A BiLSTM-CF and BiGRU-based Deep Sentiment Analysis Model to Explore Customer Reviews for Effective Recommendations

Muhammad Rizwan Rashid Rana
University Institute of Information Technology, PMAS Arid Agriculture University, Pakistan
rizwanrana315@gmail.com

Asif Nawaz
University Institute of Information Technology, PMAS Arid Agriculture University, Pakistan
asif.nawaz@uaar.edu.pk (corresponding author)

Tariq Ali
University Institute of Information Technology, PMAS Arid Agriculture University, Pakistan
tariq.ali@uaar.edu.pk

Ahmed M. El-Sherbeeny
Industrial Engineering Department, College of Engineering, King Saud University, Saudi Arabia
aelsherbeeny@ksu.edu.sa

Waqar Ali
Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Italy
waqar.ali@unive.it

Received: 13 August 2023 | Revised: 24 August 2023 | Accepted: 26 August 2023

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | *DOI: https://doi.org/10.48084/etasr.6278*

ABSTRACT

The advancement of technology has led to the rise of social media forums and e-commerce platforms, which have become popular means of communication, and people can express their opinions through comments and reviews. Increased accessibility to online feedback helps individuals make informed decisions about product purchases, services, and other decisions. This study used a sentiment analysis-based approach to improve the functionality of the recommendations from user reviews and consider the features (aspects and opinions) of products and services to understand the characteristics and attributes that influence the performance of classification algorithms. The proposed model consists of data preprocessing, word embedding, character representation creation, feature extraction using BiLSTM-CF, and classification using BiGRU. The proposed model was evaluated on different multidomain benchmark datasets demonstrating impressive performance. The proposed model outperformed existing models, offering more promising performance results in recommendations.

Keywords-sentiment analysis; reviews; classification; deep learning; recommendations

I. INTRODUCTION

The rapid and continuous growth of the Internet has resulted in the migration of a multitude of applications from the physical to the digital realm [1]. Various sectors, such as e-commerce, banking, multimedia consumption, and service bookings, have experienced this transition, altering the way people access information. In the physical world, people typically rely on a limited number of resources to search for information. For example, to purchase a mobile phone, one may visit multiple electronic stores, seek information from sales assistants about phone specifications and recommendations, and ultimately choose after a few consultations. However, the overwhelming abundance of online content, particularly unstructured text and online reviews, is overflowing the Web. Online reviews play a crucial role for users who need to make decisions among various options, such as purchasing a product, watching a movie, or selecting a restaurant [2]. Through these reviews, users express their opinions about different items and their respective aspects, including characteristics, features, attributes, or components. Typically written in free text format, reviews require users to carefully read and analyze them to identify expressed opinions and discover the strengths and weaknesses of the available items to make informed decisions [3].

Recommendation systems aim to offer personalized recommendations for most items to address the challenge of information overload and assist users in decision-making tasks. Recommendation systems, which were developed in the 1990s and rely on user preferences and ratings, have experienced significant growth and widespread adoption over the past few decades [4-5]. Today, they are essential in various industries, including e-commerce, health, telecoms, etc. For instance, Amazon, YouTube, and Twitter use recommendation systems to suggest popular products, recommend related videos, and offer connections with people and pages to follow, respectively. Recommendation systems can derive significant benefits from incorporating sentiment analysis [6]. Sentiment analysis is the process of extracting sentiments/opinions from user text [7]. There are two categories of sentiments: binary and multi-sentiments. Binary sentiments are positive or negative, while multi-sentiments include happy, angry, upset, dull, etc. Existing approaches in recommendation systems use sentiment analysis at the sentence or document level. However, the modern Aspect-Based Sentiment Analysis (ABSA) is used less in recommendation systems [8].

Existing sentiment analysis methods are domain-oriented and face the challenges of subjective sentences and slang words in user reviews [9-10]. This reduces the accuracy of sentiment analysis and, as a result, the recommendations of products. This study proposes a hybrid deep learning model for product recommendation using ABSA to overcome these challenges. Using ABSA models, recommendation systems can effectively address the data sparsity issue commonly encountered in traditional recommendation systems. Therefore, integrating sentiment analysis at the aspect level into recommendation systems can significantly enhance the quality of recommendations provided to users. The main contributions of this study are:

- Consideration of the features (aspects and opinions) of products and services that help to understand the characteristics and attributes that influence the performance of classification algorithms when dealing with product reviews.
- The proposal of a transformer-based hybrid model that uses Concurrent Neural Networks, Bidirectional Long Short-Term Memory, and Conditional Random Field (CNN-BiLSTM-CRF) for feature extraction and Bidirectional Gated Recurrent Unit (BiGRU) for classification.
- The results of the proposed method were compared with baseline approaches, demonstrating its superior results.

