IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 9, 2021, accepted August 23, 2021, date of publication August 30, 2021, date of current version September 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3108502

Early and Late Fusion of Emojis and Text to Enhance Opinion Mining

SADAM AL-AZANI AND EL-SAYED M. EL-ALFY, (Senior Member, IEEE)
College of Computing and Mathematics, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran 31261, Saudi Arabia
Interdisciplinary Research Center for Intelligent Secure Systems (IRC-ISS), KFUPM, Dhahran 31261, Saudi Arabia

Corresponding author: Sadam Al-Azani (sadam.azani@kfupm.edu.sa)

This work was supported by King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia.

ABSTRACT Opinion mining has gained increasing importance to draw insights from social media content to support decision making. Despite the explosive growth of efforts on linguistic analysis to detect and track people's opinions, more specifically when dealing with aspect-level opinion mining, the results are still away from generalization to real-world applications. Nowadays, emojis are getting excessively popular in social media communication as a complementary way to quickly express opinions and ideas in a visual manner. Two emoji-related issues are highlighted: ambiguity and misinterpretation of emojis' sentiment and tendency of persons to adopt emojis more in positive cases. This paper aims at investigating to what extent the usage of emojis can contribute to the automated detection of sentiment polarity of text messages with focus on Twitter posts in the Arabic language, a widely spoken language but has complex morphology and limited reliable resources for sentiment analysis. For this purpose, after an extensive review of the state-of-the-art of emojis-related work, a dataset is composed and several feature extraction methods are applied for both text and emojis modalities. Moreover, various early and late fusion techniques are proposed to combine both modalities at different levels including feature, score, decision and hybrid. The experimental results revealed that emojis features can significantly improve the classification results, especially when integrated with text at the score level.

INDEX TERMS Emojis, information fusion, opinion mining, sentiment analysis, social media, text mining.

I. INTRODUCTION

As the Internet is becoming an indispensable part of our lives, technologies have enabled users to generate online content to share knowledge and experience, express opinions and provide comments on various topics of interest. Consequently, opinion mining has gained importance in research as an emerging area in natural language processing (NLP), social media analysis, and web and text mining. It is defined as "the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes" [1]. Research on opinion mining covers a wide range of tasks such as subjectivity determination, polarity detection, affect analysis, opinion extraction, sentiment mining, emotion detection, and review mining, at different levels namely, documents, sentences, features or aspects [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Haiyong Zheng.

Different approaches have been proposed to address sentiment analysis tasks including: lexicons, machine learning and hybrid [3]–[10]. Additionally, several resources were designed, constructed and developed including datasets, lexicons, corpora, and software tools. However, these approaches and resources were mostly for the Western languages (e.g. English, Spanish and French) and mainly focused on text modality. Although these resources were well-validated and achieved reasonable empirical results, they are still challenging and is expected to encounter biased results when applied to morphological-rich and syntactical-rich languages such as the Arabic language or multicultural aspects.

The research on Arabic sentiment is relatively still in its early stage compared to the Western languages. With the increasing number of Arabic Internet users, the volume of online Arabic content in social media is tremendously increasing. Arabic content is evolving with its informal use to express ideas or concepts in social media in several different dialects, e.g. Egyptian, Levantine, Gulf, Iraqi, Maghrebi, and Yemeni. Dialectical Arabic lacks standardization, and is written in informal style with many variations. This results in further challenges with reiterated concerns for preprocessing and feature extraction [11]. The issue gets worse when dealing with short texts such as tweets (originally limited to a maximum of 140 characters before being doubled to 280 characters). Twitter is a universally popular example of several online microblogging platforms.

With the rapidly increasing popularity of emojis in social media to complement or condense the meaning of text, researchers studied their semantic relation to words [12]. According to emojitracker.com, which was launched in July 2013 to monitor the use of emojis in Twitter in realtime, until the mid of April 2021 there are more than 34 billion occurrences of emojis where the most popular emoji is "face with tears of joy" occurring more than 3.2 billion times. These numbers are rapidly increasing from day to day. For example, we noticed that the number of emojis available online increased by around eight billion occurrences of emojis from around 28 billion in the beginning of June 2019 to around 34 billion occurrences of emojis in the mid of April 2021. The number of "face with tears of joy" occurrences raised rapidly from 2.4 billion to 3.2 billion occurrences during the same period. Additionally, according to a survey by TalkTalk Mobile in United Kingdom,[1] 72% of young people (ages 18 to 25) found emojis more easier to communicate their thoughts. Emojipedia[2] provides an online collection of several categories of emojis reporting changes of emojis' symbols and their meanings in the Unicode Standard. The first version of Unicode standard that supported emojis is 6.0 published in October 2010 and it is planned to release Unicode 14.0 in September 2021. In the Unicode Standard 12.0, there are 3,019 symbols of emojis[3] and 13.1 contains 3,521 emojis.[4] An earlier term to emojis was coined by a computer scientist at Carnegie Mellon University in 1982 is emoticon, which is defined in Longman dictionary as "a special sign that is used to show an emotion in an email and on the Internet, often by making a picture". The first form of emojis was text-based emoticons such as:-) and:-(, and was used to express the writer's feelings and reduce language ambiguity in communication. Nowadays, emojis are becoming more general and can represent not just facial expressions but also concepts and ideas such as celebration, weather, vehicles and buildings, foods and drinks, animals and plants, emotions and feelings, and activities [13]. Though the word emoji was originally used on mobile phones in Japan in 1999 where '*e*' means picture and '*moji*' means character, the use of visual symbols for writing goes back to the ancient Egyptians' Hieroglyphics. Nowadays, it is used to describe all pictographs used in social media such as 😃, 😒, 🐎, 🏆, 🚴, 📻 and 🔋. Depending on the system or platform used, emojis can appear in different forms, sizes and colors.

Emojis allow more expressive messages with non-verbal visual elements. It was reported that users tended to use emojis with clear semantic meanings more frequently [14] and entity-related emojis were used to replace words. Emojis often play a complementary role in a message especially those with clear sentiment polarity. Additionally, emojis not just express, stress, or disambiguate sentiments or emotions but also are capable of expressing useful relational roles in conversation. They might help to clarify the intent of text and consequently contribute to reduce ambiguity in short text-based social media platforms such as Twitter. In the context of sentiment analysis and in contrary to text, emojis are domain and topic independent [15], [16] but their usage is still globally inconsistent and depends on context and culture. In automated text mining, they have been exploited to a limited extent, mainly for automated data annotation [17].

Emojis can provide helpful features to compensate textual features for opinion mining and sentiment analysis. Building upon earlier work [18]–[20], this study aims at exploring the extent and effectiveness of emojis usage to enhance sentiment polarity detection for opinion mining. Several emojis-based features are extracted and adopted to build computational sentiment detection models. The paper explores and evaluates various single- and multi-level fusion of emojis with a variety of textual features to improve polarity classification performance compared to single modality machine learning classifiers. Besides structural features, text is represented using bag-of-words term-frequency inverse-document frequency (*tf-idf*), Latent Semantic Analysis (LSA), and two methods of word embeddings. Emojis are represented in four ways: emojis frequencies, lexicon-based sentiment weighted scores, and two emoji embedding models. Intensive experimental work is conducted on a larger dataset of tweets containing emojis to evaluate and compare the performances of various models. Moreover, this study enriches the theory of emojis through providing a comprehensive review of the state-of-the art of emojis-related work in social media. The main contributions of this study are as follows:

- Provide a comprehensive literature review and classification of emojis-related studies in social media analysis.
- Propose a number of representations of emojis for computational modeling and evaluating them to predict sentiments in microblogs.
- Investigate various fusion approaches to combine emojis with Arabic textual features to improve the sentiment polarity classification results.
- Conduct intensive empirical analysis of various models using different features extraction methods, machine learning classifiers, and performance measures.

II. BACKGROUND AND RELATED WORK

In this section, we present essential background and review the state-of-the-art techniques for emojis in social media analysis. First, a summary of existing review studies for

[1] http://salesholding.talktalk.co.uk/
[2] https://emojipedia.org/
[3] https://www.unicode.org/emoji/charts-12.0/emoji-counts.html
[4] https://www.unicode.org/emoji/charts-13.1/emoji-counts.html

textual based opinion mining is presented. We also provide a taxonomy of emojis' related works. Lastly, we survey studies related to emojis' sentiment.

A. TEXT-BASED OPINION MINING

Significant attention of the research community has been attracted towards opinion mining and considerable effort has been conducted to text-based sentiment analysis in a variety of languages. This resulted in an increasing number of publications from year to year. We refer interested readers to comprehensive surveys on the opinion mining and sentiment analysis process and related tasks, algorithms and applications in [2] and [21]–[23]. For instance, Giachanou and Crestani [4] reviewed sentiment analysis approaches in Twitter and categorized them according to the utilized techniques with a discussion of research trends of the topic and its related fields. Though text-based opinion mining approaches have proven to be extremely useful in the field of sentiment analysis, they suffer from problems such as domain, topic, and temporal dependence. Therefore, this study aims at alleviating such issues through incorporating other sources of information with text such as visual modality (emojis) and evaluating its effectiveness to detect polarities of microblogs.

B. EMOJIS IN SOCIAL MEDIA

We first defined a taxonomy for emojis based on their applications, representations, issues, and approaches; as depicted in Figure 1. Next, the prior works were categorized and analyzed according to the proposed taxonomy.

1) APPLICATIONS

Emojis have been utilized to build different resources including emoji-embedding models [24], [25] and lexicons [13], [17]. Supervised machine learning opinion mining approaches require annotated datasets, which is a tedious, labor-intensive and time consuming task. Emoticons/emojis has been adopted to facilitate annotation of training datasets [26]. In [15], the authors explored a form of machine learning known as distant supervision, to study the topic-, domain- and temporal dependency of text sentiment. The dataset is weakly labeled with the help of using smile and frown emoticons.

