Aspect based multimodal sentiment analysis of product reviews using deep learning techniques

Anitha Padigapati, A.V. Praveen Krishna
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India

Article Info

Article history:

Received Sep 30, 2024
Revised Jul 8, 2025
Accepted Nov 23, 2025

Keywords:

Aspect sentiment analysis
Deep learning
Emojis
Multimodal data
Natural language processing
Sentiment analysis

ABSTRACT

Sentiment analysis plays a crucial role in understanding customer opinions, particularly in product reviews. Traditional approaches primarily focus on textual data; however, with the rise of social media, incorporating multimodal data, including text and emojis, enhances sentiment analysis accuracy. This research introduces a multimodal aspect-based sentiment analysis (MABSA) framework, integrating textual and emoji representations for Samsung M21 product reviews from Flipkart. The methodology involves data preprocessing, aspect extraction, sentiment grouping, and feature extraction using deep learning (DL) techniques. Bidirectional long short-term memory (Bi-LSTM) networks are employed for classification, leveraging Word2Vec, Emoji2Vec, and bidirectional encoder representations from transformers (BERT) embeddings. Experimental results show that BERT with Bi-LSTM outperforms Word2Vec with Bi-LSTM, achieving 95.6% accuracy in aspect prediction and 96.28% accuracy in sentiment classification. Comparative analysis with existing models highlights the superiority of the MASAT model, effectively integrating implicit aspects, emoticons, and emojis. The study demonstrates the importance of multimodal sentiment analysis for a more comprehensive understanding of user opinions, offering valuable insights for businesses to enhance customer satisfaction.

Corresponding Author:

Anitha Padigapati
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation
Vaddeswaram, Andhra Pradesh, India
Email: anitha.padigapati@gmail.com

1. INTRODUCTION

Sentiment analysis, a crucial subfield of natural language processing (NLP), aims to determine the sentiment or opinion expressed in a piece of text. Traditionally, sentiment analysis has primarily focused on text data, leveraging various machine learning (ML) and deep learning (DL) techniques to classify sentiments as positive, negative, or neutral [1]. However, with the proliferation of social media and the advent of emojis, there is an increasing need to incorporate multimodal data for a more comprehensive sentiment analysis. Emojis, being pictorial representations of emotions and expressions, complement textual data and can significantly enhance the accuracy and depth of sentiment analysis [2].

Aspect-based sentiment analysis (ABSA) delves deeper than general sentiment analysis by identifying and categorizing sentiments related to specific aspects of entities. This granularity is particularly beneficial for mobile product reviews, where consumers often discuss various aspects such as battery life, user interface, camera quality, and more [3]. By integrating ABSA with multimodal data, we can achieve a more nuanced understanding of user sentiments. This integration is essential as users frequently employ emojis to reinforce or nuance their textual expressions, thus providing additional layers of meaning that purely text-based analysis might miss [4]. Recent advancements in multimodal sentiment analysis highlight the potential of combining text and emoji data to improve sentiment classification. Studies have shown that incorporating emojis can enhance the performance of sentiment analysis models, making them more robust and accurate in capturing the users' true sentiments [5]. For instance, an emoji like 😊 can alter the sentiment of a sentence significantly, adding a layer of positive emotion that a purely text-based analysis might overlook. Therefore, a multimodal approach that seamlessly integrates textual and emoji data is indispensable for capturing the full spectrum of user sentiment in mobile product reviews [6].

This study not only contributes to the growing body of research on multimodal sentiment analysis but also offers practical implications for businesses seeking to understand and improve customer satisfaction based on detailed sentiment insights [7]. Furthermore, many existing models struggle to capture implicit sentiment expressed through emojis, limiting their effectiveness in real-world applications. This research aims to bridge this gap by introducing a multimodal framework that combines text and emoji representations for ABSA. Despite the progress made in sentiment analysis, several challenges remain unaddressed. Most existing models lack the ability to fully leverage emoji data to refine sentiment classification, often treating emojis as simple text tokens rather than distinct sentiment-bearing elements. Additionally, there is a lack of comprehensive datasets that contain both text and emoji-based sentiment annotations, making it difficult to evaluate multimodal models effectively. Another key limitation in current methodologies is their inability to associate sentiment with specific product aspects, leading to generic sentiment classification rather than aspect-based insights. Addressing these issues is crucial for improving sentiment analysis applications, particularly in mobile product reviews where users frequently combine text and emojis to express their opinions.

To overcome these challenges, this paper proposes a novel multimodal aspect-based sentiment analysis (MABSA) framework that integrates textual and emoji data for more precise sentiment classification. The proposed approach employs DL techniques, specifically bidirectional long short-term memory (Bi-LSTM) networks, to extract features from both text and emoji modalities. The model is trained and evaluated using a dataset of mobile product reviews that includes both textual and emoji-based sentiment expressions. The following sections of this paper will present a comprehensive review of existing literature, a detailed methodology outlining the experimental setup, a discussion of the obtained results, and a conclusion highlighting key findings and future research directions.

2. RELATED WORK

ABSA has gained prominence for its ability to deliver detailed insights into user opinions on specific aspects of products or services. Recent research has focused on advancing methodologies and models for ABSA, such as Alturaief *et al.* [8] AWARE dataset for app reviews and Arumugam and Nallaperumal [9] EIAASG, a graph convolutional network (GCN) incorporating emotional aspects. Graph-based approaches, like those of Chen *et al.* [10] and Dai *et al.* [11], have proven effective for improving sentiment analysis accuracy. Moreover, De Greve *et al.* [12] applied ABSA to German literary reviews, showcasing the method's versatility. Advanced models, including bidirectional encoder representations from transformers (BERT) [13] and selective attention-based GCNs [14], [15] have further enhanced ABSA by capturing contextual nuances and integrating syntactic structures [16]. However, challenges such as limited applicability across languages and domains with specialized vocabularies, along with the need for standardized evaluation benchmarks, remain [8], [13].

