Instruct-DeBERTa: A Hybrid Approach for Aspect-based Sentiment Analysis on Textual Reviews

Dineth Jayakody[1], A V A Malkith[1], Koshila Isuranda[1], Vishal Thenuwara[2], Nisansa de Silva[2], Sachintha Rajith Ponnamperuma[3], G G N Sandamali[1], K L K Sudheera[1]
[1]*Department of Electrical and Information Engineering, University of Ruhuna*
[2]*Department of Computer Science & Engineering, University of Moratuwa*
[3]*Emojot Inc.*

***Abstract*—Aspect-based Sentiment Analysis (ABSA) is a critical task in Natural Language Processing (NLP) that focuses on extracting sentiments related to specific aspects within a text, offering deep insights into customer opinions. Traditional sentiment analysis methods, while useful for determining overall sentiment, often miss the implicit opinions about particular product or service features. This paper presents a comprehensive review of the evolution of ABSA methodologies, from lexicon-based approaches to machine learning and deep learning techniques. We emphasize the recent advancements in Transformer-based models, particularly Bidirectional Encoder Representations from Transformers (BERT) and its variants, which have set new benchmarks in ABSA tasks. We focused on finetuning Llama and Mistral models, building hybrid models using the SetFit framework, and developing our own model by exploiting the strengths of state-of-the-art (SOTA) Transformer-based models for aspect term extraction (ATE) and aspect sentiment classification (ASC). O ur hybrid model `Instruct - DeBERTa` uses SOTA `InstructABSA` for aspect extraction and `DeBERTa-V3-baseabsa-V1` for aspect sentiment classification. W e u tilize d atasets f rom d ifferent domains to evaluate our model's performance. Our experiments indicate that the proposed hybrid model significantly improves the accuracy and reliability of sentiment analysis across all experimented domains. As per our findings, o ur h ybrid model `Instruct - DeBERTa` is the best-performing model for the joint task of ATE and ASC for both SemEval restaurant 2014 and SemEval laptop 2014 datasets separately. By addressing the limitations of existing methodologies, our approach provides a robust solution for understanding detailed consumer feedback, thus offering valuable insights for businesses aiming to enhance customer satisfaction and product development.**

***Index Terms*—Aspect-Based Sentiment Analysis, Aspect Extraction, DeBERTaV3, Hybrid Model, InstructABSA, Natural Language Processing, Sentiment Classification, Textual Reviews**

Correspondence: Dineth Jayakody (E-mail: dinethjkd00@gmail.com)
Received: 16-06-2024 **Revised:** 12-08-2024 **Accepted:** 09-09-2024
Dineth Jayakody, Malkith Amanda, Koshila Isuranda, Nadeesha Sandamali Gammana Guruge, Kushan Sudheera Kalupahana Liyanage are from Faculty of Engineering, University of Ruhuna (dinethjkd00@gmail.com, malkithamanda@gmail.com, koshi.isuranda@gmail.com, nadeesha@eie.ruh.ac.lk, kushan@eie.ruh.ac.lk), Vishal Thenuwara, Nisansa de Silva are from Department of Computer Science and Engineering, University of Moratuwa (vishal.thenuwara.mail@gmail.com, nisansadds@cse.mrt.ac.lk), and Sachintha Rajith is from Emojot Inc (sachintha@emojot.com).
DOI: https://doi.org/10.4038/icter.v18i2.7290
The 2025 Special Issue contains the full papers of the abstracts published at the 24th ICTer International Conference.

I. Introduction

Aspect-Based Sentiment Analysis (ABSA) has become an essential technique in Natural Language Processing (NLP) for extracting fine-grained opinions from textual data. It focuses on identifying sentiment towards specific aspects within a text, providing a detailed understanding of customer feedback and reviews. Traditional sentiment analysis techniques, while effective at determining overall sentiment, often fail to capture the nuanced opinions that consumers express about particular features or attributes of a product or service.

Over the years, ABSA methodologies have evolved significantly. Early approaches primarily relied on lexicon-based methods, which used predefined dictionaries of sentiment-laden words to infer polarity. These methods; however, struggled with context and ambiguity. The advent of machine learning approaches introduced more sophisticated techniques, including supervised learning models that could be trained on annotated datasets. Despite their advancements, these models required substantial manual effort for feature extraction and were often domain-specific.

The breakthrough in deep learning, particularly with the development of recurrent neural networks (RNNs), Long Short-Term Memory networks (LSTMs), and Convolutional Neural Networks (CNNs), marked a significant improvement in sentiment analysis. These models could automatically learn features from text, capturing context and sequential dependencies more effectively than traditional methods. LSTM and CNN-based models became popular for their ability to handle long-range dependencies and local features, respectively. However, these models still had limitations, especially in understanding long-term dependencies and complex syntactic structures.

The introduction of Transformer architectures, especially `BERT`, revolutionized the field by leveraging attention mechanisms to capture contextual relationships in both directions of a sentence. `BERT` and its variants, such as `RoBERTa` and `DeBERTa`, have set new benchmarks in various NLP tasks, including ABSA. These models have demonstrated superior performance in aspect extraction and sentiment classification tasks due to their ability to understand complex language patterns and relationships.

In our study, we focus on developing a hybrid model that

leverages the strengths of the latest Transformer-based models for ABSA. We aim to address the limitations of existing approaches by combining aspect extraction and sentiment classification into a unified framework. Our approach utilizes datasets from the hospitality domain, including SemEval 2014 (Res-14), 2015 (Res-15), and 2016 (Res-16) restaurant reviews, and extends to the laptop domain with the SemEval 2014 laptop dataset (Lap-14). By evaluating these models on imbalanced datasets using the F1 metric, we ensure a balanced and comprehensive assessment of performance.

Through our literature review, we have identified key methodologies and their respective accuracies, guiding the design of our hybrid model. We focus on models that excel in aspect term extraction (ATE) and aspect sentiment classification (ASC), aiming to develop a model that builds on the successes of past methodologies while innovating in areas where existing methods may fall short. Our goal is to enhance the accuracy and reliability of sentiment analysis in our application domain, ultimately providing a robust solution for understanding consumer feedback.

II. Literature Review

Our literature review systematically investigates various models that have demonstrated efficacy in ATE and ASC. Table II provides a comprehensive summary of the methodologies employed for ASC over the years considering four benchmark datasets while Table III provides an extension for the above table considering only SemEval 2014. Also, Table IV provides a summary for AE considering SemEwval 2014. Notably, this table exclusively considers models that do not incorporate LLMs. Table V lists models that perform the joint task, of the ATE and ASC tasks using by a single model. The model is fed only in with the relevant sentences or the reviews. Then, the model identifies the aspects by itself and classifies the polarities to the aspects that have been identified. Then, the F1 score of the whole process is reported. For a sentence S_i, the ATE, ASC, and the joint task can be visualized as below.

S_i: The price was high, but the restaurant was breathtaking.

TABLE I: Overview of tasks for ABSA with the sample sentence S_i

Task	Input	Output
Aspect Term Extraction (ATE)	S_i	price, restaurant
Aspect Sentiment Classification (ASC)	$S_i + price, S_i + restaurant$	Negative, Positive
Joint Task (ATE + ASC)	S_i	(price, Negative), (restaurant, Positive)

A. Models for a Single Task (ATE or ASC)

1) ***LSTM Based Models****:* Standard RNNs suffer from significant limitations, primarily the vanishing gradient and exploding gradient problems. To address these limitations, the LSTM network was developed by Hochreiter and Schmidhuber [1]. To effectively utilize aspect information, Wang et al. [2] proposed a model called LSTM with Aspect Embedding (AE-LSTM). However, to further leverage aspect information, Wang et al. [2] developed an enhanced model called Attention-based LSTM with Aspect Embedding (ATAE-LSTM).

Li et al. [3] proposed Target-Specific Transformation Networks (TNet), a new architecture designed to improve target sentiment classification by effectively handling multiple targets and extracting relevant features without introducing noise. TNet introduces a novel Target-Specific Transformation (TST) [3] component for generating target-specific word representations. The models LSTM-FC-CNN-LF and LSTM-FC-CNN-AS were built by Li et al. [3] incorporating a fully connected layer and context-preserving mechanisms. These models performed better with F1 scores of 70.60% and 70.23% for LSTM-FC-CNN-LF, 70.72% and 70.06% for LSTM-FC-CNN-AS for Lap-14 and Res-14, respectively.

