Team VYN at SemEval-2026 Task 3: Dimensional Aspect-Based Sentiment Analysis

Vishal Thenuwara, Yomal de Mel, and **Nisansa de Silva**
Department of Computer Science & Engineering
University of Moratuwa, Sri Lanka
{vishal.23,yomal.23,nisansa}@cse.mrt.ac.lk

Abstract

This paper describes our system for the DimABSA 2026 Shared Task (SemEval-2026 Task 3), which requires predicting continuous Valence-Arousal (VA) scores in $[1, 9]$ for aspect-level sentiments instead of categorical labels. We develop an approach to address all three Track A subtasks: aspect-level VA regression, triplet extraction. The DESS (DeBERTa Enhanced Syntactic-Semantic) model with VA Regression Head achieves 8.22% micro cF1 on the development set on the triplet extraction subtask. We analyze the challenges of adapting categorical ABSA (Aspect Based Sentiment Analysis) architectures to continuous dimensional sentiment and discuss directions for improvement. We release our code at https://github.com/VishalRepos/DimABSA2026.

1 Introduction

Aspect-Based Sentiment Analysis (ABSA) has traditionally relied on categorical sentiment labels (positive, negative, neutral) to characterize opinions toward specific aspects in text in domains such as customer reviews (Jayakody et al., 2024b; Pontiki et al., 2014; Gunathilaka and De Silva, 2022; Jayakody et al., 2024a, 2025) and law (Mudalige et al., 2020; Rajapaksha et al., 2020, 2021a,b). However, this coarse-grained representation fails to capture the nuanced intensity and emotional dimensions of human sentiment (De Mel and de Silva, 2025). The DimABSA 2026 Shared Task (Lee et al., 2026) addresses this limitation by introducing *dimensional* sentiment analysis into the ABSA framework, where sentiment is represented as continuous Valence-Arousal (VA) scores. We participate in Track A (English), addressing all three subtasks across the restaurant and laptop review domains.

Our approach adapts the DESS architecture (Thenuwara and de Silva, 2025), a state-of-the-art span-based model for ASTE (Aspect Sentiment Triplet Extraction) that leverages dual-channel graph convolutional networks (GCNs) to capture both syntactic and semantic relationships. We modify the sentiment classification head to output continuous VA predictions. MSE is used in the DESS model and used for VA regression.

Our experiments reveal that directly adapting categorical span-extraction models to continuous regression is non-trivial. The system achieves a micro F1 of 8.22% for triplet extraction on the development set, with entity extraction proving more robust than VA prediction.

2 Background

2.1 Related Work

Aspect Sentiment Triplet Extraction has been addressed through span-based methods (Xu et al., 2021), sequence tagging (Wu et al., 2020), instruction learning (Scaria et al., 2024; Jayakody et al., 2025) and generative approaches (Zhang et al., 2021). The DESS model (Thenuwara and de Silva, 2025) represents the model ASTE, using dual-channel GCNs to integrate syntactic dependency and semantic similarity information. The key challenge lies in jointly performing span extraction (a structured prediction task) and VA regression (a continuous prediction task) within a unified framework.

3 System Overview

3.1 Architecture

We have suggested three different architectures for three subtasks in Figure 1, Figure 2 and Figure 3 respectively. Common updates applied on the DESS architecture are described in Appendix A, while further information and larger scale diagrams for the specific subtasks can be found in Appendix B.

Proceedings of the The 20th International Workshop on Semantic Evaluation (2026), pages 1755–1764

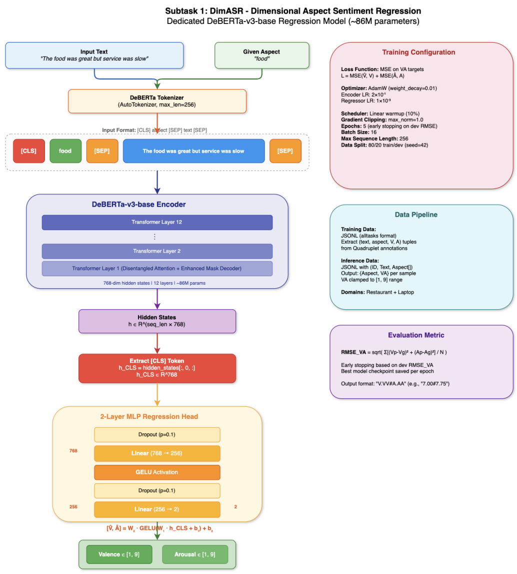

Figure 1: Architecture of the Subtask 1 (DimASR) regression model. The input text–aspect pair is encoded as `[CLS] aspect [SEP] text [SEP]` and passed through a DeBERTa-v3-base encoder. The `[CLS]` token representation is fed into a two-layer MLP regression head (768→256→2) with GELU activation and dropout, producing continuous Valence and Arousal predictions in $[1, 9]$.