A similar approach to label datasets has been followed in [27] for detecting emotions. However, it has been shown in [28] that the performance of classifiers was always higher on noisy labeled data when using emojis. It is not recommended to use the entire training dataset which is annotated automatically for building a classifier because of the noise in data. Therefore, Liu *et al.* [29] proposed a model called emoticon smoothed language model (ESLAM) for utilizing and smoothly integrating both manually and noisy labeled data for building a training dataset. ESLAM first utilizes manually labeled data to train language models and then the noisy labeled data is utilized for smoothing. Another task is to predict the most likely emoji given the text of a tweet [12], [30], [31]. Moreover, emojis have also been employed to detect polarity in tweets [18], [32], [33].

Emojis have been successfully applied for user verification task in [34] to determine whether an unknown tweet was written by a certain author/user or not. Moreover, a transfer learning approach using emojis have been proposed in [35] to detect irony in the Persian language.

2) REPRESENTATIONS

The ''representation'' concept differs from ''appearance''. While the first refers to how emojis are handled in different tasks by mapping them into features to build computational models. By contrast, the latter is platform dependent and means how emojis look/appear for users on different applications or devices. Emojis can be represented in different forms depending on the specific task in which they are used. The emoji-based feature extraction can use various syntactical or contextual relations and can use values of different types [18], [20], [36]. For example, they can use binary representation (0 or 1) referring to whether an emoji exists in a particular instance or not, integer numbers such as counting their occurrences in instances, real number such as their existence likelihood or intensities.

The number of emojis, average sentiment score of all emojis per post, number of positive emojis and number of negative emojis were used as the main features for polarity detection of Uzbek movie reviews in [33]. Similar to term/word/phrase polarity lexicons, emojis have their lexicons in which each emoji has a polarity or sentiment score. Embedding models are others form to represent emojis, in which each emoji is represented in Vector Space Model (VSM) as a vector of real numbers generated using well-known embedding tools (similar to Word2Vec word embedding [37], [38]).

3) POLYSEMY

Polysemy refers to coexistence of several meanings of a particular entity. In the context of emojis, several factors can lead to ambiguity in interpreting its meaning, e.g. diversity of users [39]–[41], genders [42]–[45], locations [46], [47], and cultures [48]. Emojis also appear differently on different platforms [39], [49], [50]. As mentioned in [39], [49], there are significant variations of people's interpretations of emojis within and across platforms. For example, an emoji appears differently in Twitter, Facebook, Apple, EmojiOne, and Samsung, and even in different versions of Android.

Figure 2 shows an example for the emoji of ''cow face'' appearance in different platforms based on version 11.0 of full emoji list.[5] Another major factor that causes misconception is the high similarity of different emojis such as octopus 🐙 and squid 🦑. As another example, users misunderstood the use of ''pile of poo'' 💩 which has negative polarity with a sentiment score of -0.116 in ESR (Emoji-Ranking Sentiment) lexicon and it is used as ''ice cream'' which is definitely positive with

[5] https://unicode.org/emoji/charts/full-emoji-list.html

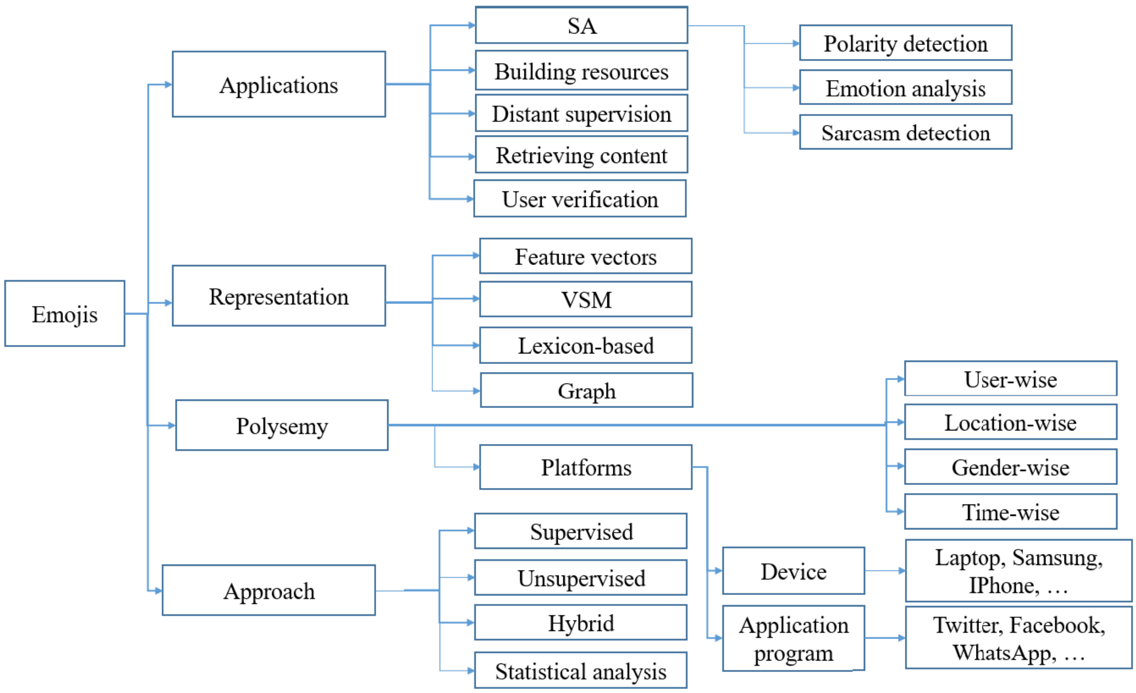


FIGURE 1. Emojis taxonomy.

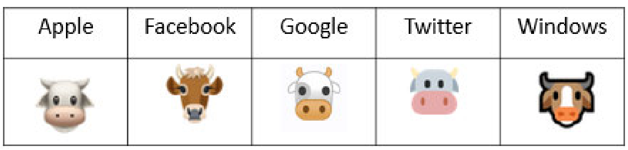


FIGURE 2. Example of "cow face" emoji appearance on different platforms.

a sentiment score of 0.212 in ESR. Table 1 shows a set of issues related to emojis polysemy and proposed solutions in the literature.

Variation of emoji sentiment perception from writers to readers viewpoints is another factor. Berengueres and Castro [51] reported that there is 82% agreement in emoji sentiment perception from writers to readers viewpoints. The disagreement concentrates in negative emojis, where the authors reported feeling 26% worse than perceived by readers. Emoji usage was not found to be correlated with author moodiness. Emoji sentiments are interpreted in a different way according to the platform. It was concluded in [39] that there is disagreement in sentiments and semantics of 22 emojis on five different platform renderings, which is a major issue especially for cross platforms. On the other hand, Cui *et al.* [52] studied the use of tweets whose sentiments conflict to some extent to emojis in training phase. The main findings were that the optimal training dataset for determining tweets' sentiment is reasonable and followed the distribution of sentiment in real tweet streams.

Users tend to use emojis with positive polarity or happy emotion more than other polarities or emotions [32], [51]. This issue is dealt with as an imbalanced class problem and was addressed through generating synthetic instances for the minority class using a bagging ensemble in [32]. Additionally, young people tend to use emojis more frequently [41], [53]. Emojis are analyzed and studied in different social media platforms including: Twitter such as [18], [19], [24], [32], [54], Facebook [55], [56], WhatsApp [41], Instagram [31], electronic mail [57], etc. Furthermore, they are studied in multiple social media in [58]. The main findings were that the most popular emojis in one social media are not as popular as in the others. For instance, emojis sentiment polarity in Twitter is high but the overall number of emojis is less than Facebook. The sentimental value of emojis is more meaningful when there are multiple emojis in one notification.

Significant attempts and efforts have been performed to clarify meanings and reduce misconception. For example, Wijeratne *et al.* [59], [60] presented the first and largest machine readable sense inventory for emoji (EmojiNet), which links Unicode emoji representations to their English meanings extracted from the Web. It is composed of a dataset of 12,904 sense labels over 2,389 emojis. Each emoji sense is associated with context words trained using Word2Vec Skip Gram model.

4) APPROACHES

Figure 1 shows various classes of approaches that have been applied in several studies related to emojis. Barbieri *et al.* [24] built several Skip Gram embedding models for emojis and words using a dataset of 10 million tweets by mapping in

TABLE 1. Polysemy issues related to emojis and proposed solutions in the literature.

Ref	Issue	Methods/strategies	Datasets/Tweets	Results	Descriptions
[59]	Emojis ambiguity	EmojiNet	3,206	Acc: 85.18%	Neighborhood-based image processing algorithm to integrate emoji resources and the most frequent sense-based sense-based word sense disambiguation algorithms to assign meanings to emoji sense labels
[60]	Emojis ambiguity	Improved EmojiNet	12,904	Acc: 83.53%	Word embedding models and vendor-specific emoji senses were used to improve EmojiNet
[52]	Sentiment inconsistent	Machine learning	9,000	Prc-Rec curve	Word embeddings + CNN
[32]	Emojis' polarity imbalance	Over-sampling + machine learning	1,101	F1: 83.73%	A method to deal with the class imbalance problem based on Bootstrap Aggregating (Bagging) algorithm and SMOTE.

the same vector space both words and emojis. The tweets were posted by the USA users. The models were then evaluated with semantic similarity experiments and compared with human assessment. Barbieri *et al.* [12] trained several supervised classifiers based on deep learning, Long Short-Term Memory (LSTM) networks, for predicting appropriate emojis from corresponding tweets. The main conclusion was that computational models can identify the underlying semantics of emojis better than humans do. Identifying the linguistic purpose of emojis is addressed in [61]. Statistical analysis studies were conducted to analyze the pattern of emojis [44], [48], [62], [63]. For example, the work in [44] provides a statistical analysis to explore the emojis usage in smart phones from gender perspectives. It was found that males and females varied in emojis usage significantly which confirms the findings of [43]. In this context, [45] reported that females trend to use emojis more than males to express their sentiments and emotions on social media. Another statistical analysis is conducted by [63] to investigate the functions of emojis from the perspective of original senders. The main finding was that the social and linguistic functions of emojis are complex and varied. It was reported in [62] that Twitter users tend to reduce their usage of emoticons and shift dramatically to emojis. The sentiment of emojis has been considered in some studies using various approaches as will be explained in the following subsection. Attention-based network models have also been proposed to improve emoji-based sentiment analysis on microblog posts in a number of studies such as [64], [65]

On the other hand, predicting emojis from text has received a significant attention nowadays such as [35], [65], [66].