2) ***GloVe Based Models****:* GloVe (Global Vectors for Word Representation) is an algorithm that generates word embeddings by aggregating word co-occurrence statistics from a corpus. It is important because it captures both local and global statistical information of words, enhancing the performance of natural language processing tasks. ASGCN introduced by Zhang et al. [4] proposed a novel aspect-specific sentiment classification framework while DualGCN by Li et al. [5] proposed a dual graph convolutional network model that considers the complementarity of syntax structures and semantic correlations simultaneously. Zhong et al. [6] proposed a new model named KGAN which uses a knowledge graph augmented network, which aims to effectively incorporate external knowledge with explicitly syntactic and contextual information. While all these models individually had their own importance, merging them with Glove as an embedding method, researchers were able to capture both local and global information to achieve higher F1 scores like 84.46% with Res-14. However, UIKA (Unified Instance and Knowledge Alignment Pretraining) by Liu et al. [7], which introduces a unified alignment pretraining framework into the vanilla pretrain-finetune pipeline, incorporates both instance and knowledge level alignments. It reported a higher F1 score of 85.53% with KGAN for Res-14. Furthermore, KGAN+UIKA achieved higher F1 scores with BERT compared to GloVE.

3) ***BERT Based Models****:* The original BERT model was introduced by Devlin et al. [8]. BERT's training process involves two main steps: pre-training and fine-tuning. BERT-DK, introduced by Zhao [9], integrates domain-specific knowledge to improve ABSA performance. By incorporating domain-specific information, BERT-DK achieved F1 scores of 77.02% and 83.55% for aspect extraction on the Res-14 and Lap-14 datasets, respectively. Similarly, BERT-SPC, developed by Song et al. [10], employs a Sentence Pair Classification framework to better understand the context of aspect-specific sentences.

Innovative approaches such as BERT-MRC, proposed by Zhao et al. [11], frame ABSA tasks as machine reading comprehension problems, while Xu et al. [12] introduced BERT-PT, which involves pre-training BERT on domain-specific data followed by fine-tuning. BAT which stands for BERT with Adversarial Training, introduced by Karimi et al. [13], enhances ABSA by generating adversarial examples during training.

Cutting-edge models like RGAT-BERT, DualGCN-BERT, TF-BERT, and dotGCN-BERT have further improved ABSA performance. RGAT-BERT, proposed by Bai et al. [14], uses relational graph attention networks to improve aspect extraction and sentiment classification abilities. In addition, DualGCN-BERT introduced by Li et al. [5], uses dual graph convolutional networks to handle both aspect extraction and sentiment classification. TF-BERT, developed by Zhang et al. [15], uses task-specific fine-tuning strategies to improve ABSA performance. In contrast to that DotGCN-BERT, proposed by Chen et al. [16], uses dot-product based graph convolutional networks to improve ABSA performance. Furthermore, keeping another step forward, DualGCN and KGAN have used BERT as the embedding methodology in order to achieve higher F1 scores. Table II shows that KGAN+UIKA with BERT reported a F1-score of 92.89%, which is the highest for the Res-16 dataset. But still, we decided to move forward with DeBERTa-V3-base-absa-v1 since it provides a higher F1 score in all datasets compared to others.

4) ***RoBERTa Based Models****:* RoBERTa [17] is an advanced language model that is built upon the foundational work of BERT. The SARL-RoBERTa model, introduced by Wang et al. [18] used span-based dependency modeling to align opinion candidates with aspects, and used an adversarial learning strategy to reduce sentiment bias in aspect embeddings. Among the compared RoBERTa based models, SARL-RoBERTa performs the best, achieving a F1-score of 82.44% and 82.97% for Res-14 and Lap-14, respectively.

However, models such as ASGCN-RoBERTa, RGAT-RoBERTa, PWCN-RoBERTa, and RoBERTa+MLP benefited by combining RoBERTa with various specialized architectures, as demonstrated by Dai et al. [19]. A strong performance is achieved by ASGCN-RoBERTa, which combines an aspect-specific graph convolutional network with dependency tree syntactic information. With a relational graph attention network integrated to collect relational information between words, RGAT-RoBERTa performs admirably. RoBERTa+MLP integrates a multi-layer perceptron with RoBERTa. It highlights the flexibility of combining RoBERTa's embeddings with simple classifiers.

Task-oriented syntactic information is well captured by pure RoBERTa based models, particularly by the fine-tuned variations (FT-RoBERTa). Research conducted by Dai et al. [19] demonstrates that FT-RoBERTa achieves a 1.56% improvement in the F1-score over standard RoBERTa induced trees, and performs better than parser-provided trees.

5) ***DeBERTa Based Models****:* DeBERTa [20] introduces a disentangled attention mechanism, which utilizes two separate vectors for each word to represent its content and position independently. Also, the model incorporates an enhanced mask decoder in its pre-training phase based on masked language modeling (MLM).

Improving from the vanilla DeBERTa model a new model, named DeBERTaV3 was introduced by He et al. [21]. By further fine tuning the model to improve its performance, Yang et al. [22] developed the DeBERTaV3-base-absa-V1 model. This was trained using Lap-14, Res-14, Res-16, and six more datasets counting up to 30k+ ABSA examples. The accuracy of this model showed an improvement of 9.35% and 10.87% for the ASC task of Res-14 and Lap-14 datasets respectively compared to the original DeBERTaV3 model.

In their independent investigations, Marcacini and Silva [23] as well as Yang and Li [24] explored the utilization of DeBERTa based models, introducing ABSA-DeBERTa and LSA-X-DeBERTa, respectively. Marcacini and Silva [23] explored disentangled learning as a method to improve BERT-based representations specifically for ABSA. On the other hand, Yang and Li [24] introduced a novel perspective in ASC by emphasizing the significance of aspect sentiment coherency. Their study revealed that neighbouring aspects usually share similar sentiments, which is known as "aspect sentiment coherency." To address this, they proposed a local sentiment aggregation paradigm (LSA) to effectively model fine-grained sentiment coherency. Consequently, the LSA-X-DeBERTa model introduced by Yang and Li [24] achieved a F1-score of 87.02% for Res-14, and 84.41% for Lap-14 under the sentiment classification task.

6) ***Other Models****:* Concerning ASC and ATE, in particular, LCF-ATEPC-CDM proposed by Yang et al. [25] and InstructABSA proposed by Scaria et al. [26] standout for their strong performances. InstructABSA, utilizing a novel instruction learning paradigm, showed exceptional abilities in obtaining pertinent aspects from the text, attaining F1 scores exceeding 92% for aspect extraction on both the Res-14 and Lap-14 datasets. Concerning ATE, LCF-ATEPC-CDM, which also employs a local context focus technique, performs fairly well. In sentiment polarity classification, InstructABSA also excels with F1 scores of 85.17% on Res-14 and 81.56% on Lap-14, outperforming many other models. The LSAT model proposed by Yang and Li [27], with its focus on aspect sentiment coherency through a local sentiment aggregation paradigm, shows impressive results, achieving a F1 score of 90.86% on Res-14. The efficacy of the BART-ABSA model in a comprehensive approach to ABSA has also been demonstrated by Yan et al. [28], which combines all ABSA subtasks into a single generative formulation.

B. Joint Task Models

Here we talk about the models that perform the joint task of the ATE and ASC tasks together by a single model. The model is fed only with the relevant sentences or the reviews. Then, the model identifies the aspects by itself and classifies the polarities to the aspects that have been identified. Then, the F1 score of the whole process is reported.

Table V presents the leading joint task models identified through our research. The InstructABSA and Grace models, previously described as single task models capable of performing both ATE and ASC tasks separately, also excel in the joint task. These models report the highest F1 scores for the joint task, achieving over 75% for both datasets, indicating their accuracy and robustness across different domains.

RACL-BERT, introduced by Chen and Qian [29], is a notable ABSA (Aspect-Based Sentiment Analysis) model that utilizes the BERT-Large model to address three subtasks simultaneously: identifying aspects, detecting sentiment words,

and classifying overall sentiment. Through multitasking and relation propagation, RACL-BERT enhances sentiment analysis accuracy. Similarly, SPAN, introduced by Hu et al. [30], employs a novel approach by focusing on key opinion points rather than tagging each word. Both RACL-BERT and SPAN achieved reasonable F1 scores, but were outperformed by InstructABSA and GRACE.