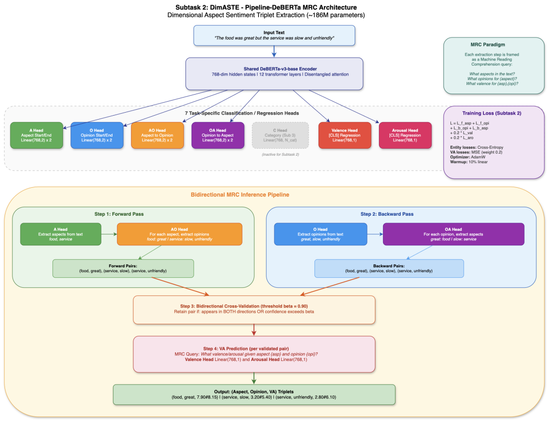

Figure 2: Architecture of the Subtask 2 (DimASTE) Pipeline-DeBERTa system. A shared DeBERTa-v3-base encoder feeds seven task-specific heads (A, O, AO, OA, C, Valence, Arousal). Bidirectional Machine Reading Comprehension (MRC) extraction identifies aspect–opinion pairs through forward (A→AO) and backward (O→OA) passes, followed by cross-validation with confidence threshold β=0.90 to filter spurious pairings. Validated pairs are scored by the Valence and Arousal regression heads to produce (Aspect, Opinion, VA) triplets.

Encoder. We use DeBERTa-v3-base (He et al., 2023) as the pretrained encoder, producing 768-dimensional contextualised token representations.

BiLSTM Layer. A 2-layer bidirectional LSTM with hidden dimension 384 processes the encoder outputs, capturing sequential dependencies.The BiLSTM output is projected back to 768 dimensions via a linear layer.

Dual-Channel GCN. Two parallel GCN channels process the encoded representations: Syntactic GCN (Syn-GCN): A 2-layer GCN operating on the dependency parse tree adjacency matrix, capturing syntactic relationships between tokens.Semantic GCN (Sem-GCN): A 2-layer GCN with multi-head self-attention that learns semantic similarity-based adjacency, capturing implicit semantic relationships.

Tree Interaction Network (TIN). The TIN module fuses the dual-channel GCN outputs through residual connections and a BiLSTM fusion layer:

$$h_{\text{fuse}} = \text{MLP}(\text{BiLSTM}([h_{\text{syn}}; h_{\text{sem}}])) \quad (1)$$

where h_{syn} and h_{sem} are the syntactic and semantic GCN outputs after residual connections with the original encoder features.

Entity Classification. Candidate entity spans (up to length 8) are classified into {`None`, `Target`, `Opinion`} using span representations formed by max-pooling token embeddings within each span, concatenated with a learned size embedding:

$$\hat{y}_{\text{ent}} = W_e[h_{\text{span}}; e_{\text{size}}] + b_e \quad (2)$$

VA Regression Head. This is our key modification. we predict Valence and Arousal scores for each aspect–opinion pair within a sentence. The model integrates the sentence-level embedding with contextual embeddings of the corresponding aspect and opinion terms. A base emotional representation derived from the sentence is refined through aspect-specific adjustments to capture localized sentiment intensity. The final Valence–Arousal scores are generated per aspect, enabling fine-grained emotion modeling within multi-aspect sentences.

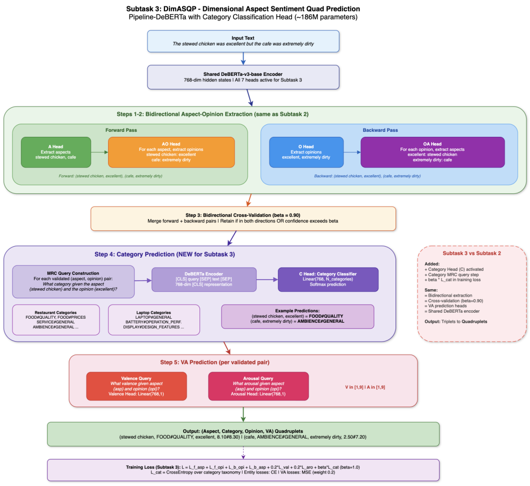

Figure 3: Architecture of the Subtask 3 (DimASQP) quad prediction system. The model extends the Subtask 2 Pipeline-DeBERTa architecture by activating the Category head (C), which classifies each validated aspect–opinion pair into a domain-specific taxonomy (e.g., `FOOD#QUALITY`, `SERVICE#GENERAL`) via an MRC query. The training loss augments the Subtask 2 objective with a weighted category cross-entropy term $\beta \cdot \mathcal{L}_{\text{cat}}$, producing (Aspect, Category, Opinion, VA) quadruplets.

3.2 Training Objective

The total loss combines entity classification loss and VA regression loss:

$$\mathcal{L} = \mathcal{L}_{\text{entity}} + \mathcal{L}_{\text{VA}} + 10 \cdot \mathcal{L}_{\text{batch}} \tag{3}$$

The entity loss uses cross-entropy:

$$\mathcal{L}_{\text{entity}} = \frac{\sum_i m_i \cdot \text{CE}(\hat{y}_i, y_i)}{\sum_i m_i} \tag{4}$$

where m_i is the sample mask.

The batch regularization loss $\mathcal{L}_{\text{batch}}$ in Eq. 5 penalizes high variance in VA predictions within each training batch, encouraging stable and consistent predictions:

$$\mathcal{L}_{\text{batch}} = \text{Var}(\hat{V}_{\text{batch}}) + \text{Var}(\hat{A}_{\text{batch}}) \tag{5}$$

The weight factor of 10 was determined through preliminary experiments on the development set; without this scaling, the batch regularization term had negligible effect on training dynamics due to its small magnitude relative to the entity and VA losses.

