A hybrid model for enhanced aspect-based sentiment analysis using large language models

Mohammed Ziaulla, Arun Biradar
School of Computer Science and Engineering, REVA University, Bengaluru, India

Article Info

Article history:

Received Aug 20, 2025
Revised Jan 12, 2026
Accepted Jan 25, 2026

Keywords:

Aspect-based sentiment analysis
BERT
Customer reviews
Gemini
Hybrid BERT
Natural language processing
Sentiment analysis

ABSTRACT

Aspect-based sentiment analysis (ABSA) is a crucial task within natural language processing (NLP), enabling fine-grained opinion mining by identifying sentiments associated with specific aspects of a product or service. While transformer-based models like bidirectional encoder representations from transformers (BERT) have improved sentiment classification, they still struggle with limited contextual adaptability, especially in customer reviews containing complex expressions. Most existing approaches rely heavily on benchmark datasets such as semantic evaluation (SemEval) and multi-aspect multi-sentiment (MAMS), which do not fully capture the diversity of real-world review scenarios. Hence, this research addresses these limitations by proposing a novel hybrid model, called as hybrid-BERT (H-BERT), that integrates span-aware BERT (SpanBERT) with bidirectional long short-term memory (BiLSTM), conditional random field (CRF), and large language models (LLMs). The objective is to enhance aspect extraction and sentiment classification performance using both annotated and synthetic data. The methodology includes preprocessing, hybrid model training, and evaluation using the SemEval 2014 dataset. Experimental results show that H-BERT achieved 90.58% accuracy and 90.56% F-score in the laptop domain and 91.21% accuracy with a 92.03% F-score in the restaurant domain. These results outperform existing models, confirming H-BERT's robustness and effectiveness. In conclusion, H-BERT improves sentiment understanding in customer reviews.

Corresponding Author:

Mohammed Ziaulla
School of Computer Science and Engineering, REVA University
Bengaluru, India
Email: mdziaulla.dr@rediffmail.com

1. INTRODUCTION

The natural language processing (NLP) is a rapidly evolving field in artificial intelligence (AI) that focuses on enabling machines to understand, interpret, and generate human language. With applications ranging from machine translation and question answering to conversational agents and document summarization, NLP serves as a critical bridge between human communication and computational understanding [1], [2]. As digital content continues to grow, the ability to automatically extract insights from text has become increasingly valuable across industries such as e-commerce, healthcare, and customer service. One of the core tasks in NLP is sentiment analysis, which involves identifying the emotional tone or opinion conveyed in a piece of text [3]. Traditional sentiment analysis techniques often classify entire sentences or documents into categories such as positive, negative, or neutral [3]. While effective for general overviews, these methods fall short in scenarios where users express mixed opinions about different components within the same review—for instance, liking the product quality but disliking the delivery service. This limitation gave rise to a more fine-grained technique known as aspect-based sentiment analysis (ABSA).

ABSA aims to identify specific aspects mentioned in a text (e.g., “battery life” and “display quality”) and determine sentiment expressed towards each aspect. This enables a deeper understanding of customer feedback by associating sentiments with particular product or service features. Despite it is growing significance, ABSA remains a challenging task due to the complexity of natural language, including explicit expressions, domain-specific terminology, and multiple sentiment-bearing components in a single sentence [4], [5]. Traditional ABSA models have increasingly employed bidirectional encoder representations from transformers (BERT) due to it is ability to capture deep contextual information [6]–[8]. However, BERT was designed for general-purpose language understanding and is not inherently designed for aspect-level sentiment analysis. BERT often fails to distinguish sentiment polarity when multiple aspects are present, and it requires explicit guidance to focus on aspect-specific content. Additionally, BERT’s pre-training on generic corpora limits its capacity to interpret domain-specific phrases, such as idiomatic expressions found in user reviews. These limitations make BERT-based models less effective for important ABSA tasks.

Another challenge in ABSA research is the reliance on traditional datasets like semantic evaluation (SemEval), multi-aspect multi-sentiment (MAMS), and domain-oriented targeted sentiment analysis (DOTSA), which are limited in size, domain diversity, and real-world variance. Such datasets often do not reflect the wide range of customer opinions found in modern digital platforms. To address this, recent research has explored using large language models (LLMs) to generate synthetic datasets that simulate real-world reviews [9], [10]. While these LLM-generated datasets provide supplemental data, the models used for ABSA still struggle to generalize well across domains and maintain accuracy in identifying aspect-sentiment pairs. To address these gaps, this work proposes a hybrid-BERT (H-BERT) framework for ABSA. H-BERT combines the contextual power of span-aware (SpanBERT) for span-based aspect detection, bidirectional long short-term memory (BiLSTM) for modeling sequential dependencies, conditional random fields (CRFs) for structured output, and LLMs for auxiliary supervision. This multi-component architecture enhances the ability to extract aspect terms, determine their associated polarities, and classify sentence-level sentiments. Additionally, the use of both human-annotated (SemEval) and synthetic datasets (ChatGPT and Gemini) improves robustness and generalization of the model in customer review scenarios. The contributions of the work are as follows.

A novel hybrid model integrating SpanBERT, BiLSTM, CRFs, and LLMs for improved ABSA is presented in this work. Synthetic customer review datasets were generated using ChatGPT-3.5-Turbo and Gemini-2.5-Flash to supplement traditional datasets and enhance generalization. A preprocessing step is included for ensuring consistency across datasets and preparing data for deep contextual and sequential modeling. The model jointly performs aspect extraction, aspect polarity detection, and sentence-level sentiment classification. The model was evaluated using accuracy and macro-F-score on both traditional (SemEval) and LLM-generated datasets, demonstrating improved performance over standard BERT-based methods. The manuscript is organized in the following manner. Section 2 presents literature survey which discusses existing ABSA approaches and LLM synthetic data generation approaches. Section 3 presents the methodology for the H-BERT model, Section 4 presents the results of H-BERT and compares with existing approaches. Section 5 presents the conclusion and future work of H-BERT.

2. LITERATURE SURVEY

This section discusses existing ABSA approaches, LLM approaches, and LLM approaches used for generating synthetic datasets. Gu *et al.* [11] aimed at enhancing aspect-level sentiment-analysis by addressing limitations in existing graph convolutional networks (GCN) based approaches, like insufficient utilization of aspect-specific information and lack of external sentiment knowledge. Hence, presented syntax-aware graph convolutional network (SAGCN), which integrated aspect-level feature into contextual representations and incorporated external sentiment lexicons for enriching sentiment perception. This work also employed multi-head self-attention (MHSA) approach along with a point-wise convolutional-transformer (PCT) for jointly capturing semantic-syntactic relationships. For evaluation of SAGCN, three datasets, i.e., ACL 14-Task Twitter dataset and SemEval2014 restaurant and SemEval2014 laptop dataset were considered, where achieved 77.97%, 87.53%, and 83.06% accuracy. Jeong and Lee [12] aimed at enhancing aspect-based analysis of hotel review by utilizing ChatGPT, an LLM model, for overcoming challenges in interpretation of ambiguous and complex customer feedback. In this work, they utilized TripAdvisor dataset, where their approach involved identification of ten key hotel attributes and generation of aspect-summarization pairs having designed prompts for efficient analysis. The ChatGPT’s outputs were evaluated qualitatively, focusing on the ability of extracting explicit keywords and summarizing sentiment-rich content. The findings showed that ChatGPT captured important information and distinguished sentiment patterns across hotel categories. The results show improved accuracy and contextual understanding.

