Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.066050

ARTICLE

# Improving Fashion Sentiment Detection on X through Hybrid Transformers and RNNs

**Bandar Alotaibi[1,*], Aljawhara Almutarie[2], Shuaa Alotaibi[3] and Munif Alotaibi[4]**

[1]Department of Information Technology, Faculty of Computers and Information Technology, University of Tabuk, Tabuk, 71491, Saudi Arabia
[2]College of Humanities and Social Sciences, Mass Communication Department, King Saud University, Riyadh, 11451, Saudi Arabia
[3]Department of Advertising and Marketing Communication, College of Media and Communication, Imam Mohammad Ibn Saud Islamic University, Riyadh, 11432, Saudi Arabia
[4]Department of Computer Science, Faculty of Computing and Information Technology, University of Shaqra, Shaqra, 11911, Saudi Arabia

*Corresponding Author: Bandar Alotaibi. Email: b-alotaibi@ut.edu.sa

Received: 28 March 2025; Accepted: 21 May 2025; Published: 30 July 2025

**ABSTRACT:** X (formerly known as Twitter) is one of the most prominent social media platforms, enabling users to share short messages (tweets) with the public or their followers. It serves various purposes, from real-time news dissemination and political discourse to trend spotting and consumer engagement. X has emerged as a key space for understanding shifting brand perceptions, consumer preferences, and product-related sentiment in the fashion industry. However, the platform's informal, dynamic, and context-dependent language poses substantial challenges for sentiment analysis, mainly when attempting to detect sarcasm, slang, and nuanced emotional tones. This study introduces a hybrid deep learning framework that integrates Transformer encoders, recurrent neural networks (i.e., Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)), and attention mechanisms to improve the accuracy of fashion-related sentiment classification. These methods were selected due to their proven strength in capturing both contextual dependencies and sequential structures, which are essential for interpreting short-form text. Our model was evaluated on a dataset of 20,000 fashion tweets. The experimental results demonstrate a classification accuracy of 92.25%, outperforming conventional models such as Logistic Regression, Linear Support Vector Machine (SVM), and even standalone LSTM by a margin of up to 8%. This improvement highlights the importance of hybrid architectures in handling noisy, informal social media data. This study's findings offer strong implications for digital marketing and brand management, where timely sentiment detection is critical. Despite the promising results, challenges remain regarding the precise identification of negative sentiments, indicating that further work is needed to detect subtle and contextually embedded expressions.

**KEYWORDS:** Sentiment analysis; deep learning; natural language processing; transformers; recurrent neural networks

## 1 Introduction

X has become a powerful platform for consumers to share their opinions and engage with brands in real time. Millions of tweets (posts) are generated on X daily, offering valuable insights into public sentiment about products, services, and brands. Digital marketers face the challenge of extracting actionable information from these unstructured data [1]. This research uses advanced text mining techniques to analyze consumer sentiment data obtained from X in order to provide insights that can help digital marketers to track sentiment and adjust their strategies accordingly. However, challenges arise due to short-form noisy data and rapidly changing sentiments due to product launches or controversies. Overcoming these challenges is essential for marketers to stay agile in the modern digital landscape.

Sentiment analysis is increasingly important due to its applications in product reviews, education, and politics [2,3]. The rapid growth of social media has further strengthened the demand for efficient sentiment analysis tools that are capable of processing large-scale user-generated content [4]. Sentiment analysis merges computer science and linguistics to evaluate the sentiment expressed in text data [5]. In particular, sentiment analysis focuses on detecting and classifying sentiments [6], generally categorizing text as positive, negative, or neutral [7], with standard methods including machine learning, lexicon-based, and hybrid approaches [8]. Sentiment analysis can be used for the assessment of product reviews, social media monitoring, and political sentiment, helping companies to better understand customer feedback and make data-driven decisions. The associated process generally includes pre-processing, feature extraction, and classification steps. Notably, challenges relating to the analysis of context, sarcasm, and multilingual content persist [9]. Future research should aim to improve the accuracy of relevant methods in these areas.

Sentiment analysis techniques can be classified into Natural Language Processing (NLP)-based approaches—which focus on text feature extraction, topic modeling, and document frequency analysis—and machine learning-based approaches—which employ supervised and unsupervised statistical models [10,11]. Recent research on sentiment analysis has highlighted various methodologies, including machine learning-based, lexicon-based, and hybrid strategies. Machine learning approaches, especially supervised learning models, have shown high classification accuracy, including algorithms such as Support Vector Machines (SVMs), Neural Networks, and Naïve Bayes. NLP approaches can also be integrated with such models to enhance their precision [12]. Machine learning models, including SVM and Naïve Bayes, have demonstrated promising classification performance [13]. However, heterogeneous datasets and model generalization pose significant challenges in this context [14]. Additionally, sentiment analysis faces issues including domain adaptation, limited availability of labeled data, and complex linguistic structures. Moreover, developing accurate sentiment models across multiple languages remains a persistent challenge. Sentiment analysis of fashion-related tweets has garnered attention. Researchers have employed lexicon-based techniques and machine learning algorithms (e.g., Naïve Bayes and SVMs) to assess public sentiment regarding brands and retailers [15,16], facilitating comparisons in terms of brand popularity and customer happiness [17]. The process starts with collecting tweets, preparing the data, and performing sentiment analysis to understand consumers' feelings. Analyzing fashion tweets is a cost-effective way to gather feedback, compared to traditional customer surveys [18].

User-generated content on social media has accelerated advancements in automated sentiment analysis [19–21]. Automated sentiment analysis methods detect and classify sentiment within textual content, including posts, comments, and reviews. Utilizing machine learning and NLP techniques, sentiment analysis can help to interpret public opinions, forecast market trends, and assess public sentiment during global events [22]. This encompasses the gathering and pre-processing of data, followed by feature extraction, labeling, and implementing NLP and machine learning algorithms [23]. Analyzing unstructured data from blogs, reviews, and social networks is challenging. Furthermore, as the volume of digital information grows, the need for effective sentiment analysis techniques becomes even more crucial.

Despite extensive work having been carried out in the field of sentiment analysis, there are still several critical gaps in the literature:

1. Limited focus on fashion-specific sentiment: Although general sentiment analysis has been studied across domains, few works have targeted the fashion industry on platforms such as X, where consumer behaviors are highly trend-driven and visually influenced.

2. Inadequate handling of short-form noisy text: Traditional methods often fail when applied to tweets, due to sarcasm, abbreviations, and lack of contextual cues.
3. Lack of hybrid model exploration: Few studies have combined Transformer-based contextual modeling with recurrent structures (e.g., Long Short-Term Memory (LSTM)/Gated Recurrent Unit (GRU)) to leverage global and sequential features for tweet classification.
4. Insufficient comparative benchmarks: Existing works often do not clearly report their performance improvements over baseline models using similar datasets or data splits, thus limiting their reproducibility and impacting the assessment results.

