Received: June 23, 2024. Revised: August 4, 2024.


International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

# Aspect Based Sentiment Analysis Using Modified Latent Dirichlet Allocation and Optimized BERT with LSTM Classification

**Nasreen Taj M B[1]*** **Girisha Gowdra Shivappa[2]**

[1]*Department of Information Science and Engineering, Gowdara Mallikarjunappa Institute of Technology, Davanagere, Karnataka, India*
[2]*Department of Computer Science and Engineering, Dayananda Sagar University, Bengaluru, Karnataka, India*
* Corresponding author's Email: nasreent@gmit.ac.in

**Abstract:** Aspect-Based Sentiment Analysis (ABSA) is essential in industries such as healthcare, automotive, and finance, where understanding customer feedback is crucial for business strategies and product development. Traditional sentiment analysis methods often fail to account for specific aspects, providing only generalized sentiment scores (polarities). This study aims to develop a robust ABSA architecture to accurately classify aspect-based sentiment polarity. We propose a Modified Latent Dirichlet Allocation (M-LDA) model that integrates Term Frequency-Inverse Document Frequency (TF-IDF) and Latent Dirichlet Allocation (LDA) to extract relevant aspects. Additionally, an Optimized Bidirectional Encoder Representations from Transformers (O-BERT) model is used to determine aspect-based sentiment polarity, categorizing sentiments as positive, negative, or neutral. A Long Short-Term Memory (LSTM) network is then employed to enhance classification accuracy. The M-LDA O-BERTLSTM for ABSA (MOL-ABSA) performance, evaluated using accuracy and macro-F-score metrics on the SemEval 2014 and a generalized dataset, demonstrates its effectiveness, achieving accuracies of 81.19%, 85.26%, and 87.53%, respectively. This integrated approach combines traditional NLP techniques with advanced machine learning models to ensure high accuracy.

**Keywords:** ABSA, LDA, TF-IDF, BERT, SemEval 2014.

## 1. Introduction

Aspect-Based Sentiment-Analysis (ABSA) is a branch of Natural-Language-Processing (NLP) that seeks to identify similarities in the way individuals feel towards specific aspects of textual information, such as reviews of products, social networking posts, or feedback from consumers [1]. ABSA is different from other sentiment evaluation techniques since it looks for scores (polarity), i.e., neutral, negative and positive polarities towards certain sections and structures in a text. Instead of providing a broad sentiment assessment for the whole text like traditional approaches do, ABSA aims to provide a more detailed evaluation by looking at how people feel about specific aspects from review [2]. For example, while reviewing a restaurant, ABSA may segregate customer sentiment based on factors such as meal quality, atmosphere, service, and price [3]. ABSA is useful in many different fields in which it is essential to know how customers feel about particular items or services. Organizations operating in the online and offline industries can benefit from ABSA's assistance in analyzing consumer input regarding product characteristics, costs, transportation, and satisfaction. This allows them to arrive at decisions based on data and enhance their client experience [4]. Moreover, to improve customer satisfaction, hotels employ ABSA to assess how guests feel about things like hotel cleanliness, staff behavior, and food options [5]. Additionally, ABSA is useful in industries such as healthcare, automobiles, banking, and others wherein sentiment research and customer input are integral to company strategy and development of products [6].

Received: June 23, 2024. Revised: August 4, 2024. 

In light of the numerous shortcomings of conventional sentiment analysis, which yields a generic sentiment score that fails to take into account the relevance of certain aspects, ABSA becomes necessary [7, 8]. With ABSA, companies may get a deeper insight into consumer sentiment, which in turn helps them pinpoint their own strengths, shortcomings, and potential for improvement. Nevertheless, there are a few issues with the conventional Deep-Learning (DL) [9, 10] and Machine-Learning (ML) [11, 12] approaches that ABSA makes use of. In addition, DL approaches in ABSA tend to be built on specific datasets and domains, which limits their adaptability to different contexts or sectors [13]. Since they are trained on a certain domain, they can fail to be able to generalize well to other datasets or new patterns, which can lead to bias and poor performance [14]. Also, a lot of the current ABSA approaches just look at the general polarity during evaluations, not the sentiments at the aspect-level, resulting in biased and incorrect results [15]. Hence, in this work we present an approach which extracts aspects from the reviews and then evaluates the polarity of the overall review. This helps to capture the aspects and overall polarity. The contribution of this work are as follows:

- This work presents a generalized model for ABSA for aspect-level sentiment analysis. This work utilizes the strengths from the topic modelling approach Latent-Dirichlet Allocation (LDA) and modifies it by inclusion of Term Frequency-Inverse Document Frequency (TF-IDF). This Modified LDA (M-LDA) helps in extraction of aspects.
- Further, for extraction of aspects, this work utilizes a Transformer-based DL approach called Bidirectional Encoder Representations from Transformers (BERT). The BERT model fails to achieve better results when the size of data is big, hence, the we present an Optimized-BERT (O-BERT). The O-BERT helps in extracting the polarities from aspect-level and review-level.
- Finally, for evaluation the classification of the O-BERT approach, this work utilizes the LSTM approach. The LSTM approach utilizes Adam optimizer and dense layer for optimizing the classification results.
- The evaluation has been done using the standard SemEval 2014 Task 4 as it provides two datasets, i.e., laptops and restaurants. Further, this work also has collected a hospital dataset for evaluation of the proposed M-LDA O-BERT LSTM approach. This helps in the evaluation and presents a generalized model for ABSA.

The manuscript is organized in the following manner. In Section II, the literature survey is discussed. Then in Section III, the M-LDA O-BERT LSTM approach is discussed in detail. In Section IV, the results for the M-LDA O-BERT LSTM are discussed. Finally, in Section V, the conclusion along with the future work is presented.