The E2E-TBSA model, proposed by Li et al. [31], addresses both ATE and ASC tasks in a single step, using a collapsed approach that combines these tasks into a unified process. Similarly, BERT-E2E-ABSA, introduced by Li et al. [32], is based on BERT models and follows the same principles. These models achieved F1 scores in the 60%-70% range, but did not outperform the leading models.

DOER, introduced by Luo et al. [33], uses a cross-shared RNN framework to generate aspect term-polarity pairs simultaneously. IMN, introduced by He et al. [34], employs an interactive architecture with multi-task learning for end-to-end ABSA tasks, including aspect term and opinion term extraction as well as aspect-level sentiment classification.

C. Selection Criteria: Dataset

After a thorough review, the following criteria were established for dataset selection:

- Relevance to ABSA: The datasets must be specifically designed for or widely used in aspect-based sentiment analysis ensuring granularity.
- Diversity of aspects and sentiments: The selected datasets should cover a wide range of aspects and sentiments ensuring generalizability.
- Quality of annotations: High-quality, manually annotated datasets are preferred to ensure the accuracy.
- Availability and accessibility: Publicly available datasets with accessibility are chosen to facilitate reproducibility.

Based on these criteria, the SemEval datasets from the years 2014, 2015, and 2016 were selected.

III. Methodology

In this section, we examine different approaches we tested in order to find the most accurate and robust solution. These approaches are listed below.

1) Fine-Tuning LLaMA 2-7B with Quantized Low Rank Adaptation (QLoRA)
2) Fine-Tuning Mistral-7B with Quantized Low Rank Adaptation (QLoRA)
3) ASGCN+UIKA+Glove for Sentiment Polarity
4) SSGCN+Glove for Sentiment Polarity
5) Span-ASTE+BERT for Aspect Extraction
6) SETFIT for efficient few-shot fine-tuning of Sentence Transformers
7) Instruct-DeBERTa (Proposed Model)

A. LLaMA 2-7B with QLoRA

This study incorporates an LLM-based analysis, building on the work done by Jayakody et al. [64] with LLaMA 2 [65], which highlighted significant advancements in natural language processing. Due to the computational challenges of traditional fine-tuning methods, the more efficient QLoRA [66] technique was employed. QLoRA enables large models to be fine-tuned on consumer hardware through four-bit quantization, a method thoroughly detailed in Jayakody et al. [64].

For this analysis, a 4-bit quantization with NF4 configuration was applied using BitsAndBytes[1]. Supervised Fine-Tuning (SFT) within the RLHF framework, as discussed by Dettmers et al. [66], was also utilized to enhance model performance. The fine-tuned model was tested with the Transformers text generation pipeline.

B. Mistral-7B with QLoRA

Jiang et al. [67] introduced Mistral 7B, a 7-billion-parameter language model designed for superior performance and efficiency. Mistral 7B surpasses the best open 13B model (Llama 2) across all evaluated benchmarks and the best released 34B model (Llama 1) in reasoning, mathematics, and code generation.

Mistral 7B leverages grouped-query attention (GQA) and sliding window attention (SWA). GQA significantly accelerates inference speed and reduces memory requirements during decoding, allowing for higher batch sizes and, consequently, higher throughput—crucial for real-time applications. Additionally, SWA is designed to handle longer sequences more effectively at a reduced computational cost, addressing a common limitation in LLMs. These attention mechanisms collectively enhance the performance and efficiency of Mistral 7B.

The Mistral 7B model was fine-tuned to perform ABSA on the Lap-14 and Res-14 datasets. We fine-tuned the base model separately for these two datasets and evaluated them separately. This involved customizing the model to better understand and analyze sentiment related to specific aspects within the review texts. The Mistral 7B model was selected from the Hugging Face hub. The mentioned datasets were processed using Pandas to ensure it was in a prompt-compatible format.

To enable efficient training, 4-bit precision loading was configured using BitsAndBytesConfig. We set float16 as the data type for the 4-bit base model, nf4 as quantization type, and nested quantization was disabled to simplify the training process. The tokenizer was loaded and configured to handle padding appropriately. The base model was then loaded with the quantization configuration, ensuring it was prepared for low-bit precision training.

In our fine-tuned models we set the Attention dimension to 64, the Scaling parameter to 64, and the Dropout probability: 0.1 for efficient Low-Rank Adaptation (LoRA) parameters. Then, the fine-tuning was conducted using the SFTTrainer with the defined training arguments. The dataset was loaded, and the model underwent supervised fine-tuning, adjusting to the specific requirements of ABSA on the mentioned datasets. As the final stage the post-training, the model was saved and reloaded in FP16 precision. The LoRA weights were merged back into the base model to create a final, streamlined

[1] https://github.com/TimDettmers/bitsandbytes

TABLE II: Accuracy (A) and F1 scores of models evaluated on the SemEval 2014 [35] both Restaurand and Laptop domains,2015 and 2016 Restaurant benchmark considering Sentiment Polarity. Note*: The F1-scores for the DeBERTa-v3-base-absa-v1 model were calculated by us separately.

Model	Accuracy and F1 Score (%)							
	Res-14		Lap-14		Res-15		Res-16	
	A	F1	A	F1	A	F1	A	F1
MCRF-SA [36]	82.86	73.78	77.64	74.23	80.82	61.59	89.51	75.92
IAN [37]	79.26	70.09	72.05	67.38	78.54	52.65	84.74	55.21
ASCNN [38]	81.73	73.10	72.62	66.72	78.47	58.90	87.39	64.56
ASGCN-DG [38]	80.77	72.02	75.55	71.05	79.89	61.89	61.89	67.48
PRET+MULT [39]	79.11	69.73	71.15	67.46	81.30	68.74	85.58	69.76
MemNet [40]	79.61	69.64	70.64	65.17	77.31	58.28	85.44	65.99
ASGCN-DT-GloVe [4]	80.86	72.19	75.55	71.05	79.34	60.78	88.69	66.64
DualGCN-GloVe [5]	84.27	78.08	78.48	74.74	79.11	62.25	87.80	70.34
DualGCN-GloVe+UIKA [7]	85.19	79.05	78.89	75.14	81.16	65.26	88.91	74.25
KGAN-GloVe [6]	84.46	77.47	78.91	75.21	83.09	67.90	89.78	74.58
KGAN-GloVe+UIKA [7]	85.53	78.00	79.31	75.53	83.89	68.52	90.92	75.74
DeBERTaV3 [20, 21]	—	83.06	—	79.45	—	73.76	—	73.59
DeBERTa-V3-base-absa-v1* [22, 41]	—	**90.94**	—	**90.32**	—	**89.55**	—	**84.91**
LSA-X-DeBERTa [24]	90.98	87.02	86.46	84.41	91.85	81.29	95.61	84.87
SK-GCN-BERT [42]	83.48	75.19	79.00	75.57	83.20	66.78	87.19	72.02
RGAT-BERT [14]	86.68	80.92	80.94	78.20	83.64	66.18	90.16	71.13
RGAT-BERT+UIKA [7]	87.25	81.98	82.03	78.83	86.40	68.11	91.87	75.28
DGEDT-BERT [43]	86.30	80.00	79.80	75.60	84.00	71.00	91.90	79.00
DualGCN-BERT [5]	87.13	81.16	81.80	78.10	84.25	69.54	89.22	72.40
DualGCN-BERT+UIKA [7]	87.90	81.97	82.43	78.71	86.82	69.80	90.81	73.13
KGAN-BERT [6]	87.15	82.05	82.66	78.98	86.21	74.20	92.34	81.31
KGAN-BERT+UIKA [7]	87.92	82.82	83.21	79.57	87.43	75.12	92.89	82.43
DualGCN-BERT [5]	87.13	81.16	81.80	78.10	84.25	69.54	89.22	72.40
dotGCN-BERT [16]	86.16	80.49	81.03	78.10	85.24	72.74	93.18	82.32
TGCN-BERT [5]	86.16	79.95	80.88	77.03	85.26	71.69	92.32	77.29
Sentic GCN-BERT [44]	87.31	81.09	81.01	77.96	85.32	71.28	91.97	79.56
SARL-RoBERTa-large [18]	90.45	85.34	85.74	82.97	91.88	78.88	95.76	84.29
TNet-AS [3]	80.69	71.27	76.54	71.75	78.47	59.47	89.07	70.43
LSTM+SynATT+TarRep [45]	79.33	69.25	70.87	66.53	78.03	58.30	**8**3.27	65.76
Sentic-LSTM [46]	79.43	70. 32	70.88	67.19	79.55	60.56	83.01	68.22

version suitable for deployment. We fine-tuned this model using a batch size of 4 for both training and evaluation over 2 epochs, enhancing its capability to extract aspects and identify sentiment polarity.

