



A survey of state-of-the-art approaches for emotion recognition in text

Nourah Alswaidan¹ · Mohamed El Bachir Menai¹

Received: 21 April 2018 / Revised: 14 February 2020 / Accepted: 15 February 2020
© Springer-Verlag London Ltd., part of Springer Nature 2020

Abstract

Emotion recognition in text is an important natural language processing (NLP) task whose solution can benefit several applications in different fields, including data mining, e-learning, information filtering systems, human–computer interaction, and psychology. Explicit emotion recognition in text is the most addressed problem in the literature. The solution to this problem is mainly based on identifying keywords. Implicit emotion recognition is the most challenging problem to solve because such emotion is typically hidden within the text, and thus, its solution requires an understanding of the context. There are four main approaches for implicit emotion recognition in text: rule-based approaches, classical learning-based approaches, deep learning approaches, and hybrid approaches. In this paper, we critically survey the state-of-the-art research for explicit and implicit emotion recognition in text. We present the different approaches found in the literature, detail their main features, discuss their advantages and limitations, and compare them within tables. This study shows that hybrid approaches and learning-based approaches that utilize traditional text representation with distributed word representation outperform the other approaches on benchmark corpora. This paper also identifies the sets of features that lead to the best-performing approaches; highlights the impacts of simple NLP tasks, such as part-of-speech tagging and parsing, on the performances of these approaches; and indicates some open problems.

Keywords Human emotion · Emotion recognition in text · Explicit emotion · Implicit emotion

1 Introduction

Humans experience different emotions on a daily basis. For example, we show our emotions when there is something to be happy about, something to be afraid of, or someone whom we are angry at [51,71]. Emotions have been studied for a very long time. Darwin [34] studied

✉ Nourah Alswaidan
nourah_swaidan@yahoo.com

Mohamed El Bachir Menai
menai@ksu.edu.sa

¹ Department of Computer Science, College of Computer and Information Sciences, King Saud University, P.O. Box 51178, Riyadh 11543, Saudi Arabia

emotion expression on the face and through body gestures. His study of emotion expression not only focused on humans but also included animals' expressions of emotions and how both have similar emotion expressions.

Plutchik [120] estimated that more than 90 definitions of emotion exist. Kleinginna and Kleinginna [84] classified these definitions into eleven categories: affective, cognitive, external emotional stimuli, physiological, emotional/expressive behavior, disruptive, adaptive, multi-aspect, restrictive, motivational, and skeptical. An emotion can simply be defined as a specific feeling that describes a person's state of mind, such as joy, love, anger, disgust, and fear [2]. In other words, emotions are intense feelings directed at something or someone [51,71]. There are two terms that are closely related to emotion and often mistaken for emotion: affect and mood. Affect is a broad range of feelings that people experience. It includes both emotions and moods [71]. Moods are feelings, but they are less intense than emotions and often lack a contextual stimulus [71,163].

To explain emotion and emotion expressions, researchers have proposed different emotion models. The circumplex model of emotion was developed by Russell [131]. This model represents emotions in a two-dimensional circular space, which contains valence as the horizontal axis and arousal as the vertical axis [121]. Plutchik [120] found that the primary emotions can be conceptualized in a manner analogous to a color wheel—similar emotions are placed close together, while opposite emotions are placed 180 degrees apart. Thus, he extended the circumplex model into a third dimension to represent the intensity of emotions; the resulting structure is shaped like a cone. Inspired by the work of Plutchik [120], Cambria et al. [24] developed the Hourglass of Emotions. This model represents affective states through labels and through four affective dimensions: pleasantness, attention, sensitivity, and aptitude.

Currently, people are increasingly relying on computers to perform their daily tasks, which have increased the need to improve human–computer interactions. The lack of commonsense knowledge makes emotion difficult for a computer to recognize and generate. Therefore, a substantial research on emotion recognition has been conducted. Emotion recognition is divided into three main categories: emotion recognition from facial expressions, emotion recognition from speech, and emotion recognition in text. Emotion recognition in text refers to the task of automatically assigning an emotion to a text selected from a set of predefined emotion labels. Emotion recognition in text is important because text is the main medium of human–computer interactions in the form of emails, text messages, chat rooms, forums, product reviews, Web blogs, and other social media platforms, including Twitter, YouTube, and Facebook. Applications of emotion recognition in text can be found in business, psychology, education, and many other fields where there is a need to understand and interpret emotions [9].

Emotion recognition in text, particularly implicit emotion recognition, is one of the difficult tasks in natural language processing (NLP), and it requires natural language understanding (NLU). There are different levels of text emotion recognition: document level, paragraph level, sentence level, and word level. The difficulty starts at the sentence level, where an emotion is expressed through the meanings of words and their relations; as the level increases, the complexity of the problem increases. Nevertheless, not all thoughts are expressed clearly, there are metaphors, sarcasm, and irony.

Different approaches have been used to recognize emotions in text. Keyword-based approaches for explicit emotion recognition have been investigated [90,116]. For example, the sentence “Sunny days always make me feel happy” explicitly expresses happiness and includes the emotion keyword “happy.” A keyword-based approach would be able to recognize the emotion successfully. However, the presence of an emotion keyword does not always match the expressed emotion. For example, the sentences “Do I look happy to you!” and “I am

not happy at all” include the emotion keyword “happy” but do not express that emotion. Additionally, a sentence can express emotion without the presence of an emotion keyword. Other approaches, namely rule-based approaches [86,155], classical learning-based approaches [4,6,105], deep learning approaches [18,48,93], and hybrid approaches [8,57,136], were specifically introduced for recognizing implicit emotions in text.

