Multimedia Tools and Applications (2024) 83:22473–22501
https://doi.org/10.1007/s11042-023-16360-3

Aspect-based sentiment analysis of drug reviews using multi-task learning based dual BiLSTM model

Somiya Rani[1] **· Amita Jain**[2]

Received: 31 March 2022 / Revised: 24 May 2023 / Accepted: 17 July 2023 /
Published online: 7 August 2023

Abstract

User-generated content on healthcare web forums, particularly drug reviews, provides valuable information on drug benefits, effectiveness, side effects, dosage, condition, cost, and overall experiences. Applying Aspect-Based Sentiment Analysis (ABSA) can help researchers categorize sentiments toward specific aspects such as drug effectiveness, side effects, and treatment experiences. These insights are highly useful for healthcare professionals, pharmaceutical companies, and researchers to assess drug efficacy and safety, utilizing the vast amount of healthcare-related user-generated content available online. However, due to scarcity of annotated data for training ABSA models in the medical domain poses challenges in accurately extracting aspect terms. Also, the identification of implicit aspects poses a huge challenge as they frequently lack explicit names or keywords that directly indicate their presence. The domain-dependent nature of ABSA and the variability of term meanings across domains necessitate the incorporation of contextual information and semantic patterns. Therefore, we propose a novel model called **M**ulti-task **L**earning based **D**ual **B**idirectional LSTM **M**odel (MLDBM) for ABSA of drug reviews. The MLDBM leverages BERT and incorporates a multi-head self-attention mechanism to produce aspect-specific representations which are further processed through the dual BiLSTM model. This enables the model to capture and analyze sentiments related to different aspects discussed in the reviews. We also introduce various modifications to the MLDBM to identify the constraints of the proposed model. The proposed model outperforms state-of-the-art models by achieving a performance gain of 8% to 12% on two benchmark datasets, demonstrating its effectiveness when compared to various baseline models. ABSA applied to drug reviews contributes to enhancing healthcare quality by considering different aspects of drugs as shared by consumers.

Keywords Aspect-based Sentiment Analysis · Attention · BERT · Deep Learning · Drug Reviews · Dual BiLSTM · Multi-Head Self-Attention · Multi-task Learning

1 Introduction

The quantity of user-generated content related to medical topics on social media and health-related forums is enormous and it continues to grow. The rapid growth in the volume of user-generated content related to medical topics on social media and health-related forums has

✉ Amita Jain
amitajain@aiactr.ac.in

Extended author information available on the last page of the article

prompted research in the field of Natural Language Processing (NLP). One such productive area of research that emerged from NLP is Aspect Based Sentiment Analysis (ABSA). ABSA is a type of sentiment analysis (SA) that focuses on identifying a sentiment category expressed towards an aspect or multiple aspects within a sentence, where an aspect refers to a word or phrase that describes any attribute related to an entity [28]. In the past, ABSA has been extensively utilized in various domains, such as products [11, 27], restaurants [5, 37], hotels [1, 17], and social media data [2, 29], but very few in the medical domain, with successful outcomes.

Patient-reported data obtained from medical social media and healthcare web forums provide significant insights that highlight various aspects of drugs, including their side effects, benefits, effectiveness, recommended dosage, impact on specific medical conditions, costs, and overall user experience. These insights can be invaluable to healthcare providers and pharmaceutical companies for understanding patient experiences with different medications, improving drug development, and enhancing patient care.

Figure 1 presents examples of drug reviews shared on pharmaceutical websites, showcasing the aspects discussed within them, and their corresponding sentiment classification. In example 1, "***Very effective***, *but side effects far outweigh the benefits.* ***Severe gastric distress*** *was the main one that got worse as time went along*." For this review, the sentiment polarity of the aspect "*very effective*" is positive, and "*severe gastric distress*" is negative. In example 2, "*This medication* ***works very well****, but unfortunately for me it is* ***expensive****"*. For this review, the aspect "*works really well*" is positive, and the aspect "*expensive"* is negative. In example 3, "*I have been on 12.5 mg of Paxil for 6 weeks. My doctor said that was the dose for treating anxiety. I have had some* ***mild side effects****. I will say the very next day after taking it my* ***anxiety is gone***." For this review, the aspect "*mild side effects*" is neutral and the aspect "*anxiety is gone*" is positive. In example 4, "*Accutane has* ***severe side effects*** *such as flaky skin, dry lips, dry eyes, and joints pain but my* ***acne*** *has improved a lot*". For this review, the sentiment polarity of the aspect "*severe side effects*" is negative, and the aspect "*acne*" is positive.

Previously, ABSA primarily relied on rule-based, machine learning (ML), and deep learning (DL) methods. Rule-based ABSA employs predefined rules and heuristics based on specific patterns and linguistic features to identify aspects and sentiments in the text [4, 30]. Sentiment lexicons or syntactic and semantic patterns can be used to assign sentiment scores. Although interpretable, designing rules for all possible patterns can be time-consuming and may not capture the complexities of language.

The ML-based ABSA automatically learns features and patterns [6, 16, 43] but requires human interventions. It relies on a huge number of annotated training examples to learn how to predict sentiment labels for new unseen texts. Another limitation of the ML models for ABSA is the lack of domain and contextual information, which can make it difficult to extract domain-specific aspects from the text.

The advent of DL has led to the development of models that can automatically learn features and train on them, without the need for handcrafted features and manual rule design. DL methods, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), have been extensively applied in ABSA in various domains such as products, restaurants, and social media data, and acquired excellent results. However, conventional DL methods have limitations in modeling entire sentences and are not able to detect important entities for fine-grained SA. Models such as Bidirectional LSTM and GRU can learn complex relationships in the input and can capture the context and semantics of the text.

Researchers have recently started utilizing DL models for ABSA in the medical social media domain and have achieved promising results [13, 42]. While ABSA has shown promising results in analyzing medical social media data, there are still some limitations to consider: (1) ABSA requires a huge amount of annotated data to train the models

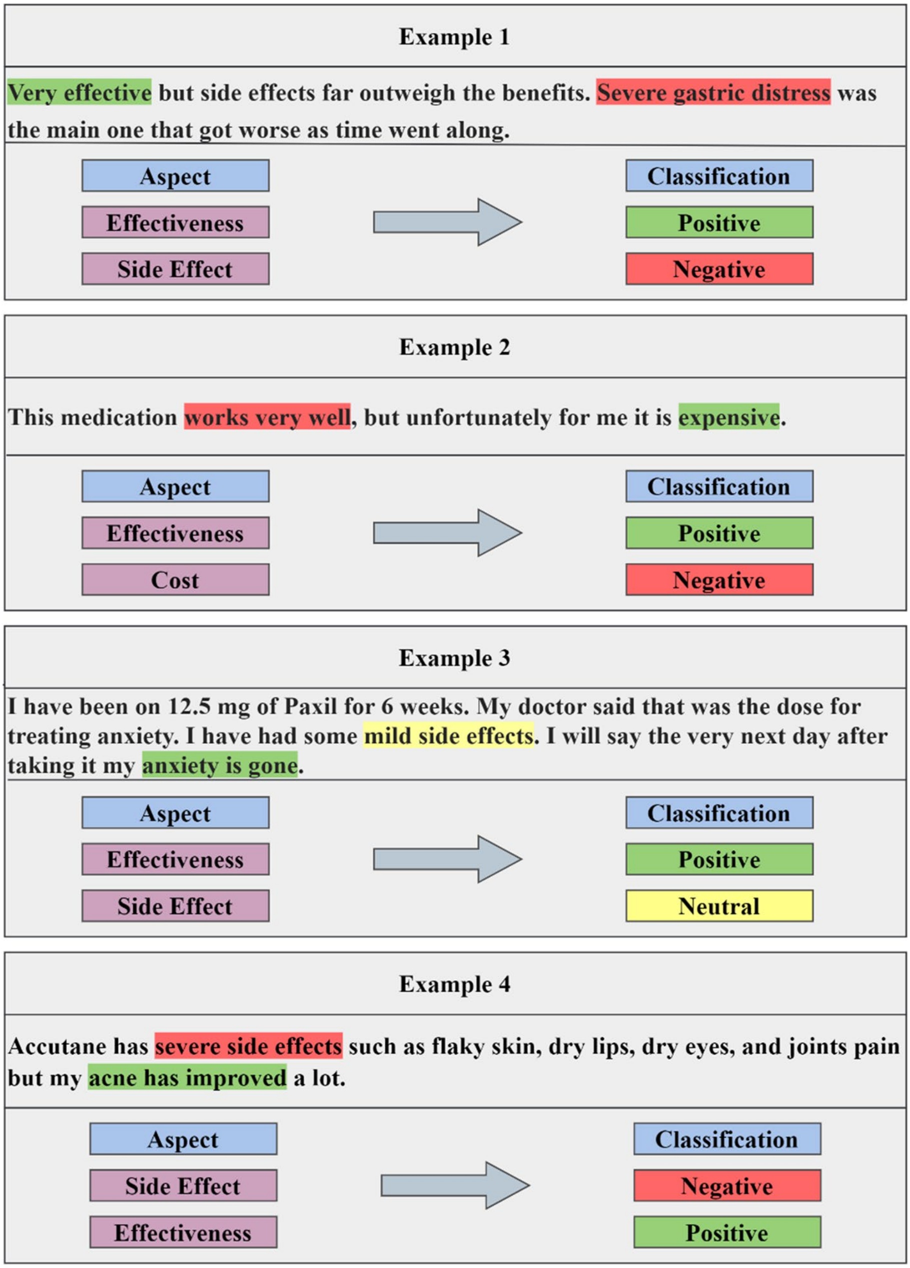

Fig. 1 Examples of drug reviews with aspects mentions and their sentiment corresponding classifications

effectively. However, obtaining annotated data in the medical domain can be challenging because of privacy concerns and ethical considerations. This limits the availability of high-quality training data, which can affect the accuracy and effectiveness of the models.; (2) Aspects can be specified explicitly or implicitly in a sentence. Explicit aspect terms refer to aspects directly stated in a sentence. For example, a sentence stating "*This medicine has **side effects** such as weight gain and hormone imbalance*" explicitly specifies the aspect of the drug that is a side effect. Implicit aspect terms are those that are not directly mentioned in the sentence but can be inferred from the context. For instance, the sentence "*After using this medicine for two months continuously, I can definitely feel the **change***" implies a benefit aspect, which is described through the word "change," but not explicitly stated in the review, as in the previous example. Identifying implicit aspects can be challenging but is necessary for accurate ABSA in the medical domain.; (3) ABSA is domain-dependent, which means that the interpretation and context of a word can vary across different domains [8]. Therefore, it is essential to consider the inclusion of external domain-specific knowledge [42]. This external knowledge can be in the form of domain-specific dictionaries, ontologies, and language models that provide additional context and meaning to words and phrases used in the domain. Incorporating such external knowledge can help ABSA models better understand the domain-specific language and accurately identify the aspects and sentiments expressed in the text.