C. EMOJIS AND OPINION MINING

In the literature, the emoticons or a limited number of emojis were considered without taking into account the full extent of the sentiment they convey [67]. Emoticons have only been considered as elementary/extra features for sentiment analysis tasks such as the number of negative or positive emoticons [68], [69] or the presence of positive and/or negative emoticons [27], [69], [70]. In addition, emoticons were converted to their textual meanings, as a preprocessing step, intuitively or using a general emoticons lexicon [68], [71]–[73]. The presence status of emoticons and their count were considered and evaluated to detect emotion in Tweets in [74]. The usage of emojis in social network is analyzed in [75].

Some attempts have considered the construction of emoji-related resources for NLP tasks such as datasets, lexicons, dictionaries and even tools. Emojis' lexicons have been constructed for sentiment analysis tasks. Hogenboom *et al.* [17] presented a lexicon-based polarity classification method to evaluate how emoticons convey sentiment. This method was evaluated on 2,080 Dutch tweets and forum messages, which all contain emoticons. They reported that the sentiment of emoticons tends to dominate the sentiment conveyed by textual cues and forms a good proxy for detecting the polarity of text. ESR is a systematic lexicon of emojis built for sentiment analysis by [13]. It is composed of 969 emojis; where for 751 of them each occurs more than four times. Each emoji is assigned a sentiment score computed from 1.6 million tweets in 13 European languages by the sentiment polarity (negative, neutral, or positive).

Another emoji sentiment lexicon with 840 emojis using an unsupervised sentiment analysis system was constructed by [76]. It was built based on the definitions given by emoji creators in Emojipedia by analyzing the sentiment of informal texts in English and Spanish. Moreover, lexicon variants were created by considering the sentiment distribution of the informal texts accompanying emojis. Donato and Paggio [77] created an annotated corpus for analyzing the informative patterns of emojis.

Emoji2Vec is another embedding method that was trained in [54], using Unicode emojis descriptions. It was found that the performance of Emoji2Vec model is better than augmented vectors used in [24]. The authors in [18] evaluated both unidirectional and bidirectional LSTM and the simplified variant Gated Recurrent Unit (GRU) models to detect sentiment polarity of Arabic microblogs using emoji-based features. The performance was compared to baseline traditional learning methods. It was concluded that LSTM and GRU based models performed significantly better than traditional models with a slight difference between them with best

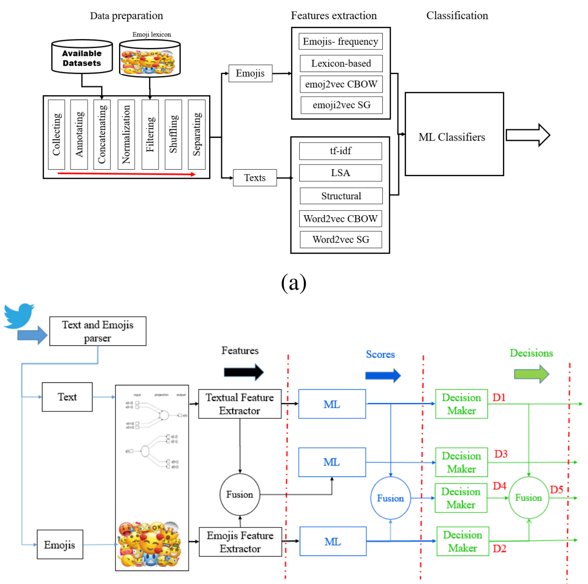


FIGURE 3. A general overview of the proposed approach (a) the basic model of single modalities of text and emojis (b) the fusion model of both modalities at different levels.

results attained when using bidirectional GRU. An approach is presented in [32] to detect polarities while handling the issue of skewed use of emojis more often in positive tweets. This issue was dealt with as an imbalanced class problem and was addressed through generating synthetic instances for the minority class using Synthetic Minority Oversampling Technique (SMOTE) with Bootstrap Aggregating (Bagging) algorithm. Emojis-based features were extracted from three datasets with different imbalance ratios such that each instance contains at least an emoji. It was concluded that the highest results are obtained using the balanced bagging classifier.

III. PROPOSED OPINION MINING FRAMEWORK

The layouts of the proposed opinion mining approach for single modality and fused modalities are illustrated in Figure 3.

A. DATASET PREPARATION

Supervised machine learning based approaches require labeled datasets in order to train the predictive model. This often requires considerable human effort to build datasets. We adopted five publicly existing sentiment related datasets [78]–[82] and augmented them with additional samples containing text and emojis. The augmented tweets are collected using Twitter API during the period from 1^{st} December 2017 to 15^{th} December 2017. They are annotated manually as positive or negative using two annotators with 100% agreements. In order to avoid the ambiguity of emojis understanding, the annotators were asked to annotate the collected instances based on both text and emojis. The annotation task was just conducted for the instances that contain emojis. A summary of the combined dataset is described in Table 2. Each sample has at least one emoji from those included in ESR. First of all, textual emoticons were normalized and transformed to their corresponding graphical symbols. A total of 2,091 instances distributed as 1,216 positive and 875 negative is finally kept. The reason of having more positive instances than negative instances is due to the observation that users likely use emojis when they have positive mode [51]. Finally, the instances are shuffled randomly in the dataset. The issues related to emojis' misconception and ambiguity, presented in the Polysemy subsection, make emoji classification more challenging than text classification.

B. PREPROCESSING

The preprocessing step is conducted at three levels relying on the type of features: emojis based features, structural-based features or text-based features. For emoji-based features (i.e., Emojis frequency, Lexicon-based, Emojis-CBOW and Emojis-Skip gram), emoticons are transformed into their corresponding graphical symbols (emojis). No preprocessing operations are conducted for structural features, due to the sensitivity of such features to the preprocessing step. For example, elongation is a structural feature but it is eliminated for other types of features. In case of textual features (*tf-idf*, LSA, word2vec CBOW and Skip grams) different operations are conducted on the text including removing noisy symbols, non-Arabic characters, diacritical and punctuation marks, links, and repeated characters. These preprocessing operations were conducted to generate the pretrained word2vec models. Using the same preprocessing operations of the pretrained models ensures including all common vocabularies. Table 3 shows the conducted preprocessing operations and corresponding types of features.

C. UNI-MODAL OPINION MINING

Feature engineering is a common and serious step in predictive analytics. A number of feature extraction methods are explored in this study utilizing a single modality, either text or emojis. For emojis, four feature extraction methods are applied and evaluated: emojis frequency, lexicon-based features and two different types of emoji embedding based features. On the other hand, five textual feature extraction methods are evaluated including *tf-idf*, LSA, structural features and two different Word2Vec embedding forms. Other textual feature extraction methods can also be explored such as non-negative matrix factorization and latent Dirichlet allocation [83]. Features extracted by various methods are separately evaluated using two machine-learning classifiers. The following subsections provide more details on the investigated feature extraction methods.

1) EMOJIS-BASED FEATURES

In this study, we investigated four different representations of emojis for opinion mining as follows.

TABLE 2. Dataset description showing various sources and number of all instances and those contain emojis.

Source	All instances			Instances with emojis		
	Negative	Positive	Total	Negative	Positive	Total
ArTwitter [79]	958	993	1951	4	3	7
ASTD [78]	812	777	1589	19	64	83
QCRI [80]	377	377	754	42	114	156
Semeval-2017 Task4 Subtask#A [81]	3364	2257	5621	350	555	905
Syria [82]	1350	448	1798	47	50	97
Additional instances:				413	430	843
Total	6861	4852	11713	875	1216	2091

TABLE 3. Type of preprocessing operations.

Features' type	Remove noisy symbols	Remove non-Arabic	Remove punctuation	Remove links	Remove elongation	Emoticons to emojis
Emojis based	✗	✗	✗	✗	✗	✓
Structural	✗	✗	✗	✗	✗	✗
$tf\text{-}idf$	✓	✓	✓	✓	✓	✗
LSA	✓	✓	✓	✓	✓	✗
Word embedding	✓	✓	✓	✓	✓	✗

a: EMOJIS FREQUENCY

The count of occurrences of each emoji in each instance in the prepared dataset is calculated. Feature vectors are prepared for the 2091 instances, each of dimension 429.

b: LEXICON-BASED FEATURES

In this method, we utilized the emojis sentiment based lexicon ESR to represent emojis using lexicon-based scores. The tweets were annotated as negative (-1), neutral (0) or positive ($+1$) by 81 annotators. Depending on the sentiment of the tweets in which each emoji occurs, the emoji is assigned three values $\{p_- = N_-/N, p_o = N_o/N, p_+ = N_+/N\}$ representing its likelihood to appear in a specific sentiment category, where N_- is the number of the emoji's occurrences in tweets with negative sentiment, N_o is the number of the emoji's occurrences in tweets with neutral sentiment, N_+ is the number of the emoji's occurrences in tweets with positive sentiment, and $N = N_- + N_o + N_+$. Then, a sentiment score (ss) of each emoji is calculated by $ss = p_+ - p_-$. Figure 4 shows the top-5 emojis in the lexicon. The feature vectors are extracted based on the scores and number of occurrences. For emoji i in tweet j, f_{ij} is computed as:

$$f_{ij} = ef_{ij} * ss_i \tag{1}$$

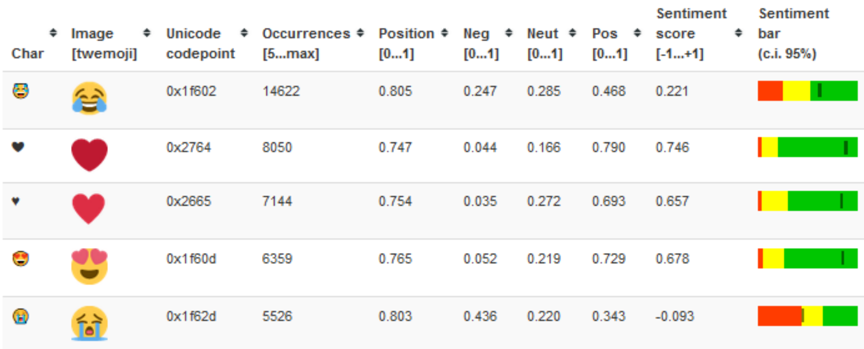
where ef_{ij} is the frequency of emoji i in tweet j while ss_i is the sentiment score of emoji i in the lexicon.

c: EMOJI EMBEDDING BASED FEATURES

Similar to word embedding, we constructed emoji embedding models using either Continuous Bag-of-Words (CBOW) or Skip Gram (SG) neural networks. These models are trained using an emoji dataset to map emojis into d-dimensional embedding. A dataset of one million tweets containing emojis[6] is used to generate the emojis embedding models. Two parameters that affect the model quality are dimensionality and size of context window. The higher dimensionality, the higher the quality until reaching some point. It is reported that the typical vector dimensionality ranges between 100 and 1000. The context window on the other hand determines how many elements before and after a given element would be included as context. Determining the size of the context window depends on several criteria including the adopted technique (i.e., CBOW or Skip Gram) and the genre of data used to learn the embedding models (tweets, paragraphs, articles, etc.). The parameters used to generate emojis embedding models are depicted in Table 4.

