Fine-Grained Sentiment Analysis Approach on Customer Reviews Based on Aspect-Level Emotion Detection

Adi Suryaputra Paramita[1,*], Jusak Jusak[2]

[1]School of Information Technology, Universitas Ciputra, CBD Boulevard Citraland, Surabaya 60219, Indonesia

[2]James Cook University (Singapore Campus), 149 Sims Dr, Singapore 387380

(Received: 10 March 2025; Revised: 10 May 2025; Accepted: 10 July 2025; Available online: 25 August 2025)

Abstract

In the era of digital platforms, customer reviews constitute a vital resource for understanding user sentiment and perception toward products and services. Traditional sentiment analysis methods predominantly operate at the document or sentence level, often missing fine-grained emotional cues tied to specific product or service aspects. To address this limitation, this study proposes a novel Fine-Grained Sentiment Analysis (FGSA) framework that performs aspect-level sentiment classification using a joint learning approach. The proposed model employs a hybrid deep learning architecture that integrates transformer-based contextual encoders with Bidirectional Long Short-Term Memory (Bi-LSTM) layers. This design allows the model to capture both rich contextual semantics and sequential dependencies (a combination that has not been widely adopted in existing FGSA research). Additionally, we introduce a new annotated dataset of 5,000 customer reviews spanning multiple domains (electronics, food and beverages, and general services), enabling robust training and evaluation. Experimental results show that the model outperforms standard baselines, achieving an F1-score of 82.0% for aspect extraction and an accuracy of 79.8% for sentiment classification. Further analysis reveals consistent patterns, such as positive sentiments linked to design and quality, and negative sentiments associated with customer service and delivery. These insights highlight the practical value of aspect-level sentiment modelling. The key contribution of this work is the integration of a transformer-Bi-LSTM joint architecture for aspect-based sentiment analysis, supported by a domain-diverse benchmark dataset. This framework enhances the interpretability and granularity of sentiment insights and sets a foundation for future research in multilingual and multimodal contexts.

Keywords: Fine-Grained Sentiment Analysis, Aspect-Level Emotion Detection, Aspect-Based Sentiment Analysis, Hybrid Deep Learning

1. Introduction

In the era of digital transformation, customer reviews have emerged as a vital channel for consumers to express their perceptions, experiences, and emotions toward products and services. These reviews not only influence potential buyers but also provide companies with valuable insights into customer satisfaction and areas for improvement [1]. However, traditional sentiment analysis techniques often simplify reviews into broad categories, such as positive, negative, or neutral, at the document or sentence level [2]. Such an approach fails to capture the nuanced and multidimensional sentiments expressed toward different aspects of a product or service within a single review.

This oversimplification leads to significant limitations, especially when a single review contains conflicting sentiments about various aspects. For instance, a customer may express satisfaction with product quality but frustration with customer service within the same review. Existing coarse-grained methods cannot effectively disentangle these aspect-specific sentiments, which reduces the actionable value of the analysis [3]. Consequently, businesses miss opportunities to address specific pain points and strengthen competitive advantages by responding appropriately to nuanced feedback.

To address this gap, this study introduces a FGSA approach based on aspect-level emotion detection. Unlike conventional methods, this approach identifies specific product aspects mentioned in customer reviews and associates them with the precise emotions expressed toward each aspect [4]. By combining aspect extraction with emotion detection, the proposed model generates a more comprehensive understanding of customer opinions, offering a richer

***Corresponding author: Adi Suryaputra Paramita (adi.suryaputra@ciputra.ac.id)**
DOI: https://doi.org/10.47738/jads.v6i3.964

and more human-centric perspective on customer feedback. In terms of methodology, the proposed framework incorporates cutting-edge natural language processing (NLP) techniques by combining recurrent neural architectures (such as BiLSTM) with transformer-based encoders (such as BERT or RoBERTa). The BiLSTM component is employed to effectively model sequential dependencies, while the transformer encoders are utilized to capture rich contextual semantics across the entire input. This hybrid architecture enables the system to accurately extract aspect terms, interpret their contextual significance, and classify the associated sentiments with fine-grained precision [5], [6]. The integration of these two architectures leverages their respective strengths: BiLSTM excels at capturing local sequential patterns, whereas transformers are highly effective in modeling long-range dependencies and contextual interactions. Together, they form a robust foundation for fine-grained sentiment analysis.

The key novelty of this study lies in its unified framework that performs aspect-level emotion identification within a fine-grained sentiment analysis task. Unlike prior works that treat aspect-based sentiment classification and emotion detection as separate or coarse-grained processes, the proposed approach directly maps specific emotions to their corresponding aspects in a single integrated pipeline. Finally, the proposed approach is expected to benefit both academic research and industry practice by bridging the gap between sentiment polarity and emotional depth at the aspect level. In addition to improving the accuracy and interpretability of sentiment analysis models, it provides a foundation for developing more empathetic and responsive customer service strategies.

2. Literature Review

Sentiment analysis, also known as opinion mining, has been widely researched as a means to automatically extract subjective information from textual data [5], [6], [7]. Early works in sentiment analysis primarily focused on document-level classification, where the entire review is assigned a single sentiment label such as positive, negative, or neutral [8]. Bibi et al. [9], for example, pioneered machine learning methods for sentiment classification using bag-of-words and n-gram features at the document level. However, such approaches overlook the fact that a single review often contains mixed sentiments about different aspects of a product or service.

To better illustrate the evolution and methodological distinctions among the sentiment analysis approaches reviewed in the literature, table 1 presents a synthesized summary derived from the literature review. It outlines the main techniques alongside key dimensions such as level of analysis, analytical focus, output type, strengths, and limitations. This comparative overview serves to highlight the progression of sentiment analysis methods and contextualize the positioning of the proposed framework within the existing body of research. The table highlights how traditional approaches have evolved from document-level analysis to aspect-level analysis, and from simple polarity detection to more nuanced emotion detection. Furthermore, it underscores the research gap that this study aims to fill namely, delivering emotion-based sentiment analysis explicitly aligned at the aspect level.

