



ARTICLE

Sentiment Analysis Using E-Commerce Review Keyword-Generated Image with a Hybrid Machine Learning-Based Model

Jiawen Li^{1,2}, Yuesheng Huang¹, Yayi Lu¹, Leijun Wang^{1,*}, Yongqi Ren¹ and Rongjun Chen¹

¹School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, 510665, China

²Hubei Province Key Laboratory of Occupational Hazard Identification and Control, Wuhan University of Science and Technology, Wuhan, 430065, China

*Corresponding Author: Leijun Wang. Email: wangleijun@gpnu.edu.cn

Received: 10 April 2024 Accepted: 14 June 2024 Published: 18 July 2024

ABSTRACT

In the context of the accelerated pace of daily life and the development of e-commerce, online shopping is a mainstream way for consumers to access products and services. To understand their emotional expressions in facing different shopping experience scenarios, this paper presents a sentiment analysis method that combines the e-commerce review keyword-generated image with a hybrid machine learning-based model, in which the Word2Vec-TextRank is used to extract keywords that act as the inputs for generating the related images by generative Artificial Intelligence (AI). Subsequently, a hybrid Convolutional Neural Network and Support Vector Machine (CNN-SVM) model is applied for sentiment classification of those keyword-generated images. For method validation, the data randomly comprised of 5000 reviews from Amazon have been analyzed. With superior keyword extraction capability, the proposed method achieves impressive results on sentiment classification with a remarkable accuracy of up to 97.13%. Such performance demonstrates its advantages by using the text-to-image approach, providing a unique perspective for sentiment analysis in the e-commerce review data compared to the existing works. Thus, the proposed method enhances the reliability and insights of customer feedback surveys, which would also establish a novel direction in similar cases, such as social media monitoring and market trend research.

KEYWORDS

Sentiment analysis; keyword-generated image; machine learning; Word2Vec-TextRank; CNN-SVM

1 Introduction

The development of e-commerce and the accelerating pace of daily life have established online shopping as a mainstream way to access services and products [1], as the e-commerce platform facilitates convenient home shopping. Despite its convenience, challenges like the differences between actual items and their descriptions, poor quality, and insufficient after-sales services persist [2]. Usually, on the e-commerce platform, consumers show a lively interest in product reviews for evaluations. These reviews, containing valuable insights into consumer experiences, serve as a vital source of cognitive value, such as considering them before purchasing decisions and seeking information on the advantages and disadvantages to refine their intentions [3]. Previous studies [4–6] have emphasized



the importance of these reviews in shaping consumer behavior, as they offer firsthand accounts of product performance and satisfaction. However, there exists a research gap in the depth and accuracy of understanding consumer emotional expressions, especially in how to effectively utilize consumer-generated content to gain insights into consumer behavior. In this regard, the sentiment analysis of the e-commerce reviews provides deep insights into consumer emotions, assisting in understanding the impact on purchasing decisions by mitigating uncertainty and risk. Meanwhile, businesses can obtain more information concerning consumer emotional expressions. This, in turn, can lead to improved customer satisfaction, enhance product offerings, and offer a more responsive approach to market dynamics. Therefore, the employment of Artificial Intelligence (AI) delivers prompt properties by analyzing the e-commerce reviews utilizing machine learning and deep learning techniques [7], which help to apply positive sentiments for marketing, optimize advertising strategies, and predict trends advance contributes to a customer-centric mode in e-commerce [8–10].

Currently, it is considered that analyzing contextual semantic information is beneficial for interpreting emotions from online reviews [11]. To this end, two typical approaches have been developed, one is the supervised method [12], and another is the lexicon-based method [13], where the first one employs labeled training data to establish a well-fitted model and adopts the machine learning classifier to test the dataset. The second one involves using predefined lists of words, referred to as lexicons or sentiment dictionaries, to recognize the emotions from a piece of text [14]. Then, each word in the lexicon is correlated with an emotional score like negative or positive. Therefore, the entire text can be computed based on the cumulative scores of its component words [15]. Besides, the topic commonly represents the essential properties conveyed in the text data, meaning its analysis aids in obtaining valuable clues. So, statistical-based approaches are also widely applied, such as word frequency co-occurrence matrix, word frequency statistics, and synonym forest [16–19].

Undoubtedly, the key to accurate sentiment analysis lies in feature extraction, which directly influences the understanding of the online review, as well as the classification ability. However, the main limitation of conventional methods is that text recognition alone improperly captures the complex emotions and subtleties within the text. This challenge has been highlighted in several previous studies [12,16], which reveal that incorporating high-level characteristics can enhance performance. In this regard, if the properties of review data are further enriched, more details that are beneficial to the classification can be found. Then, with the help of a machine learning-based model, sentiment analysis can be accomplished through such valuable features. Therefore, the motivation of this paper is to propose a sentiment analysis method that uses the text-to-image approach, where the visual data derived from the e-commerce review keyword captures additional layers of emotional context that pure text analysis might miss, which offers a novel perspective for customer experience analysis, addressing the limitations of conventional methods and paving the way for accurate classification. To achieve this goal, inspired by the existing image-based studies, we first apply Word2Vec-TextRank to generate sentence representations from the reviews and extract keyword features. Subsequently, these keywords act as the inputs to generate the corresponding images by generative AI, so the keyword-generated images are obtained, which enrich the text to the image accordingly. Finally, by feeding these images into a hybrid Convolutional Neural Network and Support Vector Machine (CNN-SVM) model, the sentiment classification can be realized, which brings an innovative solution for analyzing sentiment in online product reviews, also enhancing the reliability of feedback analysis in similar cases like social media monitoring and market trend research. For better illustration, Fig. 1 presents the overview of the proposed method, including data collection, text preprocessing, keyword feature extraction, text-to-image generation, and sentiment classification. In particular, this paper has the following contributions:

- 1) Novel text-to-image approach: proposing a sentiment analysis method that converts e-commerce reviews into images using generative AI, which helps to capture additional layers of emotional context that pure text analysis might miss.
- 2) Enhanced feature extraction: utilizing Word2Vec-TextRank for generating sentence representations and extracting keyword features, which enrich the reviews with visual properties.
- 3) Hybrid classification model: implementing a hybrid CNN-SVM model that processes the keyword-generated images for sentiment classification, offering a reliable solution for analyzing sentiment in online product reviews with a remarkable accuracy of up to 97.13%.