Therefore, we propose a novel model, the Multi-task Learning based Dual Bidirectional LSTM Model (MLDBM), for the ABSA of drug reviews. The MLDBM utilizes a dual BiLSTM architecture that learns the contextualized embeddings of the input sequence through the BERT. In addition, the model employs a Multi-Head Self Attention (MHSA) method to extract important aspects and generate aspect-specific representations. The model utilizes dual bidirectional LSTM layers for processing both BERT and MHSA representations. It then applies attention mechanisms to further refine the processed representation for sentiment classification purposes. This approach helps the model to better capture the aspects depending on the context, enabling it to achieve improved performance in ABSA. The main contributions in this paper are summarized as follows:

- Proposed a novel method, Multi-task Learning based Dual Bidirectional LSTM Model (MLDBM), for ABSA of pharmaceutical drug reviews.
- Leveraged BERT for input representations and utilized MHSA to generate aspect-specific representations. Dual BiLSTM layers are employed to process these representations, followed by an attention layer to enhance ABSA.
- Evaluated the generalizability of the proposed model and its three variations on two widely recognized benchmark datasets from drugs.com and druglib.com.
- The experimental results demonstrate the effectiveness of MLDBM compared to several baseline models, indicating its robustness and efficacy.

The remaining paper is organized as follows. In section 2, the work related to the current state-of-the-art is discussed. Section 3 presents the proposed methodology in a phase-wise manner. Section 4 discusses the experimental setup. Results and evaluation of the proposed model is discussed in section 5, along with a case study. Finally, the conclusion and future work are discussed in section 6.

2 Related work

SA has emerged as a widely utilized NLP task aimed at finding individuals' sentiments towards various products [11], services [5], and social media data [7, 29]. Notably, there has been a growing trend in employing ML, DL, and their fusion models to analyze textual data for SA, yielding remarkable levels of performance. These approaches have been substantiated by empirical evidence, demonstrating their effectiveness in accurately determining and interpreting sentiments expressed by users. For example, Basiri et al. [3] introduced a fusion model that combines DL and ML techniques to classify drug reviews. They proposed deep fusion models: 3W1DT and 3W3DT for analyzing drug reviews. Both models showed a 4% increase in accuracy and the F1-measure compared to traditional models. The 3W3DT method performed even better than the 3W1DT method by 2%. Similarly, Dubey et al. [10] proposed a method for drug review classification based on user opinions by generating aspect-based lexicons and utilized Bi-DLSTM and Bi-GRU to capture long dependencies and process sequences in both directions. Additionally, Jiménez-Zafra et al. [18] identified how individuals shared their viewpoints in medical forums. The author utilized supervised and lexicon-based techniques to analyze reviews of drugs and physicians. Furthermore, Yadav et al. [40] introduced a classification framework that used deep convolutional neural networks for analyzing sentiment in a medical context, which focused on specific medical factors, such as medications and medical conditions instead of just positive or negative emotions. Karsi et al. [19] utilized ELMo to capture sentiment information by generating distinct vectors for words with contrasting polarities. Table 1 presents a brief review of few recent works focusing on sentiment analysis in the medical domain.

2.1 Aspect-based sentiment analysis in general domain

ABSA is an approach for performing detailed SA by identifying the mentioned aspect in a sentence and determining the sentiment category associated with each aspect. However, the presence of multiple aspects in a sentence creates a challenge for models to accurately extract and assess each aspect's individual and overall polarity. Previous research in this area has relied on manually crafted features to train the algorithm. However, DL techniques have addressed this limitation by automatically detecting features, leading to promising results without the need for manual feature design.

Several studies have explored ABSA in the general domain, particularly focusing on consumer reviews. For example, Geetha and Renuka [11] conducted SA on consumer review data by employing various ML-based classification models. Miao et al. [27] developed an enhanced BiLSTM-CRF model that combined Chinese character vectors and word position features. This model allowed for the joint extraction of attribute and sentiment words, while simultaneously determining the polarity judgments of sentiment words.

In addition, some researchers have utilized external knowledge to improve semantic representation. Ma et al. [25] developed a hybrid network called Sentic LSTM for TABSA to capture aspect-specific information and sentiment polarities. Ke et al. [20] developed a framework for SA, which offered domain knowledge to develop additional labeled data for each new labeling decision. Through a rule-based semantic parser, these explanations were transformed into programmatic labeling functions that generated noisy labels for an arbitrary amount of unlabeled sentiment information, enabling the training of a SA classifier.

Recently, language models have gained considerable attention in ABSA due to their emergence as powerful tools in the field of NLP. For example, Xu et al. [38] utilized BERT in their study for question answering and ABSA in product reviews. The authors fine-tuned BERT on a large-scale review dataset and then introduced a task-specific post-training stage for model enhancement. Sun et al. [32] performed ABSA and

Table 1 Overview of research works related to sentiment analysis in the medical domain

Author	Year	Objective	Dataset	Performance
Basiri et al. [3]	2020	To develop a fusion model that combines DL and ML techniques for the classification of drug reviews.	Drugs.com	*Accuracy: 88.36%* *Precision: 88.68%* *Recall: 88.36%* *F1-score: 87.35%*
Dubey et al. [10]	2022	To propose a model to classify drug reviews by incorporating a modular lexicon and a fusion strategy.	Drugs.com and Druglib.com	*Accuracy: 93.02%* *Recall: 88.72%* *F-measure: 92.64%*
Jiménez-Zafra et al. [18]	2019	To identify the way how people express opinions in medical forums.	DOS and COPOS	DOS corpus *Accuracy: 57.4–65.3%* *Precision: 58.3–65.2%* *Recall: 55.9–64.8%* *F1-score: 53.2–64.8%* COPOS corpus *Accuracy: 77.7–89.2%* *Precision: 64.2–85.1%* *Recall: 74.3–87.6%* *F1-score: 74.3–87.6%*
Yadav et al. [40]	2018	To find the sentiments from medical text from the condition the user has, symptoms, treatment, and drug prescribed.	Patient.info	Medical Condition: *Precision: 68%* *Recall: 60%* *F1-score: 63%* Medication: *Precision: 86%* *Recall: 77%* *F1-score: 82%*
Karsi et al. [19]	2021	To enhance sentiment classification by leveraging ELMo.	Drugs.com	*F1-score: 80.60%*

converted it into a sentence-pair classification task. The author introduced a technique for ABSA utilizing BERT, where auxiliary sentences are constructed to capture aspect-specific details. Li et al. [21] utilized an end-to-end approach by leveraging BERT. Their method enabled the simultaneous identification of aspects and their corresponding sentiment polarities.

The attention mechanism also has become widely popular across domains for its capability to capture meaningful patterns and relationships in data. Yang et al. [41] introduced an alternating co-attention network to find the interdependencies between aspects and their corresponding sentiment polarities. Liu et al. [23] used sentence-level and context-level attention mechanisms for aspect-based sentiment classification (ABSC). A brief overview of recent research works related to ABSA in the general domain is presented in Table 2.

2.2 Aspect-based sentiment analysis in medical domain

Researchers have recently begun utilizing ABSA techniques in the medical domain. The promising outcomes have highlighted the potential of these techniques in enhancing SA in healthcare settings. ABSA research in the medical field has predominantly concentrated on ML and DL models. These models have received significant attention due to their potential in analyzing medical-related data. Gräßer et al. [12] analyzed drug reviews for ABSA and assessed the adaptability of their model across different disorders and data sources. The authors created two benchmark datasets, namely drugs.com, and druglib.com. Colón-Ruiz and Segura-Bedmar [9] performed a comparison of different DL models and developed multiple fusions of DL models for ABSA. Additionally, they assessed the influence of pre-trained word embeddings on the models' performance.

The attention mechanism has gained widespread popularity among authors in different domains due to its ability to capture informative patterns and relationships within data. This has led to its extensive utilization as a fundamental component in various research works, enabling improved performance and insights across a wide range of applications. Han et al. [13] introduced a Double BiGRU model enhanced with knowledge transfer to categorize drug reviews into three sentiment categories. The classification was conducted by considering the aspects that were mentioned within the reviews. Similarly, Setiawan et al. [31] developed a BiGRU model with character-enhanced token embeddings and used a multilevel attention method for targets to find the interactions between aspects and their corresponding sentiment polarities.

Recently, some studies have been done by utilizing language models for ABSA in the healthcare domain, as these models have shown their effectiveness in extracting sentiment-related insights from healthcare-related data. Sweidan et al. [33] introduced a method for classifying the sentiment of sentence-level aspects using a combination of ontology and the XLNet model. The author explored the impact of incorporating a lexicalized ontology to enhance the performance of ABSA, specifically by capturing indirect relationships within user social data. Žunić et al. [44] developed an approach for ABSA of drug reviews. They analyze signs and symptoms, which were automatically extracted using the Unified Medical Langauge System (UMLS). The extracted details are then fed into the BERT, which undergoes two additional layers for fine-tuning specifically for final sentiment classification. Table 3 presents a brief overview of the studies on ABSA applied to drug reviews.

