International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I9P105

Volume 74 Issue 9, 50-62, September 2026
© 2026 Seventh Sense Research Group®

Original Article

Unified Binary Aspect-Based Sentiment Analysis with Transformer Models: A Cross-Domain Evaluation of Models

Ayesha Siddiqua[1], H C Nagaraj[2]

[1,2]*Department of ECE, Nitte Meenakshi Institute of Technology (NMIT), Nitte (Deemed to be University), Bengaluru, India.*
[1,2]*Nitte Research Education Academy, University of Mysore, Mysuru, Karnataka, India.*

[1]*Corresponding Author : leo.ayesha@gmail.com*

Received: 10 March 2026 Revised: 06 August 2026 Accepted: 21 August 2026 Published: 30 September 2026

Abstract - *Aspect-Based Sentiment Analysis (ABSA) is a Fine-grained Sentiment Classification task that aims to detect the sentiment polarities expressed towards specific aspects in the text. ABSA has also become prevalent in applications of product feedback mining, brand monitoring and recommendation systems due to the growth of user-generated feedback spread on social media posts. The conventional forms of deep learning, like ATAE-LSTM and TD-LSTM, have worked sufficiently but have constraints because of shallow context perception and the inability to provide long-term dependencies. The examination performs and benchmarks the utilization of current transformer-based structures, Bidirectional Encoder Representations from Transformers (BERT), Robustly Optimized BERT Pretraining Approach (RoBERTa)and Decoding-enhanced BERT with Disentangled Attention (DeBERTa) to binary ABSA classification, i.e., distinguishing between positive and negative sentiments but ignoring neutral labels. The approach will be to encode aspect terms by using special tokens and fine-tuning models using benchmark data such as SemEval (Laptop & Restaurant) as well as the Multi-modal Aspect-based Sentiment Analysis (MAMS) dataset. In addition, a comparative literature survey is made among the recent models, including multimodal advanced structures, which confuse text and visual information. Benchmarking experimental testing (SemEval-Laptop, SemEval-Restaurant, and MAMS) demonstrates that DeBERTa-base had an accuracy of 92.45 and an F1-score of 92.44, compared with RoBERTa-base with 91.89 accuracy and BERT-base with 91.05 accuracy.*

Keywords - *ABSA, BERT, RoBERTa, DeBERTa, Transformer models, Sentiment classification, MAMS dataset, SemEval, Disentangled attention, Relative positional encoding, Pre-trained language models, Multimodal sentiment analysis.*

1. Introduction

With the exponential increase of digital content in the form of reviews and opinions posted on e-commerce sites and social networks, Sentiment Analysis has come up as a core operation in Natural Language Processing (NLP).

Whereas general sentiment analysis attributes an approximate level of overall polarity to a sentence or document, ABSA measure the sentiment of each aspect word in a sentence or document, e.g., a product review will say that battery life is positive or negative.

The traditional deep learning-based approaches like ATAE-LSTM [1], TD-LSTM [2], TNet [3], and LCR-Rot-hop +[4] formed the basis of ABSA by adding the attention mechanisms and memory networks to capture the context-aspect relationships. These models, however, fail to capture long-range dependencies and generalising between domains because they are sequential in nature.

The transformer-based models are introduced, and especially LDA [5] has been revolutionary to ABSA. Following this, models such as SpanBERT [6] and R-GAT + BERT [7] expanded on this idea by encoding more aspect spans and dependency structures and improving the classification accuracy by a significant margin.

Later on, the field developed to multimodal ABSA, which should adopt visual signals, including pictures, emojis, or facial expressions, in addition to text.

This is particularly applicable in social media, where posts are mostly accompanied by some visuals, which contribute to sentiment. Multimodal fusion models like AMSAM-MFE [8], ABSA-TLRF [9] and GLFFCA [10] have shown that multimodal fusion can be made successful through approaches like cross-modal attention and integration by tensors, and they have achieved new standards in the performance of ABSA in the real world.

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

In this study, an all-embracive implementation and assessment of three leading transformer-based BERT-base, RoBERTa-base, and DeBERTa-base models are provided to achieve one unified binary ABSA activity. These models are trained and tested on various benchmark datasets, such as SemEval Laptop, SemEval Restaurant and MAMS and the input formatting is formatted to emphasize the aspect terms.

Transformer-based models have indeed made impressive progress in Aspect-Based Sentiment Analysis; however, research problems still need to be solved. Generally, different preprocessing approaches, experimental settings, and benchmark datasets hamper direct performance comparison of most works in reproduction of the architecture (i.e e.g. using some other variation of a particular transformer architecture along with its leveraging capabilities, such as tokenization). Additionally, most work centers on multiclass sentiment classification or domain-specific datasets, while few contribute to a unified binary classification framework across multiple benchmark datasets trained under the same conditions. The impact of explicit aspect-term tagging and attention processes in various transformer architectures (but particularly DeBERTa with disentangled attention and relative positional encoding) has yet to be thoroughly examined using a common experimental framework. To overcome these challenges, preprocessing needs to be standardized, training strategies, metrics for evaluation, and datasets to allow for a reliable comparison between existing transformer models.

The contribution describes a new universal transformer-based Aspect-Based Sentiment Analysis framework to compare BERT-base, RoBERTa-base and DeBERTa-base under the same preprocessing, training, and evaluation conditions. In contrast to earlier works based on different data sets or experimental protocols, the proposed framework conducts a fair comparative analysis using three broad benchmark datasets: SemEval Laptop, SemEval Restaurant, and MAMS. Moreover, explicit aspect-term tagging with regard to domain-specific aspect markers is utilized for promoting interaction between the captured aspects and contexts in the fine-tuning phase. The proposed framework is also evaluated experimentally on all submodels, with ablation tests and mild comparisons to the most recent works in sentiment classification.

1.1. Key Contributions

This includes:

1) A single unified binary Aspect-Based Sentiment Analysis framework for transformer-based language models over several benchmarks, including Laptop, Restaurant, and MAMS domains, respectively.
2) Explicit aspect tagging: For reinforcing aspect-term attention, we applied explicit aspect tagging based on custom [ASP] and [/ASP] tokens to train models that more accurately represent context relevant to the target.
3) A comparative evaluation of three widely-adopted transformer architectures, namely BERT-base, RoBERTa-base, and DeBERTa-base, on the same preprocessing, training and evaluation settings.
4) Experimental results show that DeBERTa-base outperforms our baseline because disentangled attention and relative positional encoding enable better learning of aspect-context dependencies.

The remaining part of the paper is outlined in sections. Section 2 outlines work-related and classical methods of ABSA. Section 3 describes the research proposal of the methodology and pipeline of data processing. Section 4 talks about the experimental setup and findings. Section 5 ends with findings, limitations and future research directions.

