syntactic relationships within sentences, thereby enabling the incorporation of linguistic context. For model weight initialization, a Xavier normal distribution strategy is employed to ensure balanced weight settings conducive to efficient learning.

³Version 4.3.0 procured from <https://stanfordnlp.github.io/CoreNLP/> (accessed on 02 October 2024).

Table 2: The confusion matrix

Confusion		Actual values	
		Positive	Negative
Predicted values	Positive	True positive (TP)	False positive (FP)
	Negative	False negative (FP)	True negative (TP)

The experimental execution is facilitated through the PyTorch framework, providing an adaptive computational environment, and executed on an NVIDIA Tesla V100 GPU, optimizing computational efficiency and performance. This rigorous experimental setup is designed to ensure the validity and replicability of our research outcomes, aligning with the standards of scholarly research in the field.

4.3 Baseline

To thoroughly evaluate the effectiveness of the proposed CITSA model, we conducted comparative evaluations against several representative baselines.

- SVM [14]: employs Support Vector Machines and grid search techniques for classifying target text.
- CNN-biLSTM [20]: utilizes the CNN and BiLSTM model to learn classification traits and capture crucial semantic and contextual cues for sentiment polarity determination.
- CNN-Capsules [30]: employs Recurrent Neural Network (RNN) to extract text features and devise a capsule strategy to activate the module with the highest attention probability for predicting the final sentiment polarity.
- BiLSTM-Attention [31]: devises a BiLSTM layer and an Attention layer to learn the representation of textual context
- RCNNGWE [32]: proposes a bidirectional convolutional recurrent neural network model that utilizes two temporal models and a convolutional layer to capture deep-seated semantic features of the text.
- GCN [22]: integrates syntactic dependency relationships into text representations using a graph neural network and employs LSTM as a state updater to filter node noise.
- RoBERTa-LSTM [21]: proposes a hybrid deep learning model that incorporates the RoBERTa and LSTM models to ensure comprehensive text feature extraction.

4.4 Experience Results and Analysis

In this section, we present the comparison results between our model and the baseline models in Table 3. To ensure a fair comparison among all models, we have re-implemented the models for which no publicly available code was provided. All experimental results are obtained from our proprietary dataset. From Table 3, we make the following observations.

Compared to traditional machine learning-based models, our model shows significant improvements in all four evaluation metrics. The experimental results exhibit some fluctuations in these metrics, indicating that relying solely on SVM may struggle to maintain stability and reliability in sentiment analysis tasks. Furthermore, compared to CNN-based models (CNN-BiLSTM) and attention-based

models (RNN-Capsule, BiLSTM-Attention, and RCNNGWE), attention-based models achieved better performance improvements. We attribute this mainly to the attention mechanism's ability to capture relevant word phrase features more effectively.

Table 3: Experimental results (%). Comparisons for all baselines on five datasets. The best performances are shown separately in bold

Model	Accuracy	Precision	Recall	F1-score
SVM+	70.4	72.3	71.2	71.8
CNN-BiLSTM	80.7	84.2	81.1	82.7
RNN-Capsule	83.6	84.1	83.3	83.7
BiLSTM-Attention	83.3	85.2	84.8	85.0
RCNNGWE	85.7	87.7	87.3	87.5
GCN-LSTM	86.2	86.8	84.7	85.7
RoBERTa-LSTM	89.8	90.1	89.3	89.7
w/o BAI	89.7	90.3	89.3	89.7
w/o SSDG	90.8	91.0	90.3	90.6
w/o SenScore	91.2	91.5	91.0	91.3
w/o CIL	90.0	90.2	89.6	89.9
Our(CITSA)	92.5	92.8	92.0	92.4

Different from classical Deep Neural Networks (DNN) models, the GCN-based model GCN-LSTM has made significant progress due to its effective incorporation of syntactic dependency relationships. Building upon this model, our approach enhances sentiment information by constructing a sentiment-aware syntactic dependency graph, enriching the external knowledge. Experimental results demonstrate substantial improvements over the GCN-LSTM model. Moreover, we observed that our model outperforms the RoBERTa-LSTM model, which suggests that the RoBERTa-LSTM model may struggle to make accurate predictions for specific tobacco-related tasks. This also underscores the importance of our tobacco baseline information in enhancing sentiment analysis.

