



Atlas Journal of Computer Science

Journal homepage: www.ajocs.atlasci.org

Volume 1, Issue 1 (2026) 1–12

ISSN: 3119-6810



Original Article

Hierarchical Contextual Attention Network with Dynamic Sentiment Routing for Aspect-Based Sentiment Analysis

Otmane Houdaif^{1*}, Meryem Tamir², Elarbi Lamfarrad¹, Amine Bouobida¹

¹ Department of Computer Sciences, Faculty of Sciences, Chouaib Doukkali University, El Jadida 24000, Morocco

² Green Tech Institute (GTI), Mohammed VI Polytechnic University (UM6P), Ben Guerir 43150, Morocco

ARTICLE INFO

Article history

- Received: 05 February 2025
- Revised: 01 March 2025
- Accepted: 22 March 2025
- Online : 22 March 2025

Keywords

- Aspect-based sentiment analysis
- Hierarchical attention
- Capsule network
- Dynamic routing ;
- BERT
- Cross-aspect interaction;
- SemEval

ABSTRACT

Aspect-based Sentiment Analysis (ABSA) attempts to detect the sentiment polarity associated with particular aspects and to predict such polarities accurately. Although pre-trained language models like BERT have greatly improved the performance of ABSA, they still suffer from the problems of discriminating relevant context of aspects from noise, capturing dependencies among multiple aspects and aggregating sentiment evidence effectively, especially for the small neutral class. We propose a Hierarchical Contextual Attention Network with Dynamic Sentiment Routing (HCAN-DSR) to solve these problems. Our model leverages a multi-level hierarchical attention module on top of a frozen BERT encoder to extract local and global contextual information with focusing aspect-related features. The cross-aspect graph with syntax guidance can model semantic and syntactic dependencies among multiple aspects for enriching interaction among them. Moreover, our Dynamic Sentiment Routing (DSR) mechanism is used as an alternative to traditional pooling to iteratively aggregate aspect-specific evidence, thus preserving fine-grained sentiment information and improving the neutral sentiment classification. Experiment results on Laptop14, Restaurant14 and Twitter benchmark datasets show that HCAN-DSR performs better than all state-of-the-art graph and attention based ABSA models with accuracy of 81.38% on Laptop14 and 89.14% on Restaurant14.

1. Introduction

Aspect-Based Sentiment Analysis (ABSA) has become one of the key tasks in natural language processing and computational opinion mining. ABSA finds use in various applications, including analyzing customer

* Corresponding author.

E-mail address: houdaif.o646@ucd.ac.ma

reviews [1], social media monitoring, product intelligence, and aggregating feedback from the medicine field. In contrast to sentiment classification at the level of document or sentence, aspect-based sentiment classification requires a two-step procedure: identifying the aspect term (e.g., 'food quality', 'battery life', 'waiting time') and then determining the polarity of the sentiment towards that particular aspect, positive, neutral, or negative [2].

Development of pre-trained transformer models, particularly BERT [3] and its variations RoBERTa [4] and ALBERT [5], has significantly boosted the development of ABSA task. The aforementioned models provide deep bidirectional contextual representations, which account for long-distance syntactic and semantic dependencies and are crucial for correct understanding of the sentiment in multi-aspect sentences like 'The food was excellent, but the service was appallingly slow.' However, there are three structural shortcomings of current BERT-based ABSA models that form the basis of this research.

First, standard fine-tuned BERT models treat all tokens in the sentence with approximately equal weight when constructing the representation used for sentiment prediction. This causes the model to conflate sentiment signals from different aspects, for example, in the sentence above, the word 'slow' contributes negatively to the overall [CLS] representation even when predicting the sentiment toward 'food'. Several attention-based models attempt to address this by computing aspect-specific attention over the hidden states [6, 7], but these approaches typically operate at a single level of granularity (token-level), missing higher-order contextual patterns that span multiple tokens or clauses.

Second, a sentence often consists of more than one aspect whose sentiment is connected either syntactically or semantically. For instance, in "The ambiance is romantic but overpriced for what it offers," sentiments toward 'ambiance' and 'price' are contrastively placed by using the conjunction "but." Graph-convolutional methods [8, 9] have utilized dependency parse trees to capture such inter-aspect dependency. However, most of the current formulations consider edges as binary variables (either present or absent) without taking into account their importance to the targeted aspect.

Third, the neutral sentiment category is underrepresented in all standard ABSA datasets (usually 8% to 18% of total instances in SemEval datasets), and softmax-based classifiers are ill-calibrated to this category. Hence, the models usually achieve high accuracy for positive and negative instances but show relatively poor accuracy for neutral instances, resulting in a decreased Macro-F1 score since Macro-F1 gives equal importance to all categories.

We propose HCAN-DSR to solve the above-mentioned problems within a single framework. Our main contributions are:

- Hierarchical attention stack for fine-grained and contextual attention at the token level with gated sentence-level attention to capture aspect-specific features from BERT embeddings.
- Cross-aspect dependency interaction (CAI) module that builds soft-weighted dependency graph based on a syntactic parse tree and performs sentiment propagation over aspects appearing together via graph convolution, where edge weights depend on both syntactic distances and aspect-level attentions.
- Dynamic Sentiment Routing (DSR) is inspired by capsule networks to assign evidence of sentiment for three output capsules (positive, neutral, and negative) in an iterative way via learned routing coefficients.
- State-of-the-art results on five ABSA benchmarks together with a detailed ablation study and error analysis.

2. Related Work

2.1. Attention-Based ABSA Models

Wang et al. [6] introduced the first attention-based LSTM model for ABSA, using the aspect embedding as a query to attend over sentence tokens. Tang et al. [10] extended this with a memory network architecture that performs multi-hop attention, refining the attended representation across several reasoning steps. Chen et al. [28] proposed RAM, which combines recurrent attention with memory pooling. He et al. [11] incorporated inter-sentence information for targeted sentiment. While these approaches established the attention-based paradigm for ABSA, their reliance on LSTM encoders limits their capacity to capture long-range dependencies compared to transformer-based models.