Although machine learning approaches are effective, direct comparative analysis is necessary. Hybrid methodologies that combine more than one deep learning model have also shown promise for consideration in future research. Sentiment analysis is a key tool for businesses and researchers, helping them to better understand public opinion and identify trends in the digital landscape [24].

This study presents a hybrid deep learning framework for sentiment classification in fashion-related tweets which integrates Transformer encoders with Recurrent Neural Networks (RNNs)—namely, LSTM and GRU—along with attention mechanisms. While sentiment analysis has been widely explored in general-purpose datasets, there remains a lack of specialized approaches targeting short-form, fashion-centric content on social media.

**The contributions of this research are as follows:**

1. We propose a text mining framework that extracts tweets about brands for effective sentiment analysis.
2. We integrate advanced deep learning techniques (i.e., Transformers and RNNs) to classify consumer sentiment as positive, negative, or neutral, ensuring high accuracy across diverse and noisy X data.
3. We address the unique challenges posed by short, informal, and event-driven X posts by employing specialized pre-processing strategies and adaptive modeling approaches.
4. We offer insights for digital marketers by connecting shifts in sentiment to marketing events and product launches, enabling data-driven engagement strategies.

The remainder of this paper is structured as follows. Section 2 surveys the related work. Section 3 presents the information regarding the dataset. Section 4 introduces the proposed method. Section 5 discusses the results. Section 6 concludes the paper.

## 2 Related Work

Researchers have explored diverse methodologies for improved sentiment classification, ranging from traditional lexicon-based techniques to advanced deep learning and multimodal frameworks such as Bidirectional Encoder Representations from Transformers (BERT) [25]. Early Twitter sentiment analysis research relied on lexicon-based approaches, such as the study by Sarlan et al. [26], who used Python dictionaries to assign polarity scores to tweets. However, subsequent research has commonly leveraged machine learning methods. For example, initial research employed machine learning methods such as logistic regression and SVM for sentiment categorization [27]. A recent systematic literature review by Mao et al. [28] provides an extensive analysis of sentiment analysis methodologies, comparing feature-based, deep learning, and hybrid approaches. This study collectively illustrate the advancements in sentiment analysis and its expanding capability to capture nuanced consumer opinions.

Sentiment analysis in the fashion domain presents unique challenges due to its multimodal nature, requiring the integration of text, images, and domain-specific attributes. To address these challenges, Yuan and Lam [29] have proposed a framework that integrates image, text, and fashion attributes for enhanced sentiment analysis of fashion-related social media posts. Unlike existing works, which focused on general multimodal sentiment analysis, their approach leverages fashion-specific attributes to improve sentiment classification. The framework consists of three modules—a fashion-aware vision composition module, a fashion-aware text composition module, and a vision-text composition module—which are combined to more effectively capture sentiment. To support their study, they constructed a dataset of more than 12,000 fashion-related posts from social media, which were manually annotated for sentiment analysis.

Other studies have explored different aspects of fashion sentiment analysis. Abdel Fattah et al. [30] investigated image analysis techniques and proposed metrics such as social value to assess the efficacy of fashion images on platforms such as Instagram. Despite substantial progress, obstacles persist, including the identification of sarcasm, clarifying ambiguous statements, and handling multilingual analysis [25,31]. Therefore, future works should concentrate on augmenting contextual comprehension, advancing the elucidation ability of models, and fostering real-time adaptability across various domains. Sentiment analysis has extensive applications outside of fashion, for example, influencing e-commerce, healthcare, and finance.

Moreover, some studies have adopted deep learning to enhance the accuracy of sentiment analysis. Tran et al. [32] integrated deep learning with rule-based approaches to attain a competitive Root Mean Square Error (RMSE) score in a beauty–fashion review competition. Kavi Priya et al. [33] attained 93.4% accuracy utilizing an RNN model for fashion-related data. While Twitter sentiment analysis can help the fashion industry, research gaps reduce its effectiveness. In particular, it is difficult to accurately capture consumer sentiment due to the complex language and visual nature associated with fashion.

A key challenge in sentiment analysis is dealing with sarcasm, which is common in social media and poses significant obstacles for sentiment analysis [34]. Researchers have explored various approaches, including lexical, syntactic, and affective feature analyses [35], pre-processing techniques, and word frequency analysis. Emoji-based sentiment detection and machine learning models have also been applied to improve the identification of sarcasm. However, accurately detecting sarcasm remains challenging, even for humans. Enhanced sarcasm detection approaches are crucial for improving the accuracy of sentiment analysis and effectively interpreting social media conversations [36].

Fashion is visual, making multimodal sentiment analysis vital for understanding the insights gained through social media. We can better capture sentiment by merging image and video analysis with text mining. Recent studies have emphasized the importance of integrating multiple modalities—including visual imagery, textual data, and fashion-specific attributes—to enhance sentiment classification [29,37]. Researchers have also examined the social value of fashion-related Instagram content [30] and the significance of visual elements in influencing consumer sentiment. However, effectively fusing diverse data sources and extracting meaningful insights from visual content remains a developing research area [38].

Context is essential for sentiment classification in fashion, with terms often having specific meanings. Traditional sentiment analysis models struggle with the evolving language used in this industry. Research has explored various methodologies to capture this context, such as dual Gaussian visual–semantic embedding models for abstract fashion concepts [39], NLP techniques for fashion trend identification [40], and fashion entity augmentation through synonym discovery [41]. These approaches highlight the complexity of fashion-related terminology and the need for more sophisticated computational techniques to improve contextual comprehension. Addressing such research gaps is vital for improving sentiment analysis in the fashion sector. Enhancements in sarcasm detection, multimodal analysis, and contextual understanding can be expected to improve consumer insights and trend forecasting. Our research utilizes hybrid Transformers and RNNs to analyze sentiment on X better, with the aim of obtaining more accurate insights into the fashion industry.

BERTweet [42] is a transformer-based language model pre-trained specifically on English tweets. Developed as a RoBERTa-based architecture, BERTweet has demonstrated exceptional performance in sentiment analysis tasks involving social media data due to its ability to capture informal language, slang, abbreviations, and emojis. Similarly to many other general-purpose models, such as RoBERTa [43], BERTweet can be considered suitable for fashion-related sentiment analysis. However, its performance comes with increased computational costs, motivating the search for more efficient alternatives or hybrid models for scalable deployment.