4 Experimental Setup

We use the DimABSA 2026 English dataset for Track A, Subtask 2. Table 1 summarizes the dataset statistics. Further information on data is available in Appendix C. The subsequent preprocessing is described in Appendix D. Hyperparameters used in the set up is discussed in Appendix E.

Domain	Statistic	Train	Test
Restaurant	Samples	1,448	200
	Avg. entities/sample	3.35	–
	Avg. sentiments/sample	1.68	–
Laptop	Samples	2,279	200
	Avg. entities/sample	2.87	–
	Avg. sentiments/sample	1.43	–
Total samples		3,727	400

Table 1: Dataset statistics for the English training and test sets, reported per domain.

5 Results

5.1 Main Results

The DESS model shows a clear learning trajectory across training epochs. In early epochs (1–2), the model produces no valid triplet predictions as both entity extraction and VA regression are still initializing. By epoch 3, the model begins generating valid predictions with a micro F1 of 3.55% (precision 5.18%, recall 2.70%). Performance improves steadily, reaching 5.72% F1 at epoch 4 and peaking at 8.22% micro F1 at epoch 5 (precision 8.40%, recall 8.06%). We note that the limited computational budget (Kaggle T4 GPUs, batch size of 4) constrained our ability to perform extensive hyperparameter tuning and longer training, which likely contributed to the modest performance.

5.2 Subtask 1: DimASR

The Subtask 1 model generates Valence–Arousal (VA) predictions for each text–aspect pair. Table 2 presents representative predictions from the development set across both laptop and restaurant domains. Overall, the model demonstrates reasonable performance, with several predictions closely aligning with the ground truth, as reflected by low Euclidean distances (e.g., *service* and *staff*). However, performance varies across aspects and domains. While the model captures general sentiment trends for positively expressed aspects (e.g., *food*, *touchscreen*), it exhibits larger deviations for certain cases such as *battery life* and *sound*, indicating difficulty in accurately modeling nuanced or context-dependent emotional intensity. These results suggest that although the model is capable

Aspect	V#A		Euclidean Distance
	Actual	Predicted	
Laptop domain			
touchscreen	7.80#7.60	7.00#7.75	0.81
keyboard	6.88#6.62	6.33#7.19	0.79
sound	7.12#7.00	8.17#8.23	1.62
battery life	3.50#6.25	6.03#7.16	2.69
Restaurant domain			
food	7.50#7.75	7.90#8.15	0.57
service	7.75#7.75	7.55#7.67	0.22
staff	7.33#7.67	7.28#7.82	0.16
food	7.75#6.50	6.47#7.11	1.42

Table 2: Subtask 1 (DimASR) sample predictions on the development set. VA format: Valence#Arousal.

Aspect	Opinion	V#A
Laptop domain		
battery life	perfect	7.32#7.12
computer	highly recommend	8.11#7.98
mac os	flawlessly	6.98#7.32
laptop	great	7.19#7.50
Restaurant domain		
stewed chicken	excellent	7.88#7.73
service	absolutely horrendous	3.19#7.53
food	excellent	8.10#7.64
server	super attentive	5.75#6.98

Table 3: Subtask 2 (DimASTE) sample triplet extractions on the test set. Aspect-opinion pairs are shown; VA scores are predicted per pair.

of learning coarse-grained VA patterns, it struggles with fine-grained emotion estimation in more complex scenarios.

5.3 Subtask 2: DimASTE

For Subtask 2, the Pipeline-DeBERTa system extracts (Aspect, Opinion, VA) triplets. Table 3 shows sample predictions on the test set. The system successfully identifies multi-word aspects and opinions, and produces VA scores that reflect sentiment polarity. The DESS model achieves 8.22% micro cF1 on the development set (precision 8.40%, recall 8.06%), with entity extraction proving more robust than VA prediction.

5.4 Subtask 3: DimASQP

For Subtask 3, the pipeline extends triplet extraction with category prediction. Table 4 shows sample quadruplet predictions. The category classifier correctly assigns domain-specific categories in most cases (e.g., `FOOD#QUALITY` for food items, `SERVICE#GENERAL` for service aspects).

Aspect	Opinion	Category
Laptop domain		
battery life	fairly low	`BATTERY#OPERATION_PERFORMANCE.`
fans	loud	`MULTIMEDIA#OPERATION_PERFORMANCE.`
laptop	great	`LAPTOP#GENERAL`
Restaurant domain		
stewed chicken	excellent	`FOOD#QUALITY`
service	horrendous	`SERVICE#GENERAL`
cafe	extremely dirty	`AMBIENCE#GENERAL`

Table 4: Subtask 3 (DimASQP) sample quadruplet predictions on the test set, showing aspect, opinion, and predicted category.

Subtask	R^2 (Valence)	R^2 (Arousal)
Subtask 1	0.589	-0.349
Subtask 2	0.659	0.412
Subtask 3	0.681	0.423

Table 5: R^2 performance for Valence (V) and Arousal (A) prediction across Subtasks 1–3.