Zhang *et al.* [13] focused on improving ABSA by addressing limitation in existing graph-based and attention-based model, hence, proposed syntactic-dependency graph convolutional network (SD-GCN), which aimed at capturing long-range syntactic-relationships and dependency between opinion-words and aspect-terms. This work utilized Biaffine-Attention, where constructed syntactic-dependency graphs for representing connection among sentiment expressions and aspects. Further, GCN was applied for extracting rich syntactic-semantic features. Evaluations were conducted on SemEval 2014 Restaurant, SemEval 2014 Laptop, SemEval 2015 Restaurant, and Twitter dataset, where achieved 88.14%, 80.35%, 85.30%, and 77.63% accuracy respectively. Mughal *et al.* [14] aimed at addressing challenges in ABSA, mainly in data dependency, sensitivity and limited usage of LLMs. Unlike traditional sentiment-based approaches, which mainly focused on document or sentiment level, ABSA links sentiments to specific aspects. In this work, they evaluated various deep learning (DL) and LLM models, which included generative-pre training-3.5-turbo (GPR-3.5-Turbo), pathways-language models (PaLM), decoding enhanced bidirectional-encoder-representation transformers (DeBERTa), fine-tuned language-T5 (FLAN-T5) and attention-based aspect-extraction long short-term memory (ATAE-LSTM), considering SemEval2016, MAMS, and DOTSA, where DeBERTa showed better performance, while PaLM showed better performance in aspect-term sentiment analysis.

Falatouri *et al.* [15] investigated effectiveness of LLM in enhancement of service-quality and sentiment-analysis dimension extraction from user-generated content. The primary objective of this work was to evaluated Claude3 and ChatGPT-3.5 against three NLP approaches using bilingual customer review datasets in Persian and English. The methodology involved comparing model performance on sentiment-classification and structure information extraction. The results showed that ChatGPT achieved 76% accuracy and substantial agreement with human raters, whereas Claude3 achieved 68% accuracy with moderate agreement. Despite outperforming traditional approaches, both LLMs showed inconsistencies in fine-grained extraction. Liu *et al.* [16], focused on improving aspect-opinion sentiment triplet extraction for sentiment analysis by addressing limitation of traditional pipeline and tagging-based approaches, which often fail for capturing deep syntactic-semantic relationships, hence, presented syntactic-semantic aspect-sentiment term-extraction (SynSem-ASTE), a multi-encoder approach which integrated syntactic-semantic encoding for capturing structural and contextual dependencies. The approach also incorporated grid-tagging approach for enabling effective triplet extraction. SynSem-ASTE was evaluated on SemEval 2016, SemEval 2015, and SemEval 2014 datasets, where achieved macro-F-scores up to 72.23%, showing improved extraction capabilities.

Hellwig *et al.* [17] investigated application of LLMs, mainly large language model meta AI-3-70B (Meta Llama-3-70B) and GPT-3.5-Turbo for generating annotated data for ABSA, addressing challenge of limited labeled datasets. The approach involved few-shot prompting for creating synthetic training data under two low-resource setting with 25 and 500 manual-labeled examples. Using datasets for aspect-category sentiment-analysis (ACSA) and aspect category detection (ACD), the findings showed that using 25 labeled examples, F1-score reached 81.33% and 71.71% for ACSA. In 500-example setting, synthetic augmentation further improved ACSA performance from 84.54% to 86.70%. Pandey and Singh [18] addressed limitation of traditional product reviews by presenting a framework which generated detailed textual reviews from user-provided aspect-wise ratings using LLMs. The objective of this work was to enhance completeness and quality of online reviews, which often lack coverage of aspects. The methodology of the work involved mapping structure Likert-Scale rating for coherent, aspect-rich narratives. The work was evaluated considering human judgement, AI-generated reviews, which demonstrated high relevance, readability and informativeness, which was often indistinguishable from human-written content. The findings showed framework's potential in improving e-commerce review systems.

Fan *et al.* [19] aimed at improving Multimodal aspect-based sentiment classification (MABSC) by addressing limitation of existing approaches for capturing context from social-media posts. For enhancing sentiment prediction, this work presented multi-modal dual cause analysis (MDCA) architecture, which introduces two explanatory components, which included direct-cause (DC) and reasoning-cause (RC), for identifying underlying motivations behind user sentiments. Using LLMs and visual-language models, MABSC dataset enriched with RC and DC labels were constructed. A multi-task learning benchmark MABSC datasets show that MDCA outperforms existing approaches. Zhao *et al.* [20] addressed limitation in DL models for ABSA, particularly their difficulty in adapting to varying sentiment classes and modeling category-specific information, hence presented BERT approach with class-aware work MHSA (BERT-CA-WMA), which introduced aspect-projection layer for aligning aspect-embedding with contextual representation and incorporated sentiment class information using dynamic attention weighting. The BERT-CA-WMA approach was evaluated using SemEval 2016, SemEval 2015, and

SemEval 2014, across laptop and restaurant domains, where achieved high accuracy of 90.54%, showing good performance and effective sentiment classification.

From above literature survey, it is seen that several recent studies have proposed innovative approaches to improve ABSA, yet each exhibits certain limitations. For instance, Gu *et al.* [11] introduced SAGCN by integrating syntax and sentiment knowledge but relied heavily on external lexicons, limiting adaptability to informal or domain-specific reviews, Jeong and Lee [12] utilized ChatGPT for aspect analysis but focused only on qualitative outputs without quantitative benchmarking, making scalability uncertain. Further, Zhang *et al.* [13] proposed SD-GCN to capture syntactic dependencies but overlooked semantic inconsistencies that often occur in user-generated content. Mughal *et al.* [14] compared LLMs and DL models for ABSA but highlighted performance inconsistency across datasets, showing a lack of domain robustness, Falatouri *et al.* [15] found that although LLMs like ChatGPT performed well, they lacked precision in fine-grained extraction. Liu *et al.* [16] addressed structural dependencies using SynSem-ASTE but required complex multi-encoder setups, making the model computationally heavy. Similarly, Hellwig *et al.* [17] generated synthetic data using LLMs, yet relied on small annotated sets, potentially introducing bias. These limitations are addressed in the proposed H-BERT framework, which combines SpanBERT for span detection, BiLSTM for sequential context, CRFs for structured prediction, and LLMs for weak supervision. This unified approach enhances aspect detection and polarity classification across both synthetic and existing datasets.