To summarize the key findings with respect to the reviewed literature:

- Lexicon-based and traditional machine learning techniques laid the foundation for early sentiment analysis, but lack contextual depth.
- Deep learning models such as BiLSTM, BiGRU, and Convolutional Neural Network (CNN)—especially when integrated—offer superior performance in sentiment prediction tasks.
- Multimodal frameworks that combine text, image, and fashion-specific attributes significantly enhance fashion sentiment analysis.
- SVMs trained on semantically enriched document vectors showed good performance (88% accuracy) in fashion review classification.
- Sarcasm and irony detection remains a major challenge, with affective features and word/sub-word frequencies proving more effective than basic pre-processing.
- Emerging metrics such as Social Value link textual sentiment with visual popularity (likes, shares), in order to provide actionable insights for marketing.
- Several studies have underlined the importance of domain-specific lexicons and evolving fashion terminology to maintain accuracy in dynamic linguistic environments.
- visual data integration with text has been increasingly emphasized for comprehensive trend forecasting and emotion recognition in fashion media.
- Supervised machine learning methods such as Naïve Bayes and SVM are still widely used, but are limited in capturing nuanced or context-dependent sentiments.

We shortlisted RNNs and Transformers among the various deep learning architectures for several reasons. First, RNNs—particularly the LSTM and GRU variants—are proficient in handling sequential dependencies and capturing temporal patterns in short-form texts such as tweets. Second, Transformers (with their self-attention mechanisms) can effectively model long-range dependencies and contextual relationships, offering advantages over RNNs regarding parallel processing and scalability. Third, hybrid architectures that combine RNNs and Transformers leverage the strengths of both approaches, achieving higher performance in complex sentiment tasks that involve sarcasm, abbreviations, and non-standard syntax, which are typical of fashion-related social media content.

## 3 Dataset Information

We collected tweets using hashtags focused on fashion, couture, and luxury themes. Our collected dataset comprises 20,000 samples with 41 features each. These entries primarily include textual content, such as the tweet and hit sentence, along with metadata columns describing the source (e.g., Uniform Resource Locator (URL), source name, source domain), author information (e.g., author name, author handle), and various engagement metrics (e.g., shares, likes, replies). Basic descriptive statistics indicate that several columns—such as title, social echo, editorial echo, and comments—predominantly contain missing or null values, whereas others (e.g., date, time, document ID, URL, tweet) are fully populated. Numerical columns such as reach, engagement, and views display wide variability, with means, medians, and high standard deviations characteristic of user-driven online data. The sentiment column, which was fully populated manually across all rows, denotes each entry's sentiment (e.g., positive, negative, or neutral) and served as the principal label in our classification task. The dataset consists of several content types and contains a large number of null values in some metadata fields. Thus, it needed to undergo pre-processing to effectively extract relevant information for sentiment analysis. All samples are in the English language. The complete workflow of the proposed approach, from data collection to sentiment prediction, is illustrated in Fig. 1.

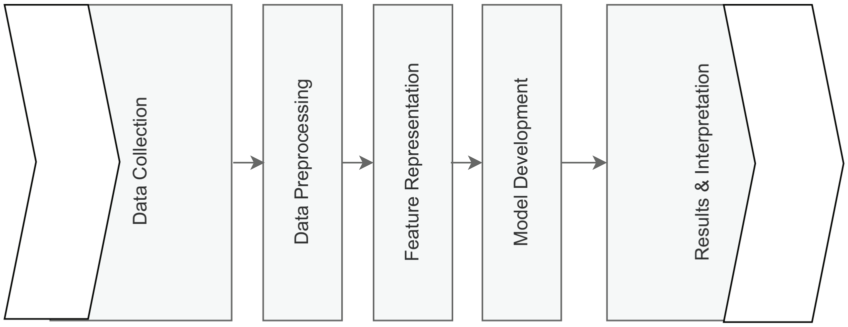

**Figure 1:** Overview of the proposed methodology for fashion sentiment analysis

### *3.1 Pre-Processing*

We removed irrelevant noise and stop words using Natural Language Toolkit (NLTK)'s stop word corpus, in order to minimize non-essential tokens during training. Each tweet was split into words, and terms appearing in the NLTK stop word list were discarded. Next, we numerically encoded the sentiment labels (e.g., positive, negative, neutral) using label encoding. We then split the data into 80% training and 20% testing sets to evaluate the generalization ability of the tested model. To handle tokenization, we specified a vocabulary limit (i.e., 10,000) and fit a tokenizer on the training set. This tokenizer converts each tweet into a sequence of integer indices representing the most frequent words. We shortened each sequence to a length of 100, thus creating a uniform input for the deep learning model.

### *3.2 Tweet Labeling*

As mentioned at the beginning of this section, we collected tweets about fashion using specific keywords and hashtags. After gathering the data, we labeled each tweet as positive, negative, or neutral. We labeled tweets that express positive feelings, such as appreciation or satisfaction, as positive. We labeled tweets that showed negative feelings, such as dissatisfaction or criticism, as negative. Tweets that neither clearly expressed positive nor negative sentiments, or which were factual and objective in nature without emotional context, were classified as neutral.

We independently labeled tweets to ensure the reliability and accuracy of the annotations, and disagreements were resolved through consensus discussion. We manually labeled our data to understand social media language, such as sarcasm, idioms, and common abbreviations, used on X. This dataset helped us to train and evaluate our sentiment analysis model, allowing for the subsequent accurate analysis of fashion-related content.

## 4 Proposed Method

This section outlines the methodological framework for the proposed Twitter sentiment analysis system. Our approach merges pre-trained word embeddings (i.e., GloVe [44]) with a Transformer Encoder [45], an LSTM layer [46], a GRU layer [47], and an Attention mechanism [45] in order to capture contextual dependencies and sequential patterns. Through combining these components, we aim to accurately classify tweets as positive, negative, or neutral.

We consider a dataset of the form

$$D = \{(t_i, c_i)\}_{i=1}^{S}, \tag{1}$$

where $S$ is the total number of instances. Each tweet $t_i \in \mathbb{N}^L$ is a sequence of token indices of length $L$, and each sentiment label $c_i \in \{1, 2, 3\}$ indicates a positive, negative, or neutral sentiment. To represent tokens, we define an embedding matrix

$$P \in \mathbb{R}^{|W| \times e_d}, \tag{2}$$

where $|W|$ is the vocabulary size and $e_d$ is the embedding dimension. For each token $w$, $P$ provides a vector $\mathbf{v}_{d,w} \in \mathbb{R}^{e_d}$. An embedding layer maps a tweet

$$t = [\, w_1, w_2, \ldots, w_L \,], \tag{3}$$

to

$$X = [\, v_{d,w_1}, v_{d,w_2}, \ldots, v_{d,w_L} \,] \in \mathbb{R}^{e_d \times L}. \tag{4}$$