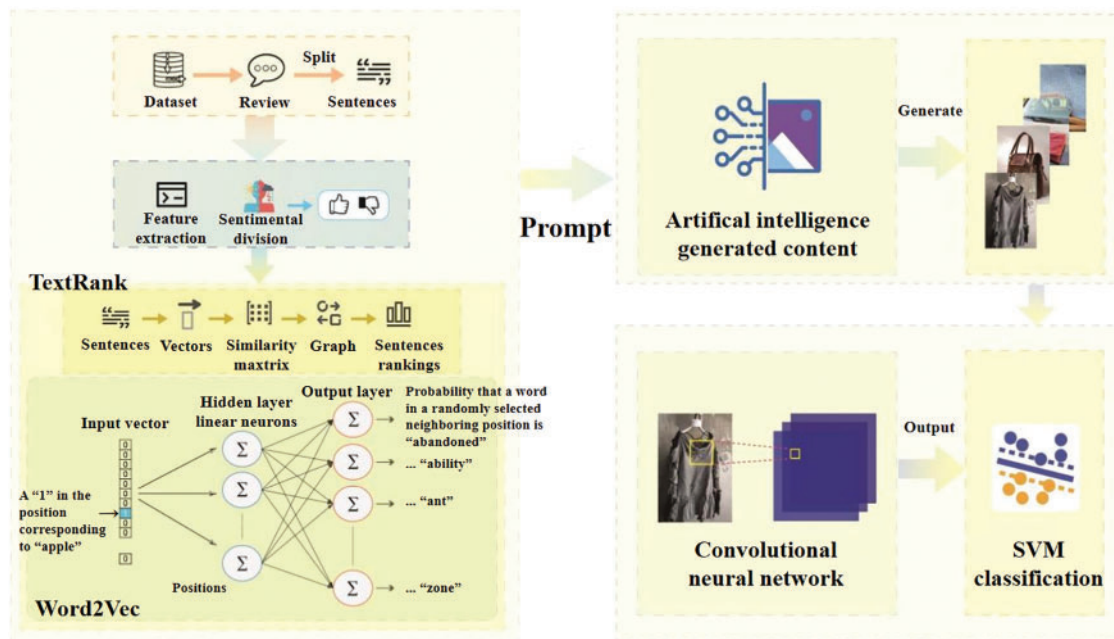


Figure 1: Overview of the proposed sentiment analysis method

The rest is organized as follows: [Section 2](#) presents the related works. [Section 3](#) describes the data collection and the proposed method. [Section 4](#) shows the experimental results from the online product reviews on Amazon. [Section 5](#) discusses the properties of the proposed method based on the results and a comparative study. Finally, the conclusions are drawn in [Section 6](#).

2 Related Works

Previously, numerous studies have integrated various computational models employing Natural Language Processing (NLP) and text mining to detect individual emotions and opinions. Mihalcea et al. [20] designed the TextRank algorithm, a graph-based text summarization method that represents words or phrases as nodes in a graph, with edge weights capturing semantic similarity. This algorithm has laid a significant foundation in NLP. However, it lacks true semantic understanding, struggles with varying text lengths and structures, and has difficulty handling synonyms or polysemy effectively due to its reliance on preprocessing quality and statistical co-occurrence without deeper linguistic analysis. Mikolov et al. [21] introduced Word2Vec, a method for learning distributed word representations by capturing semantic relationships within a continuous vector space. While powerful for generating word embeddings, Word2Vec requires substantial computational resources and large

amounts of training data to produce high-quality embeddings. Besides, it can be challenging to capture long-range dependencies in text. Huang et al. [22] developed a polymerization topic sentiment model for analyzing text in online product reviews, where semantic information is extracted and filtered from the reviews. By integrating this with machine learning-based classifiers, precise classification of emotions can be achieved, demonstrating that the topics hidden in the review data significantly influence sentiment analysis. Nevertheless, this method relies heavily on high-quality data and involves complex and computationally intensive semantic extraction. Zhang et al. [23] applied cognitive appraisal theory for sentiment classification by combining SVM with latent semantic analysis, while Obiedat et al. [24] presented a hybrid model that combines Particle Swarm Optimization (PSO) with SVM, alongside diverse oversampling methods to address imbalanced data issues in sentiment analysis. These two approaches enhance the performance of sentiment classification but introduce additional complexity and computational demands. By comparing the above techniques, it is clear that while advancements in NLP and text mining have significantly improved sentiment classification, challenges remain in accomplishing semantic understanding, managing computational complexity, and ensuring the scalability of these models across varied domains.

On the other hand, deep learning-based approaches have been extensively employed in sentiment analysis, including Deep Neural Network (DNN) [25], CNN [26], Recurrent Neural Network (RNN) [27], and attention mechanism-based network [28]. These models excel at learning intricate patterns from online review data, making them well-suited for sentiment classification tasks. CNN, in particular, has demonstrated exceptional capability in segmenting and classifying images and achieved significant success in NLP tasks due to its proficiency in processing spatial data. One well-known application of CNN in sentiment analysis was proposed by Meena et al. [29], who aimed to classify sentiment polarity in social media data. They categorized comments preferred by people of different ethnicities into sentiment polarities such as positive, negative, and neutral, achieving an impressive accuracy of 95.4%. Similarly, Kruspe et al. [30] conducted sentiment analysis on European COVID-19-related Twitter messages using a neural network with pre-trained word and sentence embeddings, incorporating skip-gram Word2Vec and multilingual Bidirectional Encoder Representations from Transformers (BERT). This model analyzed a total of 4.6 million tweets and identified 79,000 of them containing COVID-19 keywords with semantic information. Alharbi et al. [31] performed sentiment analysis on Amazon reviews using various RNN variants, including Long Short-Term Memory (LSTM), group LSTM, Gated Recurrent Unit (GRU), and Update Recurrent Unit (URU), to classify customer sentiment as negative, neutral, or positive. These RNNs were combined with different word embeddings (e.g., GloVe, Word2Vec, FastText) for feature extraction, with the group LSTM-based model and FastText achieving the highest accuracy of 93.75%. Bansal et al. [32] proposed a hybrid attribute-based sentiment classification method that integrates Optical Character Recognition (OCR) sentiment orientation by investigating implicit keyword relationships and domain-specific characteristics, validated on Amazon mobile phone reviews and TripAdvisor hotel reviews. Alzahrani et al. [33] studied LSTM and CNN-LSTM for sentiment analysis on Amazon reviews. After preprocessing the data through lowercase conversion, stop-word and punctuation removal, and tokenization, they employed these two models to train the cleaned data for sentiment classification. Mohbey [34] utilized an LSTM model to predict customer review sentiments, accomplishing an accuracy of 93.66%, which reveals the superiority of deep learning-based approaches due to their abilities to handle extensive real-time data and robust feature extraction results.

Furthermore, aspect-based sentiment analysis has a wide range of applications in consumer behavior analysis, market research, and product development. It focuses on identifying and extracting specific emotions or opinions related to particular entities or attributes, known as "aspects," within

the text. Unlike traditional sentiment analysis, which typically only determines whether the overall sentiment of the text is positive, negative, or neutral, aspect-based analysis delves deeper into recognizing the specific sentiments directed at different aspects of an entity. For instance, Hajek et al. [35] developed a fake review detection model using an aspect-based method while considering the impact of product types. Utilizing a dataset of Amazon reviews, this model revealed that two aspects, the product category, and the verified purchase attribute, are useful for detecting fake reviews, with the greatest contribution observed for credence and experience product types. In another work, Chen et al. [36] presented an attention-based deep learning approach to capture semantic information. They achieved good results by incorporating syntactic information, which is noteworthy for understanding the structure of sentences. This model integrates a Graph Convolutional Network (GCN) and a co-attention mechanism to handle aspect-based information and eliminate noise from irrelevant contextual words. It allows both semantic and syntactic information to be conveyed to the sentiment analysis, improving the overall accuracy. The aforementioned works indicated that aspect-based methods, particularly when combined with advanced deep learning techniques like attention mechanism and GCN, can provide insightful depth of sentiment analysis.

Both machine learning-based and deep learning-based methods find extensive applications in sentiment analysis, each catering to different needs and scenarios. Machine learning-based methods are well-suited for rapid development and situations where interpretability is vital, making them appropriate to smaller datasets or limited computational resources. Deep learning-based methods excel in handling large-scale datasets and intricate pattern recognition tasks, particularly in capturing the nuanced semantic information within text. However, they require substantial labeled data and computational resources, while the interpretability of deep learning models remains an ongoing challenge. The choice between these methods depends on the specific project requirements, including data size and complexity, resource availability, and the importance of model interpretability. In this regard, a hybrid model is preferred, combining machine learning for feature extraction with deep learning for intricate pattern recognition and comprehensive sentiment analysis, which is also the main purpose of this paper.