II. BACKGROUND STUDY

Sentiment analysis holds immense value across various domains, such as business, government, and education, and has been proven incredibly useful. Extensive research has investigated the use of sentiment analysis in recommendation systems. In [11], data from Twitter-based airline discussions were analyzed using various features with data mining algorithms. The findings showed that emoticons played a significant role in determining the conveyed sentiment, surpassing the influence of textual data analysis. In [12], LSTM was combined with Word2Vec representation to improve sentiment analysis performance, showing that the LSTM gating mechanism during training contributed to improved results, outperforming other baseline models. In [13], a Bi-LSTM model was presented for opinion mining of reviews. The dataset consisted of 23,485 assessments classified into three categories based on their negative, neutral, and positive rating points. The data were cleaned by tokenizing and removing special characters from the reviews. Word embeddings were also generated using a Word2Vec pre-trained model. In [14], a nearest-neighbor-based approach was used for feature selection, calculating distance metrics and cosine similarity on preprocessed data. The primary goal was to improve performance by optimizing feature selection and tuning hyperparameters, involving various feature vectorization methods to achieve higher accuracy.

In [15], the efficiency of various machine-learning models was evaluated for classifying sentiments, showing that the Bidirectional Encoder Representations from Transformers (BERT) model demonstrated the most exceptional performance, having a remarkable accuracy of 85.4% in sentiment analysis. In [16], a combined CNN and BiLSTM model, called ConvBiLSTM, was proposed for sentiment analysis. The CNN layer receives feature embeddings as input and outputs lower-level features, while BiLSTM is used for classification. In [17], a sentiment analysis was carried out on Weibo posts about travelers' experiences with commercial air travel. This study used a multitask structural arrangement to evaluate events and sentiments simultaneously. This model achieved 89% accuracy but struggled with out-of-vocabulary words, limiting its performance and adaptability to dynamic language. In [18], a mixed-fusion architecture was proposed for image-text sentiment analysis, leveraging the correlation between visual and semantic content. This model achieved 76% accuracy for Twitter data and 87% for Getty photos but was costly and time-consuming due to the search for opinionated words. In [19], extensive social information was incorporated for multimodel sentiment analysis using regional attention and a heterogeneous relation network. This model achieved 87% accuracy but lost some features during extraction, potentially losing important information and reducing predictive power.

In [20], ABSA-PER was introduced, which was a model designed for polarity estimation of reviews through ABSA. Experimental results demonstrated that ABSA-PER outperformed the baseline methods, achieving a remarkable 86.5% accuracy for polarity calculation. In [21], a novel approach was proposed, called lexical attention and aspect-oriented GCN. This method involved constructing an aspect-oriented dependency-parsed tree that is crucial in calculating sentiment scores. Then, an aspect-oriented GCN extracts weighted lexical features specific to each aspect. The proposed approach was used in three datasets and, by considering lexical properties and aspect-specific dependencies, achieved promising results in sentiment polarity prediction for different aspects within the text. In [22], a meta-ensemble deep learning approach was proposed that combined multiple deep learning models trained with different meta-learners. This approach was experimentally shown to outperform baseline deep learning models, especially when trained with probability class distributions.

III. METHODOLOGY

The proposed approach revolves around three key components: Data Processing, Feature Extraction, and Sentiment Classification. Since e-commerce platforms host user-generated content, it is necessary to preprocess textual reviews to ensure their cleanliness before extracting meaningful features. Then, advanced language representation models, namely BERT, and CNN, were used for word embedding and character representation, respectively. Then, BiLSTM with CRF was used for feature extraction, which is well-suited for sequential data processing. Lastly, BiGRU was used to classify the sentiments. This approach effectively captured contextual dependencies and performed superior sentiment analysis using bidirectional information flow in the Recurrent Neural Network (RNN). Figure 1 shows the proposed model.