3. METHOD

This section begins by presenting the overall architecture of the proposed system, followed by a detailed discussion of the datasets used and the preprocessing steps applied. It then outlines the limitations of traditional BERT models, introduces the proposed H-BERT framework, and concludes with the performance metrics used for evaluation. The architecture of the complete work is presented in Figure 1, which shows a comprehensive framework for ABSA using H-BERT approach. In this architecture, first the dataset is considered, i.e., SemEval 2014 Task 4 dataset which comprises of restaurant and laptop. In this work, two more datasets were created using LLMs, i.e., using ChatGPT-3.5-Turbo and Gemini-2.5-Flash which is discussed in detail in sub-section 3.2. These datasets separately go through preprocessing, where the null values are checked and the review sentence is tokenized. The complete preprocessing steps are discussed in detail in sub-section 3.3. Further, the pre-processed text is passed on to the H-BERT model, which is discussed in detail in sub-section 3.4. The main aim of H-BERT is to extract aspects, extract aspect-term polarity and predicting overall review sentence sentiment polarity. Parallel to this process, this work has also used LLMs (ChatGPT-3.5-Turbo and Gemini-2.5-Flash) for extracting aspects, extract aspect-term polarity and predicting overall review sentence sentiment polarity, which is discussed in detail in section 4. Finally, the performance of the H-BERT and LLMs is evaluated using two standard metrics used for evaluating ABSA, i.e., accuracy and macro-F-score which is discussed in detail in sub-section 3.6. This architecture combines machine learning (ML), DL, and rule-based approaches for providing accuracy sentiment insights.

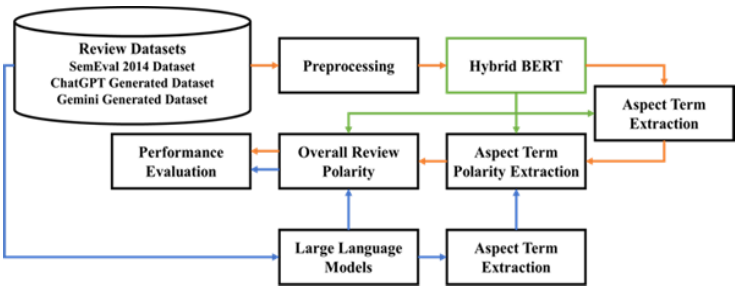

Figure 1. Proposed architecture of the work

This section discusses the datasets used in this work for evaluating H-BERT model. In this work, for evaluation of H-BERT, three distinct ABSA datasets were used. The first dataset was SemEval 2014 Task 4 dataset, which is open accessible and can be downloaded from in [21], [22]. The dataset consists of customer reviews primarily from laptop and restaurant domains. In this dataset, every aspect was labelled with one of three polarities, i.e., neutral, negative, and positive. The second and third dataset were generated using ChatGPT-3.5-Turbo and Geminin-2.5-Flash, similar to the work presented in [12], [17], [18]. The datasets were created using prompting approach, which is presented in Table 1. Both the LLMs were prompted with same context and instruction, i.e., to create 50 synthetic customer review entries in .csv format. In the .csv file, each entry included a review sentence, up to three aspects with corresponding sentiment polarities and overall sentiment classification. The output files feature columns such as ID, review, aspect 1, aspect 1 polarity, aspect 2, and aspect 2 polarity. These synthetic datasets provide valuable supplemental data for evaluating model robustness and generalization across both human-annotated and machine-generated reviews. The complete flow of the dataset collection using ChatGPT and Gemini is presented in Figure 2.

Table 1. Prompt used for querying Gemini and ChatGPT

Context	Generate a synthetic dataset of customer reviews with multiple aspect-level sentiments and overall polarity
Instruction	Create 50 review entries in .csv format where each entry includes a review sentence, multiple aspects mentioned in the sentence with their respective sentiment polarity (positive/negative/neutral), and the overall sentiment polarity of the review.
Output	CSV file with columns: ID, review, aspect 1, aspect 1 polarity, aspect 2, aspect 2 polarity, aspect 3, aspect 3 polarity, overall sentiment

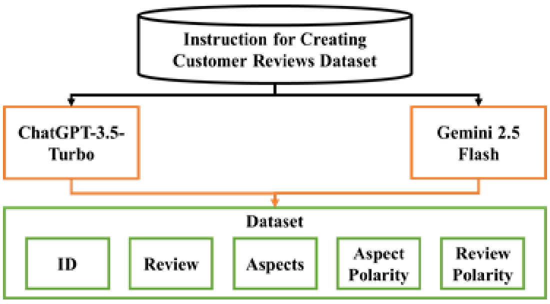

Figure 2. Process of dataset creation

In this work, the datasets went through preprocessing which is important for consistency before they are passed on to the H-BERT. Each dataset, i.e., SemEval 2104, ChatGPT dataset, and Gemini dataset were processed independently for preserving structural integrity and manage variations in data format in [23], [24]. The first step in preprocessing involved checking for null or missing values in review sentences, aspect terms, and sentiment polarities. Any entry containing null values were removed, providing clean, and reliable dataset. Following this, the review sentences were tokenized, which is important for H-BERT model. This process enables the hybrid model to understand semantic structure of each review and facilitate aspect extraction and sentiment analysis at granular level. Also, text normalization steps, which included lower casing and unnecessary whitespace were applied for maintaining uniformity across all dataset. This step ensured consistent input formats, providing accurate and robust sentiment predictions [25].

BERT is a pre-trained language representation model developed by google, which has revolutionized NLP tasks, including sentiment analysis. The BERT's architecture is based on Transformer model, which utilizes self-attention for capturing contextual relationship among words in a sentence. Unlike traditional word-embedding approaches like global-vectors (GloVe) and word-to-vector (Word2Vec), BERT approach is bidirectional, meaning it considers both right and left contexts of word simultaneously during training. This provides BERT to generate context-aware embeddings which provide better performance on tasks like sentiment classification. In sentiment analysis, BERT is being widely used for aspect-level sentiment and sentence-level classification because of its understanding of linguistic context. Nevertheless, having good performance, BERT models exhibit several limitations when applied to fine-grained tasks like ABSA. One major drawback is that BERT is designed for general language understanding and fails to inherently focus on specific aspect-terms in sentence unless explicitly guided in [26], [27]. Hence, it fails to distinguish sentiments associated with multiple aspects in same review. Additionally, BERT's pre-training on generic corpora limits BERT ability for accurately interpreting domain-specific sentiment cues, like idiomatic expression or domain-specific terms. Moreover, fine-tuning BERT for ABSA requires significant computation resources and careful hyperparameter tuning. These limitations highlight need for hybrid of modified model which incorporates BERT's contextual strengths while address shortcoming in aspect-oriented sentiment tasks. Hence, in this work, a H-BERT model is presented for solving the following issues.