Several survey papers regarding emotion recognition in text have been published. In general, these survey papers did a shallow investigation; none of them reported the results or evaluated the reviewed papers. Moreover, Kao et al. [79] did not review any papers; they only discussed the limitations of the keyword-based, learning-based, and hybrid approaches and suggested a solution to overcome these limitations. Canales and Martínez-Barco [25] classified the work published in emotion recognition based on the used emotion model and approach. For emotion models, they included the categorical approach and the dimensional approach. However, there was no mention of the componential approach. The focus of the survey papers of Jain and Sandhu [75] and Deborah et al. [36] was only on learning-based approaches. In both papers, there were no evaluations of the reviewed papers and no reporting on the features and results. In this survey paper, we review the following explicit and implicit emotion recognition approaches: keyword-based approaches, rule-based approaches, classical learning-based approaches, deep learning approaches, and hybrid approaches. Nevertheless, the main focus is on implicit emotion recognition approaches. We include the strengths and limitations of the reviewed papers, compare them within tables, and discuss some open problems. We review more papers than the previous survey papers and cover emotion recognition in different languages, including Arabic, Chinese, English, bilingual (English and Hindi), Indonesian, and Spanish. We also present emotion modeling approaches and resources (corpora and affect lexicons) available for emotion recognition in text.

Fig. 1 The structure of the paper

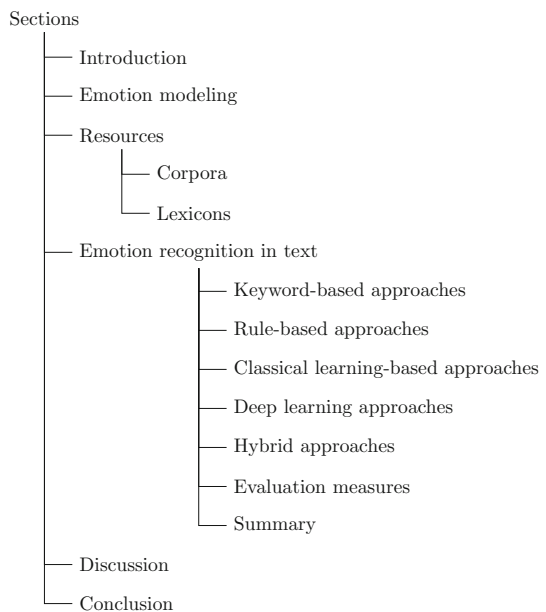


Table 1 Summary of the three dominant emotion modeling approaches [70]

Item	Categorical	Dimensional	Componential
Semantic primitives	Basic emotions	2 or 3 dimensions (valence and arousal or valence, arousal, and power)	Appraisal variables
Semantic primitives correspond to	Basic emotions	Underlying dimensions of felt mood	Interpretive features of the stimulus and the stimulus-agent relationship
Number/type of emotions possible	Small set of basic emotions	A large set of emotions	Very large set of emotions
Degree of elaboration of appraisal process	Low	Low	Very high
Degree of detail about affective dynamics	Low/qualitative	Medium/focus on arousal	Some/qualitative
Representative theorists	Tomkins [154], Ekman [44], Izard [74], Panksepp [112]	Wundt [165], Russell [121], Watson and Tellegen [161,162]	Frijda [51], Scherer [132], Smith [141,142], Ellsworth [45], Roseman [128], Ortony et al. [110]

The remainder of this paper is organized as follows. Section 2 presents the emotion modeling. Section 3 presents different resources for emotion recognition in text. Section 4 investigates prior work related to emotion recognition in English text and other languages. Section 5 discusses the advantages and limitations of the state-of-the-art approaches. Finally, the main conclusions are presented in Sect. 6. Figure 1 illustrates the paper structure.

2 Emotion modeling

Psychology research has distinguished three major approaches for emotion modeling [60,62]. Table 1 summarizes the three dominant emotion modeling approaches:

- *Categorical approach* This approach is based on the idea that there exist a small number of emotions that are basic and universally recognized [62]. The most commonly used model in emotion recognition research is that of Paul Ekman [44], which involves six basic emotions: happiness, sadness, anger, fear, surprise, and disgust.
- *Dimensional approach* This approach is based on the idea that emotional states are not independent but are related to each other in a systematic manner [62]. This approach covers emotion variability in three dimensions [20,78]:
 - Valence: This dimension refers to how positive or negative an emotion is [62].
 - Arousal: This dimension refers to how excited or apathetic an emotion is [62].
 - Power: This dimension refers to the degree of power [62].

- *Appraisal-based approach* This approach can be considered as an extension of the dimensional approach. It uses componential models of emotion based on appraisal theory [132], which states that a person can experience an emotion if it is extracted via an evaluation of events and that the result is based on a person's experience, goals, and opportunities for action. Here, emotions are viewed through changes in all significant components, including cognition, physiology, motivation, motor, reactions, feelings, and expressions [62].

In the categorical approach, the emotional states are limited to a fixed number of discrete categories, and it may be difficult to address a complex emotional state or mixed emotions [172]. However, these types of emotions can be well addressed in the dimensional approach, although the reduction in the emotion space to three dimensions is extreme and may result in information loss. Furthermore, not all basic emotions fit well in the dimensional space, some become indistinguishable, and some may lie outside the space. Regarding the advantage of componential models, they focus on the variability of different emotional states due to different types of appraisal patterns [62].

3 Resources

The following section presents the resources (corpora and lexicons) available for emotion recognition in text.

3.1 Corpora

- Alm¹ [4]: This corpus consists of approximately 185 children's stories, including those of Grimm, H.C. Andersen, and B. Potter. The annotation is performed at the sentence level with one of the following labels: neutral, anger-disgust, sadness, fear, happiness, positive surprise, and negative surprise. For the annotation, annotators work in pairs on the same stories. To avoid any bias, each annotator is trained separately and works independently. If a disagreement occurs between the annotators, the paper's first author chooses one of the selected emotion labels.
- Aman [6]: This corpus consists of blog posts that are retrieved using seed words that represent Ekman's six basic emotions. (For example, the words happy, pleased, and enjoy are selected as seed words for the happiness emotion category.) The annotation is performed at the sentence level with eight emotion categories; two new categories were recently added: mixed emotion and no emotion. Regarding the annotation, four annotators manually annotate the corpus.
- ISEAR² [133]: This corpus is the result of the international survey on emotion antecedents and reactions (ISEAR) project. A large group of psychologists from around the world worked on this project, and approximately 3000 students participated by reporting on situations in which they experienced the following emotions: joy, fear, anger, sadness, disgust, shame, and guilt.
- SemEval-2007³ [148]: This corpus consists of news headlines taken from newspapers, such as BBC News, CNN, and the New York Times, in addition to the Google News search

¹ <http://people.rc.rit.edu/~coagla/affectdata/index.html>.