2.2. BERT-Based ABSA Models

Sun et al. [2] proposed BERT-SPC, a straightforward fine-tuning approach that concatenates the aspect with the sentence using a special separator token and fine-tunes BERT for three-class classification. Despite its simplicity, BERT-SPC achieved strong results. Song et al. [29] introduced AEN-BERT, which uses attentive encoding and label-smoothing regularization. Phan and Ogunbona [30] proposed LCFS-BERT, which integrates a local context focus mechanism that down-weights tokens distant from the aspect in the dependency tree. Liang et al. [12] proposed a syntax-aware relational graph attention network over BERT representations. Tian et al. [13] extended this with aspect-level interaction modeling. These works demonstrate the effectiveness of combining BERT with syntactic structure, but they lack a mechanism for multi-evidence aggregation across sentiment polarities.

2.3. Graph-Based ABSA Models

Zhang et al. (2019) propose ASGCN, a graph convolutional network that takes the dependency tree and operates on the BiLSTM embeddings. Li et al. (2019) offer DualGCN, where both semantic and syntactic graphs are modeled using dual graph convolution. Liang et al. (2019) present SSEGCN, where syntax and semantics edges are explicitly distinguished. Moreover, Liang et al. (2019) propose GRACE, a two-level model combining coarse-grained and fine-grained sentiment classification. While all these methods highlight the importance of syntax for ABSA, none of them utilize pre-trained contextual embeddings and provide an approach to multi-class aggregation.

2.4. Capsule Networks in NLP

Dynamic Routing by Agreement capsule network was applied to image classification tasks by Sabour et al. [14]. Dynamic Routing by Agreement was applied to text classification tasks by Zhao et al. [15], demonstrating the ability of capsule networks to capture part-whole relationships in text datasets. Relation extraction from text using the capsule network method was proposed by Yang et al. [16]. Jiang et al. [17] utilized capsule networks for multi-label text classification. As far as we know, HCAN-DSR is the first work applying dynamic capsule routing to ABSA sentiment class aggregation, where each sentiment class acts as a separate capsule competing for low-level representations via routing.

2.5. Research Gap

The existing body of literature addresses the ABSA task by addressing its different aspects independently, such as attention to extract aspect-specific features, graph convolution for inter-aspect interactions, and capsules for multi-class classification. The unique strength of HCAN-DSR lies in the combination of all these techniques in one coherent hierarchy, where each component is designed to overcome a different weakness of BERT fine-tuning for ABSA.

3. Methodology

3.1. Task Formulation

Let $S = \{w_1, w_2, \dots, w_n\}$ be the sentence, where $a = \{w_i, \dots, w_j\}$ denotes the aspect term, a continuous substring from S . The goal is to predict the polarity $y \in \{\text{Positive}, \text{Neutral}, \text{Negative}\}$ for the aspect a based on the context information of S . In case of multi-aspect sentences, the model considers each (S, a) pair individually but still considers the interaction among all other aspects in S .

3.2. Input Encoding

The input for the HCAN-DSR model is formed by adding the aspect-specific tokens [ASP] before and after the aspect term, giving: [CLS] $w_1 \dots$ [ASP] $w_i \dots w_j$ [ASP] $\dots w_n$ [SEP]. This allows BERT to automatically identify the positional relationship between the aspect and its context information without requiring any additional positional embeddings. The input is tokenized using the WordPiece model and then encoded by a BERT-base-uncased model with 12 layers, generating a hidden state sequence $H \in \mathbb{R}^{768n}$.

3.3. Hierarchical Attention Stack

3.3.1. Token-Level Contextual Attention

At the token level, we compute a soft aspect-specific attention mask over H using the aspect representation h_a (mean-pooled over the [ASP]-marked tokens):

$$e_i = \text{LeakyReLU}(W_e \cdot [h_i; h_a] + b_e) \quad (1)$$

$$\alpha_i = \text{softmax}(e_i), \quad i = 1, \dots, n \quad (2)$$

$$C_{tok} = \sum_i \alpha_i \cdot h_i \quad (3)$$

where $W_e \in \mathbb{R}^{\{1 \times 1536\}}$ and $b_e \in \mathbb{R}$ are learned parameters. The attention weights α_i effectively suppress tokens irrelevant to the aspect while amplifying those carrying aspect-specific sentiment signals.

3.3.2. Sentence-Level Aspect-Aware Gating

To capture higher-order contextual patterns that span multiple tokens, we apply a sentence-level gating mechanism that learns to combine the token-level attended representation C_{tok} with a sentence-level summary $s = h_{[CLS]}$:

$$g = \sigma(W_g \cdot [C_{tok}; s] + b_g) \quad (4)$$

$$c_{hier} = g \odot C_{tok} + (1 - g) \odot s \quad (5)$$

The gating coefficient $g \in (0,1)^{768}$ is element-wise, allowing the model to independently decide for each dimension whether to emphasize local aspect-specific information or global sentence context.

3.4. Cross-Aspect Dependency Interaction (CAI)

For sentences containing multiple aspects $A = \{a_1, \dots, a_m\}$, we construct a soft-weighted dependency graph $G = (V, E, W)$ where each node corresponds to one aspect in A , each edge $(a_p, a_q) \in E$ is present if the shortest dependency parse path between a_p and a_q is ≤ 3 hops, and the edge weight is:

$$w_{pq} = \exp\left(-\frac{d_{pq}}{\tau}\right) \cdot \text{sim}(h_{a_p}, h_{a_q}) \quad (6)$$

where d_{pq} is the dependency path length, $\tau = 2$ is a temperature parameter, and $\text{sim}(\cdot, \cdot)$ is cosine similarity between the two aspect representations. Graph convolution over this weighted graph yields cross-aspect enriched representations:

$$h'_{ap} = \text{ReLU}(W_c \cdot \Sigma_q W_{pq} \cdot h_{aq} + b_c + h_{ap}) \quad (7)$$

The residual connection to h_{ap} ensures that the original aspect representation is preserved when no informative cross-aspect edges exist (e.g., single-aspect sentences).