To mitigate overfitting, we apply a spatial dropout function $\mathbb{S}(X, \gamma)$ to the embeddings $X$, where $\gamma$ is the dropout rate. This step randomly zeroes entire embedding vectors along the sequence dimension, yielding

$$X' = \mathbb{S}(X, \gamma). \tag{5}$$

We then use a Transformer Encoder $E$ that combines multi-head attention, a feed-forward network $\mu$ with a ReLU-like activation $A$, and a dense layer, alongside residual connections and layer normalization. Let

$$\rho \in \mathbb{R}^{h_d \times L}, \tag{6}$$

be the input to the multi-head attention, where $h_d$ specifies the hidden dimension of the Transformer output. Each head constructs query, key, and value matrices

$$Q_m = \rho\, W_k^Q, \quad K_m = \rho\, W_k^K, \quad V_m = \rho\, W_k^V, \tag{7}$$

and applies an attention mechanism

$$v = \mathbb{A}(Q_m, K_m, V_m). \tag{8}$$

The outputs of all heads are concatenated and passed through a function $M(\cdot)$, yielding

$$U(\rho) = M(v_1, \ldots, v_h)\, W^O. \tag{9}$$

A two-layer feed-forward network $\mu(z)$ then takes the form

$$\mu(z) = \mathbb{D}\big(A(z\, W_1 + b_1)\, W_2 + b_2\big), \tag{10}$$

where $\mathbb{D}$ includes dropout, $W_1$ and $W_2$ are weight matrices, and $b_1$ and $b_1$ are bias vectors. Residual connections and layer normalization appear at each stage:

$$\rho' = N(\rho + \mathbb{D}(U(\rho))), \quad \rho'' = N(\rho' + \mathbb{D}(\mu(\rho'))). \tag{11}$$

Consequently, the Transformer-encoded output is

$$\rho'' = E(X'). \tag{12}$$

We capture sequential dependencies with two recurrent layers (LSTM $\mathbb{L}$ and GRU $R$) applied to $\rho''$, yielding

$$h_{\mathbb{L}} = \mathbb{L}(\rho''), \quad h_R = R(\rho''). \tag{13}$$

Each recurrent layer typically applies internal dropout for additional regularization. We also use a standalone attention mechanism

$$\mathbb{A} = \mathbb{A}(\rho'', \rho''), \tag{14}$$

followed by global average pooling across the sequence dimension to obtain

$$h_{\mathbb{A}} = \frac{1}{L}\sum_{t=1}^{L} \mathbb{A}_t. \tag{15}$$

We concatenate the three feature vectors $h_{\mathbb{L}}$, $h_R$, and $h_{\mathbb{A}}$ into

$$h_F = [\, h_{\mathbb{L}}, h_R, h_{\mathbb{A}} \,] \in \mathbb{R}^{f_d}, \tag{16}$$

and pass the feature vector dimension after concatenation, $h_F$, through a dense layer, resulting in

$$h_{\mathbb{D}} = \mathbb{D}\big(A(h_F\, W_3 + b_3)\big). \tag{17}$$

Finally, an output layer $\mathbb{O}$ with a Softmax $\mathbb{S}_m$ produces class probabilities

$$\hat{c} = \mathbb{S}_m\big(h_{\mathbb{D}}\, W_4 + b_4\big) \in \mathbb{R}^3, \tag{18}$$

covering the three sentiment classes. Here, $W_3$, $W_4$, $b_3$, and $b_4 \in \mathbb{R}^3$ denote the weight and bias parameters in the final layers, corresponding to the three sentiment classes.

We define $\theta$ as the model parameters and use sparse categorical cross-entropy as the loss function:

$$\mathcal{L} = -\frac{1}{S}\sum_{i=1}^{S}\sum_{c=1}^{3} \mathbb{I}(c_i = c)\, \log\big(\hat{c}_c\big). \tag{19}$$

The optimizer uses Adam with a learning rate of $\eta$. We further employ early stopping if the validation loss fails to improve for three epochs, and reduce $\eta$ by half if the validation loss flattens for five epochs.

To implement the embedding matrix $P$, we rely on pre-trained GloVe embeddings, defaulting unknown tokens to zero vectors to handle out-of-vocabulary words. Our hybrid model integrates the embedding layer, spatial dropout, Transformer encoder, recurrent networks, attention, and dense layers. We typically set the hyperparameters as follows: $\text{max}_{\text{words}} = 10{,}000$, max_sequence_length = 100, $e_d = 100$, $\text{head}_{\text{size}} = 64$, $\text{num}_{\text{heads}} = 4$, and $\text{ff}_{\text{dim}} = 128$. After passing a tweet through these layers, we obtain $\hat{c}$, indicating the predicted sentiment class. We perform training for up to five epochs using a batch size of 64, with early stopping and learning rate scheduling in order to balance computational efficiency and model generalization to real-world applications. The architecture of our deep learning model is illustrated in Fig. 2.

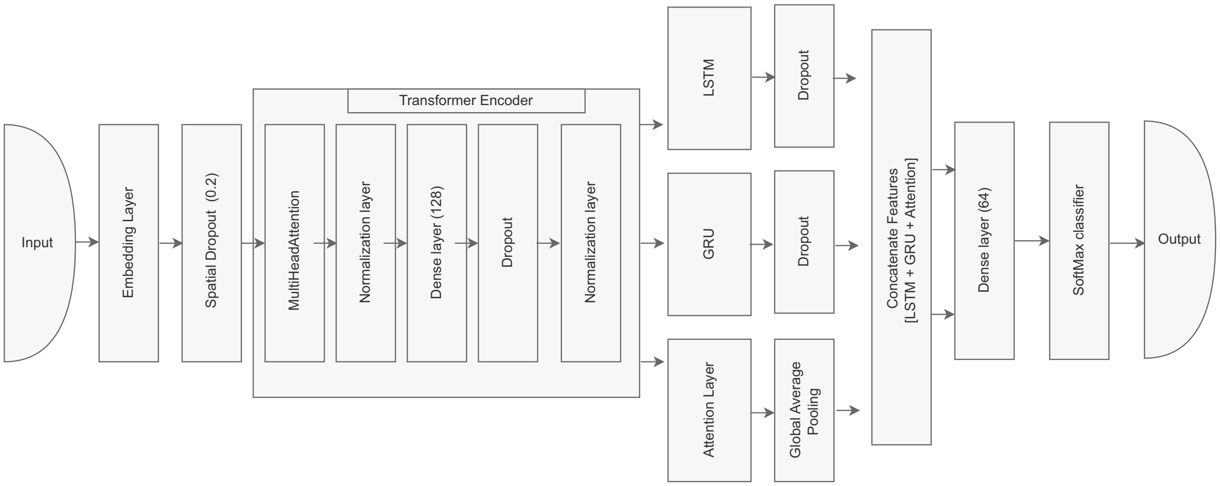

**Figure 2:** The architecture of the proposed hybrid model for sentiment classification, which combines word embeddings, spatial dropout, a Transformer encoder for contextual features, LSTM and GRU layers for sequential patterns, and an attention mechanism. Features are concatenated and passed through dense layers for sentiment prediction via a Softmax classifier