² http://www.affective-sciences.org/index.php/download_file/view/395/296.

³ <http://web.eecs.umich.edu/~mihalcea/affectivetext/#resources>.

engine. The structure of headlines allows for performing sentence-level annotations. Each headline is annotated with one or more of the following emotions: anger, disgust, fear, joy, sadness, and surprise. Two datasets are developed: one for training (with 250 annotated headlines) and one for testing (with 1000 annotated headlines).

- SemEval-2018⁴ [104]: This corpus consists of tweets. Each tweet is either neutral or expresses one or more of eleven given emotions, which are anger, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust. Separate training, trial, and test datasets for English, Arabic, and Spanish tweets are provided.
- SemEval-2019⁵ [29]: This corpus consists of textual dialogues between two individuals. The first individual starts the conversation; then, it is the second individual's turn, and the turn comes back to the first individual. Each conversation is either labeled as joy, anger, sadness, or others. The classification of emotion labels is based on the third turn of the conversation. Separate training, trial, and test datasets are provided.
- Neviarouskaya: The annotation is performed at the sentence level with the ten labels defined by Izard [73]: anger, disgust, fear, guilt, interest, joy, sadness, shame, surprise, and neutral. The annotation process is performed by three annotators [28].
 - Dataset 1 [108] consists of 1000 annotated sentences collected from stories in 13 different categories grouped by topic.
 - Dataset 2 [107] consists of 700 annotated sentences collected from diary-like blog posts.

Table 2 presents the available datasets for emotion recognition in text, the recognized emotions in each dataset, and the number of instances in each emotion class. Noting that the SemEval-2007 dataset and SemEval-2018 dataset are multi-label multiclass datasets; thus, the total number of instances may appear less than the summation of the number of instances in each emotion. Each dataset is built from a different source. The sources vary in the style of writing (formal, informal), the quality of the text (with/without spelling errors, grammatical mistakes), and the use of special symbols, such as emojis, emoticons, and hashtags. Additionally, Alm, Aman, and ISEAR only provide a single dataset, while SemEval-2007 provides two (train and test) datasets, and both SemEval-2018 and SemEval-2019 provide three (train, trial, and test) datasets. All of the datasets have anger, sadness, and joy. ISEAR is the only dataset that does not include surprise and includes shame and guilt. SemEval-2018 is the only dataset that includes anticipation, optimism, pessimism, trust, and love. Of the group, Alm and SemEval-2007 have the least number of instances, and ISEAR is the only balanced dataset. The Alm dataset joins anger and disgust in one class; although they share similar characteristics, they are different emotions. Even if the reason behind this choice is the low number of instances in each class, the emotions should have been represented separately to measure the ability of a model to recognize each one accurately. SemEval-2019 is the only dataset of textual dialogues. The first step in any NLP task is preprocessing the data, and the success of any emotion recognition model is dependent on this step. In the SemEval-2018 competition, the highest ranked teams used Twitter-specific preprocessing to accommodate the special characteristics of tweets. Standard tokenization, segmentation, and part-of-speech (POS) tagging are not implemented to handle emojis, emoticons, hashtags, and informal styles of writing with many grammatical and spelling mistakes.

⁴ <https://competitions.codalab.org/competitions/17751>.

⁵ <https://www.linkedin.com/groups/12133338/>.

Table 2 Corpora for emotion recognition in text

Emotion	Alm	Aman	ISEAR	SemEval-2007		SemEval-2018		SemEval-2019	
				Trial	Test	Train	Trial	Train	Test
Anger	–	179	1096	20	21	2544	315	5506	150
Anger-Disgust	218	–	–	–	–	–	–	–	–
Anticipation	–	–	–	–	–	978	124	–	–
Disgust	–	172	1096	9	12	2602	319	–	–
Fear	166	115	1095	32	92	1242	121	–	–
Joy	445	536	1094	35	113	2477	400	4243	142
Love	–	–	–	–	–	700	132	–	–
Neutral	–	2800	–	146	650	207	14	–	–
Optimism	–	–	–	–	–	1984	207	–	–
Others	–	–	–	–	–	–	–	–	–
Pessimism	–	–	–	–	–	795	100	14,948	2338
Quilt	–	–	1093	–	–	–	–	–	–
Sadness	264	173	1096	41	104	2008	265	5463	125
Shame	–	–	1096	–	–	–	–	–	–
Surprise	114	115	–	8	42	361	35	–	–
Trust	–	–	–	–	–	357	43	–	–
Total	1207	4090	7666	250	1000	6838	886	30160	2755
									5509

3.2 Lexicons

- WordNet⁶ [50]: This is an online English lexical database. It groups verbs, nouns, adjectives, and adverbs into sets of synonyms called synsets.
- WordNet-Affect⁷ [149]: This lexicon is an affective extension of WordNet. A subset of WordNet synsets, words that either express direct or indirect emotions, are annotated using semantic labels.
- SentiWordNet⁸ [10,47]: This lexicon assigns one of three sentiment scores, namely positivity, negativity, or objectivity, to each synset of WordNet.
- SenticNet⁹ [19]: This is a lexicon of concepts with their respective emotions.
- Multi-perspective Question Answering (MPQA) Subjectivity Lexicon¹⁰ [164]: This lexicon consists of over 8000 subjectivity single-word clues; each clue is classified as positive or negative.
- Bing Liu Lexicon¹¹ [68]: This lexicon consists of a list of positive and negative opinion words or sentiment words.
- AFINN¹² [109]: This lexicon consists of manually rated words for valence with an integer between minus five (negative) and plus five (positive).
- NRC Word-Emotion Association Lexicon¹³ [102,103]: This lexicon is manually annotated using Amazon's Mechanical Turk. Eight emotions, which are anger, anticipation, disgust, fear, joy, sadness, surprise, and trust, and two sentiments, positive and negative, are included.
- NRC Affect Intensity Lexicon¹⁴ [100]: This lexicon provides real-valued intensity scores for the emotions anger, fear, sadness, and joy.
- NRC Valence, Arousal, and Dominance (VAD) Lexicon¹⁵ [99]: This lexicon includes a list of more than 20,000 words and their valence, arousal, and dominance scores. The scores range from 0 to 1.
- NRC Hashtag Emotion Lexicon¹⁶ [98,101]: This lexicon is automatically generated from tweets that include emotion word hashtags, such as #happy. It associates the words with the emotions anger, disgust, fear, sadness, anticipation, surprise, joy, and trust.
- NRC Hashtag Sentiment Lexicon¹⁷ [83]: This lexicon is generated automatically from tweets that include sentiment-word hashtags such as #amazing. It associates words with a positive or negative sentiment.
- Sentiment140 Lexicon¹⁸ [83]: This lexicon is generated automatically from tweets with emoticons.