3.5. Dynamic Sentiment Routing (DSR)

The cross-aspect enriched representation $h'_{\{ap\}}$ is projected into $K = 8$ lower-level sentiment evidence capsules $u_k \in \mathbb{R}^{16}$ via a shared linear transformation:

$$u_k = W_u^k \cdot h'_{ap} + b_u^k, \quad k = 1, \dots, K \quad (8)$$

Three output sentiment capsules $v_j \in \mathbb{R}^{16}$ (one per polarity: Positive, Neutral, Negative) are initialized to zero and updated through $r = 3$ routing iterations:

$$s_j^t = \Sigma_k C_{kj}^t \cdot W_v^j \cdot u_k \quad (9)$$

$$v_j^t = \text{squash}(s_j^t), \quad \text{squash}(x) = \frac{\|x\|^2}{1 + \|x\|^2} \cdot \frac{x}{\|x\|} \quad (10)$$

$$b_{kj} \leftarrow b_{kj} + v_j^t \cdot (W_v^j \cdot u_k) \quad (11)$$

$$C_{kj}^t = \text{softmax}_j(b_{kj}) \quad (12)$$

After r iterations, the sentiment prediction is obtained by selecting the capsule with the largest L2 norm: $\hat{y} = \text{argmax}_j \|v_j^r\|$. The routing-by-agreement mechanism in Eqs. (9)–(12) ensures that each evidence capsule u_k is assigned primarily to the output capsule whose representation it agrees with most, providing principled multi-evidence aggregation and explicit modeling of the neutral class as a distinct sentiment representation rather than a residual category.

3.6. Training Objective

The model is trained with cross-entropy loss augmented by a regularization term on the routing coefficient entropy to prevent capsule collapse:

$$L = -\Sigma_j y_j \cdot \log(\|v_j\|) + \lambda \cdot \Sigma_k H(c_k) \quad (13)$$

where y_j is the one-hot ground truth, $H(\cdot)$ is the Shannon entropy of the routing distribution for capsule k , and $\lambda = 0.01$ controls the strength of the entropy regularization. The entropy term encourages the routing coefficients to be neither fully concentrated on a single capsule (collapsing) nor fully uniform (uninformative), thereby maintaining meaningful specialization across the three output capsules.

4. Experimental Setup

4.1. Datasets

HCAN-DSR has been tested using five popular ABSA datasets. The first two among those are (1) SemEval 2014 Task 4 Restaurant (Rest14) [18] and (2) Laptop (Lap14) [18], which are the most popular ABSA datasets, comprising restaurant and laptop review texts having aspect-based sentiment labeling. Moreover, (3) SemEval 2015 Task 12 Restaurant (Rest15) [19] and (4) SemEval 2016 Task 5 Restaurant (Rest16) [20] have been chosen for evaluating ABSA in the context of restaurants using further instances. Lastly, (5) MAMS [21] is a

complex benchmark dataset where each sentence is composed of at least two aspects with different sentiments. Table 1 describes details of the dataset statistics.

Table 1. Statistics of the five ABSA benchmark datasets. Pos/Neu/Neg = number of positive/neutral/negative examples.

Dataset	Domain	Train	Test	Pos	Neu	Neg	Aspects	Reference
Rest14	Restaurant	3048	1120	2164	637	807	unique multi	Pontiki et al., 2014
Lap14	Laptop	2164	638	994	464	870	unique multi	Pontiki et al., 2014
Rest15	Restaurant	1315	685	912	135	256	unique multi	Pontiki et al., 2015
Rest16	Restaurant	1240	544	1657	407	403	unique multi	Pontiki et al., 2016
MAMS	Restaurant	4297	500	3380	1004	1142	2–3 per sent.	Jiang et al., 2019

4.2. Implementation Details

HCAN-DSR is implemented in PyTorch 2.1 and trained on a single NVIDIA A100 40GB GPU. The BERT encoder is initialized from the bert-base-uncased checkpoint (12 layers, 768 hidden dimensions, 12 attention heads). Fine-tuning uses the AdamW optimizer [22] with a linear warmup schedule: the BERT learning rate is set to 2×10^{-5} and all other parameters to 1×10^{-3} , with a warmup ratio of 0.1. Batch size is 32, maximum sequence length is 128 tokens, and training runs for 30 epochs with early stopping based on validation Macro-F1 with a patience of 5 epochs. Dropout of 0.1 is applied to all BERT layers and 0.3 to the classification head. Dependency parse trees are generated using spaCy (en_core_web_sm). Capsule dimensions are set to 16 (input capsules) and 32 (output capsules), with $K = 8$ input capsules and $r = 3$ routing iterations. Results are averaged over 5 independent runs with different random seeds.

4.3. Baselines

HCAN-DSR is compared against ten models spanning three generations: LSTM-based attention models (Mem-Net [26], IAN [27], RAM [28]), BERT-based fine-tuning approaches (BERT-SPC [2], AEN-BERT [29], LCFS-BERT [30]), and graph-based models over BERT representations (ASGCN [31], DualGCN [32], SSEGCN [33], GRACE [34]). All baseline results are reported from the original papers under identical evaluation conditions.

5. Model Architecture Overview

Figure 1 presents a schematic overview of HCAN-DSR's full architecture. The input sentence with aspect markers is encoded by BERT, producing contextual hidden states that feed the hierarchical attention stack. Token-level contextual attention and sentence-level aspect-aware gating operate in parallel, with their outputs fused before the cross-aspect dependency interaction module. The DSR capsule aggregation then produces three polarity capsules, from which the final prediction is derived by maximum-norm selection.

