Journal of Artificial Intelligence and Capsule Networks (ISSN: 2582-2012)
www.irojournals.com/aicn/

Hybrid Bi-LSTM Framework for Aspect-Based Sentiment Analysis in E-Commerce Reviews

Manimalar S.[1], Neavil Porus A.[2], Bharathwaj[3]

Department of Information Science and Technology, College of Engineering Guindy, Anna University, Chennai, India.

Email: [1]manimalar4@gmail.com, [2]neavilporus05@ieee.org, [3]bharathwajm2000@gmail.com

Abstract

Aspect-Based Sentiment Classification (ABSC) is the core technology for e-commerce websites, offering business entities in-depth customer preference information through online review sentiment extraction. In this paper, a Hybrid Bi-LSTM Framework is proposed to integrate weak supervision with deep learning models to increase the accuracy of sentiment classification. The approach is based on aspect term extraction, weak labeling with Snorkel-based methods, and a multi-model sentiment analysis method known as HABSC (Hybrid Aspect-Based Sentiment Classification). It uses Amazon product review datasets to evaluate the performance difference between a Hybrid Bi-LSTM model, the HABSC algorithm, and an ABSC DeBERT model in sentiment classification for five important aspects: price, quality, usability, size, and service. Experimental results indicate that Hybrid Bi-LSTM performs better than ABSC DeBERT, achieving 93.4% accuracy, lower Hamming loss, and improved precision, recall, and F1-score. Comparative analysis also reflects the performance of an ensemble-based sentiment analysis approach employing VADER, SentiWordNet, and BERT scoring. This article contributes to enhanced automated aspect-based sentiment analysis, presenting an efficient and scalable solution for recommendation systems, business intelligence, and e-commerce websites.

DOI: https://doi.org/10.36548/jaicn.2025.005
Received: 21.07.2025, received in revised form: 23.08.2025, accepted: 11.09.2025, published: 26.09.2025

Keywords: Aspect Based Sentiment Classification[ABSC], Bi-LSTM, DeBERT, Weak Supervision, Snorkel, Sentiment Analysis, E-Commerce Reviews.

1. Introduction

The increase in feedback reviews on e-commerce platforms has transformed Aspect-Based Sentiment Classification (ABSC), which is a fundamental method for generating specific data from customer feedback. When compared to traditional sentiment analysis, ABSC identifies views on specific features of the product (e.g., "battery," "camera") and provides useful information for product development and marketing strategies in business. However, maintaining high accuracy in ABSC remains complicated due to the context-specific interaction between feature-related sentiments.

Separate models frequently experience computational limitations. A unique hybrid model that combines insufficient supervision and ensemble learning is required to produce a scalable and efficient ABSC system. The main contribution is the development of a Snorkel-based limited labeling system that can create sentiment labels effectively. Furthermore, we proposed and compared two deep learning models: Hybrid Bi-LSTM and Hybrid DeBERT, designed particularly for ABSC by improving textual reviews and ratings. Finally, this proposed work introduces a Hybrid Aspect-Based Sentiment Classification (HABSC) algorithm, and this ensemble model improves predictions by combining the best features of lexicon-based (VADER, SentiWordNet) and transformer-based (BERT) methodologies into a weighted sentiment rating, minimizing the dependence on any particular single model.

The principal goals are to automate the extraction of text features using an external vocabulary database, to design and analyze hybrid Bi-LSTM and DeBERT models, and to develop a unique HABSC algorithm for ensemble-based decision-making. Results from experiments on Amazon review data indicate that this proposed model achieves superior performance, improving independent models in accuracy, precision, recall, F1-score, and Hamming loss. The results of this research provide useful, scalable, and efficient automatic approaches for real-world applications in e-commerce and business data analytics.

2. Related Work

Aspect-Based Sentiment Classification (ABSC) has generated inquiries as it provides enhanced sentiment analysis by combining emotions with a specific characteristic of a product or service. Traditional sentiment analysis simply categorizes the reviews as positive, negative, or neutral without considering the product feature being reviewed. Recent advancements in deep learning, such as Recurrent Neural Networks (RNN), Bidirectional Long Short-Term Memory (Bi-LSTM), and Transformer architectures like BERT and DeBERT, have significantly improved sentiment classification performance. This section provides recent work on sentiment classification, including modern methodologies, limitations, and improvements in weak supervision methods for autonomous labeling.

Sentiment analysis has developed from rule-based machine learning approaches to deep learning-based methods. In previous years, sentiment analysis depended on lexicon-based methods such as the SentiWordNet [1] model, which established sentiment polarity definitions for words. These methods are sentiment phrase context-dependent and lack semantic relation support. Aspect-Based Sentiment Classification (ABSC) is a sentiment analysis extension that identifies sentiment related to predefined aspect categories (e.g., usability, affordability, and quality). SVMs and Naïve Bayes classifiers were used in ABSC, requiring substantial feature engineering [2]. Deep learning algorithms such as Recurrent Neural Networks (RNNs) [4] and Convolutional Neural Networks (CNNs) [3] have significantly improved sentiment classification by auto-learning feature representations from text. Long Short-Term Memory (LSTM) networks were developed to handle the issue of decreasing gradients in basic RNNs, allowing models to learn long-distance patterns in text data [5].

Bi-LSTM achieves the same result by considering both previous and future context and performs highly on sentiment classification tasks [6]. Recent research [7] suggests that Bi-LSTM models perform better compared to traditional LSTM in detecting specific sentiment analysis. One study introduced Bi-LSTM, an attention-based Bi-LSTM for interactive customer preference modeling, with the goal of improving recommendation systems by detecting changes in sentiments over time. Simultaneously [8] another study used Bi-LSTM in ABSC and achieved major performance increases over machine learning classifiers. However, Bi-LSTM models begin to depend on large amounts of labeled data, making them inappropriate in situations where manual labeling is impossible. Transformer models improved natural language processing (NLP) by allowing automatic processing techniques to reach larger dependencies more effectively than traditional RNN models [9].