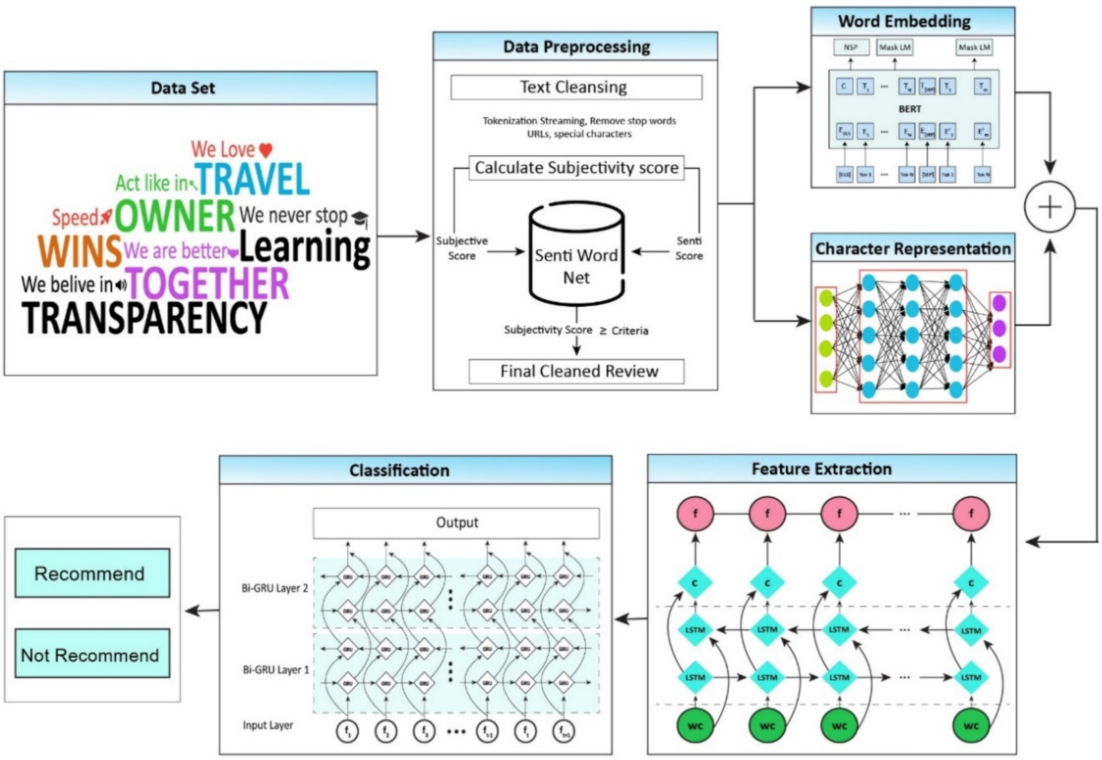

Fig. 1. The proposed model.

A. Data Preprocessing

Data preprocessing is an important step in preparing text for classification, particularly in online contexts where noise and irrelevant content can be prevalent. This noise may include HTML tags, advertisements, and other distracting elements [23]. Moreover, words within the text often have minimal impact on the classification task but still contribute to the overall complexity because each word is treated as a distinct dimension. The fundamental assumption behind effective data preprocessing is that reducing text noise can improve classifier performance and sentiment analysis results [11]. The entire approach to data preprocessing consists of several essential steps. The first step involves cleaning the online text, removing unwanted elements such as HTML tags and advertisements, and ensuring that only the relevant text remains for further analysis. Next, focus is given to removing unnecessary white spaces within the text, as they do not contribute to the meaning or the classification task. Expanding abbreviations is another important preprocessing step, as many texts contain abbreviated words or acronyms that can be difficult for classification algorithms to interpret accurately. Expanding these abbreviations to their full forms facilitates better comprehension and analysis of the text.

Sentences can be categorized into two types: subjective and objective. Subjective sentences convey positive or negative opinions, while objective sentences do not express any opinions [20]. Objective sentences are usually facts or universal truths. For example, "I love the world" is considered subjectivity, but "The sun is hot" must be considered an objective statement. Therefore, it is necessary to remove the objective sentences from the dataset. This step improves the accuracy of the sentiment analysis. A lexicon, SentiWordNET

3.0, calculates the subjectivity score and provides the negative and positive scores of words [3]. The subjectivity score is calculated using the following equation:

$$Sub\ Score = \frac{\sum_{i=0}^{n}(w_{pos} + w_{neg})}{n} \quad (1)$$

where w_{pos} and w_{neg} are the positive and negative scores of the words and n represents the total number of words present in the review.

B. Feature Extraction and Selection

The most suitable and optimal features (aspects and opinions) are important for successful sentiment analysis. The feature extraction and selection phase is carried out using the hybrid deep learning technique BiLSTM-CRF. Two different word representations are used to feed Bi-LSTM-CRF for feature extraction: word embedding using BERT and character-level representation using CNN. To generate word embeddings, the BERT tokenizer is used first to tokenize the input text into individual words or subwords, and then the tokenized input is passed through the BERT model to generate a sequence of hidden states [24]. The dot product of these concealed states and a learned weight matrix can then be used to generate word embeddings for every word in the input text. The BERT word embeddings are particularly useful because they are context-aware, meaning that the embedding for a word can vary depending on the context in which it appears. This contrasts with many other word embedding methods, which generate a fixed embedding for each word regardless of context.

Previous studies have demonstrated the effectiveness of CNN in extracting morphological information, such as prefixes or suffixes of words, from individual characters and encoding it into neural representations [25]. This study adopted a similar CNN architecture with the distinction that character embeddings were used solely as inputs to the CNN, omitting character-type features. Additionally, a dropout layer was applied before feeding the character embeddings into the CNN [26]. Finally, the neural network model was constructed by feeding the BLSTM output vectors into a CRF layer. Figure 2 shows the architecture of the framework in detail, with CNN-computed character-level representations for each word using BERT-generated character embeddings as input. The character-level representation vector and the word embedding vector are combined and provided as input to the BLSTM network.

BLSTM is an RNN variant that combines the properties of both forward and backward LSTMs [13]. In traditional LSTM, the input sequence undergoes forward processing, with each time step receiving information from the preceding time step. However, in natural language processing, information from past and future time steps can prove to be useful. To overcome the limitation of being unable to capture information from past and future contexts, BLSTM was designed with two hidden layers in its architecture: one layer operates on the input in the forward direction, and the other layer processes it in the backward direction. As the backward layer processes the sequence in reverse order, it can effectively capture information from both past and future contexts. By combining the outputs of both the forward and backward layers, BLSTM can model the dependencies in both directions, rendering it particularly useful for tasks that require understanding a sequence's context, such as speech recognition, sentiment analysis, named entity recognition, and many more.