### *4.1 Hyperparameter Details*

We set the vocabulary size to 10,000 and the maximum input length to 100 tokens. Each token is mapped to a 100-dimensional embedding, initialized from pre-trained GloVe vectors, and made trainable to allow for fine-tuning. The Transformer encoder uses a multi-head attention mechanism with a head size of 64, a number of heads of 4, and a feed-forward dimension of 128. We employ dropout rates of 0.1 or 0.2 in different layers (embedding, LSTM, GRU, and dense layers) to balance regularization with stable convergence. The LSTM and GRU layers have hidden sizes 64, each with a recurrent dropout of 0.2. We use two dense layers for the final classification: the first has 64 units with a ReLU activation, and the output layer uses Softmax to predict the sentiment class. The learning rate is initialized at 0.001 for the Adam optimizer, with early stopping triggered if the validation loss flattens for three epochs and a learning rate scheduler halving after five epochs without improvement, down to a minimum of $10^{-6}$. Training is typically run for up to six epochs with a batch size of 64, and we store the best model weights based on validation performance. Table 1 presents the hyperparameters used in the proposed model.

**Table 1:** Hyperparameters used in the proposed model

| Hyperparameter | Value |
|---|---|
| Learning rate | 0.1% |
| Batch size | 64% |
| Optimizer | Adam |
| Epochs | 6 % |
| Activation function | ReLU |
| Transformer heads | 4 |
| Dropout rate | 0.1 or 0.2 |
| Embedding dimension | 100 |

## 5 Results and Discussion

Throughout the six training epochs, as shown in Fig. 3, there was a clear downward trend in the training loss $\mathcal{L}$ (from 0.5969 in epoch 1 to 0.1066 by epoch 6), accompanied by a steady increase in the training accuracy $\alpha$ (from 0.7656 to 0.9686). These paired trends indicate effective learning: the model reduces its predictive error and more accurately classifies training examples.

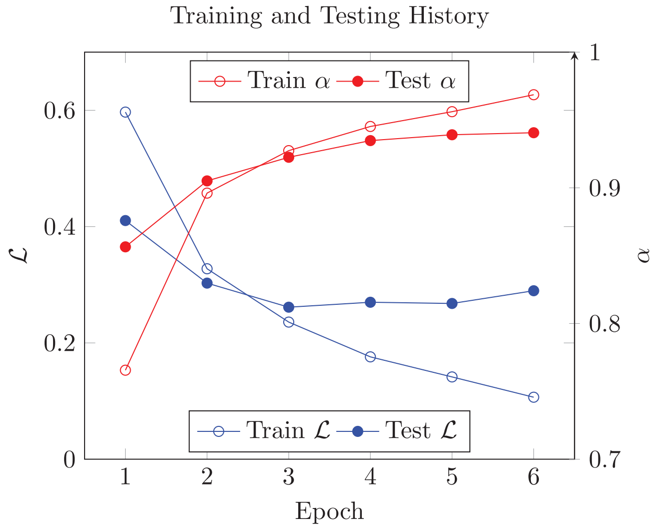

**Figure 3:** Training and testing accuracies vs. training and testing loss for the six epochs

The model showed strong performance during validation. The validation accuracy started at 0.8565 in the first epoch and rose to 0.9406 by epoch 6, which means that the model learns well. The validation loss decreased from 0.4104 to around 0.26–0.29 in the later epochs, with normal minor fluctuations. By the final epoch, the training and validation accuracies reached the mid-90% range, indicating that the model can capture patterns effectively without much overfitting. A separate test evaluation revealed a loss of 0.2614 and an accuracy of 0.9225, confirming that the model also performs well on new, unseen data.

Furthermore, the confusion matrix shown in Fig. 4 indicates that the model achieved 95% correctness for neutral tweets (1124 of 1183), misclassifying only 61 as positive and 18 as negative. For positive tweets, it reached 92% correctness (1297 of 1409), with relatively few errors (involving 19 misclassified as negative and 33 as neutral). Negative tweets were misclassified more often than others, with a 77% accuracy rate, meaning that 125 out of 163 tweets are correctly identified. Many of these tweets were incorrectly classified as positive or neutral, indicating that better methods are needed to recognize subtle negative samples.

### *5.1 Analysis of Tweet Length by Sentiment*

Fig. 5 presents the Cumulative Distribution Function (CDF) of tweet lengths, measured by the number of words, for the three sentiment categories: positive, negative, and neutral. On the horizontal axis, each point indicates a particular number of words, while the vertical axis shows the fraction of tweets (from 0 to 1) that contain up to that many words. This layout directly compares how quickly each sentiment's tweets accumulate across increasing word counts.