4.5 Parameter Study

To empirically evaluate the effectiveness and theoretical justification of the proposed model, a comprehensive parameter study is undertaken. We investigate the impact of the number of layers and the quantity of hidden units in graph convolutional networks on the performance of our proposed model.

4.5.1 Impact of GCN Layer

The GCN layer plays a crucial role in capturing the sentiment of word phrases and the syntactic dependency relationships among them. Additionally, the number of GCN layers significantly affects the degree of feature learning. In this section, we conducted experiments on our self-acquired dataset to investigate the impact of the number of GCN layers on the model's classification performance. As shown in Table 4, our model achieves the best classification results when the number of GCN layers is set to 2. This demonstrates that two GCN layers are sufficient to effectively integrate the features of consumer review texts and sentiment-based syntactic dependency graphs.

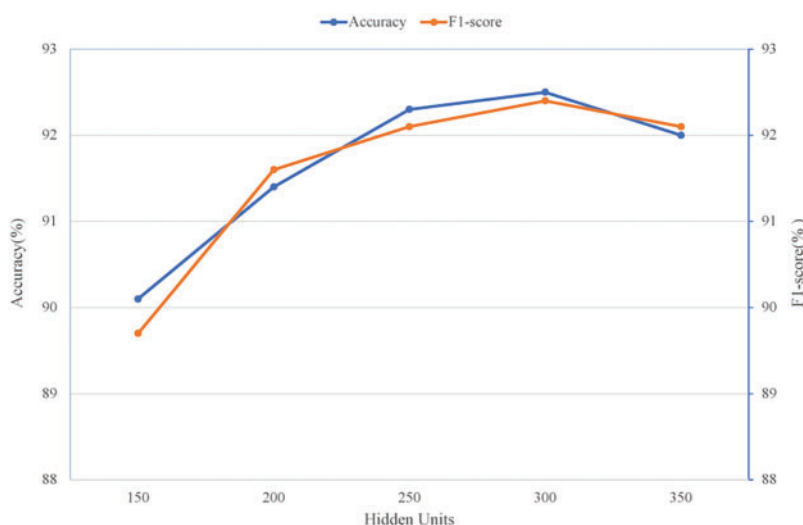
Table 4: Impact of GCN layers on our datasets

Layers	Accuracy	Precision	Recall	F1-score
1	91.7	91.9	91.7	91.8
2	92.5	92.8	92.0	92.4
3	90.6	90.9	90.2	90.5
4	89.3	89.7	89.5	89.5

Through our analysis, we found that a single layer of GCN cannot fully capture the complex relationships between the graph structure and text features, making it difficult to understand the intricate dependencies between words in sentences, leading to a decrease in performance. Additionally, as the number of GCN layers increases, the classification performance of the model also tends to decrease. We attribute this to the fact that an increase in the number of layers results in very small gradients during backpropagation, causing over-smoothing of feature extraction, and thereby making training unstable.

4.5.2 The Impact of Hidden Units

In the CITSA model, all hidden units are assigned the same quantity, rendering the selection of this value critical for the model's performance. To assess its impact while maintaining other parameters constant (with the GCN layer count fixed at 2), we incrementally varied the number of hidden units from 150 to 350 in steps of 50. The resulting performance trends are presented in Fig. 2. The data indicate that a configuration of 300 hidden units yields optimal performance across the majority of tasks. However, increasing the number of hidden units beyond this threshold leads to a significant escalation in both time and spatial complexity, suggesting that a value of 300 hidden units constitutes a well-balanced and effective choice for the model.