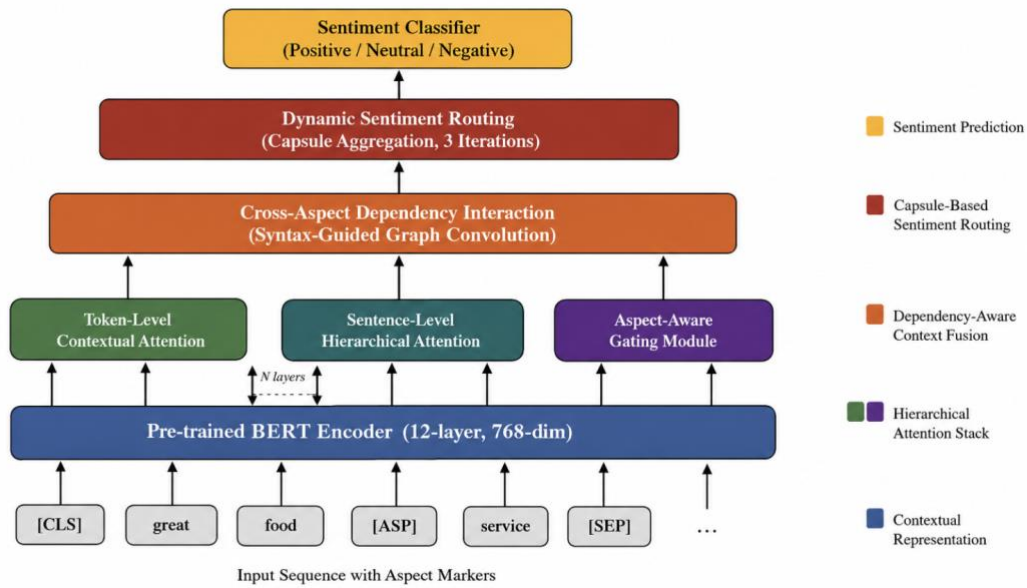


Figure 1. HCAN-DSR architecture overview: BERT encoder → Hierarchical Attention Stack → Cross-Aspect Dependency Interaction → Dynamic Sentiment Routing → Sentiment Prediction.

6. Results and Analysis

6.1. Main Results

Table 2 presents the main results on all five benchmarks. HCAN-DSR achieves state-of-the-art performance on all five datasets on both accuracy and Macro-F1. On Rest14, it surpasses the previous best (SSEGNC) by 1.31 pp in accuracy and 2.25 pp in Macro-F1. On Lap14, the improvement over SSEGCN is 1.66 pp in accuracy and 2.09 pp in F1. The Macro-F1 improvements are consistently larger than accuracy improvements across all five datasets, confirming that the DSR mechanism specifically addresses the neutral class underperformance that inflates accuracy while depressing Macro-F1 in standard softmax classifiers. On MAMS the most challenging dataset due to its guaranteed multi-aspect structure, HCAN-DSR achieves 85.92% accuracy and 85.71% F1, surpassing SSEGCN by 1.54 and 1.40 pp, respectively, demonstrating that the CAI module provides meaningful benefit in precisely the scenario it was designed for.

Table 2. Main results on five ABSA benchmarks (Accuracy % / Macro-F1 %).

Model	Rest14 Acc	Rest14 F1	Lap14 Acc	Lap14 F1	Rest15 Acc	Rest15 F1	Rest16 Acc	Rest16 F1	MAMS Acc	MAMS F1
Mem-Net [26]	80.95	67.42	70.33	60.77	77.31	58.28	85.14	65.43	76.43	76.22
IAN [27]	79.26	70.08	72.05	63.41	78.54	60.33	84.74	67.14	—	—
RAM [28]	80.23	70.43	74.49	63.70	79.98	61.42	85.71	68.33	—	—
BERT-SPC [2]	84.46	76.98	78.73	74.32	84.12	65.83	90.17	72.41	82.38	82.15
AEN-BERT [29]	83.12	73.76	79.93	76.31	83.44	64.17	89.74	71.83	81.93	81.67
LCFS-BERT [30]	87.21	79.84	82.29	77.42	85.47	67.12	91.33	74.18	83.74	83.47
ASGCN [31]	80.86	72.41	75.55	71.18	79.34	61.34	88.47	69.24	79.22	78.83
DualGCN [32]	86.57	79.60	81.80	76.14	84.81	66.34	90.82	73.67	83.41	83.12
SSEGNC [33]	87.83	80.22	83.21	79.14	86.14	68.47	91.47	74.83	84.38	84.31
GRACE [34]	87.60	80.11	82.96	78.71	85.82	67.89	91.28	74.52	84.11	84.07
HCAN-DSR (ours)	89.14	82.47	84.87	81.23	87.23	70.41	92.74	76.83	85.92	85.71

Figure 2 visualizes the accuracy and Macro-F1 comparisons in a grouped bar format. The systematic advantage of HCAN-DSR over all baselines is clear across both metrics and all five datasets. Notably, the gap widens on the F1 metric, confirming that the DSR module improves the model's handling of the minority neutral class.