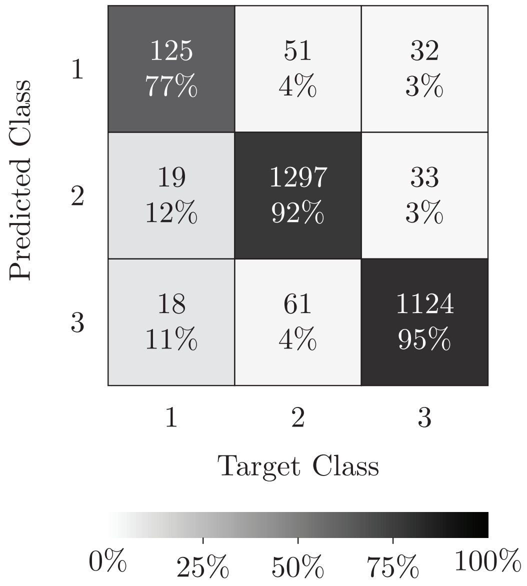

**Figure 4:** The confusion matrix for the three sentiment classes: negative (1), positive (2), and neutral (3)

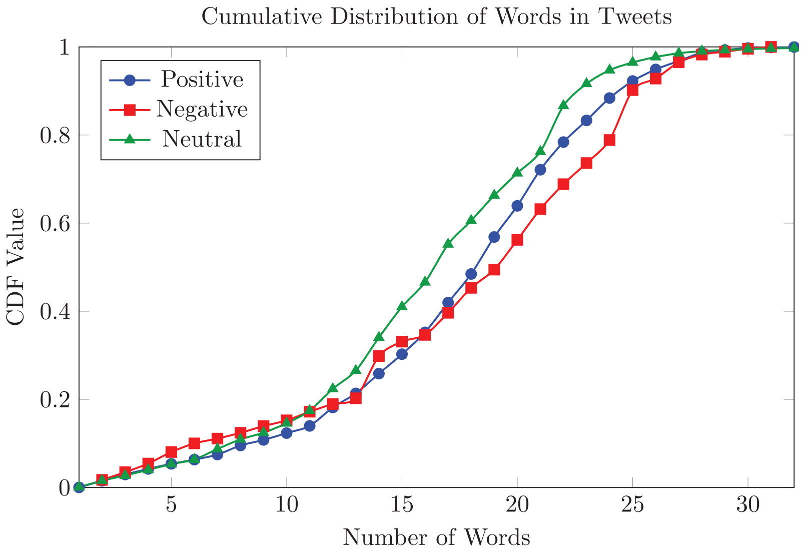

**Figure 5:** The CDFs of tweet lengths for three sentiment classes: positive, negative, and neutral

The positive class curve starts to climb at very short tweets, surpassing 50% (i.e., half of all positive tweets) around 14–15 words. When the word count reaches roughly 30–32, the positive distribution flattens, signifying that few tweets exceed this length. The negative class curve increases from low to high word counts, but initially includes a larger fraction of very short tweets (e.g., at 3–4 words) compared to the positive distribution; it likewise peaks in the upper 20s or early 30s. Meanwhile, the neutral class curve begins near zero (with minimal word counts) but catches up by around 16 words to encompass half of all neutral tweets. Although neutral tweets occupy a slightly broader middle range, they also culminate at around 30–32 words in the higher percentiles.

Negative class tweets are somewhat briefer at the lower end, indicating concise or direct expressions of negativity. In contrast, positive and neutral class tweets more commonly occupy moderate or slightly higher word counts. Comparing these three sentiment curves we see that, although there is substantial overlap in their ranges, minor differences appear at both the lower and upper extremes. Such observations reinforce that the way in which sentiment is expressed on social media (e.g., concise negativity versus more elaborated positivity) may yield additional clues for refining sentiment classifiers.

### 5.2 *Frequent Word Sequences*

In addition to examining length distributions, we analyzed the dataset's most frequent unigrams, bigrams, and trigrams. Fig. 6 shows the top 10 items in each category on a log-log scale. The horizontal axis indicates their rank (from most frequent to 10th), while the vertical axis indicates their frequency. For unigrams, the first few ranks exceed 8000 occurrences each, but the count drops sharply by the 10th rank. A similarly steep decline was observed in the highest ranks for bigrams and trigrams, illustrating a typical power-law pattern where a small set of tokens or phrases dominates in usage. At the same time, the majority occur far less frequently.

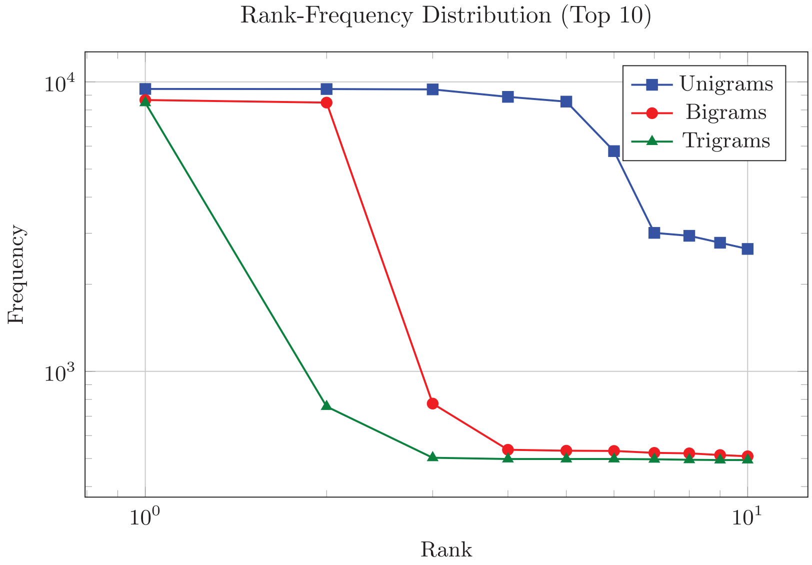

**Figure 6:** The top ten unigrams, bigrams, and trigrams on a log scale

Many social media users frequently repeat specific words, phrases, or hashtags, reflecting common language patterns. Recognizing these high-frequency terms can further assist in feature engineering or in refining token-level embeddings, ultimately improving sentiment classification, especially for context-driven patterns in positive or negative expressions.

The above findings highlight the effectiveness of our hybrid model in handling noisy, short-form Twitter data. The strong classification results and analyses of tweet length distributions and word-sequence frequencies revealed how different sentiments manifest in varying textual styles. Negative tweets, for instance, are typically shorter at the lower end, suggesting concise negativity, whereas positive and neutral content often extends to moderately longer messages. The prevalence of specific tokens across unigrams, bigrams, and trigrams further reflects the skewed nature of language use on Twitter, where a small set of terms dominates.

These observations provide support for our approach and indicate areas where we can make improvements. We should pay special attention to short and direct complaints in order to better handle negative tweets. Additionally, knowing common word patterns can help us to develop specific vocabularies or embeddings that reflect different feelings. These results support our goal of creating a strong sentiment analysis system that helps marketers to track brand discussions, spot new trends, and improve engagement strategies in a rapidly changing social media environment.

To illustrate the effectiveness of our method, we compared it with various baseline models on the same dataset. Table 2 summarizes the performance of our proposed hybrid model against several baseline methods, all trained and evaluated using the same data split. For fair comparison, the LSTM-based model employed the same number of epochs as our hybrid approach. As the results indicate, the hybrid model achieved a loss of 0.2614 and an accuracy of 92.25%, outperforming the LSTM baseline, which exhibited a loss of 0.3792 and an accuracy of 90.29%. Meanwhile, the classical machine learning methods—namely, Logistic Regression (86.85%), Linear SVM (88.95%), and Random Forest (84.71%)—all trailed further behind.

**Table 2:** Performance comparison (in terms of accuracy) between our hybrid model and baseline models

| Model | Accuracy |
|---|---|
| LSTM | 90.29% |
| Logistic regression | 86.85% |
| Linear SVM | 88.95% |
| Random forest | 84.71% |
| Our hybrid model | 92.25% |

Compared to general-purpose models such as BERT [42] or RoBERTa [43], these methods require significantly more memory and training time due to their large transformer backbones. Our hybrid model offers a more resource-efficient solution while maintaining a competitive performance.

This performance gap indicates that our hybrid design can capture text patterns better than a single LSTM or simpler classifiers. The baseline LSTM achieved a good accuracy, but the hybrid system's deeper layers and specialized feature extraction significantly improved its classification results. The fact that all models shared identical training/testing partitions and epochs ensured that the observed performance improvements derived from architectural enhancements, rather than differences in data.