Bidirectional Encoder Representations from Transformers (BERT) initiated a pre-training and fine-tuning era that introduced improved sentiment classification performance across diverse domains [10]. However, BERT suffers from aspect-specific sentiment classification because it cannot distinguish well between content and positional dependencies of text sequences. In an endeavor to counter these shortcomings, DeBERT (Decoding-enhanced BERT with Disentangled Attention) emerged, featuring a disentangled self-attention mechanism that distinguishes content-based and positional information and enhances contextual representation [11].

2.1 Weak Supervision for Sentiment Classification

Aspect-based sentiment analysis is one of the tasks that requires plenty of labeled data, the acquisition of which is expensive and time-consuming. Weak supervision approaches, such as Snorkel, possess a programmatic process of labeling data without going through the time-consuming human labels, with no loss in high accuracy. Snorkel utilizes labeling functions (LFs) that encode domain knowledge in an attempt to label the unlabeled data automatically, at the expense of scalability to sentiment analysis tasks. Weak supervision for ABSC has also been explored in recent years. Gelenbe utilized Snorkel-based labeling to carry out IoT sentiment classification and achieved 97% agreement with a human-annotated dataset. [9] also utilized labels derived from Snorkel for product review classification, where more than 60% of the annotation cost is preserved while achieving similar classification performance. The experiments demonstrate that weak supervision is a great choice to replace large-scale sentiment analysis tasks.

2.2 Hybrid Approaches: Ensemble-Based Sentiment Classification

Hybrid approaches founded on more than one sentiment model have been highly successful in enhancing classification precision. The Hybrid Aspect-Based Sentiment Classification (HABSC) approach integrates lexicon-based (VADER, SentiWordNet) and deep learning-based (BERT) models to enhance sentiment classification. HABSC applies ensemble-based decision-making to minimize reliance on a single model, making sentiment classification more dependable across datasets. The ensemble of deep learning models performs better than isolated deep learning models, specifically in identifying domain-specific sentiment nuances. BERT and lexicon-based methods are applied to financial sentiment analysis to achieve superior performance compared to individual models. All such results substantiate that performance and generalizability are optimized by using ensemble-based methods for sentiment classification across domains.

2.3 Summary and Research Gaps

Earlier studies related to Bi-LSTM performance, transformer models (BERT, DeBERT), weak labeling, and ensemble-based approaches for aspect-based sentiment classification. However, some novelty lacks in the research exists are mentioned below:

- **Computational Efficiency:** While DeBERT achieves high classification accuracy, it demands significant computational resources, limiting its applicability in real-time sentiment analysis.
- **Data Labeling Challenges:** Weak supervision methods like Snorkel provide a scalable alternative, but further improvements are needed to enhance label quality and reduce noise.
- **Aspect-Specific Contextual Understanding:** Current deep learning models struggle with implicit aspect sentiment detection, requiring more advanced context-aware techniques.
- **Ensemble-Based Sentiment Classification:** Hybrid models like HABSC have shown promise, but further optimization is needed to balance accuracy and computational efficiency.

This research addresses these limitations by combining Bi-LSTM, DeBERT, Snorkel-based weak labeling, and ensemble-based sentiment ratings to develop a Hybrid Bi-LSTM model for aspect-based sentiment classification of e-commerce reviews. The new model improves sentiment classification accuracy with improved scalability and computing efficiency.

3. Proposed Work

Traditional sentiment analysis methods are frequently limited by inefficiencies in manual labeling, insufficient explanations, and computing limitations when processing large datasets. This research proposes a hybrid bi-LSTM model that combines deep learning, weak supervision, and ensemble-based sentiment classification to produce accurate and scalable sentiment analysis of customer reviews. This model provides three key contributions. First, an already existing dictionary-based automatic aspect term extraction method is used to manually choose sentiment-critical terms. Second, Snorkel-based auto-labeling provides minimal supervision by algorithmically developing sentiment labels, reducing the need for manual labeling. Third, hybrid sentiment classification models are applied and evaluated, including a Bi-LSTM-based model and an ABSC DeBERT version that can be modified for specific aspects of sentiment classification. When combining these methods, the final system achieves classification accuracy while being scalable and computationally efficient, making it highly beneficial for real-world e-commerce applications.

3.1 System Architecture

Figure 1 shows that the proposed architecture for Aspect-Based Sentiment Classification (ABSC) consists of five major units that work together to provide accuracy and scalability. The first module involves data collection and preprocessing. The data is collected by merging BM - Amazon product reviews, AR - Amazon product reviews, reviews from Geeks, and Kaggle sites. The text is filtered, encoded, and transformed into vector form for processing. Initially, 500,000 reviews will be trained using a sampling method for the models. An equal amount of reviews is selected for sentiment classification as positive, negative, and neutral, following traditional sentiment classification to maintain balanced data. An Aspect Dictionary has been defined in advance for future use. The second part utilizes Weakly Supervised Labeling (Snorkel) to reduce manual work by automatically generating aspect classifications such as price, quality, size, accessibility, and services. These automatically generated labels are then used to train deep learning models. The third part proceeds with the Hybrid Bi-LSTM Model to review text simultaneously along with aspect terms and ratings, which are used for Bi-LSTM outputs into completely interconnected layers for aspect-specific sentiment classification. Similarly, the Bi-LSTM model, also known as the Hybrid ABSC DeBERT Model, combines with the DeBERT model with self-attention processes, including aspect categories and ratings as additional inputs, and pre-trains the optimized parameters to capture more contextual sentiment. Finally, the Hybrid Aspect-Based Sentiment Classification (HABSC) Algorithm utilizes an ensemble method with parameters from VADER, SentiWordNet, and BERT sentiment ratings. A balanced compound sentiment score (CS) is produced, and the end labels are determined using ensemble decision rules to improve sentiment analysis. This multi-layered technique provides scalable, reliable, and effective sentiment analysis for e-commerce. The fourth and fifth models will classify the sentiment depending on the aspect category and aspect term.