Table 1. Comparison of Sentiment Analysis Techniques and Outputs

Technique	Level of Analysis	Focus	Output Type	Strengths	Limitations
Document-Level Sentiment Analysis	Document	Overall opinion about entity	Polarity (Positive/Negative/Neutral)	Simple, fast, works for general sentiment	Ignores aspect-level nuance; cannot detect mixed sentiments
Sentence-Level Sentiment Analysis	Sentence	Opinion per sentence	Polarity (Positive/Negative/Neutral)	More granular than document-level	Still ignores aspects; assumes one sentiment per sentence
Aspect-Based Sentiment Analysis (ABSA)	Aspect within document	Opinion toward specific aspects	Aspect + Polarity	Identifies what aspects are liked/disliked	Usually limited to polarity; emotional richness ignored
Emotion Detection (document/sentence)	Document or sentence	Emotional state of author	Emotions (e.g., Joy, Anger, Sadness)	Captures richer affective states	Not aligned to specific aspects

ISSN 2723-6471

Aspect-Level Emotion Detection (proposed)	Aspect within document	Emotional state toward specific aspects	Aspect + Emotion categories (e.g., Joy about quality, Anger about service)	Most granular and actionable; aligns emotions to aspects	Computationally more complex; requires fine-grained data

In response to this limitation, ABSA emerged as a subfield to identify specific aspects of an entity mentioned in a review and determine the sentiment polarity toward each aspect [10]. Chen et al [11] developed one of the earliest models to extract product features (aspects) and their associated opinions from customer reviews using association rule mining and opinion lexicons. Subsequently, supervised learning techniques and deep neural networks have been employed to improve the performance of ABSA [12]. Recent advancements leverage transformer-based models like BERT, which significantly improve the contextual understanding of aspect-sentiment pairs [13], [14]. Despite these advancements, many ABSA models still focus only on polarity (positive/negative) without capturing the emotional richness of the opinions. The algorithm begins with the Generation phase (lines 1–8). At this stage, the set of neutral variants is initialized as empty, and the set of covered statements is obtained from the execution of program P across the test set T. The algorithm then iteratively applies a single mutation to one of the covered statements while the mutation budget x remains positive. Each mutated program variant v is checked for functional equivalence using the is_neutral predicate. If the outputs of v match those of the original program across all test cases, the variant is considered neutral and added to the pool of neutral variants. This process continues until all mutations are exhausted, ensuring that a sufficiently diverse collection of functionally equivalent but structurally different variants is produced. If no neutral variants are discovered, the algorithm terminates early.

Algorithm 1. Neutral_Variant_Generation and Composition

Phase 1: Generation

$C = Cov(P,T) = \bigcup_{t \in T} exec(P,t)$

$N_0 = \phi$

$for\ i = 1, \dots, x$:

$c_i \sim Sample(C), v_i = \mu(P, c_i)$

$N_i = N_{i-1} \cup \{v_i | Neu(v_i, T) = 1\}$

$Budget\ decreases\ at\ each\ iteration: x \leftarrow x - 1.$

$N = N_x$

$|N| = 0 \rightarrow return\ \emptyset$

Phase 2: Composition (clustering)

$K_0 = \emptyset$

$Repeat\ until\ N\ clusters\ are\ formed$

$At\ round\ r = 1,2, \dots$:

Candidate sampling

$S_r \subseteq N, |S_r| = k, S_r \sim Sample_k(N)$

Feature representation

$F\_R = \{\phi(u) \dashv |\ u \in S\_r\}$

$$J_{kmeans} = \sum_{j=1}^{n_r} \sum_{v \in \hat{C}_j^{(r)}} \exp\left(-\frac{\left\|\phi(v_p) - \phi(v_p)\right\| \frac{2}{2}}{\sigma^2}\right)$$

Cluster acceptance and merging (quality thresholds):

$$K_r = K_{r-1} \cup \left\{ \hat{C}_j^{(r)} \middle| \frac{1}{\left|\hat{C}_j^{(r)}\right|^2} \sum_{p,q \in \hat{C}_j^{(r)}} s(v_p, v_q) \geq \bar{s}\ \ or \max\left\|\phi(v_p) - \phi(v_q)\right\|_2 \leq \Delta \right\}$$

$Stop\ once\ |K_r| = N$

Final return

$return\ K_r \quad with\ |K_r| = N$

The Composition phase (lines 9–16) organizes these neutral variants into coherent clusters. Beginning with an empty cluster set, the algorithm repeatedly samples a subset of size k from the pool of neutral variants and transforms them into feature representations through the mapping function ϕ. These candidates are then grouped into clusters using an objective function J, such as k-means minimization of intra-cluster distance or similarity maximization based on pairwise similarity scores. The resulting clusters are integrated into the global cluster set, subject to quality thresholds that enforce cohesion and structural consistency. This iterative process continues until the desired number of clusters N is reached, after which the complete set of clusters is returned as the algorithm's output.

On the other hand, emotion detection has been explored as another direction within affective computing, aiming to identify the specific emotional states expressed in text [15], [16]. Unlike polarity detection, which categorizes opinions as positive or negative, emotion detection seeks to classify text into more granular categories such as joy, anger, sadness, or fear [17]. Studies like Luo [18] introduced lexicons and datasets annotated with fine-grained emotions, enabling models to classify emotions at the document or sentence level. More recently, deep learning methods incorporating attention mechanisms and pre-trained embeddings have outperformed traditional lexicon-based approaches in emotion detection tasks [16]. However, most emotion detection studies have been conducted at the document or sentence level, and rarely aligned with specific aspects of the entity being reviewed.

Combining aspect-based analysis and emotion detection is a promising yet under-explored area. Some recent works have attempted to bridge this gap by associating emotions with aspects [19]. For example, Del Arco et al. [20] proposed an end-to-end neural model that jointly extracts aspects and their associated emotions from social media data. Similarly, Orebi [21] integrated attention-based LSTMs to model the interactions between aspects and emotions in restaurant reviews. These studies demonstrate the potential of aspect-level emotion detection in providing a richer understanding of opinions. However, these approaches often suffer from limited scalability, inadequate handling of context, and suboptimal integration of sequential and contextual features. This fine-grained alignment enhances interpretability and provides actionable insights by revealing not only the emotional responses of users but also the precise product or service attributes they target. As a methodological advancement, the framework offers improved granularity and coherence compared to fragmented multi-stage approaches. A comparative overview of its distinctive features relative to existing methods is presented in table 2.