### *5.3 Limitations of the Study*

Despite the promising performance of the proposed hybrid model, several limitations warrant consideration. First, the used dataset included numerous missing values in non-textual metadata fields (e.g., title, comments, and social echo), which were excluded during pre-processing. This exclusion may have limited the model's ability to leverage potentially valuable contextual cues. Future research could explore imputation techniques or metadata-aware architectures to better utilize this information. Second, although sentiment labels were manually annotated—thus contributing to higher labeling accuracy—this process is inherently subjective and may have introduced bias or inconsistency. To enhance the reliability of labels, future studies could adopt more advanced methods. Third, the model was trained exclusively on fashion-related tweets in the English language, which restricts its generalizability across other languages. Extending the model to include multilingual datasets could broaden its global applicability.

Additionally, the confusion matrix revealed that the model underperformed in detecting negative sentiment, which is often expressed through subtlety, irony, or sarcasm. Incorporating sarcasm detection modules or affective feature extraction techniques may enhance the model's sensitivity to such nuanced emotional expressions. Finally, while the proposed hybrid architecture is more efficient than large-scale transformer-only models such as BERTweet, it still introduces greater complexity when compared to lightweight models. Future work may consider model distillation, pruning, or quantization to reduce the model's computational overhead without compromising its predictive performance, making it more suitable for real-time applications.

## 6 Conclusion

This study presented and evaluated a deep learning-based sentiment analysis model tailored for short, noisy Twitter data, a critical source of brand and consumer engagement signals. Integrating pre-trained word embeddings, a Transformer encoder, recurrent neural networks, and attention mechanisms, the proposed model achieved a strong performance (over 92% accuracy on the test set). The obtained results underscore its effectiveness in classifying tweets into positive, neutral, or negative sentiment, with particular strength in recognizing neutral and positive posts. However, the analysis also revealed the model's weaker performance on negative content, highlighting the nuanced language often used to express dissatisfaction. For digital marketers, these insights are paramount: timely recognition of negative trends can mitigate reputational risks, while understanding consumers' positive responses guides the amplification of successful campaigns. Future work could explore domain-specific vocabularies or custom lexicons to capture the slang and emotional cues typical in fashion- or product-related discussions on Twitter. Advanced text mining techniques, supported by context-aware models, offer a scalable solution to monitor brand sentiment, identify emerging trends, and empower data-driven marketing strategies in an era where consumer feedback unfolds continuously and publicly on social networks. In addition, there is still room for improvement in terms of the model's accuracy, especially regarding the classification of negative emotions, which remains more challenging due to the nature of negative expressions often seen on social media.

**Acknowledgement:** Not applicable.

**Funding Statement:** The authors received no specific funding for this study.

**Author Contributions:** The authors confirm their contributions to the paper as follows: Conceptualization, Bandar Alotaibi, Aljawhara Almutarie, and Shuaa Alotaibi; methodology, Bandar Alotaibi and Munif Alotaibi; software, Bandar Alotaibi; validation, Bandar Alotaibi and Munif Alotaibi; writing—original draft preparation, Bandar Alotaibi, Aljawhara Almutarie, and Shuaa Alotaibi; writing—review and editing, Munif Alotaibi and Shuaa Alotaibi; visualization, Shuaa Alotaibi; supervision, Bandar Alotaibi; project administration, Bandar Alotaibi. 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 available from the corresponding author, B.A., upon reasonable request.

**Ethics Approval:** Not applicable.

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

## References

1. Al Montaser MA, Ghosh BP, Barua A, Karim F, Das BC, Shawon RER, et al. Sentiment analysis of social media data: business insights and consumer behavior trends in the USA. Edelweiss Appl Sci Technol. 2025;9(1):545–65. doi:10.55214/25768484.v9i1.4164.
2. Mirzh N, Ali ZH. Sentiment analysis techniques—survey. Wasit J Pure Sci. 2023;2(2):282–90. doi:10.31185/wjps.152.
3. Madhoushi Z, Hamdan AR, Zainudin S. Sentiment analysis techniques in recent works. In: 2015 Science and Information Conference (SAI); 2015 Jul 28–30; London, UK. p. 288–91. doi:10.1109/SAI.2015.7237157.
4. Kathuria A, Upadhyay S. A novel review of various sentimental analysis techniques. Int J Comp Sci Mobile Comput. 2017;6(4):17–22.
5. Taboada M. Sentiment analysis: an overview from linguistics. Annu Rev Linguist. 2016;2(1):325–47. doi:10.1146/annurev-linguistics-011415-040518.
6. Balahur A. Sentiment analysis in social media texts. In: Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis; 2013 Jun 14; Atlanta, GA, USA. p. 120–8.
7. Tan KL, Lee CP, Lim KM. A survey of sentiment analysis: approaches, datasets, and future research. Appl Sci. 2023;13(7):4550. doi:10.3390/app13074550.
8. Alharbi A, Aljurbua R, Gupta S, Obradovic Z. TriLex: a fusion approach for unsupervised sentiment analysis of short texts. PLoS One. 2025;20(4):e0317100. doi:10.1371/journal.pone.0317100.
9. Nalawat M. An overview of sentiment analysis: concept, techniques, and challenges. Indian J Comput Sci. 2019;4(4):24–31.
10. Jurek A, Mulvenna MD, Bi Y. Improved lexicon-based sentiment analysis for social media analytics. Secur Inform. 2015;4(1):9. doi:10.1186/s13388-015-0024-x.
11. Kim HD, Ganesan K, Sondhi P, Zhai C. Comprehensive review of opinion summarization. Technical Report. Champaign, IL, USA: University of Illinois at Urbana-Champaign; 2011.
12. Yaakub MR, Latiffi MIA, Zaabar LS. A review on sentiment analysis techniques and applications. IOP Conf Ser: Mater Sci Eng. 2019;551:012070. doi:10.1088/1757-899X/551/1/012070.
13. Tandon V, Mehra R. An integrated approach for analysing sentiments on social media. Informatica. 2023;47(2):4390. doi:10.31449/inf.v47i2.4390.
14. Read J. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In: Proceedings of the ACL Student Research Workshop; 2005 Jun 27; Ann Arbor, MI, USA. p. 43–8.
15. Rasool A, Tao R, Marjan K, Naveed T. Twitter sentiment analysis: a case study for apparel brands. J Phys: Conf Ser. 2019;1176:022015. doi:10.1088/1742-6596/1176/2/022015.
16. Shinde GK, Lokhande VN, Kalyane RT, Gore VB, Raut UM. Sentiment analysis using hybrid approach. Int J Res Appl Sci Eng Technol. 2021;9:282–5.
17. Sharma V, Manocha T. Comparative analysis of online fashion retailers using customer sentiment analysis on Twitter. In: Proceedings of the International Conference on Innovative Computing & Communication (ICICC); 2022; Singapore: Springer.
18. Younis EM. Sentiment analysis and text mining for social media microblogs using open source tools: an empirical study. Int J Comput Appl. 2015;112(5):44–8.
19. Chen X, Xie H, Tao X, Wang FL, Zhang D, Dai HN. A computational analysis of aspect-based sentiment analysis research through bibliometric mapping and topic modeling. J Big Data. 2025;12(1):40. doi:10.1186/s40537-025-01068-y.
20. Zhao T, Li C, Li M, Ding Q, Li L. Social recommendation incorporating topic mining and social trust analysis. In: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management; 2013 Oct 27–Nov 1; San Francisco, CA, USA. p. 1643–8. doi:10.1145/2505515.2505592.
21. Kaur G, Malik K. A comprehensive overview of sentiment analysis and fake review detection. In: Mobile Radio Communications and 5G Networks: Proceedings of MRCN 2020; 2020; Singapore: Springer. p. 293–304. doi:10.1007/978-981-15-7130-5_22.