## 2. Literature survey

ABSA has gained significant attention in recent years due to its ability to extract sentiments from text, particularly in domains such as healthcare. This section delves into several notable studies and methodologies in ABSA, shedding light on the evolution of sentiment analysis techniques in healthcare and related domains. A. Bansal et al., [16] were pioneers in ABSA within the realm of hospital reviews. Their groundbreaking study involved collecting data from 500 hospitals, comprising over 30,000 reviews. These reviews underwent rigorous preprocessing and subsequent polarity evaluation. The study focused on four main aspects of hospital services, evaluating sentiment ratings associated with each aspect. This foundational work laid the groundwork for subsequent advancements in hospital review sentiment analysis. P. Han et al., [17] presented an end-to-end generative approach for aspect and opinion extraction from hospital reviews, similar to Bansal et al. [16]. Their methodology involved collecting and preprocessing hospital review datasets, focusing on precision, recall, and f-score as evaluation metrics. Their results showcased impressive precision scores of 85.46%, 75.33%, and 69.05% for aspect-sentiment, operation-sentiment, and aspect-sentiment-opinion triplet extraction, respectively. K. Denecke et al., [18] conducted a comprehensive review of recent works in healthcare sentiment analysis. Their analysis revealed a trend towards utilizing built-in libraries such as SentiWordNet and TextBlob, with a predominant focus on ML techniques. Findings from their review indicated that recent studies achieved accuracy levels ranging from 71.5% to 88.2%, showcasing the efficacy of ML approaches in healthcare sentiment analysis tasks. C. Golz et al., [19] focused on understanding sentiment patterns among healthcare professionals regarding technological changes. Utilizing the SentimentWortschatz dataset, a German-language sentiment evaluation dataset, their findings revealed that approximately 73% of reviews expressed positive sentiments towards technological advancements in healthcare.

Moving beyond specific studies, advancements in DL techniques have significantly enhanced ABSA capabilities. B. Yu et al., [20] introduced the Multi-Weight Graph-Convolutional-Network (MWGCN), leveraging DL for aspect-related sentiment detection. MWGCN utilizes weighted adjacency graphs and Graph Convolutional Networks (GCN) for context feature extraction, achieving competitive accuracy levels of 79.98% and 86.36% on the SemEval 2014 laptop and restaurant datasets, respectively. Y. Ma et al., [21] presented a Multiple-GCN DL technique for context feature extraction in sentiment analysis, enhancing classification accuracy through a combined similarity loss and traditional loss functions. Their methodology, evaluated on the SemEval 2014 dataset, achieved notable accuracy rates of 78.8% and 83.82% for laptop and restaurant datasets, respectively. T. Lin et al., [22] proposed a masked attention approach with local and global embeddings for aspect term position evaluation, leveraging datasets such as SemEval 2014, SemEval 2016, and Multi-Aspect Multi-Sentiment (MAMS). Their evaluation based on macro-f-score yielded impressive results, highlighting the efficacy of their approach in feature extraction and sentiment classification across different datasets. Q. Zhao et al., [23] introduced a Structured-Dependency Tree-Based GCN (SDTGCN), a DL approach considering positional information for aspect-pivotal word connections. Their evaluation across various datasets, including SemEval and Twitter, demonstrated the versatility and effectiveness of SDTGCN in sentiment analysis tasks, achieving accuracies ranging from 76.25% to 91.53%.

From the above review, recent advancements in ABSA primarily focus on leveraging DL techniques for improved accuracy and feature extraction. The absence of standard datasets in specific domains, such as healthcare, prompts researchers to collect domain-specific data for training and evaluation. The SemEval dataset has emerged as a standard benchmark for evaluating ABSA models due to its diverse domains and well-annotated data. Hence, this work aims to contribute to the evolving landscape of ABSA by addressing gaps in existing approaches, utilizing DL methodologies, and leveraging domain-specific datasets for accurate sentiment analysis. The proposed methodology and its detailed discussion will be presented in the subsequent section, providing insights into the innovative approach adopted for sentiment analysis in healthcare contexts.

## 3. Methodology

### 3.1 Architecture

The architecture for ABSA in this approach begins with a comprehensive consideration of the dataset, as presented in Fig. 1, ensures that it is representative and suitable for the task at hand. Further, the preprocessing follows, where the data undergoes cleaning, normalization, and tokenization to prepare it for further analysis. The core analytical process involves a M-LDA model, which integrates TF-IDF and LDA. TF-IDF helps in emphasizing important words in the documents while LDA is used to uncover the hidden thematic structure of the dataset, enabling the extraction of relevant aspects from the text. Once the aspects are identified using M-LDA, an O-BERT model comes into play. O-BERT, leveraging the power of the BERT architecture, fine-tunes the ABSA polarity classification by analyzing the extracted aspects to determine whether they carry positive, negative, or neutral aspect-based sentiments. The aspect-based sentiments and the review sentences are then assigned a score based on their polarity. To enhance the classification of the aspect-based sentiments and the review sentences, a LSTM network is employed. LSTM, known for its capability to handle sequential data and capture long-term dependencies, further refines the ABSA, ensuring a more accurate classification by learning from the sequence of words in the text. The final evaluation of the aspect sentiment polarity classification is carried out using metrics accuracy and macro-F-score. Accuracy measures the overall correctness of the classification, while the macro-F-score provides a balanced evaluation of the precision and recall across all classes (positive, negative, and neutral), ensuring that

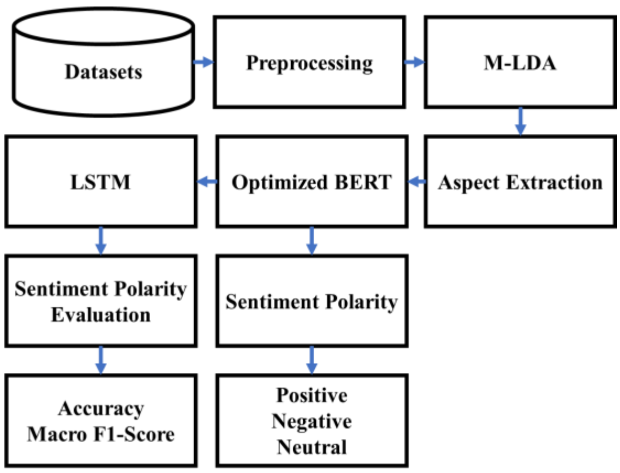


Figure. 1 Architecture of proposed M-LDA O-BERT-LSTM for ABSA (MOL-ABSA)

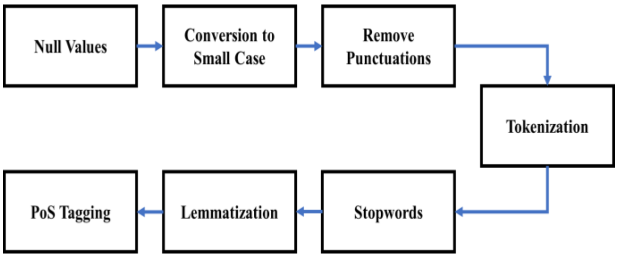

Figure. 2 Steps of preprocessing used

the model performs well across different sentiment categories. This comprehensive architecture thus provides a robust framework for detailed ABSA.