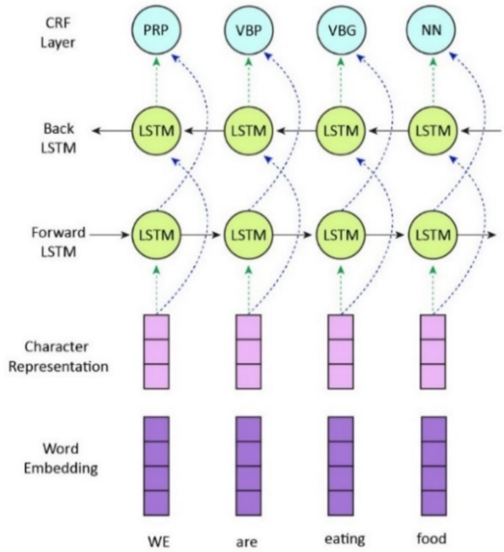

Fig. 2. Feature extraction framework.

The output vectors produced by BiLSTM are subsequently fed into the CRF layer to determine the most suitable features. CRF is a probabilistic graphical model particularly used in sequence labeling problems [27]. In sequence labeling, the goal is to assign a label or category to each element in a sequence of observations. For example, in natural language processing, sequence labeling can involve assigning part-of-speech tags to each word in a sentence or identifying entities such as person names, locations, or dates. The output of the CRF layers is the suitable features, i.e., aspects and opinions.

C. Classification

GRU exhibits a simpler structure compared to LSTM, making it relatively easier to train. Specifically, BiGRU integrates gates that suppress information loss. It consolidates the output and forget gates in LSTM to create an update gate, streamlining the structure of BiGRU and significantly saving disk space [28]. The BiGRU model effectively highlights significant information present in a text by generating output in both forward and backward directions.

The GRU model automatically learns which resources are essential and which can be eliminated during training, making it potentially more effective in handling extended texts. However, a one-way GRU network can miss important information in text sentiment analysis because its states are always from the front to the back. Research evidence strongly suggests that BiGRU outperforms GRU [28]. BiGRU consists of two unidirectional GRUs that operate in opposite directions, forward and backward, as shown in Figure 3. Both GRU layers in the BiGRU model simultaneously use all resources that flow through the network. At each moment, the input is fed into both GRUs simultaneously but in opposite directions. The result is jointly determined to improve accuracy.

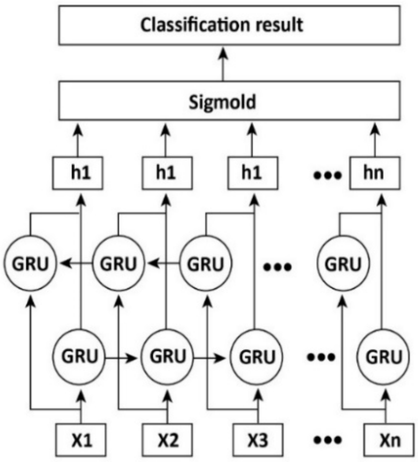

Fig. 3. BiGRU framework.

The BiGRU network consists of an input layer, a forward GRU, a reverse GRU, and a forward GRU output layer. This intricate network is designed to transmit data to both the forward GRU and the reverse GRU from the input layer, and these two powerful GRUs work together to determine the output sequence jointly. The following expression represents the BiGRU layer:

$$\overrightarrow{h_t} = \mathrm{Gru}(\mathrm{xt}, \overrightarrow{h_t} - 1)$$

$$\overleftarrow{h_t} = \mathrm{Gru}(\mathrm{xt}, \overleftarrow{h_t} - 1)$$

$$h_t = f\ (W\ \overrightarrow{h_t}\,\overrightarrow{h_t} + W\overleftarrow{h_t}\overleftarrow{h_t}\ b_t)$$

where t is time, $\overrightarrow{h_t}$ is positive layers' state at time t, $\overleftarrow{h_t}$ is negative layers' state at time t, $W\overrightarrow{h_t}$ is the weight of the positive hidden layer, $W\overleftarrow{h_t}$ is the weight of the negative hidden layer, and b_t is the hidden layer bias at time t.

IV. RESULTS AND DISCUSSION

A series of experiments were carried out to evaluate the accuracy and efficacy of the proposed model. The experimental evaluation demonstrated that the proposed method significantly outperformed the existing cutting-edge methods.

A. Dataset

The proposed model was evaluated using three benchmark datasets from different domains: products, hotels, and movie reviews. The product reviews dataset consists of reviews of five distinct products: Canon, Nikon, MP3 Players, Nokia 6610, and Apex AD 2600 [29]. The hotel review dataset originated from tripadvisor.com [30]. Lastly, the movie reviews dataset was obtained from the large movie review dataset used by Stanford University's AI department [31]. Table I shows the details of the datasets.