**Figure 2:** The impact of hidden units on the CITSA model

4.5.3 Convergence Analysis of Model Accuracy

As shown in Fig. 3, the accuracy trends indicate a solid progression toward the model's optimal performance, with training accuracy reaching close to 98% and testing accuracy stabilizing around 92% by the end of training. Notably, accuracy improves rapidly in the initial epochs, particularly on the training set, demonstrating efficient learning. Beyond approximately 100 epochs, the incremental gains in accuracy diminish, indicating that the model is nearing convergence. This plateau suggests that the model has effectively balanced fitting the training data while maintaining generalizability, as seen in the alignment of training and testing accuracies. The stable trajectory in the final epochs, without substantial fluctuations, underscores a low risk of overfitting, confirming the model's suitability for reliable sentiment analysis applications.

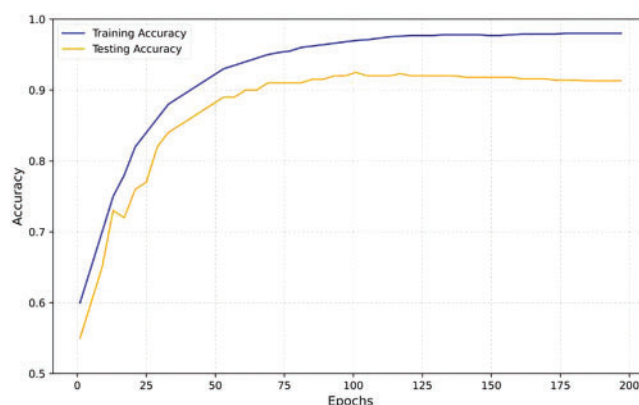


Figure 3: Training and testing accuracy curves of the CITSA model

4.6 Ablation Study

To further analyze and understand the impact of different components of our model on the sentiment analysis task, we conducted an ablation study with three ablation models in Table 3. We systematically removed or modified specific components of the model to assess and analyze the changes in results. The ablation models include: 1) CITSA without tobacco basic information (w/o BAI). 2) CITSA without sentiment-based syntactic dependency graph (w/o SSDG). 3) CITSA without sentiment score (w/o SenScore). 4) CITSA without cross-infusion layer (w/o CIL).

Effect of tobacco basic information: In the absence of tobacco basic information, we relied solely on consumer review text to make sentiment judgments. As shown in Table 3, all four evaluation metrics experienced a noticeable decrease in performance without this component. We attribute this decline to the complexity of consumer review texts, which often involve opinions about multiple cigarette brands. Tobacco basic information serves as an anchor, allowing the model to center its analysis on the target cigarette, thus reducing ambiguity. This background information provides critical details about the product, helping the model differentiate between brands and focus its sentiment analysis more accurately. In essence, it complements consumer reviews by providing specificity, making the analysis more robust and context-aware.

Effect of sentiment-based syntactic dependency graph: To compare the impact of external sentiment scores and syntactic dependency relationships, we directly bypassed the Text Attention Enhancement Layer and used the output of the Cross Infusion Layer as the final feature representation. Compared to our full model, this ablation model experienced performance decreases of 1.84%, 1.94%, 1.85%,

and 1.95% across the four evaluation metrics, demonstrating the significance of the sentiment-based syntactic dependency graph in our model. Without this layer, the model struggles to capture these complexities, showing the syntactic dependency graph's instrumental role in analyzing mixed or layered sentiments within consumer reviews.

Effect of sentiment score: When we considered only syntactic information to construct the syntactic dependency graph, our model's performance experienced a significant decrease of approximately 1.1% in terms of Accuracy (ACC) metric. This indicates that sentiment scores play a crucial role in assisting tobacco sentiment analysis, particularly affecting judgments related to neutral sentiments.

Effect of cross-infusion layer: We replaced the cross-infusion layer with a feature concatenation approach, and the experimental results showed a significant decrease compared to our model, with a reduction of 2.70%, 2.80%, 2.61%, and 2.71% across the four evaluation metrics. This substantial decrease underscores the critical role of the cross-infusion strategy in effectively fusing deep textual and tobacco basic information. Specifically, the cross-infusion layer is crucial for a synergistic fusion of text-based and tobacco-specific information, far surpassing the superficial combination achieved by simple feature concatenation. By facilitating dynamic interactions between these information sources, the cross-infusion layer captures deeper correlations and context-specific patterns, enhancing coherence in the final representation.