Several semantics and syntactical relations can be obtained using the generated emojis embedding models. Figure 5 shows four different information types that can be obtained. For each query the 10 (or less based on the availability) highest probability answers are retrieved. They are ordered based on their probabilities from left to right and from up to down. The first query is to retrieve the relation (King + man-woman) with the 10 highest probability using CBOW and Skip Gram. Both techniques agree in the first answer which is 👑 crown. This is similar to the well-known Word2Vec example: King − Man + Woman ≈ Queen. CBOW and Skip Gram differ in the order of some other retrieved answers while others are common. The second query is to retrieve the most related emojis for a concept or word; three examples are shown for this query. Another query is to retrieve the most likelihood emojis related semantically to a certain emoji and two examples are shown in the same

[6]https://github.com/jiali-ms/emoji2vec/tree/master/data



Char	Image [twemoji]	Unicode codepoint	Occurrences [5...max]	Position [0...1]	Neg [0...1]	Neut [0...1]	Pos [0...1]	Sentiment score [-1...+1]	Sentiment bar (c.i. 95%)
😂	😂	0x1f602	14622	0.805	0.247	0.285	0.468	0.221	
❤	❤	0x2764	8050	0.747	0.044	0.166	0.790	0.746	
♥	♥	0x2665	7144	0.754	0.035	0.272	0.693	0.657	
😍	😍	0x1f60d	6359	0.765	0.052	0.219	0.729	0.678	
😭	😭	0x1f62d	5526	0.803	0.436	0.220	0.343	-0.093	

FIGURE 4. Top-five emojis in ESR.

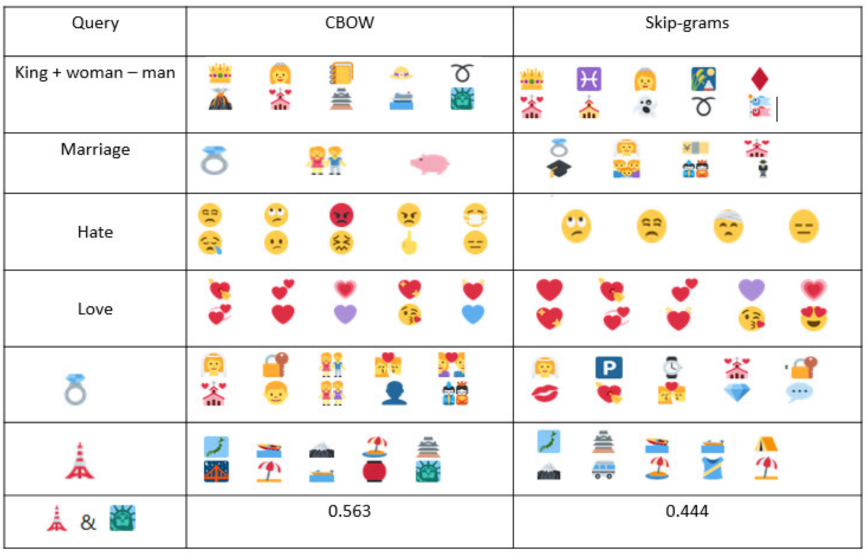


FIGURE 5. Different queries for empji-embedding CBOW and Skip-Gram models.

TABLE 4. Adopted parameters for training the emojis embedding models.

Model	Dimension	Window size	Sample	Negative	Min count	Iterations
CBOW or Skip Gram	300	5	$1 \times e^{-3}$	10	10	10

figure. The last query is to represent the similarity of two emojis.

Now, assume a tweet T has n emojis after filtering words, $T = \{e_1, e_2, \ldots, e_n\}$, where e_i is the i^{th} emoji in T. Let $x_i \in \mathbb{R}^d$ be the d-dimensional emoji vector from Emoj2Vec model corresponding to the i^{th} emoji in T. To compute the feature vector for T, the feature vectors of emojis are arranged in a matrix column-wise then the row-wise average is computed as illustrated in Figure 7 to obtain the feature vector:

$$f_i = \frac{1}{n}\sum_{k=1}^{n} x_{ki}, \quad i = 1, 2, \ldots, d \tag{2}$$

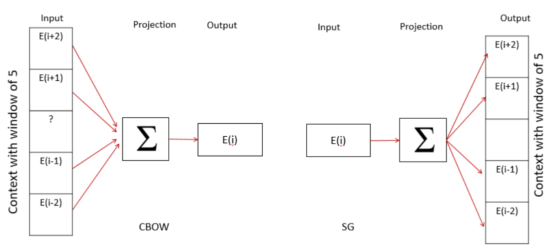


FIGURE 6. CBOW and Skip Gram architectures.

2) TEXT MINING FOR STRUCTURAL AND TEXTUAL FEATURES

a: STRUCTURAL FEATURES

We considered the following structural features as well:

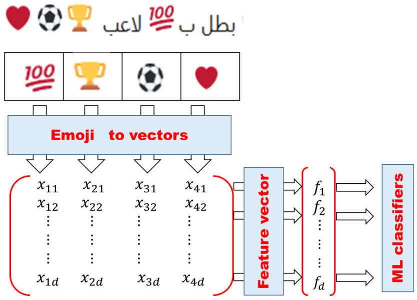


FIGURE 7. Sentence representation using emojis embedding.

- Count of links: to determine how many links in a certain tweet. Its value is equal to the number of links in the tweet otherwise zero value is assigned.
- Count of mentioned accounts: if a tweet mentions twitters' accounts, the number is assigned otherwise it takes zero value.
- Count of hashtags: if a tweet contains hashtags, the count of hashtags is assigned otherwise it takes zero value.
- Count of emojis in the tweet.
- Does the tweet have elongation words? elongation words mean some characters are repeated such as Nooooo! Hiiiiiiiiii!.
- Does the tweet contain dialectical marks?
- Length of the tweet in words.
- Length of the tweet in characters.

b: *tf-idf* FEATURE EXTRACTION

This is a standard term-weighting scheme in text mining to reflect how important a word to a document in a collection. It can be calculated in slightly different ways but in this study we computed it as described next. A given corpus needs first to be represented as a sparse matrix in which each row represents a unique term (word level uni-gram) and each column represents a document (or tweet in our case). The matrix content represents the number of times term i appears in document j is referred to as f_{ij}. Then, the corresponding *tf-idf* is computed as $tf_{ij} \times idf_i$, where tf_{ij} is computed by dividing f_{ij} by f_{*j} which is the number of terms in document j:

$$tf_{ij} = f_{ij}/f_{*j} \tag{3}$$

and idf_i is computed as follows:

$$idf_i = log_2 \frac{N}{n_i} + 1 \tag{4}$$

where N is the number of documents and n_i is number of documents containing the term i. Finally, the *tf-idf* of term i in document j, is:

$$tf\text{-}idf_{ij} = (f_{ij}/f_{*j}) \times log_2 \frac{N}{n_i} + 1 \tag{5}$$

c: LSA FEATURE EXTRACTION

Latent Semantic Analysis (LSA) [84], [85] is a fully automated statistical approach for analyzing relations between terms and documents by combining terms that are highly correlated to produce a reduced set of concepts. It is based on an unsupervised learning technique (clustering), and assumes that terms with common meaning occur in similar paragraphs. It builds a term-document matrix from a corpus, and aims at exposing some useful similarity structures for related text-analysis tasks and information retrieval. LSA starts with *tf-idf* as an initial step and then applies singular value decomposition (SVD) to perform dimensionality reduction on the *tf-idf* vectors. If we denote to matrix generated in *tf-idf* as X, then X can be decomposed using SVD into the product of three unique matrices. If we assume there are m terms and n document and we need to obtain r concepts. Then $X = U\Sigma V^T$, where X is $m \times n$ matrix representing the original term-document matrix, U is $m \times r$ matrix representing the left singular vectors of the r concepts, V is $n \times r$ matrix representing the right singular vectors of the r concepts, and Σ is $r \times r$ diagonal matrix containing scalar positive values representing the singular values or strengths of various concepts. Both U and V are orthonormal.

d: WORD-EMBEDDING BASED FEATURES

Embedding techniques are recognized as being powerful for natural language processing. They can learn high-quality compact vector representations of words/terms/phrases from a large amount of unstructured text data. The resulting vectors are close in the feature space for elements with similar meanings or used in similar contexts. Word2Vec [37], [38] is an example approach developed by a team at Google in 2013 and has two neural network architectures: CBOW and Skip Gram as illustrated in Fig. 6. CBOW and Skip Gram have similar algorithms but the former predicts a word given its context whereas the latter predicts a context given a word. In this study, we adopted pretrained CBOW and Skip Gram models generated from an Arabic Corpus of 77,600,000 tweets written in modern standard Arabic and dialectal Arabic [86]. A dimensionality of 100 and a window size of 3 were used for generating both models. Having a feature vector for each word, a matrix is created for each tweet where each column is a feature vector for one word, then the row-wise average vector is computed to represent the tweet (similar to Fig. 7 but the input will be words instead of emojis).