### 3.2 Preprocessing

In the preprocessing phase of the architecture, as presented in Fig. 2, the dataset is initially examined to ensure its suitability for further analysis. This involves evaluating the dataset for any null values, which are identified and appropriately handled, by removing null values to maintain data integrity. Once the dataset is free of null values, it is standardized by converting all text to lowercase, ensuring uniformity and reducing redundancy in textual data. Following this, punctuation marks are removed from the dataset to eliminate any extraneous characters that do not contribute to the semantic meaning of the text. The next step is tokenization, where the text is split into individual words or tokens, facilitating more granular analysis. Stopwords, which are common words that do not add significant meaning (such as “and,” “the,” and “in”), are then removed from the tokenized text to streamline the data and focus on more meaningful words. Lemmatization is then applied to the dataset, transforming words to their base or root forms. This process helps in reducing word variations and ensures that different forms of a word are treated as a single entity. Finally, PoS tagging is performed, which assigns grammatical categories (such as nouns, verbs, adjectives) to each token, providing further context and aiding in the understanding of the syntactic structure of the sentences. After these preprocessing steps, the refined dataset is ready for the M-LDA model. This model, which integrates TF-IDF and LDA, utilizes the pre-processed text to identify and extract thematic structures and relevant aspects from the data, forming the foundation for subsequent sentiment analysis and classification stages in the architecture. The process of M-LDA is discussed in the next section.

### 3.3 M-LDA

When it comes to text analysis, the TF-IDF plays an important role, especially when utilizing LDA, an

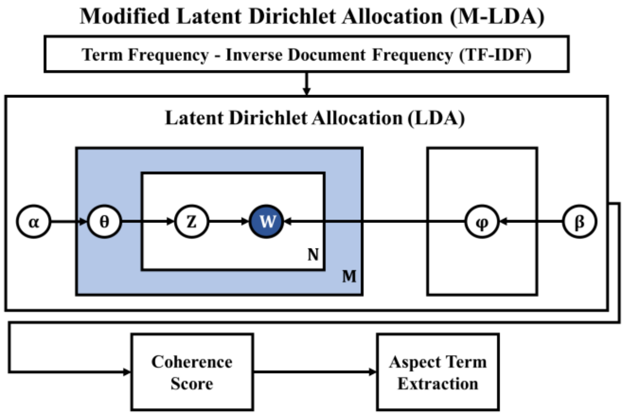

Figure. 3 M-LDA

important ML clustering technique [24]. In the context of this work, LDA helps in extracting aspects from the given review data. LDA is a topic modelling approach which extracts related words with respect to the given topic. To enhance the LDA, this work combines TF-IDF and LDA and presented a M-LDA. The complete structure of M-LDA is presented in Fig. 3.

The complete process of M-LDA is performed in three steps. In Step 1, the TF-IDF is constructed. For the construction of the TF-IDF, initially, the cleaned processed data is considered into a dictionary. The dictionary consists of only pre-preprocessed words which has various words. Hence to keep a total count of the overall occurences of the frequently occuring words, the dictionary is converted into Bag-of-Words (BoW). The BoW helps in creating a vector set which contains the overall occurences of the words in review data. The BoW is further considered as input for TF-IDF. The TF-IDF helps to extract the most important words along with less important words from the BoW using the review data. After the extraction of words from the TF-IDF, the process goes to Step 2, where the LDA is used to assign the words a given topic. Further, in Step 3, after the process of LDA, the coherence score is evaluated to understand which topic has highest relevant score. In this work, the Coherence-Validation (CV) is considered which is calculated using below equation [25].

$$\phi S_i(\vec{u},\vec{w}) = \frac{\sum_{i=1}^{|W|} u_i \cdot w_i}{||\vec{u}||_2 \cdot ||\vec{w}||_2} \quad (1)$$

After the evaluation of the coherence scores, the number of topics is selected by analyzing the coherence scores. The number of topics getting the highest coherence can be selected, but as the corups contains more than 50,000 reviews, it is impossible to gather all the topics. Hence, in this work, the number of topics has kept as 5. Further, after

Received: June 23, 2024. Revised: August 4, 2024. 

selecting the top 5 topics, the top keywords for the given topic were considered as aspects. After the extraction of the aspects, some of the words which were not relevent were discarded. Further, after the extraction of aspects, the initial sentence was further considered for extracting more deeper information using proposed optimized BERT model.

### 3.4 Optimized BERT

In the transformer-based approaches, the most utilized approach is BERT [26]. For understanding how much important an aspect is in the given review, the transformer-based approach, a kind of DL approach, uses self-attention process. When comparing neural networks, with transformer-based approaches, the transformer-based approaches provide parallel processing because of the self-attention process which provides contextual information using the positional embedding for the given review data. Hence, the BERT model has been considered. Further, as the dataset considered in this work is large, i.e., more than 50,000 reviews, the BERT model cannot provide better outcome. Hence, this work optimizes the BERT model, called as O-BERT. The architecture of the O-BERT is presented in Fig. 4.