Jangid *et al.* [17] introduced a DL approach for aspect-based financial sentiment analysis, demonstrating the effectiveness of domain-specific adaptations in capturing nuanced sentiments. Karimi *et al.* [18] enhanced BERT for ABSA, improving performance in diverse domains. Liang *et al.* [19] extended ABSA with affective knowledge-enhanced GCNs, while Liang *et al.* [20] employed few-shot learning for adapting to new aspect categories. Mewada and Dewang [21] developed SA-ASBA, a hybrid model combining BERT and extreme gradient boosting for improved sentiment classification. Ma *et al.* [22] utilized multiple GCNs to capture aspect-sentiment relationships, and Qi *et al.* [23] proposed aspect-sensitive word embeddings to refine sentiment analysis. Alqaryouti *et al.* [24] applied ABSA to smart government reviews, highlighting its versatility in public service analysis.

Li *et al.* [25] introduced a part-of-speech (POS)-based label update network for aspect sentiment triplet extraction, improving aspect term extraction precision by incorporating syntactic information. Ahmed *et al.* [26] emphasized systematic approaches to handle linguistic variations in ABSA, while Chen and Qian [27] refined aspect identification using soft prototypes for better adaptability. Devlin *et al.* [28] revolutionized ABSA with BERT, a transformer-based model that set a new benchmark for language understanding. Wang *et al.* [29] addressed post-processing errors in aspect term extraction, enhancing model accuracy, while Yuan *et al.* [30] integrated syntactic dependencies into transformers for improved sentiment analysis. Zhu *et al.* [31] leveraged rich structural features for ABSA, though challenges remain, particularly in integrating multimodal data sources like text and emojis.

This gap necessitates further research into developing hybrid models that can effectively fuse textual and emoji modalities for more holistic sentiment analysis in mobile product reviews. The integration of text and emojis poses unique challenges in ABSA, requiring novel methodologies to effectively capture and analyze multimodal sentiment expressions. The objective of this work is to develop a MABSA framework that integrates textual and emoji data for enhanced sentiment understanding in mobile product reviews.

3. PROPOSED WORK

The Figure 1 presents a comprehensive framework for MABSA using text and emojis for mobile product reviews. Below is a detailed description of each component along with the relevant mathematical formulas to support the processes.

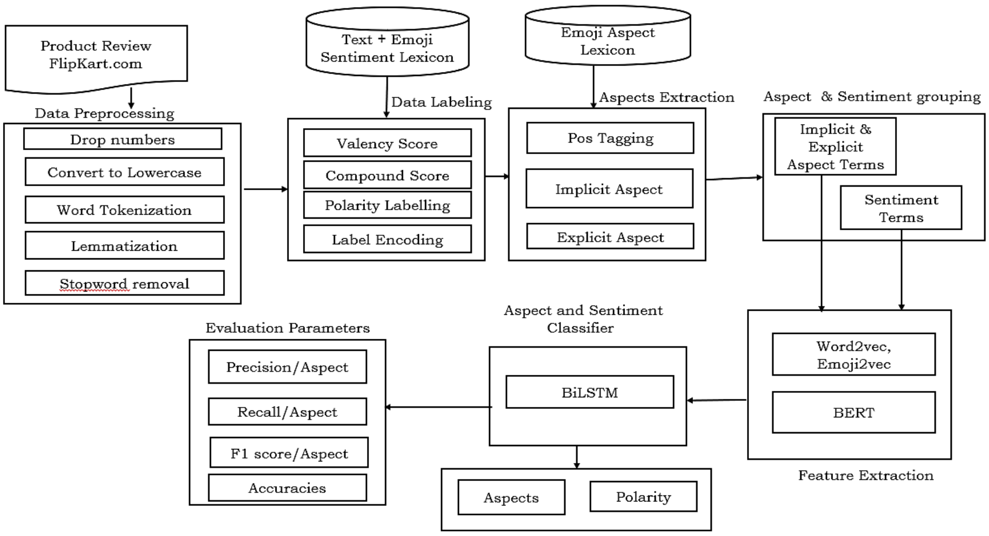

Figure 1. Proposed framework for MABSA

3.1. Dataset

The model uses input data comprising customer reviews from Flipkart for the Samsung product M21 smartphone. These reviews, which include both text and emojis, serve as the primary dataset. Customer reviews are essentially feedback provided by user's post-purchase on an online marketplace, and they can reflect positive, negative, or neutral sentiments. A total of 9,003 reviews were collected for this analysis.

3.2. Data preprocessing

The data preprocessing stage is crucial for preparing raw product reviews for sentiment analysis. The reviews are sourced from Flipkart.com, a popular e-commerce platform. The preprocessing steps involve several transformations as (i) dropping numbers, (ii) converting text to lowercase, (iii) tokenizing words, (iv) lemmatizing words, and (v) removing stopwords. These steps are designed to clean and standardize the text, ensuring that the subsequent analysis is based on meaningful and consistent data. For instance, removing numbers and stopwords eliminates irrelevant information, while lemmatization reduces words to their base forms, aiding in more accurate sentiment analysis. The preprocessing steps involve several transformations to clean and standardize the text, ensuring that the subsequent analysis is based on meaningful and consistent data.

3.3. Data labeling

After preprocessing, the next step is data labeling, which involves assigning sentiment scores to the text and emojis in the reviews. This process utilizes a sentiment lexicon that includes both text and emoji sentiments. The data labeling step calculates several scores, such as valency score, compound score, and polarity labeling. The valency score measures the sentiment intensity of individual words or emojis, while the compound score provides an overall sentiment score for the review. Polarity labeling then categorizes these scores into positive, negative, or neutral sentiments. Finally, label encoding transforms these categorical labels into numerical values suitable for machine and DL models. This step ensures that the sentiment information is quantitatively represented, facilitating further analysis.

3.4. Aspects extraction

Aspects extraction focuses on identifying specific features or components of the product that are mentioned in the reviews. This step uses techniques such as POS tagging to identify nouns and other relevant parts of speech that typically represent product aspects. Both explicit and implicit aspects are extracted, explicit aspects are directly mentioned in the text, while implicit aspects are inferred from the context. For example, in a review stating "The battery life is amazing," "battery life" is an explicit aspect. This dual approach ensures a comprehensive identification of product features discussed in the reviews. The extracted aspects are crucial for linking sentiments to specific product attributes, allowing for a more nuanced sentiment analysis.