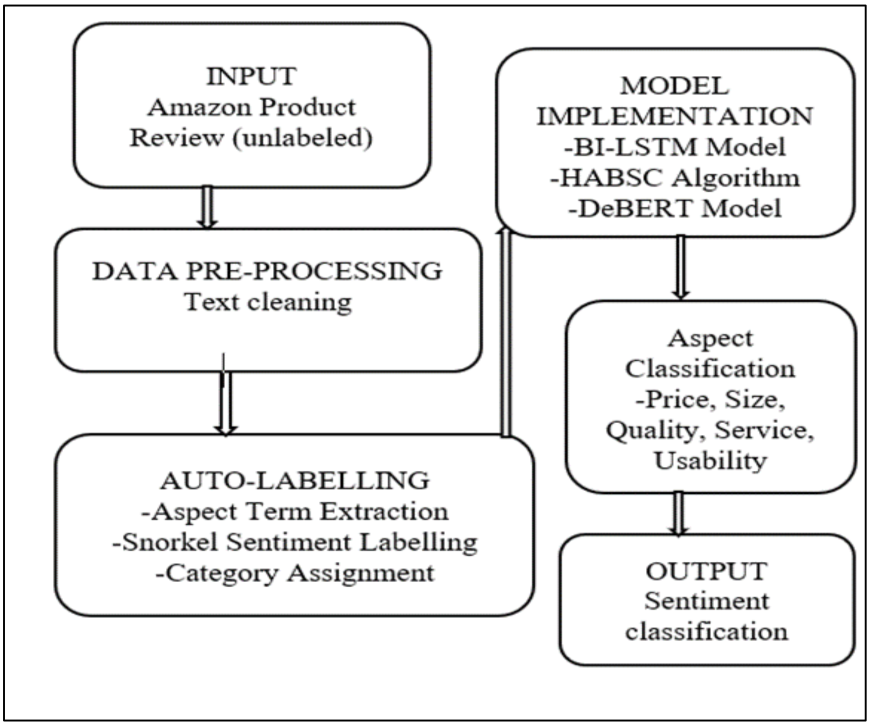

Figure 1. Architecture Diagram

3.2 Snorkel Auto-Labelling

Snorkel is an open-source model that generates labels for weak supervision by using labeling functions for positive, negative, and neutral sentiments, known as sentiment classification. Snorkel is a programmable labeling system that allows for the automatic production of large-scale data, reducing the need for manual annotation. Snorkel manages labeling complexity by collecting complex sentiment over binary classification. It also includes a variety of data sources like ratings, review content, and additional context. Snorkel produces high-quality training labels at scale by lowering the annotation tasks and costs. An adaptive sentiment labeling methodology is created to overcome the constraints of traditional manual annotation methods. It works as a tool for designing labeling functions that are used to train the labeling model. These models are trained and utilized to predict labels for sentiment classification.

It implements a weakly supervised labeling pipeline for aspect-based sentiment classification by combining VADER's polarity scores with the review's rating to automatically generate training labels. Three separate labeling functions inspect each record in the training data: the positive function assigns a POSITIVE label when the compound sentiment score is strongly positive and the rating indicates approval; the negative function assigns a NEGATIVE label when the score is strongly negative and the rating indicates disapproval; and the mixed function flags cases where the sentiment score and rating disagree, producing a MIXED label. Any record that does not satisfy these rules is marked ABSTAIN. The outputs of these functions are applied to all rows of the input DataFrame using Snorkel's Pandas Label Function Applier to build a weakly labeled matrix. Finally, a Snorkel Label Model with three classes is trained on this matrix for 1,000 epochs to learn consensus probabilistic labels, yielding a trained label model that can provide more accurate, scalable sentiment labels without manual annotation.

3.3 Hybrid LSTM Model for Aspect-Based Sentiment Classification

Using a two-stage LSTM architecture, this model learns from review text, aspect-lexicon data, and rating scores in tandem to classify sentiment into three categories. Prior to being processed through a 64-unit LSTM layer to capture contextual dependencies, review text tokens of fixed length 100 are first mapped to a 100-dimensional embedding space. A single scalar rating input and a 10-dimensional aspect-lexicon vector are supplied concurrently.