In the O-BERT architecture, the CLS represents classification token, TOK represents the token and SEP represents separator. The review sentence goes through the token and position embedding where tokens and positions are assigned for each word in the review. Further, in this work we only utilize the encoder section of the BERT and no decoder has been used as the encoder part helps to create context-embeddings for the input review. The encoder section has a self-attention network along with a feed-forward-network which helps in mapping of input review sequences to context-encodings. Also, as the BERT approach utilizes static masking, the O-BERT utilizes dynamic masking for generating a representation which is context specific and diversified. Instead of using the Character-Level Byte-Pair-Encoding (CL-BPE), as used in BERT, this work considers Byte-Level Byte-Pair-Encoding (BL-BPE) for faster computation.

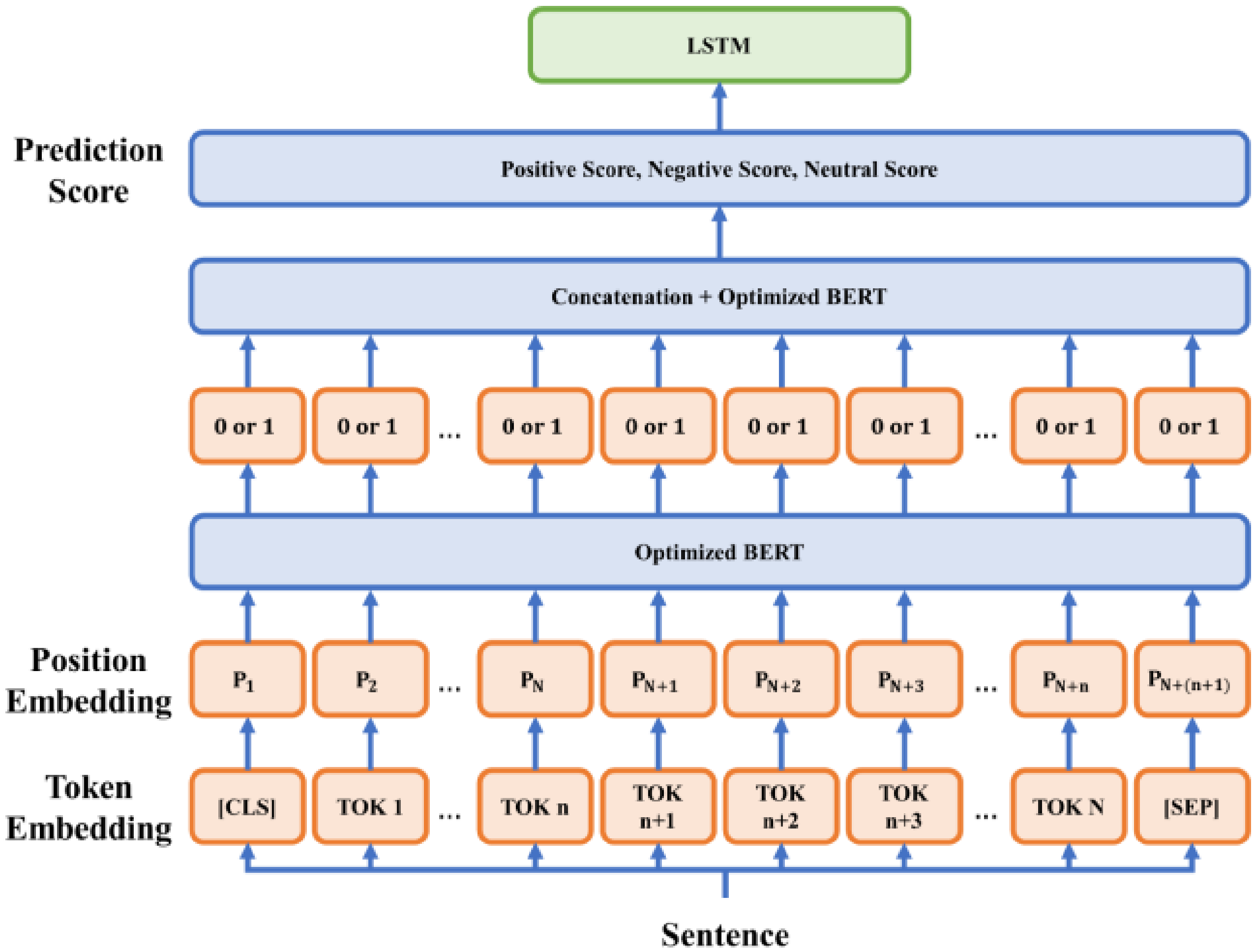


Figure. 4 O-BERT architecture

By making these optimizations, the O-BERT can consider large datasets with a greater number of batches and more sequences when compared with BERT. Prior to implementation of the attention-mask and input-ids, the input sequence was tokenized. During the concatenation process, the attention-mask helps in representing which token has more importance, whereas the input-ids helps in encoding the text into numerical form in a sequential way. After the concatenation process, both the attention-mask and input-ids are used as input for O-BERT. The final O-BERT has 768 hidden-states and 12 encoding-layers. The final layer of O-BERT provides the classification for the given review in negative, neutral or positive. After the classification, the classification is evaluated using the LSTM approach which is discussed in the next section.

### 3.5 LSTM

For solving the problems of neural networks, LSTM was introduced. The LSTM solves the problem of vanishing-gradients in neural networks. In LSTM network, the process of gates is crucial for encoding the long-range dependency of input, as presented in Fig. 5.
The LSTM usually consists of three gates, i.e., input, forget and output gate. The input-gate $(i_t)$ is utilized to determine which value holds significance for updating of the present stage. The forget-gate $(f_t)$ is utilized for determining whether the important information has to be dropped or stored. Finally, the output-gate $(o_t)$ is used for determining which information has to be transmitted for the subsequent hidden-state. The complete evaluation of the gates as presented in Fig. 5 is done using below presented equation.

$$f_t = \sigma(W_f X_t + U_f h_{t-1} + b_f \quad (2)$$

$$i_t = \sigma(W_i X_t + U_i h_{t-1} + b_i) \quad (3)$$

$$o_t = \sigma(W_o X_t + U_o h_{t-1} + b_o) \quad (4)$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (5)$$

$$\tilde{c}_t = \tanh(W_c X_t + U_c h_{t-1} + b_c) \quad (6)$$

$$h_t = o_t * \tanh(c_t) \quad (7)$$

The $h_t$ and $c_t$ represent hidden and cell-states for the given time-step $t$ having the input review sentence as $X_t$, $\sigma$ represents sigmoid-function which helps in controlling the input data in gates, i.e., either to block or pass the input data for the time-step $t$.