5.5 Analysis

Table 5 summarizes the R^2 scores for Valence and Arousal prediction across the three subtasks. Subtask 1 achieves reasonable performance for Valence prediction (0.589), but the negative R^2 value for Arousal (-0.349) indicates that the model performs worse than a mean-based baseline for this dimension. This suggests that arousal signals in Subtask 1 are more difficult to model, potentially due to weaker intensity cues or higher distributional variance. Subtask 2 demonstrates clear improvements in both Valence (0.659) and Arousal (0.412), reflecting the benefit of aspect-aware modeling and better contextual alignment. The strongest overall performance is observed in Subtask 3, which achieves the highest R^2 values for both Valence (0.681) and Arousal (0.423), indicating improved stability and more effective fine-grained emotion prediction.

Table 6 compares our system's performance against the official baselines provided by the task organizers (Yu et al., 2026) on the DimABSA 2026 dataset (Lee et al., 2026). Our system scores below both LLM-based baselines, which highlights the difficulty of adapting span-based categorical architectures to continuous VA regression compared to general-purpose large language models. Further analysis on errors can be found in Appendix F.

System	Score
Ours (Team VYN)	1.7978
Baseline (Kimi-K2 Thinking)	2.1461
Baseline (Qwen-3 14B)	2.6427

Table 6: Comparison with official baselines on the DimABSA 2026 shared task.And this is specifically for Restaurant domain English dataset

6 Discussion

Lessons Learned. Adapting a categorical span-extraction model to continuous regression is more challenging than a simple head replacement. The entire training pipeline (including negative sampling, loss balancing, and evaluation) needs to be redesigned for the regression setting.

Future work should consider:

- Decoupling entity extraction from VA regression, training the entity extractor first and then fine-tuning the VA head
- Using a dedicated regression loss such as Smooth L1 or Huber loss instead of MSE
- Implementing the continuous F1 metric directly as a training signal or for model selection
- Exploring binned VA prediction as an intermediate approach between classification and regression

7 Conclusion

We presented Team VYN's system for the DimABSA 2026 Shared Task, addressing all three Track A subtasks. Our approach combines a dedicated DeBERTa regression model for Subtask 1, an MRC-based pipeline for Subtasks 2 and 3, and DESS as an alternative for Subtask 2. Our experiments highlight the fundamental challenges of transitioning from categorical to continuous dimensional sentiment in structured ABSA tasks. Key difficulties include VA label space explosion in span-based models, loss balancing between extraction and regression objectives, and biased VA distributions in training data. Future work should explore architectures specifically designed for joint span extraction and continuous regression, decoupled training strategies, and data augmentation to address VA distribution imbalance. The system achieves 8.22% micro F1 for triplet extraction (Subtask 2) on the development set, indicating substantial room for improvement. We believe that architectures specifically designed for joint span extraction and continuous regression, rather than adapted classification models, will be necessary to advance performance on this challenging task.

Limitations

Our work has several limitations. First, we only evaluate on English data, while the DimABSA task covers multiple languages. Second, our DESS adaptation treats VA prediction as a simple head replacement without deeper architectural changes to accommodate regression. Third, computational constraints limited our hyperparameter search, and the model was trained on Kaggle T4 GPUs with small batch sizes. Finally, the evaluation framework inherited from D2E2S uses exact string matching for sentiment types, which is fundamentally incompatible with continuous VA predictions and likely underestimates actual model performance. Limitation in the contect of LLMs in discussed in Appendix G.

Ethics Statement

This work uses publicly available datasets from the DimABSA 2026 shared task. The restaurant and laptop review data does not contain personally identifiable information. Our models are trained for research purposes and are not intended for deployment in production systems without further validation.

Acknowledgments

We thank the DimABSA 2026 organizers for providing the dataset and evaluation framework. We also acknowledge the authors of the D2E2S model for making their code publicly available. Computational resources were provided by Kaggle.

References

Yomal De Mel and Nisansa de Silva. 2025. GeeSanBhava: Sentiment Tagged Sinhala Music Video Comment Data Set. In *International Conference on Computational Collective Intelligence*, pages 157–171. Springer.

Sadeep Gunathilaka and Nisansa De Silva. 2022. Aspect-based sentiment analysis on mobile application reviews. In *2022 22nd International Con-*

ference on Advances in ICT for Emerging Regions (ICTer), pages 183–188. IEEE.

Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2023. DeBERTaV3: Improving DeBERTa using ELECTRA-style pre-training with gradient-disentangled embedding sharing. In *International Conference on Learning Representations*.

Dineth Jayakody, Koshila Isuranda, A V A Malkith, Nisansa de Silva, Sachintha Rajith Ponnamperuma, G G N Sandamali, and K L K Sudheera. 2024a. Aspect-Based Sentiment Analysis Techniques: A Comparative Study. In *2024 Moratuwa Engineering Research Conference (MERCon)*, pages 205–210. IEEE.

Dineth Jayakody, Koshila Isuranda, Ava Malkith, Nisansa de Silva, Sachintha Rajith Ponnamperuma, Ggn Sandamali, Klk Sudheera, and Kashnika Gimhani Sarathchandra. 2024b. Enhanced aspect-based sentiment analysis with integrated category extraction for instruct-DeBERTa. In *Proceedings of the 38th Pacific Asia Conference on Language, Information and Computation*, pages 665–674. Tokyo University of Foreign Studies.