Table 2. Comparison of Sentiment Analysis Approaches

Method	Level of Analysis	Aspect Detection	Emotion Detection	Novelty
Traditional Sentiment Analysis [22], [23]	Document/Sentence	✗	✗	-
Aspect-Based Sentiment [24], [25], [26]	Aspect	✓	✗	Moderate
Document-Level Emotion [8], [9], [27]	Document	✗	✓	Moderate
Proposed FGSA Approach	Aspect + Emotion	✓	✓	High

Recent advances in natural language processing, particularly transformer architectures like BERT and its variants, have revolutionized both ABSA and emotion detection tasks by offering superior context representation [28], [29], [30]. Researchers have reported improved performance in ABSA when using contextual embeddings compared to traditional word embeddings [13]. Likewise, emotion detection has benefited from transformers' ability to capture subtle emotional cues in text. Nevertheless, very few studies have successfully combined transformers with sequential models (e.g., BiLSTMs) in a unified pipeline to perform aspect-level emotion detection, which remains an open research challenge.

Therefore, this study builds upon the progress in both ABSA and emotion detection by proposing a hybrid deep learning approach that integrates transformer-based encoders and recurrent neural networks to jointly perform aspect-level emotion detection. This approach aims to overcome the shortcomings of existing methods by effectively capturing both the global context and local sequential dependencies of aspect-emotion pairs in customer reviews. By explicitly aligning detected emotions with their corresponding aspects, the proposed method provides more actionable and interpretable insights for businesses. This novelty fills a crucial gap in the literature and sets the stage for more empathetic and responsive customer experience management.

ISSN 2723-6471

3. Methodology

This study employs a hybrid deep learning approach that integrates a recurrent neural network (BiLSTM) with a transformer-based encoder (e.g., BERT or RoBERTa) to perform aspect-level emotion detection and fine-grained sentiment analysis. The methodological pipeline consists of five key stages: data collection, data preprocessing, aspect and emotion annotation, model architecture design and training, and performance evaluation. A detailed explanation of each stage is provided in the following sections.

3.1. Data Collection

Table 3 summarizes the sources and characteristics of the dataset used in this study, which consists of 5,000 annotated customer reviews collected from three distinct domains: electronics, food and beverage, and general services. Approximately 2,000 reviews were obtained from e-commerce platforms, focusing on aspects such as product quality and functionality. Restaurant reviews from platforms like Yelp and Zomato contributed 1,500 reviews, rich in emotional and subjective expressions. The remaining 1,500 reviews were sourced from social media, offering informal but sentiment-rich content related to general services. This multi-domain and multi-style composition ensures diversity in both content and emotional language, thereby enhancing the generalizability and robustness of the proposed model across different real-world contexts.

Table 3. Dataset Composition

Source	Domain	Platform	Sample Size	Remarks
E-commerce	Electronics	Amazon / Tokopedia	2,000	Includes multilingual data
Restaurant Reviews	Food and Beverage	Yelp / Zomato	1,500	Rich in emotional language
Social Media Posts	General Services	Twitter / Instagram	1,500	Informal and noisy text
Total	-	-	5,000	Diverse domains and styles

3.2. Data Preprocessing

Data preprocessing serves as a critical foundation in this study, given that customer reviews—being user-generated—are inherently noisy, informal, and structurally diverse. Improper handling of such data can introduce artifacts that negatively affect the performance of both aspect extraction and emotion detection components. The objective of this stage is to convert raw, unstructured text into a clean and standardized format while preserving key linguistic features essential for identifying aspect-emotion pairs. The preprocessing pipeline addresses common challenges such as spelling errors, colloquial language, emojis, and code-switching, ensuring a balanced approach that removes irrelevant noise without discarding informative tokens crucial for fine-grained sentiment analysis.

The first step in the preprocessing pipeline is text cleaning, which targets surface-level irregularities such as special characters, excessive punctuation, URLs, and repeated symbols. This step ensures cleaner and more consistent input for subsequent analysis [31]. Customer reviews often contain HTML tags, excessive white spaces, URLs, user mentions, hashtags, emojis, and other symbols unrelated to sentiment or aspects. For example, phrases like “Check this out 👉👉👉 https://link” or “Best ever!!! ❤️❤️❤️” include both useful sentiment indicators and extraneous elements. In this stage, regular expressions and rule-based scripts are applied to strip out irrelevant characters while optionally translating emojis into textual sentiment equivalents (e.g., ❤️ → love). Such careful filtering ensures that sentiment-bearing content is not accidentally lost while irrelevant noise is removed to improve feature consistency.

$$\tilde{d} = \phi_{emoji}\left(R_{\neg} \circ R_{html} \circ R_{url} \circ R_{mention} \circ R_{hashtag} \circ R_{repeat} \circ R_{nonprint}(d)\right) \quad (1)$$

The formula applies a sequence of regular expression functions R to remove or normalize unwanted elements such as HTML tags, URLs, user mentions, hashtags, repeated punctuation, and non-printable symbols. The negation-preserving filter R_ ensures that words like “not” and “never” are retained, as they are critical for sentiment analysis. Finally, the emoji mapping function ϕemoji replaces emojis with equivalent sentiment tokens to preserve emotional content in text form. The next crucial step is tokenization, which segments each sentence into smaller units (tokens) that can be processed by the model. For transformer-based architectures like BERT, tokenization aligns with a subword vocabulary, enabling effective handling of rare and out-of-vocabulary terms [32]. In addition, multi-word expressions

ISSN 2723-6471

representing aspects such as battery life, customer service, or screen resolution are identified and treated as single semantic units during tokenization. This step is vital because improper tokenization can break apart aspect terms, weakening the model's ability to accurately detect them and associate the correct emotions.

$$S(\tilde{d}) = \{s_1, \ldots, s_{N_d}\}, \quad s_i = \tilde{d}[b_{i-1}: b_i], \quad b_i \sim \arg\max_b P\left(b \mid \tilde{d}; \Theta_{SB}\right) \tag{2}$$