Table 2 Overview of research works related to aspect-based sentiment analysis in the general domain

Author	Year	Objective	Dataset	Performance
Geetha and Renuka [11]	2021	To utilize fine-tuned Bert Base Uncased model for improving ABSA of product reviews.	Amazon Reviews	*Accuracy: 88.48%* *Precision: 88.09%* *Recall: 86.22%* *F1-score: 89.84%*
Miao et al. [27]	2021	To improve ABSA using BiLSTM-CRF combining Chinese character and word position features.	Mobile Reviews	*Precision: 89.20%* *Recall: 88.29%* *F1-score: 88.74%*
Ma et al. [25]	2018	To use common-sense knowledge in the DL neural sequential model for TABSA.	SentiHood and SemEval 2015	Sentihood: Aspect Categorization: *Accuracy: 69.38%* *Macro-F1: 80%* *Micro-F1: 79.79%* Sentiment Categorization: *Accuracy: 88.80%* Semeval 2015: Aspect Categorization: *Accuracy: 69.19%* *Macro-F1: 77.55%* *Micro-F1: 75.00%* Sentiment Categorization: *Accuracy: 76.47%*
Ke et al. [20]	2021	To present a SA framework using natural language explanations to generate additional labeled data for each new labeling decision.	SemEval 2014	Restaurant: *Precision: 69.1%* *Recall: 81.5%* *F1-score: 74.7%* Laptop: *Precision: 54.7%* *Recall: 72.6%* *F1-score: 62.3%*

Table 2 (continued)

Author	Year	Objective	Dataset	Performance
Xu et al. [38]	2019	To utilize the BERT for question answering and ABSA on product reviews.	E-commerce Reviews	Aspect Extraction: F1 -score: *Laptop: 84.26%* *Rest: 77.97%* Aspect Sentiment Category: *Laptop: Accuracy: 78.08%* *Macro-F1: 75.08% Rest:* *Accuracy: 84.95%)* *Macro-F1: 76.96%*
Sun et al. [32]	2019	To perform ABSA and convert it into sentence-pair classification tasks such as question answering and NLI, by constructing auxiliary sentences from the aspect terms.	SentiHood and SemEval-2014 Task 4	ABSA: Aspect Extraction: *Accuracy: 79.8%* Sentiment classification: *Accuracy: 93.6%* ABSA: *Precision: 93.57%* *Recall: 90.83%* *F1-score: 92.18%*
Li et al. [21]	2019	To utilize BERT embeddings for end-to-end ABSA.	SemEval-2014	F1-score *Laptop: 61.12%* *Rest: 74.72%*
Yang et al. [41]	2019	To develop an alternating co-attention network for ABSA.	SemEval 2014 and Twitter	Accuracy: *Rest: 79.7%* *Laptop: 73.5%* *Twitter: 71.5%*
Liu et al. [23]	2018	To enhance the ABSC using sentence-level and context-level attention mechanisms by embedding them into the model.	SemEval 2014 and Twitter	Accuracy: *Rest: 80.89%* *Laptop:75.07%* *Twitter: 71.53%*

Table 2 (continued)

Author	Year	Objective	Dataset	Performance
Bensoltane and Zaki [5]	2022	To leverage BERT for ABSA on the Arabic dataset.	News posts and Facebook comments	Aspect Term Extraction: *Precision: 87.7%* *Recall: 88.5%* *F1-Score: 88.1%* Aspect Category Detection: *Precision: 84.1%* *Recall: 84.1%* *F1-Score: 84.1%*
Xue and Li [39]	2018	To develop a CNN-based model with gating mechanisms.	SemEval-2014	Aspect-Term Accuracy *Rest: 77.38%* *Laptop: 69.14%* Aspect-Categorization Accuracy: Rest.: 79.35–85.92%

Table 3 Overview of research works related to aspect-based sentiment analysis in the medical domain

Author	Year	Objective	Dataset	Performance
Gräßer et al. [12]	2018	To perform ABSA on cross-domain drug reviews.	Drugs.com and Druglib.com	In-Domain Drugs.com: *Accuracy: 92.24%* Druglib.com: Accuracy: *Overall Rating (all): 75.19%* *Benefits: 77.70%* *Side effects: 76.93%* Cross-Data: *Overall Rating (all): 70.06%* *Side effects: 49.75%*
Han et al. [13]	2020	To perform ABSA on drug reviews using a double BiGRU and knowledge transfer.	Druglib.com	*Accuracy: 78.26%* *Macro-F1: 77.75%*
Setiawan et al. [31]	2020	To propose a BiGRU model based on character-enhanced token embeddings and use multilevel attention for specific targets.	Google Reviews, Careopinion and Semeval 2015 task 12	Macro-F1-score: *Hospital: 86.137%* *Sentihood: 82.356%* *Indonesian Rest.: 76.260%* Micro-F1-Score: *Hospital: 84.105%* *Sentihood: 80.101%* *Indonesian Rest.: 76.130%*
Sweidan et al. [33]	2021	To utilize lexicalized ontology to extract indirect relationships in social data.	Askapatient, WebMD, DrugBank, n2c2 2018, TAC 2017, Twitter	*Precision: 94.2% to 98.2%* *Recall: 93.5% to 98.2%* *F-measure: 94.1% to 98%*
Žunić et al. [44]	2022	To perform ABSA by extracting signs and symptoms through UMLS and utilizing fine-tuned BERT to classify the aspect.	Drugs.com	*Accuracy: 94.74%*

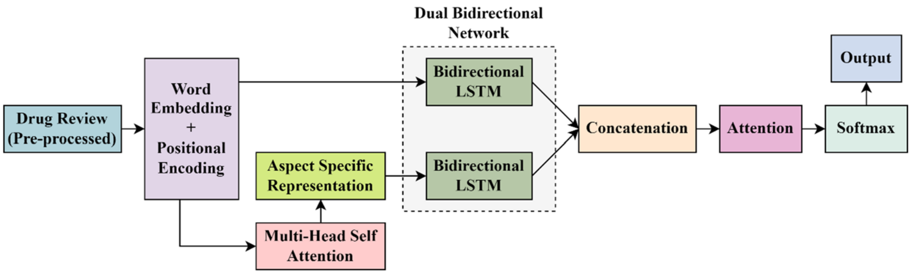

Fig. 2 Block representation of Multi-task Learning based Dual Bidirectional LSTM Model

3 Proposed methodology

This section discusses the proposed model, Multi-task Learning based Dual Bidirectional LSTM Model (MLDBM) for ABSA of drug reviews. A block representation of the proposed model to provide a visual overview of how the input tokens are processed by each component and how the model generates the final predictions is shown in Fig. 2.

The proposed methodology consists of three phases: preprocessing, feature extraction, and sentiment classification. In the first phase, drug reviews are processed to eliminate irrelevant information. In the second phase, feature extraction, contextual embeddings are initialized using BERT, and aspect-specific representations are generated by incorporating MHSA to emphasize different parts of the input sequence, focusing on important aspects.

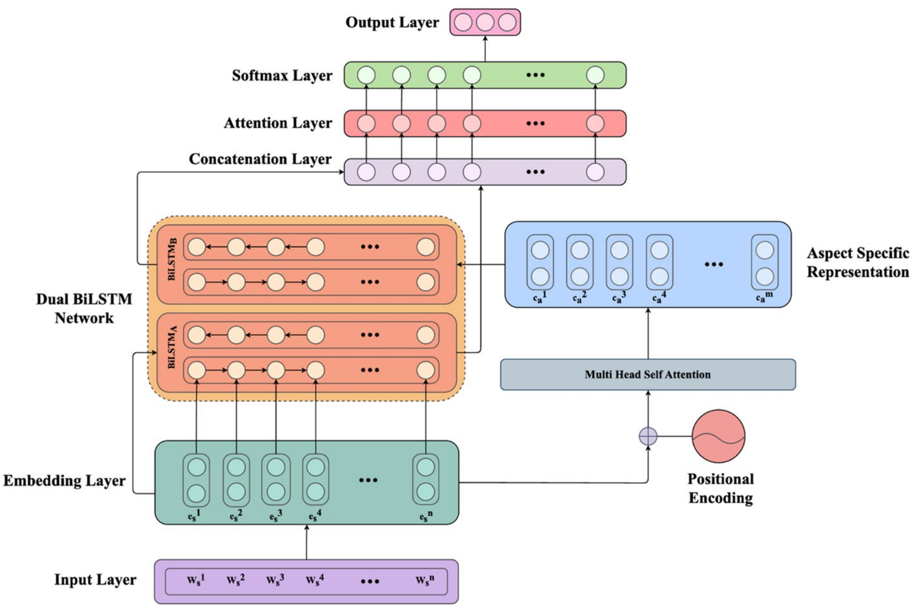

Fig. 3 Architecture of the proposed Multi-task Learning based Dual Bidirectional LSTM Model

In the third phase, sentiment classification, an attention-based dual bidirectional LSTM network that consists of two layers of bidirectional LSTM is proposed where one layer takes the input sequence embeddings generated by BERT, while the other layer takes the aspect-specific representations generated by the MHSA layer. The outputs of the two layers are concatenated at the concatenation layer, and an additional attention layer is applied on top of the concatenation layer. This attention layer extracts crucial sentiment information by leveraging the captured aspect-specific representations. Finally, a SoftMax layer is used to make predictions and categorize the reviews into positive, neutral, or negative classes. The classification is based on the final representations obtained from the attention layer.

The proposed architecture of MLDBM is depicted in Fig. 3. Algorithm 1 provides a step-by-step outline of the procedure for feature extraction and aspect-based sentiment classification.

Algorithm 1 Procedure for feature extraction and aspect-based sentiment classification

Input: Sentence $S = [w_s^1, w_s^2, w_s^3, ...w_s^n]$, h_f: $BiLSTM_A$, h_s: $BiLSTM_B$, MHSA: Multi-head self-attention

Output: Sentiment classification based on aspects.

1. WordPiece tokenizer(S)

2. Add special tokens [CLS] at the start and [SEP] at the end of S.

3. Procedure MLDBM(S)

4. BERT Embedding Layer:

a. $E_s \leftarrow$ *BuildEmbeddings(S)*

b. Compute positional encoding using Eqs. (1) and (2).

c. Compute MHSA(E) using Eqs. (3) and (4).

d. Obtain aspect-specific representation using eq. (5).