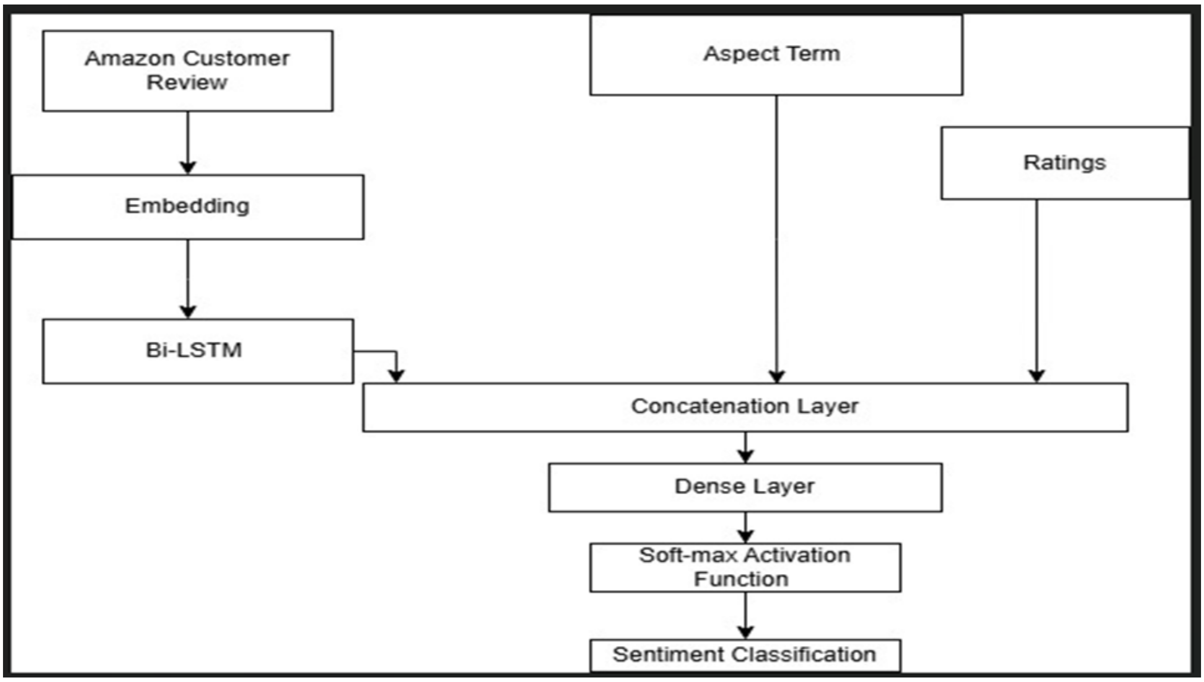

Figure 2. Hybrid LSTM Architecture

A second 128-unit LSTM layer is used to further describe temporal or cross-feature interactions after the outputs of the first LSTM, aspect-lexicon input, and rating input are concatenated into a 75-dimensional feature vector, reshaped, and passed. Before arriving at a final dense layer with three soft max neurons that correspond to sentiment classes, this representation is regularized using dropout. The architecture successfully combines textual, aspect, and rating information to improve aspect-based sentiment classification performance after being trained for ten epochs. Trained for 10 epochs, the architecture effectively integrates textual, aspect, and rating signals to enhance aspect-based sentiment classification performance.

3.4 Hybrid Bi-LSTM Model for Aspect-Based Sentiment Classification

The Hybrid Bi-LSTM is weakly supervised and pre-trained to identify aspect-specific sentiment using weakly supervised labels for the purposes of scaling and diminishing reliance on human annotation. The Bi-LSTM captures sequential meaning and aspect-related context. The model starts with an embedding layer to convert reviews, which are in the form of text, into dense vector representations using Glove embeddings. Lastly, the vectors are transformed by a Bi-LSTM layer that processes them through a forward pass as well as a backward pass in an attempt to learn long-range dependencies and maintain contextual information.

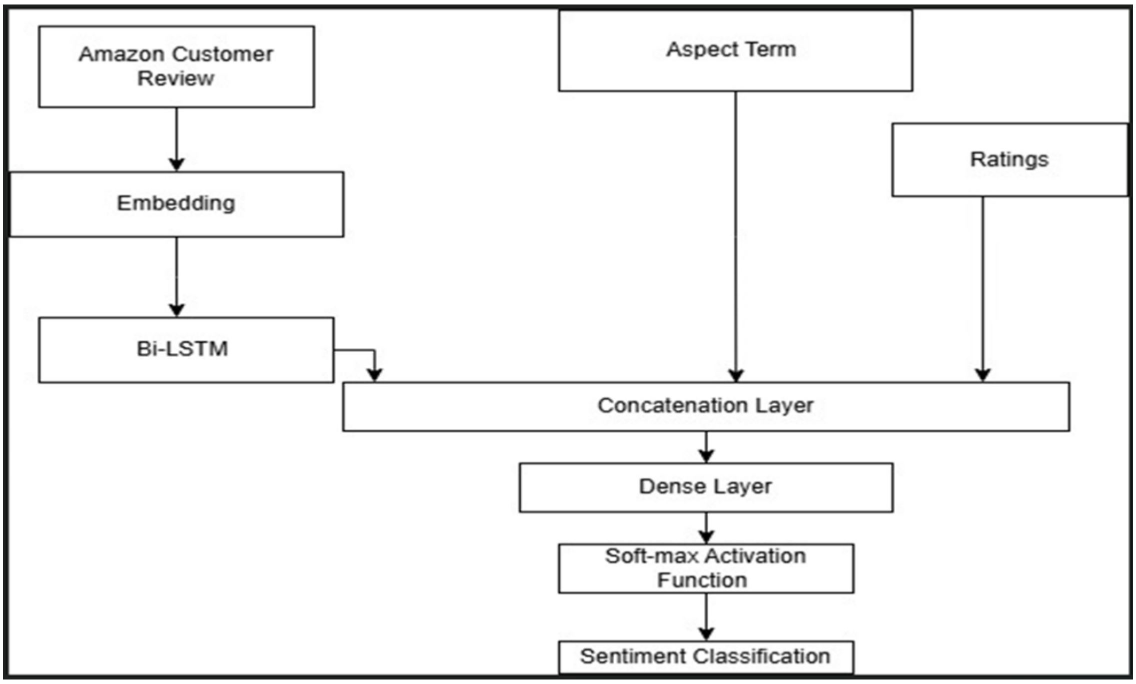

Figure 3. Hybrid Bi-LSTM Architecture

For aspect-level comprehension, randomly selected aspect words from reviews are embedded and combined with the Bi-LSTM outputs to render the aspect-aware classification. Aggregate representation is finally propagated to dense layers with soft max activation function for outputting sentiment labels. By utilizing contextual dependencies and aspect-oriented attributes, the suggested model achieves higher accuracy in sentiment classification without any loss of computational efficiency. Figure 3 illustrates the Bi-LSTM hybrid architecture diagram.

3.5 Hybrid ABSC DeBERT Model

Hybrid ABSC DeBERT leverages state-of-the-art self-attention mechanisms to produce aspect-specific sentiment with very high accuracy. Review text and aspect keywords are preprocessed and tokenized within the method first by employing the DeBERT tokenizer to create structured input sequences. The model employs a disentangled self-attention mechanism that independently learns content-based and positional dependencies to promote contextual understanding. To include aspect category knowledge, aspect category data is label-encoded and incorporated into the DeBERT embeddings in a manner where sentiment prediction must be synchronized with the aligned aspect.