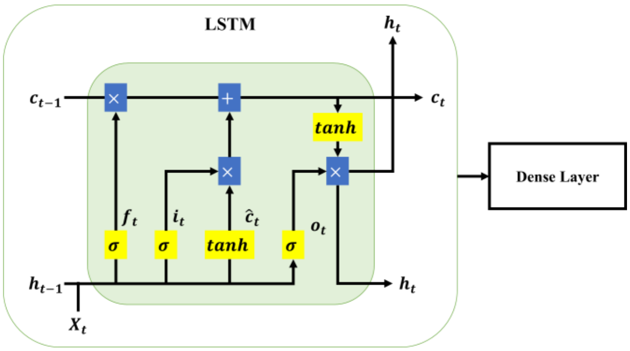


Figure. 5 LSTM

When the value of $\sigma$ is 0, it blocks the data for the given gate, i.e., the data will not be fed into the subsequent stage. When the value of $\sigma$ is 1, it does not block the data for the given gate, i.e., the data will be allowed to fed into the subsequent stage. Moreover, $(b_f, b_i, b_o, b_c)$ represent the bias values and $(W_f, W_i, W_o, W_c, U_f, U_i, U_o, U_c)$ represent the weight matrix for the respective gates. An overall of 256 units were considered for LSTM. Further, for better optimization, this work utilized the Adam Optimizer [27]. After the optimization, the LSTM goes to the dense layer for achieving better accuracy for classification.

### 3.6 Dense layer

In this work, two dense-layers were considered. In first layer, a total of 256 hidden-neurons were considered to capture the correct classification using LSTM. Further, in second layer, a Softmax activation-function was utilized. The Softmax helps in computation of probability distribution of the sentiment classification. The probability in the Softmax function is evaluated using the below equation.

$$Y = Softmax(CR_{e1} W_{e1} + b_{e1}) \quad (8)$$

In above Eq. (8),$CR_e = \{r_{e1}, r_{e2}, \dots, r_{el}\}$, where $r_{el}$ denotes every single token from the O-BERT and LSTM models, $W_{e1}$ denotes weight-matrix which is utilized to change the context tokens into classification form and $b_{e1}$ denotes bias-vector which helps in transforming the output for achieving better accuracy using LSTM. Further, the classification of ABSA is evaluated using two metrics, i.e., accuracy and macro-f-score which is evaluated using below equations.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

$$F - Score = \frac{2\times(\frac{TP}{TP+FP})\times(\frac{TP}{TP+FN})}{(\frac{TP}{TP+FP})+(\frac{TP}{TP+FN})} \quad (10)$$

$$Macro\ F - Score = Average\ of\ F - score \quad (11)$$

Where, $TP, TN, FP$ and $FN$ denote True-Positive, True-Negative, False-Positive and False-Negative respectively. In the next section, the results achieved using the MOL-ABSA is presented below.

## 4. Results and discussions

### 4.1 System requirements, parameter settings and dataset

The proposed MOL-ABSA approach was exeucted on AMD Ryzen 5 processor having 4 cores and 8 logical processors with 16 GB of RAM, running on Windows 11. Python was used for coding and execution process was done using the Annaconda Jupyter Notebook. The parameters considered for the MOL-ABSA is presented in Table 1.

For evaluation of this work, SemEval 2014 Task 4 [28] was utilized. The SemEval 2014 Task 4 consists of two different domains, i.e., laptop and restaurant, both having more than 1000 of reviews and labelled sentiments. The SemEval 2014 also has labelled polarities as negative, neutral and positive. Further, this work also considers another dataset which has been collected from two websites, Practo and Mouthshut which has hospital reviews [29, 30]. The dataset was collected by using Web-Scrapping Application-Programming-Interface (WS-API). The dataset did not contain any labelled sentiments and polarites. Hence, this MOL-ABSA approach was tested using a standard dataset, i.e., SemEval 2014 and one custom dataset, i.e., hospital reviews dataset to evaluate the performance.

Table 1. Parameter Settings

| Parameter | Value |
|---|---|
| Dropout-Rate | 0.1 |
| Optimizer | Adam |
| Learning Rate | 1.00E-05 |
| Batch Size | 16 |
| Max Sequence Length | 128 |
| Max Epoch | 5 |
| O-BERT Hidden States | 768 |
| O-BERT Encoding Layers | 12 |
| LSTM units | 256 |
| LSTM Dense Layer Hidden States | 256 |

### 4.2 Aspect extraction

In this section, the aspects extracted from the reviews using M-LDA are discussed. The results are presented in Table 2. Table 2 presents a collection of reviews along with their corresponding aspects. The first review mentions a “Strong” as the aspect with a positive polarity, emphasizing durability. The second review discusses satisfaction with the “performance” of a computer, indicating a positive sentiment towards this aspect. The third review, however, does not mention any specific aspect, hence, labelled as “No Aspect Term.” Moving on, the fourth review highlights “Doctors” as the aspect, praising their courteousness, professionalism, and sympathetic approach, indicating a positive sentiment. In contrast, the fifth review addresses the “Patients” aspect negatively, expressing dissatisfaction with poor service, lack of pro-activity among staff, and miscommunication leading to long wait times. Finally, the last review mentions “Members” of staff at Manipal Hospital, portraying a positive sentiment towards their cheerful demeanor, pleasantness, and neat appearance. Table 2 provides a structured overview of reviews, their associated aspects, and the sentiments expressed towards those aspects, offering valuable insights into customer experiences and perceptions.