Let's denote the set of review sentences as $S = \{s1, s2, \dots, sk\}$. For each sentence s, POS tagging provides a tagged set,

$$POS(s) = \{(w_1, p_1), (w_2, p_2), \dots, (w_n, p_n)\} \quad (1)$$

where w_i represents a word and p_i represents its corresponding POS tag. Nouns (NN), noun phrases (NP), and adjectives (JJ) are typically extracted as potential aspects A,

$$A(s) = \{w_i \mid p_i \in \{NN, NP, JJ\}\} \quad (2)$$

explicit aspects are directly mentioned in the text. For example, in the review "The battery life is amazing," "battery life" is identified as an explicit aspect. Formally, explicit aspects $A_{\ explicit}$ can be defined as,

$$A_{\ explicit} = \{a \mid a \in A(s)\ and\ is\ directly\ mentioned\ in\ the\ text\} \quad (3)$$

implicit aspects, on the other hand, are inferred from the context. For instance, if a review says, "It lasts all day," the aspect "battery life" is inferred. Implicit aspects $A_{\ implicit}$ can be defined as,

$$A_{\ implicit} = \{a \mid a \in A(s)\ and\ inferred\ from\ context\} \quad (4)$$

for emojis, let $E = \{e_1, e_2, \dots, e_m\}$ represent the set of emojis in the review. Emojis are mapped to aspects using an emoji lexicon that links emojis to product features. The aspect extraction from emojis A_e is represented as,

$$A_e = \{map(e) \mid e \in E\} \quad (5)$$

where $map(e)$ returns the product aspect associated with emoji e. This combined approach ensures comprehensive identification of product features discussed in the reviews,

$$A_{total} = A_{explicit} \cup A_{implicit} \cup A_{\ e} \quad (6)$$

the extracted aspects A_{total} are crucial for linking sentiments to specific product attributes, allowing for a more nuanced sentiment analysis. By incorporating both text and emoji data, the analysis captures a fuller range of user sentiments and product features.

3.5. Aspect and sentiment grouping

Once the aspects are extracted, the next step is to group these aspects with their corresponding sentiments. This involves associating each aspect term A with the sentiment terms S identified in the review. The process ensures that both implicit and explicit aspect terms are considered, and they are paired with the sentiment terms derived from both text and emojis.

For each review, let $A(d) = \{a_1, a_2, \dots, a_m\}$ be the set of extracted aspect terms, and let $S(d) = \{s_1, s_2, \dots, s_n\}$ be the set of sentiment terms. The sentiment score for each aspect a_i is computed by averaging the sentiment scores of the sentiment terms associated with a_i. If $V(s)$ represents the valency score for a sentiment term, then the sentiment score $S(a_i)$ for aspect a_i is given by:

$$S(a_i) = \frac{1}{|S(a_i)|} \sum_{s_j \in S(a_i)} V(s_j) \quad (7)$$

where $S(a_i) \subseteq S(d)$ is the subset of sentiment terms associated with aspect a_i. This calculation is performed separately for text and emojis.

For text-based sentiment terms S_{text}.

$$S_{text}\ (a_i) = \frac{1}{|S_{text}\ (a_i)|} \sum_{s_j\ \in S_{text}\ (a_i)} V_{text}(s_j) \tag{8}$$

For emoji-based sentiment terms S_{emoji}.

$$S_{emoji}\ (a_i) = \frac{1}{|S_{emoji}\ (a_i)|} \sum_{s_j\ \in S_{emoji}\ (a_i)} V_{emoji}(s_j) \tag{9}$$

The grouped aspect-sentiment pairs provide a detailed mapping of how specific product features are perceived by users. This step integrates the different components of the review, creating a holistic view of user sentiments towards various product aspects. The outcome is a set of aspect-sentiment pairs that form the basis for further analysis and classification. Each aspect a_i is associated with a sentiment score $S(a_i)$ that combines both text and emoji sentiments,

$$S(a_i)\ = \alpha S_{text}(a_i) + \beta S_{emoji}(a_i) \tag{10}$$

where α and β are weights that balance the contributions of text and emoji sentiments, respectively. This integrated sentiment score ensures a comprehensive understanding of user opinions, aiding in more accurate sentiment analysis and classification.

3.6. Feature extraction

The next step in the process involves feature extraction. Feature extraction transforms the labeled and grouped data into numerical representations that can be fed into DL models. This step employs techniques such as Word2Vec, Emoji2Vec, and BERT to generate embeddings for text and emojis, capturing their semantic meanings and relationships.

3.6.1. Text feature extraction

For text data, Word2Vec is used to convert words into vectors. Let $T = \{t_1\ , t_2, \ldots , t_n\}$ represent the tokenized words from a review. Each word t_i is transformed into a vector v_{t_i} using the Word2Vec model.

$$v_{t_i}\ = Word2Vec(t_i) \tag{11}$$

The review R can then be represented as a matrix V_R.

$$V_R\ = [v_{t_1}\ , v_{t_2}\ , \ldots , v_{t_n}\ ] \tag{12}$$

Similarly, BERT embeddings capture contextual relationships between words. For a sequence of tokens, BERT generates contextual embeddings B_{t_i} for each token t_i.

$$B_{t_i} = BERT(t_i \mid T) \tag{13}$$

The entire review R can be represented as a matrix B_R.

$$B_R\ = [B_{t_1}\ , B_{t_2}\ , \ldots , B_{t_n}\ ] \tag{14}$$

3.6.2. Emoji feature extraction

For emoji data, Emoji2Vec is used to create vector representations. Let $E = \{e_1\ , e_2, \ldots , e_n\}$ denote the set of emojis in a review. Each emoji e_1 is converted into a vector v_{e_i} using the Emoji2Vec model.

$$v_{e_i} = Emoji2Vec(e_i\ ) \tag{15}$$

The review R can then be represented as a matrix V_E.

$$V_E\ = [v_{e_1}\ , v_{e_2}\ , \ldots , v_{e_n}\ ] \tag{16}$$

3.6.3. Combined feature representation

To form a comprehensive representation of the review that includes both text and emoji embeddings, the text and emoji matrices are concatenated. If V_R represents the text embedding matrix and V_E represents the emoji embedding matrix, the combined feature matrix can be F_R expressed as $F_R = [V_R\ + V_E\ ]$, where "+" denotes concatenation.