The architecture of the H-BERT is presented in Figure 3, which starts by taking input, i.e., pre-processed review texts (tokens and indices). These reviews are passed onto the H-BERT which consists of SpanBERT, a pre-trained BERT model which is designed for better representing and predicting spans of text, which in this work is utilized for capturing aspect spans in [28]. The output of SpanBERT is passed on to the BiLSTM for sequential contextualized learning, further passed on to dropout for regularization. After dropout, a linear-layer is added, which acts as fully-convolutional layer (FCL). The outputs from FCL are passed on to CRFs for structured sequence prediction and LLMs (ChatGPT-3.5-Turbo and Gemini-2.5-Flash) to support generalization. The processed embeddings are then classified using SoftMax classifier into three tasks, i.e., aspect-extraction, aspect-polarity, and sentence-polarity.

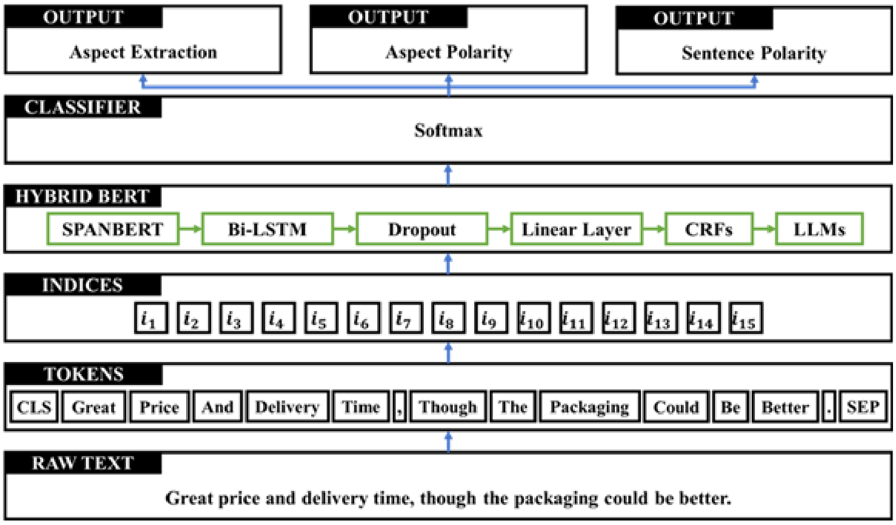

Figure 3. H-BERT architecture

SpanBERT is a span-based extension of BERT, which is specifically trained for predicting spans of text rather than individual masked tokens. Consider a tokenized sentence as $S = \{w_1, w_2, \dots w_n\}$ be input to SpanBERT. The model encodes contextual embeddings as $E = \{e_1, e_2, \dots, e_n\}$ and let d denote embedding dimension. Hence, from this, the SpanBERT captures representations as $s_{i,j}$ as presented in (1).

$$s_{i,j} = [e_i; e_j; e_i \odot e_j; Pooling(e_i, \dots, e_j)] \quad (1)$$

In (1), $[\cdot]$ denotes concatenation, $\odot$ denotes element-wise multiplication, and $Pooling$ denotes mean pooling over span (i, j). When compared with BERT approaches, which focusses on masked-token prediction, the SpanBERT is fine-tuned on masked-span prediction, making it more suitable for aspect extraction where multi-tokens spans need identification. The SpanBERT provides understanding of syntactic-semantic dependencies within spans, providing improved ABSA. To capture the sequential dependency of sentence, this work has utilized BiLSTM which is discussed in detail in next section.

For capturing sequential-dependency, a BiLSTM is used on output of SpanBERT. Given input embeddings $E = \{e_1, e_2, \dots, e_n\}$, the BiLSTM computes backward and forward hidden states using (2) and (3). The final representation at each token is represented as (4).

$$\vec{h}_t = LSTM(e_t, \vec{h}_{t-1}) \quad (2)$$

$$\overleftarrow{h}_t = LSTM(e_t, \overleftarrow{h}_{t+1}) \quad (3)$$

$$h_t = [\vec{h}_t; \overleftarrow{h}_t] \in \mathbb{R}^{2h} \quad (4)$$

In (4), h denotes size of LSTM hidden units. The BiLSTM adds temporal understanding by utilizing both future and past context, which is important for polarity disambiguation in phrases like "thought the packaging could be better". Once the BiLSTM has encoded the complete sequence, the hidden representations $\{h_1, h_2, \ldots, h_n\}$ are passed through dropout layer for regularization. The dropout is a stochastic regularization approach where during training, a fraction p of output units are randomly set to zero with probability $p \in [0,1]$. This helps preventing overfitting by not allowing the model from becoming too reliant on specific neurons. The process is mathematical represented using (5). Following dropout, the representations are passed into linear layer, i.e., FCL. This layer performs an affine transformation which projects BiLSTM outputs into new space for classification using (6).

$$h'_t = Dropout(h_t), where\ h'_t \in \mathbb{R}^{2h} \tag{5}$$

$$z_t = Wh'_t + b \tag{6}$$

In (6), W denotes weight matrix and b denotes bias vector. The FCL acts as bridge among deep sequential features extracted by BiLSTM and higher-level classification which later performed by CRFs and SoftMax, discussed in next sections respectively. By learning linear combinations of hidden features, this work prepares data for structured sequence decoding and final sentiment decision-making.

For ensuring structured prediction of aspect terms and labels, this work has used CRFs on top of the BiLSTM outputs. Let $H = \{h_1, h_2, \ldots, h_n\}$ denote hidden states from BiLSTM and $Y = \{y_1, y_2, \ldots, y_n\}$ denote the sequence of predicted tags. In this work, the CRF layer models the conditional probability using (7). In (7), A denotes transition-matrix, $P_t(y_t)$ denotes score from linear layer for tag y_t at position t. The CRFs ensure that output sequences are valid and semantically consistent.

$$P(Y|H) = \frac{exp\left(\sum_{t=1}^{n}\left(A_{y_{t-1},y_t} + P_t(y_t)\right)\right)}{\sum_{\tilde{Y}} exp\left(\sum_{t=1}^{n}\left(A_{\tilde{y}_{t-1},\tilde{y}_t} + P_t(\tilde{y}_t)\right)\right)} \tag{7}$$

The H-BERT framework also incorporates LLMs (ChatGPT-3.5-Turbo and Gemini-2.5-Flash) as auxiliary support for knowledge integration. The LLMs in this work are utilized for weak supervision, enhancing model's generalization. Consider R be a review passed to LLM. The output from the LLMs achieved is as presented in (8). The outputs from LLMs are used for fine-tuning the final prediction outcome, enhancing hybrid model's robustness. The final representation from LLM is passed to SoftMax classifier for multi-task learning, i.e., aspect extraction, aspect polarity, and sentence-level sentiment. The SoftMax probability for class c at position t is defined using (9).

$$LLM(R) = \{Aspect_i, Polarity_i, Overall\ Polarity\} \tag{8}$$

$$P(y_t = c|h_t) = \frac{\exp(W_c^T h_t + b_c)}{\sum_{j=1}^{C} \exp\left(W_j^T h_t + b_j\right)} \tag{9}$$

In (9), W_c and W_j denotes weight for class c and j, b_c and b_j denotes bias and C denotes number of classes, i.e., positive, negative, and neutral. The SoftMax layer ensures probabilistic and interpretable outputs for each token and overall sentence sentiment.