⁶ <https://wordnet.princeton.edu>.

⁷ <http://wndomains.fbk.eu/wnaffect.html>.

⁸ <http://sentiwordnet.isti.cnr.it>.

⁹ <http://sentic.net>.

¹⁰ http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/.

¹¹ <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon>.

¹² <https://github.com/fnielsen/afinn>.

¹³ <http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>.

¹⁴ <http://www.saifmohammad.com/WebPages/AffectIntensity.htm>.

¹⁵ <http://saifmohammad.com/WebPages/nrc-vad.html>.

¹⁶ <http://saifmohammad.com/WebPages/lexicons.html#HashEmo>.

¹⁷ <http://saifmohammad.com/WebPages/lexicons.html#NRCTwitter>.

¹⁸ <http://saifmohammad.com/WebPages/lexicons.html#NRCTwitter>.

AffectiveTweets¹⁹ is a WEKA²⁰ (Waikato Environment for Knowledge Analysis) package for analyzing emotion and sentiment of tweets. What follows is the most used filters by the participants who achieved high ranking in the SemEval-2018 competition:

- TweetToLexiconFeatureVector: calculates features from a tweet using the lexicons:
 - MPQA: counts the number of positive and negative words from the Multi-perspective Question Answering (MPQA) Subjectivity Lexicon.
 - Bing Liu: counts the number of positive and negative words from the Bing Liu Lexicon.
 - AFINN: calculates positive and negative variables by aggregating the positive and negative word scores provided by the AFINN lexicon.
 - Sentiment140: calculates positive and negative variables by aggregating the positive and negative word scores provided by the Sentiment140 Lexicon.
 - NRC Hashtag Sentiment Lexicon: calculates positive and negative variables by aggregating the positive and negative word scores provided by the NRC Hashtag Sentiment Lexicon.
 - NRC Word-Emotion Association Lexicon: counts the number of words matching each emotion from the NRC Word-Emotion Association Lexicon.
 - NRC-10 Expanded²¹ [23]: adds the emotion associations of the words matching the Twitter-specific expansion of the NRC Word-Emotion Association Lexicon.
 - NRC Hashtag Emotion Association Lexicon: adds the emotion associations of the words matching the NRC Hashtag Emotion Association Lexicon.
 - SentiWordNet: calculates positive and negative scores using SentiWordNet.
 - Emoticons: calculates a positive and a negative score by aggregating the word associations provided by a list of emoticons.
 - Negations: counts the number of negating words in the tweet.
- TweetToInputLexiconFeatureVector: calculates features from a tweet using a given list of affective lexicons in arff format. The NRC affect intensity lexicon is used by default.
- TweetToSentiStrengthFeatureVector: calculates positive and negative sentiment strengths for a tweet using SentiStrength²² [152].

4 Emotion recognition in text

The following section investigates prior work related to both explicit and implicit emotion recognition in English and other languages. Our comprehensive review of the literature has lead us to distinguish between five classes of approaches for recognizing emotions in text, including keyword-based approaches, rule-based approaches, classical learning-based approaches, deep learning approaches, and hybrid approaches. The articles are classified based on the proposed approach. Using such classification will help evaluate these approaches based on their performance, strengths, and limitations and draw a comparison between them. Explicit emotion is mainly recognized with keyword-based approaches. The rule-based approaches, classical learning-based approaches, deep learning approaches, and hybrid

¹⁹ <https://affectivetweets.cms.waikato.ac.nz>.

²⁰ <https://www.cs.waikato.ac.nz/~ml/weka/>.

²¹ <https://www.cs.waikato.ac.nz/ml/sa/lex.html#emolextwitter>.

²² <http://senticstrength.wlv.ac.uk/>.

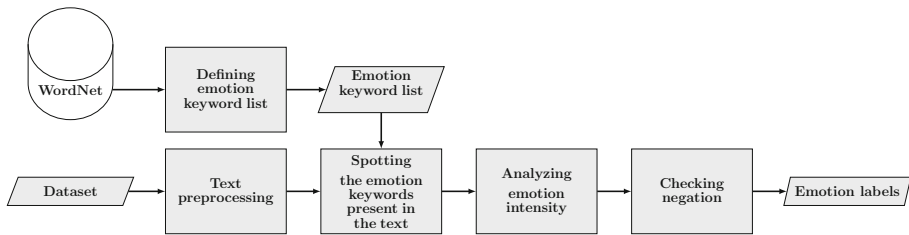


Fig. 2 Main steps of a keyword-spotting technique

approaches have mainly been introduced to recognize implicit emotions in text, even though they have been used for explicit emotion recognition.

4.1 Keyword-based approaches

A keyword-based approach relies on finding occurrences of keywords in a given text and assign an emotion label based on the detected keyword. The most used approach is the keyword-spotting technique; Fig. 2 outlines the main technique. First, a list of emotional words for each emotion label is defined using lexicons such as WordNet or WordNet-Affect. Then, text preprocessing, which includes tokenization, stop words removal, and lemmatization, is performed on the emotion dataset. The next step is to spot the emotion keywords present in the text using the predefined emotion keyword list. After that, the intensity of the emotion is analyzed. Then, negation checking is performed. Finally, the emotion label for each sentence is determined.