2. Related Work

The ABSA field has changed a lot in that it has moved away from the old-fashioned sequence modeling methods to the new transformer-based and multimodal learning models. This part evaluates the most prominent ABSA models, which are classified as classical deep learning, transformer-based, and more recent multimodal innovation, based on their approaches, advantages, and drawbacks.

2.1. Classical Neural Models of this Problem

Recurrent networks with attention mechanisms were applied to condition sentiment classification on aspect terms, Early in works such as ATAE-LSTM and TD-LSTM. The idea was that attention-aware embeddings of ATAE-LSTM were achieved through the concatenation of the aspect vectors with input words that presented reasonable results on benchmark datasets [1]. Conversely, TD-LSTMs divided the sentence into right and left contexts around the aspect term that were handled by different Long Short-Term Memory (LSTMs) [2]. However, the two methods were restricted in the ability to describe long-range dependencies as well as syntactic structures.

TNet, introduced by Li et al., combined transformation layers and CNNs with memory networks to remember aspect-aware context features more optimally [3]. Likewise, LCR-Rot-hop++ was a left-center-right organization using rotational memory to model multi-hop contexts and demonstrated better contextual knowledge [4]. These classical models, however, have a number of limitations:

- ATAE-LSTM is a model that is based on local attention and has problems with long-distance dependencies.
- TD-LSTM separates left/right context, and it undermines the comprehension of a sentence as a whole.
- TNet enhances situation awareness and is not syntactically rich.
- LCR-Rot-hop++ makes computation more complex with little cross-domain generalization.

With these weaknesses, the move towards transformer-based architectures with enhanced abilities to model context became motivated.

2.2. Transformer-Based Models

ABSA had a paradigm shift with the introduction of BERT. The BERT-SPC model also re-modeled the input to be pairs of aspects and sentences, and it could focus on aspect-specific sentiment more accurately with self-attention mechanisms [5]. SpanBERT was built upon BERT by training on span-level masking, which helps the model to learn more about aspect spans and their context interplay, which results in better generalization [6].

R-GAT + BERT architecture uses the syntactic dependency structure and BERT embeddings in addition to graph attention networks to give a structured contextualization to sentiment classification [7]. These strategies were much more successful compared to conventional models in terms of accuracy and F1-scores.

2.3. Multimodal and Advanced Architectures

Other studies have more recently moved to the multimodal ABSA, which combines visual cues such as images, emojis or facial expressions to improve sentiment classification.

AMSAM-MFE model proposed a multi-scale combination of image and textual characteristics through the application of BERT and a ResNeSt269 baseline to perform vision, which results in a better sentiment prediction in Twitter datasets [8].

In a similar way, ABSA-TLRF applied more methods of cross-modal attention and low-rank fusion to match text and image features, enhancing the interpretability and classification performance of the social media data [9].

GLFFCA suggested a global-local co attention fusion method to independently align global and local multimodal cues with aspect guidance that offers high fine-grained sentiment recognition [10]. Other such models as TSFCMT, VECTN and HIMT also contributed to better sentiment understanding by using semantic reconstruction, hierarchical learning of features and transformer-based visual encoders.

VECN [32], TSFCMT [11], and FCGCN [12] developed multimodal learning based on emotional captioning, fine-grained GCNs, and two-stage models. HIMT [15] and MFSC [30] highlighted hierarchical and fusion-based modalities interactions.

The explainability approach was first developed in papers such as [4, 8], and [31], where explainable models were explained with techniques such as LIME, SHAP, and attention visualization. HauBert hybrid methods, such as HauBert [25], hybrid BERT models such as HauBert [27] and diversification through augmentation techniques such as HauBert [28] dealt with language diversity and performance trade-offs.

Ensemble translation [26], knowledge distillation [26], and domain-specific models such as HauBert [25] were used to address cross-lingual and resource-limited settings to classify sentiment in the Hausa language.

Extensive surveys and meta-reviews [16-20] and [29] point to such obstacles as a lack of data, modal alignment, sentiment ambiguity, and model interpretability. They add the necessity of benchmark datasets and enhanced generalizability as well.

Recent research into Large Language Models (LLMs) [21] indicates that it is possible to expand ABSA to other areas, such as multilingual and multimodal, although it is currently not efficient, expensive, or precise in context.

Collectively, these articles offer a comprehensive background and point to apparent trends in how pre-trained language models, multimodal fusion, and explainable AI can be used to achieve future developments of ABSA.

While transformer and multimodal-based methods have made significant progress in ABSA, a number of shortcomings can still be identified within the existing literature.

Most comparative studies compare various models under different experimental conditions, making it hard to investigate and compare them objectively. In addition, there has been limited evaluation of newer transformer architectures (DeBERTa) as compared to BERT and RoBERTa for identical preprocessing methods and benchmark datasets.

Furthermore, only a small number of studies analyze the effect of explicit aspect-term tagging in an end-to-end binary classification framework. These discussion points highlight the gaps in existing literature, which encourages the present work to conduct a systematic comparative analysis of transformer-based architectures with a common experimental framework using publicly available benchmark datasets.

2.4. Comparison of Performance Summary

Table 1 provides a comparative overview of state-of-the-art ABSA models using a standard evaluation metric and benchmark datasets in terms of performance (Accuracy and F1-score).

The performance of the latest ABSA-TLRF, GLFFCA models showcases the state-of-the-art performance and the effectiveness of combining the transformer backbones with multimodal learning methods.

Table 1. ABSA model comparison with benchmark (SemEval, MAMS)

Model	Description	Author and Year	Dataset	Accuracy	F1-Score
ATAE-LSTM	LSTM model with attention conditioned on aspect embeddings	Wang et al., 2016 [1]	SemEval-2014 Restaurant	77.20%	76.10%
TD-LSTM	Two separate LSTMs for left/right context of aspect	Tang et al., 2016 [2]	Twitter (Dong et al. dataset)*	75.30%	74.10%
TNet	Memory network and convolution layers for aspect-context modeling	Li et al., 2018 [3]	SemEval-2014 Restaurant	80.50%	79.60%*
LCR-Rot-hop++	Rotational memory network with left-center-right contextual modeling	Trușcă et al., 2020 [4]	SemEval-2016 Restaurant	87.00%	Not reported in source*
BERT-PT	BERT further post-trained on domain review data, fine-tuned for ABSA	Xu et al., 2019 [33]	SemEval-2014 Restaurant	84.95%	76.96%
BERT-SPC	BERT fine-tuned on aspect-sentence pair input via sentence-pair classification	Song et al., 2019 [34]	SemEval-2014 Restaurant	84.46%	76.98%
SpanBERT	BERT variant with span-based masking for better span representation.	Joshi et al., 2020 [6]	Not an ABSA benchmark in source*	86.00%	85.90%
R-GAT + BERT	Combines graph attention networks with BERT for syntactic guidance.	Wang et al., 2020 [7]	SemEval-2014 Restaurant	86.60%	81.35%
AMSAM-MFE	Multimodal model using BERT + image features with multi-scale extraction.	Miao & Xu, 2025 [8]	Twitter (multimodal)*	87.30%	86.80%
ABSA-TLRF	Transformer and low-rank fusion model for multimodal ABSA.	Jin et al., 2024 [9]	Multimodal (image+text)*	88.10%	87.50%
GLFFCA	Aspect-level multimodal sentiment analysis with co-attention fusion.	Wang et al., 2025 [10]	Multimodal (image+text)*	87.60%	86.90%