IV. TEXTUAL-VISUAL OPINION MINING

To integrate both emojis and text based features, several fusion methods are investigated at different levels as described in the following subsections.

A. SINGLE-LEVEL FUSION APPROACHES

Information fusion for predictive modeling is mainly carried out at three levels: feature level, score level and decision level. Feature-level fusion is also known as an early-fusion technique whereas both score-level and decision-level fusion are known as late-fusion techniques. Feature-level fusion provides more information than score level which also provides more information than decision level. In the following, we discuss each approach and highlight its strengths and shortcomings.

1) FEATURE-LEVEL FUSION

In this approach, features extracted from different sources are combined by concatenation to form an augmented feature vector in a higher-dimensional feature space. Integrating information from various sources at this level is straightforward and can lead to improved prediction. Hence, it is widely used in the literature. However, some features might be redundant or noisy. Moreover, this fusion level may suffer from the curse of dimensionality (too many features) and the large variability in the features scales. This requires the application of some feature a reduction technique such as principal component analysis (PCA), and a normalization scheme such as min-max, z-score, or tanh-estimator normalization.

Feature-level fusion, in this study, is carried out through simply concatenating the extracted textual and emojis features. Mathematically, let $F = \{f_1, f_2, \ldots, f_n\}$, $F \in R^n$ represents the textual feature vector with length n and $E = \{e_1, e_2, \ldots, e_m\}$, $E \in R^m$ represents the emojis feature vector with size m. Combining F and E results in a new feature vector $C = \{f_1, f_2, \ldots, f_n, \ldots, e_1, e_2, \ldots, e_m\}$, $C \in R^k$, with size of $k = n + m$. Features are then normalized using min-max scheme to produce $C_{norm} = \{f'_1, f'_2, \ldots, f'_n, e'_1, e'_2, \ldots, e'_m\}$ as follows:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{6}$$

where x is the original feature value in the range (x_{min}, x_{max}) and x' is the corresponding normalized value in the range (0, 1). For dimensionality reduction, we applied PCA with the criteria of selecting a number of components such that the amount of variance that needs to be explained is greater than 0.99

2) SCORE-LEVEL FUSION

Score level is a common late-fusion technique in data mining. Multiple prediction models are built one for type of feature vectors resulting in normalized scores in the range (0, 1). Intuitively, these scores represent the similarity or likelihood of each category for the input instance. After normalization, scores are combined by various methods and in this study we investigate three rules. Assume that there is k matchers, $\{M_1, \ldots, M_k\}$, where s_i is the normalized score of matching the input instance to a particular category using matcher M_i, then the overall score (S) is computed by one of the following rules:

Sum rule:

$$S = \sum_{i=1}^{k} s_i \tag{7}$$

Product rule:

$$S = \prod_{i=1}^{k} s_i \tag{8}$$

Max rule:

$$S = max\{s_1, s_2 \ldots, s_k\} \tag{9}$$

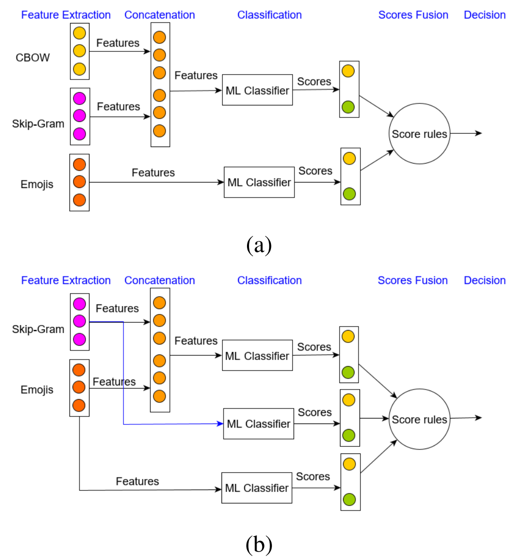


FIGURE 8. Hybrid fusion of feature and score levels.

3) DECISION-LEVEL FUSION

Decision-level fusion is another late-fusion technique in which the final/global decision is determined based on the local decisions of each individual prediction model. This level is easy to be implemented but similarly to score level it is often computationally expensive due to the various classification methods. If we assume k decision makers (models) denoted as $\{DM_1, \ldots, DM_k\}$, such that d_i is the decision made by DM_i for a given input instance, the final decision can be made from the local decisions by voting, i.e. the final decision is the most frequent decision (mode operation) of local decisions: $\hat{y} = mode\{d_1, d_2 \ldots, d_n\}$.

B. MULTI-LEVEL HYBRID FUSION

Features, scores and decisions can be mixed together at multiple levels. This may result in an improved model but it requires lots of efforts and computational complexity. In this study, we investigate combing features at the first level then combining the resulting scores from the prediction models at the second level as illustrated in Figure 8.

V. EXPERIMENTS

A. EVALUATION MEASURES

There is no single measure that can serve as a reliable measure for performance evaluation, especially due to the nature of uneven distribution of the sentiment classes in most real-world sentiment datasets [87]. Accuracy alone is a biased measure toward the majority class in imbalanced classification problems. Therefore, several evaluation measures are considered in addition to the accuracy (Acc) for assessing and comparing the performance of various models. In this sense, Precision (Prc), Recall (Rec), F_1 score, and Geometric Mean (GM) score are more suitable. The reported measures, except accuracy, are the weighted average of those computed for each class. The receiver operating characteristic curve (ROC)

TABLE 5. Comparison of SVC and LR machine-learning approaches using textual and emojis features extractors; highest results are in bold.

Features	Classifiers	Rec	Prc	F_1	GM	Acc
(a) Textual based features						
tf-idf	SVC	72.55 ±2.11	73.59 ±1.71	72.63 ±2.13	72.74 ±1.92	72.55 ±2.11
	LR	**74.41** ±2.31	**75.61** ±1.84	**74.49** ±2.29	**74.75** ±2.01	**74.41** ±2.31
LSA	SVC	70.83 ±2.48	71.98 ±2.85	70.94 ±2.53	71.10 ±2.93	70.83 ±2.48
	LR	**71.78** ±2.51	**72.22** ±2.70	**71.86** ±2.55	**71.50** ±2.83	**71.78** ±2.51
Structural	SVC	55.90 ±4.59	42.80 ±14.15	41.95 ±3.31	50.30 ±1.61	55.90 ±4.59
	LR	**59.16** ±2.82	**61.28** ±2.54	**59.32** ±2.85	**59.99** ±2.63	**59.16** ±2.82
Text CBOW	SVC	**81.49** ±2.43	82.10 ±2.51	**81.58** ±2.43	81.76 ±2.59	**81.49** ±2.43
	LR	81.44 ±2.50	**82.22** ±2.57	81.54 ±2.50	**81.84** ±2.63	81.44 ±2.50
Text Skip-Gram	SVC	**82.64** ±3.03	83.31 ±3.12	**82.73** ±3.02	83.02 ±3.22	**82.64** ±3.03
	LR	82.54 ±2.64	**83.41** ±2.90	82.64 ±2.64	**83.04** ±2.91	82.54 ±2.64
(b) Emojis based features						
Emojis frequency	SVC	77.96 ±2.12	79.11 ±1.81	78.07 ±2.12	78.50 ±1.95	77.96 ±2.12
	LR	**78.53** ±2.33	**79.75** ±2.41	**78.64** ±2.33	**79.12** ±2.42	**78.53** ±2.33
Lexicon-based	SVC	**77.96** ±1.88	**79.70** ±1.77	**78.08** ±1.89	**78.91** ±1.82	**77.96** ±1.88
	LR	76.33 ±1.65	79.61 ±1.71	76.38 ±1.71	77.99 ±1.58	76.33 ±1.65
Emoji-CBOW	SVC	78.67 ±1.39	79.31 ±1.52	78.77 ±1.39	78.84 ±1.53	78.67 ±1.39
	LR	**79.10** ±2.17	**79.77** ±2.32	**79.20** ±2.18	**79.29** ±2.41	**79.10** ±2.17
Emojis Skip-Gram	SVC	78.62 ±1.57	79.24 ±1.47	78.72 ±1.56	78.77 ±1.54	78.62 ±1.57
	LR	**79.53** ±2.03	**80.17** ±2.16	**79.62** ±2.03	**79.71** ±2.22	**79.53** ±2.03

and area under the curve (AUC) are also used as an evaluation measures.

B. EXPERIMENTAL SETTINGS

In this study, we have a binary class sentiment polarity detection problem, where the class labels are negative and positive. 10-fold cross-validation mode is used for evaluation. For single modality classification, Support Vector Machines (SVC) and Linear Regression (LR) are adopted. In order to avoid biased models and over-fitting, the classifiers' parameters are chosen empirically by experimenting on the Syrian dataset. The linear kernel support vector machine is among the top ranked classifiers, hence it is used in the fusion models. Since the dataset is imbalanced, the classifier is trained with balancing class weight. Our system is developed in Python. Gensim package [88] was used for semantic analysis and feature extraction and Scikit-Learn package [89] was used for machine-learning approaches.

C. RESULTS AND DISCUSSIONS

a: TEXT-BASED FEATURES RESULTS

The top section of Table 5 shows the results obtained for SVC and LR using five textual feature extraction methods: *tf-idf*, LSA, structural, CBOW and Skip Gram. For the *tf-idf* features, LR classifier shows the best performance in terms of all evaluation measures, with the highest accuracy of 74.41%. For LSA features, the highest results are obtained also when using LR classifier. Except for structural features, there is a minor difference between LR and SVC. These results match the findings of an empirical study conducted by [90] between support vector machines and logistic regression. However, both classifiers have the worst results in the case of structural features.