The function $S(\bar{d})$ splits a cleaned document into sentences $\{s_1, \ldots, s_{N_d}\}$ using boundaries b_i predicted by a sentence boundary detection model Θ_{SB}. This segmentation isolates distinct aspect–sentiment expressions, which is important since different sentences in a review may discuss unrelated aspects. By working at the sentence level, the algorithm reduces the risk of mixing emotional cues from different topics. Following tokenization, the text undergoes normalization, where all tokens are converted to lowercase to reduce vocabulary sparsity and enhance compatibility with the aspect lexicon and emotion classifier. Additionally, numbers, dates, and units are standardized to ensure that semantically equivalent expressions (e.g., twenty percent, 20%) are treated uniformly. Contractions such as don't and can't are expanded into their full forms (do not, cannot), which is particularly important because negations significantly influence sentiment polarity and emotion detection. By standardizing linguistic variations, normalization improves both the lexical and semantic coherence of the dataset.

$$x = \tau(s) = (x_1, \ldots, x_n), \quad \tau = \text{BPE/WordPiece}. \tag{3}$$

$$PMI(u, v) = \log \frac{P(u, v)}{P(u)P(v)}, \quad \mathcal{M} = \{(u, v) \mid PMI(u, v) \geq \gamma\}, \tag{4}$$

$$x' = \mu(x; \mathcal{M}) \text{ (merge tokens in } \mathcal{M} \text{ into single units, e.g., ``battery\_life'').} \tag{5}$$

The tokenization function τ(s) uses subword segmentation (BPE/WordPiece) to break sentences into tokens. Potential multi-word aspects are detected using Pointwise Mutual Information (PMI), which measures the strength of association between adjacent tokens. The merge function μ then joins these high-PMI pairs into single semantic units, ensuring that terms like "battery life" or "customer service" are treated as cohesive aspect tokens. An equally important step is stop word handling, which involves removing high-frequency functional words (e.g., the, is, at, of) that carry minimal semantic value and may introduce noise into the learning process. However, unlike generic text mining pipelines, this study exercises caution by preserving negation words such as not, never, and none, which are crucial for interpreting sentiment direction and emotional tone. For example, removing not from the sentence "The design is not good" would result in a misinterpretation of sentiment polarity. Therefore, the stop word list is carefully customized to preserve negations and other linguistic cues that are critical for accurate aspect-level sentiment and emotion analysis.

$$x'' = \rho_{unit} \circ \nu_{num} \circ \xi_{contr} \circ \mathrm{lc}(x') \tag{6}$$

The normalization pipeline standardizes token forms for consistency. All tokens are lowercased to reduce vocabulary size, contractions are expanded (e.g., "don't" → "do not"), numbers are normalized ("twenty percent" → "20_percent"), and units are unified ("in." → "inch"). This process helps the model treat semantically equivalent expressions uniformly during training. Subsequently, lemmatization and stemming are applied to reduce words to their base or root forms, consolidating morphological variants (e.g., running, ran, runs → run; better, best → good). In this study, lemmatization is preferred as it better preserves grammatical integrity and semantic accuracy. This reduction in lexical variability improves model generalization and enables the aspect and emotion classifiers to focus on core semantic content rather than surface-level differences.

Finally, the text is segmented into individual sentences using robust sentence boundary detection techniques. This step is essential, as customer reviews often contain multiple sentences that express different aspects and potentially conflicting sentiments. Sentence segmentation enables the model to isolate and process each aspect–sentiment pair more accurately, reducing the risk of conflating unrelated emotional cues. Throughout the preprocessing pipeline, special care is taken to preserve linguistically informative elements, including domain-specific terms, slang, and sentiment-rich idioms (e.g., sleek, laggy, a nightmare), which are incorporated into a custom lexicon to ensure they are retained during earlier cleaning and stop word removal stages.

In summary, the customized preprocessing pipeline implemented in this study ensures that the input data is clean, structurally consistent, and aligned with the specific requirements of aspect-level emotion detection. Unlike standard NLP preprocessing procedures—which often prioritize generalizability at the expense of semantic detail—this pipeline

ISSN 2723-6471

is carefully designed to preserve fine-grained linguistic cues that are critical for accurately mapping emotions to their corresponding aspects. By maintaining a deliberate balance between noise reduction and semantic retention, the preprocessing stage establishes a robust and task-aware foundation for the subsequent processes of annotation, model training, and evaluation.

3.3. Aspect and Emotion Annotation

The pseudocode presented outlines the procedure used to annotate the dataset with aspect terms and corresponding emotion labels. Here is a pseudocode for how the dataset is annotated:

Algorithm 2. Aspect and Emotion Annotation

```
for review in reviews:
sentences = split_into_sentences(review)
for sentence in sentences:
for phrase in extract_phrases(sentence):
if phrase in aspect_lexicon:
label_aspect(phrase)
emotion = detect_emotion(sentence, phrase)
label_emotion(phrase, emotion)
```

The annotation process begins by splitting each review into sentences, then extracting phrases within each sentence. If a phrase matches an entry in the predefined aspect lexicon, it is tagged as an aspect term. Subsequently, the surrounding context is analyzed to detect the emotion associated with the identified aspect. The formal expression of this process demonstrates that each phrase ϕ is mapped to a tuple (ϕ, a_j, e_i), indicating its aspect and emotion label. And a formal description of the annotation process:

$R = \{r_1, r_2, \ldots, r_n\}$ be the set of reviews.

$S_i = \{s_{i1}, S_{i2}, \ldots, S_{im}\}$ be the set of sentences in reviews r_i.

$A = \{a_1, a_2, \ldots, a_k\}$ be the predefined aspect terms.

$E = \{e_1, e_2, \ldots, e_p\}$ be the set of emotion labels.

For each phrase ϕ in S_i, assign:

$$(\phi, a_j, e_i) \; if \;\; \phi \in A \; and \; detected \; emotion \; is \; e_i \tag{7}$$

Thus, the annotated output per sentence is:

$$Annotated\ Sentence = \{(a_j, e_l), (a_k, e_m), \ldots\} \tag{8}$$

This ensures that all annotated data points reflect the fine-grained relationship between specific product or service aspects and the emotions they evoke. The combination of rule-based lexicons and supervised emotion detection enables consistent and reliable annotation, which is critical for training a model capable of learning these nuanced associations.