3 Materials and Methods

3.1 Data Collection

The data applied in this paper was collected from Amazon from September 2021 to April 2023, totaling 179,673, including customer ID, review time, overall rating, customer name, product review data, and ASIN, where the ASIN is a unique identifier assigned by Amazon to its products, consists of 10 characters containing letters and numbers. For sentiment analysis, the aim is to categorize the review data as positive or negative. To this end, those reviews with ratings of 5 and 4 were regarded as positive, indicating that customers perceive the product can satisfy their expectations and needs. Conversely, the reviews with ratings of 1 and 2 were marked as negative, meaning that customers found the product unsatisfactory, resulting in a dissatisfied view. Based on that, 162,243 reviews were labeled. In order to achieve the subsequent training and testing through a time-saving manner, 2500 reviews were randomly chosen from the positive and negative labels, respectively, i.e., 5000 reviews were applied for method validation in this paper.

3.2 Text Preprocessing

Text preprocessing is a vital step to eliminate redundant vocabulary and noise interference. In this paper, the first stage is text segmentation, which involves splitting the acquired data into JavaScript

Object Notation (JSON) format to extract the review text required. To this end, we utilize Python's built-in JSON module to parse and process the JSON data. Then, the JSON string is converted to a Python dictionary and segmented using the slicing method. The second stage is case folding, which entails converting all uppercase and lowercase letters to lowercase. We employ Python's built-in lower() to achieve this goal. The third stage is deactivated word filtering, which refers to removing common words that regularly occur in the text but are meaningless to sentiment analysis. For instance, words like "a", "the", and "is" often lack meaningful information. Hence, the size of the review data can be reduced to improve sentence representation, enabling the selection of more valuable keyword feature from the text. We use a list of English stopwords downloaded from the nltk library for this operation. The next stage is tokenization, which segments text into tokens, providing finer-grained input for text-processing tasks. This process aids in better understanding and processing of review data. The last stage is lexical stemming, which reduces words to their base forms, addressing morphological variations such as plurals and tenses. For example, "dogs" is reduced to "dog". WordNet, an English vocabulary database derived from the nltk library, is involved in performing the lexical stemming.

3.3 Keyword Feature Extraction

Word2Vec is a neural word embedding model that learns representations by predicting target words from surrounding words. It can be accomplished through two ways: the Continuous Bag of Words (CBOW) model and the continuous skip-gram model. Both of them focus on reducing the dimensionality of the data and making dense word vectors [37]. Besides, the jump grid assigns more weight to nearer context words compared to more distant context words. Therefore, it can predict the center word with the help of a weighted window of surrounding words [38]. If a sequence of words $W = \{w_1, w_2, \dots, w_N\}$, its learning objective function can be expressed by maximizing the log-likelihood probability of (1) and (2):

$$L_{CBOW}(W) = \frac{1}{N} \sum_{i=1}^N \log \Pr(w_i | w_{ctx}) \quad (1)$$

$$L_{Skip-gram}(W) = \frac{1}{N} \sum_{i=1}^N \sum_{-c \leq j \leq c, j \neq 0} \log \Pr(w_{i+j} | w_i) \quad (2)$$

where w_{ctx} denotes the context word of the current word w_i , c refers to the window size of the context, and the conditional probability $\Pr(w_i | w_j)$ can be obtained by (3):

$$\Pr(w_i | w_j) = \frac{\exp(w_i \cdot w_j)}{\sum_{w \in W} \exp(w \cdot w_j)} \quad (3)$$

However, a main limitation of Word2Vec is that it determines a word to be more semantically similar to its neighboring words, although the surrounding words may not exhibit semantic similarity to the current word. To solve it, the TextRank algorithm, an unsupervised automatic keyword extraction technique, is applied. Its primary aim is to decide the most relevant keyword by assessing the importance of each vocabulary node in the text. During the construction of the weighted undirected graph, the edge weights between vocabulary nodes are computed employing various methods, such as the co-occurrence matrix and the cosine similarity [39]. Meanwhile, through the iterative calculation process of PageRank, the score value of each vocabulary node is continuously updated until it

converges to a stable state. Mathematically, the TextRank is represented as (4):

$$WS(V_i) = (1 - d) + d * \sum_{V_j \in \ln^*(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j) \quad (4)$$

where $WS(V_i)$ denotes the weight of node V_i , $\ln(V_i)$ means the set of nodes pointing to V_i , w_{ji} is the weight of the edge between the two nodes, $WS(V_j)$ refers to the weight of node V_j , and d is the damping coefficient, which can be usually set at 0.85.

When a keyword appears in the vocabulary list, its weight will be multiplied by 1.5 on top of the original weight. The construction of the edge set in the graph considers both the positional information of the words, as well as the semantic similarity between them, to construct the transformation matrix. The weight transfer between any two nodes in the graph is realized by (5):

$$w_{ij} = ft(i, j) + rev(i, j) \quad (5)$$

where the eigenvalue $ft(i, j)$ is shown as (6):

$$ft(i, j) = vsim(i, j) * coc(i, j) \quad (6)$$

with $coc(i, j)$ refers to the co-occurrence frequency of two words, $vsim(i, j)$ indicates the semantic similarity between V_i and V_j , which can be computed using the cosine similarity presented in (7):

$$vsim(i, j) = \cos(\theta) = \frac{v_i \cdot v_j}{\|v_i\| \|v_j\|} \quad (7)$$

In constructing a word graph model, co-occurrence frequency is an important parameter. The word graph model organizes the words in the text into a network structure, with words as nodes and the semantic relationships between them represented by boundaries. Typically, co-occurrence frequency measures the connection weights between words based on the number of co-occurrences, which helps to construct a more accurate word graph model.

Besides, $rev(i, j)$ in (5) represents the importance of the words in the text, which can be calculated by the following (8) and (9):

$$rev(i, j) = \frac{p_j}{\sum_{k \in Out(v_j)} p_k} \quad (8)$$

$$p_j = \begin{cases} 1.0, & j \text{ appears at the beginning of the paragraph} \\ 0.8, & \text{else} \end{cases} \quad (9)$$

where p_j is the weight of the occurrence position. When j appears at the beginning of a paragraph, its weight will be assigned as 1.0, and in other positions, it will be 0.8. This is because the opening word is likely to be the beginning of a paragraph topic or keyword. Conversely, in other positions, it is relatively less important.

The Word2Vec-TextRank integrates the representation of word vectors from Word2Vec with the graph-theoretic algorithm of TextRank to discern pivotal concepts within texts. In particular, Word2Vec generates dense vectors that encapsulate the contextual meaning of words through extensive learning from textual data. Then, TextRank constructs a directed graph where nodes denote words and edge weights are determined by the similarity between these vectors. Through the iterative computation, TextRank assigns an importance score to each word and delineates the keywords. As a result, the main advantage of Word2Vec-TextRank for keyword extraction is that it can take into

account both the semantic relationships and contextual information. That means this combination transcends mere word frequency, showing the semantic roles and contextual significance of words within the text, which allows for the precision of keyword identification by incorporating semantic and contextual cues.

3.4 Text-to-Image Generation

After the feature extraction, we can obtain a series of keywords related to emotions and products that provide the consumer's point of view. In order to enrich the properties of keywords and represent them as image data, a stable diffusion generative AI is trained using the pre-labeled images derived from the keywords. Subsequently, we input the extracted keywords into this generative AI to gain a series of point-to-point images, realizing text-to-image generation.

