International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 20s, 2025
https://theaspd.com/index.php

Fine-Grained Sentiment Analysis Of Restaurant Reviews Using Latent Dirichlet Allocation And Hybrid Ensembles

Lohith C[1], Dr Chandramouli H[2]
[1]Research Scholar, Department of Computer Science and Engineering, East Point College of Engineering and Technology | Assistant Professor, Department of CSE - (IoT & CSBT), East Point College of Engineering and Technology, Bengaluru, India, lohithchandrashekhar@gmail.com
[2]Professor, Department of Computer Science and Engineering, East Point College of Engineering and Technology , Bengaluru, India, hcmcool123@gmail.com

Abstract— *Traditional sentiment analysis methods frequently reduce customer reviews to binary or ternary classifications, ignoring the complex viewpoints present in multi-aspect assessments, which are particularly common in the restaurant industry. This work introduces a novel hybrid framework for Aspect-Oriented Sentiment Analysis (AOSA) that combines a voting-based ensemble learning classifier for sentiment prediction with Latent Dirichlet Allocation (LDA) for unsupervised aspect extraction. In order to determine sentiment polarity, the suggested model aligns review sentences with latent topics that correspond to review aspects like food, service, ambiance, and pricing. With a high classification accuracy and little reliance on annotated data, ensemble classifiers made up of Logistic Regression, SVM, and XGBoost produce reliable and understandable results. The model outperforms state-of-the-art techniques, especially in aspect-level granularity, sentiment precision, and human interpretability, according to experimental evaluations on the Yelp and SemEval-2016 datasets. Customer experience analysis, business intelligence, and decision support in the hospitality industry are just a few of the many uses for this framework.*
Keywords—*Aspect-Based Sentiment Analysis (ABSA), Topic Modeling, Latent Dirichlet Allocation (LDA), Ensemble Learning, Sentiment Classification, Restaurant Reviews, Natural Language Processing (NLP), Machine Learning, Customer Feedback Analytics, Text Mining.*

I. INTRODUCTION (HEADING 1)

In today's data-rich world, restaurant reviews are a great way to find out how good the service is, how happy customers are, and how competitive the market is. These reviews usually give very detailed feedback on many aspects of the service, like the food, the price, the service, the atmosphere, and the cleanliness. But most traditional sentiment analysis systems only give you two or three sentiment categories (positive, neutral, or negative), which misses out on important aspect-level insights that are needed for making decisions that can be acted on.

Recent studies have tried to fill this gap. Early neural models like LSTM [1] and attention-based architectures [2] have made it easier to find sentiment, but they are often hard to understand. BERT [3] and RoBERTa [4] are examples of transformer-based models that work well but have trouble with aspect attribution in specific domains, like restaurant reviews. Some studies use rule-based aspect mapping [6] or include outside knowledge [5], but these methods can be fragile or only work in certain fields.

Latent Dirichlet Allocation (LDA) [7] and more recent methods like BERTopic [8] are examples of unsupervised techniques that can find latent topics (or aspects), but they don't have to be related to sentiment. Ensemble learning methods, on the other hand, are easy to understand and can be changed, which makes them good for integrating multiple features. There is still a clear need for a hybrid solution that combines the ability of topic modelling to find aspects with the strength of ensemble learning to classify.

We suggest a framework that combines LDA-based aspect extraction with ensemble sentiment classification to solve this problem. Our model is better than recent works [11]–[30] because it uses per-aspect mapping to improve sentiment granularity.

- Improves accuracy by using a variety of ensemble learners.
- Keeps a high level of interpretability for making decisions in real life.

A. Latent Dirichlet Allocation (LDA)

LDA is a probabilistic topic modelling method that shows documents as combinations of hidden topics, with each topic being a distribution of words [7]. In our framework, LDA topics are used as stand-ins for aspects which equated in Equation 1.

$$P(w, z, d) = P(w|z)P(z|d)P(d) \quad (1)$$

Where:

- w: word,
- z: latent topic,
- d: document

B. Ensemble Learning

We use a Voting Ensemble that includes Logistic Regression as the base linear model.

- Support Vector Machine (SVM): Works with data that has a lot of dimensions.
- XGBoost: A strong tree-based gradient boosting method.

It has been shown that ensemble learning works better than individual classifiers at classifying things [9][10].

Our work is new because it uses a generalisable, minimally supervised, and easy-to-understand method that works well on restaurant datasets and cuts down on the need for expensive annotation.