5. $BiLSTM_A$ *Layer:*

a. Apply $BiLSTM_B$ *on* E_s *to obtain* $\overrightarrow{h_s}$ *and* $\overleftarrow{h_s}$.

b. Construct h_f *using Eqs. (11) and (12).*

6. $BiLSTM_B$ *Layer:*

a. Apply $BiLSTM_B$ *on* C_a *to obtain* $\overrightarrow{h_s}$ *and* $\overleftarrow{h_s}$.

b. Construct h_s *using Eqs. (14) and (15).*

7. Concatenation Layer:

a. $h_{DBiLSTM} \leftarrow$ *Concatenate(* h_f*,*h_s*)*

8. Compute attention:

a. $z_s = attention(\alpha_i, h_{DBiLSTM})$

9. $Output = softmax(W_s z_{DBiLSTM} + b_s)$

10. $loss = compute\_loss(L^{fe} + L^{sc})$

11. c = argmax(output)

12. end procedure

13. end

3.1 Preprocessing

The first phase in the proposed methodology involves preprocessing. The first task in this stage is sentence tokenization, where each sentence is divided into individual words using the BertTokenizer. The subsequent step involves the elimination of stop words, such as “a”, “an”, and “the,” which do not provide any relevant information for the classification task. Additionally, each token is converted to lowercase and reduced to its stem using a rule-based method. Further, NLTK lemmatization is applied to reduce each token to its base form. The sentences are then examined for repeating patterns, which are replaced with white spaces. Lastly, part-of-speech tagging is employed to label the words based on their function in the sentence, such as verbs, nouns, adjectives, adverbs, etc., which is dependent on the sentence’s context and definition.

3.2 Feature Extraction

The second phase of the proposed model is feature extraction, which involves processing each token of the input text and extracting the aspects from the reviews. This is accomplished by utilizing a pre-trained learning model, BERT, to generate contextualized embeddings for the words in the input sequence. The hidden representation of BERT encodes the semantic information. In addition, the MHSA mechanism is trained to encode the significant contextual words required for identifying the aspect, which is typically needed by ABSA. The resulting embeddings from this process are subsequently used as input to the first bidirectional LSTM layer, $BiLSTM_A$, while the aspect-specific representations are input to the second bidirectional LSTM layer, $BiLSTM_B$.

Given a sentence S, represented as $S = \left[w_s^1, w_s^2, w_s^3, \ldots w_s^n\right]$ and the aspect $a = \left[w_a^1, w_a^2, w_a^3, \ldots w_a^m\right]$ where w denotes word, and a denotes aspect term. To generate an embedding vector matrix, represented as $\left[e_s^1, e_s^2, e_s^3, \ldots, e_s^n\right] \in \mathbb{R}^{n\times \mathrm{d}}$, the BERT generates embeddings for each word in the input sequence and maps them to the d-dimensional vector that the model learns during training. This process results in obtaining a matrix representation $E(w)$ of dimensions $n \times d$. To determine the word order, the model then incorporates positional encoding using eqs. (1) and (2), where pos refer to the position of a specific token in the input sequence, i denotes the dimension of each embedding, and d denotes the dimensions of word embedding.

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10,000^{2i/d_{model}}}\right) \tag{1}$$

$$\mathrm{PE(pos, 2i+1)} = \cos\left(\frac{\mathrm{pos}}{10,000^{2\mathrm{i}/\mathrm{d}_{\mathrm{model}}}}\right) \tag{2}$$

The MHSA is then utilized to compute multiple-scale dot-product attention by using eq. (3) where $E(w)$ is the embedding matrix for a single sentence with n words, $W^Q \in \mathbb{R}^{d_h \times d_q}$

is the query matrix, $W^k \in \mathbb{R}^{d_h \times d_k}$ is the key matrix and $W^V \in \mathbb{R}^{d_h \times d_v}$ is the value matrix. These three weight matrices are then used to calculate the multiple-scale dot-product attention in parallel and concatenated to output features. The features are then transformed by multiplying W^{MH} weight matrix as shown in eq. (4). h is the attention head. The "; " is the concatenation of each head.

$$Softmax\left(\frac{\left(X \cdot W^Q\right) \cdot \left(X \cdot W^K\right)^T}{\sqrt{d_k}}\right) \cdot \left(E \cdot W^V\right) \tag{3}$$

$$MHSA(E) = \tan h\left(\left\{H_1;H_2; \ldots ;H_h\right\}.W^{MH}\right) \tag{4}$$

$$\gamma_{term} = \frac{\exp\left(C_i\right)}{\sum_{k=1}^{N} \exp\left(C_i\right)} \tag{5}$$

To obtain the final context representation of the extracted aspect terms, eqs. (3), (4), and (5) are used, which involve a weighted summation of the context representations. Here, the context feature of the token T is denoted by C_i, where i corresponds to the position of the token. The contextual information from each token is combined to create an overall representation of the aspect terms matric as $\left[c_a^1, c_a^2, c_a^3, \ldots, c_a^m\right] \in \mathbb{R}^{m \times \mathrm{d}}$.

3.3 Dual BiLSTM network

LSTM models are primarily designed to capture sequential dependencies between inputs, and they are not explicitly trained to identify and extract specific aspects or features from a text. As a result, they are not able to adequately detect the important words or phrases that are relevant to ABSA. To overcome this limitation, additional techniques or models are often employed, such as incorporating attention mechanisms or using neural networks specifically designed for aspect extraction. A typical LSTM cell structure is shown in Fig. 4. The mathematical formulations of LSTM are as follows:

$$i_t = \sigma\left(w_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{6}$$

$$f_t = \sigma\left(w_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{7}$$

$$o_t = \sigma\left(w_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{8}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh\left(w_c \cdot \left[h_{t-1}, x_t\right] + b_c\right) \tag{9}$$

$$h_t = o_t \odot \tan h\left(c_t\right) \tag{10}$$

where x_i is the input to the LSTM cell at time t. LSTM generates a sequence of hidden outputs, denoted as $H = \{h_1, h_2, h_3, \ldots, h_n\}$, where each h_t represents the hidden state vector. The input gate is denoted as i_t, representing the information that is considered for inclusion

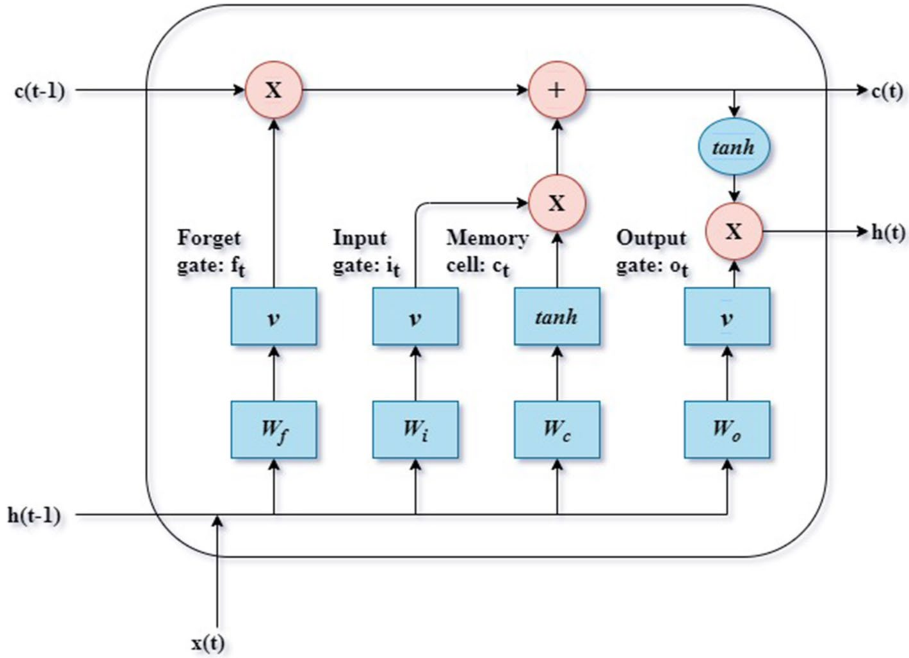

Fig. 4 Basic structure of LSTM cell

in the cell state. The forget gate is denoted as f_t, is responsible for determining which information from the previous cell state should be discarded. The output from the LSTM is represented by o_t, indicating the selected information is passed on to the next step. The cell state is denoted as c_t, which serves as a memory unit that carries information across time steps. The sigmoid function σ, hyperbolic tangent function *tanh*, and product operation ⊙ are utilized in the computations. The weight matrices w_i, w_f, w_o, $w_c \in \mathbb{R}^{d\times 2d}$, along with the bias vectors b_i, b_f, b_o, $b_c \in \mathbb{R}^{d}$, are employed during the training process.

Building upon the work of Wang et al. [35], we propose a new dual BiLSTM model constructed with two independent BiLSTM layers, which are designed to capture more contextual information and enhance the model's overall capability. A basic bidirectional LSTM cell structure is shown in Fig. 5.