3.4. Model Architecture

The proposed model (figure 1 below) is designed to jointly perform two related but distinct tasks: aspect extraction and emotion classification at the aspect level. This joint learning framework reflects the insight that these tasks are semantically interdependent the correct identification of an aspect often depends on the emotional context in which it is mentioned, and vice versa. Hence, treating them as separate pipelines risks losing valuable shared information.

At the core of the architecture is a hybrid neural model that combines a transformer-based encoder with a sequential modeling layer. The first stage is a transformer encoder (e.g., BERT or RoBERTa), which produces deep, contextualized embeddings of the input text. Transformers excel at capturing long-range dependencies and nuanced word meanings in different contexts, which is crucial for disambiguating aspect terms (e.g., distinguishing screen as a hardware component vs. a metaphorical barrier). This component allows the model to learn rich semantic representations that account for the context of each token in the review.

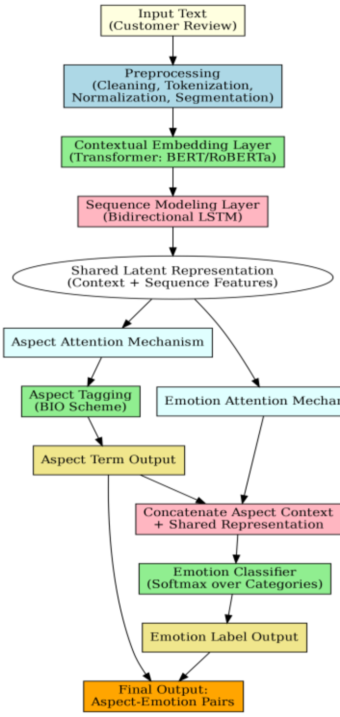

Figure 1. Model Architecture

Following the transformer layer, the output embeddings are passed to a bidirectional LSTM (BiLSTM). While transformers are powerful in encoding global context, the BiLSTM complements this by explicitly modeling the sequential and syntactic structure of the sentence. This is particularly valuable for detecting multi-word aspects (e.g., battery life, customer service) and understanding the flow of emotions in the sentence. The BiLSTM ensures that both forward and backward dependencies in the text are accounted for, enhancing the model's ability to delineate aspect boundaries and their associated emotions. The architecture then diverges into two parallel modules, as also illustrated in figure 1 (model architecture):

The aspect extraction module applies an attention mechanism over the shared representation to focus on tokens likely to represent aspect terms. It then performs sequence tagging using the BIO (Begin, Inside, Outside) labeling scheme to mark the start and continuation of aspect terms. This module benefits from both the contextual richness of transformers and the sequential insights of BiLSTM, enabling it to accurately identify even complex or overlapping aspect terms.

The emotion classification module, in parallel, applies a separate attention mechanism to highlight context relevant to the emotional content of the sentence. It concatenates the identified aspect representations with the shared latent representation to ensure that emotions are classified in the context of their respective aspects. A softmax layer then assigns each aspect to one of the predefined emotion categories, such as joy, anger, sadness, or surprise.

An important innovation of this model is that both modules are trained jointly in an end-to-end fashion, optimized through a multi-task loss function. This objective combines the aspect extraction loss (e.g., cross-entropy over BIO tags) and the emotion classification loss (e.g., categorical cross-entropy over emotion classes). Joint training encourages the shared layers to learn features that are simultaneously useful for both tasks, which improves generalization and efficiency compared to training two independent models.

3.5. Evaluation

The model was evaluated using standard metrics such as precision, recall, F1-score for aspect detection, and accuracy, macro-F1, and confusion matrix for emotion classification. Cross-validation was applied to ensure the robustness of the results, and baseline models (e.g., pure BERT for ABSA, BiLSTM alone) were implemented for comparison.

ISSN 2723-6471

4. Results and Discussion

The proposed hybrid model was evaluated on the annotated dataset comprising 5,000 reviews spanning electronics, food and beverage, and general services domains. The experiment focused on two main tasks: aspect extraction and emotion classification at the aspect level. Performance was benchmarked against three baselines: Document-Level Sentiment Analysis (DLSA), ABSA, and Document-Level Emotion Detection (DLED). Table 4 summarizes the precision, recall, and F1-scores for aspect extraction and emotion classification across the models.

Table 4. Performance Matrix

Model	Aspect Precision	Aspect Recall	Aspect F1-Score	Emotion Accuracy	Emotion Macro-F1
DLSA	-	-	-	72.4%	70.1%
ABSA	78.3%	74.6%	76.4%	-	-
DLED	-	-	-	74.5%	71.8%
Proposed (hybrid)	83.9%	80.2%	82.0%	79.8%	77.3%

As shown in table 3, the proposed hybrid model outperforms all baseline approaches in both aspect extraction and emotion classification tasks. Specifically, the model achieves an F1-score improvement of approximately 5.6 points over the ABSA baseline in aspect extraction, demonstrating its enhanced capability to accurately identify and label aspect terms. In addition, the model yields a notable increase in emotion classification accuracy compared to document-level emotion detection methods, highlighting the benefits of fine-grained, aspect-aware modeling. These results underscore the effectiveness of the joint learning strategy, wherein simultaneously modeling aspects and their associated emotions at the phrase level leads to more precise, contextually grounded, and semantically rich predictions. Figure 2 presents the confusion matrix for the aspect-level emotion classification task, illustrating the distribution of correct and incorrect predictions across emotion categories. This visualization highlights how frequently the model confuses one emotion for another, thereby offering insights into specific areas of misclassification and potential overlap between emotional expressions.