Tao [151] created a lexicon in which each word was classified as either a content word or an emotion functional word (EFW). The EFWs were then classified as an emotion keyword, modifier word, or metaphor word. The emotion keyword class consists of six labels of emotions and their corresponding weights. The modifier word class consists of words that emphasize the emotion by making it stronger or weaker. The metaphor word class consists of words that either show spontaneous expressions or show personal character. A coefficient is assigned to each word that was classified as a modifier word or a metaphor word. To obtain the relationship between the content word and the EFWs, POS tagging, a semantic tree, and HowNet [40], which is a Chinese knowledge database, were used. To recognize emotions, the first step is to apply the POS tagger, check for EFWs, and assign an emotion rating. The second step is to assign the weights for each emotion keyword and construct the link between the EFWs. The final step is to sum the weights of the emotion keywords across all sentences, run the scores through a fuzzy-logic process to determine the overall score, and assign each sentence a suitable emotion. Although their approach was able to classify the emotion conveyed in sentences correctly, many mislabeled emotions still occurred.

Ma et al. [90] proposed a model to recognize emotions in text messages in a chat system. First, they defined emotion keywords. Then, WordNet and WordNet-Affect were used to find the synonyms of the selected keyword. Each word was assigned a weight based on its sense. After building the affective lexicon, the overall emotion estimation was calculated by summing the weights of the matched keywords. Finally, sentence-level processing, which includes sentence splitting, POS tagging, and negation detection, was applied. The strategy used to address negation, which involved flipping the polarity of an emotion word, is not practical and will cause errors.

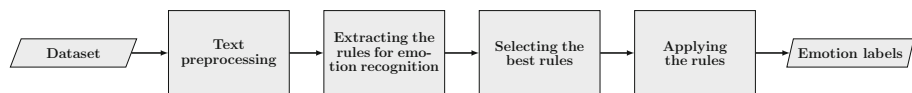


Fig. 3 Main steps of a rule-based approach

Perikos and Hatzilygeroudis [116] utilized NLP techniques, including POS tagging and parsing, to analyze the structure of a sentence. The emotion words were recognized using WordNet-Affect. The overall emotion of a sentence was selected based on the sentence dependency graph. The performance was tested on a corpus created by the authors. The corpus consists of 180 sentences, 120 of which convey emotion. Although the results were promising, the model must be tested on a known emotion corpus to truly measure its performance. Shivhare et al. [137] proposed a model that used an ontology with the keyword-spotting technique. The emotion ontology consists of three levels based on the emotion hierarchy presented by Parrott [114], and the Protégé²³ application was used to create it. The results showed that adding the ontology improves the accuracy but does not overcome all of the limitations of the keyword-spotting technique.

4.2 Rule-based approaches

A rule-based approach is based on the manipulation of knowledge to interpret information in a useful way; Fig. 3 outlines the main steps of a rule-based approach. First, text preprocessing is performed on the emotion dataset. The preprocessing steps may include tokenization, stop words removal, lemmatization, POS tagging, and dependency parsing. Then, the emotion rules are extracted using linguistic, statistics, and computational concepts. The best rules are then selected. Finally, the rules are applied to the emotion dataset to determine the emotion labels.

Lee et al. [86] proposed a rule-based model for recognizing emotion cause events in Chinese. Cause events refer to the explicitly expressed opinions or events that evoke a corresponding emotion. First, an annotated emotion causes corpus is constructed. Second, the distribution of cause event types, the position of cause events relative to emotional experiences, and keywords are calculated for each emotion class. Then, seven groups of linguistic cues are identified, and two sets of linguistic rules for recognizing emotion causes are generalized. Finally, based on the linguistic rules, a system that recognizes the causes of emotions is developed. The experiments showed that the system has promising performance in terms of cause occurrence recognition and cause event recognition.

Udochukwu and He [155] proposed an emotion recognition model based on the emotion model that was created by Ortony et al. [110] and modified by Steunebrink et al. [146]. The Ortony, Clore, and Collins (OCC) model consists of five variables: direction, tense, overall sentence polarity, event polarity, and action polarity. To fill the OCC model variables, the data must first be preprocessed. The following techniques were used for this purpose: sentence splitting and tokenization, POS tagging, word sense disambiguation (WSD), dependency parsing, sentence tense detection based on the POS tags, and polarity detection using the majority vote based on the lexicon matching results obtained from SentiWordNet [46], AFINN, and the subjectivity lexicon [164]. Because the goal was to recognize implicit emotions, sentences that express explicit emotions via emotion words were filtered. Thus, any sentence that contains an emotion word found in WordNet-Affect was deleted. The results showed that their approach is very sensitive to the text quality.

²³ <http://protege.stanford.edu>.

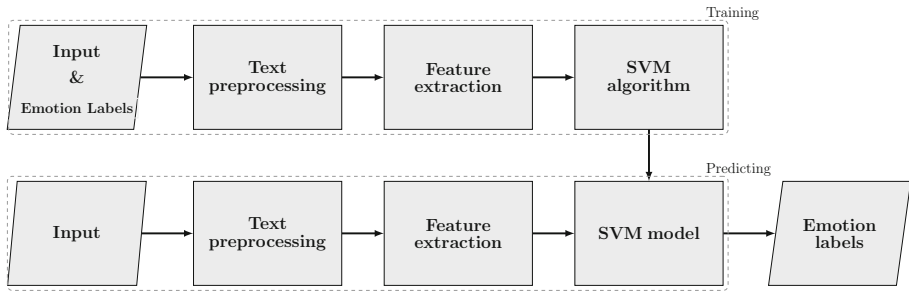


Fig. 4 Main steps of SVM

4.3 Classical learning-based approaches

A classical learning-based approach provides systems the ability to automatically learn and improve from experience. Machine learning algorithms are often categorized as supervised or unsupervised. The most used classification algorithm in the reviewed papers is SVM, which is a supervised machine learning algorithm. Figure 4 outlines the main steps of SVM for emotion recognition. First, text preprocessing is performed on the emotion dataset. The preprocessing steps may include tokenization, stop words removal, lemmatization, and POS tagging. The next step is to extract useful features. Then, the features with the most information gain are selected. Given the feature set and emotion labels, the SVM algorithm outputs an optimal hyperplane. Finally, the trained SVM model is used to classify emotions in unseen text.

One of the earliest works on emotion recognition in text was conducted by Alm et al. [4]. They proposed a supervised machine learning approach using a variation of the Winnow update rule implemented in the SNoW learning architecture [26]. Three experiments were performed. The first experiment tested whether a sentence was neutral or emotional. The second experiment tested whether the sentence was neutral or conveyed a positive or negative emotion. The results were affected by the size of the dataset. The worst result was obtained when classifying a positive emotion because only 9.87% of the sentences were annotated as this class. The third experiment tested the performance when different configurations of features were selected. The authors concluded that features interact with each other and that none of the features are independent of each other; hence, selecting the best feature set is challenging.