To test the fine-tuned model, we developed a function to process user input and generate corresponding aspects and sentiments. The function takes an input sentence and utilizes a text generation pipeline where we set the prompt and specific parameters for sampling.

Using Hugging Face libraries like `transformers`, `accelerate`, `peft`, `trl`, and `bitsandbytes`, we successfully fine-tuned and evaluated both the 7B parameter `LLaMA 2` and `Mistral` models on a consumer GPU.

C. SetFit

Few-shot learning has become essential for handling label-scarce scenarios where data annotation is costly and time-consuming. Traditional methods often rely on large-scale pre-trained language models (PLMs), requiring substantial computational resources and specialized infrastructure. Moreover, the need for manually crafted prompts introduces variability and complexity, limiting accessibility.

To address these challenges, Tunstall et al. [68] introduced `SETFIT` (Sentence Transformer Fine-tuning), a prompt-free framework designed for efficient few-shot learning. `SETFIT` eliminates the need for manual prompts and achieves high accuracy with significantly fewer parameters.

As explained in Jayakody et al. [64], `SETFIT` involves two main steps: fine-tuning a pre-trained Sentence Transformer (ST) using a contrastive loss function, and subsequently training a classification head on the fine-tuned ST. This separation of fine-tuning and classification allows `SETFIT` to be both efficient and scalable, making it suitable for various applications with limited labeled data.

D. Instruct-DeBERTa (Proposed Model)

In this study, we developed an aspect-based sentiment analysis pipeline utilizing transformer-based models to automatically extract aspects and analyze sentiments in textual data. The pipeline is composed of two primary stages: aspect extraction and sentiment classification. For these two stages, we utilized the best models for each task that we found through our thorough literature review which is summarized in Table II, Table III and Table IV. When the analysis, it is clear that `InstructABSA` [26] performs the best in all the analyzed datasets irrespective of the domain. For the Res-14 dataset, it recorded an F1 score of 92.10%, which was higher than all the other reported models. It still remained the highest on the Res-15 data set and was only 1.67%, less than the highest recorded accuracy under the Res-16 dataset. But

TABLE III: Accuracy (A) and F1 scores of models evaluated on the SemEval 2014 [35] benchmark considering Sentiment Polarity. Note*: The F1-scores for the DeBERTa-v3-base-absa-v1 model were calculated by us separately.

Model	Res-14		Lap-14	
	A	F1	A	F1
LCF-ATEPC-CDM [25]	90.18	85.88	83.02	79.84
KaGRMN-DSG [47]	87.68	81.98	82.13	79.42
MGAN [32]	**81.49**	71.48	**76.21**	71.42
RAM [48]	80.23	70.80	74.49	71.35
TN [32]	77.91	65.75	70.58	65.34
CNN-ASP [49]	77.82	65.11	72.46	65.31
RGAT-BERT [14]	86.68	80.92	80.94	78.20
DeBERTaV3 [20, 21]	—	83.06	—	79.45
ABSA-DeBERTa[23]	89.46	83.42	82.76	79.36
DeBERTa-V3-base-absa-v1* [22, 41]	—	**90.94**	—	**90.32**
BERT [8]	81.54	71.94	75.29	71.91
BERT-DK [9]	77.02	75.45	83.55	73.72
BERT-SPC [10] [42]	.84.46	76.98	**78.99**	75.03
BERT-MRC [11]	74.21	74.97	81.06	74.10
BERT-PT [12] [13]	84.95	76.96	84.26	75.08
BAT [13]	**79.35**	76.50	85.57	79.24
P-SUM [42]	86.30	79.68	79.55	76.81
H-SUM [42]	86.37	79.67	79.40	76.52
SDGCN-BERT [50]	83.57	76.47	81.35	78.34
BERT-ADA [51]	87.14	80.05	78.60	74.09
TF-BERT [15]	87.09	81.15	81.80	78.46
Dual-MRC [52]		82.04		75.97
DPL-BERT [53]	89.54	84.86	81.96	78.58
SSEGCN-BERT [54]	87.31	81.09	81.01	77.96
GCAE +GLoVe [55]	79.27	67.66	73.56	67.84
TransCap [56]	79.29	70.85	73.87	70.10
ASGCN-RoBERTa [19]	86.87	80.59	83.33	80.32
RGAT-RoBERTa [19]	87.52	81.29	83.33	79.95
PWCN-RoBERTa [19]	87.35	80.85	84.01	81.08
RoBERTa+MLP [19]	87.37	80.96	83.78	80.73
MN [57]	75.30	64.34	68.90	62.89
MN(+AS) [58]	78.75	69.15	70.53	65.24
TNet [3]	80.69	71.27	76.54	71.75
TNet-LF [3]	80.79	70.84	76.01	71.47
TNet-ATT [58]	81.53	72.90	77.62	73.84
TNet-AS [3]	80.69	71.27	76.54	71.75
TNet-ATT(+AS) [58]	81.53	72.90	77.62	73.84
AE-LSTM [2]	76.25	64.32	68.97	62.50
ATAE-LSTM [2]	77.23	64.95	68.65	62.45
TD-LSTM [59]	75.63	64.16	68.18	62.28
BILSTM-ATT-G [60]	80.38	70.78	74.37	69.90
LSTM-ATT-CNN [3]	78.95	68.71	73.37	68.03
LSTM-FC-CNN-LF [3]	80.41	70.23	75.59	70.60
LSTM-FC-CNN-AS [3]	80.23	70.06	75.78	70.72
AEN-BERT [10]	83.12	73.76	79.93	76.31
AEN-GloVe [10]	80.98	72.14	73.51	69.04

TABLE IV: F1 scores of models evaluated on the SemEval 2014 [35] benchmark for Aspect Extraction.

Model	Res-14	Lap-14
InstructABSA [26]	**92.10**	**92.30**
LCF-ATEPC-CDW [25]	88.65	81.61
LCF-ATEPC-CDM [25]	89.78	85.88
LCF-ATEPC-Fusion [25]	89.02	83.82
GPT2(med) [61]	75.94	82.04
GRACE [62]	85.45	87.93
BART-ABSA [28]	87.07	83.52
BERT-DK [9]	77.02	83.55
BERT-MRC [11]	74.21	81.06
IMN-BERT [11]	84.06	77.55
RACL-BERT [11]	86.38	81.79
SPAN-BERT [11]	86.71	74.97
Span-ASTE [63]	79.36	67.02

TABLE V: F1 scores of different models which perform the joint task

Model	Lap-14	Res-14
InstructABSA [26]	79.34	79.47
GRACE [62]	70.71	77.26
SPAN [30]	68.06	74.92
RACL-BERT [29]	63.40	75.42
BERT-E2E-ABSA [32]	61.12	74.72
DOER [33]	60.35	72.78
IMN [34]	58.37	69.54
E2E-TBSA [31]	57.90	69.80

P-SUM [42] which reported the highest F1 score for the Res-16 dataset performed significantly less than InstructABSA in the previous datasets. Hence, InstructABSA still remained the best option for aspect extraction. Moreover, for the Lap-14 dataset, InstructABSA topped the charts again showcasing the models' adaptability and robustness regardless of the relevant domain. Hence InstructABSA was selected as the best performing model for aspect extraction. When looking at the performances on the sentiment polarity task, DeBERTa-V3-baseabsa-V1 [22, 41] is the best overall performing model across all the datasets. It showcases an accuracy of 90.94% for the Res-14 dataset which is the highest overall. It shows the same promising results in the Res-15 and Res-16 datasets. In addition, to that DeBERTa-V3-baseabsa-V1 also shows the adaptability of the model by recording the highest accuracy for the Lap-14 dataset as well which falls into a completely different domain.

Since we identified the best overall performing models for individual tasks of aspect extraction and sentiment polarity detection, we tried to exploit the performances of these models and build up a hybrid model that performs the joint task of aspect extraction and sentiment polarity detection by itself. Then, we created a **novel hybrid model** utilizing these models to build up a model which as per our research gives the best performance as a pipelined hybrid model, which in fact makes this the SOTA model for the pipelined aspect extraction and sentiment polarity classification task, also known as the joint task for ABSA.