TABLE I. DATASET BREAKDOWN

#	Dataset	Review type	Reviews about	Reviews
1	DS 1	Products	Canon, Nikon, MP3 Players, Nokia 6610, Apex AD 2600	3,552
2	DS 2	Hotels	Hotels	20,490
3	DS 3	Movies	Movies	50,000

B. Results

K-fold cross-validation was used to validate the results of the proposed model. At first, the dataset was divided into K bins of identical sizes, where one bin was allocated for testing and the other K–1 bins were used for training purposes [3]. This process was executed on a GeForce GTX 1080 TiGPU. The parameters used were a batch size of 64, a learning rate of $3e^{-5}$, a dropout rate of 0.3, and 16 epochs. Performance was evaluated using three key metrics: accuracy, recall, and precision. Table II presents the results. When applied to DS 1, the proposed model achieved 92.56% accuracy, 91.32% precision, and 91.82% recall. In DS 2, the proposed BERT-CNN model achieved 91.8% accuracy, 91.05% precision, and 91.34% recall. Finally, in DS 3, it obtained an accuracy of 90.85%, an exceptional precision of 98.22%, and a recall of 89.45%. These results indicate that the proposed model accurately identified sentiments and aspects across all three datasets, showing remarkably high precision while maintaining a respectable recall rate, thereby affirming its potential as an efficient tool for sentiment analysis in a variety of domains.

TABLE II. EXPERIMENTAL RESULTS

#	Dataset	Accuracy (%)	Precision (%)	Recall (%)
1	DS 1 (product reviews)	92.56	91.32	91.82
2	DS 2 (hotel reviews)	91.82	91.05	91.34
3	DS 3 (movie reviews)	90.85	89.22	89.45

Another experiment was carried out to assess the performance of BERT-CNN on the three datasets, using standard word embedding models, namely Global Vectors for word representation (GloVe) and Word2Vec, which are well-known methods for constructing word embeddings [32]. GloVe is a word embedding model that learns word vectors by analyzing co-occurrence statistics in a corpus, thereby encoding global and local word relationships. Word2Vec, is a word embedding model that learns word vectors by predicting adjacent words or contexts in a corpus, producing vector representations that encode semantic relationships and word similarities [33]. Table III provides a comparison of the proposed BERT-CNN approach with GloVe and Word2Vec.

TABLE III. COMPARISON OF THE PROPOSED APPROACH WITH GLOVE AND WORD2VEC

#	Word Embeddings	Accuracy	Precision	Recall
DS 1 (Product Reviews)				
1	Glove	82.79%	82.12	82.53
2	Word2Vec	78.28%	77.56	77.86
3	BERT-CNN (proposed)	92.56%	91.32	91.82
DS 2 (Hotel Reviews)				
1	Glove	80.65%	80.01	80.42
2	Word2Vec	74.28%	73.51	73.81
3	BERT-CNN (proposed)	91.82%	91.05	91.34
DS 3 (Movie Reviews)				
1	Glove	81.32%	80.48	81.03
2	Word2Vec	75.36%	74.65	74.32
3	BERT-CNN (proposed)	90.85%	89.22	89.45

The results show that the proposed BERT-CNN model consistently outperformed the baseline word embedding models across the three different domain datasets. This implies that the BERT-CNN model performs well and effectively captures contextual information and features from user reviews. These results highlight its potential to enhance various natural language processing tasks, indicating its superiority over traditional word embedding approaches.

The effectiveness of the proposed model was assessed in contrast to some of the latest baseline models. Table IV shows the comparisons in all datasets in terms of accuracy. BERT-SKG [34] integrates sentiment domain knowledge with BERT and generates embedding vectors that align entities in the sentiment knowledge graph. LSIBA-ENN [35] introduces a novel model designed to analyze online product reviews. ARM-BERT [36] proposed POS-ARM with BERT for ABSA. Finally, LeBERT [37] integrates N-grams, BERT lexicon, and CNN for sentiment classification in reviews.

TABLE IV. COMPARISON OF THE PROPOSED APPROACH WITH BASELINE APPROACHES IN TERMS OF ACCURACY

#	Approaches	DS 1	DS 2	DS 3
		Accuracy (%)		
1	BERT-SKG	89.24	88.12	88.34
2	LSIBA-ENN	90.31	89.31	89.16
3	ARM-BERT	91.80	91.23	90.07
4	LeBERT	88.20	88.72	88.42
5	Proposed	92.56	91.82	90.85

The BERT-SKG approach achieved an average accuracy of 88.90%, having its highest accuracy of 89.24% in DS 1, followed by 88.34% in DS 3, and 88.12% in DS 2. Although BERT-SKG performed reasonably well, it had slightly lower accuracy than the other approaches. The LSIBA-ENN approach showed better performance with an average accuracy of 89.26%, showing its highest accuracy of 90.31% in DS 1, followed by 89.31% in DS 2, and 89.16% in DS 3. These results indicate that LSIBA-ENN outperformed BERT-SKG in all three datasets. The ARM-BERT approach demonstrated even higher accuracy, averaging 91.03%, having an impressive accuracy of 91.80% in DS 1, followed by 91.23% in DS 2, and a slightly lower accuracy of 90.07% in DS 3. ARM-BERT outperformed both BERT-SKG and LSIBA-ENN, indicating superior performance in sentiment analysis tasks. LeBERT achieved an average accuracy of 88.45%, having the highest accuracy of 88.72% in DS 2, followed by 88.20% in DS 1, and 88.42% in DS 3. Although LeBERT performed reasonably well, it fell behind the ARM-BERT approach, which achieved higher accuracies across all datasets. Finally, the proposed model outperformed all other approaches, achieving an impressive average accuracy of 91.41%, having the highest accuracy of 92.56% in DS 1, followed by 91.82% in DS 2, and 90.85% in DS 3.