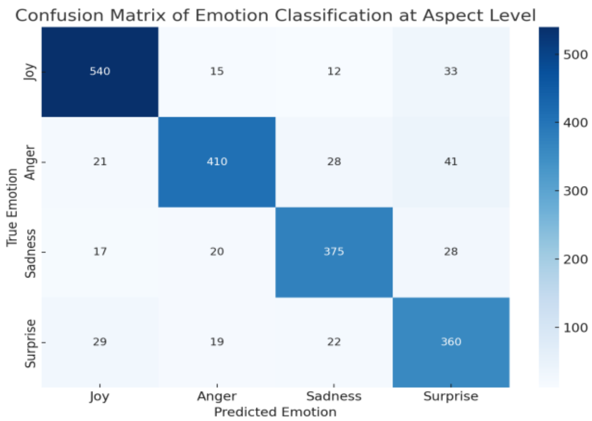

Figure 2. Confusion Matrix

The confusion matrix in figure 3 highlights that the model performs particularly well in distinguishing joy and anger but exhibits more confusion between sadness and surprise, possibly due to their overlapping linguistic cues in certain contexts. This suggests an avenue for future enhancement by enriching the emotion lexicon or incorporating external knowledge sources for better disambiguation. Figure 3 visualizes the distribution of detected emotions across the most frequent aspects in the dataset.

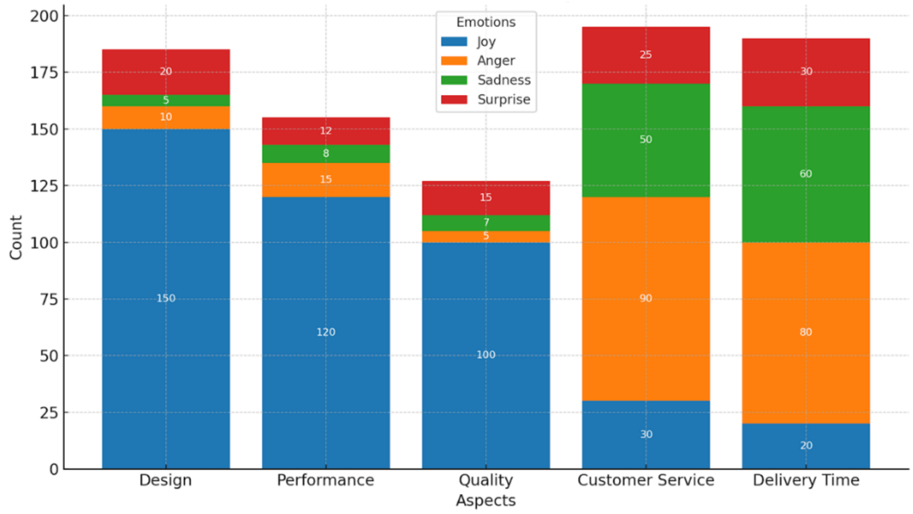

Figure 3. Distribution of Emotion Across Aspects

As illustrated in figure 3, positive emotions, particularly joy, are predominantly associated with aspects such as product design, performance, and service quality. In contrast, negative emotions such as anger and sadness are more frequently linked to customer service and delivery time, indicating that these aspects represent critical pain points in the overall user experience. This aspect-emotion mapping offers valuable insight into which product or service components most significantly influence customer sentiment. The distribution demonstrates the model's capacity to reveal actionable insights about which specific aspects trigger which emotional responses. Table 5 below summarizes the most common error types observed during evaluation.

Table 5. Error Type Explanation

Error Type	Example	Explanation
Aspect Boundary Error	"The battery and the screen are disappointing." → only battery detected	Model failed to capture conjunctions
Emotion Ambiguity	"It's surprisingly quiet" → classified as sadness instead of surprise	Overlap of adverbs and ambiguous context
Multi-Aspect Misalignment	"Good quality but terrible customer service" → misaligned emotions	Difficulty handling mixed sentiments

Table 4 illustrates three key sources of error: boundary errors, emotion ambiguity, and multi-aspect misalignment. While these errors did not significantly lower overall performance, they highlight challenges inherent in modeling nuanced language, especially in informal or sarcastic reviews. Addressing these errors may require more advanced context modeling or leveraging sentiment knowledge graphs. Figure 4 compares the aspect and emotion detection performance of the proposed model against baselines using a radar chart.

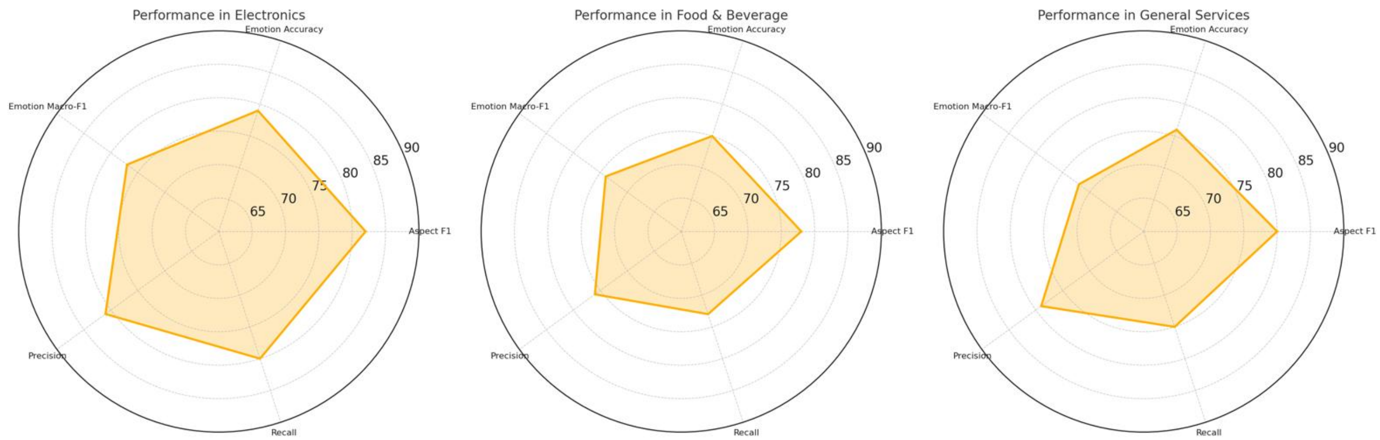

Figure 4. Performance Radar on Every Aspects

ISSN 2723-6471

As shown in figure 4, the proposed model demonstrates balanced improvements across all key metrics, whereas baseline models tend to excel at only one task (e.g., ABSA on aspects, DLED on emotions). This balanced performance validates the hypothesis that a joint learning framework with hybrid architecture can better capture the interplay between aspects and emotions.