To this end, first, the generative AI model gradually expands on the low-resolution images to form high-resolution images, in which the model fine-tunes the details of the generated images according to the inputted keywords so that they can correctly express the meanings represented by the keywords. Second, we compare the generated images with the inputted keywords to assess the performance in terms of text-to-image generation, i.e., well-matched with the keywords or not. This step requires not only examining the correspondences between the images and the keywords but also assessing the quality of the generated images. Finally, we accomplish the transformation from natural language to visualized images, providing the text-to-image source for subsequent sentiment classification. In short, the phase of generating images is the core of the proposed method, where the keywords extracted by Word2Vec-TextRank are employed as the inputs, and the generative AI is adopted to generate images associated with the inputs.

3.5 Sentiment Classification

CNN is a feed-forward neural network triggered by biological visual cognitive mechanisms. By accepting raw graphical data and avoiding complex preprocessing steps, CNN has a wide scope of applications. In this paper, it is the fundamental component utilized within the neural network for processing the keyword-generated images, as displayed in Fig. 2.

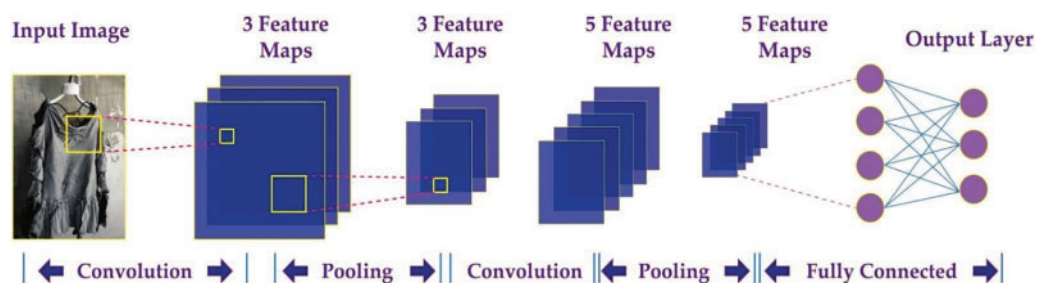


Figure 2: The CNN used for processing the keyword-generated images

In Fig. 2, the keyword-generated images act as inputs to the neural network. Concerning the convolutional layer, a kernel is used to perform a convolutional operation on the input image to obtain the features of this image. As for the outcome of the convolutional layer, a nonlinear activation function is necessary, and we adopt the ReLU function to increase the nonlinearity of the network. Next, regarding the pooling layer, downsampling is performed and maximum pooling is included to decrease the dimensionality of the feature map and retain vital semantic information. Finally, the

feature maps are spread in the fully connected layer, and the outputs are denoted as feature vectors, which can be used to train the subsequent SVM model for conducting the sentiment classification.

Regarding the SVM, its principle is to determine an optimal hyperplane that maximizes the separation between samples of different types. For linearly divisible datasets (i.e., perceptrons), there are infinitely hyperplanes. However, the separating hyperplane with the largest geometric spacing is unique. Therefore, the SVM is to identify the support vectors, i.e., the sample points closest to the hyperplane. These support vectors determine the location and orientation of the hyperplane and play a vital role in determining the optimal hyperplane. That means the hyperparameter optimization is crucial for SVM. To this end, it is essential to identify the optimal combination of hyperparameters tailored to the characteristics of different feature vectors, ensuring a better fit to the training data for improving predictive capability. Meanwhile, optimizing hyperparameters maximizes model performance while mitigating the risks of overfitting or underfitting, which helps to enhance generalization to unseen data and saves time and computational resources. In this regard, fine-tuning key hyperparameters, such as the penalty parameters (C values) and kernel function parameters, can lead to optimal results, which improve the robustness and classification performance in sentiment analysis [40]. To achieve this goal, we implement hyperparameter optimization by the GridSearchCV in the sklearn library, which forms a parameter grid by exhaustively enumerating the given candidate hyperparameters and then cross-validates each set of hyperparameters to assess the performance. Hence, by comparing different combinations, the best one can be selected as the optimal configuration.

In short, we capitalize on the image feature extraction capabilities from CNN and the discriminative power from SVM to enhance accuracy. This hybrid CNN-SVM architecture offers a synergistic approach to sentiment classification, combining the strengths of both models to realize improved performance.

4 Experimental Results

4.1 Review Data Results

The data initially collected 179,673 product reviews and then performed preprocessing, where the reviews with a rating of 3 were removed, resulting in 162,243 data. These reviews were subsequently labeled as positive and negative, respectively. For training and testing through a time-saving way, we randomly selected 2500 reviews each for the positive and negative emotions, resulting in 5000 data for method validation, which also guarantees a balanced sample distribution. Table 1 lists the examples of the collected product reviews from Amazon.

Table 1: The examples of the collected product reviews from Amazon

Customer ID	Review data	Overall rating	Category
A14BTJRH9VNLJJ	One CD, . . . , I think that you will like this one as well. Buy it!	5	Positive
A3DL686B8JEM8A	People tend to glorify their rock and roll idols, . . . , Overall 10 out of 10.	5	Positive

(Continued)

Table 1 (continued)

Customer ID	Review data	Overall rating	Category
AVOKN4NDZCN78	Don't last but a couple of weeks,..., Will never buy it again even if only 5 bucks.	1	Negative
AHF7SNSZPRNCE	Second one I bought, . . . , but these are junk.	1	Negative

Word clouds are commonly used as a tool for analyzing qualitative sentiment data, helping to visualize which words are most frequent in consumer reviews. Now, we concentrate on the preprocessed review data (i.e., de-deactivated words) and the words after word-splitting for frequency counting. The results of word clouds are depicted in Fig. 3.

In Fig. 3, the common words in the word clouds like “show”, “season”, “one”, “like”, “good”, “great”, and “series” suggest that consumers focus on aspects like product presentation, seasonal demand, individual product ratings, as well as their satisfaction and liking of the product in their reviews. Thus, it indicates that the data collection for sentiment classification is proper.

**Figure 3:** Word clouds of sentiment data from the online product review on Amazon

4.2 Keyword Extraction Results

Regarding the 5000 review data extracted, 70% is utilized as the training set, while the rest 30% serves as the test set, i.e., 3500 reviews as training data and 1500 as test data. Then, keywords are extracted by Word2Vec-TextRank, which is beneficial for acquiring the most representative keywords from the reviews. These keywords are adopted for establishing the word vectors, facilitating a comprehensive reflection of contextual semantic information. This process enables each review to bring its corresponding list characterized by representativeness. The keywords extracted from the examples of online product reviews are displayed in Table 2.

Table 2: The examples of online product reviews after keyword extraction

Customer ID	Review data	Overall rating	Keyword extraction
A2I6KBFBFIORH7	We bought,..., them continually ever since,..., have been well worth the cost.	Positive	'good', 'product', 'excellent', 'customer', 'service'
A30S5MTFNNOH0	My daughter,..., the kids, and the dog are now getting their own.	Positive	'great', 'idea', 'keep', 'animals', 'kids'
A1TJ393OFP21I8	Total flimsy junk, . . . , DO NOT BUY THIS PIECE OF JUNK.	Negative	'terrible', 'wouldn't', 'guitar', 'stand', 'forget'
A3400KQV4QXWNQ	I don't know who designed the leg feature of this thing,..., just don't like the features of this stand at all.	Negative	'locking', 'legs', 'place', 'caused', 'scuffed'

4.3 Sentiment Analysis Results

The performance evaluation is presented by the confusion matrix assessed by True Negative (TN), True Positive (TP), False Negative (FN), and False Positive (FP). We can also employ them to calculate the typical evaluation metrics used in the classification task, including accuracy, precision, recall, F1-score, and specificity. It is improper to evaluate the robustness of a classification model solely by a specific metric. For example, in an unbalanced dataset, relying only on accuracy fails to indicate the performance adequately. In this regard, hyperparameter tuning becomes imperative to identify an optimal scenario across various evaluation metrics. That is why we need hyperparameter optimization. Besides, we evaluate various classifiers, including Random Forest (RF), Logistic Regression Classification (Logi), Decision Tree (DT), Naive Bayesian (NB), and SVM, aiming to validate the advances of using the hybrid CNN-SVM model as an optimized classifier in the proposed method. [Table 3](#) shows the evaluation metrics of each classifier, and [Fig. 4](#) illustrates the results of confusion matrices.