Figure 4 shows the comparison results of the proposed method with other baseline approaches in terms of accuracy, precision, and recall. The proposed model outperformed the other approaches in different domain datasets, showcasing its effectiveness in sentiment analysis and recommendation systems. ARM-BERT also showed strong performance, outperforming BERT-SKG, LSIBA-ENN, and LeBERT. LSIBA-ENN and LeBERT showed decent results, while BERT-SKG lagged slightly behind. The performance of the proposed model stems from its effective feature extraction and selection methods, as it uses word embedding with BERT, character representation through CNN, feature extraction with BiLSTM-CRF, and classification with BiGRU. BiGRU operates in both forward and backward directions in the input sequence, allowing it to capture contextual information from preceding and subsequent contexts. This comprehensive approach facilitates understanding of word dependencies and capturing the overall sentiment expressed in sentences. Therefore, the proposed model acts as an efficient system for recommending products, movies, and hotels.

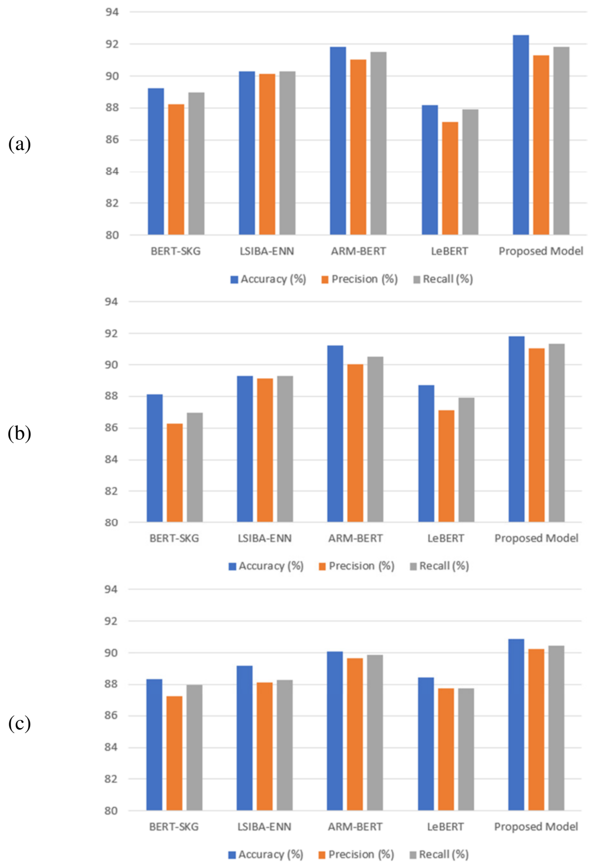

Fig. 4. (a) DS 1, (b) DS 2, and (c) DS 3 results comparison.

V. CONCLUSION

Sentiment analysis is important in offering recommendations for various products and services, as it empowers individuals to perceive and use customers' opinions, emotions, and attitudes toward them. This study presented an innovative approach to sentiment analysis in recommendation systems by implementing hybrid deep-learning models. The proposed method extracted features using BiLSTM-CRF and performed classification using BiGRU. An experimental evaluation was conducted on three multidomain benchmark datasets on product, hotel, and movie reviews. The proposed model demonstrated superior performance compared to the BERT-SKG, ARM-BERT, and LeBERT approaches, since it achieved accuracy of 92.56% in the product reviews dataset, 91.82% in the hotel reviews dataset, and 90.85% in the movie reviews dataset. The results demonstrate that the proposed hybrid deep learning-based sentiment analysis model performed better than the previous ones by utilizing additional information from user reviews/comments data with better feature extraction and classification. In the future, sentiment analysis and evaluation are necessary to confirm its superiority and applicability in real-world scenarios, particularly across different cultures.

ACKNOWLEDGMENT

The authors express their appreciation to King Saud University, Saudi Arabia, for funding this work through Researchers Supporting Project Number RSP2023R133.

REFERENCES

[1] M. A. Kausar, S. O. Fageeri, and A. Soosaimanickam, "Sentiment Classification based on Machine Learning Approaches in Amazon Product Reviews," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10849–10855, Jun. 2023, https://doi.org/10.48084/etasr.5854.

[2] M. R. R. Rana, S. U. Rehman, A. Nawaz, T. Ali, and M. Ahmed, "A conceptual model for decision support systems using aspect based sentiment analysis," *Proceedings of the Romanian Academy*, vol. 22, no. 4, pp. 371–380, Apr. 2021.