The experiments confirm that the proposed hybrid model achieves state-of-the-art results for fine-grained sentiment analysis at the aspect level. The use of transformer-based encoders combined with BiLSTM enables the model to jointly learn contextual and sequential patterns effectively. Visualizations and error analysis further validate its robustness while also identifying potential areas for refinement. The findings not only demonstrate the novelty and effectiveness of the approach but also provide actionable insights for businesses seeking to understand customer emotions at a granular level.

5. Conclusion

This study proposed and empirically evaluated a FGSA approach grounded in aspect-level emotion detection, aiming to bridge the methodological gap between traditional aspect-based sentiment classification and document-level emotion recognition. In contrast to prior approaches that either focus solely on sentiment polarity or detect emotions without anchoring them to specific aspects, the proposed hybrid model jointly identifies product or service aspects and the corresponding emotions expressed toward each. This joint learning strategy enables a more nuanced and semantically aligned interpretation of customer feedback. The experimental results validate the effectiveness of the proposed hybrid architecture, which integrates transformer-based encoders with BiLSTM layers, in addressing fine-grained sentiment analysis. The model achieved an F1-score of 82.0% for aspect extraction and an accuracy of 79.8% for emotion classification, outperforming established baselines such as ABSA and DLED. In addition to its overall performance gains, the model demonstrated balanced accuracy across diverse emotion categories and aspect types, reinforcing the advantage of treating aspect extraction and emotion detection as a joint learning task rather than isolated components.

Furthermore, the analysis of aspect-emotion distributions revealed actionable insights with practical relevance for businesses. Specifically, aspects such as design and product quality were identified as primary drivers of positive emotions, whereas customer service and delivery time were frequently associated with negative emotions, particularly anger and sadness. This fine-grained sentiment perspective allows organizations to move beyond generic feedback analysis and strategically prioritize improvements in areas that exert the greatest emotional influence on the overall customer experience. Nevertheless, this study acknowledges several limitations that present opportunities for further research. These include occasional inaccuracies in aspect boundary detection, confusion between semantically similar emotions, and challenges in processing multi-aspect sentences where overlapping sentiments may occur. Such issues underscore the need for future work to explore more sophisticated attention mechanisms, the integration of external knowledge bases to enhance semantic understanding, and strategies to mitigate class imbalance among emotion categories, thereby improving the model's robustness and generalizability.

In summary, this research offers both methodological and practical contributions. Methodologically, it advances the field of sentiment analysis through the introduction of a novel joint-learning framework for aspect-level emotion detection. Practically, it enables businesses to derive a more nuanced, empathetic, and actionable understanding of customer feedback. These findings lay the groundwork for future studies to further enhance model performance and extend applicability across multiple domains, languages, and modalities.

6. Declarations

6.1. Author Contributions

Conceptualization: A.S.P., J.J.; Methodology: A.S.P., J.J.; Software: A.S.P.; Validation: J.J.; Formal Analysis: A.S.P.; Investigation: A.S.P.; Resources: J.J.; Data Curation: A.S.P.; Writing – Original Draft Preparation: A.S.P.; Writing – Review and Editing: A.S.P., J.J.; Visualization: A.S.P.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

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

References

[1] E. Cerdá-Mansilla, N. Rubio, and S. Campo, "Viralization keys of messages in unofficial accounts during crisis periods: the case of Covid-19 on Twitter," *Spanish J. Mark. - ESIC*, vol. 25, no. 1, pp. 137–155, Jan. 2021, doi: 10.1108/SJME-09-2020-0170.

[2] A. Gour, S. Aggarwal, and M. Erdem, "Reading between the lines: analyzing online reviews by using a multi-method Web-analytics approach," *Int. J. Contemp. Hosp. Manag.*, vol. 33, no. 2, pp. 490–512, Jan. 2021, doi: 10.1108/IJCHM-07-2020-0760.

[3] F. Tang, L. Fu, B. Yao, and W. Xu, "Aspect based fine-grained sentiment analysis for online reviews," *Inf. Sci. (Ny).*, vol. 488, no. 1, pp. 190–204, 2019, doi: 10.1016/j.ins.2019.02.064.

[4] A. Gupta, S. Priyani, and R. Balakrishnan, "Customized reputation generation of entities using sentiment analysis," *World J. Eng.*, vol. 18, no. 4, pp. 596–605, Jan. 2021, doi: 10.1108/WJE-09-2020-0470.

[5] R. Inomata, N. Kobayashi, M. Nakada, and S. Shirasaka, "Proposal of Customer Value Consistency Canvas, Using an Ontology of Value Proposition with Service Dominant Logic," *Rev. Integr. Bus. Econ. Res.*, vol. 9, no. 4, pp. 47–63, 2020.

[6] M. G. Majumder, S. D. Gupta, and J. Paul, "Perceived usefulness of online customer reviews: A review mining approach using machine learning and exploratory data analysis," *J. Bus. Res.*, vol. 150, no. 1, pp. 147–164, 2022.

[7] M. Godovykh and A. D. A. Tasci, "Customer experience in tourism: A review of definitions, components, and measurements," *Tour. Manag. Perspect.*, vol. 35, n o. 1, pp. 10-24, 2020.

[8] D. Sunitha, R. K. Patra, N. V Babu, A. Suresh, and S. C. Gupta, "Twitter Sentiment Analysis Using Ensemble based Deep Learning Model towards COVID-19 in India and European Countries," *Pattern Recognit. Lett.*, vol. 2022, no. 1, pp. 1-12, 2020, doi: 10.1016/j.patrec.2022.04.027.

[9] M. Bibi, W. A. Abbasi, W. Aziz, S. Khalil, M. Uddin, C. Iwendi, and T. R. Gadekallu, "A novel unsupervised ensemble framework using concept-based linguistic methods and machine learning for Twitter sentiment analysis," *Pattern Recognition* Letters, vol. 158, no. 1, pp. 80-86, 2022, doi: 10.1016/j.patrec.2022.04.004.

[10] F. Mehraliyev, I. C. C. Chan, and A. P. Kirilenko, "Sentiment analysis in hospitality and tourism: a thematic and methodological review," *Int. J. Contemp. Hosp. Manag.*, vol. 34, no. 1, pp. 46–77, Jan. 2022, doi: 10.1108/IJCHM-02-2021-0132.