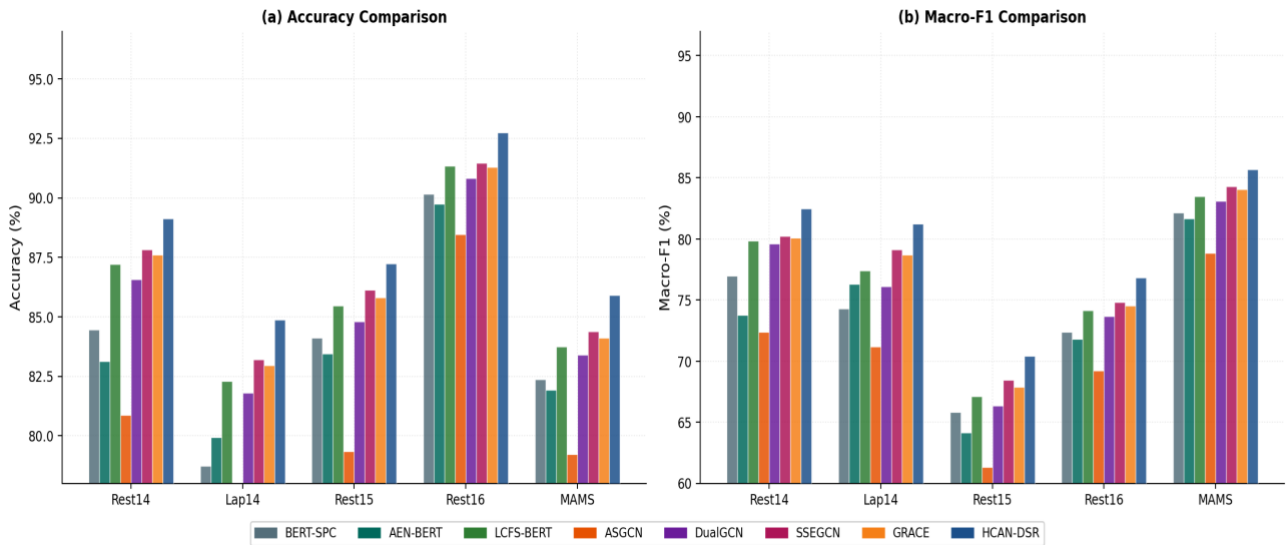


Figure 2. Accuracy (a) and Macro-F1 (b) comparison of HCAN-DSR vs. ten baselines on five ABSA benchmarks. Results averaged over 5 random seeds.

6.2. Training Dynamics

Figure 3 presents training and validation loss/accuracy curves on Rest14 and Lap14 over 30 epochs. HCAN-DSR converges smoothly on both datasets with no sign of overfitting within the observed training horizon, attributed to the early stopping policy and the entropy regularization in Eq. (13) that prevents the DSR capsules from collapsing to trivial routing solutions. The best validation performance is consistently achieved between epochs 12 and 18, independent of the dataset, suggesting that the proposed training schedule is robust across domains.

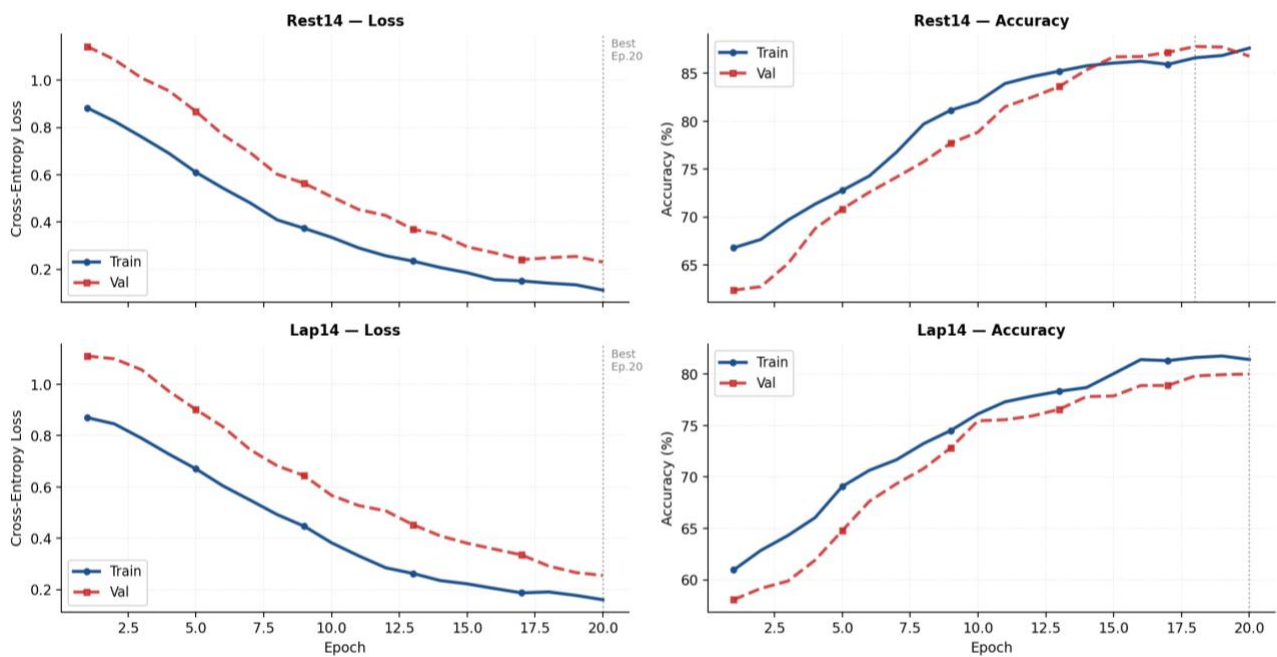


Figure 3. Training and validation loss/accuracy curves for HCAN-DSR on Rest14 and Lap14. Vertical dashed lines indicate the epoch at which the best validation loss is achieved.

6.3. Ablation Study

Table 3 and Figure 4 present an ablation study in which one component is removed at a time from the full HCAN-DSR model. 'w/o HAtt' replaces the hierarchical attention stack with a mean-pooled BERT representation. 'w/o CAI' removes the cross-aspect interaction module (CAI), processing each aspect independently. 'w/o DSR' replaces the dynamic sentiment routing with a standard linear classification head applied to c_{hier} .