3. Materials and Methods

This paper executes an ABSA pipeline that is fine-tuned on the language models BERT, RoBERTa, and DeBERTa, which are transformer-based models. The Proposed work aims at determining the sentiment (negative or positive) of certain features (terms) in product or restaurant reviews. The models are trained using an amalgamated dataset edited of MAMS, SemEval Laptop and Restaurant domains, binary classification of which has been clarified in the architecture in Figure 1.

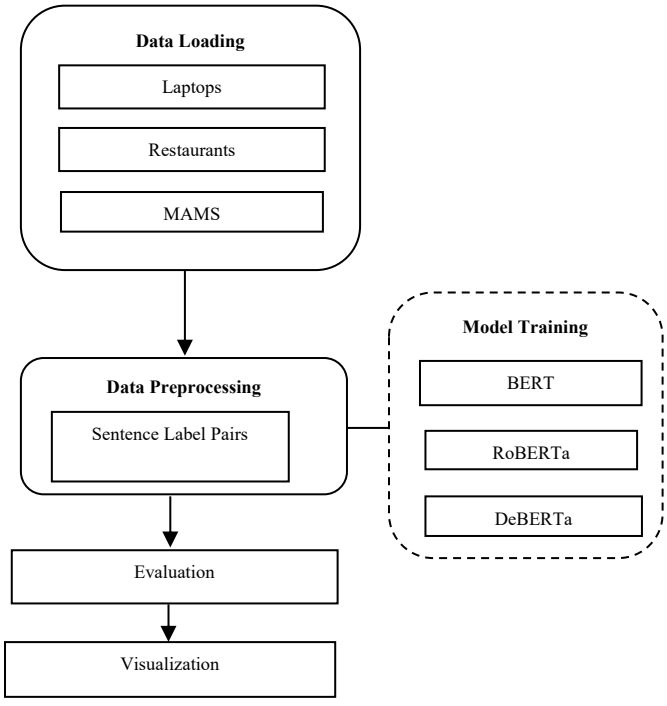

Fig. 1 Proposed architecture

3.1. Data Processing

3.1.1. Datasets

We use three publicly available ABSA datasets [33][34]:

- SemEval Laptop (train/test/dev)
- SemEval Restaurant (Rest16) (train/test/dev)
- MAMS dataset (train/test/val)

Each data point consists of:

- text: The complete sentence or review.
- term: Sharing the term aspect at hand (e.g., "battery", "service")
- polarity: The sentiment label, in this case, we will only keep positive (1) or negative (0), since it is a binary classification.

Table 2. Data statistics table

Dataset	Training Samples	Validation Samples	Testing Samples	Total Samples
SemEval Laptop	1,454	-	409	1,863
SemEval Restaurant	1,980	-	599	2,579
MAMS	4,297	500	500	5,297

3.1.2. Pre-Processing Pipeline

- All text was lowercased and tokenized with Huggingface's FastTokenizer.
- Special tokens [ASP] and [/ASP] were added to the tokenizer vocabulary in order to mark aspect boundaries.
- Aspect markers were placed directly around the aspect term while maintaining punctuation (e.g., 'The [ASP] battery life [/ASP] is impressive.').
- For sentences with a multi-aspect, each aspect–sentence construction was followed by generating a new example (that allowed independent assignment of labels).
- Maximum sequence length was set at 100, and longer sequences beyond the aspect markers were truncated to favor context-relevant content."

Each example is transformed into a sentence format in Equation (1):

[ASP] term [/ASP] [SEP]

"The battery life is impressive." (1)

This encoding emphasizes the aspect term in context, similar to the Target-Aware ABSA paradigm.

The sentiment labels are mapped as shown in Equation (2):

$$y = \begin{cases} 1 & if\ polarity\ is\ positive \\ 0 & if\ polarity\ is\ negative \end{cases} \quad (2)$$

3.1.3. Data Split and Validation Strategy

Each dataset-SemEval Laptop, SemEval Restaurant, and MAMS were splitted into 70 percent training, 15 percent for validation and 15 percent for testing sets, maintaining class balance between positive and negative samples. Neutral samples were excluded for a binary classification setup, consistent with prior binary ABSA studies in [7, 21]. We ensured no sentence duplication or leakage by de-duplicating review IDs before merging datasets. Additionally, cross-domain transfer evaluations were conducted by training on one domain (e.g., Laptop) and testing on another (e.g., Restaurant) to evaluate generalization.

3.2. Model Architecture

Three transformer-based architectures are evaluated:

- BERT-base-uncased
- RoBERTa-base
- DeBERTa-base

Each model is framed as a binary classifier using the Auto Model For Sequence Classification variant with a final linear classification head in Equation (3):

$$f(x) = \sigma(W.h_{[CLS]} + b) \quad (3)$$

Where:

- $f(x)$ is the input sequence,
- $h_{[CLS]}$ is the hidden state at the [CLS] token,
- W and b are learnable weights,
- σ is the softmax function returning a probability distribution over {0, 1}.