The input sequence embeddings and aspect-specific representations generated by the BERT and MHSA are fed into $BiLSTM_A$ and $BiLSTM_B$, respectively. The forward hidden representations, $\overrightarrow{h_f}$ and backward hidden representations, $\overleftarrow{h_f}$ for the $BiLSTM_A$ layer are computed using Eqs. (11) and (12), respectively. The final output hidden representation, h_f, is achieved by concatenating $\overrightarrow{h_f}$ and $\overleftarrow{h_f}$, as per eq. (13).

$$\overrightarrow{h_f} = \left[\overrightarrow{h_f^1}, \overrightarrow{h_f^2}, \overrightarrow{h_f^3}, \ldots \overrightarrow{h_f^n}\right] = \overrightarrow{LSTM}\left(\left[e_s^1, e_s^2, e_s^3, \ldots, e_s^n\right]\right) \tag{11}$$

$$\overleftarrow{h_f} = \left[\overleftarrow{h_f^1}, \overleftarrow{h_f^2}, \overleftarrow{h_f^3}, \ldots \overleftarrow{h_f^n}\right] = \overleftarrow{LSTM}\left(\left[e_s^1, e_s^2, e_s^3, \ldots, e_s^n\right]\right) \tag{12}$$

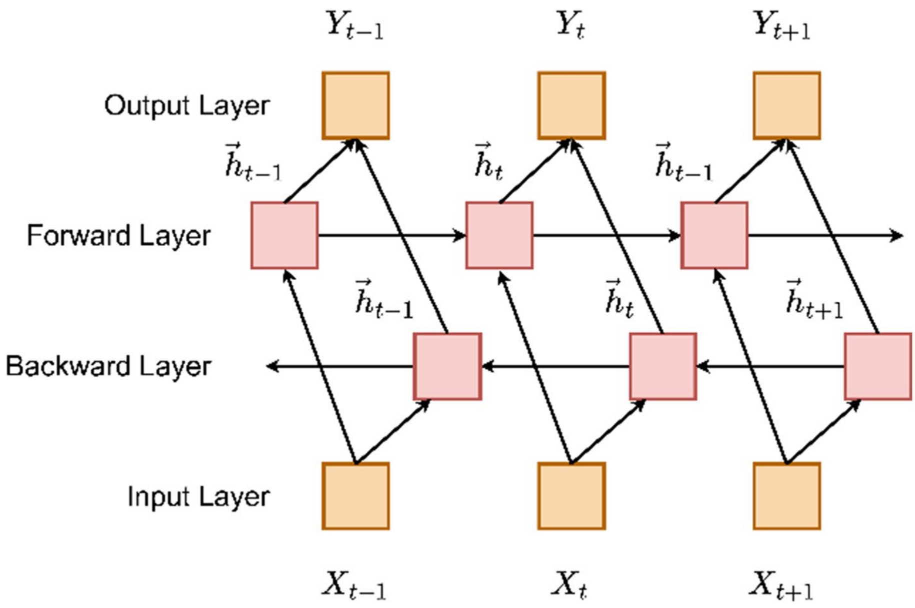

Fig. 5 Basic structure of Bidirectional LSTM cell

$$h_f = \left[h_f^1, \ldots, h_f^n\right] = \left[\overrightarrow{h_f^1} + \overleftarrow{h_f^1} \ldots \overrightarrow{h_f^n} + \overleftarrow{h_f^n}\right] \tag{13}$$

Similarly, the forward hidden representations, $\overrightarrow{h_s}$ and backward hidden representations, $\overleftarrow{h_s}$ for the $BiLSTM_B$ layer are computed using eqs. (14) and (15). The final output hidden representation, h_s, for the $BiLSTM_B$ is obtained by concatenating $\overrightarrow{h_s}$ and $\overleftarrow{h_s}$, as given in eq. (16).

$$\overrightarrow{h_s} = \left[\overrightarrow{h_s^1},\overrightarrow{h_s^2},\overrightarrow{h_s^3}, \ldots \overrightarrow{h_s^m}\right] = \overrightarrow{LSTM}\left(\left[c_a^1, c_a^2, c_a^3, \ldots, c_a^m\right]\right) \tag{14}$$

$$\overleftarrow{h_s} = \left[\overleftarrow{h_s^1},\overleftarrow{h_s^2},\overleftarrow{h_s^3}, \ldots \overleftarrow{h_s^m}\right] = \overleftarrow{LSTM}\Big(\left[c_a^1, c_a^2, c_a^3, \ldots, c_a^m\right] \tag{15}$$

$$h_s = \left[h_s^1, \ldots, h_s^m\right] = \left[\overrightarrow{h_s^1} + \overleftarrow{h_s^1} \ldots \overrightarrow{h_s^m} + \overleftarrow{h_s^m}\right] \tag{16}$$

3.4 Concatenation layer

The concatenation layer performs an element-wise summation of the representations from the $BiLSTM_A$ and $BiLSTM_B$. The resulting representations contain the combined information from the first and second BiLSTM layers. The concatenated output, $h_{DBiLSTM}$ is then used as input for subsequent layers.

$$h_{DBiLSTM} = \left[h_f + h_s\right] \tag{17}$$

3.5 Attention layer

The inclusion of an attention layer assists the model in comprehending the significant portions of the input that are crucial for capturing sentiment-related information. Specifically, the attention layer focuses on the aspect-specific representations that were captured by the previous layers. The attention layer assigns weight to each element, indicating how important that element is for capturing the overall sentiment of the sentence. Specifically, let the matrix of hidden vectors be denoted as $h_{DBiLSTM} \in \mathbb{R}^{d \times n}$, where d is the dimensionality of the hidden vectors and n is the length of the sentence. The attention score α_i is computed as follows:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} h_a^i \tag{18}$$

$$\kappa_i = f_{score}\left(h_{DBiLSTM}^i, \mu\right) = \tanh\left(h_{DBiLSTM}^{i^T} W_\alpha \mu\right) \tag{19}$$

$$\alpha_i = \frac{\exp\left(\kappa_i\right)}{\sum_{i=1}^{n} \exp\left(\kappa_i\right)} \tag{20}$$

where h_a^i is the hidden state vector of the aspect, μ denotes the average of aspect hidden states h_a, and f_{score} represents a content-based function utilized to capture the semantic correlation between the word and the aspect. $W_\alpha \in \mathbb{R}^{d \times d}$ represents a parameter matrix. The final representation of the aspect-based review sentences S is then given by:

$$z_{DBiLSTM} = \sum_{i=1}^{n} \alpha_i h_{DBiLSTM}^i \tag{21}$$

3.6 Softmax Layer

The final layer is a SoftMax layer, which is used to make predictions about the sentiment polarity of the input sentence. The SoftMax layer takes the weighted sum of the hidden vectors from the attention layer as input and produces a probability distribution. The final representation $z_{DBiLSTM}$ processed by SoftMax is computed as:

$$\hat{\gamma}_s = softmax\left(W_s z_{DBiLSTM} + b_s\right) \tag{22}$$

where the probability distribution $\hat{\gamma}_s \in \mathbb{R}^c$ and c denotes the sentiment classes viz., positive, neutral, and negative. $W_s \in \mathbb{R}^{c \times d_h}$ and $b_s = \mathbb{R}^c$ are the softmax parameter.

3.7 Model Training

In the proposed model, two loss functions are used to train the feature extraction and sentiment classification tasks: L^{fe} and L^{sc}. L^{fe} and L^{sc} are formulated as:

$$L^{fe} = \sum_1^N \sum_1^K \hat{t}_i \log t_i + \lambda \sum_{\theta \in \Theta} \theta^2 \tag{23}$$

$$L^{sc} = \sum_1^C \hat{y}_i \log y_i + \lambda \sum_{\theta \in \Theta} \theta^2 \tag{24}$$

where N represents the count of training examples, K signifies the dimensionality of the output representation, $y_{i,j}$ is a binary label indicating the presence of the j^{th} feature in the i^{th} sentence, $t_{i,j}$ represents the predicted output value of the j^{th} feature for the i^{th} sentence as generated by the model, λ denotes the $L2$ regularization parameter, and Θ represents the collection of all trainable parameters within the model. The final loss function is formulated as:

$$L^{absa} = L^{fe} + L^{sc} \tag{25}$$

4 Experimental setup

This section provides an overview of the datasets utilized and the evaluation metrics employed in the study. Additionally, we assessed the performance of various baseline models and conducted a comprehensive analysis of the proposed model and its variants. Furthermore, a detailed investigation was conducted, including a case study, to evaluate the performance of the proposed model.

4.1 Dataset

We have used two benchmark datasets drugs.com and druglib.com, which are commonly utilized in the literature [3, 12, 13, 44]. Drugs.com and druglib.com are pharmaceutical websites that offer information on medicines to healthcare professionals and consumers. The dataset from drugs.com contains 215,063 drug reviews, whereas the druglib.com dataset consists of 4143 drug reviews. The reviews shared by the patients drugs.com have a 10-star rating that denotes overall user satisfaction. For example, "*Tylenol does nothing for my sciatica. The pain is there all the time.*" In this review, a patient with sciatica pain is describing their experience with the drug Tylenol. The reviews on druglib.com has a 5-star rating for side effects reflecting the range from no side effects to severe side effects, and a 5-star rating for effectiveness that ranges from ineffective to highly effective.

4.2 Evaluation Metrics

For evaluating the proposed model's performance and effectiveness, this study employed Accuracy and *Macro-F1* as the evaluation metrics. Accuracy measures the proportion of accurately predicted reviews out of the total number of reviews.

$$Accuracy = \frac{T}{N} \tag{26}$$

where T denotes correctly predicted samples and N denotes the total number of samples.

As the proposed framework performs multiclass classification, we have used *Macro-F1* as an additional evaluation metric. *Macro-F1* is the harmonic mean of macro precision and macro recall and provides a comprehensive evaluation of multi-classification problems. *Macro-F1* is calculated as

$$Macro - F1 = \sum_{k \in C} \frac{F_{1,k}}{|C|} \tag{27}$$

where $F_{1,\,k}$ represents F_1 of the kth category, and $|C|$ represents the number of classes.

4.3 Training Details

All models are evaluated on a single NVIDIA RTX2060 GPU on a Linux machine. The implementation of the model was done using the Keras framework. The pre-training of the BERT model involved 12 layers of transformer encoders, 12 attention heads, and a hidden dimension of 768. During the training process, the model utilized the ADAM optimizer with a learning rate set to 0.001. A batch size of 40 was employed for each iteration, and the model was trained for a total of 25 epochs. To regularize the model, dropout was utilized with a dropout rate of 0.5. Furthermore, we conducted an analysis of how different values of various parameters, such as learning and dropout rates, impact the performance of both the model and its variant in section 5.3.

5 Results and evaluation

The proposed model is compared with several traditional and state-of-the-art models and evaluated for its effectiveness.