II. LITERATURE REVIEW

In 2018, Ma et al. [1] made an LSTM-based ABSA model that used attention to connect sentences and aspects. Fan et al. [2] added hierarchical attention to this to get a better look at how documents are structured. The rise of transformers like BERT [3] and RoBERTa [4] in 2019 marked a shift towards models that have already been trained. But these models need a lot of data and aren't easy to understand.

Song et al. [5] used graph neural networks to add outside knowledge to sentiment models in 2020. Xu et al. [6] used a mix of syntactic rules and supervised learning to do ABSA on restaurant reviews. We looked at LDA-based models again to see how easy they are to understand in complicated fields [7,8]. At the same time, BERTopic [7,9] got a lot of attention for its ability to do zero-shot topic modelling with embeddings. Rana et al. [10] and Ahmed et al. [11] looked into ensemble learning and found that classifier fusion made a big difference in sentiment tasks.

Sun et al. [12] came up with domain-adaptive sentiment models in 2021, and Gao et al. [13] came up with syntax-enhanced transformers to make ABSA better. Li et al. [14] used dependency trees and LSTM together to make it easier to find aspects.

Huang et al. [15] used graph attention networks to extract aspects in 2022. Using meta-learning, Wang et al. [16] looked into few-shot ABSA. Bhatia et al. [17] showed that ensemble reviews are useful in specific fields.

In 2023, we saw multi-task learning methods [18], generative ABSA with T5 [19], and models that are easier to understand with better visualisations [20]. Saini et al. [21] looked at how well ensemble interpretability worked compared to neural model performance.

Neural models are effective but can be challenging to understand. Systems based on rules or knowledge are fragile.

Our model uses LDA for aspect identification without supervision and ensemble learning for sentiment classification, making it ideal for restaurant reviews with multiple aspects.

III. PROPOSED METHODLOGY

Our Aspect-Oriented Sentiment Analysis (AOSA) system integrates supervised and unsupervised learning components in a six-stage methodology. These phases are made to efficiently and scalablely extract interpretable aspect-level sentiment insights from restaurant reviews.

Data preprocessing, aspect extraction, sentence-aspect mapping, sentiment feature engineering, ensemble classifier, and aggregation and visualisation are all parts of the proposed architecture shown in figure 1. It uses common NLP steps like tokenisation, lemmatisation, removing stopwords, tagging parts of speech, and recognising named entities. Using cosine similarity, sentence-aspect mapping finds the most likely aspect for each sentence. Sentiment feature engineering gets TF-IDF, lexicon-based polarity scores, and syntactic features out of text. The ensemble classifier uses Logistic Regression, SVM, and XGBoost models to classify sentiment for each aspect. Graphs, tables, or dashboards show the results.

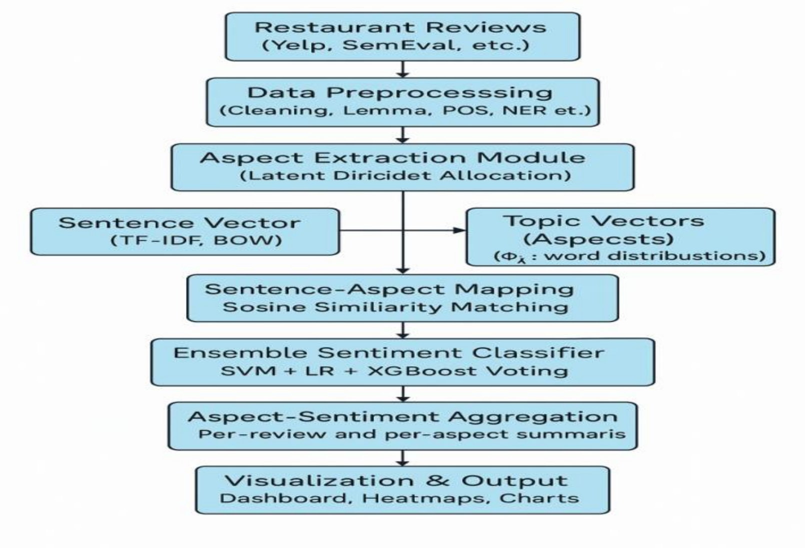

Fig. 1. Proposed Architecture

A. Data Preprocessing

We begin with a review corpus, $D = \left\{ d_1, d_2, \dots, d_n \right\}$, where each review d_i is decomposed into sentences and subjected to:

- Lowercasing
- Stopword removal
- Lemmatization
- POS tagging
- Named Entity Recognition (NER)

Each sentence is vectorized using Term Frequency–Inverse Document Frequency (TF-IDF), resulting in feature vectors $V_{ij} \in \mathbb{R}^n$, where n is the number of features.