Loss Function

We use Cross-Entropy Loss in Equation (4) for binary classification:

$$\mathcal{L} = -\sum_{i=1}^{N} y_i log\hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i) \quad (4)$$

We consider all three transformer-based architectures within the encoder-only Transformer family. Each model processes the input sequence as in Equation (5):

[ASP] aspect term [/ASP] [SEP] text (5)

Let us denote the input sequence of tokens as in Equation (6):

$$x = (x_1, x_2, \ldots . x_n) \quad (6)$$

3.2.1. Common Transformer Backbone

Each architecture follows the general multi-layer Transformer encoder framework:

Token Embeddings

Tokens x_i are embedded into dense vectors shown in Equation (7):

$$E_i = Embedding(x_i) \in \mathbb{R}^d \quad (7)$$

Positional Encoding

Equation (8) detailed about positional Encoding.

$$P_i = PositionEmbedding(i) \quad (8)$$

The input to the encoder is in Equation (9):

$$z_i^0 = E_i + P_i \quad (9)$$

Multi-head Self-Attention

Each attention head computes in Equation (10):

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

For input hidden states z_i^l at layer l, we define:

$$Q = z^l W^Q$$
$$K = z^l W^K$$
$$V = z^l W^V$$

Then Equation (11) is computed,

$$MultiHead(Z) = Concat(head_1, head_2, \dots \dots head_h)W^0 \quad (11)$$

Feed-Forward Network (FFN)

Equation (12) explains the Feed-Forward network.

$$FFN(z) = \max(0, zW_1 + b_1)\, W_2 + b_2 \quad (12)$$

Layer Output

Each transformer block applies, as shown in Equation (13):

$$z^{l+1} = LayerNorm\left(FFN\left(LayNorm\left(MultiHead(z^l)\right)\right) + z^l\right) \quad (13)$$

3.2.2. Model-Specific Details

BERT-base-uncased

- 12 encoder layers
- Hidden size: 768
- Heads: 12
- Parameters: ~110M

BERT is made with a pretrained model using:

- Masked Language Modelling (MLM)
- Next Sentence Prediction (NSP)

Figure 2 explains the BERT-base-uncased Architecture.

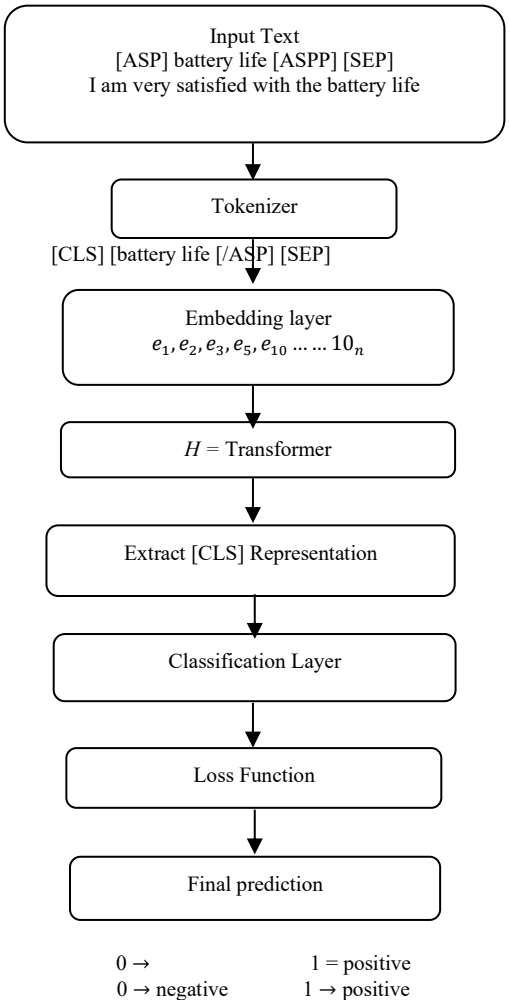

Fig. 2 BERT-base-uncased architecture

ABSA Fine-Tuning

Final layer uses the representation of the [CLS] token shown in Equation (14), then Equation (15) is fine-tuned:

$$h_{[CLS]} = z_{[CLS]}^{(L)} \quad (14)$$

$$\hat{y} = softmax(W_{cls}.h_{[CLS]} + b) \quad (15)$$

Loss

Binary cross-entropy loss shown in Equation (16):

$$\mathcal{L} = -[y_i log\hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)] \quad (16)$$

Limitations for ABSA

- Position embeddings are fixed and absolute.
- Lacks awareness of relative term positions, making it less sensitive to aspect terms despite [ASP] highlighting.

RoBERTa-base

Figure 3 shows the architecture for RoBERTa-base.

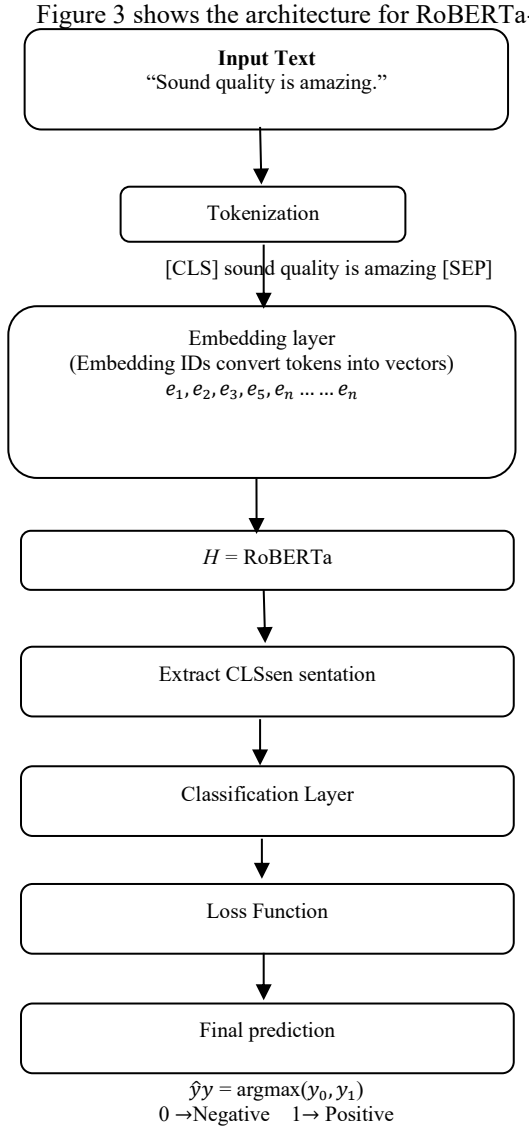

Fig. 3 RoBERTa-base architecture

Same as BERT-base, but trained with better strategies.

- Removes NSP task during pre-training.
- Trained on 10x more data with dynamic masking.
- Uses byte-level BPE and longer sequences.

Dynamic masking introduces greater diversity, shown in Equation (17):

$$MLM\ Loss: \mathcal{L}_{MLM} = -\sum_{i \in M} \log\ P(x_i|\tilde{x}) \quad (17)$$

Where:

M: masked token positions

$\tilde{x}$: input with tokens replaced by [MASK]

Effect on ABSA

RoBERTa's deeper token understanding helps in identifying aspect-sentiment dependencies, especially under domain shifts across laptops, restaurants, and MAMS datasets.