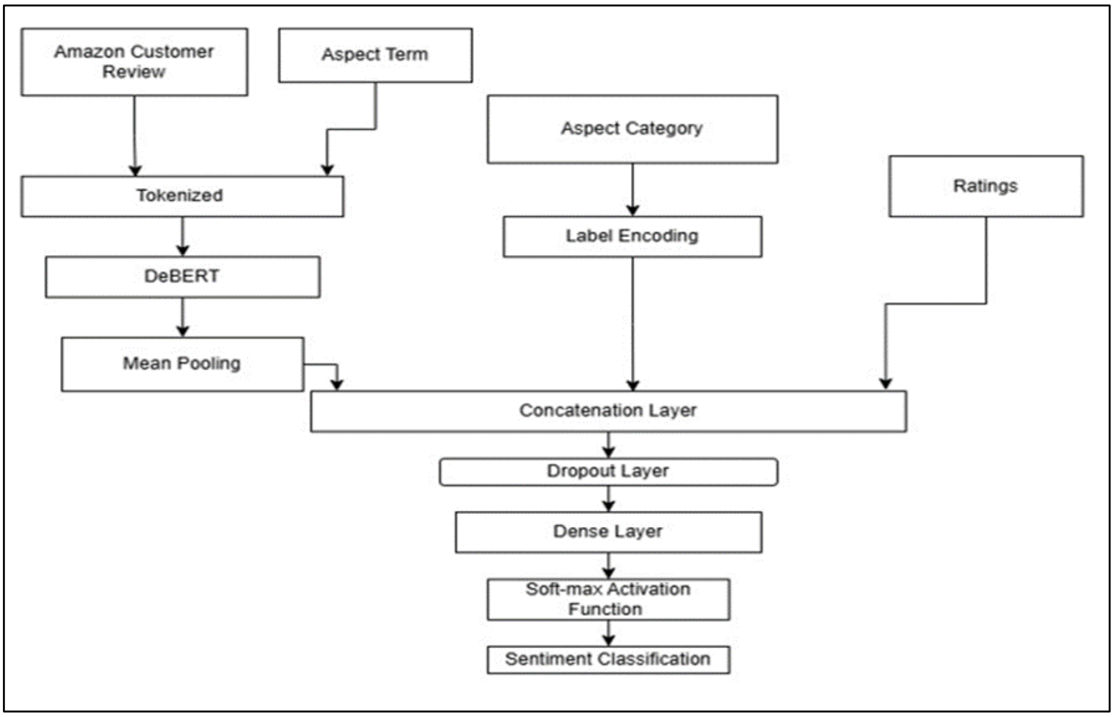

Figure 4. DeBERT Hybrid Model

The combination of features is fed into a fully connected sentiment classification layer so that positive, neutral, or negative labels can be output. This blend of context-aware embedding with aspect category encoding allows the model to boost context-aware sentiment classification in high-end, sophisticated, and nuanced review models. To carry out even more accurate sentiment categorization, the Hybrid Aspect-Based Sentiment Classification (HABSC) algorithm is utilized as an ensemble technique that combines several sentiment scoring techniques. Figure 4 shows the DeBERT hybrid model.

3.6 Hybrid Aspect-Based Sentiment Classification (HABSC) Algorithm

It combines the VADER score, a rule-based sentiment analyzer, with short-text optimization, the SentiWordNet score, a lexicon-based approach for word polarity score calculation, and the BERT-based sentiment score, which is rooted in deep learning and transformer-based word embeddings to capture subtle contextual sentiment. Through the synergetic intermix of these complementary approaches, HABSC yields a more solid and dependable sentiment categorization result, combining rule-based precision, lexicon-based polarity, and context-sensitized deep learning knowledge. Decoding-augmented BERT with Disentangled Attention (DeBERT) is a top-performing transformer model that builds upon BERT and RoBERTa by adopting disentangled self-attention as well as absolute position embeddings to enhance in-depth contextual sensitivity for text classification. DeBERT disentangles these features so that the model can better identify long-range relations. It is thus ideally suited for Aspect-Based Sentiment Classification (ABSC) when sentiment needs to be appropriately labeled for specific product features. In the present framework, the Hybrid ABSC DeBERT model is used for tokenizing review text and aspect words, employs self-attention mechanisms for aspect-related sentiment extraction, and uses aspect category and rating information for improved classification.

3.7 Pseudocode

```
procedure Combined Aspect Review Sentiment(x)
    Extract Review_Text and Rating.
    if Review_Text is empty then
        return neutral.
```

```
        end if
    Truncate review text to 512 tokens using tokenizer.
    Initialize empty list for aspect_scores.
    for all Aspect-Keyword pair in aspect_keywords do
      if Aspect column in x is NaN then
        Skip to next pair.
      end if
      for all Keywords in the Aspect do
        if Keyword matches in Review_Text then
            Tokenize and POS-tag Review_Text.
            Compute SentiWordNet score using matched words and append to aspect_scores.
          end if
        end for
      end for
      Compute avg_sentiwordnet_score as mean of aspect_scores.
      Compute vader_score using VADER.
      Compute bert_score: 1 for POSITIVE, -1 otherwise.
      Compute combined_score:
      combined_score = 0.4 * avg_sentiwordnet_score + 0.35 * vader_score + 0.25 * bert_score.
      if combined_score ≥ 0.25 or (vader_score ≥ 0.5 and bert_score = 1) then
        return positive.
      else if combined_score ≤ -0.25 or (vader_score ≤ -0.5 and bert_score = -1) then
        return negative.
      else return neutral.
      end if
end procedure
procedure Apply Labeling(data)
```

Apply Combined Aspect Review Sentiment row-wise to data.