5.1 Baselines

The baselines methods used in the study include **LSTM**, **GRU**, **BiLSTM**, **BiGRU**, **Dual LSTM**, **Dual GRU**, Interactive Attention Network (**IAN**) [24], Attention Over Attention (**AOA**) [15], Pretraining+Multi-task Learning (**PRET + MULT**) [14], **Memnet** [34], Pretraining and Multi-task learning model based on Double BiGRU (**PM-DBiGRU**) [13], and Multi-Head Self-Attention Transformation (**MSAT-BERT**) [22]. **LSTM** and **GRU** are unidirectional models that model sentences and generate the final representation as the average of all the hidden states. **BiLSTM** is a bidirectional model that generates representations in both the forward and backward directions. Similarly, **BiGRU** uses GRU units to capture contextual information from both directions. **Dual LSTM** is similar to LSTM but uses two individual LSTM networks where the first LSTM network takes the input representations, and the second LSTM network takes the aspect representations. The output representations are concatenated and processed by the attention layer for the final prediction using Softmax. Similarly, **Dual GRU** is similar to GRU but uses two individual GRU networks to model the input and aspect representations which are then concatenated and processed by the attention layer for prediction. The **IAN** uses interactive attention to learn the coarse-grained attention of the sentence and target, which are then concatenated for ABSC**.** **AOA** uses two bidirectional LSTM networks to model sentences and targets and applies attention over attention mechanisms for sentiment classification. **PRET + MULT** enhances ABSC by transferring knowledge from document-level data. **Memnet** uses multiple-hop attention to important context words for ABSC. **PM-DBiGRU** incorporates domain knowledge from short-text-level drug reviews through pretraining and multitasks training with a double BiGRU model that includes an attention mechanism. **MSAT-BERT** utilizes BERT and BiLSTM to encode context words incorporating a target-specific MHSA for capturing global dependency features and integrating a target-sensitive Transformation.

5.2 Comparative analysis with state-of-the-art models

To assess the performance of the proposed model, we conducted evaluations on both datasets, comparing it against various traditional and state-of-the-art models. The results obtained for the drugs.com and druglib.com dataset are provided in Table 4.

Our findings indicate that LSTM exhibited poor performance in both datasets, achieving an accuracy of 75.12% and a Macro-F1 score of 74.01% on the drugs.com dataset, and an accuracy of 70.63% and a Macro-F1 score of 68.60% on the druglib.com dataset. This can be attributed to LSTM's limited ability to model word dependencies in both directions. On the other hand, GRU yielded similar results but with shorter training time. The reason for BiGRU achieving slightly higher accuracy compared to unidirectional GRU is due to its ability to process input in both forward and backward directions, enabling it to capture a more comprehensive context. LSTM and GRU models were outperformed by dual LSTM and dual GRU, which utilize separate networks for aspect representations, indicated improved performance. The inclusion of an additional layer of LSTM in dual LSTM and an extra layer of GRU in dual GRU offered an added level of abstraction, resulting in somewhat comparable outcomes between dual LSTM and dual GRU.

Table 4 Performance comparison of different models on the Drugs.com and Druglib.com dataset

Model	Drugs.com		Drugslib.com	
	Accuracy	Macro-F1	Accuracy	Macro-F1
LSTM	0.7512	0.7401	0.7063	0.6860
GRU	0.7537	0.7434	0.7318	0.7026
BiLSTM	0.7621	0.7591	0.7156	0.7031
BiGRU	0.7685	0.7374	0.7224	0.7068
Dual LSTM	0.7633	0.7515	0.6933	0.6650
Dual GRU	0.7657	0.7594	0.7079	0.6803
IAN	0.7879	0.7824	0.7311	0.7126
AOA	0.7881	0.7857	0.7348	0.7031
PRET+MULT	0.7893	0.7872	0.7472	0.7397
Memnet	0.7673	0.7512	0.7091	0.6765
PM-DBiGRU	0.7994	0.7863	0.7574	0.7318
MSAT-BERT	0.8543	0.7618	0.7781	0.7292
MLM	0.7712	0.7643	0.7312	0.7227
MLDM	0.7821	0.7809	0.7359	0.7199
MLBM	0.8136	0.807	0.7531	0.7398
MLDBM	0.8771	0.8597	0.7897	0.7623

The IAN demonstrated a higher level of accuracy in comparison to both traditional models and their dual counterparts. The reason behind the superior performance of the IAN on the drugs.com dataset can be attributed to its interactive learning approach and the distinct generation of contexts and targets. AOA outperformed IAN by employing two BiLSTMs to generate representations for target and input sentences, followed by the computation of a pair-wise interaction matrix to capture their interaction. Memnet, employing an LSTM, exhibited inferior performance compared to the other models due to its inability to generate semantic representations necessary for fine-grained ABSA. Although the PM-DBiGRU achieved superior results compared to other models, it faced limitations in detecting implicit expressions towards aspects and occasionally failed to identify explicit opinion words. Additionally, the accuracy of their approach was hindered by manual annotation and a limited amount of training data. MSAT performed best among all the baselines and achieved comparable performance than MLDBM. The utilization of the target-sensitive transformation component effectively captured sentiment expressions that are specifically influenced by the target.

MLDBM outperformed all the baseline approaches with outstanding results. On the drugs.com dataset, MLDBM achieved an impressive accuracy of 87.71% and a Macro-F1 score of 85.97%. Similarly, on the druglib.com dataset, MLDBM attained a notable accuracy of 78.97% and a Macro-F1 score of 76.23%. It is important to note that MLDBM's performance on the druglib.com dataset was slightly lower compared to the drugs.com dataset, which can be attributed to the limited availability of training data.

Figure 6 presents a comparison of the accuracies of different models on both datasets. Additionally, Fig. 7 shows the comparison of the Macro-F1 scores of the different models on both datasets. These figures highlight the superior performance of MLDBM compared to the other models. Furthermore, we explored variations of the MLDBM model, discussed in detail in the next section.

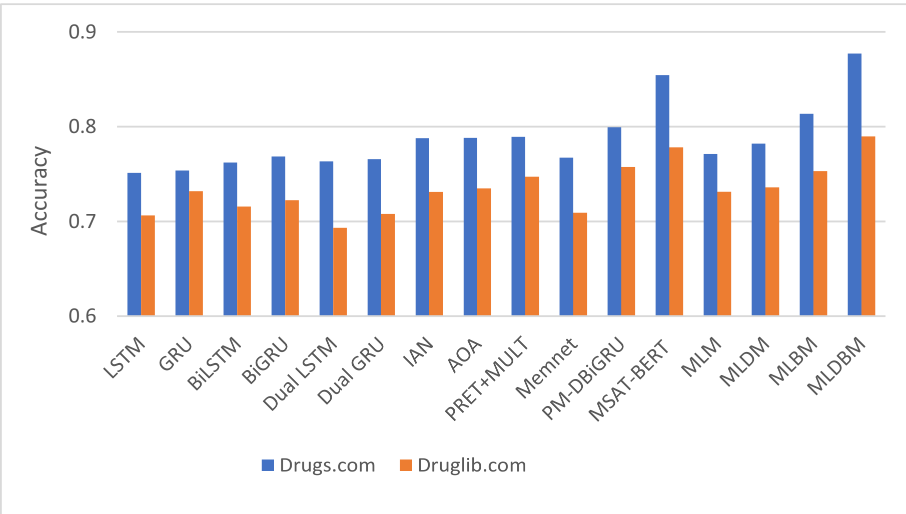

Fig. 6 Accuracy comparison of different models on the Drugs.com and Druglib.com datasets

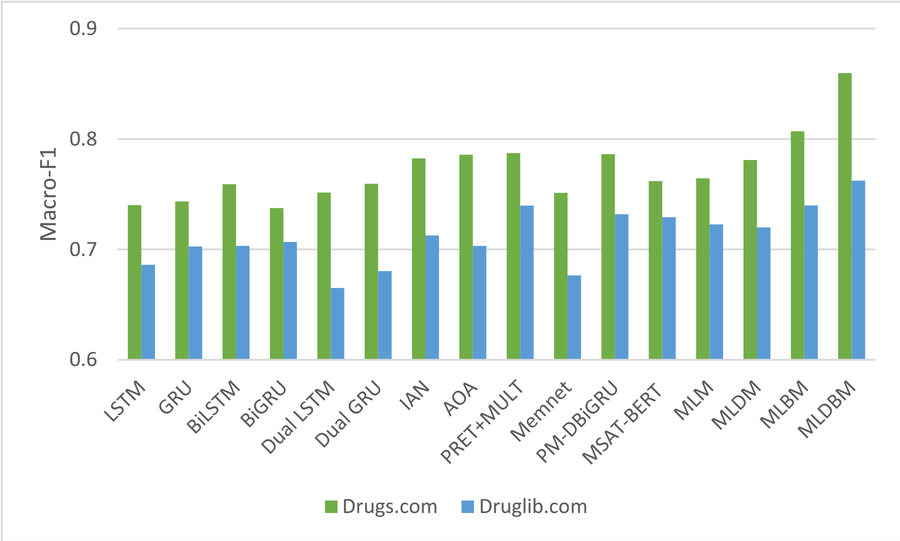

Fig. 7 Macro-F1 comparison of different models on the Drugs.com and Druglib.com datasets

5.3 Proposed Model Analysis

An analysis has been conducted for various variations of MLDBM. The variations are Multi-task Learning based LSTM Model (MLM), Multi-task Learning based Dual LSTM Model (MLDM), Multi-task Learning based Bidirectional LSTM model (MLBM), and Multi-task Learning based Dual Bidirectional LSTM model (MLDBM).

The MLM model consists of a single LSTM layer that learns input and aspect-specific representations simultaneously. An attention mechanism is then employed to generate the final aspect-specific representations, which are subsequently used for sentiment polarity prediction through a softmax classifier. On the other hand, MLDM incorporates two LSTM layers. The first LSTM layer processes the input sequence embeddings, while the second LSTM layer specifically handles aspect-specific representations. The outputs from both LSTM networks are combined through concatenation. Subsequently, the attention mechanism is applied, and the softmax classifier predicts the sentiment polarity.

In comparison, the MLBM model has a similar architecture to the MLM but differs by utilizing a single bidirectional LSTM network. The other layers in the MLBM remain consistent with the previously discussed variations. On the other hand, the MLDBM model consists of two bidirectional LSTM layers, which facilitate the inclusion of additional contextual information essential for ABSA, as discussed in section 3.