4. PERFORMANCE EVALUATION

For evaluation of sentiment classification using H-BERT, this work utilized accuracy and macro-F-score. The accuracy measures proportion of correctly predicted labels over total number of predictions. It is evaluated using (10). In ABSA sentiment classification, the macro-F-score is important metrics as it calculates F1-score for each class independently and then averages them, providing equal weight for all classes regardless of their frequencies. It is evaluated using (11). In (11), C denotes class. The metrics provide ABSA evaluation. The performance of the H-BERT is evaluated in the next section and discussed in detail.

$$Accuracy = \frac{Total\ Number\ of\ Predictions}{Number\ of\ Correct\ Predictions} \tag{10}$$

$$Macro - F1 = \frac{1}{C}\sum_{c=1}^{C} \frac{2 \times Precision, Recall}{Precision + Recall} \tag{11}$$

5. RESULTS AND DISCUSSION

The H-BERT model was implemented and tested on a Windows 11 system featuring an AMD Ryzen 5 processor, 16 GB RAM, and a 4 GB NVIDIA GeForce GTX 1650 GPU. Development was carried out using Python within a Python 3.11 environment. Table 2 displays the sample distribution of the SemEval 2014 dataset used for evaluation. Sample reviews generated by ChatGPT and Gemini are shown in Tables 3 and 4, respectively, while samples from the SemEval 2014 dataset are provided in Table 5.

This section presents the performance of the H-BERT model in accurately identifying aspect terms and their associated sentiment polarities. The model was evaluated on a range of review samples, and the predicted results were compared with the actual aspect-polarity annotations. As shown in Table 6, H-BERT demonstrated a strong ability to correctly extract multiple aspects and their sentiments within complex and multi-sentiment sentences. For example, in sentences containing both positive and negative sentiments about different aspects, H-BERT was able to distinguish and classify them appropriately. While the model achieved high accuracy in several instances, minor mismatches were observed in a few predictions, such as classifying "service" as neutral instead of positive. These discrepancies highlight the challenges of contextual importance in sentiment classification. The results indicate that H-BERT is effective in handling multi-aspect sentiment analysis and offers robust performance in aspect extraction tasks across both general and nuanced review contexts.

Table 2. SemEval 2014 dataset

Dataset	Train			Test		
	Negative	Neutral	Positive	Negative	Neutral	Positive
Laptop	870	464	994	128	169	341
Restaurant	807	637	2164	196	196	728

Table 3. Samples of ChatGPT generated reviews

ID	Review	Aspect 1	Aspect 1 polarity	Aspect 2	Aspect 2 polarity	Aspect 3	Aspect 3 polarity	Overall sentiment
1	The battery life is excellent but the screen quality is poor and the customer support was average.	Battery life	Positive	Screen quality	Negative	Customer support	Neutral	Neutral
2	I love the camera and the design, but the performance is sluggish.	Camera	Positive	Design	Positive	Performance	Negative	Neutral
3	The food was delicious, the service was quick, but the ambiance was lacking.	Food	Positive	Service	Positive	Ambiance	Negative	Positive
4	Great price and delivery time, though the packaging could be better.	Price	Positive	Delivery time	Positive	Packaging	Neutral	Positive
5	The staff was rude, the room was dirty, and the check-in process was chaotic.	Staff	Negative	Room	Negative	Check-in process	Negative	Negative

Table 4. Samples of Gemini generated reviews

ID	Review	Aspect 1	Aspect 1 polarity	Aspect 2	Aspect 2 polarity	Aspect 3	Aspect 3 polarity	Overall sentiment
1	The food was amazing, especially the pasta, but the service was a bit slow.	Food Quality	Positive	Service	Negative	-	-	Positive
2	Great ambiance and comfortable seating. The coffee was decent, but nothing special.	Ambiance	Positive	Coffee quality	Neutral	-	-	Positive
3	The delivery was super fast, however, the pizza arrived cold and soggy.	Delivery speed	Positive	Food quality	Negative	-	-	Negative
4	Friendly staff and a wide variety of menu options. Prices are a bit high though.	Service	Positive	Menu variety	Positive	Price	Negative	Positive
5	The wait time for a table was excessive, but once seated, the waiter was very attentive.	Wait time	Negative	Service	Positive	-	-	Neutral

Table 5. Samples of SemEval 2014 Task 4 dataset

ID	Review	Aspect 1	Aspect 1 polarity	Aspect 2	Aspect 2 polarity	Aspect 3	Aspect 3 polarity	Aspect 4	Aspect 4 polarity	Overall sentiment
1	I took it back for an Asus and same thing-blue screen which required me to remove the battery to reset.	B	N	-	-	-	-	-	-	N
2	When I finally had everything running with all my software installed, I plugged in my droid to recharge and the system crashed.	S	N	Syst	Neg	-	-	-	-	Neg
3	However, the multi-touch gestures and large tracking area make having an external mouse unnecessary (unless you're gaming).	MTG	P	TA	P	EM	N	G	N	P
4	The food is uniformly exceptional, with a very capable kitchen which will proudly whip up whatever you feel like eating, whether it's on the menu or not.	F	P	K	P	M	N	-	-	P
5	They did not have mayonnaise, forgot our toast, left out ingredients (i.e. cheese in an omelet), below hot temperatures and the bacon was so over cooked it crumbled on the plate when you touched it.	T	Neg	Mn	Neg	B	N	C	N	Neg
6	The seats are uncomfortable if you are sitting against the wall on wooden benches.	S	Neg	-	-	-	-	-	-	Neg

Notes: B = battery, N = neutral, S = software, Syst = system, Neg = negative, MTG = multi-touch gestures, P = positive, TA = tracking area, EM = external mouse, G = gaming, F = food, M = menu, T = toast, Mn = mayonnaise, B = bacon, C = cheese, and S = seats

Table 6. H-BERT aspect extraction

Samples	Actual aspects (polarity)	Predicted aspects (polarity)
The battery life is excellent but the screen quality is poor and the customer support was average.	Battery life (positive), screen quality (negative), customer support (neutral)	Battery life (positive), screen quality (negative), customer support (neutral)
The food was delicious, the service was quick, but the ambiance was lacking.	Food (positive), service (positive), ambiance (negative)	Food (positive), service (positive), ambiance (negative)
Friendly staff and a wide variety of menu options. Prices are a bit high though.	Service (positive), menu (neutral), variety (positive), price (negative)	Service (positive), price (neutral)
The wait time for a table was excessive, but once seated, the waiter was very attentive.	Wait time (negative), service (positive)	Service (neutral)
I took it back for an Asus and same thing-blue screen which required me to remove the battery to reset.	Battery (neutral)	Battery (neutral)
The food is uniformly exceptional, with a very capable kitchen which will proudly whip up whatever you feel like eating, whether it's on the menu or not.	Food (positive), kitchen (positive), menu (neutral)	Food (positive), menu (neutral)