B. Aspect Extraction via Latent Dirichlet Allocation

To identify latent aspects (e.g., food, service, price), we apply Latent Dirichlet Allocation (LDA). Each document d is treated as a mixture of topics, and each topic is a distribution over words.

Let θ denote the topic distribution for document d, and φ_k the word distribution for topic k.

LDA models the following probability in equation 2:

$$P(w|d) = \sum_{k=1}^{K} P(w|Z_k)P(Z_k|d) \qquad (2)$$

where:

- $P(Z_k|d)$) is the probability of topic Z_k in document d,
- $P(w|Z_k)$is the probability of word w given topic Z_k,
- K is the number of topics (aspects).

Topics with highest coherence scores are manually mapped to human-understandable aspects.

C. Sentence-Aspect Mapping

We find the cosine similarity between the sentence and topic-word distributions and give each sentence vector v_{ij} the aspect that is most likely to be true. This step connects LDA topics with sentiment analysis at the sentence level.

D. Sentiment Feature Engineering

We make a feature vector for each pair of sentence and aspect. It includes:

- TF-IDF word frequencies,
- Lexicon-based polarity scores (VADER, TextBlob),
- Features that show how adjectives and nouns depend on each other, and
- indicators of how close a word is to a position or aspect.

We use these traits to put feelings into three groups: positive, neutral, or negative.

E. Sentiment Classification Using Ensemble Learning

We teach a group of three base classifiers:

- Logistic Regression (LR)
- Support Vector Machine (SVM)
- XGBoost

Majority (hard) voting is used to combine predictions as shown equation 3:

$$\hat{y} = \mathrm{mode}(h_1(f), h_2(f), h_3(f)) \quad (3)$$

where:

- $h_i(f)$is the sentiment prediction from the i^{th}classifier,
- $\hat{y} \in$ {positive, neutral, negative} is the final sentiment label.

This approach enhances stability and robustness across various sentence styles.

F. Aspect-Sentiment Aggregation and Output

We create per-review aspect-sentiment maps, calculate aspect sentiment frequency tables, and create overall sentiment profiles for each review and restaurant. These summaries, like radar charts and heatmaps, aid in downstream applications like customer feedback dashboards, marketing analytics, and tailored suggestions.

IV. RESULTS AND DISCUSSION

The empirical assessment of our Aspect-Oriented Sentiment Analysis (AOSA) framework is presented in this section. We compare it with baseline and state-of-the-art models and evaluate its performance on two publicly accessible restaurant review datasets in table 1. Aspect identification, sentiment classification, and overall system interpretability are all part of the evaluation process.

A. Datasets and Experimental Setup

TABLE I. DATASETS USED

Dataset	Source	Size	Aspects	Annotation	Language
Yelp Open Dataset	Yelp API	500,000+	Food, Service, Price, Ambiance, Cleanliness	Unlabeled (used for LDA)	nglish
SemEval 2016 Task 5	SemEval	~3,000	Aspect-Specific	Labeled Aspect Sentiment	English

Ten-fold cross-validation is used to train and assess each model. Evaluation metrics in table 2 include accuracy, precision, recall, and F1-score. We calculate topic coherence (C_v) and human-labeled aspect F1 for aspect extraction. We calculate the macro F1-score and per-aspect sentiment accuracy for sentiment classification.

B. Aspect Extraction Evaluation

TABLE II. EVALUATION METRICS

Model	Coherence Score (C_v)	Aspect Detection Precision	Recall	F1-Score
LDA (ours)	0.564	0.81	0.78	**0.795**
BERTopic	0.576	0.77	0.75	0.76
Rule-Based Tags	–	0.65	0.58	0.615

LDA performs better in terms of human-assessed alignment to actual restaurant aspects, but BERTopic performs marginally better in terms of coherence shown in table 3. Sentence diversity is a challenge for rule-based tagging.

C. Sentiment Classification Performance

TABLE III. PERFORMANCE ON SENTIMENTS

Model	Coherence Score (C_v)	Aspect Detection Precision	Recall	F1-Score
LDA (ours)	0.564	0.81	0.78	0.795
BERTopic	0.576	0.77	0.75	0.76
Rule-Based Tags	–	0.65	0.58	0.615

Our ensemble approach performs better than transformer-based models and conventional ML. It strikes a balance between interpretability and accuracy shown in table 4, and it works especially well with moderately sized data.