[11] Y. Y. Chen, N. Wiratunga, and R. Lothian, "Integrating selection-based aspect sentiment and preference knowledge for social recommender systems," *Online Inf. Rev.*, vol. 44, no. 2, pp. 399–416, 2020, doi: 10.1108/OIR-02-2017-0066.

[12] P. Ray and A. Chakrabarti, "A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis," *Appl. Comput. Informatics*, vol. 18, no. 1/2, pp. 163–178, 2022.

[13] M. Venugopalan and D. Gupta, "An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis," *Knowledge-Based Syst.*, vol. 246, no. 1, pp. 1-20, 2022, doi: 10.1016/j.knosys.2022.108668.

[14] Ł. Augustyniak, T. Kajdanowicz, and P. Kazienko, "Comprehensive analysis of aspect term extraction methods using various text embeddings," *Comput. Speech Lang.*, vol. 69, no. 1, pp. 10-27, 2021, doi: 10.1016/j.csl.2021.101217.

[15] A. Montoyo, P. Martínez-Barco, and A. Balahur, "Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments," *Decis. Support Syst.*, vol. 53, no. 4, pp. 675–679, 2012, doi: 10.1016/j.dss.2012.05.022.

[16] E. Öhman, M. Pàmies, K. Kajava, and J. Tiedemann, "XED: A Multilingual Dataset for Sentiment Analysis and Emotion Detection," *COLING 2020 - 28th Int. Conf. Comput. Linguist. Proc. Conf.*, vo. 2020, no. 2007, pp. 6542–6552, 2020, doi: 10.18653/v1/2020.coling-main.575.

[17] A. S. Imran, S. M. Daudpota, Z. Kastrati, and R. Batra, "Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets," *IEEE Access*, vol. 8, no. 1, pp. 181074–181090, 2020.

[18] H. Luo, "Emotion Detection for Spanish with Data Augmentation and Transformer-Based Models," *Proc. Iber. Lang. Eval. Forum (IberLEF 2021)*, vol. 2021, no. 1, pp. 1-21, 2021.

[19] H. H. Do, P. W. C. Prasad, A. Maag, and A. Alsadoon, "Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review," *Expert Syst. Appl.*, vol. 118, no. 1, pp. 272–299, 2019, doi: 10.1016/j.eswa.2018.10.003.

[20] F. M. Del Arco, J. Collado-Montañez, L. A. Ureña, and M.-T. Martín-Valdivia, "Empathy and Distress Prediction using Transformer Multi-output Regression and Emotion Analysis with an Ensemble of Supervised and Zero-Shot Learning Models," in Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, *Sentiment and Social Media Analysis*, vol. 2022, no. 1, pp. 239–244, 2022.

[21] S. M. Orebi and A. M. Naser, "Opinion Mining in Text Short by Using Word Embedding and Deep Learning," *J. Appl. Data Sci.*, vol. 6, no. 1, pp. 526–636, Dec. 2024.

[22] M. K. Lim, Y. Li, and X. Song, "Exploring customer satisfaction in cold chain logistics using a text mining approach," *Ind. Manag. Data Syst.*, vol. 121, no. 12, pp. 2426–2449, Jan. 2021, doi: 10.1108/IMDS-05-2021-0283.

[23] S. Hu, A. Kumar, F. Al-Turjman, S. Gupta, and S. Seth, "Reviewer credibility and sentiment analysis based user profile modelling for online product recommendation," *IEEE Access*, vol. 8, no. 1, pp. 26172–26189, 2020.

[24] C. Su, L. Kong, F. Ciabuschi, and U. Holm, "Demand and willingness for knowledge transfer in springboard subsidiaries of Chinese multinationals," *J. Bus. Res.*, vol. 109, no. March 2019, pp. 297–309, 2020, doi: 10.1016/j.jbusres.2019.12.014.

[25] V. Yadav, P. Verma, and V. Katiyar, "E-Commerce Product Reviews Using Aspect Based Hindi Sentiment Analysis," 2021 *Int. Conf. Comput. Commun. Informatics, ICCCI 2021*, vol. 2021, no. 1, pp. 1-7, 2021, doi: 10.1109/ICCCI50826.2021.9402365.

[26] R. Wadawadagi and V. Pagi, "Polarity enriched attention network for aspect-based sentiment analysis," *Int. J. Inf. Technol.*, vol. 14, no. 6, pp. 2767–2778, 2022.

[27] R. H. Hama Aziz and N. Dimililer, "SentiXGboost: enhanced sentiment analysis in social media posts with ensemble XGBoost classifier," *J. Chinese Inst. Eng. Trans. Chinese Inst. Eng. A*, vol. 44, no. 6, pp. 562–572, 2021, doi: 10.1080/02533839.2021.1933598.

[28] A. N. Azhar and M. L. Khodra, "Fine-tuning pretrained multilingual bert model for indonesian aspect-based sentiment analysis," in 2020 7th *International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA)*, IEEE, vol. 2020, no. 1, pp. 1–6, 2020.

[29] P. R. Amalia and E. Winarko, "Aspect-Based Sentiment Analysis on Indonesian Restaurant Review Using a Combination of Convolutional Neural Network and Contextualized Word Embedding," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 15, no. 3, pp. 285-292, 2021, doi: 10.22146/ijccs.67306.

[30] H. Jafarian, A. H. Taghavi, A. Javaheri, and R. Rawassizadeh, "Exploiting BERT to improve aspect-based sentiment analysis performance on Persian language," in 2021 7th *International Conference on Web Research (ICWR)*, IEEE, vol. 2021, no. 1, pp. 5–8, 2021.

[31] R. Riyanto, "Application of the Vector Machine Support Method in Twitter Social Media Sentiment Analysis Regarding the Covid-19 Vaccine Issue in Indonesia," *J. Appl. Data Sci.*, vol. 2, no. 3, pp. 102–108, 2021, doi: 10.47738/jads.v2i3.40.

[32] B. Sabur, N. Ashraf, G. Sidorov, and A. Gelbukh, "Sexism identification using BERT and data augmentation-EXIST2021," *Int. Conf. Spanish Soc. Nat. Lang. Process. Sepln* 2021, vol. 2021, no. 1, pp. 1-7, 2021.