This section evaluates the performance of the H-BERT model in determining the overall sentiment polarity of customer reviews. Table 7 compares the predicted review polarities with the actual labels. The results show that H-BERT performs well in most cases, correctly identifying the overall sentiment as positive or neutral. However, there are instances where the model slightly misjudges the review's tone. For example, in the first sample, where both positive and negative aspects are present, H-BERT classified the review as negative, while the actual label was neutral. Similarly, in another case, a review labeled positive was predicted as neutral, likely due to the mention of a negative aspect such as high pricing. These variations suggest that while H-BERT effectively handles general sentiment classification, it occasionally struggles with balancing conflicting sentiments across multiple aspects. Despite this, the model demonstrates strong potential in capturing overall sentiment in customer feedback in [29], [30].

This section highlights the classification performance of the H-BERT model, evaluated using accuracy, and macro-F-score on SemEval 2014 dataset across both laptop and restaurant domains. As illustrated in Figure 4, H-BERT achieved an accuracy of 90.58% and macro-F-score of 90.56% in the laptop domain, and an even higher accuracy of 91.21% and macro-F-score of 92.03% in the restaurant domain. These metrics indicate that the hybrid approach significantly improves both precision and recall in sentiment classification tasks. The inclusion of LLM-generated datasets appears to enhance the model’s generalization capabilities, allowing it to perform consistently across domains.

Table 7. H-BERT review polarity outcome

Samples	Actual review polarity	Predicted review polarity
The battery life is excellent but the screen quality is poor and the customer support was average.	Neutral	Negative
The food was delicious, the service was quick, but the ambiance was lacking.	Positive	Positive
Friendly staff and a wide variety of menu options. Prices are a bit high though.	Positive	Neutral
The wait time for a table was excessive, but once seated, the waiter was very attentive.	Neutral	Neutral
I took it back for an Asus and same thing- blue screen which required me to remove the battery to reset.	Neutral	Neutral
The food is uniformly exceptional, with a very capable kitchen which will proudly whip up whatever you feel like eating, whether it's on the menu or not.	Positive	Positive

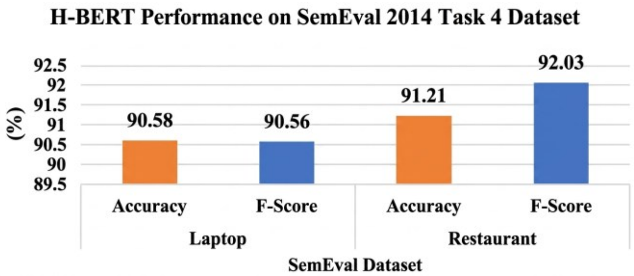

Figure 4. H-BERT performance evaluation on SemEval 2014 dataset

A comparative analysis was conducted to evaluate the performance of proposed H-BERT model against several existing state-of-the-art approaches in ABSA, as presented in Table 8. Traditional graph-based models such as SAGCN [11] and SD-GCN [13] achieved moderate accuracy and macro-F-score, with SAGCN scoring 83.06% accuracy and 79.69% macro-F-score in the laptop domain. SynSem-ASTE [16] and its transformer-based variants like BERT, robustly optimized BERT pretraining approach (RoBERTa), and DeBERTa integration showed lower macro-F-scores, particularly in fine-grained ABSA tasks, with macro-F-score ranging from 59.73% to 74.73%. BERT-CA-WMA [20] performed better, achieving an accuracy of 89.72% and macro-F-score of 89.81% in the laptop domain. However, the proposed H-BERT outperformed all models, achieving 90.58% accuracy and 90.56% macro-F-score for laptops, and 91.21% accuracy with 92.03% macro-F-score for restaurants. The integration of SpanBERT, BiLSTM, CRF, and LLM-generated data makes H-BERT more robust and effective in handling both human-annotated and synthetic review data, leading to superior performance in ABSA.

This section presents the ablation study which investigates the impact of incorporating LLM-generated data (from ChatGPT and Gemini) into the H-BERT framework for both aspect extraction and review polarity tasks. As shown in Table 9, a qualitative comparison across models highlights that while ChatGPT and Gemini perform well in identifying most aspects and polarities, they occasionally misclassify the overall sentiment or fail to extract all relevant aspects. For instance, in the third review, both ChatGPT and Gemini correctly identified several aspects but misjudged the polarity of "price," while H-BERT identified fewer aspects but aligned better with the actual sentiment polarity.

Further, Table 10 presents a detailed quantitative comparison of the experimental results, summarizing the performance of H-BERT under different data-augmentation settings. The base H-BERT model, trained without any integration of LLMs, already demonstrates strong performance, indicating that the architecture itself is effective for ABSA. However, small but consistent improvements are observed when either ChatGPT-generated or Gemini-generated data are added independently, suggesting that synthetic data from each LLM contributes positively to the model's learning. The best overall performance is achieved when both ChatGPT- and Gemini-generated data are combined and used together, with H-BERT attaining the highest accuracy and macro-F-scores (90.58% and 90.56% for laptops, 91.21% and 92.03% for restaurants). This result clearly confirms that incorporating multiple LLMs within the H-BERT framework not only improves the robustness and accuracy of ABSA but also enhances the model's generalization capability across diverse review formats, including varying writing styles, domain-specific expressions, and different levels of polarity.

Table 8. comparative study

Ref	Model	Accuracy	Macro-F-score	Accuracy	Macro-F-score
[11]	SAGCN	83.06	79.69	87.53	81.28
[13]	SD-GCN	80.35	75.87	88.14	81.42
[16]	SynSem-ASTE	-	59.73	-	72.23
	BERT+SynSem-ASTE	-	59.73	-	72.23
	RoBERTa+SynSem-ASTE	-	63.13	-	74.67
	DeBERTa+SynSem-ASTE	-	63.37	-	74.73
[20]	BERT-CA-WMA	89.72	89.81	90.54	91.49
Proposed	H-BERT	90.58	90.56	91.21	92.03