In general for textual features, Word2Vec embedding models demonstrate the highest results, with Skip Gram achieving a higher accuracy of 82.64% ± 3.03 when using SVC classifier, i.e. 8.23% increase than the traditional *tf-idf*. Word2Vec CBOW comes next with an accuracy of 81.49% ± 2.43 using SVC classifier.

b: EMOJIS-BASED FEATURES RESULTS

The emojis-based results for the same classifiers are reported in the lower section of Table 5. Four feature extraction methods are evaluated: emojis frequencies, lexicon-based features, emojis CBOW and Skip-Gram models. Compared to the well-known textual features namely *tf-idf*, LSA and structural features, the basic form of emojis based features with similar machine learning approaches, i.e. emojis frequencies, performs significantly better. The highest performance of emojis based features is achieved when using LR classifier with Skip Gram features, reaching an accuracy of 79.53% ± 2.03. Although it is lower by 3.11% than the best approach for textual features, it has lower computational complexity than the extraction method of textual features.

c: SINGLE-LEVEL FUSION RESULTS

As found above, text Skip Gram achieves the highest results followed by Text CBOW and then emojis. These individual feature extraction approaches are considered here as baseline uni-modal predictive models of tweets sentiment. We ran several experiments to evaluate different early and late fusion methods to improve the results. We considered fusing emojis features in their basic form (emoji frequencies), which only requires counting, with textual features using Text CBOW and Text Skip Gram. Table 6 illustrates the attained results

TABLE 6. Fusion of Text CBOW, Text Skip Gram and Emojis Frequencies models at feature, score and decision levels using support vector machine with linear kernel (where Ψ refers to sum, prod or max score fusion function).

Features	Fusion	Rec	Prc	F_1	GM	Acc
CBOW-Emojis	C-E	83.83 ±2.64	84.00 ±2.78	83.85 ±2.65	83.57 ±2.90	83.83 ±2.64
	Ψ(C,E)	**84.31** ±2.09	**84.31** ±2.10	**84.25** ±2.13	**83.58** ±2.31	**84.31** ±2.09
SG-Emojis	S-E	83.74 ±2.78	83.88 ±2.79	83.75 ±2.77	83.45 ±2.87	83.74 ±2.78
	Ψ(S,E)	**85.41** ±2.59	**85.43** ±2.60	**85.37** ±2.62	**84.80** ±2.83	**85.41** ±2.59
CBOW-SG-Emojis	C-S-E	83.41 ±3.21	83.48 ±3.18	83.41 ±3.20	83.01 ±3.26	83.41 ±3.21
	sum(C,S,E)	84.60 ±2.56	84.64 ±2.59	84.58 ±2.58	84.14 ±2.74	84.60 ±2.56
	prod(C,S,E)	84.74 ±2.29	84.76 ±2.30	84.72 ±2.30	84.23 ±2.43	84.74 ±2.29
	max(C,S,E)	**85.08** ±2.34	**85.07** ±2.36	**85.03** ±2.37	**84.44** ±2.53	**85.08** ±2.34
	mode(C,S,E)	83.07 ±2.77	83.14 ±2.82	83.06 ±2.80	82.62 ±2.99	83.07 ±2.77

Note: C: Text CBOW, S: Text Skip Gram (SG), E: Emojis frequencies, C-E: Feature-level fusion of C and E, S-E: Feature-level fusion of S and E, C-S-E: Feature-level fusion of C, S and E

TABLE 7. Performance of hybrid fusion schemes of feature and score levels.

Feature	Score	Fig.	Rec	Prc	F_1	GM	Acc
CS	Ψ(CS, E)	8-a	85.08 ±2.24	85.07 ±2.24	85.03 ±2.26	84.43 ±2.38	85.08 ±2.24
SE	sum(SE, S, E)	8-b	84.98 ±2.79	85.03 ±2.79	84.97 ±2.81	84.54 ±2.94	84.98 ±2.79
	prod(SE, S, E)	8-b	85.03 ±2.72	85.09 ±2.73	85.01 ±2.75	84.58 ±2.89	85.03 ±2.72
	max(SE, S, E)	8-b	**85.22** ±2.84	**85.23** ±2.85	**85.18** ±2.86	**84.59** ±2.99	**85.22** ±2.84
CSE	sum(CSE, C, S, E)	—	84.60 ±2.73	84.66 ±2.75	84.59 ±2.75	84.16 ±2.90	84.60 ±2.73
	prod(CSE, C, S, E)	—	84.69 ±2.79	84.75 ±2.80	84.68 ±2.82	84.24 ±2.98	84.69 ±2.79
	max(CSE, C, S, E)	—	84.89 ±2.46	84.88 ±2.47	84.85 ±2.47	84.28 ±2.59	84.89 ±2.46

for feature-level, score-level and decision-level fusions of two and three feature representations. We can observe that the performance has remarkably improved, with a highest accuracy of 85.41% ± 2.59 when using two representations (Text Skip Gram + Emojis Frequencies) at the score level, using any fusion rule sum, prod or max (which is indicated in the table by Ψ; since the results are similar when fusing two scores). We also noticed that although combining the three feature representations (Text CBOW, Text Skip Gram, Emojis) has slightly lower accuracy than combining only Skip Gram and Emojis, this may be due to the confusion resulting from using two word embeddings CBOW and Skip Gram. For decision-level fusion, combining the three representations (Text CBOW + Text Skip Gram + Emojis) has better accuracy than the single modalities.

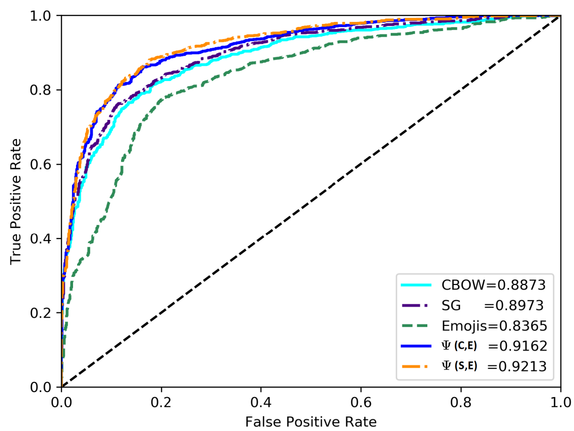


FIGURE 9. Performance comparison using ROC curves and area under the curves (AUCs) for five sentiment classification models.

d: MULTI-LEVEL HYBRID FUSION RESULTS

A number of multi-level hybrid fusion techniques is investigated using two and three feature representations. Table 7 shows the results for two-level hybrid fusion of feature and scores using two and three types of features. The first level is feature level and three possibilities are tested CS (Text CBOW + Text Skip Gram), SE (Text Skip Gram + Emojis Frequencies), CSE (Text CBOW + Text Skip Gram + Emojis Frequencies). Then, the resulting scores are combined with scores from other single feature vector models using sum, prod and max rules. For instance, the notation sum (SE, S, E) means the sum of the scores resulting from three models: one trained on augmented feature vector SE (Text Skip Gram + Emojis Frequencies), one trained on S (Text Skip Gram) and one trained on E (Emojis Frequencies). The highest performance is achieved when using max(SE, S, E), with an accuracy of 85.22% ± 2.84.

e: ROC CURVES

Figure 9 compares the performance in terms of ROC and AUC for five sentiment classification models using SVC. The feature extractors covers three baseline methods each using a single modality: Text CBOW (CBOW), Text Skip Gram (SG), and Emojis Frequencies (Emojis). It also shows the best two score-level fusion methods using two modalities: Text CBOW with Emojis Frequencies (Ψ(C, E)) and Text Skip Gram with Emojis Frequencies (Ψ(S, E)).

VI. CONCLUSION

This study presents a comprehensive review of work related to emojis-based opinion mining and sentiment analysis for social media. Previous work indicates that emojis will have a profound role in analyzing opinions and detecting sentiments in social platforms since they provide a quick visual way to express ideas and provide critique or appraisal reviews. However, the research in this area is still in its early stages and more innovations are expected in near future. Moreover, this paper demonstrates how effective combining emojis with text to detect sentiment polarity. Four methods have been investigated to extract features from emojis and build predictive models to detect sentiment polarity. The results are comparable to text-alone related features and even better than some of the traditional textual feature extraction methods. Additionally, in order to determine the best way to combine emojis features with textual features, various single-level and multi-level fusion techniques are evaluated. The experimental results show that fusing emojis with text has improved the performance at all fusion levels in terms of recall, precision, F_1 score, geometric mean and accuracy. The highest performance is achieved when using Skip Gram with Emojis at the score level. As future work, the proposed methodology can be explored for other languages and more feature extraction methods can be evaluated. Moreover, it is important to investigate the performance with other methods of topic modeling such as non-negative matrix factorization and latent Dirichlet allocation. Another point worth of exploring is the variations of impact and interpretation of emojis across multiple languages.

ACKNOWLEDGMENT

The authors would like to thank King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia, for support during this work.

REFERENCES

[1] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.

[2] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowl.-Based Syst.*, vol. 89, pp. 14–46, Nov. 2015.

[3] H. Tang, S. Tan, and X. Cheng, "A survey on sentiment detection of reviews," *Expert Syst. Appl.*, vol. 36, no. 7, pp. 10760–10773, 2009.

[4] A. Giachanou and F. Crestani, "Like it or not: A survey of Twitter sentiment analysis methods," *ACM Comput. Surv.*, vol. 49, no. 2, p. 28, 2016.

[5] Z. Hailong, G. Wenyan, and J. Bo, "Machine learning and lexicon based methods for sentiment classification: A survey," in *Proc. IEEE Web Inf. Syst. Appl. Conf. (WISA)*, Sep. 2014, pp. 262–265.

[6] O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications," *Expert Syst. Appl.*, vol. 77, pp. 236–246, Jul. 2017.

[7] C. Catal and M. Nangir, "A sentiment classification model based on multiple classifiers," *Appl. Soft Comput.*, vol. 50, pp. 135–141, Jan. 2017.

[8] V. García-Díaz, J. P. Espada, R. G. Crespo, B. C. Pelayo G-Bustelo, and J. M. Cueva Lovelle, "An approach to improve the accuracy of probabilistic classifiers for decision support systems in sentiment analysis," *Appl. Soft Comput.*, vol. 67, pp. 822–833, Jun. 2018.

[9] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: A review," *Artif. Intell. Rev.*, vol. 53, no. 6, pp. 4335–4385, Aug. 2020.

[10] K. Jindal and R. Aron, "A systematic study of sentiment analysis for social media data," *Mater. Today, Proc.*, Feb. 2021.