Figure 1 shows the proposed model of our study. The model structure used for **ATE** is **InstructABSA**, while the model used for **ASC** is **DeBERTa-V3-baseabsa-V1**. The collective model is named as **Instruct-DeBERTa**, which stands for **Instruct**ABSA for aspect term extraction and **DeBERTa**-V3-baseabsa-V1 for aspect sentiment classification. When looking on to Figure 1 it shows how these two independent models are being pipelined to build up a single joint task model.

The algorithm for our proposed model can be presented as below for further clarification.

Sets:

- X: Review represented as a word sequence ($X = \{x_1, x_2, \ldots, x_n\}$)
- A: Set of extracted aspect terms ($A \subseteq X$)
- S: Set of sentiment labels for aspect terms ($S \subseteq \{positive, negative, neutral\}$)

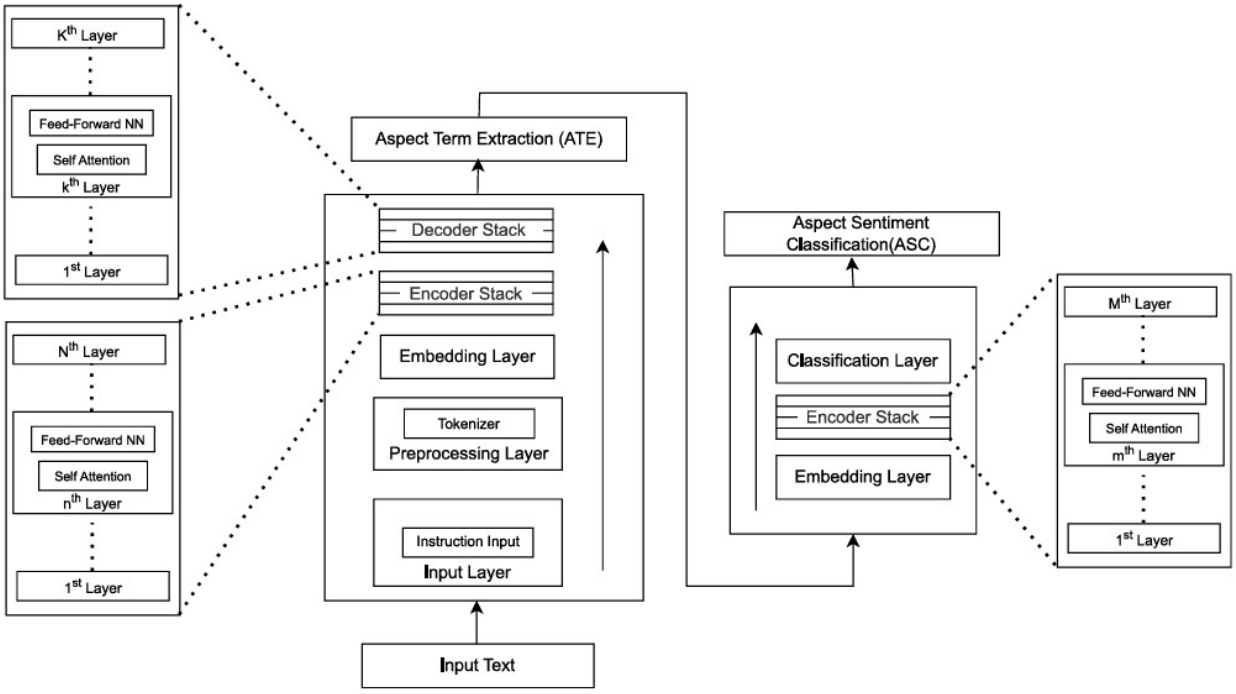

Fig. 1: Proposed model

Algorithm 1: Aspect-based Sentiment Analysis (ABSA)

Require: Review X

Aspect Term Extraction (ATE):
  $A_c = f_{ATE}(X)$
Target Aspect Filtering:
  $A = f_{filter}(A_c)$
Aspect Sentiment Classification (ASC):
  $S = \{\}$
for each aspect term a in A **do**
    $s = f_{ASC}(X, a)$
    Add (a, s) to S: $S = S \cup \{(a, s)\}$
end for
return Final aspect terms A with sentiment labels S: $\{(a, s) \mid a \in A, s = f_{ASC}(X, a)\}$

Functions:

- $f_{ATE}(X)$: Function for Aspect Term Extraction. Takes review X and returns candidate terms (A_c). ($A_c \subseteq X$)
- $f_{filter}(A_c)$:A filtering function. Takes candidate terms and returns refined aspects (A). ($A \subseteq A_c$)
- $f_{ASC}(X, a)$: Function for Aspect Sentiment Classification. Takes review X and an aspect term a, returns sentiment label s. ($s \in \{positive, negative, neutral\}$)

IV. Results

Table VI presents the F1 scores for the models being built and evaluated by ourselves. In addition to that we have presented Figure 4 and Figure 5 which give a view on the robustness of each model for the aspect term extraction task and the sentiment polarity task separately based on the accuracies. According to Figure 4 and Figure 5, if the model shows high accuracy for each task and the accuracies for the two domains do not exhibit drastic deviations, then the relevant model will be selected.

A. LLaMA 2-7B with QLoRA

The first section of Table VI shows the performance of `Llama-2` with QLoRA. It emphasizes that this model shows notable performance in both aspect extraction and sentiment polarity tasks. On *Res-14* and *Lap-14* datasets, `Llama 2` shows a slight edge in aspect extraction compared to sentiment polarity. These performances were obtained using the L4 GPU emphasizing the model's efficiency and effectiveness in computational performance.

B. Mistral 7B with QLoRA

According to the values in Table VI, the model `Mistral 7B` exhibits superior performance in both aspect extraction and sentiment polarity tasks compared to the `Llama 2` model. Specifically, `Mistral 7B` achieves higher F1 scores across both *Res-14* and *Lap-14* datasets, indicating its greater capability in accurately identifying aspects and determining sentiment polarity within the text. These results were achieved using an L4 GPU, similar to the `Llama 2` model.

When comparing the two models, `Mistral 7B` demonstrates a clear advantage in both tasks. While `Llama 2` performs competently, `Mistral 7B` consistently outperforms it,

showcasing its enhanced effectiveness and reliability in handling aspect extraction and sentiment polarity analysis. This comparison highlights Mistral 7B's performance, making it a more capable model for these specific natural language processing tasks.

C. Some models with BERT and GloVe

As a part of the comparative study for the survey, we conducted experiments using several advanced models for aspect-based sentiment analysis. We experimented three main models: SSGCN+Glove, ASGCN+UIKA+Glove[4], and Span-ASTE+BERT [63] both in local and Colab environments. The focus was on to evaluating their performance on the SemEval 2014 dataset. These results are stated in Table VI. But still we observed that Instruct-DeBERTa outperforms all.

D. SetFit

In the second section of Table VI, F1-scores for various sentence models using the SETFIT framework [68] are presented. A model listed alone indicates its use for both aspect extraction and sentiment polarity identification (e.g., BGE [69]), while combinations are marked with a "+". For instance, Paraphrase-MiniLM-L6-v2 [70, 71] is used for both tasks in one row, and with +MpNet [72] indicates that Paraphrase-MiniLM-L6-v2 was used for the aspect extraction component and MpNet was used for the sentiment polarity identification component.

Differences in aspect extraction accuracy, even with the same model, arise from the end-to-end fine-tuning process, where the sentiment polarity model can impact final results—either positively, as with Paraphrase-MiniLM-L6-v2 and MpNet, or negatively, as with LaBSE.

LaBSE [73] consistently excels, likely due to its subtle information capture, performing well in both aspect-based sentiment analysis and sentiment polarity classification. Mpnet and RoBERTa-STSb-v2 [74] also enhance performance in multiple setups.

To give a better overview of how various models perform, we include Figure 4 and Figure 5, which visualize the SETFIT accuracy values for the models that we evaluated.

In Figure 4, we provide a detailed analysis of aspect extraction accuracy across various models. It's clear that LaBSE stands out as the top performer on both datasets. Additionally, ALBERT+DistilRoBERTa and LaBSE+RoBERTa-STSb also perform well, with accuracies approaching 90%. Similarly, in Figure 5, we analyze sentiment polarity identification accuracy for the same models and datasets. LaBSE+RoBERTa-STSb achieves the highest accuracy for *Res-14*, while LaBSE+MpNet leads in accuracy for *Lap-14*.

E. Instruct-DeBERTa (Proposed Model)

From all the evaluated models, our model performs the best with the highest accuracies and F1 scores. It is also robust for both domains which makes it the best performing model. As discussed in the methodology, we selected the best performing models from Table III and Table IV to create our own hybrid model. When looking at Table VI, Figure 4 and Figure 5 it is clear that Instruct-DeBERTa outperforms our finetuned Llama, Mistral, and all the Setfit based models.