Dineth Jayakody, A V A Malkith, Koshila Isuranda, Vishal Thenuwara, Nisansa de Silva, Sachintha Rajith Ponnamperuma, G G N Sandamali, and K L K Sudheera. 2025. Instruct-DeBERTa: A Hybrid Approach for Aspect-based Sentiment Analysis on Textual Reviews. *ICTer*, 18(2):39–50.

Lung-Hao Lee, Liang-Chih Yu, Natalia Loukashevich, Ilseyar Alimova, Alexander Panchenko, Tzu-Mi Lin, Zhe-Yu Xu, Jian-Yu Zhou, Guangmin Zheng, Jin Wang, et al. 2026. Dimabsa: Building multilingual and multidomain datasets for dimensional aspect-based sentiment analysis. *arXiv preprint arXiv:2601.23022*.

Chanika Ruchini Mudalige, Dilini Karunarathna, Isanka Rajapaksha, Nisansa de Silva, Gathika Ratnayaka, Amal Shehan Perera, and Ramesh Pathirana. 2020. SigmaLaw-ABSA: Dataset for Aspect-Based Sentiment Analysis in Legal Opinion Texts. In *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, pages 488–493. IEEE.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35. Association for Computational Linguistics.

Isanka Rajapaksha, Chanika Ruchini Mudalige, Dilini Karunarathna, Nisansa de Silva, Gathika Rathnayaka, and Amal Shehan Perera. 2020. Rule-Based Approach for Party-Based Sentiment Analysis in Legal Opinion Texts. In *2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer)*, pages 284–285. IEEE.

Isanka Rajapaksha, Chanika Ruchini Mudalige, Dilini Karunarathna, Nisansa de Silva, Amal Shehan Perera, and Gathika Ratnayaka. 2021a. Sigmalaw PBSA-A Deep Learning Model for Aspect-Based Sentiment Analysis for the Legal Domain. In *International Conference on Database and Expert Systems Applications*, pages 125–137. Springer.

Isanka Rajapaksha, Chanika Ruchini Mudalige, Dilini Karunarathna, Nisansa de Silva, Gathika Ratnayaka, and Amal Shehan Perera. 2021b. Sigmalaw PBSA-A Deep Learning Approach for Aspect-Based Sentiment Analysis in Legal Opinion Texts. *Journal of Data Intelligence*, 3(1):101–115.

Kevin Scaria, Himanshu Gupta, Siddharth Goyal, Saurabh Sawant, Swaroop Mishra, and Chitta Baral. 2024. InstructABSA: Instruction learning for aspect based sentiment analysis. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pages 720–736. Association for Computational Linguistics.

Vishal Thenuwara and Nisansa de Silva. 2025. Dess: Deberta enhanced syntactic-semantic aspect sentiment triplet extraction. In *International Conference on Computational Collective Intelligence*, pages 185–199. Springer.

Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. 2020. Grid tagging scheme for aspect-oriented fine-grained opinion extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2576–2585. Association for Computational Linguistics.

Lu Xu, Yew Ken Chia, and Lidong Bing. 2021. Learning span-level interactions for aspect sentiment triplet extraction. 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)*, pages 4755–4766. Association for Computational Linguistics.

Liang-Chih Yu, Lung-Hao Lee, et al. 2026. SemEval-2026 task 3: DimABSA – dimensional aspect-based sentiment analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, Vienna, Austria. Association for Computational Linguistics. To appear.

Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, and Wai Lam. 2021. Aspect sentiment quad prediction as paraphrase generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9209–9219. Association for Computational Linguistics.

A DESS Architecture Adaptation Details

The DESS model is adapted from the D2E2S architecture (Thenuwara and de Silva, 2025), originally designed for categorical ASTE. Below we