Table 3: The evaluation metrics using various classifiers and the proposed hybrid CNN-SVM model

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Specificity (%)
RF	94.00	94.84	93.36	94.09	94.67
Logi	94.07	95.57	92.71	94.12	95.49
DT	92.27	90.95	94.27	92.58	90.16
NB	88.53	84.49	95.05	89.46	81.69
SVM	94.93	95.77	94.27	95.01	95.63
CNN-SVM	97.13	97.73	96.57	97.15	97.71

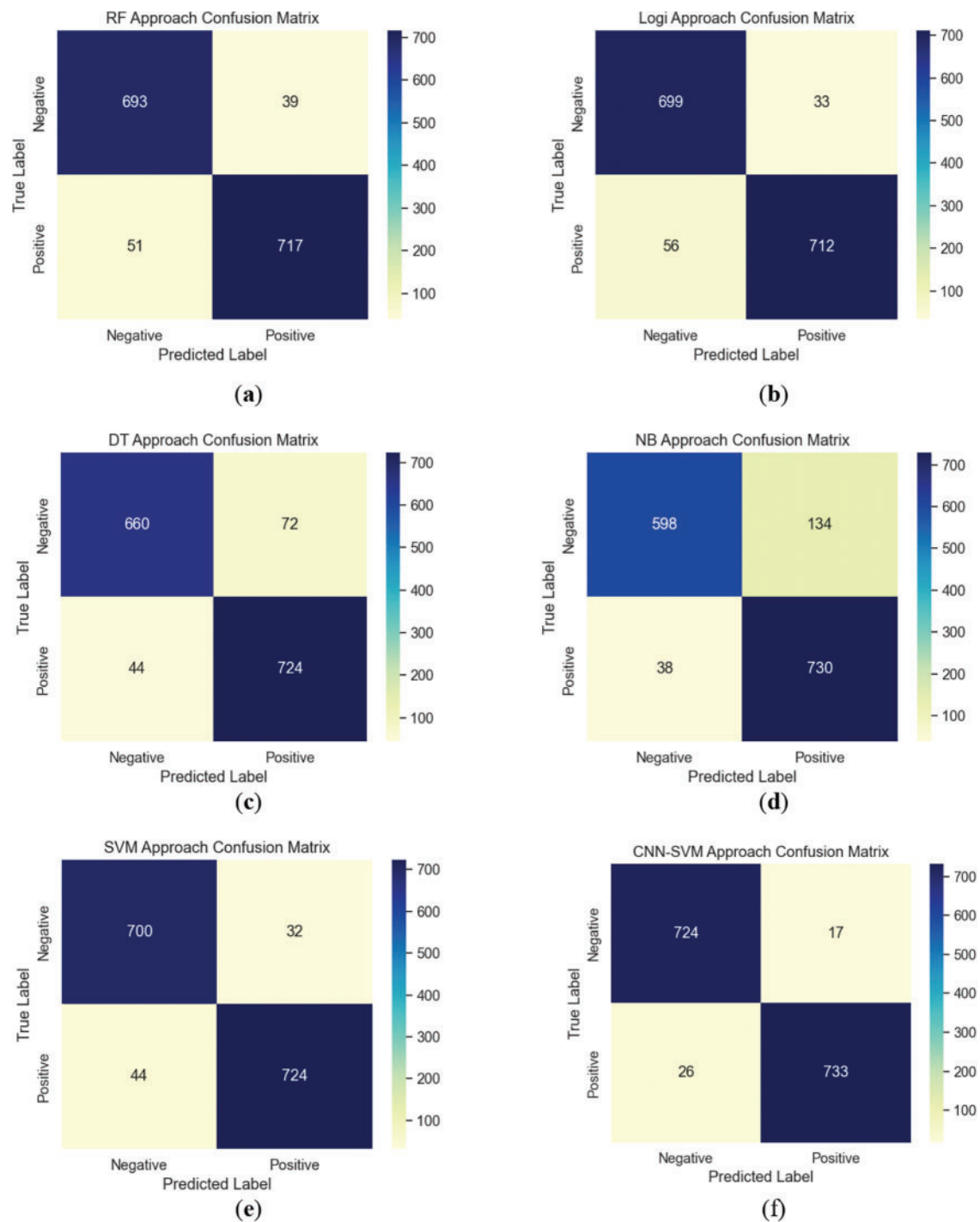


Figure 4: The confusion matrices using different classification methods: (a) RF; (b) Logi; (c) DT; (d) NB; (e) SVM; (f) CNN-SVM

In this evaluation, various classifiers adopt consistent feature extraction procedures, and the comparisons in Table 3 and Fig. 4 demonstrate that our method performs best, achieving 97.13%, 97.73%, 96.57%, 97.15%, and 97.71% on the accuracy, precision, recall, F1-score, and specificity, respectively. Such performances indicate that all the metrics of sentiment analysis are improved when CNN-SVM is employed in conjunction with the Word2Vec-TextRank.

Next, it is crucial to emphasize that hyperparameter tuning significantly impacts the results, and hyperparameter optimization can enhance the predictive capability and generalizability across various cases, so it is vital to assess them in detail. The hyperparameter optimization employing Python yields different results varied with penalty parameters (C values), as shown in Table 4. Here, by adjusting the C value, we can make a trade-off between overfitting and underfitting the model. Table 4 displays that the CNN-SVM model accomplishes the best when the C value is 1, which indicates that the proposed method offers remarkable ability in sentiment analysis for both positive and negative categories. It also reveals the importance of keyword-generated image, which can supply trustworthy features for sentiment classification.

Table 4: The results of CNN-SVM under various penalty parameters (C values)

Penalty parameters	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Specificity (%)
0.1	96.67	97.20	96.17	96.68	97.17
1	97.13	97.73	96.57	97.15	97.71
10	96.27	96.80	95.78	96.29	96.77
100	94.87	95.47	94.33	94.90	95.41

4.4 Training Loss

The hybrid CNN-SVM model is performed to train keyword-generated image derived from the Words2Vec-TextRank and generative AI. Then, as for the SVM, its kernel function is the Radial Basis Function (RBF) with penalty parameter $C = 1$. After 100 rounds of training, the loss values are stabilized. The loss and accuracy iteration curves are drawn in Fig. 5.

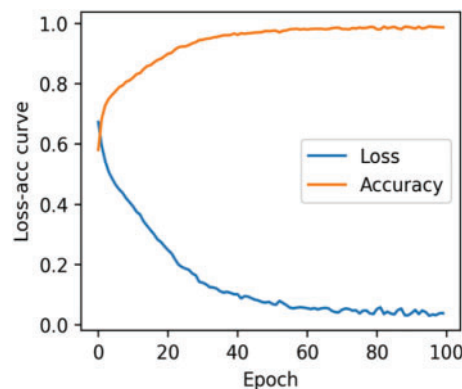


Figure 5: The loss and accuracy iteration curves of the proposed model

After the optimization, the parameters are determined, and the variation in the loss value is mostly affected by the structure of the model. In particular, the dynamic variation of the loss value reveals how well the model fits the training data, where a higher loss value implies that the model improperly describes the data accurately enough, while a lower loss value indicates higher accuracy. Fig. 5 demonstrates that the hybrid CNN-SVM model exhibits stability when the training iterations approach 50, and almost remains unchanged until 100. Consequently, it reaches convergence on the current training data and provides effectiveness for sentiment analysis.