22. Shukla A, Shukla S. A survey on sentiment classification and analysis using data mining. Int J Adv Res Comp Sci. 2015;6(7):20.
23. Redhu S, Srivastava S, Bansal B, Gupta G. Sentiment analysis using text mining: a review. Int J Data Sci Technol. 2018;4(2):49–53. doi:10.11648/j.ijdst.20180402.12.
24. Smith-Mutegi D, Mamo Y, Kim J, Crompton H, McConnell M. Perceptions of STEM education and artificial intelligence: a Twitter (X) sentiment analysis. Int J Stem Educ. 2025;12(1):1–18. doi:10.1186/s40594-025-00527-5.
25. Sushma S, Nayak SK, Krishna MV. A comprehensive review of sentiment analysis: trends, challenges, and future directions. In: 2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI); 2024 Nov 18–20. p. 1175–81. doi:10.1109/ICDICI62993.2024.10810919.
26. Sarlan A, Nadam C, Basri S. Twitter sentiment analysis. In: Proceedings of the 6th International Conference on Information Technology and Multimedia; 2014 Nov 18–20; Putrajaya, Malaysia. p. 212–6. doi:10.1109/ICIMU.2014.7066632.
27. Lee DY, Jo JC, Lim HS. User sentiment analysis on Amazon fashion product review using word embedding. J Korea Converg Soc. 2017;8(4):1–8. doi:10.15207/JKCS.2017.8.4.001.
28. Mao Y, Liu Q, Zhang Y. Sentiment analysis methods, applications, and challenges: a systematic literature review. J King Saud Univ Comput Inf Sci. 2024;36(4):102048. doi:10.1016/j.jksuci.2024.102048.
29. Yuan Y, Lam W. Sentiment analysis of fashion related posts in social media. In: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining; 2022 Feb 21–25; Online. p. 1310–8. doi:10.1145/3488560.3498423.
30. AbdelFattah M, Galal D, Hassan N, Elzanfaly DS, Tallent G. A sentiment analysis tool for determining the promotional success of fashion images on Instagram. Int J Interact Mob Technol. 2017;11(2):67–73. doi:10.3991/ijim.v11i2.6563.
31. Wankhade M, Rao ACS, Kulkarni C. A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev. 2022;55(7):5731–80. doi:10.1007/s10462-022-10144-1.
32. Tran LQ, Van Duong B, Nguyen BT. Sentiment classification for beauty-fashion reviews. In: 2022 14th International Conference on Knowledge and Systems Engineering (KSE); 2022 Oct 19–21; Nha Trang, Vietnam. p. 1–6. doi:10.1109/KSE56063.2022.9953782.
33. Kavi Priya S, Porkodi J, Kaviya Sri AN, Shweatha K. Empowering sentiment analysis for improved fashion choices. Int J Eng Technol Manag Sci. 2023;7(5):319–24. doi:10.46647/ijetms.2023.v07i05.037.
34. Dimovska J, Angelovska M, Gjorgjevikj D, Madjarov G. Sarcasm and irony detection in English tweets. In: ICT Innovations 2018. Engineering and Life Sciences: 10th International Conference, ICT Innovations 2018; 2018 Sep 17–19; Ohrid, Macedonia. p. 120–31. doi:10.1007/978-3-030-00825-3_11.
35. Farías DIH, Patti V, Rosso P. Irony detection in twitter: the role of affective content. ACM Trans Internet Technol. 2016;16(3):1–24. doi:10.1145/2930663.
36. Tingre Y, Pingale R, Choudhary N, Kale A, Hajare N. Sarcasm detection of emojis using machine learning algorithm. In: 2024 IEEE International Conference on Contemporary Computing and Communications (InC4); 2024 Mar 15–16; Bangalore, India. p. 1–4. doi:10.1109/InC460750.2024.10649041.
37. Angeli A, Piccolomini EL, Marfia G. Learning about fashion exploiting the big multimedia data. In: 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC); 2018 Sep 9–12; Bologna, Italy. p. 48–51. doi:10.1109/PIMRC.2018.8581037.
38. Sunil KV, Vedashree CR, Sowmyashree S. Image sentimental analysis: an overview. Int J of Adv Res. 2022;10:361–70.
39. Shimizu R, Kimura M, Goto M. Fashion-specific attributes interpretation via dual Gaussian visual-semantic embedding. arXiv:2210.17417. 2022. doi:10.48550/arXiv.2210.17417.
40. Sleiman R, Tran KP, Thomassey S. Natural language processing for fashion trends detection. In: 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET); 2022 Jul 20–22; Prague, Czech Republic. p. 1–6. doi:10.1109/ICECET55527.2022.9872832.
41. Sarkar S, Parmar M, Sandhu S. Entity set expansion for detecting fashion trends. In: 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA); 2019 Dec 16–19; Boca Raton, FL, USA. p. 162–7. doi:10.1109/ICMLA.2019.00033.

42. Nguyen DQ, Vu T, Nguyen AT. BERTweet: a pre-trained language model for English Tweets. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations; 2020 Nov 16–20; Online. p. 9–14.
43. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: a robustly optimized BERT pretraining approach. arXiv:1907.11692. 2019.
44. Pennington J, Socher R, Manning CD. Glove: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); 2014 Oct 25–29; Doha, Qatar. p. 1532–43.
45. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;30:1–11.
46. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. doi:10.1162/neco.1997.9.8.1735.
47. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv:1412.3555. 2014. doi:10.48550/arXiv.1412.3555.