Aman and Szpakowicz [6] annotated a corpus for text emotion recognition. In their experiment, only the sentences for which all annotators agreed on regarding their emotion categories were selected. However, the focus was to recognize emotional sentences regardless of their emotion category. Thus, there were two classes: one representing all nonemotional sentences and one containing all the sentences labeled with one of Ekman's emotions. Different feature sets were tested, including features from the General Inquirer [147] only, features from WordNet-Affect only, and combined features from the previous lexicons with and without emoticons, exclamation marks and question marks. The authors used naïve Bayes and SVM for the classification. The best result was achieved using the SVM, and although the non-lexical features did not improve the results of the SVM, they did improve the accuracy of the naïve Bayes classifier. Moreover, Aman and Szpakowicz [7] conducted an experiment using different sets of features, including corpus-based unigram features, features derived from an emotion lexicon constructed based on Roget's Thesaurus [77] structure, and features

extracted from WordNet-Affect. The best result was obtained when all three features were combined.

Danisman and Alpkocak [33] proposed a vector space model (VSM), where each class of emotion is represented by a set of documents. To classify an input text, the similarity between each emotion class document and the input text is calculated by considering the cosine angle between them. The emotion class with the maximum similarity value is selected to be the label of the input text. The model was compared to ConceptNet [89], naïve Bayes, and SVM classifiers [64]. The experimental results showed that the VSM classifier performs better than all three classifiers.

Ghazi et al. [55] proposed a novel hierarchical approach to emotion recognition. First, the input text is classified into two categories: emotion sentences or nonemotion sentences. Second, the polarity of the emotion sentences is taken. Positive polarity represents the happiness emotion, while negative polarity represents the other five emotions: sadness, fear, anger, surprise, and disgust. The final step is to classify the emotions of the sentences with a negative polarity. Two experiments were performed: one that involved two-level classification and one that involved three-level classification. The main focus was to compare the hierarchical and flat classifications. The experimental results showed that this hierarchical approach outperforms the flat classification approach.

Ghazi et al. [56] aimed to take the context of a sentence into consideration by using different emotion lexicons and NLP techniques to extract meaningful feature sets. The performances of the features were tested using two classification algorithms: SVM and logistic regression. The logistic regression performed better than the SVM, and both performed better than the baseline, which used the SVM with BOW. Additionally, the features were grouped by similarity to test their contribution and significance. The results showed that lexical, POS, dependency, and negation features significantly improve the performance.

Xu et al. [167] proposed a hierarchical emotion classification for a Chinese microblog. In the first level, the input text is classified as neutral or emotional. The second level finds the polarity of the emotional sentences. The third level classifies the sentences with the negative polarities as either distress, surprised, fearful, angry, or disgusted and classifies the positive sentences as either fond or joyful. Then, each emotion class of the third level is divided into a number of emotion classes, resulting in 19 different classes of emotions in the fourth level. The support vector regression (SVR) algorithm is used for classification. Moreover, Zhang et al. [174] proposed a knowledge-based topic model (KTM) to identify implicit emotion features. Additionally, a hierarchical emotion structure was employed to classify emotions into 19 classes of four levels using an SVM. The authors achieved good results. However, the tree-structure classification was a time-consuming process.

As mentioned in Sect. 2, there are three major approaches for emotion modeling. Kim et al. [81] presented an evaluation of the categorical model and the dimensional model. For the categorical model, features were derived from WordNet-Affect, and the VSM was used for text representation. To reduce the VSM representation, three dimensionality reduction techniques were used: latent semantic analysis (LSA), probabilistic latent semantic analysis (PLSA), and nonnegative matrix factorization (NMF). Regarding the dimensional model, features were derived from the affective norm for English words (ANEW) [22]. The experimental results showed that the categorical NMF model and the dimensional model achieved the best results.

Chaffar and Inkpen [28] investigated the use of a heterogeneous emotion-annotated dataset, which included the SemEval-2007 dataset, Alm's dataset, Aman's dataset, and the Neviarouskaya dataset. The best results were obtained using the SVM classifier. Moreover, the results showed that using n-gram features for the SemEval-2007 dataset yields better

results than those obtained using BOW. However, the opposite is true for the Neviarouskaya dataset. Moreover, using features extracted from WordNet-Affect does not improve the accuracy.

Ho and Cao [66] developed an emotion recognition model based on two ideas: emotions depend on the mental state of humans [118], and emotions are caused by emotional events [69], which means that when a certain event occurs, the mental state of a human transitions from one state to another. The authors implemented this idea using a hidden Markov model (HMM), where each sentence consists of multiple sub-ideas and each idea is considered to be an event that causes a transition to a certain state. The states of the HMM were automatically generated based on the dataset, and its parameters were estimated during training. Compared to the other models, the results were not promising. The results could be improved by using a better dimensionality reduction method and including more linguistic information.

Bandhakavi et al. [15] created an emotion lexicon from a labeled emotion corpus. To show that a domain-specific emotion lexicon (DSEL) is more suitable than a general-purpose emotion lexicon (GPEL), the authors tested the quality of features extracted from their emotion lexicon against features extracted from a GPEL, such as WordNet-Affect, the NRC Word-Emotion Association Lexicon, and a lexicon learned using a point mutual information (PMI). The results showed that their features outperform the GPEL features and those of BOW. Moreover, the BOW features were better than the GPEL features, thus revealing that the use of the GPEL is not sufficient for a specific domain such as Twitter.

Anusha and Sandhya [9] developed a model that uses NLP techniques to improve the performance of learning-based approaches by including the syntactic and semantic features of text. Two classification algorithms were trained: naïve Bayes and SVM. The authors performed two experiments: one classified the emotions as either positive or negative and tested the sentence polarity, and the other tested the model performance on emotion recognition. Each experiment was repeated twice: once with the use of NLP techniques to preprocess the data and once without this step. The difference in the results when using the NLP techniques was significant. This outcome showed that applying methods that select the important part of a sentence is essential for improving the results.