3.7. Classification

In this proposed work, classifiers such as Bi-LSTM networks are employed to predict the aspects and associated sentiments based on the extracted features from both text and emojis. The Bi-LSTM model processes sequences in both forward and backward directions, capturing contextual dependencies more effectively than unidirectional LSTMs. For text, embeddings like Word2Vec are used to transform words into continuous vector representations. Let w_i be the word vector for the $i-th$ word in a review. The review d consisting of n words is represented as a sequence of word vectors $d = [w_1, w_2, \dots, w_n]$. Similarly, emojis are represented using Emoji2Vec embeddings. Let e_i be the emoji vector for the $i-th$ emoji in a review. If a review contains m emojis, it can be represented as $E = [e_1, e_2, \dots, e_n]$.

3.7.1. Bi-LSTM processing

The Bi-LSTM processes the concatenated sequence of text and emoji embeddings. Let x_t be the combined feature vector at time step, consisting of both word and emoji vectors $x_t = w_t \oplus e_t$, where $\oplus$ denotes concatenation. The Bi-LSTM maintains two sets of hidden states forward $(\overrightarrow{h_t})$ and backward $(\overleftarrow{h_t})$. The forward pass updates the forward hidden state is represented as $\overrightarrow{h_t} = LSTM(x_t, \overrightarrow{h_{t-1}})$. The backward pass updates the backward hidden state is represented $\overleftarrow{h_t} = LSTM(x_t, \overleftarrow{h_{t+1}})$. The final hidden state at each time step is the concatenation of the forward and backward states $h_t = \overrightarrow{h_t} \oplus \overleftarrow{h_t}$.

3.7.2. Aspect and sentiment prediction

The concatenated hidden states are then used to predict aspects and sentiments. Let h_T be the final hidden state after processing the entire sequence. A fully connected layer with a SoftMax activation function is applied to predict the probability distribution over the aspect labels A.

$$P(A \mid h_T) = softmax(W_A h_T + b_A) \tag{17}$$

Where W_A and b_A are the weights and biases for the aspect classification layer.

Similarly, for sentiment prediction, another fully connected layer with a SoftMax activation function is used to predict the probability distribution over sentiment labels S,

$$P(S \mid h_T) = softmax(W_S h_T + b_S) \tag{18}$$

where W_S and b_S are the weights and biases for the sentiment classification layer.

3.7.3. Loss function

The model is trained using cross-entropy loss for both aspect and sentiment classification. The total loss L is the sum of the aspect loss L_A and sentiment loss L_S.

$$L = L_A + L_S \tag{19}$$

The aspect loss L_A is defined as,

$$L_A = -\sum_{i=1}^{N} log\, P(A_i \mid h_T) \tag{20}$$

the sentiment loss L_S is defined as,

$$L_S = -\sum_{i=1}^{N} log\, P(S_i \mid h_T) \tag{21}$$

where N is the number of training samples. This combined loss function ensures that the model learns to predict both aspects and sentiments effectively.

3.8. Evaluation parameters

The final output consists of the multiclass confusion matrix and classification report, which display the precision $P(A_i)$, recall $R(A_i)$, F1-score $F1S(As_i)$, and accuracy $A(A_i)$. The formulae for each aspect A_i are given as follows;

$$P(A_i) = \frac{TP_{a_i}}{TP_{a_i} + FP_{a_i}} \tag{22}$$

$$R(A_i) = \frac{TP_{a_i}}{TP_{a_i} + FN_{a_i}} \tag{23}$$

$$F1S(As_i) = 2 * \frac{P(A_i) * R(A_i)}{P(A_i) + R(A_i)} \tag{24}$$

$$A(A_i) = \frac{FPa_i + FNa_i}{TPa_i + TNa_i + FPa_i + FNa_i} \tag{25}$$

4. RESULTS AND DISCUSSION

This section outlines the experiments conducted during the implementation phase. The system used had an 11th Gen i9 processor, 32 GB DDR4 RAM, 1 TB HDD, 256 GB SSD, and ran Windows 11 (64-bit). The proposed ABSA framework integrates textual and emoji data for improved sentiment accuracy. Using Python, the framework handles data preprocessing, feature extraction, model training, and evaluation. For the experiment, 9,003 product reviews from Flipkart.com were analyzed. Preprocessing involved removing numbers, converting text to lowercase, tokenizing words, lemmatization, and stopword removal using the nltk package.

POS tagging is performed using the spacy library (spacy.load('en_core_web_sm')) to identify explicit and implicit aspects through POS tags and dependency parsing. Figures 2 and 3 illustrate practical applications, tagging words and emojis in a sample review like "Camera quality is 😒 and bad screen quality," where each element is classified (e.g., NOUNs, ADJectives), including emojis as sentiment indicators. The dependency parser reveals relationships, such as the adjectival modifier (amod) role of 😒. Figure 4 demonstrates similar tagging for emoji-based reviews like "📷😃," identifying the camera as a NOUN and the happy face as an ADJective. In Figure 5, the review "Good🔋📷" shows positive sentiment towards the battery emoji (amod) and punctuation dependency (punct) on the camera emoji.

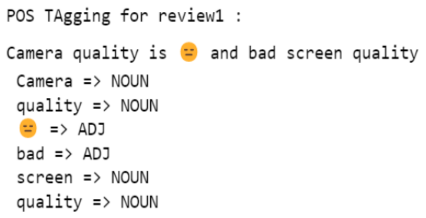

Figure 2. POS tagging for review 1 with text and emoji

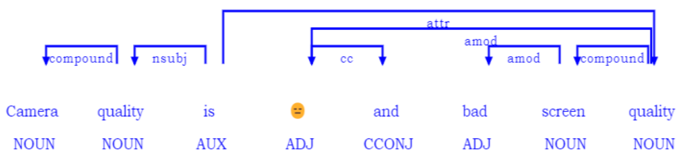

Figure 3. Dependency parser of POS tagging for review 1

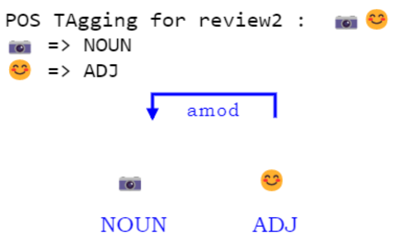

Figure 4. POS tagging and dependency parser for review2 with aspect and sentiment as emoji's