[11] O. F. Zaidan and C. Callison-Burch, "Arabic dialect identification," *Comput. Ling.*, vol. 40, no. 1, pp. 171–202, Mar. 2014.

[12] F. Barbieri, M. Ballesteros, and H. Saggion, "Are emojis predictable?" 2017, *arXiv:1702.07285*. [Online]. Available: https://arxiv.org/abs/1702.07285

[13] P. K. Novak, J. Smailović, B. Sluban, and I. Mozetič, "Sentiment of emojis," *PLoS ONE*, vol. 10, no. 12, Dec. 2015, Art. no. e0144296.

[14] W. Ai, X. Lu, X. Liu, N. Wang, G. Huang, and Q. Mei, "Untangling emoji popularity through semantic embeddings," in *Proc. 11th Int. AAAI Conf. Web Social Media*, 2017, pp. 2–11.

[15] J. Read, "Using emoticons to reduce dependency in machine learning techniques for sentiment classification," in *Proc. ACL Student Res. Workshop*, 2005, pp. 43–48.

[16] P. Han, S. Li, and Y. Jia, "A topic-independent hybrid approach for sentiment analysis of Chinese microblog," in *Proc. IEEE 17th Int. Conf. Inf. Reuse Integr. (IRI)*, Jul. 2016, pp. 463–468.

[17] A. Hogenboom, D. Bal, F. Frasincar, M. Bal, F. De Jong, and U. Kaymak, "Exploiting emoticons in polarity classification of text," *J. Web Eng.*, vol. 14, nos. 1–2, pp. 22–40, 2015.

[18] S. Al-Azani and E.-S. El-Alfy, "Emojis-based sentiment classification of Arabic microblogs using deep recurrent neural networks," in *Proc. Int. Conf. Comput. Sci. Eng. (ICCSE)*, Mar. 2018, pp. 1–6.

[19] S. Al-Azani and E.-S.-M. El-Alfy, "Combining emojis with Arabic textual features for sentiment classification," in *Proc. 9th Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2018, pp. 139–144.

[20] S. Al-Azani and E.-S.-M. El-Alfy, "Emoji-based sentiment analysis of Arabic microblogs using machine learning," in *Proc. 21st Saudi Comput. Soc. Nat. Comput. Conf. (NCC)*, Apr. 2018, pp. 1–5.

[21] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.

[22] M. Al-Ayyoub, A. A. Khamaiseh, Y. Jararweh, and M. N. Al-Kabi, "A comprehensive survey of Arabic sentiment analysis," *Inf. Process. Manag.*, vol. 56, no. 2, pp. 320–342, Mar. 2019.

[23] O. Oueslati, E. Cambria, M. B. HajHmida, and H. Ounelli, "A review of sentiment analysis research in Arabic language," *Future Gener. Comput. Syst.*, vol. 112, pp. 408–430, Nov. 2020.

[24] F. Barbieri, F. Ronzano, and H. Saggion, "What does this emoji mean? A vector space skip-gram model for Twitter emojis," in *Proc. Lang. Resour. Eval. Conf. (LREC)*, Portoroz, Slovenia, May 2016, pp. 1–6

[25] S. Wijeratne, L. Balasuriya, A. Sheth, and D. Doran, "A semantics-based measure of emoji similarity," 2017, *arXiv:1707.04653*. [Online]. Available: https://arxiv.org/abs/1707.04653

[26] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proc. LREc*, vol. 10, 2010, pp. 1320–1326.

[27] E. Refaee and V. Rieser, "Evaluating distant supervision for subjectivity and sentiment analysis on Arabic Twitter feeds," in *Proc. EMNLP Workshop Arabic Natural Lang. Process. (ANLP)*, 2014.

[28] W. A. Hussien, Y. M. Tashtoush, M. Al-Ayyoub, and M. N. Al-Kabi, "Are emoticons good enough to train emotion classifiers of Arabic tweets?" in *Proc. 7th Int. Conf. Comput. Sci. Inf. Technol. (CSIT)*, Jul. 2016, pp. 1–6.

[29] K.-L. Liu, W.-J. Li, and M. Guo, "Emoticon smoothed language models for Twitter sentiment analysis," in *Proc. 26th AAAI Conf. Artif. Intell.*, vol. 12, 2012, pp. 22–26.

[30] F. Barbieri, J. Camacho-Collados, F. Ronzano, L. E. Anke, M. Ballesteros, V. Basile, V. Patti, and H. Saggion, "Semeval 2018 task 2: Multilingual emoji prediction," in *Proc. 12th Int. Workshop Semantic Eval.*, 2018, pp. 24–33.

[31] F. Barbieri, M. Ballesteros, F. Ronzano, and H. Saggion, "Multimodal emoji prediction," in *Proc. NAACL, Short Papers, New Orleans, US. Assoc. Comput. Linguistics*, 2018.
[32] S. Al-Azani and E.-S.-M. El-Alfy, "Imbalanced sentiment polarity detection using emoji-based features and bagging ensemble," in *Proc. 1st Int. Conf. Comput. Appl. Inf. Secur. (ICCAIS)*, Apr. 2018, pp. 1–5.
[33] I. Rabbimov, I. Mporas, V. Simaki, and S. Kobilov, "Investigating the effect of emoji in opinion classification of uzbek movie review comments," in *Proc. Int. Conf. Speech Comput.*, in Lecture Notes in Computer Science, vol. 12335, A. Karpov and R. Potapova, Eds. Cham, Switzerland: Springer, 2020, pp. 435–445.
[34] C. Suman, S. Saha, P. Bhattacharyya, and R. S. Chaudhari, "Emoji helps! A multi-modal Siamese architecture for tweet user verification," *Cognit. Comput.*, vol. 13, no. 2, pp. 1–16, 2020.
[35] P. Golazizian, B. Sabeti, S. A. A. Asli, Z. Majdabadi, O. Momenzadeh, and R. Fahmi, "Irony detection in Persian language: A transfer learning approach using emoji prediction," in *Proc. 12th Lang. Resour. Eval. Conf.*, 2020, pp. 2839–2845.
[36] R. Wahyuni and I. Budi, "Combining linguistic, semantic and lexicon feature for emoji classification in Twitter dataset," *Procedia Comput. Sci.*, vol. 135, pp. 194–201, Jan. 2018.
[37] T. AMikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proc. Workshop Int. Conf. Learn. Represent.*, 2013, pp. 1–12.
[38] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.
[39] H. Miller, J. Thebault-Spieker, S. Chang, I. Johnson, L. Terveen, and B. Hecht, "'Blissfully happy' or 'ready to fight': Varying interpretations of emoji," in *Proc. 10th Int. Conf. Web Social Media (ICWSM)*, 2016, pp. 1–10.
[40] H. J. Miller, D. Kluver, J. Thebault-Spieker, L. G. Terveen, and B. J. Hecht, "Understanding emoji ambiguity in context: The role of text in emoji-related miscommunication," in *Proc. 11th Int. AAAI Conf. Web Social Media*, 2017, pp. 152–161.
[41] S. Annamalai, U. U. Malaysia, S. N. A. Salam, and U. U. Malaysia, "Undergraduates' interpretation on WhatsApp smiley emoji," *Jurnal Komunikasi, Malaysian J. Commun.*, vol. 33, no. 4, pp. 89–103, Dec. 2017.
[42] A. Wolf, "Emotional expression online: Gender differences in emoticon use," *CyberPsychol. Behav.*, vol. 3, no. 5, pp. 827–833, Oct. 2000.
[43] C. C. Tossell, P. Kortum, C. Shepard, L. H. Barg-Walkow, A. Rahmati, and L. Zhong, "A longitudinal study of emoticon use in text messaging from smartphones," *Comput. Hum. Behav.*, vol. 28, no. 2, pp. 659–663, Mar. 2012.
[44] Z. Chen, X. Lu, W. Ai, H. Li, Q. Mei, and X. Liu, "Through a gender lens: Learning usage patterns of emojis from large-scale Android users," 2017, *arXiv:1705.05546*. [Online]. Available: https://arxiv.org/abs/1705.05546
[45] S. Gupta, A. Singh, and J. Ranjan, "Sentiment analysis: Usage of text and emoji for expressing sentiments," in *Advances in Data and Information Sciences* (Lecture Notes in Networks and Systems), vol. 94, M. Kolhe, S. Tiwari, M. Trivedi, and K. Mishra, Eds. Singapore: Springer, 2020, pp. 477–486.
[46] N. Ljubešić and D. Fišer, "A global analysis of emoji usage," in *Proc. 10th Web Corpus Workshop*, 2016, pp. 82–89.
[47] F. Barbieri, G. Kruszewski, F. Ronzano, and H. Saggion, "How cosmopolitan are emojis?: Exploring emojis usage and meaning over different languages with distributional semantics," in *Proc. 24th ACM Int. Conf. Multimedia*, Oct. 2016, pp. 531–535.
[48] X. Lu, W. Ai, X. Liu, Q. Li, N. Wang, G. Huang, and Q. Mei, "Learning from the ubiquitous language: An empirical analysis of emoji usage of smartphone users," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput.*, Sep. 2016, pp. 770–780.
[49] G. W. Tigwell and D. R. Flatla, "Oh that's what you meant!: Reducing emoji misunderstanding," in *Proc. 18th Int. Conf. Hum.-Comput. Interact. Mobile Devices Services Adjunct*, Sep. 2016, pp. 859–866.
[50] F. Morstatter, K. Shu, S. Wang, and H. Liu, "Cross-platform emoji interpretation: Analysis, a solution, and applications," 2017, *arXiv:1709.04969*. [Online]. Available: https://arxiv.org/abs/1709.04969
[51] J. Berengueres and D. Castro, "Differences in emoji sentiment perception between readers and writers," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2017, pp. 4321–4328.
[52] H. Cui, Y. Lin, and T. Utsuro, "Sentiment analysis of tweets by CNN utilizing tweets with emoji as training data," in *Proc. Workshop Issues Sentiment Discovery Opinion Mining (WISDOM)*, 2018, pp. 1–8.
[53] L. K. Kaye, S. A. Malone, and H. J. Wall, "Emojis: Insights, affordances, and possibilities for psychological science," *Trends Cognit. Sci.*, vol. 21, no. 2, pp. 66–68, Feb. 2017.
[54] B. Eisner, T. Rocktäschel, I. Augenstein, M. Bošnjak, and S. Riedel, "Emoji2vec: Learning emoji representations from their description," 2016, *arXiv:1609.08359*. [Online]. Available: https://arxiv.org/abs/1609.08359
[55] Y. Tian, T. Galery, G. Dulcinati, E. Molimpakis, and C. Sun, "Facebook sentiment: Reactions and emojis," in *Proc. 5th Int. Workshop Natural Lang. Process. Social Media*, 2017, pp. 11–16.
[56] T. Tran, D. Nguyen, A. Nguyen, and E. Golen, "Sentiment analysis of marijuana content via Facebook emoji-based reactions," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.
[57] J. B. Walther and K. P. D'Addario, "The impacts of emoticons on message interpretation in computer-mediated communication," *Social Sci. Comput. Rev.*, vol. 19, no. 3, pp. 324–347, Aug. 2001.
[58] C. Tauch and E. Kanjo, "The roles of emojis in mobile phone notifications," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput., Adjunct*, Sep. 2016, pp. 1560–1565.
[59] S. Wijeratne, L. Balasuriya, A. Sheth, and D. Doran, "EmojiNet: Building a machine readable sense inventory for emoji," in *Social Informatics* (Lecture Notes in Computer Science), vol. 10046, E. Spiro and Y. Y. Ahn, Eds. Cham, Switzerland: Springer, 2016, pp. 527–541.
[60] S. Wijeratne, L. Balasuriya, A. Sheth, and D. Doran, "EmojiNet: An open service and API for emoji sense discovery," 2017, *arXiv:1707.04652*. [Online]. Available: https://arxiv.org/abs/1707.04652
[61] N. Na'aman, H. Provenza, and O. Montoya, "Mojisem: Varying linguistic purposes of emoji in (Twitter) context," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics, Student Res. Workshop*, 2017, pp. 136–141.
[62] U. Pavalanathan and J. Eisenstein, "Emoticons vs. emojis on Twitter: A causal inference approach," in *Proc. AAAI Spring Symp. Observational Stud. Through Social Media Other Hum.-Generated Content*, 2016.
[63] H. Cramer, P. de Juan, and J. Tetreault, "Sender-intended functions of emojis in us messaging," in *Proc. 18th ACM Int. Conf. Hum.-Comput. Interact. Mobile Devices Services*, 2016, pp. 504–509.
[64] Y. Lou, Y. Zhang, F. Li, T. Qian, and D. Ji, "Emoji-based sentiment analysis using attention networks," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 19, no. 5, pp. 1–13, Aug. 2020.
[65] T. Tomihira, A. Otsuka, A. Yamashita, and T. Satoh, "Multilingual emoji prediction using BERT for sentiment analysis," *Int. J. Web Inf. Syst.*, vol. 16, no. 3, pp. 265–280, Sep. 2020.
[66] H. W. Raj and S. Balachandran, "Future emoji entry prediction using neural networks," *J. Comput. Sci.*, vol. 16, no. 2, pp. 150–157, Feb. 2020.
[67] G. Guibon, M. Ochs, and P. Bellot, "From emojis to sentiment analysis," in *Proc. Workshop Affect Compagnon Artificiel Interact. (WACAI)*, 2016.
[68] S. R. El-Beltagy, M. El Kalamawy, and A. B. Soliman, "NileTMRG at SemEval-2017 task 4: Arabic sentiment analysis," in *Proc. 11th Int. Workshop Semantic Eval. (SemEval)*, 2017, pp. 790–795.
[69] R. Baly, G. Badaro, A. Hamdi, R. Moukalled, R. Aoun, G. El-Khoury, A. Al Sallab, H. Hajj, N. Habash, K. Shaban, and W. El-Hajj, "OMAM at SemEval-2017 task 4: Evaluation of English state-of-the-art sentiment analysis models for Arabic and a new topic-based model," in *Proc. 11th Int. Workshop Semantic Eval. (SemEval)*, 2017, pp. 603–610.
[70] E. Refaee and V. Rieser, "An Arabic Twitter corpus for subjectivity and sentiment analysis," in *Proc. 4th Int. Conf. Lang. Resour. Eval. (LREC)*, 2014, pp. 2268–2273.
[71] R. M. Duwairi, R. Marji, N. Sha'ban, and S. Rushaidat, "Sentiment analysis in Arabic tweets," in *Proc. 5th Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2014, pp. 1–6.
[72] M. El-Masri, N. Altrabsheh, H. Mansour, and A. Ramsay, "A web-based tool for Arabic sentiment analysis," *Procedia Comput. Sci.*, vol. 117, pp. 38–45, Jan. 2017.
[73] S. Alowaidi, M. Saleh, and O. Abulnaja, "Semantic sentiment analysis of Arabic texts," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 2, pp. 256–262, 2017.
[74] R. Meo and E. Sulis, "Processing affect in social media: A comparison of methods to distinguish emotions in tweets," *ACM Trans. Internet Technol.*, vol. 17, no. 1, pp. 1–25, Mar. 2017.