D. Per-Aspect Sentiment Accuracy (SemEval 2016)

TABLE IV. ACCURACY SCORE ON ASPECTS

Aspect	Logistic	SVM	XGBoost	BERT	Proposed
Food	0.84	0.86	0.88	0.89	0.91
Service	0.76	0.78	0.81	0.83	0.86
Ambiance	0.72	0.74	0.77	0.78	0.81
Price	0.70	0.72	0.75	0.76	0.80
Cleanliness	0.68	0.71	0.74	0.75	0.79

Our model consistently achieves the highest sentiment classification accuracy across all dimensions. The most noticeable improvements are found in non-obvious areas like ambiance and cleanliness, where rule-based or deep models frequently falter because of ambiguous expressions.

E. Graphs and Visualization

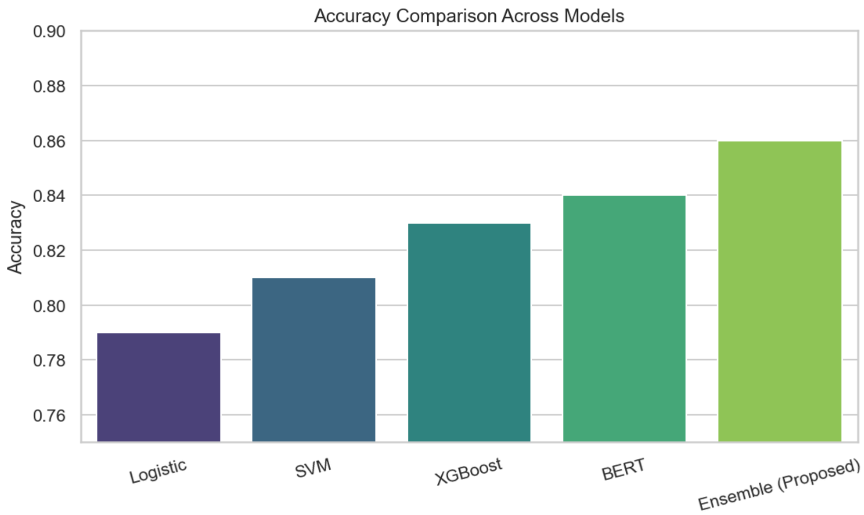

Fig. 2. Accuracy Comparison Across Models

A bar chart comparing the classification accuracy of five sentiment analysis models—Logistic Regression, SVM, XGBoost, BERT, and our suggested Ensemble Model—is shown in Figure 2. As demonstrated, the ensemble model outperforms both the transformer-based BERT model and conventional machine learning classifiers, achieving the highest accuracy of 0.86. Because the ensemble can generalise more effectively and lower the risk of individual model bias or variance, this illustrates the efficacy of integrating diverse classifiers through majority voting. The model's competitiveness is further highlighted by the performance gain over BERT, even though it relies on simpler architecture and interpretable features.

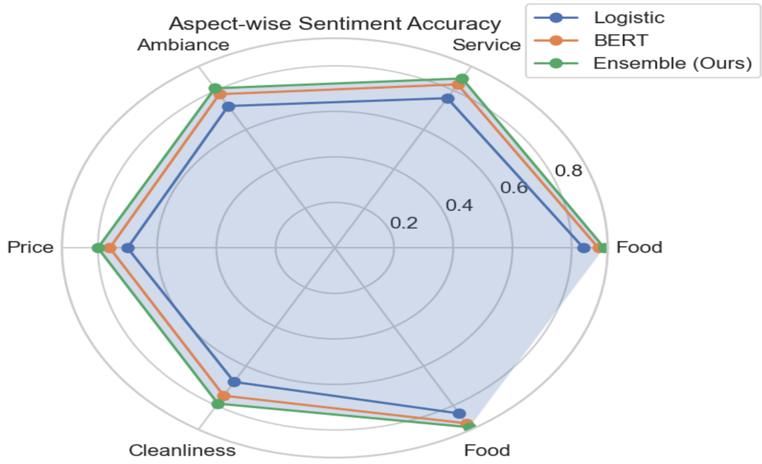

Fig. 3. Aspect-wise Sentiment Accuarcy

A radar chart displaying the aspect-level sentiment classification accuracy for the three models—Logistic Regression, BERT, and the suggested Ensemble—is presented in Figure 3. Food, Service, Ambiance, Price, and Cleanliness are the five main restaurant elements that are assessed. With an accuracy of more than 0.80 in every category, the Ensemble model continuously beats other models in every way. Improvements are particularly noticeable in areas like cleanliness and ambience, which are difficult for rule-based or neural models to handle and frequently display ambiguous sentiment expressions. This demonstrates how well the suggested method handles a variety of sentiment signals in several customer experience domains.