5 Discussion

First, the choice of the hybrid CNN-SVM architecture for sentiment classification stems from the complementary strengths of both models, where CNN is adept at capturing local patterns and spatial relationships within the keyword-generated image data. On the other side, SVM exhibits its ability to handle high-dimensional feature spaces and non-linear decision boundaries. As a result, by integrating CNN and SVM, we employ the feature extraction capabilities of CNN to capture meaningful data representations, and SVM offers a robust mechanism for classifying the feature vectors into sentiment categories. It can be said that the hierarchical feature extraction process of CNN followed by the discriminative classification of SVM is beneficial for finding the sentiment patterns within the e-commerce review keyword-generated image, supplying more trustworthy features for improving sentiment classification accuracy accordingly. That is why the CNN-SVM model shows better than others.

Second, a comparative study with previous works is presented in Table 5. It indicates that the proposed model yields an impressive performance in the relevant tasks compared to traditional text-based sentiment analysis approaches, where almost all of them achieved the sentiment analysis by considering NLP and text mining techniques. The reason may be the limitation of text features in sentiment analysis based on online product review data since their susceptibility to noise and ambiguity inherent in natural language. Online product reviews often contain informal language, spelling errors, sarcasm, and other linguistic nuances that can affect the accuracy of sentiment analysis. Besides, text features may struggle to capture the sentiment expressed in images, videos, or other non-textual elements presented in online reviews. Now, we convert text into image, which aids in deep insights into emotions within online reviews. This advance can be attributed to the keyword-generated image based on Word2Vec-TextRank and generative AI, as the text-to-image feature effectively represents valuable details from large amounts of review data and reflects its semantic information during training, so improving the understanding and prediction of online product reviews.

Table 5: A comparative study with previous works

Work	Main methodology	Dataset	Highest accuracy (%)
Hammou et al. [27]	FastText with LSTM, BiLSTM, and GRU	Yelp and Twitter	93.28
Meena et al. [29]	Keras embedding with CNN	Twitter	95.40
Alharbi et al. [31]	FastText with LSTM	Amazon reviews	93.75

(Continued)

Table 5 (continued)

Work	Main methodology	Dataset	Highest accuracy (%)
Alzaharani et al. [33]	Keras text tokenizer with LSTM, and CNN-LSTM	Amazon reviews	94.00
Mohbey [34]	NB, SVM, decision tree, Logistic regression, and LSTM	Amazon reviews	93.66
Hajek et al. [35]	W2VLDA with DNN	Amazon reviews	82.89
Chen et al. [36]	GCN with co-attention mechanism	Twitter, Lap14, Rest14, Rest15, and Rest16	89.94
Labhsetwar [41]	Extra trees classifier, XGBoosting, and SVM	Telecommunication reviews	89.87
Li et al. [42]	Word2Vec with BERT + Fully connected (FC) layer	Stock reviews	92.65
Gaye et al. [43]	Regression Vector-stochastic gradient descent classifier (RV-SGDC)	Employee reviews	97.00
This work	Word2Vec-TextRank with hybrid CNN-SVM model	Amazon reviews	97.13

Furthermore, to demonstrate the superiority of the proposed approach, benchmarking using other state-of-the-art sentiment analysis methods has been conducted on the same 5000 reviews from Amazon evaluated in this paper, as summarized in Table 6. As seen, the proposed approach shows superiority by combining lightweight feature extraction with text-to-image generation and the hybrid model. Concerning the Word2Vec-TextRank applied in the feature extraction, it not only accomplishes lightweight text processing but also appropriately captures the core semantic information of the text. Compared with large-scale language models such as BERT [30,42] and Sentiment Knowledge Enhanced Pre-training (SKEP) [44], the Word2Vec-TextRank is resource-saving and performs well in keyword extraction. This is evident from the fact that the proposed method outperforms the two models in the evaluation metrics. Besides, SenticNet-based GCN [45] and Graph Attention Network Model Incorporating Syntactic, Semantic, and Knowledge (SSK-GAT) [46] have achieved impressive results previously, where SenticNet-based GCN shows good results in aspect-level sentiment analysis by utilizing sentiment knowledge enhancement through an augmented GCN, while SSK-GAT employs graph attention networks and focuses on aspect-level sentiment classification. In this regard, the proposed approach outperforms them, mainly due to the innovative feature extraction that performs on keyword-generated image, which helps the model to concentrate on learning the most representative and distinguishing information in both visual and textual sides. Consequently, it can be said that text-to-image is a key point to improving sentiment analysis in the proposed method, especially when the dataset has a small size of 5000.

Table 6: Benchmarking using other state-of-the-art sentiment analysis methods

	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Specificity (%)
BERT [30,42]	90.68	90.37	89.32	89.47	90.68
SKEP [44]	88.52	88.97	87.84	88.12	89.20
SenticNet-based GCN [45]	90.84	91.43	90.16	90.43	91.52
SSK-GAT [46]	92.08	92.65	91.40	91.97	92.76
This work	97.13	97.73	96.57	97.15	97.71

Finally, although we have achieved relatively remarkable results, the proposed method still has potential for improvement in the future. On one hand, the use of generative AI models to convert keywords into images faces challenges in consistency and quality control, as its process is dependent on the size of pre-labeled images derived from the keywords. On the other hand, while the dataset of 5,000 reviews from Amazon provides a solid foundation for this study, its specificity limits the generalizability of the findings. Thus, we would like to collect a large dataset of reviews from a variety of e-commerce platforms, such as Flipkart, Alibaba, Shopify, and eBay, to comprehensively assess the results in recognizing more complex sentiments concerning online feedback of product reviews.

6 Conclusions

In this paper, we contribute a sentiment analysis method of online product reviews using a hybrid machine learning-based model with text-to-image features, where TextRank is applied to acquire the importance of sentences throughout the text, and Word2Vec is used to extract the representations by predicting the target word from the surrounding words. The keyword features are then adopted to generate images through generative AI, so the keyword-generated images can be made into the inputs for training the CNN-SVM model for sentiment classification, along with determining an optimal hyperplane that maximizes the margins between samples of different sentiment categories.

As for the experiments, the reviews of various products on Amazon from September 2021 to April 2023, totaling 179,673 reviews, are collected at first. Considering the sample balance and time-saving issues, we randomly used 2500 positive and negative reviews each, totaling 5000 data for method validation. Besides, Python is used for programming based on the keras library and the sklearn library to process the reviews data and classify them into positive and negative categories. The results demonstrate that the proposed method holds the advantage of recognizing the keyword-generated image, accomplishing 97.13%, 97.73%, 96.57%, 97.15%, and 97.71% on the accuracy, precision, recall, F1-score, and specificity, respectively, which realizes the robustness to analyze the e-commerce reviews. As a result, it can be concluded that the text-to-image manner is beneficial for data representations of the e-commerce reviews, and the hybrid CNN-SVM model is appropriate to improve the classification performance through the keyword-generated image. This brings an innovative solution to enhance the accuracy and depth of sentiment analysis, and also establish a direction in conducting similar cases, such as social media monitoring and market trend research. In the future, several state-of-the-art models in emotion recognition will be employed to test the keyword-generated image, which expects high performance while reducing computational time and cost.

Acknowledgement: The authors would like to appreciate the special support from Digital Content Processing and Security Technology of Guangzhou Key Laboratory.