Table 9. Qualitative comparison of H-BERT with ChatGPT and Gemini

Samples	Actual aspects (polarity)	Actual review polarity	ChatGPT aspect extraction (polarity)	ChatGPT review polarity	Gemini aspect extraction (polarity)	Geminin review polarity	H-BERT aspect extraction (polarity)	H-BERT review polarity
The battery life is excellent but the screen quality is poor and the customer support was average.	Battery life (positive), screen quality (negative), customer support (neutral)	Neutral	Battery life (Positive), Screen quality (Negative), Customer support (Neutral)	Neutral	Battery life (positive), screen quality (negative), customer support (neutral)	Negative	Battery life (positive), screen quality (negative), customer support (neutral)	Negative
The food was delicious, the service was quick, but the ambiance was lacking.	Food (positive), service (positive), ambiance (negative)	Positive	Food (Positive), Service (Positive), Ambiance (Negative)	Positive	Food (positive), service (positive), ambiance (negative)	Positive	Food (positive), service (positive), ambiance (negative)	Positive
Friendly staff and a wide variety of menu options. Prices are a bit high though.	Service (positive), menu (neutral), variety (positive), price (negative)	Positive	Staff (Positive), Menu options (Positive), Prices (Negative)	Positive	Staff (positive), menu options (positive), prices (negative)	Positive	Service (positive), price (neutral)	Neutral
The wait time for a table was excessive, but once seated, the waiter was very attentive.	Wait time (negative), service (positive)	Neutral	Wait time (Negative), Waiter attentiveness (Positive)	Neutral	Wait time (negative), waiter (positive)	Neutral	Service (neutral)	Neutral
I took it back for an Asus and same thing- blue screen which required me to remove the battery to reset.	Battery (neutral)	Neutral	Product reliability (Negative), Battery reset (Negative)	Negative	Blue screen (negative), battery (negative)	Negative	Battery (neutral)	Neutral

Table 10. H-BERT performance study using ChatGPT and Gemini

	Laptop		Restaurant	
Model	Accuracy	Macro-F-score	Accuracy	Macro-F-score
H-BERT without LLM	90.51	90.49	91.14	92.1
H-BERT+ChatGPT	90.54	90.52	91.17	92.07
H-BERT+Gemini	90.53	90.51	91.16	92.08
H-BERT+ChatGPT+Gemini	90.58	90.56	91.21	92.03

6. CONCLUSION

This study explored the field of NLP, focusing on ABSA, a critical task in understanding fine-grained customer feedback. Traditional BERT-based models, although widely used, face challenges in accurately extracting aspect-level sentiments, especially in the presence of ambiguous language or domain-specific expressions. A significant research gap exists in their limited ability to generalize across diverse review structures and the reliance on static datasets such as SemEval 2014, MAMS, and others. These datasets often lack the variability and contextual depth of real-world customer reviews, making existing models less robust. The core objective of this research was to address these limitations by proposing a hybrid model, called as H-BERT, that combines the strengths of SpanBERT, BiLSTM, CRF, and LLMs to improve aspect and sentiment extraction accuracy. The methodology involved integrating both human-annotated and LLM-generated reviews, preprocessing the data, and using the H-BERT framework for classification. Experimental results using the SemEval 2014 dataset demonstrated that H-BERT achieved an accuracy of 90.58% and macro-F-score of 90.56% in the laptop domain, and 91.21% accuracy with a 92.03% macro-F-score in the restaurant domain, outperforming existing models. These improvements highlight the model's effectiveness in handling both synthetic and real-world reviews. The study concludes that H-BERT offers a more reliable and flexible approach to ABSA. For future work, the model can be extended to multilingual datasets, incorporate multimodal inputs like images or speech, and explore reinforcement learning to further enhance sentiment understanding in dynamic, real-time applications.

ACKNOWLEDGMENTS

Authors would like to express our heartfelt thanks to our guide for his unwavering guidance, invaluable insights, and encouragement throughout the research process.

FUNDING INFORMATION

The authors state no funding is involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mohammed Ziaulla	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	
Arun Biradar	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓		

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

The author declares that there are no known conflicts of interest associated with this publication. There are no financial or personal relationships that could inappropriately influence or bias the content of this work.

INFORMED CONSENT

Not applicable. This study did not involve human participants, human data, or any personally identifiable information. All data used were either publicly available, fully anonymized, or derived from non-human sources, and therefore no informed consent was required from individuals.

ETHICAL APPROVAL

Not applicable. This research did not involve human subjects, human biological materials, or experimental procedures on animals. The work was conducted solely on computational models, publicly available datasets, or non-sensitive data that did not require intervention with living organisms. Therefore, ethical approval from an institutional review board or animal ethics committee was not necessary for this study.

DATA AVAILABILITY

No dataset is utilized in this research.

REFERENCES

[1] N. Ahmed, A. K. Saha, M. A. Al Noman, J. R. Jim, M. F. Mridha, and M. M. Kabir, "Deep learning-based natural language processing in human–agent interaction: applications, advancements and challenges," *Natural Language Processing Journal*, vol. 9, Dec. 2024, doi: 10.1016/j.nlp.2024.100112.

[2] Supriyono, A. P. Wibawa, Suyono, and F. Kurniawan, "Advancements in natural language processing: implications, challenges, and future directions," *Telematics and Informatics Reports*, vol. 16, Dec. 2024, doi: 10.1016/j.teler.2024.100173.

[3] J. R. Jim, M. A. R. Talukder, P. Malakar, M. M. Kabir, K. Nur, and M. F. Mridha, "Recent advancements and challenges of NLP-based sentiment analysis: a state-of-the-art review," *Natural Language Processing Journal*, vol. 6, Mar. 2024, doi: 10.1016/j.nlp.2024.100059.

[4] K. Aziz, D. Ji, P. Chakrabarti, T. Chakrabarti, M. S. Iqbal, and R. Abbasi, "Unifying aspect-based sentiment analysis BERT and multi-layered graph convolutional networks for comprehensive sentiment dissection," *Scientific Reports*, vol. 14, no. 1, Jun. 2024, doi: 10.1038/s41598-024-61886-7.

[5] M. M. Hossain, M. S. Hossain, M. F. Mridha, M. Safran, and S. Alfarhood, "Multi task opinion enhanced hybrid bert model for mental health analysis," *Scientific Reports*, vol. 15, no. 1, Jan. 2025, doi: 10.1038/s41598-025-86124-6.

[6] C. Zhao, R. Feng, X. Sun, L. Shen, J. Gao, and Y. Wang, "Enhancing aspect-based sentiment analysis with bert-driven context generation and quality filtering," *Natural Language Processing Journal*, vol. 7, Jun. 2024, doi: 10.1016/j.nlp.2024.100077.

[7] N. M. Gardazi, A. Daud, M. K. Malik, A. Bukhari, T. Alsahfi, and B. Alshemaimri, "BERT applications in natural language processing: a review," *Artificial Intelligence Review*, vol. 58, no. 6, Mar. 2025, doi: 10.1007/s10462-025-11162-5.

[8] A. Musa, F. M. Adam, U. Ibrahim, and A. Y. Zandam, "HauBERT: a transformer model for aspect-based sentiment analysis of Hausa-language movie reviews," in *Engineering Proceedings*, vol. 87, no. 1, Apr. 2025, doi: 10.3390/engproc2025087043.

[9] M. Goyal and Q. H. Mahmoud, "A systematic review of synthetic data generation techniques using generative AI," *Electronics*, vol. 13, no. 17, Sep. 2024, doi: 10.3390/electronics13173509.