Table 2. Review, Aspect

| Review | Aspect |
|---|---|
| I can say that I am fully satisfied with the performance that the computer has supplied. | Computer |
| All the doctors also sympathetic and professional in their approach. | Doctors |
| The members of staff of Manipal Hospital are very cheerful, pleasant and neat in their dressing. | Staff |

### 4.3 Review level polarity

In this section, the complete review has been considered and the polarity has been evaluated for the complete sentence using O-BERT. The results are presented in Table 3 which presents a breakdown of review-level polarity scores, indicating the distribution of negative, neutral, and positive sentiments for each review. The first review, praising the “Strong build” for durability, demonstrates a high positive polarity score of 0.895, indicating strong positive sentiment. The second review expressing satisfaction with computer performance also shows a high positive polarity score of 0.947. The third review about keyboard responsiveness shows a balanced distribution of sentiment, with a relatively high neutral polarity score of 0.39.

Received: June 23, 2024. Revised: August 4, 2024. 

Table 3. Review Level Polarity Scores

| Review | Negative | Neutral | Positive |
|---|---|---|---|
| Strong build though which really adds to its durability. | 0.003014 | 0.101324 | 0.895662 |
| All the doctors also sympathetic and professional in their approach. | 0.004468 | 0.048415 | 0.947117 |
| The members of staff of Manipal Hospital are very cheerful, pleasant and neat in their dressing. | 0.001363 | 0.015192 | 0.983444 |

Further, the fourth review highlights positive aspects about the hospital and doctors, such as courteousness and professionalism, receives a significantly high positive polarity score of 0.971. On the other hand, the fifth review, expressing dissatisfaction with hospital service, staff pro-activity, and miscommunication issues, exhibits a notably high negative polarity score of 0.952. Lastly, the sixth review praising the cheerful and pleasant demeanor of staff at Manipal Hospital attains a remarkably high positive polarity score of 0.983. Table 3 provides a quantitative insight into the sentiment distribution across various reviews, highlighting the varying degrees of positivity, negativity, and neutrality expressed in each review. This analysis aids in understanding customer perceptions and sentiments towards different aspects of the experiences under review.

### 4.4 Overall aspect based sentiment analysis

In this section, the ABSA classification results are discussed using the SemEval 2014 Task 4 dataset utilizing the LSTM. The results are presented in two metrics, i.e., accuracy and macro-f-score. The results for the accuracy and macro-f-score achieved by the SemEval 2014 Task 4 Laptop dataset are compared with existing works [20-23] and presented in Fig. 6. The models evaluated include MWGCN [20], MultiGCN [21], CDM-CDW [22], SDTGCN [23], and proposed MOL-ABSA. The findings show that MWGCN achieves an accuracy of 79.78% and a macro-f-score of 76.68%, indicating its ability to classify sentiments accurately on the SemEval Laptop dataset. Similarly, MultiGCN achieves a slightly lower accuracy of 78.8% but maintains a respectable macro-f-score score of 74.97%. The CDM-CDW model demonstrates improved

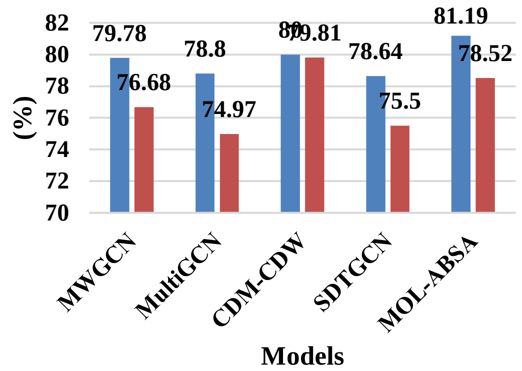


Figure. 6 Performance Evaluation using Laptop-2014 dataset

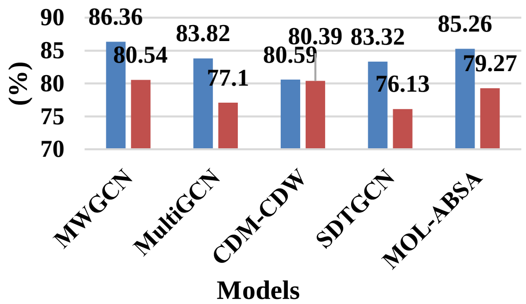


Figure. 7 Performance Evaluation using Laptop-2014 dataset

performance with an accuracy of 80% and a high macro-f-score of 79.81%, showcasing its effectiveness in sentiment analysis on this dataset. SDTGCN achieves an accuracy of 78.64% and a macro-f-score of 75.5%, demonstrating competitive performance compared to other models. Notably, the MOL-ABSA surpasses all other models, achieving the highest accuracy of 81.19% and a commendable macro-f-score of 78.52%. This indicates that the MOL-ABSA model offers superior sentiment analysis capabilities on the SemEval Laptop dataset, outperforming existing state-of-the-art models in terms of accuracy and overall sentiment classification performance.

Further, the results for the SemEval 2014 Task 4 Restaurant dataset compared with the existing approaches and MOL-ABSA is presented in Fig. 7. From Fig. 7, it is seen that MWGCN achieves an accuracy of 86.36% and a respectable Macro-F-Score of 80.54%, showcasing its ability to accurately classify sentiments on the SemEval Restaurant dataset. MultiGCN follows closely with an accuracy of 83.82% and a macro-f-score of 77.1%, indicating competitive performance in sentiment analysis. The CDM-CDW model exhibits an accuracy of 80.59% but achieves a high macro-f-score of 80.39%, highlighting its effectiveness in accurately capturing

Received: June 23, 2024. Revised: August 4, 2024. 

Table 4. Comparative Study

| Models | Laptop | | Reataurant | | Hospital | |
|---|---|---|---|---|---|---|
| | Accuracy | Macro-F-Score | Accuracy | Macro-F-Score | Accuracy | Macro-F-Score |
| MWGCN, 2023[20] | 79.78 | 76.68 | 86.36 | 80.54 | - | - |
| MultiGCN, 2023[21] | 78.8 | 74.97 | 83.82 | 77.1 | - | - |
| CDM-CDW, 2023[22] | 80 | 79.81 | 80.59 | 80.39 | - | - |
| SDTGCN, 2024[23] | 78.64 | 75.50 | 83.32 | 76.13 | - | - |
| MOL-ABSA(Proposed) | 81.19 | 78.52 | 85.26 | 79.27 | 87.53 | 85.15 |

sentiments on this dataset. SDTGCN achieves an accuracy of 83.32% and a macro-f-score of 76.13%, demonstrating robust sentiment analysis capabilities. The MOL-ABSA achieves an accuracy of 85.26% and a commendable Macro-F-Score of 79.27%, showcasing its competitive performance in sentiment classification on the SemEval Restaurant dataset. Although not the highest in accuracy, the MOL-ABSA model maintains a strong macro-f-score, indicating its ability to handle sentiment analysis tasks effectively and capture sentiment nuances across different aspects of restaurant reviews. Furthermore, when assessed using the hospital reviews dataset, the MOL-ABSA attained an accuracy of 87.14% and a macro-f-score of 85.15%. The comprehensive study results are presented in Table 4.