[75] M. Shiha and S. Ayvaz, "The effects of emoji in sentiment analysis," *Int. J. Comput. Electr. Eng.*, vol. 9, no. 1, pp. 360–369, 2017.
[76] M. Fernández-Gavilanes, J. Juncal-Martínez, S. García-Méndez, E. Costa-Montenegro, and F. J. González-Castaño, "Creating emoji lexica from unsupervised sentiment analysis of their descriptions," *Expert Syst. Appl.*, vol. 103, pp. 74–91, Aug. 2018.
[77] G. Donato and P. Paggio, "Investigating redundancy in emoji use: Study on a Twitter based corpus," in *Proc. 8th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal.*, 2017, pp. 118–126.
[78] M. Nabil, M. Aly, and A. Atiya, "ASTD: Arabic sentiment tweets dataset," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 2515–2519.
[79] N. A. Abdulla, N. A. Ahmed, M. A. Shehab, and M. Al-Ayyoub, "Arabic sentiment analysis: lexicon-based and corpus-based," in *Proc. IEEE Jordan Conf. Appl. Electr. Eng. Comput. Technol. (AEECT)*, Dec. 2013, pp. 1–6.
[80] A. Mourad and K. Darwish, "Subjectivity and sentiment analysis of modern standard Arabic and Arabic microblogs," in *Proc. 4th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal.*, 2013, pp. 55–64.
[81] S. Rosenthal, N. Farra, and P. Nakov, "SemEval-2017 task 4: Sentiment analysis in Twitter," in *Proc. 11th Int. Workshop Semantic Eval. (SemEval)*, 2017, pp. 502–518.
[82] S. M. Mohammad, M. Salameh, and S. Kiritchenko, "How translation alters sentiment," *J. Artif. Intell. Res.*, vol. 55, pp. 95–130, Jan. 2016.
[83] R. Albalawi, T. H. Yeap, and M. Benyoucef, "Using topic modeling methods for short-text data: A comparative analysis," *Frontiers Artif. Intell.*, vol. 3, p. 42, Jul. 2020.
[84] S. T. Dumais, "Latent semantic analysis," *Annu. Rev. Inf. Sci. Technol.*, vol. 38, no. 1, pp. 188–230, 2004.
[85] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," *Discourse Process.*, vol. 25, nos. 2–3, pp. 259–284, 1998.
[86] A. B. Soliman, K. Eissa, and S. R. El-Beltagy, "Aravec: A set of Arabic word embedding models for use in Arabic NLP," in *Proc. 3rd Int. Conf. Arabic Comput. Linguistics (ACLing)*, vol. 117, 2017, pp. 256–265.
[87] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.
[88] R. Řehůřek and P. Sojka, "Software framework for topic modelling with large corpora," in *Proc. LREC Workshop Challenges NLP Frameworks*. Valletta, Malta: ELRA, May 2010, pp. 45–50.
[89] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.
[90] A. B. Musa, "Comparative study on classification performance between support vector machine and logistic regression," *Int. J. Mach. Learn. Cybern.*, vol. 4, no. 1, pp. 13–24, Feb. 2013.

SADAM AL-AZANI received the B.Sc. degree (Hons.) in computer science from Thamar University, Dhamar, Yemen, in 2004, and the M.Sc. and Ph.D. degrees in computer science from the King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, in 2014 and 2019, respectively. He was a full-time Teaching Assistant with Thamar University, for five years. He is currently a Research Scientist III with the Interdisciplinary Research Center for Intelligent Secure Systems (IRC-ISS), KFUPM. His research interests include artificial intelligence, natural language processing, video analytics, machine/deep learning, social network analysis, and pattern recognition. He has served as a Technical Program Committee Member and a Reviewer for a number of international conferences and journals.

EL-SAYED M. EL-ALFY (Senior Member, IEEE) is currently a Professor with the Information and Computer Science Department, King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia. He is also affiliated with the University Interdisciplinary Research Center for Intelligent Secure Systems (IRC-ISS). He has over 25 years of experience in industry and academia, involving research, teaching, supervision, curriculum design, program assessment, and quality assurance in higher education. He is an approved ABET/CSAB Program Evaluator (PEV), a Reviewer, and a Consultant for NCAAA and several universities and research agents in various countries. He is an active researcher with interests in fields related to machine learning and nature-inspired computing and applications to data science and cybersecurity analytics, pattern recognition, multimedia forensics, and security systems. He has published numerously in peer-reviewed international journals and conferences, edited a number of books published by reputable international publishers, attended and contributed to the organization of many world-class international conferences, and supervised master and Ph.D. students. He was a member of ACM, the IEEE Computational Intelligence Society, the IEEE Computer Society, the IEEE Communication Society, and the IEEE Vehicular Technology Society. His work has been internationally recognized and received a number of awards. He has served as the Guest Editor for a number of special issues in international journals and been on the Editorial Board of a number of premium international journals, including IEEE/CAA JOURNAL OF AUTOMATICA SINICA, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, *International Journal of Trust Management in Computing and Communications*, and *Journal of Emerging Technologies in Web Intelligence* (JETWI).