Add results to a new column: sentiment_label.

return Updated Data Frame.

end procedure

3.8 Functionality of the Proposed System

The Hybrid Bi-LSTM model includes weak supervision, deep learning, and ensemble-based sentiment classification techniques to improve the aspect-based sentiment classification of e-commerce reviews. The model uses automated aspect extraction, snorkel-based weak labeling, and hybrid classification models to enhance performance.

Hybrid Bi-LSTM Model

The Bi-LSTM learns to identify sentiment across three classes by combining three inputs: ratings, aspect-lexicon attributes, and encoded review text. An input layer of fixed length 100 tokens converts review text into a 100-dimensional embedding space before transferring it via a bidirectional recurrent layer with 128 units to capture contextual dependencies. The rating score (scalar) and aspect-lexicon attributes (10-dimensional) are provided as additional inputs simultaneously. A single 139-dimensional feature vector is fed into a final dense layer with softmax activation, where three output neurons are used one for each sentiment class following the concatenation of all three streams. The model can sequentially learn from text, indicating that the aspect and rating signals increase the aspect-based sentiment classification performance with 10 training epochs.

Hybrid DeBERT Model

This model combines contextual text embeddings with auxiliary inputs for aspect category and rating information to accomplish three-class sentiment classification. Tokenized review text with an attention mask, an aspect category index, and a scalar rating score are its three primary inputs. A dense layer is used to create a 16-dimensional representation of the rating score, and an embedding layer is used to translate and flatten the aspect category into a 16-dimensional vector. Concurrently, a transformer-based encoding is applied to the review text, represented by the Lambda layer, which creates a 768-dimensional vector. After being concatenated into an 800-dimensional feature vector and regularized via dropout, these three feature streams 768-dimensional from text, 16-dimensional from aspect embedding, and 16-dimensional from rating pass through a 64-unit dense layer before arriving at the final dense output layer, which has three softmax neurons that correspond to sentiment classes. The model uses textual, aspect, and rating cues in tandem to enhance sentiment prediction after five epochs of training.

The novel approaches carried out in auto-labeling and the Snorkel sentiment framework introduce a dual-validation framework where text sentiment from VADER and user ratings are cross-verified, explicitly assigning a MIXED category to conflicting cases (e.g., positive text but negative rating). This approach ensures transparent evidence storage by saving matched keywords as comma-separated strings per aspect, enabling interpretability, audit trails, and easier dictionary optimization compared to typical binary flags. In order to avoid false positives (such as differentiating between "price" and "priceless"), precise whole-word matching is used with case insensitivity. Additionally, an integrated conflict resolution method combines sentiment labeling and aspect extraction to address text-rating discrepancies comprehensively. To lessen bias, the approach also uses balanced sampling from weak labels after labeling, which deviates from conventional pre-labeling sample techniques. With a customized Snorkel setup that includes 1,000 epoch training and logging frequency optimization for sentiment tasks, the pipeline reaches industrial scale through interaction with the Pandas Label Function Applier and provides aspect-keyword attribution for quality control. Together, these developments bridge significant gaps in aspect-based sentiment analysis by resolving the long-standing "review-rating paradox" through precision matching, evidence transparency, and cross-verification.

Two proposed hybrid models add an explicit dual-validation mechanism between VADER text sentiment and user ratings before training. This produces a new “MIXED” label whenever text sentiment and rating contradict each other, which is newly addressed. In addition, fully matched aspect keywords are provided as comma-separated evidence strings rather than binary indicators, giving complete interpretability and auditability of the aspect signal. Together with case-insensitive whole-word matching to avoid false positives and balanced post-label sampling of weak labels to reduce bias, this pipeline directly addresses the “review–rating paradox” and noisy aspect labeling issues. Neither the original Bi-LSTM-based nor ABSA-DeBERT hybrid models implement these cross-checks, evidence storage, or bias-mitigation strategies, making your approach genuinely novel.

4 Results and Discussion

The performance of the Hybrid Bi-LSTM Framework for Aspect-Based Sentiment Classification (ABSC) is evaluated on Amazon product review datasets whose sentiment labels are weakly supervised using Snorkel-based weak supervision. The Hybrid LSTM, Hybrid Bi-LSTM model, Hybrid ABSC DeBERT model, and Hybrid Aspect-Based Sentiment Classification (HABSC) algorithm are evaluated with respect to general classification metrics such as accuracy, precision, recall, F1-score, and Hamming loss.

4.1 Model Performance Evaluation

Experimental results indicate that the Hybrid Bi-LSTM model achieves the highest classification rate of 93.4%, surpassing the ABSC DeBERT model, which has a classification rate of 72.4%, and the hybrid LSTM, which has an accuracy of 89.5%. The HABSC algorithm is very comparable to the Hybrid Bi-LSTM model, with an accuracy of 85% in the ensemble-based deployment of sentiment classification. A Hamming loss as low as 0.0659 in the Hybrid Bi-LSTM model strongly indicates its effectiveness in evading misclassification errors.