The results showed that the MLM demonstrated promising accuracy for both datasets. Specifically, MLM achieved an accuracy of 77.12% for the drugs.com dataset and 73.12% for the druglib.com dataset. The performance achieved by MLM is better than that of most of the baselines such as LSTM, GRU, Dual LSTM, and Dual GRU, used in this study for comparison purposes. Although the MLM effectively captures sentiment information, the analysis indicates that the attention mechanism does not contribute significantly to performance improvement.

MLDM outperformed MLM with an accuracy of 78.21% on the drugs.com dataset. This improvement can be attributed to the inclusion of individual networks which introduce an additional level of abstraction to the model. However, when evaluating the duglib.com dataset, MLDM achieved a similar accuracy of 73.59% compared to MLM. This is due to fewer training examples in the druglib.com dataset which limited the ability of MLDM to effectively utilize an additional level of abstraction.

MLBM outperformed MLDM by an accuracy of 81.36% for the drugs.com dataset and 75.31% for the druglib.com dataset. This is due to the utilization of a bidirectional LSTM, which allows it to capture more comprehensive semantics. Additionally, the incorporation of an attention method further enhances the performance of the MLBM.

However, the MLDBM outperformed all previous variations, demonstrating its superiority in capturing important aspects relevant to ABSA. By encoding the input in both directions using bidirectional LSTM layers, MLDBM effectively captures a wide range of contextual information. The results highlight the outstanding performance of MLDBM, achieving an accuracy of 87.71% and a Macro-F1 of 85.97% for the drugs.com dataset. On the druglib.com dataset, MLDBM achieved an accuracy of 78.97% and Macro-F1 of 76.23%. The findings emphasize that MLDBM excels in capturing the aspects leveraging dual bidirectional LSTM layers, resulting in significantly improved accuracy and Macro-F1 compared with the previous variations.

To further analyze the proposed model, we examined the MLDBM and its variants with different learning rates and dropout rates. Different learning rates were tested, and their corresponding results are presented in Table 5. The most optimal learning rate was found to be 0.001, which resulted in the highest accuracy and Macro-F1 score for both datasets.

Different dropout rates were also evaluated, and the results are shown in Table 6. The most optimal outcomes were attained by gradually increasing the dropout rate from 0.2 to 0.5. The highest accuracy and Macro-F1 were obtained when the dropout rate was set to 0.5 for both datasets.

Table 5 Analysis of results with different learning rates

Learning Rate	Model	Drugs.com		Druglib.com	
		Accuracy	Macro-F1	Accuracy	Macro-F1
0.1	MLM	0.7341	0.7175	0.7034	0.6956
	MLDM	0.7483	0.7232	0.7061	0.6813
	MLBM	0.7757	0.7341	0.7228	0.7065
	MLDBM	0.8424	0.8182	0.7353	0.7171
0.01	MLM	0.7583	0.7495	0.7303	0.7138
	MLDM	0.7623	0.7329	0.7183	0.7091
	MLBM	0.7976	0.7623	0.7265	0.7184
	MLDBM	0.8652	0.8248	0.7486	0.7373
0.001	MLM	0.7712	0.7643	0.7312	0.7227
	MLDM	0.7821	0.7809	0.7359	0.7199
	MLBM	0.8136	0.807	0.7531	0.7398
	MLDBM	0.8771	0.8597	0.7897	0.7623
0.0001	MLM	0.7601	0.7436	0.7294	0.7015
	MLDM	0.7583	0.7247	0.7267	0.7135
	MLBM	0.7797	0.7554	0.7352	0.7186
	MLDBM	0.8507	0.8143	0.7389	0.7294

Table 6 Analysis of results with different dropout rates

Dropout Rate	Model	Drugs.com		Druglib.com	
		Accuracy	Macro-F1	Accuracy	Macro-F1
0.2	MLM	0.7368	0.7237	0.6934	0.6783
	MLDM	0.7589	0.7382	0.7268	0.6953
	MLBM	0.7956	0.7774	0.7286	0.7012
	MLDBM	0.8128	0.8057	0.7227	0.7039
0.3	MLM	0.7591	0.7372	0.7042	0.6817
	MLDM	0.7742	0.7581	0.7313	0.7085
	MLBM	0.8049	0.7943	0.7425	0.7256
	MLDBM	0.8515	0.8173	0.7473	0.7345
0.4	MLM	0.7424	0.7201	0.6971	0.6734
	MLDM	0.7613	0.7483	0.7045	0.6873
	MLBM	0.7686	0.7257	0.7147	0.7038
	MLDBM	0.8339	0.8116	0.7232	0.7184
0.5	MLM	0.7712	0.7643	0.7312	0.7227
	MLDM	0.7821	0.7809	0.7359	0.7199
	MLBM	0.8136	0.807	0.7531	0.7398
	MLDBM	0.8771	0.8597	0.7897	0.7623

Furthermore, the MLDBM model requires approximately 134 seconds per epoch when running on a CPU, and around 63 seconds per epoch when utilizing a GPU, representing a significant reduction in processing time. However, it is important to note that hardware capabilities have been advancing rapidly, and optimization techniques like parallel computing can further enhance efficiency and accelerate the computational process.

To evaluate the statistical significance of the proposed approach, we conducted the Wilcoxon signed-rank test on the classification results, as described in [36]. On comparing MLDBM with various models, the analysis demonstrated a significant overall p value below 0.05 for the Macro-F1 metric. These findings indicate that MLDBM holds the potential for further advancement in ABSA and related domains. Additionally, the results suggest that MLDBM can be effectively integrated with other classifiers to enhance their performance in SA and other relevant areas.

5.4 Case study

To gain a better understanding of the proposed methodology, a case study was conducted using a set of drug reviews, focusing on identifying words that contribute significantly to ABSA. The attention weights assigned to individual words in the sentences were visualized in Fig. 8, where the intensity of the color reflected the importance of each word. Words with more intense colors were deemed more significant, while less intense colors indicated less important words.

The analysis revealed that the model ignored common words like “an,” “,”, “the,” “and” and “of” as they didn’t have a significant impact on sentiment classification. However, the model placed considerable emphasis on aspect-related words such as “pain,” “swelling,”

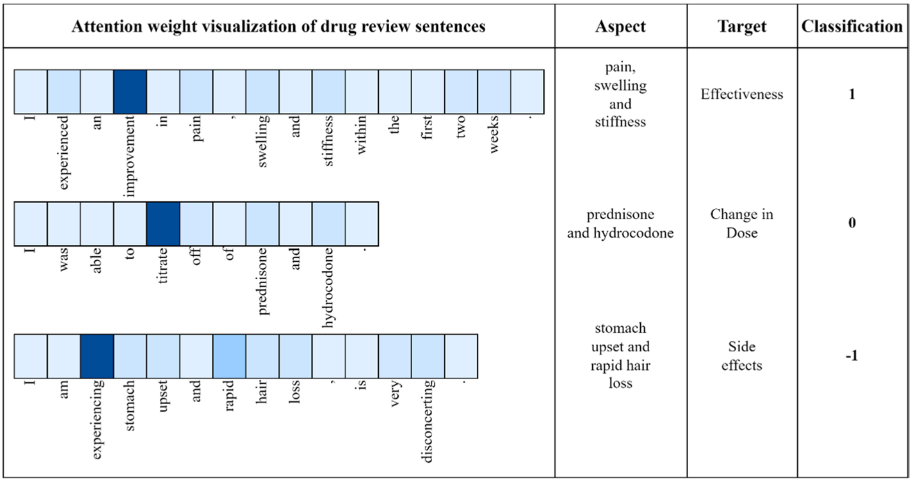

Fig. 8 Attention weight visualization of drug review sentences

"stiffness," "prednisone," "hydrocodone," "stomach upset," and "rapid hair loss." These words were given high attention weights, indicating their relevance in ABSA.

The results obtained from the proposed methodology demonstrate the model's capability to assess sentiment polarity by considering both aspects and semantics. For example, in the first instance, in Fig. 8, the model accurately identifies the aspect of "pain, swelling, and stiffness" and associates it with a positive sentiment conveyed through the word "improvement." Likewise, in the second example, the model correctly recognizes the phrase "titrate off" which refers to gradually reducing medication, classifying it as a neutral sentiment. Furthermore, the model successfully identifies the negative sentiment expressed in the phrase "experiencing stomach upset and rapid hair loss."

The model showcases the superiority of BERT as a word embedding model over other state-of-the-art approaches. Additionally, the combination of contextual embeddings, dual bidirectional LSTM network, and attention mechanism allows for effective classification of each word, leading to enhanced feature extraction capabilities by considering the most appropriate semantic meaning.

6 Conclusion

This paper presents Multi-task Learning based Dual Bidirectional LSTM Model (MLDBM) for ABSA which can categorize the extracted aspects into predefined sentiment categories. Specifically, the MLDBM integrates BERT and attention which jointly perform aspect extraction and sentiment analysis. The proposed model draws attention to the aspect terms in the sentence using multi-head self attention for generating comprehensive and enhanced aspect-specific representations. The dual BiLSTM architecture efficiently processes those representations and employs attention for sentiment classification. The experimental results validate that the proposed approach attains the highest performance and demonstrates the potential of MLDBM as a robust model for handling complex ABSA tasks i.e., aspect extraction and sentiment classification. Furthermore, MLDBM outperformed the state-of-the-art models on widely used benchmark datasets for ABSA. In the future, we will explore the effectiveness of integrating the latest language models. It is also valuable to analyze and compare alternative neural network architectures. The incorporation of external domain-specific knowledge is another avenue for future research. Furthermore, extending the proposed model to handle multiple languages will broaden its applicability and impact.

Funding The authors received no financial support for the research.

Data availability The datasets analyzed during the current study are available on reasonable request.