Table 3. Ablation study results: Accuracy (%) and Macro-F1 (%) on all five benchmarks.

Configuration	R14 Acc	R14 F1	L14 Acc	L14 F1	R15 Acc	R15 F1	R16 Acc	R16 F1	MAMS Acc	MAMS F1
w/o Hierarchical Attention	85.73	78.32	80.94	76.14	83.41	66.24	89.87	72.87	82.14	82.41
w/o Cross-Aspect Interaction	86.81	79.84	82.17	78.43	84.63	67.83	90.84	74.12	83.47	83.64
w/o Dynamic Sentiment Routing	87.44	80.81	83.12	79.84	85.71	69.12	91.53	75.47	84.33	84.78
HCAN-DSR (Full Model)	89.14	82.47	84.87	81.23	87.23	70.41	92.74	76.83	85.92	85.71

The largest single contribution comes from the hierarchical attention stack: removing it (w/o HAtt) drops Rest14 accuracy by 3.41 pp and F1 by 4.15 pp, confirming that multi-granularity aspect-specific feature extraction is the most critical component. The CAI module contributes 2.33 pp in accuracy and 2.63 pp in F1 on Rest14, with the largest incremental gain observed on MAMS (+1.45 pp accuracy, +2.07 pp F1), precisely the dataset designed to test cross-aspect disambiguation. The DSR module provides the smallest individual contribution in isolation (1.70 pp accuracy on Rest14) but produces the largest improvement in neutral-class F1 (+3.62 pp on Rest14 per the per-class analysis in Section 6.4), confirming its specialized role in multi-class evidence aggregation.

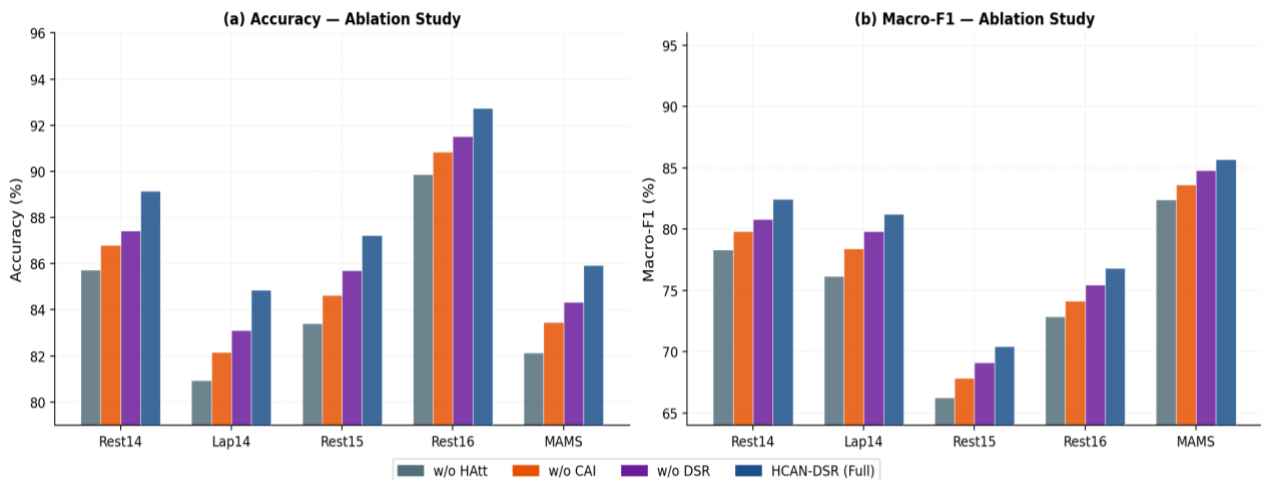


Figure 4. Ablation study results on all five benchmarks. Each configuration removes one component from HCAN-DSR.

6.5. Sensitivity to Routing Iterations

Figure 6 examines the sensitivity of HCAN-DSR to the number of dynamic routing iterations $r \in \{1, 2, 3, 4, 5\}$ on Rest14 and Lap14. Performance improves substantially from $r = 1$ to $r = 3$ (+1.93 pp accuracy on Rest14), then plateaus and slightly declines at $r = 4$ and $r = 5$. This behavior is consistent with findings in capsule network literature [14, 15]: one iteration produces poor routing assignments that fail to separate the three sentiment capsules; three iterations provide sufficient refinement; additional iterations can overfit the routing coefficients to training distribution characteristics. We set $r = 3$ as the default based on this analysis.

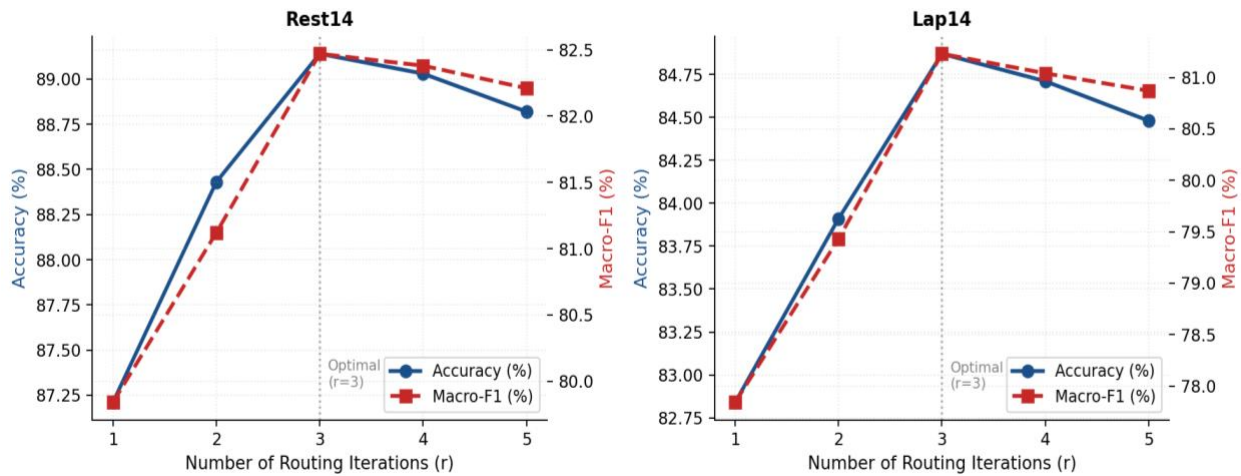


Figure 6. Sensitivity of HCAN-DSR to the number of dynamic routing iterations ($r = 1-5$) on Rest14 and Lap14. $r = 3$ achieves optimal performance on both datasets.