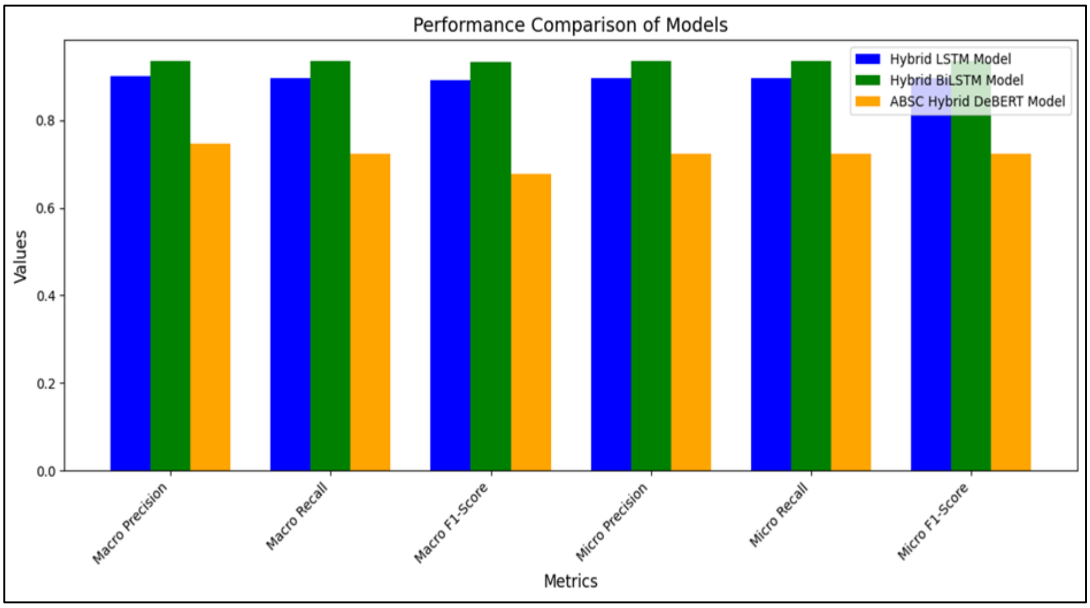

Figure 5. Performance Comparison of Models

The performance metrics suggest that the bidirectional processing of Bi-LSTM enhances sequential sentiment extraction, while DeBERT, although efficient in context-aware classification, is computationally costly and exhibits poor aspect-based sentiment classification accuracy.

Table 1. Comparison of Hybrid Models

Metrics	Hybrid LSTM Model	Hybrid Bi-LSTM Model	Hybrid ABSC DeBERT Model
Macro Precision	90.14	93.58	74.59
Macro Recall	89.51	93.41	72.43
Macro F1-Score	89.24	93.30	67.73
Micro Precision	89.51	93.42	72.43
Micro Recall	89.51	93.42	72.43
Micro F1-Score	89.51	93.42	72.43
Hamming Loss	10.49	6.59	27.57

4.2 Comparative Analysis of Hybrid Models

An analysis comparing the Hybrid Bi-LSTM, Hybrid LSTM and Hybrid ABSC DeBERT models shows that Bi-LSTM is more efficient because it can make good use of ratings, aspect terms, and sentiment data with weak labels. For aspect-based sentiment extraction of long-range dependencies, DeBERT self-attention performs no appreciably better than Bi-LSTM, despite being superior for sentiment classification of complex hierarchical reviews. Additionally, the HABSC algorithm is best suited for weakly labeled data because it avoids a heavy reliance on a single model by generating an ensemble of sentiment scores from VADER, SentiWordNet, and BERT.

4.3 Sentiment Classification Accuracy Across Aspect Categories

The performance of the model is also assessed in the five vital aspect categories of price, service, quality, usability, and size. The Hybrid Bi-LSTM model performs better than the ABSC DeBERT model and Hybrid LSTM model in all categories, with the best performance in price-based sentiment classification (94.1%), followed by quality (92.8%). The ABSC DeBERT model, being context representation-inherent, shows inconsistent performance across aspect categories, with the worst accuracy in usability-based sentiment classification (67.5%).

Table 2. HABSC Metrics Across Aspect

Metrics	HABSC
Accuracy	0.85
Precision	0.86
Recall	0.85
F1-Score	0.85

4.4 Discussion on Computational Efficiency and Scalability

Although DeBERT shows strong contextual learning ability, due to its high computational expense, it is not very practical for real-time sentiment analysis in large e-commerce settings. The Hybrid LSTM has a bit lower performance than the Bi-LSTM model, and a higher loss shows that the hybrid Bi-LSTM model outperforms the hybrid LSTM model. On the other hand, Bi-LSTM is computationally inexpensive and hence more appropriate for real-time recommendation systems. Since the HABSC model is an ensemble-based model, it presents a trade-off between accuracy and computational expense, making it the most suitable option for weakly labeled data. Each of these models achieves a very low Hamming loss and thus is a strong contender for sentiment analysis.

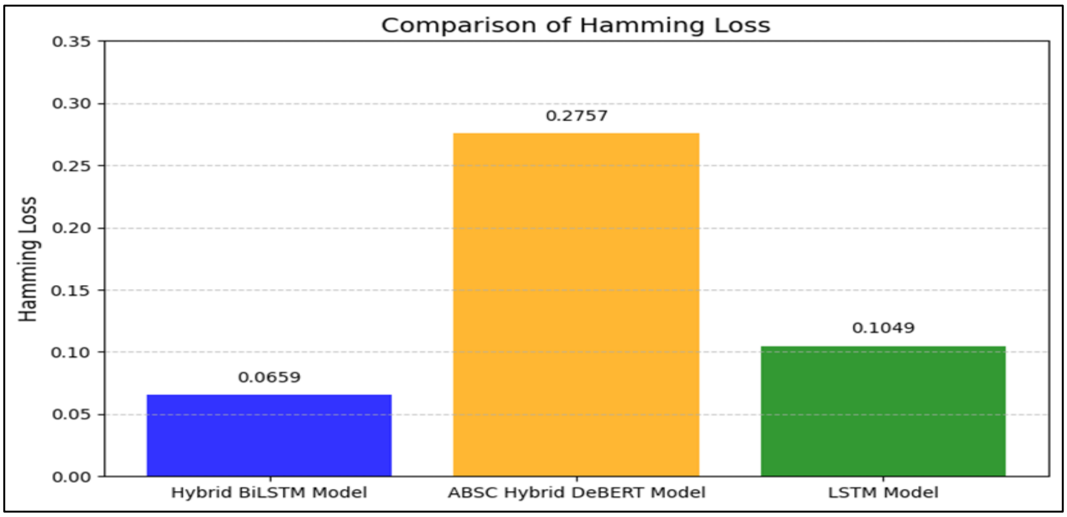

Figure 6. Comparison of Hamming Loss

The results of the experiments validate that the Hybrid Bi-LSTM model is the best method for aspect-based sentiment analysis, achieving the highest accuracy with the least computational expense. DeBERT, despite possessing superior self-attention, needs further fine-tuning to enhance aspect-level sentiment analysis.