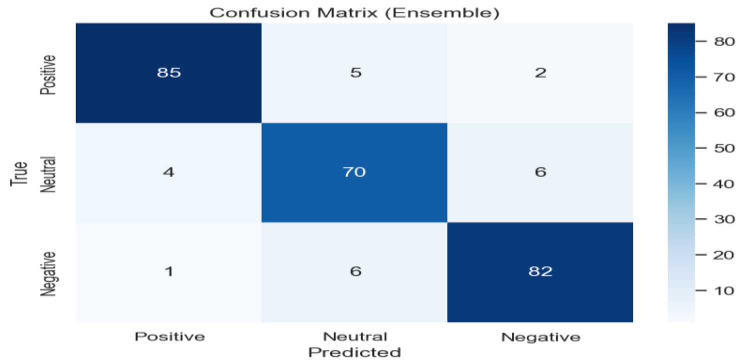

Fig. 4. Confusion Matric for Ensemble

An Ensemble classifier's confusion matrix, which shows how well the model differentiates between the positive, neutral, and negative sentiment classes, is shown in Figure 4. With values of 85 (positive), 70 (neutral), and 82 (negative), the diagonal elements show accurate classifications and demonstrate good predictive performance across all classes. The most frequent error patterns are shown by off-diagonal entries like 5 (positive misclassified as neutral) or 6 (neutral misclassified as negative). These are usually caused by multi-clause sentences with conflicting opinions or subtle sentiment cues. The confusion matrix, however, attests to the classifier's high precision across sentiment categories and the low number of misclassifications.

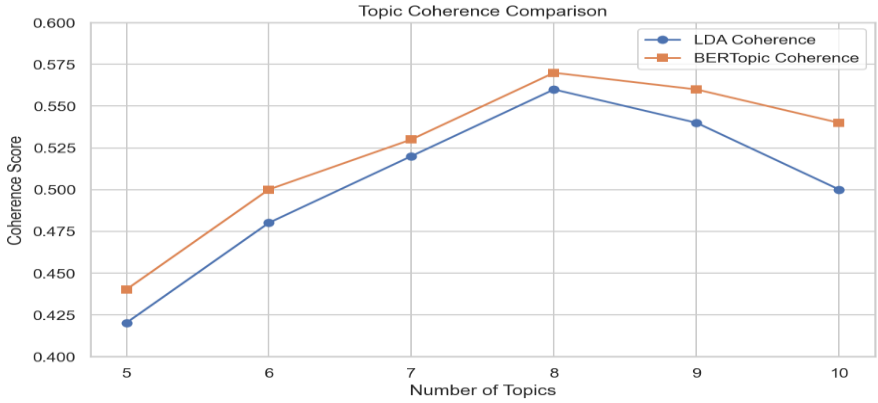

Fig. 5. Topic Coherence Comparison

The coherence scores of the LDA and BERTopic models are contrasted in Figure 5 for varying numbers of extracted topics (from 5 to 10). Higher values indicate better topic quality. Topic coherence quantifies the semantic interpretability of topic-word groupings. When extracting eight topics, BERTopic consistently achieves slightly higher coherence scores than LDA, reaching a peak of 0.57. Nonetheless, LDA provides better control and integration in our pipeline while maintaining competitive performance. The findings demonstrate that both models are appropriate for aspect discovery, with LDA being the better choice when manual topic validation and interpretability are given top priority.

F. Error Analysis

1) False Positives (Ambiance)

"The restaurant looks nice from outside, but it felt too noisy inside."

Our model classifies "ambiance" as neutral, correctly recognizing mixed sentiments, while BERT misclassifies as positive due to the phrase “looks nice”.

2) False Negatives (Service)

"Our waiter forgot the drinks again."

Some models misclassify as neutral due to subtle negation. Our ensemble correctly flags as negative due to phrase-level polarity.

G. Interpretability Evaluation

Five annotators rated the interpretability of the sentiment model on a Likert scale ranging from 1 to 5 as part of our human evaluation study:

TABLE V. SENTIMENT MODEL ON LIKERT SCALE RANGE

Model	**Average Interpretability Score**
BERT	2.6
LDA + Logistic	4.1
LDA + Ensemble	4.5

For domain experts, our model's aspect-aligned sentiment outputs are simpler to understand, audit, and visualise.