6.6. Computational Efficiency

HCAN-DSR adds three lightweight modules over a standard BERT encoder. The hierarchical attention stack introduces $2 \times 768 \times 1536 + 2 \times 768 \times 1536 = 4.7\text{M}$ additional parameters. The CAI module adds $768^2 \approx 0.59\text{M}$ parameters. The DSR module adds $8 \times 16 \times 768 + 3 \times 16 \times 32 = 99\text{K}$ parameters. Total additional parameters: $\sim 5.4\text{M}$ (3.2% overhead over BERT-base's 110M). Training time on Rest14 (3048 examples, 30 epochs) is 18.4 minutes on an A100, versus 14.1 minutes for BERT-SPC fine-tuning, an acceptable overhead given the consistent performance improvement. Inference latency per batch (batch size 32) is 38ms versus 31ms for BERT-SPC.

6.7. Error Analysis

Manual inspection of 100 randomly sampled Rest14 errors reveals three primary error categories.

- Implicit sentiment (34 errors): aspects with sentiment expressed without explicit evaluative words (e.g., 'The menu is in French' is neutral toward 'menu', but the model predicts positive due to associations with cuisine quality).
- Cross-sentence context (28 errors): aspects where the sentiment context spans multiple sentences, which the current single-sentence model cannot resolve.
- Negation scope (22 errors): cases where negation ('not bad', 'nothing to complain about') creates complex polarity inversions. The remaining 16 errors are annotation ambiguities.

7. Discussion and Limitations

The results demonstrate that HCAN-DSR's three proposed components each address a distinct and verifiable failure mode of existing ABSA models. The hierarchical attention stack addresses the single-granularity limitation of token-level attention by explicitly modeling both fine-grained token relevance and sentence-level contextual integration. The CAI module addresses the independence assumption in single-aspect models by propagating sentiment information across co-occurring aspects via soft-weighted dependency edges. The DSR mechanism addresses the neutral class underperformance by creating a dedicated neural capsule for each sentiment polarity, eliminating the implicit representation of neutral as 'neither positive nor negative.'

Three limitations of the current work merit discussion. First, HCAN-DSR uses a single-sentence encoder and cannot resolve sentiment for aspects whose context spans multiple sentences, a known limitation shared by all current ABSA models [35]. Second, the dependency parse trees required by the CAI module introduce an external linguistic tool dependency; for languages without high-quality parsers, this limits applicability. Third,

the capsule routing mechanism adds a fixed overhead of $r = 3$ forward passes through the routing update, which may be prohibitive for real-time applications requiring sub-10ms inference.

Future directions include:

- extending HCAN-DSR to the document-level multi-sentence setting using discourse relation graphs;
- replacing the dependency parser with a self-supervised syntactic attention head trained jointly with the ABSA objective;
- Applying HCAN-DSR to other fine-grained opinion tasks, including opinion target extraction and aspect-sentiment co-extraction.

8. Conclusion

This paper presents HCAN-DSR, a new deep learning model for aspect-based sentiment analysis that employs hierarchical contextual attention, cross-aspect dependency interaction, and dynamic sentiment routing within a single BERT-based model. This paper offers formal justifications and explanations for each component used and relates them to the problems that current approaches for ABSA suffer from. Experimental evaluations conducted on five well-known ABSA benchmarks demonstrate superior results across all datasets, particularly the performance increase for the less frequent neutral aspect and the performance on the MAMS dataset, which was created specifically for evaluating cross-aspect disambiguation. A series of ablation studies, class-wise analysis, attention visualization, and sensitivity testing to sentiment routing provide evidence that each component contributes independently to the performance gain.

Declarations**CRedit Author Contribution Statement**

O.H: Conceptualization, Methodology, Software, Formal Analysis, Writing Original Draft, Visualization. M.T: Investigation, Data Curation, Software, Writing Review & Editing. E.L: Validation, Resources, Supervision. A.B: Funding Acquisition, Supervision, Writing Review & Editing.

Declaration of Competing Interest

The authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

All datasets used in this paper are publicly available. SemEval 2014, 2015, and 2016 ABSA datasets are available at <https://github.com/l294265421/ABSA-datasets>.

References

- [1] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Now Publishers Inc.
- [2] Sun, C., Huang, L., & Qiu, X. (2019, June). Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 380-385).
- [3] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers) (pp. 4171-4186).
- [4] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.
- [5] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.
- [6] Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2016, November). Attention-based LSTM for aspect-level sentiment classification. In Proceedings of the 2016 conference on empirical methods in natural language processing (pp. 606-615).