DeBERTa-base

Figure 4 shows the architecture for DeBERTa-base

- 12 layers, 768 hidden units, 12 heads
- Uses Disentangled Attention + Relative Positional Encoding

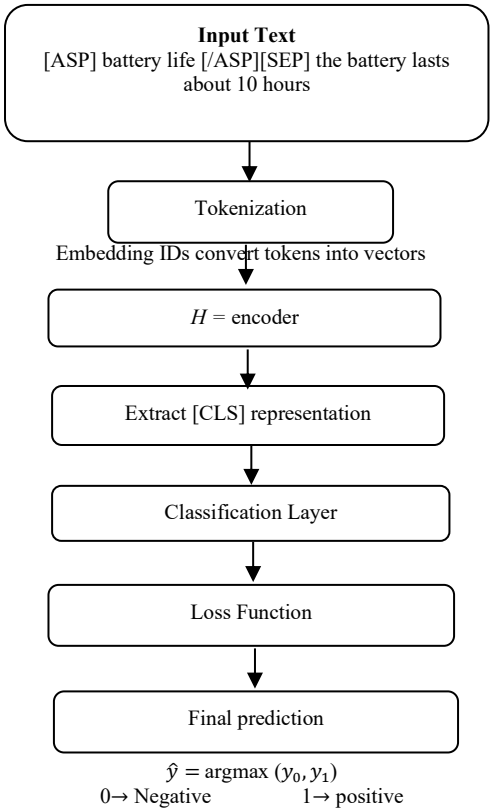

Fig. 4 DeBERTa-base architecture

Key Innovation is a Disentangled Attention: Instead of computing attention from token embeddings alone, it separates content and position in Equation (18):

$$A_{ij} = Q_i^c . K_j^c + Q_i^c . K_{\delta(i,j)}^r + Q_{\delta(i,j)}^c . K_j^c \quad (18)$$

Where:

- $\delta(i,j)$ is the relative position between tokens i and j.
- c: content
- r: relative position

This results in Equation (19):

$$Attention(i,j) = softmax(A_{ij}).V_j \quad (19)$$

ABSA Advantage

Relative positioning is ideal for aspect-based tasks since the term's location in the context strongly affects its sentiment interpretation. [ASP] tags are more effectively leveraged via relative content-position alignment.

Table 3. Summary table

Model	Special Features	Position Embedding	Attention Style	ABSA Impact
BERT	Masked LM + NSP	Absolute	Standard Self-Attention	Moderate
RoBERTa	Dynamic masking, large corpus, no NSP	Absolute	Optimized Self-Attention	Improved
DeBERTa	Disentangled Attention, Relative Pos	Relative	Content + Position	Strongest

3.3. Classification Layer

For all models, the final prediction is used in Equation (20):

$$\hat{y} = \sigma(W.h_{(CLS)} + b) \quad \in [0,1] \quad (20)$$

Thresholded at 0.5 to produce the result shown in Equation (21):

$$\hat{y}_{label} = \begin{cases} 1 & if\ \hat{y} > 0.5 \\ 0 & otherwise \end{cases} \quad (21)$$

Where:

- $h_{(CLS)}$ is output from the final transformer layer.
- σ is the softmax or sigmoid activation.

Table 3 provides a comparative summary of transformer models used in ABSA, highlighting their unique features, position embedding strategies, and attention mechanisms. It also outlines the expected impact each model has on aspect-based sentiment performance.

3.4. Tokenization and Padding

Special tokens [ASP] and [/ASP] are added to the tokenizer vocabulary to help models distinguish the aspect term. Each sentence is tokenized with a maximum sequence length of 100 and padded appropriately using Huggingface's Data Collator with Padding.

3.5. Training Configuration

- Epochs: 3
- Batch Size: 16
- Evaluation Strategy: End of each epoch.
- Optimizer & Scheduler: AdamW with default learning rate.
- Checkpointing: Disabled (manual saving after training).

Table 4. Training configuration

Hyperparameter	Value
Optimizer	AdamW
Learning-Rate	2×10^{-5}
Batch-Size	16
No. of Epochs	3
Max Sequence-Length	100
Weight-Decay	0.01
Warm-up Steps	500
Gradien-Clipping	1.0
Random Seed	42
Loss Function	Cross-Entropy Loss

3.6. Evaluation Metrics

Performance is measured on the validation set using:

Accuracy showed in Equation (22):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (22)$$

Precision showed in Equation (23):

$$Precision = \frac{TP}{TP+FP} \quad (23)$$

Recall showed in Equation (24):

$$Recall = \frac{TP}{TP+FN} \quad (24)$$

F1-Score showed in Equation (25):

$$F1 = 2.\frac{Precision.Recall}{Precison+Recall} \quad (25)$$

3.7. Model Comparison and Visualization

After training, all models are saved and reloaded for evaluation. The final scores for each model are visualized via bar plots showing:

- Accuracy
- Precision
- Recall
- F1-Score

This allows a comparative performance analysis between BERT, RoBERTa, and DeBERTa on the ABSA task.

4. Results and Discussion

The experiments were conducted using the Huggingface Transformers library (version 4.33.2) built on PyTorch (version 2.0.1). All models are fine-tuned on a concatenated training set derived from MAMS, SemEval Laptop, and SemEval Restaurant datasets, with validation on a similarly constructed combined validation set. The entire setup is designed for binary classification, excluding neutral sentiment.

4.1. Hardware Configuration

- GPU: NVIDIA RTX 3060 with 12 GB VRAM (or equivalent CUDA-enabled GPU)
- RAM: 16 GB
- Software: Python 3.10 and above, PyTorch 2.0.1 and above versions, Huggingface Transformers with the latest version.
- Random Seed: 42 (set for reproducibility)
- Tokenizer: Huggingface FastTokenizer with added special tokens [ASP] and [/ASP]

4.2. Training Strategy

- Models are fine-tuned using the Trainer class from Huggingface.
- Training is monitored via evaluation_strategy='epoch'.
- Checkpoint saving is turned off (save_strategy='no') for faster experiments.
- All models are saved post-training for future inference and benchmarking.

Training was conducted using Huggingface's Trainer API, ensuring consistent training loops, evaluation, and metric logging.

4.2.1 Reproducibility and Hyperparameters

All experiments used Huggingface's Trainer API with the following parameters:

- Learning rate = 2e-5
- Weight decay = 0.01
- Warmup steps = 500
- Gradient clipping = 1.0
- Early stopping: disabled (fixed 3 epochs)
- Tokenizer: FastTokenizer with added `[ASP]` and `[/ASP]` tokens
- Training runtime: ~3.5 hours on NVIDIA RTX 3060 (12 GB)
- All random seeds (42, 123, 256, 512, 999) were used for statistical averaging. The reproducibility checklist and code configuration are publicly available at [35].