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

1. Abdelgwad MM, Soliman TH, Taloba AI, Farghaly MF (2022) Arabic aspect based sentiment analysis using bidirectional GRU based models. J King Saud Univ-Comput Inf Sci 34(9):6652–6662. https://doi.org/10.48550/arXiv.2101.10539
2. Alsayat A (2022) Improving sentiment analysis for social media applications using an ensemble deep learning language model. Arab J Sci Eng 47(2):2499–2511. https://doi.org/10.1007/s13369-021-06227-w
3. Basiri ME, Abdar M, Cifci MA, Nemati S, Acharya UR (2020) A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques. Knowl-Based Syst 198:105949. https://doi.org/10.1016/j.knosys.2020.105949
4. Bayraktar K, Yavanoglu U, Ozbilen A (2019) A rule-based holistic approach for Turkish aspect-based sentiment analysis. In: 2019 IEEE International Conference on Big Data (Big Data) pp 2154–2158. https://doi.org/10.1109/BigData47090.2019.9005473
5. Bensoltane R, Zaki T (2022) Towards Arabic aspect-based sentiment analysis: A transfer learning-based approach. Soc Netw Anal Min 2:1–6. https://doi.org/10.1007/s13278-021-00794-4
6. Bhatti UA, Wu G, Bazai SU, Nawaz SA, Baryalai M, Bhatti MA, Hasnain A, Nizamani MM (2022) A Pre-to Post-COVID-19 change of air quality patterns in anhui province using path analysis and regression. Pol J Environ Stud 31(5). https://doi.org/10.15244/pjoes/148065
7. Bonifazi G, Breve B, Cirillo S, Corradini E, Virgili L (2022) Investigating the COVID-19 vaccine discussions on Twitter through a multilayer network-based approach. Inf Process Manag 59(6):103095. https://doi.org/10.1016/j.ipm.2022.103095
8. Chen J, Tam D, Raffel C, Bansal M, Yang D (2013) An empirical survey of data augmentation for limited data learning in NLP. Trans Assoc Comput Linguist 11:191–211. https://doi.org/10.1162/tacl_a_00542
9. Colón-Ruiz C, Segura-Bedmar I (2020) Comparing deep learning architectures for sentiment analysis on drug reviews. J Biomed Inform 110:103539. https://doi.org/10.1016/j.jbi.2020.103539
10. Dubey G, Singh HP, Sheoran K, Dhand G, Malik P (2023) Drug review sentimental analysis based on modular lexicon generation and a fusion of bidirectional threshold weighted mapping CNN-RNN. Concurr Comput Pract Exp 35:e7512. https://doi.org/10.1002/cpe.7512
11. Geetha MP, Renuka DK (2021) Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model. Int J Intell Netw 2:64–69. https://doi.org/10.1016/j.ijin.2021.06.005
12. Gräßer F, Kallumadi S, Malberg H, Zaunseder S (2018) Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning. In: Proceedings of the 2018 international conference on digital health pp 121–125 https://doi.org/10.1145/3194658.3194677
13. Han Y, Liu M, Jing W (2020) Aspect-level drug reviews sentiment analysis based on double BiGRU and knowledge transfer. IEEE Access 8:21314–21325. https://doi.org/10.1109/ACCESS.2020.2969473

14. He R, Lee WS, Ng HT, Dahlmeier D (2018) Exploiting document knowledge for aspect-level sentiment classification. arXiv preprint arXiv:1806.04346. https://doi.org/10.48550/arXiv.1806.04346
15. Huang B, Ou Y, Carley KM (2018) Aspect level sentiment classification with attention-over-attention neural networks. In: Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, Proceedings 11 2018 (pp 197–206). Springer International Publishing. https://doi.org/10.48550/arXiv.1804.06536
16. Huang S, Huang M, Zhang Y, Chen J, Bhatti U (2020) Medical image segmentation using deep learning with feature enhancement. IET Image Process 14(14):3324–3332. https://doi.org/10.1049/iet-ipr.2019.0772
17. Jayanto R, Kusumaningrum R, Wibowo A (2022) Aspect-based sentiment analysis for hotel reviews using an improved model of long short-term memory. Int J Adv Intell Inform 8(3):391–403. https://doi.org/10.26555/ijain.v8i3.691
18. Jiménez-Zafra SM, Martín-Valdivia MT, Molina-González MD, Ureña-López LA (2019) How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for the medical domain. Artif Intell Med 93:50–57. https://doi.org/10.1016/j.artmed.2018.03.007
19. Karsi R, Zaim M, El Alami J (2021) Leveraging Pre-Trained Contextualized Word Embeddings to Enhance Sentiment Classification of Drug Reviews. Rev d'Intelligence Artif 35(4):307–314. https://doi.org/10.18280/ria.350405
20. Ke Z, Sheng J, Li Z, Silamu W, Guo Q (2021) Knowledge-guided sentiment analysis via learning from natural language explanations. IEEE Access 9:3570–3578. https://doi.org/10.1109/ACCESS.2020.3048088
21. Li X, Bing L, Zhang W, Lam W (2019) Exploiting BERT for end-to-end aspect-based sentiment analysis https://doi.org/10.48550/arXiv.1910.00883
22. Lin Y, Wang C, Song H, Li Y (2021) Multi-head self-attention transformation networks for aspect-based sentiment analysis. IEEE Access 9:8762–8770. https://doi.org/10.1109/ACCESS.2021.3049294
23. Liu Q, Zhang H, Zeng Y, Huang Z, Wu Z (2018) Content attention model for aspect based sentiment analysis. In: Proceedings of the 2018 world wide web conference pp 1023–1032. https://doi.org/10.1145/3178876.3186001
24. Ma D, Li S, Zhang X, Wang H (2017) Interactive attention networks for aspect-level sentiment classification. arXiv preprint arXiv:1709.00893. https://doi.org/10.48550/arXiv.1709.00893
25. Ma Y, Peng H, Cambria E (2018) Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In: Proceedings of the AAAI conference on artificial intelligence, vol. 32, no. 1. https://doi.org/10.1609/aaai.v32i1.12048
26. Ma Y, Peng H, Khan T, Cambria E, Hussain A (2018) Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis. Cogn Comput 10:639–650. https://doi.org/10.1007/s12559-018-9549-x
27. Miao YL, Cheng WF, Ji YC, Zhang S, Kong YL (2021) Aspect-based sentiment analysis in Chinese based on mobile reviews for BiLSTM-CRF. J Intell Fuzzy Syst 40(5):8697–8707. https://doi.org/10.3233/JIFS-192078
28. Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends® Inf Retr 2(1–2):1–35. https://doi.org/10.1561/1500000011
29. Patil RS, Kolhe SR (2022) Supervised classifiers with TF-IDF features for sentiment analysis of Marathi tweets. Soc Netw Anal Min 12(1):51. https://doi.org/10.1007/s13278-022-00877-w
30. Ray P, Chakrabarti A (2022) A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis. Appl Comput Inform 18(1/2):163–178. https://doi.org/10.1016/j.aci.2019.02.002
31. Setiawan EI, Ferry F, Santoso J, Sumpeno S, Fujisawa K, Purnomo MH (2020) Bidirectional GRU for targeted aspect-based sentiment analysis based on character-enhanced token-embedding and multi-level attention. Computing. 1:2. https://doi.org/10.22266/ijies2020.1031.35
32. Sun C, Huang L, Qiu X (2019) Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence https://doi.org/10.48550/arXiv.1903.09588
33. Sweidan AH, El-Bendary N, Al-Feel H (2021) Sentence-Level Aspect-Based Sentiment Analysis for Classifying Adverse Drug Reactions (ADRs) Using Hybrid Ontology-XLNet Transfer Learning. IEEE Access 9:90828–90846. https://doi.org/10.1109/ACCESS.2021.3091394
34. Tang D, Qin B, Liu T (2016) Aspect level sentiment classification with deep memory network. In: Proceedings of the 2016 conference on empirical methods in natural language processing pp 214–224, Austin, Texas. Association for Computational Linguistics. https://doi.org/10.48550/arXiv.1605.08900
35. Wang Y, Huang M, Zhu X, Zhao L (2016) Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of the 2016 conference on empirical methods in natural language processing pp 606–615. https://doi.org/10.18653/v1/D16-1058

36. Wilcoxon F (1945) Individual Comparisons by Ranking Methods. Biom Bull 1(6):80–83. https://doi.org/10.2307/3001968
37. Xiao L, Xue Y, Wang H, Hu X, Gu D, Zhu Y (2022) Exploring fine-grained syntactic information for aspect-based sentiment classification with dual graph neural networks. Neurocomputing 471:48–59. https://doi.org/10.1016/j.neucom.2021.10.091
38. Xu H, Liu B, Shu L, Yu PS (2019) BERT post-training for review reading comprehension and aspect-based sentiment analysis. https://doi.org/10.48550/arXiv.1904.02232
39. Xue W, Li T (2018) Aspect based sentiment analysis with gated convolutional networks https://doi.org/10.48550/arXiv.1805.07043
40. Yadav S, Ekbal A, Saha S, Bhattacharyya P (2018) Medical sentiment analysis using social media: Towards building a patient assisted system. In: Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)
41. Yang C, Zhang H, Jiang B, Li K (2019) Aspect-based sentiment analysis with alternating coattention networks. Inf Process Manag 56(3):463–478. https://doi.org/10.1016/j.ipm.2018.12.004
42. Zhao A, Yu Y (2021) Knowledge-enabled BERT for aspect-based sentiment analysis. Knowl-Based Syst 227:107220. https://doi.org/10.1016/j.knosys.2021.107220
43. Zhao H, Liu Z, Yao X, Yang Q (2021) A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. Inf Process Manag 58(5):102656. https://doi.org/10.1016/j.ipm.2021.102656
44. Žunić A, Corcoran P, Spasić I (2022) The case of aspect in sentiment analysis: seeking attention or co-dependency? Mach Learn Knowl Extr 4(2):474–487

Authors and Affiliations

Somiya Rani[1] · Amita Jain[2]

Somiya Rani
somiya1093@gmail.com

[1] Department of Computer Science and Engineering, Netaji Subhas University of Technology East Campus (erstwhile A.I.A.C.T.R.), Guru Gobind Singh Indraprastha University, Delhi, India

[2] Department of Computer Science and Engineering, Netaji Subhas University of Technology, Delhi, India