Funding Statement: This work was supported in part by the Guangzhou Science and Technology Plan Project under Grants 2024B03J1361, 2023B03J1327, and 2023A04J0361, in part by the Open Fund Project of Hubei Province Key Laboratory of Occupational Hazard Identification and Control under Grant OHIC2023Y10, in part by the Guangdong Province Ordinary Colleges and Universities Young Innovative Talents Project under Grant 2023KQNCX036, in part by the Special Fund for Science and Technology Innovation Strategy of Guangdong Province (Climbing Plan) under Grant pdjh2024a226, in part by the Key Discipline Improvement Project of Guangdong Province under Grant 2022ZDJS015, and in part by the Research Fund of Guangdong Polytechnic Normal University under Grants 22GPNUZDJS17 and 2022SDKYA015.

Author Contributions: Study conception and design: Jiawen Li, Yuesheng Huang, Yayi Lu; data collection: Yuesheng Huang, Yongqi Ren; analysis and interpretation of results: Jiawen Li, Leijun Wang, Rongjun Chen; draft manuscript preparation: Jiawen Li, Yuesheng Huang, Yayi Lu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available in <https://github.com/Yorkson-huang/Text-to-image-SA> (accessed on 17/05/2024).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment analysis for e-commerce product reviews in Chinese based on sentiment lexicon and deep learning," *IEEE Access*, vol. 8, pp. 23522–23530, 2020. doi: [10.1109/ACCESS.2020.2969854](https://doi.org/10.1109/ACCESS.2020.2969854).
- [2] R. Liang and J. Q. Wang, "A linguistic intuitionistic cloud decision support model with sentiment analysis for product selection in e-commerce," *Int. J. Fuzzy Syst.*, vol. 21, no. 3, pp. 963–977, 2019. doi: [10.1007/s40815-019-00606-0](https://doi.org/10.1007/s40815-019-00606-0).
- [3] P. Ji, H. Y. Zhang, and J. Q. Wang, "A fuzzy decision support model with sentiment analysis for items comparison in e-commerce: The case study of <http://PConline.com>," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 49, no. 10, pp. 1993–2004, 2018. doi: [10.1109/TSMC.2018.2875163](https://doi.org/10.1109/TSMC.2018.2875163).
- [4] P. Puengwattanapong and A. Leelasantham, "A holistic perspective model of plenary online consumer behaviors for sustainable guidelines of the electronic business platforms," *Sustainability*, vol. 14, no. 10, pp. 6131, 2022. doi: [10.3390/su14106131](https://doi.org/10.3390/su14106131).
- [5] E. G. Dias, L. K. de Oliveira, and C. A. Isler, "Assessing the effects of delivery attributes on e-shopping consumer behaviour," *Sustainability*, vol. 14, no. 1, pp. 13, 2022. doi: [10.3390/su14010013](https://doi.org/10.3390/su14010013).
- [6] N. Punetha and G. Jain, "Bayesian game model based unsupervised sentiment analysis of product reviews," *Expert. Syst. Appl.*, vol. 214, no. 4, pp. 119128, 2023. doi: [10.1016/j.eswa.2022.119128](https://doi.org/10.1016/j.eswa.2022.119128).
- [7] U. Singh, A. Saraswat, H. K. Azad, K. Abhishek, and S. Shitharth, "Towards improving e-commerce customer review analysis for sentiment detection," *Sci. Rep.*, vol. 12, no. 1, pp. 21983, 2022. doi: [10.1038/s41598-022-26432-3](https://doi.org/10.1038/s41598-022-26432-3).
- [8] S. Saeed, "A customer-centric view of e-commerce security and privacy," *Appl. Sci.*, vol. 13, no. 2, pp. 1020, 2023. doi: [10.3390/app13021020](https://doi.org/10.3390/app13021020).
- [9] E. Deniz, H. Erbay, and M. Coşar, "Multi-label classification of e-commerce customer reviews via machine learning," *Axioms*, vol. 11, no. 9, pp. 436, 2022. doi: [10.3390/axioms11090436](https://doi.org/10.3390/axioms11090436).

- [10] N. Shafiabady, N. Hadjinicolaou, F. U. Din, B. Bhandari, R. M. Wu and J. Vakilian, "Using artificial intelligence (AI) to predict organizational agility," *PLoS One*, vol. 18, no. 5, pp. e0283066, 2023. doi: [10.1371/journal.pone.0283066](https://doi.org/10.1371/journal.pone.0283066).
- [11] A. Kumar, "Contextual semantics using hierarchical attention network for sentiment classification in social internet-of-things," *Multimed. Tools Appl.*, vol. 81, no. 26, pp. 36967–36982, 2022. doi: [10.1007/s11042-021-11262-8](https://doi.org/10.1007/s11042-021-11262-8).
- [12] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif. Intell. Rev.*, vol. 55, no. 7, pp. 5731–5780, 2022. doi: [10.1007/s10462-022-10144-1](https://doi.org/10.1007/s10462-022-10144-1).
- [13] N. Mukhtar and M. A. Khan, "Effective lexicon-based approach for Urdu sentiment analysis," *Artif. Intell. Rev.*, vol. 53, no. 4, pp. 2521–2548, 2019. doi: [10.1007/s10462-019-09740-5](https://doi.org/10.1007/s10462-019-09740-5).
- [14] M. Kamyab, G. Liu, and M. Adjeisah, "Attention-based CNN and Bi-LSTM model based on TF-IDF and GloVe word embedding for sentiment analysis," *Appl. Sci.*, vol. 11, no. 23, pp. 11255, 2021. doi: [10.3390/app112311255](https://doi.org/10.3390/app112311255).
- [15] K. Hu *et al.*, "A domain keyword analysis approach extending term frequency-keyword active index with Google Word2Vec model," *Scientometrics*, vol. 114, no. 3, pp. 1031–1068, 2018. doi: [10.1007/s11192-017-2574-9](https://doi.org/10.1007/s11192-017-2574-9).
- [16] P. Mehta and S. Pandya, "A review on sentiment analysis methodologies, practices and applications," *Int. J. Sci. Technol. Res.*, vol. 9, no. 2, pp. 601–609, 2020.
- [17] A. Kumar and A. Jaiswal, "Systematic literature review of sentiment analysis on Twitter using soft computing techniques," *Concurr. Comput. Pract. Exp.*, vol. 32, no. 1, pp. e5107, 2020. doi: [10.1002/cpe.5107](https://doi.org/10.1002/cpe.5107).
- [18] S. D. Cardoso, M. da Silveira, and C. Pruski, "Construction and exploitation of an historical knowledge graph to deal with the evolution of ontologies," *Knowl. Based Syst.*, vol. 194, no. 2, pp. 105508, 2020. doi: [10.1016/j.knosys.2020.105508](https://doi.org/10.1016/j.knosys.2020.105508).
- [19] M. Birjali, M. Kasri, and A. B. Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowl. Based Syst.*, vol. 226, pp. 107134, 2021. doi: [10.1016/j.knosys.2021.107134](https://doi.org/10.1016/j.knosys.2021.107134).
- [20] R. Mihalcea and P. Tarau, "Textrank: Bringing order into text," in *Proc. EMNLP*, Barcelona, Spain, 2004, pp. 404–411.
- [21] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.
- [22] L. Huang, Z. Dou, Y. Hu, and R. Huang, "Textual analysis for online reviews: A polymerization topic sentiment model," *IEEE Access*, vol. 7, pp. 91940–91945, 2019. doi: [10.1109/ACCESS.2019.2920091](https://doi.org/10.1109/ACCESS.2019.2920091).
- [23] W. Zhang, S. Kong, Y. Zhu, and X. Wang, "Sentiment classification and computing for online reviews by a hybrid SVM and LSA based approach," *Clust. Comput.*, vol. 22, no. S5, pp. 12619–12632, 2018. doi: [10.1007/s10586-017-1693-7](https://doi.org/10.1007/s10586-017-1693-7).
- [24] R. Obiedat *et al.*, "Sentiment analysis of customers' reviews using a hybrid evolutionary SVM-based approach in an imbalanced data distribution," *IEEE Access*, vol. 10, no. 1, pp. 22260–22273, 2022. doi: [10.1109/ACCESS.2022.3149482](https://doi.org/10.1109/ACCESS.2022.3149482).
- [25] A. Onan, "Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks," *Concurr. Comput. Pract. Exp.*, vol. 33, no. 23, pp. e5909, 2021. doi: [10.1002/cpe.5909](https://doi.org/10.1002/cpe.5909).
- [26] M. Giménez, J. Palanca, and V. Botti, "Semantic-based padding in convolutional neural networks for improving the performance in natural language processing. A case of study in sentiment analysis," *Neurocomputing*, vol. 378, no. 7, pp. 315–323, 2020. doi: [10.1016/j.neucom.2019.08.096](https://doi.org/10.1016/j.neucom.2019.08.096).
- [27] B. A. Hammou, A. A. Lahcen, and S. Mouline, "Towards a real-time processing framework based on improved distributed recurrent neural network variants with fastText for social big data analytics," *Inf. Process. Manag.*, vol. 57, no. 1, pp. 102122, 2020. doi: [10.1016/j.ipm.2019.102122](https://doi.org/10.1016/j.ipm.2019.102122).
- [28] M. E. Basiri, S. Nematy, M. Abdar, E. Cambria, and U. R. Acharya, "ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis," *Future Gener. Comput. Syst.*, vol. 115, no. 3, pp. 279–294, 2021. doi: [10.1016/j.future.2020.08.005](https://doi.org/10.1016/j.future.2020.08.005).