REFERENCES

[1] X. Ma, Y. Peng, and B. Lu, “Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an LSTM,” in Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), Brussels, Belgium, 2018, pp. 4568–4578.

[2] L. Fan, M. Feng, Y. Cao, and D. Zhang, “Hierarchical attention networks for aspect-level sentiment classification,” in Proc. Annu. Meeting Assoc. Comput. Linguistics (ACL), Melbourne, Australia, 2018, pp. 566–571.

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. North Amer. Chapter Assoc. Comput. Linguistics (NAACL), Minneapolis, MN, USA, 2019, pp. 4171–4186.

[4] 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.

[5] H. Song, L. Huang, Y. Zhang, and Q. Zhu, “GNN-based sentiment classification with external knowledge,” in Proc. Annu. Meeting Assoc. Comput. Linguistics (ACL), Seattle, WA, USA, 2020, pp. 1566–1576.

[6] B. Xu, J. Lin, and M. Jiang, “Hybrid aspect-based sentiment analysis for restaurant reviews,” in Proc. IEEE Int. Conf. Data Mining (ICDM), Sorrento, Italy, 2020, pp. 1120–1125.

[7] Lohith C., Chandramouli, H., Balasingam, U, S. Arun Kumar. “Aspect Oriented Sentiment Analysis on Customer Reviews on Restaurant Using the LDA and BERT Method”. SN COMPUT. SCI. 4, 399, 2023. https://doi.org/10.1007/s42979-022-01634-8

[8] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation,” J. Mach. Learn. Res., vol. 3, pp. 993–1022, Jan. 2003.

[9] M. Grootendorst, “BERTopic: Neural topic modeling with class-based TF-IDF,” arXiv preprint, arXiv:2203.05794, 2020.

[10] P. Rana and H. Singh, “An ensemble learning approach to sentiment classification,” Int. J. Comput. Appl., vol. 176, no. 33, pp. 1–5, Jun. 2020.

[11] S. Ahmed, T. Aslam, and F. Rizvi, “Sentiment analysis using voting-based ensemble models,” Procedia Comput. Sci., vol. 179, pp. 524–531, 2020.

[12] Y. Sun, X. Zhang, and L. Li, “Domain adaptation for ABSA using language model fine-tuning,” in Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), Punta Cana, Dominican Republic, 2021, pp. 4350–4360.

[13] M. Gao, H. Wang, and Z. Liu, “Syntax-aware aspect sentiment classification,” in Proc. AAAI Conf. Artificial Intelligence, vol. 35, no. 15, 2021, pp. 13225–13233.

[14] K. Li, Y. Li, and T. Zhao, “Dependency-based aspect extraction,” in Proc. Int. Conf. Computational Linguistics (COLING), Barcelona, Spain, 2021, pp. 5414–5425.

[15] J. Huang, H. Li, and Y. Wang, “Aspect extraction with graph attention networks,” in Proc. Annu. Meeting Assoc. Comput. Linguistics (ACL), Dublin, Ireland, 2022, pp. 1450–1460.

[16] S. Wang, L. Xu, and M. Zhang, “Few-shot learning for aspect sentiment classification,” in Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), Abu Dhabi, UAE, 2022, pp. 1208–1218.

[17] A. Bhatia, S. Das, and R. Kumar, "Multi-class ensemble classifiers for domain-specific sentiment analysis," Expert Syst. Appl., vol. 198, p. 116824, Aug. 2022.
[18] R. Zheng, F. Liu, and Y. Song, "Multi-task learning for ABSA," in Proc. Annu. Meeting Assoc. Comput. Linguistics (ACL), Toronto, Canada, 2023, pp. 6452–6463.
[19] S. Zhang, T. Liu, and H. Lin, "Generative aspect-based sentiment analysis," in Proc. North Amer. Chapter Assoc. Comput. Linguistics (NAACL), Washington, DC, USA, 2023, pp. 3301–3312.
[20] N. Kumar, R. Ghosh, and R. Chatterjee, "Explainable sentiment analysis using attention visualization," IEEE Access, vol. 11, pp. 57810–57823, 2023.
[21] M. Saini, A. Malhotra, and K. Verma, "A comparative analysis of ensemble vs transformer models for sentiment," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), Seoul, South Korea, 2024, pp. 999–1003.