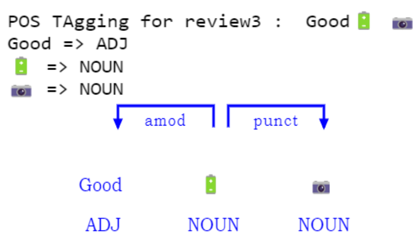

Figure 5. POS tagging and dependency parser for review3 only with aspect as emoji and sentiment as text

The extraction of aspects and sentiments involves identifying both implicit and explicit aspect terms. The Figure 6 demonstrates a key-pair like structure representing different aspects of a product, such as "camera quality," "delivery," and "smartphone,". Each aspect is associated with a description that includes both text and emojis, indicating the sentiment expressed by the user. For instance, "camera quality" is paired with a 😒 emoji, indicating a negative sentiment, while "delivery" is described as "good," suggesting a positive sentiment. The set of aspects and descriptions comprises both implicit and explicit aspects, organized into various categories. For the purposes of experimentation, we have focused on the ten most frequently occurring aspects. Top ten aspects are chosen from a total of 386 "phone", "battery", "camera", "charge", "display", "price", "weight", "sound", "processor", "touch".

[{'aspect': 'camera quality', 'description': '😔'},
{'aspect': 'delivery', 'description': 'good'},
{'aspect': 'smartphone', 'description': '😔'},
{'aspect': 'touch', 'description': 'awesome'},
{'aspect': ' 📱 ', 'description': '❤️'},
{'aspect': 'mobile', 'description': 'perfect'},
{'aspect': 'charger', 'description': 'willing'},
{'aspect': 'price', 'description': 'good'},
{'aspect': 'backup', 'description': 'best'},
{'aspect': 'colors', 'description': 'very dull'},
{'aspect': 'screen quality', 'description': 'bad'},
{'aspect': 'cameralens', 'description': 'superb'},
{'aspect': ' 📷', 'description': 'okay'},
{'aspect': 'panel', 'description': ' 👌 '},
{'aspect': ' 🔋 ', 'description': 'better'},
{'aspect': 'performance', 'description': 'good'},
{'aspect': '🔉', 'description': 'good'}]

Figure 6. Aspect terms and descriptive term extraction

Feature extraction combines textual and emoji data to create a rich representation of the reviews. The proposed work uses Word2Vec and Emoji2Vec for embedding textual and emoji data, respectively, and BERT for contextual embeddings. The gensim package is used to create word embeddings (gensim.models.Word2Vec) and emoji embeddings. The transformers library from hugging face provides pre-trained BERT models for generating embeddings (transformers.BertModel and transformers.BertTokenizer). Word2Vec, trained with a window size of 5, a vector dimension of 300, and a minimum word frequency of 5, generates dense vector representations capturing semantic relationships between words. BERT embeddings are extracted using the pre-trained BERT model with 12 layers, 768 hidden units, and a maximum sequence length of 128 tokens, leveraging its bidirectional context to capture deep contextual information. For classification, 80% of the reviews are considered for training and 20% of the reviews are taken for testing. A Bi-LSTM network is employed with 100 hidden units in each direction and a dropout rate of 0.5 to prevent overfitting. The Bi-LSTM processes the sequence embeddings to classify sentiment based on the learned context and aspect-specific features, ensuring robust and nuanced sentiment predictions. Lastly the precision, recall, f1_score, and accuracy are used to evaluate the performance of the models. The Figures 7 and 8 illustrate the performance comparison between two feature extraction methods, Word2Vec and BERT, when paired with a Bi-LSTM classifier for ABSA. The Figure 7 focuses on Word2Vec with Bi-LSTM, presenting both the confusion matrix and evaluation metrics in Table 1, while the Figure 8 details the same for BERT with Bi-LSTM.

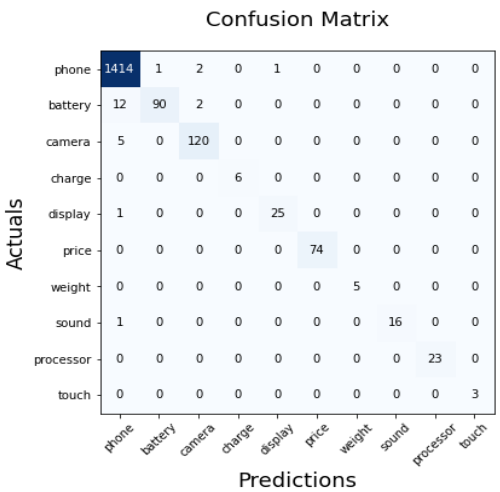

Figure 7. Confusion matrix for Word2Vec with Bi-LSTM model

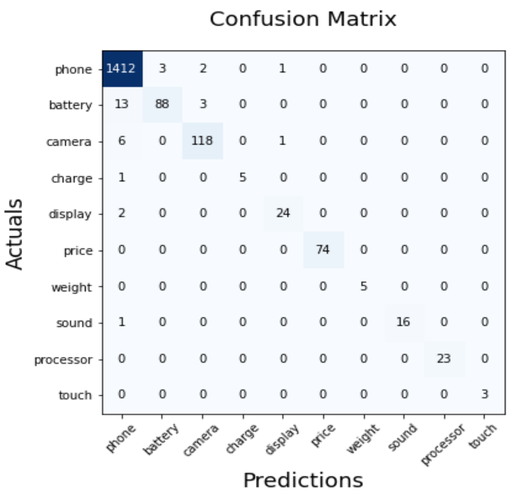

Figure 8. Confusion matrix for BERT with Bi-LSTM model

In the Word2Vec with Bi-LSTM model, the confusion matrix highlights the distribution of correct and incorrect predictions across various aspects such as phone, battery, camera, charge, display, and others. The diagonal elements represent correctly classified instances, showing strong performance with high values such as 1,414 for the phone and 120 for the camera. The off-diagonal elements are sparse, indicating few misclassifications. The evaluation metrics in Table 1 further corroborate this performance, showing high precision, recall, F1-scores, and accuracy across all aspects. For instance, the phone aspect achieved a precision of 1.00, recall of 0.98, and F1-score of 0.99, resulting in an overall accuracy of 0.98.