- [29] G. Meena, K. K. Mohbey, and A. Indian, "Categorizing sentiment polarities in social networks data using convolutional neural network," *SN Comput. Sci.*, vol. 3, no. 2, pp. 116, 2022. doi: [10.1007/s42979-021-00993-y](https://doi.org/10.1007/s42979-021-00993-y).
- [30] A. Kruspe, M. Häberle, I. Kuhn, and X. X. Zhu, "Cross-language sentiment analysis of European Twitter messages during the COVID-19 pandemic," arXiv preprint arXiv:2008.12172, 2020.
- [31] N. M. Alharbi, N. S. Alghamdi, E. H. Alkhamash, and J. F. Al Amri, "Evaluation of sentiment analysis via word embedding and RNN variants for Amazon online reviews," *Math. Probl. Eng.*, vol. 2021, pp. 5536560, 2021. doi: [10.1155/2021/5536560](https://doi.org/10.1155/2021/5536560).
- [32] B. Bansal and S. Srivastava, "Hybrid attribute based sentiment classification of online reviews for consumer intelligence," *Appl. Intell.*, vol. 49, no. 1, pp. 137–149, 2019. doi: [10.1007/s10489-018-1299-7](https://doi.org/10.1007/s10489-018-1299-7).
- [33] M. E. Alzahrani, T. H. Aldhyani, S. N. Alsubari, M. M. Althobaiti, and A. Fahad, "Developing an intelligent system with deep learning algorithms for sentiment analysis of e-commerce product reviews," *Comput. Intell. Neurosci.*, vol. 2022, pp. 3840071, 2022. doi: [10.1155/2022/3840071](https://doi.org/10.1155/2022/3840071).
- [34] K. K. Mohbey, "Sentiment analysis for product rating using a deep learning approach," in *Proc. ICAIS*, Coimbatore, India, 2021, pp. 121–126.
- [35] P. Hajek, L. Hikkerova, and J. M. Sahut, "Fake review detection in e-commerce platforms using aspect-based sentiment analysis," *J. Bus. Res.*, vol. 167, no. 5, pp. 114143, 2023. doi: [10.1016/j.jbusres.2023.114143](https://doi.org/10.1016/j.jbusres.2023.114143).
- [36] Z. Chen, Y. Xue, L. Xiao, J. Chen, and H. Zhang, "Aspect-based sentiment analysis using graph convolutional networks and co-attention mechanism," in *Proc. ICONIP*, Bali, Indonesia, 2021, pp. 441–448.
- [37] F. Mehmood, M. U. Ghani, M. A. Ibrahim, R. Shahzadi, W. Mahmood and M. N. Asim, "A precisely xtreme-multi channel hybrid approach for roman Urdu sentiment analysis," *IEEE Access*, vol. 8, pp. 192740–192759, 2020. doi: [10.1109/ACCESS.2020.3030885](https://doi.org/10.1109/ACCESS.2020.3030885).
- [38] D. Baishya, J. J. Deka, G. Dey, and P. K. Singh, "SAFER: Sentiment analysis-based fake review detection in e-commerce using deep learning," *SN Comput. Sci.*, vol. 2, no. 6, pp. 479, 2021. doi: [10.1007/s42979-021-00918-9](https://doi.org/10.1007/s42979-021-00918-9).
- [39] M. B. A. Miah, S. Awang, M. M. Rahman, A. S. M. S. Hosen, and I. H. Ra, "A new unsupervised technique to analyze the centroid and frequency of key phrases from academic articles," *Electronics*, vol. 11, no. 17, pp. 2773, 2022. doi: [10.3390/electronics11172773](https://doi.org/10.3390/electronics11172773).
- [40] I. Jamaleddyn, R. El ayachi, and M. Biniz, "An improved approach to Arabic news classification based on hyperparameter tuning of machine learning algorithms," *J. Eng. Res.*, vol. 11, no. 2, pp. 100061, 2023. doi: [10.1016/j.jer.2023.100061](https://doi.org/10.1016/j.jer.2023.100061).
- [41] S. R. Labhsetwar, "Predictive analysis of customer churn in telecom industry using supervised learning," *ICTACT J. Soft Comput.*, vol. 10, no. 2, pp. 2054–2060, 2020. doi: [10.21917/ijsc.2020.0291](https://doi.org/10.21917/ijsc.2020.0291).
- [42] M. Li, L. Chen, J. Zhao, and Q. Li, "Sentiment analysis of Chinese stock reviews based on BERT model," *Appl. Intell.*, vol. 51, no. 7, pp. 5016–5024, 2021. doi: [10.1007/s10489-020-02101-8](https://doi.org/10.1007/s10489-020-02101-8).
- [43] B. Gaye, D. Zhang, and A. Wulamu, "Sentiment classification for employees reviews using regression vector-stochastic gradient descent classifier (RV-SGDC)," *PeerJ Comput. Sci.*, vol. 7, pp. e712, 2021.
- [44] P. Savci and B. Das, "Prediction of the customers' interests using sentiment analysis in e-commerce data for comparison of Arabic, English, and Turkish languages," *J. King Saud. Univ.–Comput. Inf. Sci.*, vol. 35, pp. 227–237, 2023.
- [45] H. T. Phan and N. T. Nguyen, "A fuzzy graph convolutional network model for sentence-level sentiment analysis," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 5, pp. 2953–2965, 2024. doi: [10.1109/TFUZZ.2024.3364694](https://doi.org/10.1109/TFUZZ.2024.3364694).
- [46] S. Zhang, H. Gong, and L. She, "An aspect sentiment classification model for graph attention networks incorporating syntactic, semantic, and knowledge," *Knowl. Based Syst.*, vol. 275, pp. 110662, 2023. doi: [10.1016/j.knosys.2023.110662](https://doi.org/10.1016/j.knosys.2023.110662).