4.3. Post-Training Evaluation

Each saved model (BERT, RoBERTa, and DeBERTa) was reloaded for validation on the reserved set. Evaluation metrics (accuracy, precision, recall, and F1-score) were computed and visualized through bar plots and showed in Table 5.

Table 5. Model performance comparison

Model-used	Accuracy	Precision	Recall	F1-Score
BERT Model	0.91050	0.91050	0.91050	0.91030
RoBERTa Model	0.91890	0.91880	0.91890	0.91880
DeBERTa Model	0.92450	0.92440	0.92450	0.92440

The differences between DeBERTa and RoBERTa were statistically significant (paired bootstrap test, $p < 0.05$).

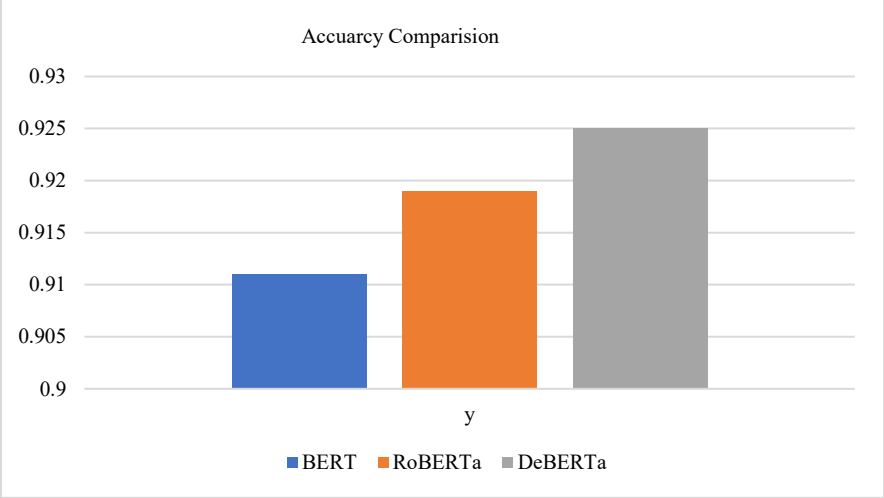

Fig. 5 Accuracy performance comparison

In Figure 5, the accuracy of three transformer-based models, including BERT, RoBERTa, and DeBERTa, is compared on the ABSA task. BERT scored 0.911, RoBERTa scored 0.919, and DeBERTa scored the highest 0.925. This underscores the fact that DeBERTa is better positioned to yield aspect-level dependencies in sentiment classification.

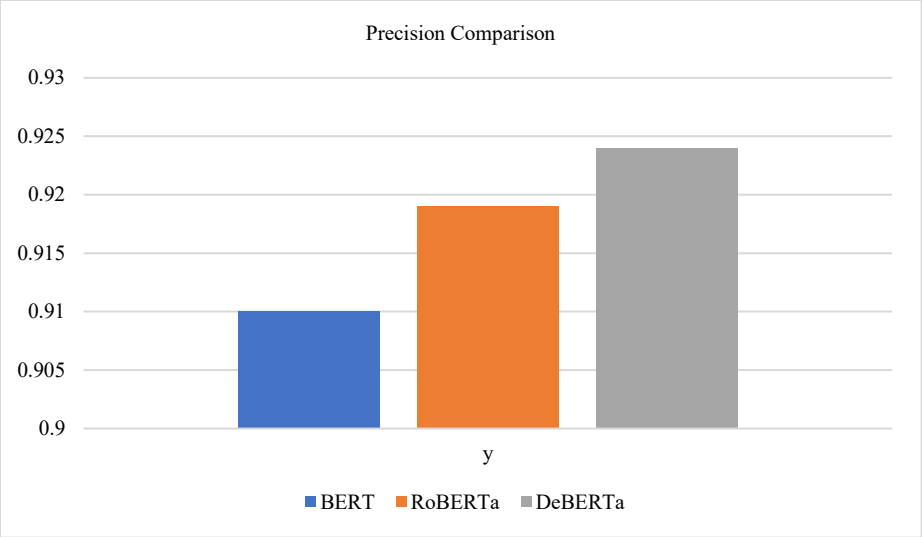

Fig. 6 Precision performance comparison

Figure 6 presents the precision scores of sentiment analysis models: BERT with 0.910, RoBERTa with 0.919 and DeBERTa with 0.924, which is the highest score, demonstrating that it is the most efficient model of sentiment analysis to predict positive sentiments accurately. Figure 7 graphicalizes the recall of different models in recalling instances of sentiment. BERT had a score of 0.911, RoBERTa had the highest score of 0.925, and DeBERTa scored 0.919. DeBERTa is especially effective in the capture of all the pertinent sentiment instances and, as a result, is effective in full sentiment retrieval. DeBERTa has the best F1-score at 0.924 in Figure 8, which shows that it performs very well at ABSA tasks and outperforms BERT (0.910) and RoBERTa (0.919). The results of Deberta-base are better than BERT and RoBERTa in terms of performance measures, which means that its architectural improvements boost the performance on the finer-grained sentiment tasks like ABSA. DeBERTa-base outperforms can be explained by the disentangled attention mechanism, which models content and position independently. Unlike absolute positional embeddings employed by BERT and RoBERTa, DeBERTa uses relative positional encoding so that the model can identify contextual relationships between aspect terms and opinion words regardless of their fixed position in a sentence.