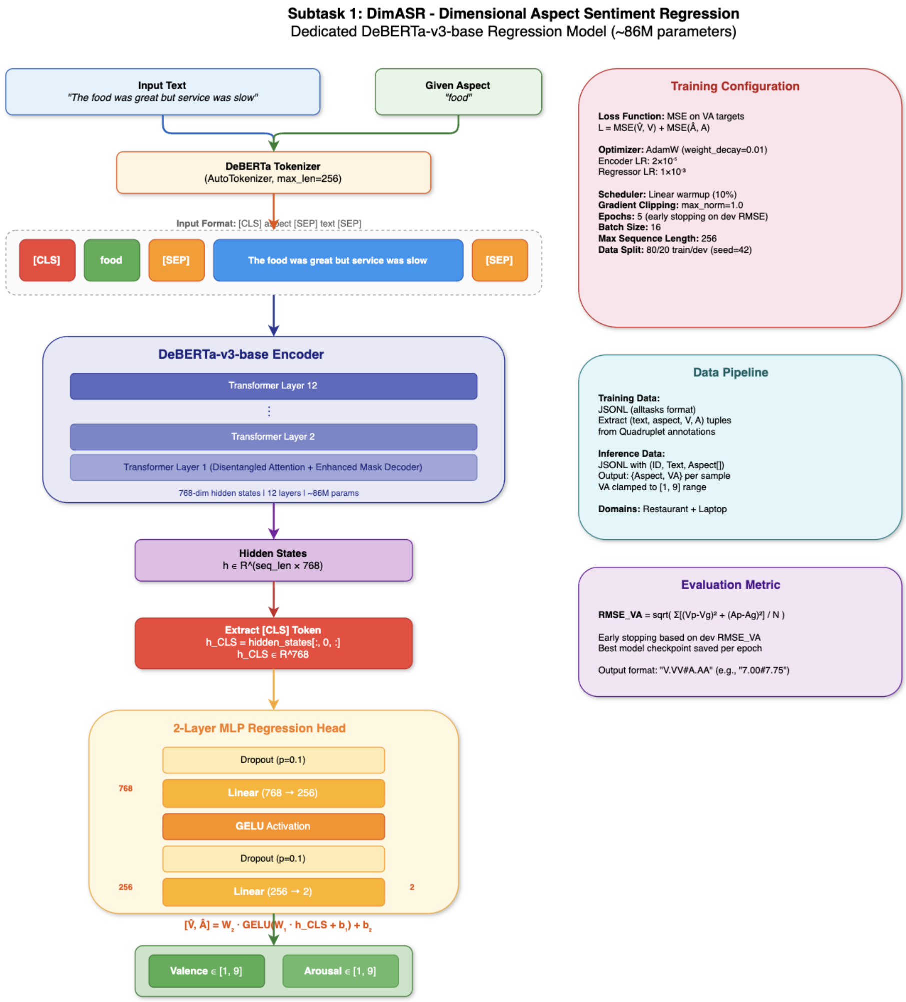

Figure 4: Enlarged: Architecture of the Subtask 1 (DimASR) regression model (same as Figure 1).

detail the specific modifications made for continuous VA regression.

GCN Layer Modifications. The dual-channel GCN layers (Syntactic and Semantic) remain structurally unchanged from D2E2S, as their role is to produce enriched token representations rather than directly predict sentiment. Both channels use 2-layer GCNs with 768-dimensional hidden states. The Syntactic GCN operates on the spaCy-generated dependency parse adjacency matrix, while the Semantic GCN uses multi-head self-attention (4 heads) to learn a soft adjacency matrix capturing implicit semantic relationships.

Output Layer Modifications. The key architectural change is in the sentiment prediction head. In the original D2E2S, this head is a linear classifier that maps a concatenated representation of the aspect span, opinion span, context (CLS token), and size embeddings to one of 3–4 discrete sentiment classes (positive, negative, neutral). For VA regression, we replace this classifier with a two-layer MLP regression head:

$$[\hat{V}, \hat{A}] = W_2 \cdot \mathrm{GELU}(W_1 \cdot [h_{\text{asp}}; h_{\text{opn}}; h_{\text{ctx}}; e_s] + b_1) + b_2 \tag{6}$$

where h_{asp} and h_{opn} are max-pooled span representations of the aspect and opinion entities, h_{ctx} is the CLS token representation, e_s is a concatenation of learned size embeddings for both spans, $W_1 \in \mathbb{R}^{256 \times d}$ and $W_2 \in \mathbb{R}^{2 \times 256}$ are learned weight matrices, and GELU is the activation function. The output is clamped to $[1, 9]$ during inference.

Loss Function Adaptation. The original D2E2S uses cross-entropy loss for the sentiment head. In the DESS adaptation, the sentiment head retains cross-entropy over string-encoded

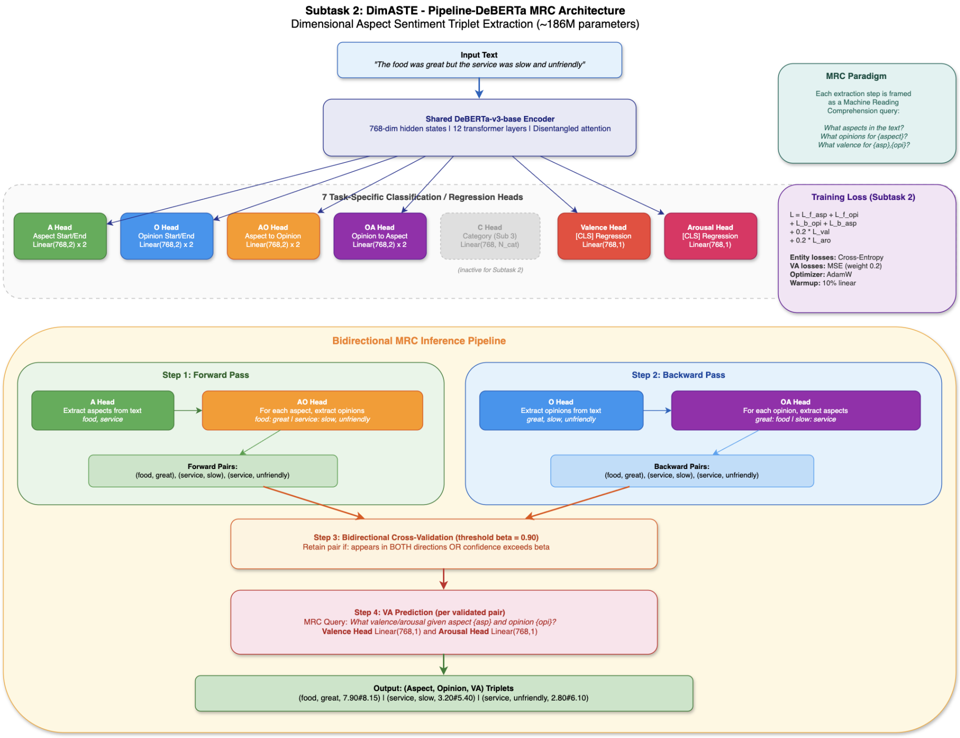

Figure 5: Enlarged: Architecture of the Subtask 2 (DimASTE) Pipeline-DeBERTa system (same as Figure 2).