4.7 Limitations of the Dataset

In this paper, the data was collected from [Yanyue.com](https://www.yanyue.com), a single online platform, which may not be fully representative of all consumer sentiments toward tobacco products. Users who choose to post reviews on this platform may have different characteristics and biases compared to the general population of tobacco product users. In addition, the dataset is specific to the Chinese market and language, which limits the model's applicability to other cultural and linguistic contexts.

Regarding the threats to validity, the model may struggle to accurately capture complex sentiments, such as sarcasm or nuanced opinions, which are not well-captured by simple positive, neutral, and negative categories.

4.8 Other Languages

Even though CITSA is primarily designed for Chinese language sentiment analysis, it can also work well with other languages for the following reasons: 1) The model leverages a deep learning architecture that includes BERT, which has been pre-trained on a large corpus of text and is known for its ability to capture contextual relationships between words. This makes it potentially adaptable to other languages, especially if those languages are supported by BERT's multilingual versions. 2) The cross-attention mechanism used in CITSA is a general approach that can be applied to any language, as it focuses on aligning and fusing different types of information (e.g., consumer comments and product information), which is a language-agnostic process. 3) The GCN module, which integrates syntactic dependencies, can be adapted to other languages as long as the syntactic dependency structures for those languages are known and can be parsed.

Nevertheless, there is some space to improve the model to better adapt to other languages. For example, we need to replace the Chinese-specific BERT with a multilingual BERT variant that has been pre-trained on a diverse range of languages, allowing the model to understand and process different languages effectively.

5 Conclusion

In this paper, we propose a Cross Infusion Attention Fusion Model (CITSA), which marks a significant stride in the field of sentiment analysis, particularly within the nuanced domain of tobacco industry consumer reviews. Our model's integration of external knowledge and cross-attention mechanisms has yielded substantial improvements in accuracy, precision, recall, and F1-score, demonstrating its effectiveness in capturing the subtleties of consumer sentiment. CITSA's ability to fuse consumer comments with fundamental tobacco information through interactive attention mechanisms provides a more comprehensive understanding of sentiment inclinations. This approach surpasses traditional methods by offering a deeper, context-aware analysis. By incorporating sentiment-syntactic relationships into consumer comments via a graph convolution network module, CITSA delivers a nuanced understanding of the text's structure and sentiment, which is crucial for languages like Chinese with complex syntactic structures.

Although our model has achieved promising results, there are still some limitations. Firstly, due to the specificity and sensitivity of tobacco-related discussions, consumers often express their sentiments with a tone of irony or sarcasm, making it challenging for sentiment analysis alone to fully capture consumers' attitudes toward cigarettes. We plan to further research the integration of sentiment analysis and sarcasm detection to address this challenge. Additionally, we solely focused on sentence-level tobacco sentiment analysis, overlooking the multi-dimensional sentiment attitudes that consumers may have towards cigarettes. In the future, we intend to reorganize the dataset, shifting our focus to aspect-level sentiment analysis of cigarettes. Leveraging the foundational framework and approach of this model, we aim to expand its applicability.

Acknowledgement: The authors gratefully acknowledge the constructive feedback and insightful comments provided by the reviewers and editors, which significantly contributed to the improvement of this manuscript.

Funding Statement: This work was supported by the Global Research and Innovation Platform Fund for Scientific Big Data Transmission (Grant No. 241711KYSB20180002) and National Key Research and Development Project of China (Grant No. 2019YFB1405801).

Author Contributions: The authors confirm their contribution to the paper as follows: methodology and writing draft preparation, Lihua Xie; writing review and editing, Ni Tang; supervision, Qing Chen and Jun Li. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

- [1] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
- [2] H. Li, Q. Chen, Z. Zhong, R. Gong, and G. Han, "E-word of mouth sentiment analysis for user behavior studies," *Inform. Process. Manag.*, vol. 59, 2022, Art. no. 102784. doi: [10.1016/j.ipm.2021.102784](https://doi.org/10.1016/j.ipm.2021.102784).