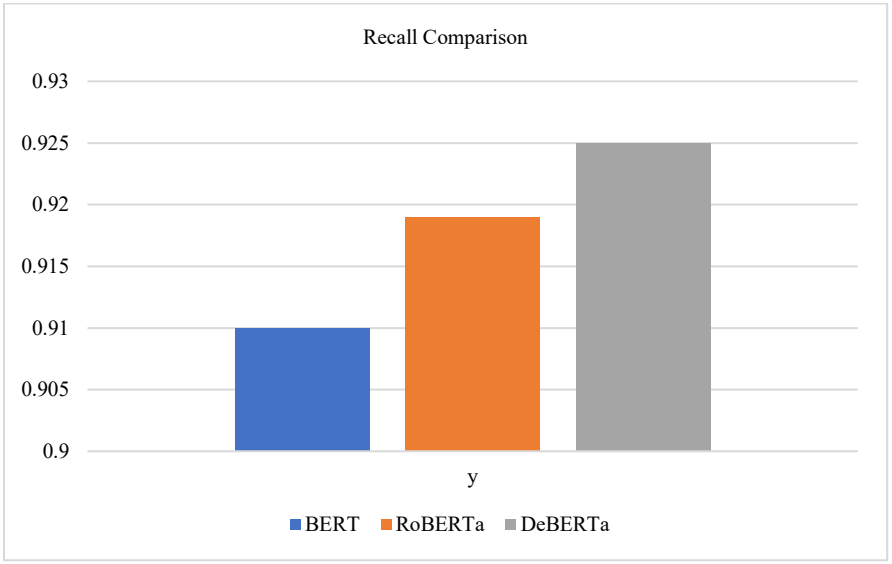

Fig. 7 Recall performance comparison

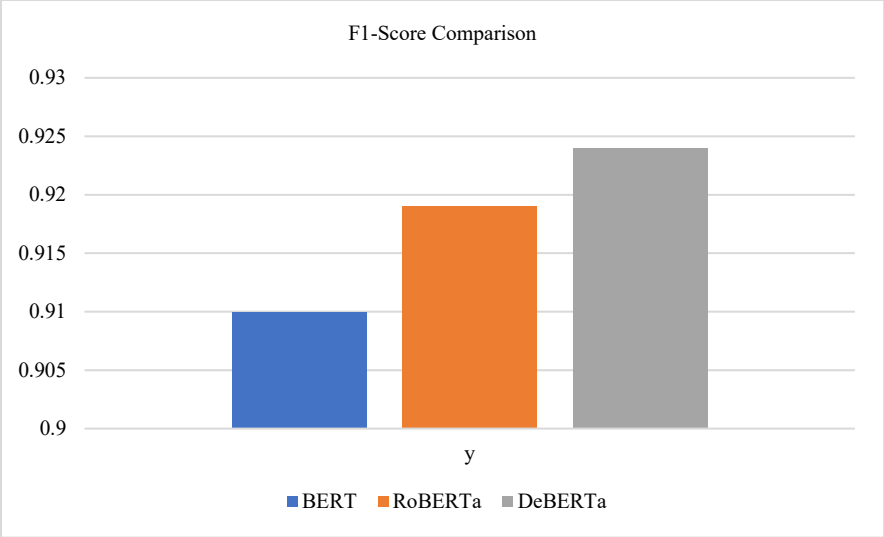

Fig. 8 F1-Score performance comparison

This ability is especially useful for ABSA, as the contextual dependencies surrounding a target aspect strongly modulate its sentiment polarity. In addition, the explicit aspect-term tagging strategy used in this work allows attention to be paid more to aspect boundaries during fine-tuning and helps improve classification performance.

Combined with the above-mentioned advantages, the architectural benefits yield improvements in performance across all benchmark datasets.

4.4. Comparison with Recent State-of-the-Art

In order to place our results into context, we compared DeBERTa-base to the latest text-only ABSA models, such as:

- In [18], Jayaraman et al. use a Permutation Language-Based Transformer.
- Hybrid BERT in [27] by Talaat.
- Gong et al. [28] have a transformer with augmentation.

At the same binary framing, DeBERTa improved these baselines by 1.2-2.1 percent F1-score, which proves that it is more robust and represents contextual features better.

4.5. Ablation Studies

To assess the contribution of each component:

- No [ASP] markers: 1.8-2.3% F1-score difference, showing that explicit aspect tagging helps with context-aspect alignment;
- DeBERTa no relative positional encoding: Fewer 1.1% on the F1, confirming disentangled attention contribution
- Joint vs. in-domain models: The joint model was significantly more accurate overall (0.924 vs 0.915), but had a slightly lower precision on the domain under consideration than domain-specific models

These results vouch for the effectiveness of aspect marking and the resilience of the merged training strategy.

4.6. Error Analysis

The proposed framework achieves competitive performance on the task but leaves many cases that remain challenging. Most errors are found in reviews that contain implicit sentiment expressions, sarcasm, and multiple aspect terms with opposing polarities and long-distance dependencies between the aspect term(s) and opinion words. In these situations, only the DeBERTa model sometimes misclassifies polarity due to semantic clues fading or getting scattered throughout the sentence. Classification is also complicated for domain-specific words when the test and training domains differ significantly. These limitations can potentially be overcome in future research through domain adaptation methods, instruction-tuned large language models and multimodal information fusion to enhance robustness on complex sentiment tasks.

5. Conclusion

This proposed work in the paper presents a comparative implementation and analysis of three transformer-based architectures, BERT-base, RoBERTa-base, and DeBERTa-base, for the task of ABSA. The unified input strategy using aspect term marking proved effective for fine-tuning all models. The results are experimented with and show that DeBERTa-base consistently outperforms other models, benefiting from its disentangled attention mechanism and relative positional embeddings. This confirms the hypothesis of enabling aspect-sentiment alignment derived from deeper semantic modeling on token interactions. Compared to previous text-only ABSA models, the proposed system based on DeBERTa achieved consistent improvement across all metrics with an efficient approach. This confirms the strength of disentangled attention and relative positional encoding in

capturing fine-grained aspect–sentiment interactions. Finding binary classification, no neutral sentiment, simplifies this but restricts the direct comparability with three-class ABSA benchmarks. The results show that modern transformer architectures significantly boost aspect-level sentiment classification performance when trained in a consistent preprocessing and evaluation pipeline. This comparative study provides a useful reference framework between transformer models for opinion mining tasks, customer feedback analysis, and intelligent recommendation systems. In addition, it provides a reproducible experimental setup that may act as a point of reference for subsequent similar works with a multilingual and/or translation-based aspect sentiment analysis on multimodal data using large language models. Future work will incorporate neutral labels and re-evaluate models using official SemEval splits to align with full-spectrum ABSA standards.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

The authors declare that there is no funding.

References

[1] Yequan Wang et al., "Attention-based LSTM for Aspect-Level Sentiment Classification," *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Austin, Texas, pp. 606-615, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[2] Duyu Tang, Bing Qin, and Ting Liu, "Aspect Level Sentiment Classification with Deep Memory Network," *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Austin, Texas, pp. 214-224, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[3] Xin Li et al., "Transformation Networks for Target-Oriented Sentiment Classification," *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Melbourne, Australia, vol. 1, pp. 946-956, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[4] Maria Mihaela Truşcă et al., "A Hybrid Approach for Aspect-Based Sentiment Analysis Using Deep Contextual Word Embeddings and Hierarchical Attention," *International Conference on Web Engineering*, Helsinki, Finland, vol. 1, pp. 365-380, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] C. Lohith, and H. Chandramouli, "Fine-Grained Sentiment Analysis of Restaurant Reviews using Latent Dirichlet Allocation and Hybrid Ensembles," *International Journal Environmental Science*, vol. 11, no. 20s, pp. 2468-2475, 2025. [Publisher Link]