Table VII shows how the two best models we selected for each subtask perform individually on their relevant task. These F1 scores are for the combined task, which means our model is capable of performing both the aspect extraction and the sentiment polarity tasks. For the first task which is extracting the aspects, our model gives closer accuracies for what has been reported by Scaria et al. [26] for the InstructABSA model. Different ways of splitting a dataset can affect the reported accuracies. Also for the sentiment polarity classification task the original model, DeBERTa-V3-base-absa-v1 by Yang et al. [22, 41] which is specialized only for detecting polarities gives slightly higher accuracies than our hybrid model. As seen for the Res-14 dataset the sentiment polarity accuracy for the individual task by DeBERTa-V3-base-absa-v1 is reported as 90.94% while our reported 88.63%. This is due to the models being pipelined and the extracted aspects from the first model is being fed to the second model rather than calculating the accuracies separately for individual tasks. Hence the slight deduction in the hybrid model is justified.

So, looking on to all the past models in Table II ,Table III, Table IV and the models that we worked on in Table VI, it is clear that our model, Instruct-DeBERTa is the best performing hybrid model designed for the combined task of aspect extraction and sentiment polarity detection. Moreover, our hybrid model shows promising results in the laptop domain as well. Our model gives an F1-score of 91.56% and 89.65% for aspect extraction and sentiment polarity respectively.

Also Table V, in the literature review lists out the joint models which are equivalent to the model we built. These perform the joint task of ABSA. Here in order for the F1 score to be counted the aspect and the respective sentiment in the original dataset needs to be correct. The F1 scores of these models along with our model can be visualized in Figure 2 and Figure 3. It is clear that our model clearly out performs the currently available joint task hybrid models. It gives a pair extraction F1 score of 80.78% and 80.94% which exceed the current reported highest accuracy for the rest-14 and lap-14 datasets. From Table V, Figure 2 and Figure 3 it is clear that our model is the best performing joint task model. Our model outperforms all other hybrid models in both domains which again proves that it is not only accurate but also robust to different domains as well.

V. Conclusion

In this paper, we presented a comprehensive review and detailed experimental analysis of ABSA methodologies, focusing on the latest advancements in Transformer-based models.

Our hybrid model, Instruct-DeBERTa, was designed to harness the specific advantages of two best performing models. InstructABSA is known for its accuracy in identifying and extracting relevant aspects from text, while DeBERTa-V3-baseabsa-V1 excels in classifying the sentiment associated with these aspects. By integrating these

TABLE VI: F1 scores of models evaluated by this study on the SemEval 2014 [35] benchmark

Model		F1 Score (%)			
		Res-14		Lap-14	
		Aspect Extraction	Sentiment Polarity	Aspect Extraction	Sentiment Polarity
Llama-2-7b [65] with QLoRA [66]		71.94	69.29	71.66	66.53
Mistral-7b [67] with QLoRA [66]		81.33	76.46	77.65	72.40
Instruct-DeBERTa (Proposed Model)		**91.39**	**88.63**	**91.56**	**89.65**
ASGCN+GloVe+UIKA [7]		–	76.93	–	75.21
SSGCN+Glove [54]		–	76.43	–	76.17
SPAN-ASTE [63]		67.54	–	61.12	–
SETFIT [68]	BGE [69] (Small)	72.24	75.59	64.79	75.59
	Sentence-T5 [75] (Base)	56.82	78.74	63.29	74.01
	RoBERTa-STSb-v2 [71, 74] (Base)	82.37	77.95	84.26	67.71
	Paraphrase-MiniLM-L6-v2 [70, 71]	84.58	71.65	83.14	64.56
	+MpNet [72]	84.58	78.74	79.40	70.07
	CLIP-ViT-B-32-multilingual-v1 [71, 76]	81.49	59.05	73.03	53.54
	facebook-dpr-ctx-encoder-multiset-base [77]	76.54	78.74	75.52	77.45
	SPECTER [78]	81.93	71.65	77.52	55.11
	GTR [79] (Base)	81.85	74.80	84.70	74.80
	SBERT [71] (Base)	83.18	70.86	84.32	61.41
	TinyBERT [71, 80]	78.76	73.22	82.83	63.77
	ALBERT [71, 81]	80.08	74.80	80.22	66.92
	+DistilRoBERTa [82]	81.49	74.81	79.40	68.50
	DistilRoBERTa [82]	84.95	75.59	80.97	69.29
	+All-MiniLM-L6-v2 [70, 71]	85.46	71.65	81.27	63.77
	MpNet [72]	86.28	77.95	88.80	73.22
	LaBSE [73]	89.38	73.23	90.30	64.57
	+MpNet [72]	88.55	74.80	89.51	75.59
	+GTR [79] (Base)	88.55	74.02	87.27	73.22
	+RoBERTa-STSb-v2 [71, 74] (Base)	90.30	77.17	89.51	70.08

TABLE VII: F1 scores for the selected models individually and when pipe-lined

Model	F1 Score (%)							
	Res-14		Lap-14		Res-15		Res-16	
	AE	SP	AE	SP	AE	SP	AE	SP
InstructABSA [26]	**92.10**	–	**92.30**	–	**76.64**	–	**80.32**	–
DeBERTa-V3-base-absa-v1.1* [22, 41]	–	**90.94**	–	**90.32**	–	**89.55**	–	**83.71**
Instruct-DeBERTa (Proposed Model)	91.39	88.63	91.56	89.65	75.13	81.26	77.79	79.35

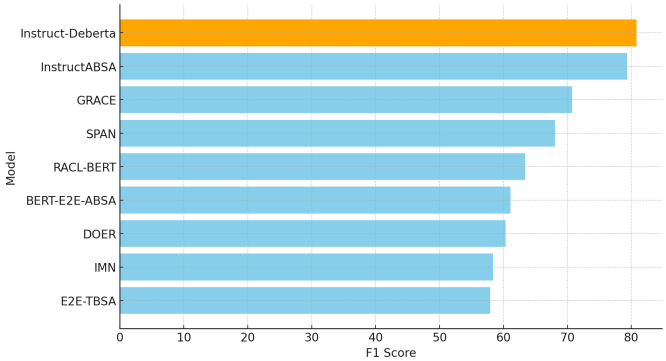

Fig. 2: F1 scores of models for the joint task of ASC and ATE for lap-14

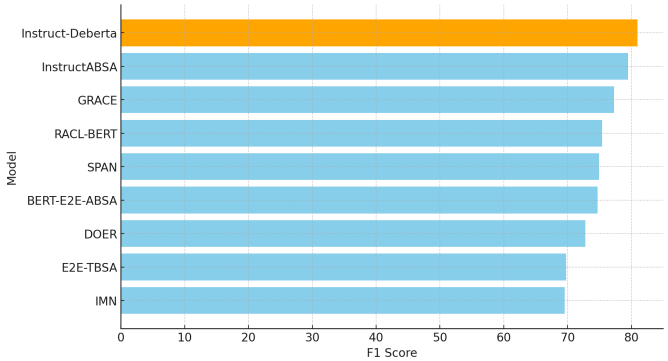

Fig. 3: F1 scores of models for the joint task of ASC and ATE for res-14

models, we aimed to create a comprehensive tool that could perform both tasks with high precision and reliability.

Our model achieved the highest accuracy and F1 scores across multiple datasets, showcasing its ability to effectively handle diverse textual data and consistently deliver high-quality results. This performance can be attributed to the synergistic integration of `InstructABSA` and `DeBERTa-V3-baseabsa-V1`, which allows our hybrid model to maintain a delicate balance between precision in aspect extraction and accuracy in sentiment classification.