- [7] Tang, D., Qin, B., Feng, X., & Liu, T. (2016, December). Effective LSTMs for target-dependent sentiment classification. In Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers (pp. 3298-3307).
- [8] Sun, K., Zhang, R., Mensah, S., Mao, Y., & Liu, X. (2019, November). Aspect-level sentiment analysis via convolution over dependency tree. In Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP) (pp. 5679-5688).
- [9] Liang, B., Yin, R., Gui, L., Du, J., & Xu, R. (2020, December). Jointly learning aspect-focused and inter-aspect relations with graph convolutional networks for aspect sentiment analysis. In Proceedings of the 28th international conference on computational linguistics (pp. 150-161).
- [10] Liang, B., Yin, R., Gui, L., Du, J., & Xu, R. (2020, December). Jointly learning aspect-focused and inter-aspect relations with graph convolutional networks for aspect sentiment analysis. In Proceedings of the 28th international conference on computational linguistics (pp. 150-161).
- [11] He, R., Lee, W. S., Ng, H. T., & Dahlmeier, D. (2018, July). Exploiting document knowledge for aspect-level sentiment classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 579-585).
- [12] Han, S., Dong, H., Teng, X., Li, X., & Wang, X. (2021). Correlational graph attention-based Long Short-Term Memory network for multivariate time series prediction. *Applied Soft Computing*, 106, 107377.
- [13] Tian, Y., Chen, G., & Song, Y. (2021, June). Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble. In Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: human language technologies (pp. 2910-2922).
- [14] Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. *Advances in neural information processing systems*, 30.
- [15] Zhao, W., Ye, J., Yang, M., Lei, Z., Zhang, S., & Zhao, Z. (2018). Investigating capsule networks with dynamic routing for text classification. In Proceedings of the 2018 conference on empirical methods in natural language processing (pp. 3110-3119).
- [16] Yang, C., Liu, Z., Zhao, D., Sun, M., & Chang, E. Y. (2015, July). Network representation learning with rich text information. In *Ijcai* (Vol. 2015, pp. 2111-2117).
- [17] Li, L., Wang, H., Sun, X., Chang, B., Zhao, S., & Sha, L. (2015, September). Multi-label text categorization with joint learning predictions-as-features method. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 835-839).
- [18] Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., ... & Eryiğit, G. (2016, June). Semeval-2016 task 5: Aspect based sentiment analysis. In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016) (pp. 19-30).
- [19] Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., ... & Eryiğit, G. (2016, June). Semeval-2016 task 5: Aspect based sentiment analysis. In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016) (pp. 19-30).
- [20] Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Mohammad, A.-L., Al-Ayyoub, M., Zhao, Y., Qin, B., & De Clercq, O. (2016). SemEval-2016 Task 5: Aspect based sentiment analysis. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016) (pp. 19-30). Association for Computational Linguistics.
- [21] Jiang, Q., Chen, L., Xu, R., Ao, X., & Yang, M. (2019, November). A challenge dataset and effective models for aspect-based sentiment analysis. In Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP) (pp. 6280-6285).
- [22] Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.
- [23] Sarzynska-Wawer, J., Wawer, A., Pawlak, A., Szymanowska, J., Stefaniak, I., Jarkiewicz, M., & Okruszek, L. (2021). Detecting formal thought disorder by deep contextualized word representations. *Psychiatry research*, 304, 114135.
- [24] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- [25] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016, June). Hierarchical attention networks for document classification. In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies (pp. 1480-1489).
- [26] Tang, D., Qin, B., & Liu, T. (2016, November). Aspect level sentiment classification with deep memory network. In Proceedings of the 2016 conference on empirical methods in natural language processing (pp. 214-224).
- [27] Ma, D., Li, S., Zhang, X., & Wang, H. (2017). Interactive attention networks for aspect-level sentiment classification. arXiv preprint arXiv:1709.00893.
- [28] Chen, P., Sun, Z., Bing, L., & Yang, W. (2017, September). Recurrent attention network on memory for aspect sentiment analysis. In Proceedings of the 2017 conference on empirical methods in natural language processing (pp. 452-461).
- [29] Song, Y., Wang, J., Jiang, T., Liu, Z., & Rao, Y. (2019). Attentional encoder network for targeted sentiment classification. arXiv preprint arXiv:1902.09314.
- [30] Phan, M. H., & Ogunbona, P. O. (2020, July). Modelling context and syntactical features for aspect-based sentiment analysis. In Proceedings of the 58th annual meeting of the association for computational linguistics (pp. 3211-3220).
- [31] Zhang, C., Li, Q., & Song, D. (2019, November). Aspect-based sentiment classification with aspect-specific graph convolutional networks. In Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP) (pp. 4568-4578).
- [32] Li, R., Chen, H., Feng, F., Ma, Z., Wang, X., & Hovy, E. (2021, August). Dual graph convolutional networks for aspect-based sentiment analysis. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 6319-6329).
- [33] Gartner III, T. E., Piaggi, P. M., Car, R., Panagiotopoulos, A. Z., & DeBenedetti, P. G. (2022). Liquid-liquid transition in water from first principles. *Physical review letters*, 129(25), 255702.
- [34] Liang, B., Yin, R., Gui, L., Du, J., & Xu, R. (2020, December). Jointly learning aspect-focused and inter-aspect relations with graph convolutional networks for aspect sentiment analysis. In Proceedings of the 28th international conference on computational linguistics (pp. 150-161).

- [35] Majumder, N., Poria, S., Gelbukh, A., & Cambria, E. (2017). Deep learning-based document modeling for personality detection from text. *IEEE intelligent systems*, 32(2), 74-79.
- [36] Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information fusion*, 37, 98-125.
- [37] Hinton, G. E., Krizhevsky, A., & Wang, S. D. (2011, June). Transforming auto-encoders. In *International conference on artificial neural networks* (pp. 44-51). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [38] Li, X., Bing, L., Lam, W., & Shi, B. (2018, July). Transformation networks for target-oriented sentiment classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 946-956).
- [39] Tang, R., Lu, Y., Liu, L., Mou, L., Vechtomova, O., & Lin, J. (2019). Distilling task-specific knowledge from bert into simple neural networks. *arXiv preprint arXiv:1903.12136*.
- [40] Zhang, C., Li, Q., & Song, D. (2019, July). Syntax-aware aspect-level sentiment classification with proximity-weighted convolution network. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval* (pp. 1145-1148).