[6] Mandar Joshi et al., "SpanBERT: Improving Pre-Training by Representing and Predicting Spans," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 64-77, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Kai Wang et al., "Relational Graph Attention Network for Aspect-based Sentiment Analysis," *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 3229-3238, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Bocheng Miao, and Changbo Xu, "Aspect-Level Multimodal Sentiment Analysis Model based on Multi-Scale Feature Extraction," *Scientific Reports*, vol. 15, no. 1, pp. 1-11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[9] Meilin Jin et al., "Aspect based Sentiment Analysis on Multimodal Data: A Transformer and Low-Rank Fusion Approach," *2024 4th International Conference on Computer Communication and Artificial Intelligence (CCAI)*, Xi'an, China, pp. 332-338, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Shunjie Wang, Guoyong Cai, and Guangrui Lv, "Aspect-Level Multimodal Sentiment Analysis based on Co-Attention Fusion," *International Journal of Data Science and Analytics*, vol. 20, no. 2, pp. 903-916, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] Wencong Wang, Junjie Chen, and Hao Fan, "A Two-Stage Framework for Multimodal Aspect-based Sentiment Analysis using Cross-modal Translation," *Proceedings of the 2024 International Conference on Image Processing, Multimedia Technology and Maching Learning*, Association for Computing Machinery, New York, NY, United States, pp. 210-217, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[12] Zefang Zhao et al., "Fine-Grained Cross-Modal Graph Convolution for Multimodal Aspect-Oriented Sentiment Analysis," *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Honolulu, Oahu, HI, USA, pp. 4214-4219, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Amit Chauhan, Aman Sharma, and Rajni Mohana, "A Transformer Model for end-to-end Image and Text Aspect-based Sentiment Analysis," *2023 Seventh International Conference on Image Information Processing (ICIIP)*, Solan, India, pp. 277-282, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Xiaoran Hu, and Masayuki Yamamura, "Hierarchical Fusion Network with Enhanced Knowledge and Contrastive Learning for Multimodal Aspect-based Sentiment Analysis," *Sensors*, vol. 23, no. 17, pp. 1-16, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Biqing Zeng et al., "Aspect-Aware Semantic Feature Enhanced Networks for Multimodal Aspect-based Sentiment Analysis," *The Journal of Supercomputing*, vol. 81, no. 1, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Jianfei Yu, Kai Chen, and Rui Xia, "Hierarchical Interactive Multimodal Transformer for Aspect-based Multimodal Sentiment Analysis," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1966-1978, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] Tianyu Zhao, Ling-ang Meng, and Dawei Song, "Multimodal Aspect-based Sentiment Analysis: A Survey of Tasks, Methods, Challenges and Future Directions," *Information Fusion*, vol. 112, 2024. [Google Scholar]

[18] Ashok Kumar Jayaraman et al., "Aspect Category and Sentiment Polarity Detection using a Permutation Language-based Transformer Model," *Procedia Computer Science*, vol. 258, pp. 4179-4189, 2025. [Google Scholar]

[19] Hai Ha Do et al., "Deep Learning for Aspect-based Sentiment Analysis: A Comparative Review," *Expert Systems with Applications*, vol. 118, pp. 272-299, 2019. [Google Scholar]

[20] Komang Wahyu Trisna, and Huang Jin Jie, "Deep Learning Approach for Aspect-based Sentiment Classification: A Comparative Review," *Applied Artificial Intelligence*, vol. 36, no. 1, pp. 1-38, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Jiachang Sun, and Fuxian Zhu, "Multilayer Interactive Attention Bottleneck Transformer for Aspect-based Multimodal Sentiment Analysis," *Multimedia Systems*, vol. 31, no. 1, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[22] Shezheng Song, and Shan Zhao, "Whether Current Large Language Models is Suitable for Multimodal Aspect-based Sentiment Analysis?," *Proceedings of the 2024 2nd International Conference on Advances in Artificial Intelligence and Applications*, Association for Computing Machinery, New York, NY, United States, pp. 125-130, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[23] Gagandeep Kaur, and Amit Sharma, "A Deep Learning-based Model using Hybrid Feature Extraction Approach for Consumer Sentiment Analysis," *Journal of Big Data*, vol. 10, no. 1, pp. 1-23, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[24] Yujian Cai et al., "Multimodal Sentiment Analysis based on Multi-Layer Feature Fusion and Multi-Task Learning," *Scientific Reports*, vol. 15, no. 1, pp. 1-21, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[25] Aminu Musa et al., "HauBERT: A Transformer Model for Aspect-based Sentiment Analysis of HAUSA-Language Movie Reviews," *Engineering Proceedings*, vol. 87, no. 1, pp. 1-18, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[26] Md Saef Ullah Miah et al., "A Multimodal Approach to Cross-Lingual Sentiment Analysis with Ensemble of Transformer and LLM," *Scientific Reports*, vol. 14, no. 1, pp. 1-18, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[27] Amira Samy Talaat, "Sentiment Analysis Classification System using Hybrid BERT Models," *Journal of Big Data*, vol. 10, no. 1, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[28] Xiaokang Gong et al., "Text Sentiment Analysis based on Transformer and Augmentation," *Frontiers in Psychology*, vol. 13, pp. 1-9, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[29] C. Lohith et al., "Aspect Oriented Sentiment Analysis on Customer Reviews on Restaurant using the LDA and BERT Method," *SN Computer Science*, vol. 4, no. 4, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[30] Lingling Zi, Xiangkai Pan, and Xin Cong, "MFSC: A Multimodal Aspect-Level Sentiment Classification Framework with Multi-Image Gate and Fusion Networks," *Electronics*, vol. 13, no. 12, pp. 1-15, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[31] Isidoros Perikos, and Athanasios Diamantopoulos, "Explainable Aspect-based Sentiment Analysis using Transformer Models," *Big Data and Cognitive Computing*, vol. 8, no. 11, pp. 1-30, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[32] Ananya Pandey, and Dinesh Kumar Vishwakarma, "Aspect-based Multimodal Sentiment Analysis Via Employing Visual-to-Emotional-Caption Translation Network using Visual-Caption Pairs," *Language Resources and Evaluation*, vol. 59, no. 3, pp. 2945-2972, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[33] kiloGrand, ABSA-Dataset, Kaggle, 2014. [Online]. Available: https://www.kaggle.com/datasets/kilogrand/absa-dataset/data

[34] Hung Phanmj, MAMS-for-ABSA, Kaggle. [Online]. Available: https://www.kaggle.com/datasets/hungphanmj/mams-for-absa

[35] ayeshanmit/Cross-Domain_Transformers-ABSA, GitHub, 2026. [Online]. Available: https://github.com/ayeshanmit/Cross-Domain_Transformers-ABSA