### 4.5 Discussion

The results of the sentiment analysis model, as reflected in Figs. 6-7 and Table 4, reveal interesting insights into their performance on different datasets. Starting with Table 3, which evaluates the models on the SemEval Laptop dataset, we observe varying levels of accuracy and macro-f-score across different models. MWGCN, MultiGCN, and SDTGCN exhibit competitive performance with accuracy scores ranging from 78.64% to 79.78% and macro-f-score ranging from 74.97% to 76.68%. These models demonstrate their effectiveness in sentiment classification tasks with slight variations in performance metrics. The CDM-CDW model stands out with an accuracy of 80% and a high macro-f-score of 79.81%, showcasing its robustness in capturing sentiments and accurately classifying sentiments on the SemEval Laptop dataset. However, the MOL-ABSA surpasses all other models with an accuracy of 81.19% and an impressive macro-f-score of 78.52%. This indicates that the MOL-ABSA model offers superior sentiment analysis capabilities, outperforming existing state-of-the-art models in terms of accuracy and overall sentiment classification performance on the SemEval Laptop dataset. From Table 4, which evaluates the models on the hospital reviews dataset, we see a notable improvement in the performance of the proposed MOL-ABSA model. With an accuracy of 87.14% and a macro-F-score of 85.15%, the model demonstrates its ability to handle sentiment analysis tasks effectively, particularly in the domain of hospital reviews. The higher accuracy and macro-f-score achieved by the MOL-ABSAmodel on the hospital reviews dataset compared to the SemEval datasets suggest that the model is well-suited for sentiment analysis tasks in specific domains. It effectively captures sentiments and accurately classifies sentiments, making it a promising choice for sentiment analysis applications, especially in domains like healthcare and customer reviews. The results highlight the importance of selecting appropriate models designed for specific datasets and domains for optimal sentiment analysis performance. The MOL-ABSA model emerges as a robust and effective solution for sentiment analysis tasks, showcasing its versatility and accuracy across different datasets and domains.

## 5. Conclusion

The architecture presents a systematic and comprehensive approach to sentiment analysis, beginning with meticulous preprocessing of the dataset to ensure its quality and readiness for analysis. The preprocessing steps, including handling null values, converting text to lowercase, removing punctuation and stopwords, lemmatization, and PoS tagging, play a crucial role in refining the textual data and preparing it for the subsequent analytical stages.

The utilization of MOL-ABSA as the core analytical model further enhances the architecture by extracting relevant aspects and thematic structures from the pre-processed dataset. This extracted information serves as valuable input for sentiment analysis, where O-BERT is employed to classify sentiment polarity and assign scores to positive, negative, and neutral sentiments. The integration of LSTM networks enhances the sentiment classification process, leveraging sequential data analysis to improve accuracy and capture sentiments. The evaluation of sentiment polarity classification using metrics like accuracy and macro-f-score ensures the robustness and effectiveness of the architecture in accurately discerning and categorizing sentiments within textual data. Overall, this architecture provides a structured and efficient framework for sentiment analysis, capable of handling diverse generalized datasets and delivering insightful sentiment insights for various applications across industries. For the future work, different datasets can be used for further improving the proposed MOL-ABSA model.

## Conflicts of Interest

The authors declare no conflict of interest.

## Author Contributions

Conceptualization, Methodology, Nasreen Taj M B and Girisha G S; software, validation,formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, Nasreen Taj M B; visualization, supervision, project administration, Girisha G S.

## Acknowledgments

This work was supported by Dayananda Sagar University and GMIT.

## References


[1] Georgios Kontonatsios *et al.*, "FABSA: An aspect-based sentiment analysis dataset of user reviews", *Neurocomputing*, Vol. 562, pp. 126867-126867, 2023, doi: 10.1016/j.neucom.2023.126867.

[2] S. Datta and S. Chakrabarti, "Aspect based sentiment analysis for demonetization tweets by optimized recurrent neural network using fire fly-oriented multi-verse optimizer", *Sādhanā*, Vol. 46, No. 2, 2021, doi: 10.1007/s12046-021-01608-1.

[3] G. Kaur and A. Sharma, "HAS: Hybrid Analysis of Sentiments for the perspective of customer review summarization", *Journal of Ambient Intelligence and Humanized Computing*, 2022, doi: 10.1007/s12652-022-03748-6.

[4] H. Taherdoost and M. Madanchian, "Artificial Intelligence and Sentiment Analysis: A Review in Competitive Research", *Computers*, Vol. 12, No. 2, p. 37, 2023, doi: 10.3390/computers12020037.

[5] N. Jeong and J. Lee, "An Aspect-Based Review Analysis Using ChatGPT for the Exploration of Hotel Service Failures", *Sustainability*, Vol. 16, No. 4, pp. 1640-1640, 2024, doi: 10.3390/su16041640.

[6] Yan Cathy Hua, P. Denny, Katerina Taškova, and J. Wicker, "A Systematic Review of Aspect-based Sentiment Analysis (ABSA): Domains, Methods, and Trends", *arXiv (Cornell University)*, 2023, doi: 10.48550/arxiv.2311.10777.

[7] G. D'Aniello, M. Gaeta, and I. La Rocca, "KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis", *Artificial Intelligence Review*, 2022, doi: 10.1007/s10462-021-10134-9.