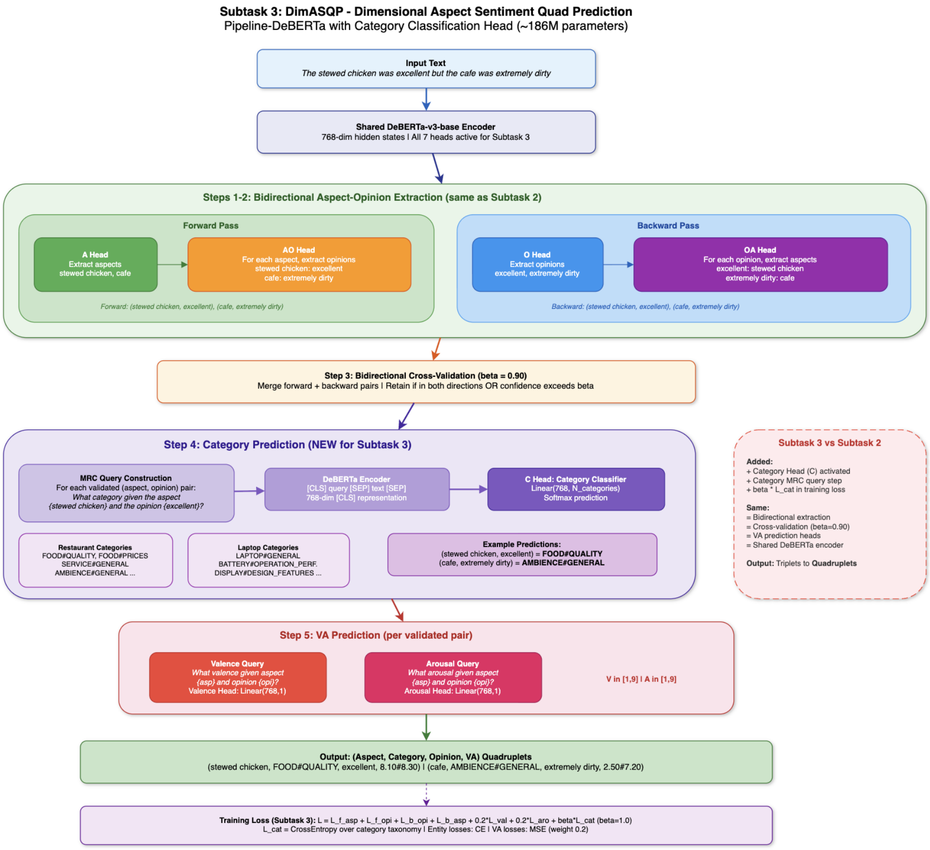

Figure 6: Enlarged: Architecture of the Subtask 3 (DimASQP) quad prediction system (same as Figure 3).

VA labels (e.g., "7.50#7.62" treated as a discrete class), which led to the label space explosion discussed in the main text. In contrast, the Pipeline-DeBERTa system uses MSE loss for VA regression, avoiding this issue.

B Subtask Architecture Summaries

Table 7 provides a structured comparison of the three subtask architectures, including their inputs, outputs, loss functions, and prediction pipelines.

Subtask 1 (DimASR). This subtask receives pre-identified aspects and predicts VA scores. The architecture is a straightforward regression model: a DeBERTa-v3-base encoder processes the concatenated aspect–text input, and the CLS token representation is passed through a two-layer MLP (768→256→2) with GELU activation and dropout (0.1) to produce Valence and Arousal predictions. Figure 4, which is a scaled up representation of Figure 1, shows the Subtask 1 architecture.

Subtask 2 (DimASTE). This subtask requires extracting (Aspect, Opinion, VA) triplets from raw text. The Pipeline-DeBERTa system uses a shared encoder feeding seven task-specific heads. Bidirectional MRC extraction identifies aspect–opinion pairs: a forward pass extracts opinions given aspects, and a backward pass extracts aspects given opinions. Cross-validation with a confidence threshold (β=0.90) filters spurious pairings. Validated pairs are then scored by dedicated Valence and Arousal regression heads. Figure 5, which is a scaled up representation of Figure 2, shows the Subtask 2 architecture.

Subtask 3 (DimASQP). This subtask extends Subtask 2 by additionally predicting aspect categories. The architecture activates a Category classification head that assigns each validated aspect–opinion pair to a domain-specific taxonomy (e.g., `FOOD#QUALITY`, `SERVICE#GENERAL`) via an MRC query. Figure 6, which is a scaled up representation of Figure 3, shows the Subtask 3 architecture.