Thomas et al. [153] investigated the use of multinomial naïve Bayes (MNB) with unigram, bigram, and trigram features from English text. The unigram features provided better results compared to the other two features. Later, Yuan, and Purver [173] utilized an SVM with high-order n-gram features for Chinese text. Character-based 4-gram features were the most effective. The results showed that the performance in terms of classifying the emotions varies among the emotions; the highest accuracy was achieved for happiness. Note that the size of the labeled data was not the same for each emotion and that happiness had the largest size.

Due to the importance of features and how they can affect the results, Gao et al. [52] proposed a feature extraction method that takes the syntactic and grammar structures of a sentence written in Chinese into account. First, they expanded the standard emotion lexicon, which was manually annotated by three annotators, using a Chi-square test and PMI with word2vec²⁴. Second, the quality of the selected feature was improved by using POS tagging and dependency parsing. The SVM was used for classification because it performs well and has been widely used. The results showed that using features with syntactic and grammar structures improves the accuracy.

Emotions are not limited to 6, 7, or 8 emotions. People express themselves using a wide range of emotions. Desmet and Hoste [38] proposed a binary SVM to recognize 15 emotions. They defined seven feature groups, and to determine the optimal feature combination, they

²⁴ <https://code.google.com/archive/p/word2vec>.

combined the seven feature groups into 17 feature sets and used bootstrap resampling. The input text was checked for spelling errors, and these errors were corrected. The experiments showed that applying spelling correction improves the results. The results also showed that the performance improves if the number of emotions is reduced. Through experimentation, the best result was obtained when retaining the following seven emotions: blame, guilt, hopelessness, information, instruction, love, and thankfulness. Yan and Turtle [169] proposed two learning-based approaches to recognize 28 emotions. The experiments demonstrated that the SVM and Bayesian networks consistently provide good performances.

Douji et al. [41] proposed an unsupervised machine learning algorithm based on the previous work of Agrawal and An [2]. YouTube comments were used as the data corpus because of the similarity between the writing styles of YouTube comments and instant messages. To recognize the emotion of a text entry, the similarity between the text and each target emotion was computed using the normalized version of PMI. Then, the average PMI values were computed, and the emotion category with the highest average value was assigned to a sentence. Since an unsupervised approach was used, the corpus required no labeling.

Muljono et al. [105] proposed a model to recognize emotions from Indonesian text. The following preprocessing steps were performed to extract the features: tokenization, case normalization, stop word removal, stemming, and term frequency-inverse document frequency (TF-IDF). Four classification methods, i.e., naïve Bayes, J48, k-nearest neighbor (KNN), and support vector machine-sequential minimal optimization (SVM-SMO), which were performed using WEKA, were evaluated. The best result was achieved using SVM-SMO. Jain et al. [76] proposed a multilingual English-Hindi emotion recognition framework, and two classification methods were tested: naïve Bayes and SVM. The best results were obtained using the SVM.

Mulki et al. [106] formulated emotion recognition as a binary classification problem. Different preprocessing steps were tested. The preprocessing pipeline used in their highest achieved result for Arabic was replacing an emoji with an emotion tag, stemming, and stop word removal. For English and Spanish, the preprocessing pipeline included replacing an emoji with an emotion tag, lemmatization, and stop word removal. TF-IDF was used to generate features. The classification was performed using one-vs-all SVM classifier with a linear kernel. Their model achieved 3rd rank for Arabic, 14th rank for English, and 3rd rank for Spanish among the teams in the SemEval-2018 competition.

Xu et al. [168] proposed a model to recognize multi-label emotion recognition in English tweets. The proposed model used different types of features including linguistic features, sentiment lexicon features, emotion lexicon features, and domain-specific features. Additionally, different classification algorithms were tested: logistic regression, SVR, bagging regressor (BR), AdaBoost regressor (ABR), gradient boosting regressor (GBR), and XGBoost regressor (XGB). The combination of all five types of features and logistic regression obtained the highest results. Their model achieved 13th rank among the teams in the SemEval-2018 competition. Deborah et al. [37] proposed a simple multilayer perceptron (MLP) for multi-label emotion recognition in English tweets. The MLP had an input layer, two hidden layers with 128 and 64 neurons, and an output layer. The model used a Nadam optimizer with 0.01 as the learning rate. Their model achieved 18th rank among the teams in the SemEval-2018 competition.

Plaza-del-Arco et al. [119] proposed a model for multi-label emotion recognition in English and Spanish tweets. First, text preprocessing was performed. The natural language

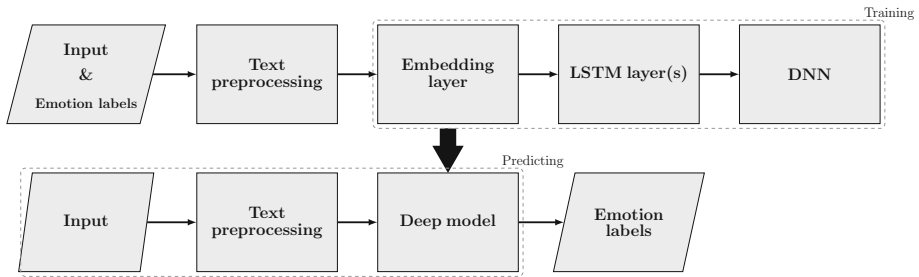


Fig. 5 Main steps of LSTM

toolkit (NLTK) TweetTokenizer²⁵ was used for tokenization, NLTK Snowball stemmer²⁶ was used for stemming, stop words were removed (only for English), and all letters were converted to lowercase. Then, different lexicons were tested including Spanish emotion lexicon [139], NRC Word-Emotion Association Lexicon [102], and WordNet-Affect. The information extracted from these lexicons with the TF-IDF representation of the tweets was used as the features. Finally, the authors used the random forest (RF) algorithm for classification. Their model achieved 25th and 5th ranks in the SemEval-2018 competition for English and Spanish, respectively. Although they ranked high in Spanish, a small number of teams participated in this language compared to English.