Table 1. Evaluation metrics for Word2Vec embedding with Bi-LSTM classification

Aspects	True positive	False positive	False negative	True negative	Precision	Recall	F1-score	Accuracy
Phone	1414	4	26	357	1.00	0.98	0.99	0.98
Battery	85	19	1	1696	0.82	0.99	0.89	0.99
Camera	117	8	5	1671	0.94	0.96	0.95	0.99
Charge	6	0	0	1795	1.00	1.00	1.00	1.00
Display	24	2	0	1775	0.92	1.00	0.96	1.00
Price	74	0	1	1726	1.00	0.99	0.99	1.00
Weight	5	0	0	1796	1.00	1.00	1.00	1.00
Sound	17	0	0	1784	1.00	1.00	1.00	1.00
Processor	23	0	0	1778	1.00	1.00	1.00	1.00
Touch	3	0	0	1798	1.00	1.00	1.00	1.00

The BERT with Bi-LSTM model exhibits strong performance, as shown in its confusion matrix (Figure 8) and evaluation metrics (Table 2). It performs comparably to the Word2Vec model, with the phone aspect correctly classified 1,412 times and the camera aspect 118 times. The model maintains high precision, recall, F1-scores, and accuracy across aspects. For instance, the phone aspect has a precision of 1.00, recall of 0.98, and F1-score of 0.99, with an overall accuracy of 0.98, similar to the Word2Vec model.

Table 2. Evaluation metrics for BERT embedding with Bi-LSTM model

Aspects	True positive	False positive	False negative	True negative	Precision	Recall	F1-score	Accuracy
Phone	1,412	6	23	360	1.00	0.98	0.99	0.98
Battery	88	16	3	1,694	0.85	0.97	0.90	0.99
Camera	118	7	5	1,671	0.94	0.96	0.95	0.99
Charge	5	1	0	1,795	0.83	1.00	0.91	1.00
Display	24	2	2	1,773	0.92	0.92	0.92	1.00
Price	74	0	0	1,727	1.00	1.00	1.00	1.00
Weight	5	0	0	1,796	1.00	1.00	1.00	1.00
Sound	16	1	0	1,784	0.94	1.00	0.97	1.00
Processor	23	0	0	1,778	1.00	1.00	1.00	1.00
Touch	3	0	0	1,798	1.00	1.00	1.00	1.00

The Word2Vec and BERT with Bi-LSTM perform well in ABSA, with minor differences in metrics. Word2Vec slightly excels in precision for some aspects, while BERT is competitive in recall and F1-scores. Both methods effectively capture aspect-specific sentiments, with Word2Vec offering efficiency and BERT providing contextual depth. The choice between them depends on specific needs like interpretability and resources. Overall, both embeddings are effective for sentiment analysis, with strong precision, recall, and accuracy in practical applications.

The graphs in Figures 9 and 10 illustrate the testing accuracies for multimodal aspect and sentiment prediction using two different models Word2Vec + Bi-LSTM and BERT + Bi-LSTM. For aspect prediction, the Word2Vec + Bi-LSTM model achieves an accuracy of 92.8%, while the BERT + Bi-LSTM model outperforms it with an accuracy of 95.6%. Similarly, in sentiment prediction, the Word2Vec + Bi-LSTM model attains an accuracy of 93.41%, whereas the BERT + Bi-LSTM model reaches a higher accuracy of 96.28%. These results indicate that the BERT + Bi-LSTM model consistently provides better performance in both aspect and sentiment prediction tasks compared to the Word2Vec + Bi-LSTM model, demonstrating the effectiveness of BERT embeddings in capturing richer contextual information for classification purposes.

Comparison of various models of aspect based sentimental analysis for online product reviews: presents a comparative analysis of various sentiment analysis models to evaluate their effectiveness. The proposed multimodal aspect-based sentiment analysis technique (MASAT) model demonstrates the ability to identify aspects and sentiments in reviews, extract both explicit and implicit aspects, analyze sentiment, and determine aspect polarity. The EAABSAM model, enhances aspect-based sentiment analysis by incorporating implicitly mentioned aspects and emoticons. In contrast, the FSBA model integrates a limited number of manually selected emoticons but does not account for implicit aspects in product reviews. The traditional aspect-based model lacks both emoticon incorporation and implicit aspect detection, limiting its sentiment analysis capabilities. To determine the most effective approach and assess the impact of emojis, emoticons, and implicit aspects on aspect-based sentiment analysis, these three models were systematically compared. Table 3 presents a detailed comparison of their performance.

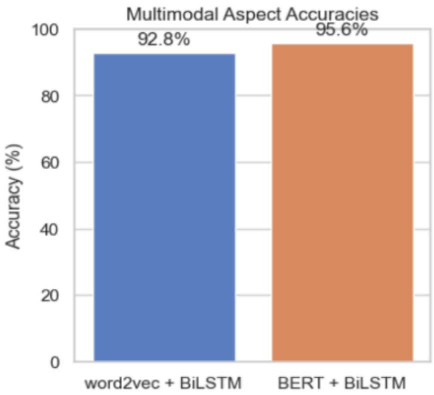

Figure 9. Testing accuracies of multimodal aspect prediction

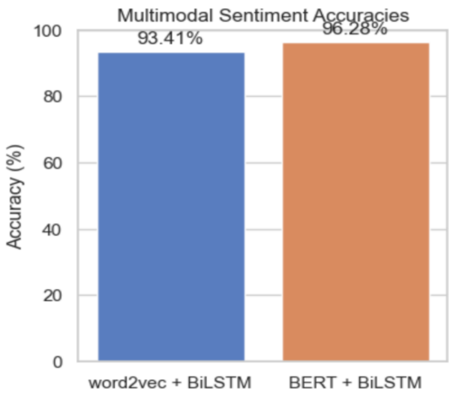

Figure 10. Testing accuracies of multimodal sentiment prediction

Table 3. Comparison of product reviews analyzed by ABM, EAABSAM, MASAT (proposed)

Types of product reviews	Models		
	Existing methods		Proposed
	ABM	EAABSAM	ABM
Only text	✓	✓	✓
Text+Emoticons	✗	✓	✓
Text+Emoji	✗	✗	✓
Only emoticons	✗	✓	✓
Only emoji's	✗	✗	✓
Text+Emoticon's+Emoji's	✗	✗	✓

Table 3 compares product reviews analyzed by existing models ABM and EAABSAM with the proposed MASAT model. The table categorizes reviews based on text, emoticons, and emojis, highlighting that while all models analyze text-based reviews, only MASAT effectively integrates text, emoticons, and emojis. Unlike ABM and EAABSAM, MASAT processes all variants, emphasizing text and emojis, which are increasingly prevalent in user-generated content. The aspect-based approach in ABM neglects implicit aspects and emoticons, while EAABSAM incorporates emoticons for improved sentiment analysis. Table 4 presents performance metrics, showing EAABSAM's superior precision over ABM, yet highlighting the benefits of including implicit details and emoticons. MASAT, utilizing SVM and Bi-LSTM with BERT embeddings, achieves the highest accuracy of 96.3%, outperforming EAABSAM. This demonstrates that incorporating multiple data types significantly enhances sentiment analysis, making MASAT a more comprehensive and effective model.