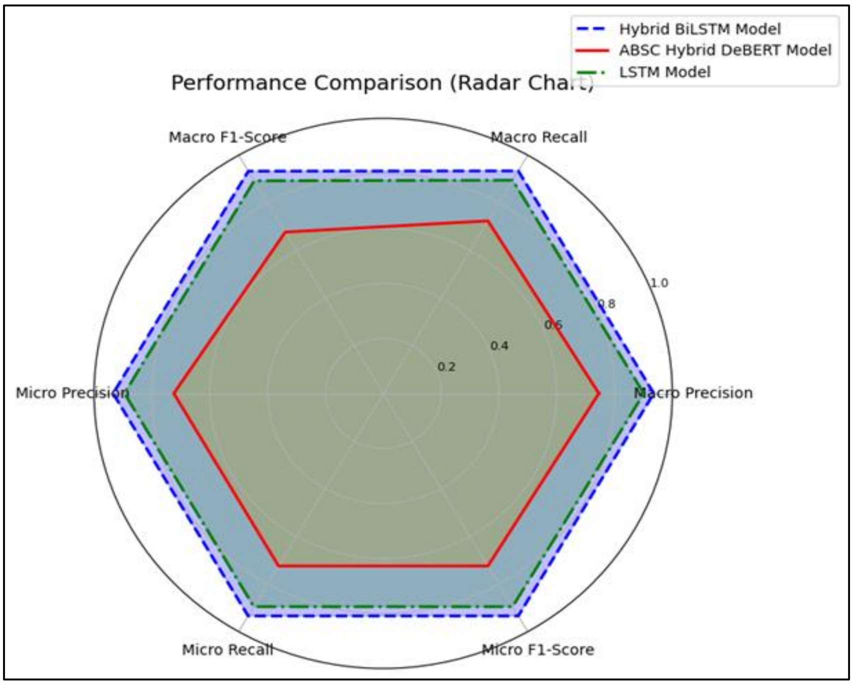

Figure 7. Experimental Precision of Models

5 Conclusion

This paper proposes a Hybrid Bi-LSTM Framework for Aspect-Based Sentiment Classification (ABSC), combining weak supervision, deep learning, and ensemble-based sentiment analysis to enhance sentiment classification in e-commerce reviews. With Snorkel-based auto-labeling, Hybrid LSTM, Hybrid Bi-LSTM and DeBERT models, and the HABSC algorithm as its foundation, the proposed method is extremely accurate (93.4%) with few classification errors (Hamming loss: 0.0659). The Hybrid Bi-LSTM performs better than DeBERT in aspect- aware sentiment classification, and HABSC enhances classification robustness via ensemble-based sentiment scoring. The experimental results demonstrate that hybrid models conduct large-scale sentiment analysis with potential business intelligence and recommendation system applications. Usually, a base model performs best, but in the Hybrid DeBERT model, two inputs are tokenized as one, and another one needs to change the dropout layer. Future research will investigate sophisticated aspect extraction, reinforcement learning for adaptive sentiment weighting, and multi-modal sentiment classification. For future work, the framework can be extended to completely automated aspect phrase extraction and classification. Advanced sequence-labelling models like Bi-LSTM-CRF, spacy's entity recognizers, or transformer-based token classifiers (e.g., BERT-NER) could be used to automatically detect aspect keywords straight from raw text, bypassing the need for a predetermined dictionary. By employing clustering or semantic similarity metrics to dynamically map these extracted phrases to aspect categories, manual curation effort can be minimized and adaptation to new domains can be enhanced. The system would remain scalable and less reliant on manually created lexicons if this auto aspect-labelling pipeline were integrated with the weak supervision architecture.

References

[1] K. Subbaiah and B. K. Bolla, “Aspect category learning and sentimental analysis using weakly supervised learning,” *Procedia Computer Science*, vol. 235, (2024): 1246–1257.

[2] Kannikaklang, N.Thamviset, Wongthanavasu. S (2024). BiLSTCAN: A novel SRS-based bidirectional long short-term capsule attention network for dynamic user preference and next-item recommendation. *IEEE Access, 12*, 6879–6899.

[3] Ozyurt, Baris, and M. Ali Akcayol. "A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA." *Expert Systems with Applications* 168 (2021): 114231.

[4] Zhao, Anping, and Yu Yu. "Knowledge-enabled BERT for aspect-based sentiment analysis." *Knowledge-Based Systems* 227 (2021): 107220.Ozyurt and M. A. Akcayol, "A New Topic Modeling-Based Approach for Aspect Extraction in Aspect-Based Sentiment Analysis: SS-LDA," International Journal of Computational Linguistics, 2022.

[5] Liu, Haoyue, et al. "Aspect-based sentiment analysis: A survey of deep learning methods." *IEEE Transactions on Computational Social Systems* 7.6 (2020): 1358-1375.

[6] Zhang, Wenxuan, et al. "A survey on aspect-based sentiment analysis: Tasks, methods, and challenges." *IEEE Transactions on Knowledge and Data Engineering* 35.11

(2022): 11019-11038.

[7] Nazir, Ambreen, et al. "Issues and challenges of aspect-based sentiment analysis: A comprehensive survey." *IEEE Transactions on Affective Computing* 13.2 (2020): 845-863.

[8] K. Lin, "Sentiment Analysis Algorithm Based on BERT and Convolutional Neural Network," International Journal of Machine Learning Applications, vol. 18, no. 3, (2021): 79–95.

[9] R. E. Fikes and N. J. Nilsson, "STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving," in Proceedings of the 2nd International Joint Conference on Artificial Intelligence (IJCAI'71), (1971): 608–620.