In conclusion, our comprehensive review and experimental analysis highlight the significant advancements made possible

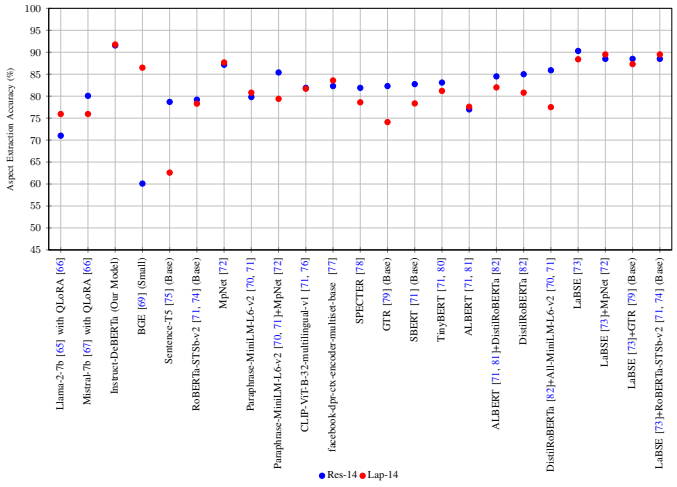

Fig. 4: Aspect extraction accuracy of models

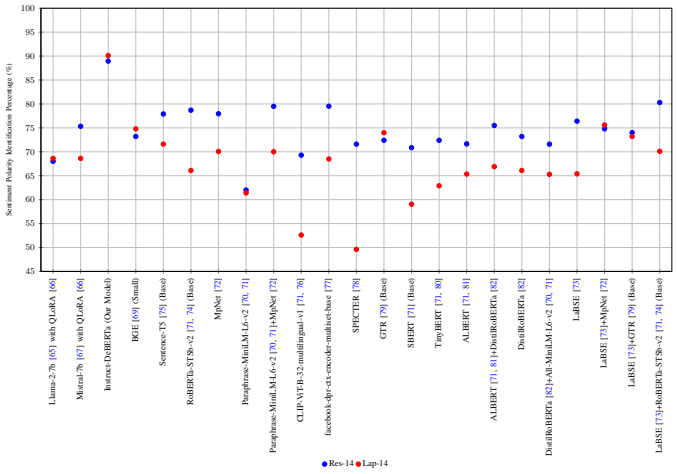

Fig. 5: Sentiment polarity percentage of models

by Transformer-based models in the field of ABSA. The development of `Instruct-DeBERTa` represents a notable contribution, offering a powerful and versatile solution for accurately extracting aspects and classifying sentiment in diverse textual data. The superior performance of our hybrid model sets a new benchmark for future research and applications in ABSA, underscoring the potential of integrating state-of-the-art models to enhance the effectiveness of sentiment analysis methodologies.

References

[1] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[2] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based lstm for aspect-level sentiment classification," in *EMNLP*, 2016, pp. 606–615.

[3] X. Li, L. Bing, W. Lam, and B. Shi, "Transformation networks for target-oriented sentiment classification," *arXiv preprint arXiv:1805.01086*, 2018.

[4] C. Zhang, Q. Li, and D. Song, "Aspect-based sentiment classification with aspect-specific graph convolutional networks," in *EMNLP-IJCNLP*, 2019, pp. 4568–4578.

[5] R. Li, H. Chen, F. Feng, Z. Ma, X. Wang, and E. Hovy, "Dual graph convolutional networks for aspect-based sentiment analysis," in *ACL-IJCNLP*, 2021, pp. 6319–6329.

[6] Q. Zhong, L. Ding, J. Liu, B. Du, H. Jin, and D. Tao, "Knowledge graph augmented network towards multi-view representation learning for aspect-based sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–14, 2023.

[7] J. Liu, Q. Zhong, L. Ding, H. Jin, B. Du, and D. Tao, "Unified instance and knowledge alignment pretraining for aspect-based sentiment analysis," *IEEE/ACM transactions on audio, speech, and language processing*, vol. 31, pp. 2629–2642, 2023.

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[9] e. a. Zhao, "Bert-dk: Incorporating domain knowledge into bert for knowledge-intensive tasks," *arXiv preprint arXiv:2002.00000*, 2020.

[10] Y. Song, J. Wang, T. Jiang, Z. Liu, and Y. Rao, "Attentional encoder network for targeted sentiment classification," *arXiv preprint arXiv:1902.09314*, 2019.

[11] X. Zhao, Y. Liu, J. Wang, and L. Zhang, "Bert-mrc: A bert-based model for machine reading comprehension," *Journal of Artificial Intelligence Research*, vol. 68, pp. 345–361, 2020.

[12] H. Xu, B. Liu, L. Shu, and P. S. Yu, "Bert post-training for review reading comprehension and aspect-based sentiment analysis," *arXiv preprint arXiv:1904.02232*, 2019.

[13] A. Karimi, L. Rossi, and A. Prati, "Adversarial Training for Aspect-Based Sentiment Analysis with BERT," in *ICPR*. IEEE, 2021, pp. 8797–8803.

[14] X. Bai, P. Liu, and Y. Zhang, "Investigating typed syntactic dependencies for targeted sentiment classification using graph attention neural network," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 503–514, 2020.

[15] M. Zhang, Y. Zhu, Z. Liu, Z. Bao, Y. Wu, X. Sun, and L. Xu, "Span-level aspect-based sentiment analysis via table filling," in *ACl*, 2023, pp. 9273–9284.

[16] C. Chen, Z. Teng, Z. Wang, and Y. Zhang, "Discrete opinion tree induction for aspect-based sentiment analysis," in *ACl*, 2022, pp. 2051–2064.

[17] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[18] B. Wang, T. Shen, G. Long, T. Zhou, and Y. Chang, "Eliminating sentiment bias for aspect-level sentiment classification with unsupervised opinion extraction," *arXiv preprint arXiv:2109.02403*, 2021.

[19] J. Dai, H. Yan, T. Sun, P. Liu, and X. Qiu, "Does syntax matter? a strong baseline for aspect-based sentiment analysis with roberta," *arXiv preprint arXiv:2104.04986*, 2021.

[20] P. He, X. Liu, J. Gao, and W. Chen, "DeBERTa: Decoding-enhanced BERT with Disentangled Attention," in *ICLR*, 2020.

[21] P. He, J. Gao, and W. Chen, "Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing," *arXiv preprint arXiv:2111.09543*, 2021.

[22] H. Yang, C. Zhang, and K. Li, "Pyabsa: A modularized framework for reproducible aspect-based sentiment analysis," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023*, I. Frommholz, F. Hopfgartner, M. Lee, M. Oakes, M. Lalmas, M. Zhang, and R. L. T. Santos, Eds. ACM, 2023, pp. 5117–5122.

[23] R. Marcacini and E. Silva, "Aspect-based sentiment analysis using bert with disentangled attention," in *LatinX in AI at ICML*, 2021, pp. 1–4.

[24] H. Yang and K. Li, "Modeling Aspect Sentiment Coherency via Local Sentiment Aggregation," in *Findings of EACL*, 2024, pp. 182–195.

[25] H. Yang, B. Zeng, J. Yang, Y. Song, and R. Xu, "A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction," *Neurocomputing*, vol. 419, pp. 344–356, 2021.

[26] K. Scaria, H. Gupta, S. Goyal, S. A. Sawant, S. Mishra, and C. Baral, "Instructabsa: Instruction learning for aspect based sentiment analysis," *arXiv preprint arXiv:2302.08624*, 2023.

[27] H. Yang and K. Li, "Modeling aspect sentiment coherency via local sentiment aggregation," *arXiv preprint arXiv:2110.08604*, 2021.

[28] H. Yan, J. Dai, X. Qiu, Z. Zhang *et al.*, "A unified generative framework for aspect-based sentiment analysis," *arXiv preprint arXiv:2106.04300*, 2021.

[29] Z. Chen and T. Qian, "Relation-aware collaborative learning for unified aspect-based sentiment analysis," in *ACL*, 2020, pp. 3685–3694.

[30] M. Hu, Y. Peng, Z. Huang, D. Li, and Y. Lv, "Open-domain targeted sentiment analysis via span-based extraction and classification," *arXiv preprint arXiv:1906.03820*, 2019.

[31] X. Li, L. Bing, P. Li, and W. Lam, "A unified model for opinion target extraction and target sentiment prediction," in *AAAI*, vol. 33, no. 01, 2019, pp. 6714–6721.

[32] X. Li, L. Bing, W. Zhang, and W. Lam, "Exploiting bert for end-to-end aspect-based sentiment analysis," *arXiv preprint arXiv:1910.00883*, 2019.

[33] H. Luo, T. Li, B. Liu, and J. Zhang, "Doer: Dual cross-shared rnn for aspect term-polarity co-extraction," *arXiv preprint arXiv:1906.01794*, 2019.

[34] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "An interactive multi-task learning network for end-to-end aspect-based sentiment analysis," *arXiv preprint arXiv:1906.06906*, 2019.

[35] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 task 4: Aspect based sentiment analysis," in *SemEval 2014*, 2014, pp. 27–35.