[8] R. K. Yadav, L. Jiao, M. Goodwin, and O.-C. Granmo, "Positionless aspect based sentiment analysis using attention mechanism", *Knowledge-Based Systems*, Vol. 226, p. 107136, 2021, doi: 10.1016/j.knosys.2021.107136.

[9] Shraddha Kalbhor and D. Goyal, "Survey on ABSA based on machine learning, deep learning and transfer learning approach", *AIP Conference Proceedings*, 2023, doi: 10.1063/5.0154549.

[10] B. Harjo, Muljono, R. Abdullah, "Attention-based Sentence Extraction for Aspect-based Sentiment Analysis with Implicit Aspect Cases in Hotel Review Using Machine Learning Algorithm, Semantic Similarity, and BERT", *International Journal of Intelligent Engineering and Systems*, Vol. 16, No. 3, pp. 189-200, 2023, doi: https://doi.org/10.22266/ijies2023.0630.15.

[11] K. W. Trisna and H. J. Jie, "Deep Learning Approach for Aspect-Based Sentiment Classification: A Comparative Review", *Applied Artificial Intelligence*, pp. 1-37, 2022, doi: 10.1080/08839514.2021.2014186.

[12] H. H. Do, P. Prasad, A. Maag, and A. Alsadoon, "Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review", *Expert Systems with Applications*, Vol. 118, pp. 272-299, 2019, doi: 10.1016/j.eswa.2018.10.003.

[13] W. G. Hatcher and W. Yu, "A Survey of Deep Learning: Platforms, Applications and Emerging Research Trends", *IEEE Access*, Vol. 6, pp. 24411-24432, 2018, doi: 10.1109/ACCESS.2018.2830661.

[14] S. Razavi, "Deep Learning, Explained: Fundamentals, Explainability, and Bridgeability to Process-based Modelling", *Environmental Modelling & Software*, p. 105159, 2021, doi: 10.1016/j.envsoft.2021.105159.

[15] J. Sun, P. Han, Z. Cheng, E. Wu, and W. Wang, "Transformer Based Multi-Grained Attention Network for Aspect-Based Sentiment Analysis", *IEEE Access*, Vol. 8, pp. 211152-211163, 2020, doi: 10.1109/access.2020.3039470.

[16] A. Bansal and N. Kumar, "Aspect-Based Sentiment Analysis Using Attribute Extraction of Hospital Reviews", *New Generation Computing*, 2021, doi: 10.1007/s00354-021-00141-3.

[17] P. Han, Ye Dongyu, G. Liang, W. Chen, H. Wei, and X. Li, "Generative Aspect Based Sentiment Analysis for Online Health Community with Data Augmentation", *Non-reviewed preprint research paper*, 2023, doi: 10.2139/ssrn.4489332.

[18] K. Denecke and D. Reichenpfader, "Sentiment analysis of clinical narratives: A scoping review", *Journal of Biomedical Informatics*, Vol. 140, p. 104336, 2023, doi: 10.1016/j.jbi.2023.104336.

[19] C. Golz, S. Aarts, C. Hacking, S. Hahn, and S.M.G. Zwakhalen, "Health professionals' sentiments towards implemented information technologies in psychiatric hospitals: a text-mining analysis", *BMC Health Services Research*, Vol. 22, No. 1, 2022, doi: 10.1186/s12913-022-08823-4.

[20] B. Yu and S. Zhang, "A novel weight-oriented graph convolutional network for aspect-based sentiment analysis", *The Journal of Supercomputing*, Vol. 79, No. 1, pp. 947-972, 2022, doi: 10.1007/s11227-022-04689-9.

[21] Y. Ma, R. Song, X. Gu, Q. Shen, and H. Xu, "Multiple graph convolutional networks for aspect-based sentiment analysis", *Applied Intelligence*, Vol. 53, No. 10, pp. 12985-12998, 2022, doi: 10.1007/s10489-022-04023-z.

[22] T. Lin and I. Joe, "An Adaptive Masked Attention Mechanism to Act on the Local Text in a Global Context for Aspect-Based Sentiment Analysis", *IEEE Access*, Vol. 11, pp. 43055-43066, 2023, doi: 10.1109/ACCESS.2023.3270927.

[23] Q. Zhao, F. Yang, D. An, and J. Lian, "Modeling Structured Dependency Tree with Graph Convolutional Networks for Aspect-Level Sentiment Classification", *Sensors*, Vol. 24, No. 2, p. 418, 2024, doi: 10.3390/s24020418.

[24] C. Y. Sy, L. L. Maceda, T. D. Palaoag and M. B. Abisado, "Transforming Education Policy: Evaluating UAQTE Program Implementation Through LDA, BoW and TF-IDF Techniques", In: *Proc. of 2024 26th International Conference on Advanced Communications Technology (ICACT)*, Pyeong Chang, Korea, Republic of, pp. 01-07, 2024, doi: 10.23919/ICACT60172.2024.10471958.

[25] S. Syed and M. Spruit, "Full-text or abstract? Examining topic coherence scores using latent dirichlet allocation", In: *Proc. of 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE*, 2017. doi: 10.1109/dsaa.2017.61.

[26] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", *arXiv.org*, 2019, doi: 10.48550/arXiv.1810.04805

[27] R. Fu, M. A. Al-Absi, K. -H. Kim, Y. -S. Lee, A. A. Al-Absi and H. -J. Lee, "Deep Learning-Based Drone Classification Using Radar Cross Section Signatures at mmWaveFrequencies", *IEEE Access*, Vol. 9, pp. 161431-161444, 2021, doi: 10.1109/ACCESS.2021.3115805.

[28] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis", *ACLWeb*, 2014, doi: 10.3115/v1/S14-2004.

[29] Practo, "Practo | Book Doctor Appointments Online, Order Medicine, Diagnostic Tests, Consult", *Practo.com*, 2017. https://www.practo.com/

[30] "Consumer reviews on Movies, Cars, Bikes, Mobile Phones, Music, Books, Airlines, Restaurants, Hotels & more - MouthShut.com", www.mouthshut.com. https://www.mouthshut.com/ (accessed Feb. 13, 2024).