[3] D. Elangovan and V. Subedha, "Adaptive Particle Grey Wolf Optimizer with Deep Learning-based Sentiment Analysis on Online Product Reviews," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10989–10993, Jun. 2023, https://doi.org/10.48084/etasr.5787.

[4] P. Resnick and H. R. Varian, "Recommender systems," *Communications of the ACM*, vol. 40, no. 3, pp. 56–59, Mar. 1997.

[5] D. Jannach, A. Manzoor, W. Cai, and L. Chen, "A Survey on Conversational Recommender Systems," *ACM Computing Surveys*, vol. 54, no. 5, Feb. 2021, Art. no. 105, https://doi.org/10.1145/3453154.

[6] A. Nawaz, T. Ali, Y. Hafeez, S. ur Rehman, and M. R. Rashid, "Mining public opinion: a sentiment based forecasting for democratic elections of Pakistan," *Spatial Information Research*, vol. 30, no. 1, pp. 169–181, Feb. 2022, https://doi.org/10.1007/s41324-021-00420-7.

[7] M. Mahyoob, J. Algaraady, M. Alrahiali, and A. Alblwi, "Sentiment Analysis of Public Tweets Towards the Emergence of SARS-CoV-2 Omicron Variant: A Social Media Analytics Framework," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8525–8531, Jun. 2022, https://doi.org/10.48084/etasr.4865.

[8] W. M. S. Yafooz, E. A. Hizam, and W. A. Alromema, "Arabic Sentiment Analysis on Chewing Khat Leaves using Machine Learning and Ensemble Methods," *Engineering, Technology & Applied Science Research*, vol. 11, no. 2, pp. 6845–6848, Apr. 2021, https://doi.org/10.48084/etasr.4026.

[9] H. Wu, Z. Zhang, S. Shi, Q. Wu, and H. Song, "Phrase dependency relational graph attention network for Aspect-based Sentiment Analysis," *Knowledge-Based Systems*, vol. 236, Jan. 2022, Art. no. 107736, https://doi.org/10.1016/j.knosys.2021.107736.

[10] A. I. Saad, "Opinion Mining on US Airline Twitter Data Using Machine Learning Techniques," in *2020 16th International Computer Engineering Conference (ICENCO)*, Cairo, Egypt, Sep. 2020, pp. 59–63, https://doi.org/10.1109/ICENCO49778.2020.9357390.

[11] M. A. Ullah, S. M. Marium, S. A. Begum, and N. S. Dipa, "An algorithm and method for sentiment analysis using the text and emoticon," *ICT Express*, vol. 6, no. 4, pp. 357–360, Dec. 2020, https://doi.org/10.1016/j.icte.2020.07.003.

[12] N. K. Gondhi, U. Chaahat, E. Sharma, A. H. Alharbi, R. Verma, and M. A. Shah, "Efficient Long Short-Term Memory-Based Sentiment Analysis of E-Commerce Reviews," *Computational Intelligence and Neuroscience*, vol. 2022, Jun. 2022, Art. no. e3464524, https://doi.org/10.1155/2022/3464524.

[13] J. S. Vimali and S. Murugan, "A Text Based Sentiment Analysis Model using Bi-directional LSTM Networks," in *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatre, India, Jul. 2021, pp. 1652–1658, https://doi.org/10.1109/ICCES51350.2021.9489129.

[14] R. Hegde and S. Seema, "Nearest neighbour-based feature selection and classification approach for analysing sentiments," *International Journal of Bioinformatics Research and Applications*, vol. 18, no. 1–2, pp. 16–29, Jan. 2022, https://doi.org/10.1504/IJBRA.2022.121757.

[15] K. Dhola and M. Saradva, "A Comparative Evaluation of Traditional Machine Learning and Deep Learning Classification Techniques for Sentiment Analysis," in *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Jan. 2021, pp. 932–936, https://doi.org/10.1109/Confluence51648.2021.9377070.

[16] R. Vatambeti, S. V. Mantena, K. V. D. Kiran, M. Manohar, and C. Manjunath, "Twitter sentiment analysis on online food services based on elephant herd optimization with hybrid deep learning technique," *Cluster Computing*, Feb. 2023, https://doi.org/10.1007/s10586-023-03970-7.

[17] L. Wang, W. Guo, X. Yao, Y. Zhang, and J. Yang, "Multimodal Event-Aware Network for Sentiment Analysis in Tourism," *IEEE MultiMedia*, vol. 28, no. 2, pp. 49–58, Apr. 2021, https://doi.org/10.1109/MMUL.2021.3079195.

[18] F. Huang, X. Zhang, Z. Zhao, J. Xu, and Z. Li, "Image–text sentiment analysis via deep multimodal attentive fusion," *Knowledge-Based Systems*, vol. 167, pp. 26–37, Mar. 2019, https://doi.org/10.1016/j.knosys.2019.01.019.

[19] J. Xu, Z. Li, F. Huang, C. Li, and P. S. Yu, "Social Image Sentiment Analysis by Exploiting Multimodal Content and Heterogeneous Relations," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2974–2982, Apr. 2021, https://doi.org/10.1109/TII.2020.3005405.