Table 4. Performance metrics of existing and proposed models

Performance metric	ABM	EAABSAM	MASAT-Bi-LSTM+BERT
Precision (%)	62.6	88.1	92.4
Recall (%)	62.4	84.6	97.0
F1-score (%)	NA	NA	90.8
Accuracy (%)	62.1	88.5	96.3

The EAABSAM model surpasses the aspect-based model in precision for polarity classification. However, incorporating implicit details and emoticons further enhances sentiment analysis accuracy. The MASAT model, utilizing Bi-LSTM with BERT embeddings, demonstrates superior performance, achieving a 96.3% accuracy. Unlike EAABSAM, MASAT effectively analyzes all types of product reviews, including text, emoticons, and emojis, making it a more comprehensive sentiment analysis approach. Its ability to integrate multiple data types ensures a more accurate and robust classification, outperforming EAABSAM in overall sentiment detection.

5. CONCLUSION

In conclusion, this work underscores the potential of multimodal sentiment analysis in capturing comprehensive user sentiments, setting a new standard for sentiment analysis in mobile product reviews through DL techniques. Aspect extraction identifies specific product features in reviews, linking sentiments to these aspects. Feature extraction employs Word2Vec, Emoji2Vec, and BERT to generate embeddings, capturing semantic meanings of text and emojis. Bi-LSTM networks in the classification stage predict aspects and associated sentiments, capturing contextual dependencies. We explored the application of advanced NLP models for multimodal sentiment analysis comparing Word2Vec + Bi-LSTM and BERT + Bi-LSTM. The BERT + Bi-LSTM model, with accuracies of 95.6% in aspect prediction and 96.28% in sentiment prediction, outperformed the Word2Vec + Bi-LSTM model, which achieved 92.8% and 93.41%, respectively. The MASAT model, surpassing ABM, EAABSAM by effectively analyzing all product review types, including text, emoticons, and emojis, enhancing sentiment analysis through implicit detail integration and robust classification. The integration of emoji sentiment analysis demonstrated the enhanced performance of sentiment classification when combining text and emoji.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

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

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Anitha Padigapati	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
A.V. Praveen Krishna		✓				✓	✓	✓	✓	✓		✓	✓	

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

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data set that supports the findings of this study are openly available in Zenodo at https://zenodo.org/records/17833486

REFERENCES

[1] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, May 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.

[2] P. K. Novak, J. Smailović, B. Sluban, and I. Mozetič, "Sentiment of emojis," *PLOS ONE*, vol. 10, no. 12, p. e0144296, Dec. 2015, doi: 10.1371/journal.pone.0144296.

[3] M. Pontiki *et al.*, "SemEval-2016 task 5: aspect-based sentiment analysis," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 19–30, doi: 10.18653/v1/S16-1002.

[4] B. Felbo, A. Mislove, A. Søgaard, I. Rahwan, and S. Lehmann, "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 1615–1625, doi: 10.18653/v1/D17-1169.

[5] Y. Chen, C. K. Hsieh, J. Mahmud, and J. Nichols, "Predicting audience engagement for online news articles based on image content," in *Proceedings of the 27th ACM international conference on information and knowledge management*, 2018, pp. 1113–1122.

[6] B. Eisner, T. Rocktäschel, I. Augenstein, M. Bosnjak, and S. Riedel, "Emoji2vec: learning emoji representations from their description," in *Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media*, 2016, pp. 48–54, doi: 10.18653/v1/W16-6208.

[7] H. Xu, B. Liu, L. Shu, and P. S. Yu, "BERT post-training for review reading comprehension and aspect-based sentiment analysis," *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, pp. 2324–2335, 2019.

[8] N. Alturaief, H. Aljamaan, and M. Baslyman, "AWARE: aspect-based sentiment analysis dataset of apps reviews for requirements elicitation," in *2021 36th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW)*, Nov. 2021, pp. 211–218, doi: 10.1109/ASEW52652.2021.00049.

[9] C. Arumugam and K. Nallaperumal, "EIAASG: emotional intensive adaptive aspect-specific GCN for sentiment classification," *Knowledge-Based Systems*, vol. 260, p. 110149, Jan. 2023, doi: 10.1016/j.knosys.2022.110149.

[10] J. Chen, H. Hou, Y. Ji, J. Gao, and T. Bai, "Graph-based attention networks for aspect level sentiment analysis," in *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, Nov. 2019, vol. 2019-Novem, pp. 1188–1194, doi: 10.1109/ICTAI.2019.00164.

[11] A. Dai, X. Hu, J. Nie, and J. Chen, "Learning from word semantics to sentence syntax by graph convolutional networks for aspect-based sentiment analysis," *International Journal of Data Science and Analytics*, vol. 14, no. 1, pp. 17–26, Jun. 2022, doi: 10.1007/s41060-022-00315-2.

[12] L. De Greve, P. Singh, C. Van Hee, E. Lefever, and G. Martens, "Aspect-based sentiment analysis for German: analyzing 'talk of literature' surrounding literary prizes on social media," *Computational Linguistics in the Netherlands Journal*, vol. 11, pp. 85–104, 2021.

[13] M. Hoang, O. A. Bihorac, and J. Rouces, "Aspect-based sentiment analysis using BERT," *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, pp. 187–196, 2019.

[14] X. Hou, J. Huang, G. Wang, P. Qi, X. He, and B. Zhou, "Selective attention based graph convolutional networks for aspect-level sentiment classification," in *Proceedings of the Fifteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-15)*, 2021, pp. 83–93, doi: 10.18653/v1/2021.textgraphs-1.8.