C Data

We use the DimABSA 2026 English dataset (Lee et al., 2026) for Track A. Table 8 summarizes the dataset statistics.

D Preprocessing

For the DESS model, we convert the DimABSA JSONL data to the DESS JSON format using spaCy (v4.0) for tokenization, POS tagging, and dependency parsing. Entity spans are aligned to token boundaries. The combined restaurant and laptop training data yields 3,712 successfully converted samples (15 samples lost due to tokenization alignment failures). The following shows the conversion from DimABSA JSONL format to DESS JSON format:

DimABSA Input:


```
{"ID": "R001",
 "Text": "The food was great.",
 "Triplet": [
   {"Aspect": "food",
    "Opinion": "great",
    "VA": "7.50#7.62"}]}
```


DESS Output:


```
{"tokens": ["The","food","was","great"],
 "entities": [
   {"type":"target","start":1,"end":2},
   {"type":"opinion","start":3,"end":4}],
 "sentiments": [
   {"type":"7.50#7.62",
    "head":0,"tail":1}],
 "pos": ["DET","NOUN","AUX","ADJ"],
 "dependency": [[0,1],[1,2],...]}
```


E Hyperparameters

Table 9 lists the key hyperparameters for the DESS model. We train on Kaggle T4 GPUs.

F Error Analysis

We observe several recurring error patterns. For the DESS model, the primary challenge is the mismatch between the architecture's original design for categorical classification and the continuous regression required for VA prediction. Encoding VA scores as strings (e.g., "7.50#7.62") causes each unique VA pair to become a separate "class," resulting in thousands of unique types versus the original 3–4 sentiment categories. This label space explosion fundamentally limits the span-based model's ability to learn meaningful VA representations.

For the Pipeline-DESS approach, error propagation across pipeline stages is the main concern: incorrect aspect or opinion extraction in early stages cascades into incorrect VA predictions. Additionally, arousal prediction proves consistently harder than valence across all subtasks, as reflected by the negative R^2 for arousal in Subtask 1. This may stem from arousal being a less

	Model	Input / Output	Loss Function	Pipeline
Subtask 1 (DimASR)	DeBERTa-v3-base + MLP regression head	Input: `[CLS] aspect [SEP] text [SEP]` Output: (Valence, Arousal) $\in [1, 9]^2$	MSE loss on VA predictions	Single forward pass: encode text–aspect pair $\rightarrow$ CLS representation $\rightarrow$ MLP $\rightarrow$ VA scores
Subtask 2 (DimASTE)	Pipeline-DeBERTa with 7 task-specific heads	Input: raw text Output: (Aspect, Opinion, VA) triplets	Cross-entropy for extraction heads + MSE for VA heads	Bidirectional MRC extraction (A$\rightarrow$AO, O$\rightarrow$OA) $\rightarrow$ cross-validation (β=0.90) $\rightarrow$ VA regression
Subtask 3 (DimASQP)	Pipeline-DeBERTa + Category head	Input: raw text Output: (Aspect, Category, Opinion, VA) quadruplets	Subtask 2 loss + $\beta \cdot \mathcal{L}_{\text{cat}}$	Subtask 2 pipeline + category classification via MRC query

Table 7: Structured comparison of architectures across the three subtasks.

Statistic	**Train**	**Test**
Samples	3,727	400
Domains	Restaurant + Laptop	
Avg. entities/sample	3.06	–
Avg. sentiments/sample	1.53	–

Table 8: Dataset statistics for the combined English training and test sets.

Parameter	**Value**
Encoder	DeBERTa-v3-base
Encoder dim	768
BiLSTM layers	2
BiLSTM hidden dim	384
GCN layers	2
Max span size	8
Batch size	4
Epochs	10
Learning rate	5×10^{-5}
LR warmup	0.2
Weight decay	0.01
Max grad norm	1.0
Dropout	0.1
GCN dropout	0.2
Neg. entity samples	100
Neg. triple samples	100
Seed	42

Table 9: Hyperparameters for the DESS model.

lexically grounded dimension than valence in review text.

G Discussion on LLM-Based Approaches

A limitation of our work is the absence of comparison with modern LLMs based approaches. Recent work has shown that LLMs such as GPT-4, Llama-3, and Qwen can achieve competitive performance on ABSA tasks through zero-shot and few-shot prompting, without task-specific fine-tuning. The official DimABSA 2026 baselines included LLM-based systems (Qwen, KIMI), which we did not benchmark against.

Given that our work focused on English data, zero-shot or few-shot prompting with accessible LLMs (e.g., GPT-4o-mini, Llama-3-8B) would have been a feasible and informative baseline. Such approaches could potentially bypass the label space explosion issue entirely by generating VA scores as continuous numbers in natural language output, rather than requiring architectural modifications to classification models.

Our decision to focus on span-based architectures was motivated by the desire to explore whether established ASTE models could be adapted to dimensional sentiment. However, we acknowledge that a brief LLM comparison would have provided valuable context for interpreting our results. Future work should include LLM baselines to better understand the trade-offs between fine-tuned specialized models and general-purpose LLMs for dimensional ABSA.