[10] M. Nadăş, L. Dioşan, and A. Tomescu, "Synthetic data generation using large language models: advances in text and code," *IEEE Access*, vol. 13, pp. 134615–134633, 2025, doi: 10.1109/ACCESS.2025.3589503.

[11] Q. Gu, Z. Wang, H. Zhang, S. Sui, and R. Wang, "Aspect-level sentiment analysis based on syntax-aware and graph convolutional networks," *Applied Sciences*, vol. 14, no. 2, Jan. 2024, doi: 10.3390/app14020729.

[12] N. Jeong and J. Lee, "An aspect-based review analysis using ChatGPT for the exploration of hotel service failures," *Sustainability*, vol. 16, no. 4, Feb. 2024, doi: 10.3390/su16041640.

[13] F. Zhang, W. Zheng, and Y. Yang, "Graph convolutional network with syntactic dependency for aspect-based sentiment analysis," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, Feb. 2024, doi: 10.1007/s44196-024-00419-6.

[14] N. Mughal, G. Mujtaba, S. Shaikh, A. Kumar, and S. M. Daudpota, "Comparative analysis of deep natural networks and large language models for aspect-based sentiment analysis," *IEEE Access*, vol. 12, pp. 60943–60959, 2024, doi: 10.1109/ACCESS.2024.3386969.

[15] T. Falatouri, D. Hrušecká, and T. Fischer, "Harnessing the power of LLMs for service quality assessment from user-generated content," *IEEE Access*, vol. 12, pp. 99755–99767, 2024, doi: 10.1109/ACCESS.2024.3429290.

[16] L. Liu, T. Qin, Y. Zhou, C. Wang, and X. Guan, "SynSem-aste: an enhanced multi-encoder network for aspect sentiment triplet extraction with syntax and semantics," *IEEE Transactions on Affective Computing*, vol. 15, no. 4, pp. 2097–2111, Oct. 2024, doi: 10.1109/TAFFC.2024.3397961.

[17] N. C. Hellwig, J. Fehle, and C. Wolff, "Exploring large language models for the generation of synthetic training samples for aspect-based sentiment analysis in low resource settings," *Expert Systems with Applications*, vol. 261, Feb. 2025, doi: 10.1016/j.eswa.2024.125514.

[18] P. Pandey and J. P. Singh, "Generating product reviews from aspect-based ratings using large language models," *Journal of Retailing and Consumer Services*, vol. 84, May 2025, doi: 10.1016/j.jretconser.2025.104244.

[19] R. Fan, T. He, M. Chen, M. Zhang, X. Tu, and M. Dong, "Dual causes generation assisted model for multimodal aspect-based sentiment classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 5, pp. 9298–9312, May 2025, doi: 10.1109/TNNLS.2024.3415028.

[20] C. Zhao *et al.*, "Aspect-level sentiment classification of consumer reviews utilizing bert and category-aware multi-head attention," *IEEE Transactions on Consumer Electronics*, vol. 71, no. 2, pp. 3329–3339, May 2025, doi: 10.1109/TCE.2025.3563150.

[21] C. Chugh, "SemEval 2014 task 4: aspectbasedsentimentanalysis," *Kaggle*. Accessed: Jul. 18, 2025. [Online]. Available: https://www.kaggle.com/datasets/charitarth/semeval-2014-task-4-aspectbasedsentimentanalysis

[22] A. Ilmania, Abdurrahman, S. Cahyawijaya, and A. Purwarianti, "Aspect detection and sentiment classification using deep neural network for indonesian aspect-based sentiment analysis," in *2018 International Conference on Asian Language Processing (IALP)*, Nov. 2018, pp. 62–67, doi: 10.1109/IALP.2018.8629181.

[23] R. A. Khachfeh, I. El Kabani, and Z. Osman, "An enhanced hybrid bert-bilstm learning model for arabic news classification," in *2025 International Conference on Machine Intelligence and Smart Innovation (ICMISI)*, May 2025, pp. 201–206, doi: 10.1109/ICMISI65108.2025.11115581.

[24] N. M. Cural, C. Karabulut, and P. Karagöz, "BERT and svm integration for fake news detection in turkish: evaluation with a new dataset," in *2025 33rd Signal Processing and Communications Applications Conference (SIU)*, Jun. 2025, pp. 1–4, doi: 10.1109/SIU66497.2025.11112123.

[25] R. Raj, N. Kaur, and P. Goel, "Design hybrid model to improve stock selection and market timing by integrating deep learning and sentiment analysis," in *2025 International Conference on Computing Technologies (ICOCT)*, Jun. 2025, pp. 1–8, doi: 10.1109/ICOCT64433.2025.11118343.

[26] J. R. Tewell and W. H. Tobey, "Gemini orbital tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-2, no. 3, pp. 346–352, May 1966, doi: 10.1109/TAES.1966.4501771.

[27] A. E. Piatti, S. C. Keller, A. D. Mackey, and G. S. Da Costa, "Gemini/GMOS photometry of intermediate-age star clusters in the large Magellanic cloud," *Monthly Notices of the Royal Astronomical Society*, vol. 444, no. 2, pp. 1425–1441, Oct. 2014, doi: 10.1093/mnras/stu1535.

[28] R. Y. Kim, "Using online reviews for customer sentiment analysis," *IEEE Engineering Management Review*, vol. 49, no. 4, pp. 162–168, Dec. 2021, doi: 10.1109/EMR.2021.3103835.

[29] S. Abbas, W. Boulila, M. Driss, G. A. Sampedro, M. Abisado, and A. Almadhor, "Active learning empowered sentiment analysis: an approach for optimizing smartphone customer's review sentiment classification," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 4470–4477, Feb. 2024, doi: 10.1109/TCE.2023.3328878.

[30] R. Obiedat, D. Al-Darras, E. Alzaghoul, and O. Harfoushi, "Arabic aspect-based sentiment analysis: a systematic literature review," *IEEE Access*, vol. 9, pp. 152628–152645, 2021, doi: 10.1109/ACCESS.2021.3127140.

BIOGRAPHIES OF AUTHORS

Mohammed Ziaulla is currently pursuing Ph.D. from REVA University and a research scholar of School of Computer Science and Engineering, REVA University, Bengaluru, India. Presently he is pursuing research works in natural language processing, aspect-based sentiment analysis, artificial intelligence, and machine learning algorithms, under REVA University, Bengaluru, India. He can be contacted at email: mdziaulla.dr@rediffmail.com.

Arun Biradar received his Bachelor of Engineering degree in Computer Science and Engineering from Gulbarga University in 1994 and Master of Technology in Software Engineering from Mysore University in 1999, Kamataka, India. He is currently working as professor in the School of Computer Science and Engineering, REVA University, Bengaluru. He has close association to the industrial training domain for last 14 years and has more than 20 years of teaching experience. His research areas are computer networks, ad-hoc networks, software engineering and genetic algorithms. He can be contacted at email: arun.biradar@reva.edu.in.