[15] L. Huang, X. Sun, S. Li, L. Zhang, and H. Wang, "Syntax-aware graph attention network for aspect-level sentiment classification," in *Proceedings of the 28th International Conference on Computational Linguistics*, 2020, pp. 799–810, doi: 10.18653/v1/2020.coling-main.69.

[16] P. K. Jain, R. Pamula, and G. Srivastava, "A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews," *Computer Science Review*, vol. 41, p. 100413, Aug. 2021, doi: 10.1016/j.cosrev.2021.100413.

[17] H. Jangid, S. Singhal, R. R. Shah, and R. Zimmermann, "Aspect-based financial sentiment analysis using deep learning," in *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*, 2018, vol. 2018-Janua, pp. 1961–1966, doi: 10.1145/3184558.3191827.

[18] A. Karimi, L. Rossi, and A. Prati, "Improving BERT performance for aspect-based sentiment analysis," *ICNLSP 2021 - Proceedings of the 4th International Conference on Natural Language and Speech Processing*, pp. 39–46, 2021.

[19] B. Liang, H. Su, L. Gui, E. Cambria, and R. Xu, "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks," *Knowledge-Based Systems*, vol. 235, p. 107643, Jan. 2022, doi: 10.1016/j.knosys.2021.107643.

[20] B. Liang *et al.*, "Few-shot aspect category sentiment analysis via meta-learning," *ACM Transactions on Information Systems*, vol. 41, no. 1, pp. 1–31, Jan. 2023, doi: 10.1145/3529954.

[21] A. Mewada and R. K. Dewang, "SA-ASBA: a hybrid model for aspect-based sentiment analysis using synthetic attention in pretrained language BERT model with extreme gradient boosting," *Journal of Supercomputing*, vol. 79, no. 5, pp. 5516–5551, Mar. 2023, doi: 10.1007/s11227-022-04881-x.

[22] 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, May 2023, doi: 10.1007/s10489-022-04023-z.

[23] Y. Qi, X. Zheng, and X. Huang, "Aspect-based sentiment analysis with enhanced aspect-sensitive word embeddings," *Knowledge and Information Systems*, vol. 64, no. 7, pp. 1845–1861, Jul. 2022, doi: 10.1007/s10115-022-01688-3.

[24] O. Alqaryouti, N. Siyam, A. A. Monem, and K. Shaalan, "Aspect-based sentiment analysis using smart government review data," *Applied Computing and Informatics*, vol. 20, no. 1–2, pp. 142–161, Jan. 2024, doi: 10.1016/j.aci.2019.11.003.

[25] Y. Li, Q. He, and L. Yang, "Part-of-speech based label update network for aspect sentiment triplet extraction," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 1, p. 101908, Jan. 2024, doi: 10.1016/j.jksuci.2023.101908.

[26] K. Ahmed *et al.*, "Breaking down linguistic complexities: a structured approach to aspect-based sentiment analysis," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 8, p. 101651, Sep. 2023, doi: 10.1016/j.jksuci.2023.101651.

[27] Z. Chen and T. Qian, "Enhancing aspect term extraction with soft prototypes," in *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2020, pp. 2107–2117, doi: 10.18653/v1/2020.emnlpmain.164.

[28] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 2019, vol. 1, pp. 4171–4186, doi: 10.18653/v1/N19-1423.

[29] R. Wang *et al.*, "Post-processing method with aspect term error correction for enhancing aspect term extraction," *Applied Intelligence*, vol. 52, no. 14, pp. 15751–15763, 2022, doi: 10.1007/s10489-022-03380-z.

[30] L. Yuan, J. Wang, L. C. Yu, and X. Zhang, "Encoding syntactic information into transformers for aspect-based sentiment triplet extraction," *IEEE Transactions on Affective Computing*, vol. 15, no. 2, pp. 722–735, Apr. 2024, doi: 10.1109/TAFFC.2023.3291730.

[31] L. Zhu, X. Zhu, J. Guo, and S. Dietze, "Exploring rich structure information for aspect-based sentiment classification," *Journal of Intelligent Information Systems*, vol. 60, no. 1, pp. 97–117, Feb. 2023, doi: 10.1007/s10844-022-00729-1.

BIOGRAPHIES OF AUTHORS

Anitha Padigapati received B.Tech. degree from JNTU Hyderabad in the year 2006 and also received M.Tech. degree in the specialization of CSE from JNTU Hyderabad in the year 2016. Currently she is a Ph.D. scholar in the Department of Computer Science and Engineering with specialization in data mining from Koneru Lakshmaiah Education Foundation, deemed to be University. She worked as Assistant Professor in reputed Engineering Colleges under JNTU Hyderabad. She acted as Incharge Head of the Department and also acted as placement officer for the period of 3 years. She was Deputed as Deputy Presiding Officer for GATE Exam and several Exams conducted by TCS collaboration with IIT Hyderabad, IIS Bangalore. She has supervised more than 15 bachelor and masters' students. She has authored or coauthored for 5 publications. There is 1 patent with her name. Her research interests include data mining, NLP, ML, and DL. She can be contacted at email: anitha.padigapati@gmail.com.

A.V. Praveen Krishna currently working as associate professor in Koneru Lakshmaiah Education Foundation, deemed to be University. Actively engaged in research and publications in the areas of network security, intrusion detections, AI, ML, and IoT. Had active research publications in reputed Springer, Elsevier, IEEE explorer, SCI and Scopus based international journals. Recognized by Association of Computing Machinery (ACM) as one of the top "Brand Ambassador" for the year 2016 for his contributions. Appointed as the Sectional Committee Member of 105th Indian Science Congress for Information Science and Technology (including Computer Science's) for the year 2017. He was awarded with "Best Computer Science Faculty" for year 2012 in all India basis by Government of Puddcherry and ASDF, Techno Forum. Awarded recently in the category of "Icon of Academic Innovation Award" by Artificial Intelligence Medical and Engineering Researchers Society (AIMER Society) for the year 2024. He was awarded with "Significant Contribution Award" for his outstanding contribution, by India's prestigious professional society "Computer Society of India (CSI)" for the year 2011 in Ahmedabad. Acted as a panel of judge for South East Asia Regional Computer Confederation and Computer Society of India. He can be contacted at email: praveenkrishna@kluniversity.in.