[36] L. Xu, L. Bing, W. Lu, and F. Huang, "Aspect sentiment classification with aspect-specific opinion spans," in *EMNLP*, 2020, pp. 3561–3567.

[37] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," *arXiv preprint arXiv:1709.00893*, 2017.

[38] C. Zhang, Q. Li, and D. Song, "Aspect-based sentiment classification with aspect-specific graph convolutional networks," *arXiv preprint arXiv:1909.03477*, 2019.

[39] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "Exploiting document knowledge for aspect-level sentiment classification," *arXiv preprint arXiv:1806.04346*, 2018.

[40] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," *arXiv preprint arXiv:1605.08900*, 2016.

[41] H. Yang, B. Zeng, M. Xu, and T. Wang, "Back to reality: Leveraging pattern-driven modeling to enable affordable sentiment dependency learning," *arXiv preprint arXiv:2110.08604*, 2021.

[42] A. Karimi, L. Rossi, and A. Prati, "Improving bert performance for aspect-based sentiment analysis," *arXiv preprint arXiv:2010.11731*, 2020.

[43] H. Tang, D. Ji, C. Li, and Q. Zhou, "Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification," in *Proceedings of the 58th annual meeting of the ACL*, 2020, pp. 6578–6588.

[44] B. Liang, H. Su, L. Gui, E. Cambria, and R. Xu, "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks," *Knowledge-Based Systems*, vol. 235, p. 107643, 2022.

[45] H. T. Nguyen and M. Le Nguyen, "Effective attention modeling for aspect-level sentiment classification," in *2018 10th International conference on knowledge and systems engineering (KSE)*. IEEE, 2018, pp. 25–30.

[46] J. Zhou, J. X. Huang, Q. V. Hu, and L. He, "Sk-gcn: Modeling syntax and knowledge via graph convolutional network for aspect-level sentiment classification," *Knowledge-Based Systems*, vol. 205, p. 106292, 2020.

[47] B. Xing and I. W. Tsang, "Understand me, if you refer to aspect knowledge: Knowledge-aware gated recurrent memory network," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, no. 5, pp. 1092–1102, 2022.

[48] P. Chen, Z. Sun, L. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *EMNLP*, 2017, pp. 452–461.

[49] H. Ge, X. Tu, M. Xie, and Z. Ma, "Asp-cnn: aligning semantic parts for fine-grained image classification," *Journal of Electronic Imaging*, vol. 28, no. 2, pp. 023 024–023 024, 2019.

[50] P. Zhao, L. Hou, and O. Wu, "Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification," *Knowledge-Based Systems*, vol. 193, p. 105443, 2020.

[51] A. Rietzler, S. Stabinger, P. Opitz, and S. Engl, "Adapt or get left behind: Domain adaptation through bert language model finetuning for aspect-target sentiment classifica-

tion," *arXiv preprint arXiv:1908.11860*, 2019.

[52] Y. Mao, Y. Shen, C. Yu, and L. Cai, "A joint training dual-mrc framework for aspect based sentiment analysis," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 15, 2021, pp. 13 543–13 551.

[53] Y. Zhang, M. Zhang, S. Wu, and J. Zhao, "Towards unifying the label space for aspect- and sentence-based sentiment analysis," *arXiv preprint arXiv:2203.07090*, 2022.

[54] Z. Zhang, Z. Zhou, and Y. Wang, "Ssegcn: Syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis," in *NAACL*, 2022, pp. 4916–4925.

[55] W. Xue and T. Li, "Aspect based sentiment analysis with gated convolutional networks," in *ACL*, 2018, pp. 2514–2523.

[56] Z. Chen and T. Qian, "Transfer capsule network for aspect level sentiment classification," in *ACL*, 2019, pp. 547–556.

[57] S. Wang, S. Mazumder, B. Liu, M. Zhou, and Y. Chang, "Target-sensitive memory networks for aspect sentiment classification," in *ACL*, 2018.

[58] J. Tang, Z. Lu, J. Su, Y. Ge, L. Song, L. Sun, and J. Luo, "Progressive self-supervised attention learning for aspect-level sentiment analysis," *arXiv preprint arXiv:1906.01213*, 2019.

[59] D. Tang, B. Qin, X. Feng, and T. Liu, "Effective lstms for target-dependent sentiment classification," *arXiv preprint arXiv:1512.01100*, 2015.

[60] J. Liu and Y. Zhang, "Attention modeling for targeted sentiment," in *EACL*, 2017, pp. 572–577.

[61] E. Hosseini-Asl and W. Liu, "Generative language model for few-shot aspect-based sentiment analysis," Dec. 26 2023, uS Patent 11,853,706.

[62] H. Luo, L. Ji, T. Li, N. Duan, and D. Jiang, "Grace: Gradient harmonized and cascaded labeling for aspect-based sentiment analysis," *arXiv preprint arXiv:2009.10557*, 2020.

[63] L. Xu, Y. K. Chia, and L. Bing, "Learning span-level interactions for aspect sentiment triplet extraction," *ACL*, 2021.

[64] D. Jayakody, K. Isuranda, A. Malkith, N. de Silva, S. R. Ponnamperuma, G. Sandamali, and K. Sudheera, "Aspect-based sentiment analysis techniques: A comparative study," *arXiv preprint arXiv:2407.02834*, 2024.

[65] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale *et al.*, "Llama 2: Open Foundation and Fine-Tuned Chat Models," *arXiv preprint arXiv:2307.09288*, 2023.

[66] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, "QLoRA: Efficient Finetuning of Quantized LLMs," in *NeurIPS*, vol. 36, 2023.

[67] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. d. l. Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier *et al.*, "Mistral 7b," *arXiv preprint arXiv:2310.06825*, 2023.

[68] L. Tunstall, N. Reimers, U. E. S. Jo, L. Bates, D. Korat, M. Wasserblat, and O. Pereg, "Efficient few-shot learning without prompts," *arXiv preprint arXiv:2209.11055*, 2022.

[69] S. Xiao, Z. Liu, P. Zhang, and N. Muennighof, "C-Pack: Packaged Resources To Advance General Chinese Embedding," *arXiv preprint arXiv:2309.07597*, 2023.

[70] W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, and M. Zhou, "MiniLM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers," *NeurIPS*, vol. 33, pp. 5776–5788, 2020.

[71] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in *EMNLP-IJCNLP*, 2019, pp. 3982–3992.

[72] K. Song, X. Tan, T. Qin, J. Lu, and T.-Y. Liu, "MPNet: Masked and Permuted Pre-training for Language Understanding," *NeurIPS*, vol. 33, pp. 16 857–16 867, 2020.

[73] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, and W. Wang, "Language-agnostic BERT Sentence Embedding," in *ACL*, 2022, pp. 878–891.

[74] D. Cer, M. Diab, E. Agirre, I. Lopez-Gazpio, and L. Specia, "SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Crosslingual Focused Evaluation," in *SemEval*, 2017.

[75] J. Ni, G. H. Abrego, N. Constant, J. Ma, K. Hall, D. Cer, and Y. Yang, "Sentence-T5: Scalable Sentence Encoders from Pre-trained Text-to-Text Models," in *Findings of ACL*, 2022, pp. 1864–1874.

[76] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *ICML*. PMLR, 2021, pp. 8748–8763.

[77] V. Karpukhin, B. Oguz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W.-t. Yih, "Dense passage retrieval for open-domain question answering," in *Proceedings of the 2020 Conference on EMNLP*. Online: ACL, Nov. 2020, pp. 6769–6781.

[78] A. Cohan, S. Feldman, I. Beltagy, D. Downey, and D. S. Weld, "SPECTER: Document-level Representation Learning using Citation-informed Transformers," in *ACL*, 2020.

[79] J. Ni, C. Qu, J. Lu, Z. Dai, G. H. Abrego, J. Ma, V. Zhao, Y. Luan, K. Hall, M.-W. Chang *et al.*, "Large Dual Encoders Are Generalizable Retrievers," in *EMNLP*, 2022, pp. 9844–9855.

[80] X. Jiao, Y. Yin, L. Shang, X. Jiang, X. Chen, L. Li, F. Wang, and Q. Liu, "TinyBERT: Distilling BERT for Natural Language Understanding," in *Findings of EMNLP*, 2020, pp. 4163–4174.

[81] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," *arXiv preprint arXiv:1909.11942*, 2019.

[82] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *ArXiv*, vol. abs/1910.01108, 2019.