[20] A. Banjar, Z. Ahmed, A. Daud, R. Abbasi, and H. Dawood, "Aspect-Based Sentiment Analysis for Polarity Estimation of Customer Reviews on Twitter," *Computers, Materials & Continua*, vol. 67, no. 2, pp. 2203–2225, 2021, https://doi.org/10.32604/cmc.2021.014226.

[21] W. Li, S. Yin, and T. Pu, "Lexical attention and aspect-oriented graph convolutional networks for aspect-based sentiment analysis," *Journal of Intelligent & Fuzzy Systems*, vol. 42, no. 3, pp. 1643–1654, Jan. 2022, https://doi.org/10.3233/JIFS-211045.

[22] R. Kora and A. Mohammed, "An enhanced approach for sentiment analysis based on meta-ensemble deep learning," *Social Network Analysis and Mining*, vol. 13, no. 1, Mar. 2023, Art. no. 38, https://doi.org/10.1007/s13278-023-01043-6.

[23] A. Nawaz, A. A. Awan, T. Ali, and M. R. R. Rana, "Product's behaviour recommendations using free text: an aspect based sentiment analysis approach," *Cluster Computing*, vol. 23, no. 2, pp. 1267–1279, Jun. 2020, https://doi.org/10.1007/s10586-019-02995-1.

[24] R. Catelli, S. Pelosi, and M. Esposito, "Lexicon-Based vs. Bert-Based Sentiment Analysis: A Comparative Study in Italian," *Electronics*, vol. 11, no. 3, Jan. 2022, Art. no. 374, https://doi.org/10.3390/electronics11030374.

[25] C. D. Santos and B. Zadrozny, "Learning Character-level Representations for Part-of-Speech Tagging," in *Proceedings of the 31st International Conference on Machine Learning*, Beijing, China, Jun. 2014, pp. 1818–1826.

[26] X. Wu, "A Dropout Optimization Algorithm to Prevent Overfitting in Machine Learning," *Machine Learning Theory and Practice*, vol. 4, no. 1, Mar. 2023, https://doi.org/10.38007/ML.2023.040103.

[27] H. Xia, Y. Yang, X. Pan, Z. Zhang, and W. An, "Sentiment analysis for online reviews using conditional random fields and support vector machines," *Electronic Commerce Research*, vol. 20, no. 2, pp. 343–360, Jun. 2020, https://doi.org/10.1007/s10660-019-09354-7.

[28] Z. Gao, Z. Li, J. Luo, and X. Li, "Short Text Aspect-Based Sentiment Analysis Based on CNN + BiGRU," *Applied Sciences*, vol. 12, no. 5, Jan. 2022, Art. no. 2707, https://doi.org/10.3390/app12052707.

[29] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, Seattle, WA, USA, May 2004, pp. 168–177, https://doi.org/10.1145/1014052.1014073.

[30] Md. H. Alam, W.-J. Ryu, and S. Lee, "Joint multi-grain topic sentiment: modeling semantic aspects for online reviews," *Information Sciences*, vol. 339, pp. 206–223, Apr. 2016, https://doi.org/10.1016/j.ins.2016.01.013.

[31] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, Portland, OR, USA, Mar. 2011, pp. 142–150.

[32] Y. Sharma, G. Agrawal, P. Jain, and T. Kumar, "Vector representation of words for sentiment analysis using GloVe," in *2017 International Conference on Intelligent Communication and Computational Techniques (ICCT)*, Jaipur, India, Sep. 2017, pp. 279–284, https://doi.org/10.1109/INTELCCT.2017.8324059.

[33] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, "Sentiment Analysis Using Word2vec And Long Short-Term Memory (LSTM) For Indonesian Hotel Reviews," *Procedia Computer Science*, vol. 179, pp. 728–735, Jan. 2021, https://doi.org/10.1016/j.procs.2021.01.061.

[34] A. Zhao and Y. Yu, "Knowledge-enabled BERT for aspect-based sentiment analysis," *Knowledge-Based Systems*, vol. 227, Sep. 2021, Art. no. 107220, https://doi.org/10.1016/j.knosys.2021.107220.

[35] H. Zhao, Z. Liu, X. Yao, and Q. Yang, "A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach," *Information Processing & Management*, vol. 58, no. 5, Sep. 2021, Art. no. 102656, https://doi.org/10.1016/j.ipm.2021.102656.

[36] M. Rizwan *et al.*, "Aspect-Based Sentiment Analysis for Social Multimedia: A Hybrid Computational Framework," *Computer Systems Science and Engineering*, vol. 46, no. 2, pp. 2415–2428, 2023, https://doi.org/10.32604/csse.2023.035149.

[37] J. Mutinda, W. Mwangi, and G. Okeyo, "Sentiment Analysis of Text Reviews Using Lexicon-Enhanced Bert Embedding (LeBERT) Model with Convolutional Neural Network," *Applied Sciences*, vol. 13, no. 3, Jan. 2023, Art. no. 1445, https://doi.org/10.3390/app13031445.