- [3] Z. B. Nezhad and M. A. Deihimi, "Twitter sentiment analysis from iran about COVID 19 vaccine," *Diabet. Metabol. Syndr.: Clinic. Res. Rev.*, vol. 16, no. 1, 2022, Art. no. 102367. doi: [10.1016/j.dsx.2021.102367](https://doi.org/10.1016/j.dsx.2021.102367).
- [4] C. Singh, T. Imam, S. Wibowo, and S. Grandhi, "A deep learning approach for sentiment analysis of COVID-19 reviews," *Appl. Sci.*, vol. 12, no. 8, 2022, Art. no. 3709. doi: [10.3390/app12083709](https://doi.org/10.3390/app12083709).
- [5] E. C. Resende and A. Culotta, *A Demographic and Sentiment Analysis of E-Cigarette Messages on Twitter*. Computer Science Department, Illinois Institute of Technology; 2015.
- [6] J. Lin and Z. Jian, "Emotional analysis of cigarette consumers based on cnn and bilstm deep learning model," *J. Phy.: Conf. Series*, vol. 1651, no. 1, 2020, Art. no. 012102. doi: [10.1088/1742-6596/1651/1/012102](https://doi.org/10.1088/1742-6596/1651/1/012102).
- [7] A. Alshamsi, R. Bayari, and S. Salloum, "Sentiment analysis in english texts," *Adv. Sci., Technol. Eng. Syst. J.*, vol. 5, no. 6, pp. 1683–1689, 2021.
- [8] C. Yang, H. Zhang, J. Huang, and J. Wan, "Text sentiment analysis of online cigarette reviews," *Acta Tabacaria Sinica*, vol. 26, no. 2, pp. 92–100, 2020.
- [9] W. Rui *et al.*, "An emotion classification method for cigarette consumers' evaluation based on combination of BILSTM and attention mechanism," *Tobacco Sci. Technol.*, vol. 55, no. 11, pp. 106–112, 2022.
- [10] Y. Hao, T. Yang, C. Shi, R. Wang, and D. Xiao, "An effective sentiment analysis model for tobacco consumption," in *Proc. 2022 11th Int. Conf. Comput. Pattern Recognit.*, 2022, pp. 496–502.
- [11] Y. Li, L. Wang, Y. Chai, and Z. Liu, "Research on construction method of dynamic sentiment dictionary based on bidirectional LSTM," *J. Chin Comput. Syst.*, vol. 40, no. 3, pp. 503–509, 2019.
- [12] Y. Wang, M. Wang, and H. Fujita, "Word sense disambiguation: A comprehensive knowledge exploitation framework," *Knowl. Based Syst.*, vol. 190, 2020, Art. no. 105030. doi: [10.1016/j.knosys.2019.105030](https://doi.org/10.1016/j.knosys.2019.105030).
- [13] Z. Jiawa, L. Wei, W. Sili, and Y. Heng, "Review of methods and applications of text sentiment analysis," *Data Anal. Knowl. Disc.*, vol. 5, no. 6, pp. 1–13, 2021.
- [14] M. Ahmad, S. Aftab, M. S. Bashir, N. Hameed, I. Ali and Z. Nawaz, "SVM optimization for sentiment analysis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 4, 2018. doi: [10.14569/IJACSA.2018.090455](https://doi.org/10.14569/IJACSA.2018.090455).
- [15] M. Alzyout, E. A. Bashabsheh, H. Najadat, and A. Alaiad, "Sentiment analysis of arabic tweets about violence against women using machine learning," in *2021 12th Int. Conf. Inform. Commun. Syst. (ICICS)*, IEEE, 2021, pp. 171–176.
- [16] F. Jemai, M. Hayouni, and S. Baccar, "Sentiment analysis using machine learning algorithms," in *2021 Int. Wireless Commun. Mobile Comput. (IWCMC)*, IEEE, 2021, pp. 775–779.
- [17] B. AlBadani, R. Shi, and J. Dong, "A novel machine learning approach for sentiment analysis on twitter incorporating the universal language model fine-tuning and SVM," *Appli. Syst. Innov.*, vol. 5, no. 1, 2022, Art. no. 13. doi: [10.3390/asi5010013](https://doi.org/10.3390/asi5010013).
- [18] W. Li, F. Qi, M. Tang, and Z. Yu, "Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification," *Neurocomputing*, vol. 387, no. 4, pp. 63–77, 2020. doi: [10.1016/j.neucom.2020.01.006](https://doi.org/10.1016/j.neucom.2020.01.006).
- [19] P. Wu, X. Li, C. Ling, S. Ding, and S. Shen, "Sentiment classification using attention mechanism and bidirectional long short-term memory network," *Appl. Soft Comput.*, vol. 112, 2021, Art. no. 107792. doi: [10.1016/j.asoc.2021.107792](https://doi.org/10.1016/j.asoc.2021.107792).
- [20] P. Bhuvaneshwari, A. N. Rao, Y. H. Robinson, and M. Thippeswamy, "Sentiment analysis for user reviews using Bi-LSTM self-attention based CNN model," *Multimed. Tools Appl.*, vol. 81, no. 9, pp. 12 405–12 419, 2022. doi: [10.1007/s11042-022-12410-4](https://doi.org/10.1007/s11042-022-12410-4).
- [21] K. L. Tan, C. P. Lee, K. S. M. Anbananthen, and K. M. Lim, "Roberta-LSTM: A hybrid model for sentiment analysis with transformer and recurrent neural network," *IEEE Access*, vol. 10, no. 4, pp. 21 517–21 525, 2022. doi: [10.1109/ACCESS.2022.3152828](https://doi.org/10.1109/ACCESS.2022.3152828).
- [22] Y. Li and N. Li, "Sentiment analysis of weibo comments based on graph neural network," *IEEE Access*, vol. 10, no. 2, pp. 23 497–23 510, 2022. doi: [10.1109/ACCESS.2022.3154107](https://doi.org/10.1109/ACCESS.2022.3154107).
- [23] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in *Proc. 2014 Conf. Empiric. Meth. Nat. Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.
- [24] J. Devlin, M. -W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

- [25] B. Jiang, J. Hou, W. Zhou, C. Yang, S. Wang and L. Pang, “METNet: A mutual enhanced transformation network for aspect-based sentiment analysis,” in *Proc. 28th Int. Conf. Comput. Linguist.*, 2020, pp. 162–172.
- [26] D. Ma, S. Li, X. Zhang, and H. Wang, “Interactive attention networks for aspect-level sentiment classification,” 2017, *arXiv:1709.00893*.
- [27] C. Zhang, Q. Li, and D. Song, “Aspect-based sentiment classification with aspect-specific graph convolutional networks,” 2019, *arXiv:1909.03477*.
- [28] Z. Zhao, Y. Liu, J. Gao, H. Wu, Z. Yue and J. Li, “Multi-grained syntactic dependency-aware graph convolution for aspect-based sentiment analysis,” in *2022 Int. Joint Conf. Neural Netw. (IJCNN)*, IEEE, 2022, pp. 1–8.
- [29] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, *arXiv:1412.6980*.
- [30] Y. Wang, A. Sun, J. Han, Y. Liu, and X. Zhu, “Sentiment analysis by capsules,” in *Proc. 2018 World Wide Web Conf.*, 2018, pp. 1165–1174.
- [31] L. Xiaoyan and R. C. Raga, “BiLSTM model with attention mechanism for sentiment classification on Chinese mixed text comments,” *IEEE Access*, vol. 11, no. 6, pp. 26199–26210, 2023. doi: [10.1109/ACCESS.2023.3255990](https://doi.org/10.1109/ACCESS.2023.3255990).
- [32] A. Onan, “Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification,” *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 5, pp. 2098–2117, 2022.