Singh et al. [140] proposed a two-stage text feature selection method to identify significant features for emotion recognition. First, they extracted meaningful words, namely nouns, verbs, adverbs, and adjectives, using a POS tagger. Then, a Chi-square method was employed to compute the statistical significance score for each word. The words with a low statistical score were removed. They used an SVM with radial basis kernel function to build the classification model. The results show that there is a significant improvement with the proposed approach compared with only using POS or statistical method.

4.4 Deep learning approaches

Deep learning is a branch of machine learning in which programs learn from experience and understand the world in terms of a hierarchy of concepts, where each concept is defined in terms of its relation to simpler concepts. This approach allows a program to learn complicated concepts by building them based on simpler ones [59]. The most used deep learning model here is long short-term memory (LSTM). LSTM is a special form of recurrent neural network (RNN) with the capability of handling long-term dependencies. LSTM overcomes the vanishing or exploding gradient problem common in RNNs. Figure 5 outlines the main steps of LSTM for emotion recognition in text. First, text preprocessing is performed on the emotion dataset. The preprocessing steps may include tokenization, stop words removal, and lemmatization. After that, the embedding layer is built and is fed into one or more LSTM layers. Then, the output is fed into a dense neural network (DNN) with units equal to the number of emotion labels and a sigmoid activation function to perform the classification.

Wang et al. [159] utilized a convolutional neural network (CNN) to solve multi-label emotion recognition. The experiments were conducted on the NLPCC 2014 Emotion Analysis

²⁵ <http://www.nltk.org/api/nltk.tokenize.html>.

²⁶ https://www.nltk.org/_modules/nltk/stem/snowball.html.

in Chinese Weibo Text (EACWT) task²⁷ [158] and the Chinese blog dataset Ren_CECps [122]. The experimental results showed that the CNN with the help of word embedding outperforms strong baselines and achieves excellent performance.

Baziotis et al. [18] proposed a deep learning model for multi-label emotion recognition English in tweets. Their model consisted of a two-layer bidirectional long short-term memory (Bi-LSTM) equipped with multilayer self-attention mechanism. They utilized the ekphrasis²⁸ [17] tool to process the text. The preprocessing steps included Twitter-specific tokenization, spell correction, word normalization, word segmentation, and word annotation. Due to the limited amount of training data, they utilized a transfer learning approach by pre-training the Bi-LSTMs on the SemEval-2017, Task 4A [129] dataset. They also collected a dataset of 550 million English tweets to be used for calculating word statistics necessary in the text preprocessing, training word2vec embeddings [96], and affective word embeddings. The experimental results showed that transfer learning did not outperform the random initialization model. Their model achieved 1st rank among the teams in the SemEval-2018 competition.

Meisheri and Dey [93] proposed a robust representation of a tweet. Two parallel architectures were designed to generate the representation using various pretrained embeddings. The first architecture generated the embedding matrix from emoji2vec [43], GloVe [115], and Character-level embeddings²⁹. The resulted matrix was fed into a Bi-LSTM [61]. The output of each time step was then fed into an attention layer [14]. The second architecture generated the embedding matrix using pretrained GloVe embeddings trained on a Twitter corpus. This matrix was fed into another Bi-LSTM, and max-pooling was applied to the output of the Bi-LSTM. The outputs of the two architectures were concatenated and then fed into two fully connected networks. Their model achieved second among the teams in the SemEval-2018 competition.

Du and Nie [42] proposed a deep learning model that uses pretrained word embeddings for the tweets representation. The embeddings were fed into a gated recurrent unit (GRU), and the classification was obtained using a dense neural network (DNN). Their model achieved 15th rank among the teams in the SemEval-2018 competition. Abdullah and Shaikh [1] formulated emotion recognition in tweets as a binary classification problem. Word embeddings were used for tweet representation. The embeddings were fed into four DNNs. The output of the fourth DNN was normalized to either one or zero based on a threshold, which was 0.5. Their model achieved 4th rank for Arabic and 17th rank for English among the teams in the SemEval-2018 competition. Li et al. [87] proposed a deep learning model that uses word embeddings for tweet representation. The embeddings were fed into an LSTM. For the classification, the model calculated a score for each emotion label and selected the ones with the top three scores. Their model achieved 23rd rank in the SemEval-2018 competition.

Ezen-Can and Can [48] proposed to formulate a multi-label emotion recognition problem as a binary classification problem. This approach allowed different model architectures and parameters for each emotion label. The authors utilized three GRU layers, two of which were bidirectional. Due to the size of the training dataset, they built an autoencoder and used unlabeled tweets to learn weights that could be used in the classifiers. They used pretrained embeddings for the representation of emojis [43], words, and hashtags [115]. Their results were better than the baseline but not as high as those of the other participants. Their model achieved 24th rank among the teams in the SemEval-2018 competition.

²⁷ http://tcci.ccf.org.cn/conference/2014/pages/page04_eva.html.

²⁸ github.com/cbaziotis/ekphrasis.

²⁹ <https://github.com/minimaxir/char-embeddings>.

Basile et al. [16] proposed a deep learning model for emotion recognition in textual conversation. The model consists of four submodels, which are three-input submodel (INP3), two-output submodels (OUT2), sentence-encoder submodel, and the bidirectional encoder representations from transformers (BERT) [39] submodel. The INP3 submodel takes each part of the conversation and fed it into two Bi-LSTM layers, followed by an attention layer [170]. The outputs are concatenated and fed into three DNNs. The OUT2 submodel has the same architecture as the INP3 submodel. However, the three parts conversation are concatenated and used as one input. Also, there is an additional DNN inserted after the attention layer that produces an additional output, a classification of the conversation as emotional or others. The purpose of this submodel is to reduce the effect of having an imbalanced dataset. In the sentence-encoder submodel, they built a feed-forward network with a fine-tuned universal sentence encoder (USE) [27] and only used the first and third part of the conversation. In the BERT submodel, they modeled the problem as a sentence-pair classification problem using only the first and third turn of the conversation. This submodel is combined with a lexical normalization system [156]. Different classification algorithms were tested including SVM, SVM with normalization (SVM-n), logistic regression, naive Bayes, JRip rule learner [31], random forest, and J48. The results show that the features learned by INP3 and OUT2 submodels lead to better performance than the features learned by USE and BERT submodels. However, an ensemble of the four submodels leads to the best performance result with SVM-n.

Xiao [166] proposed a deep learning model for emotion recognition in textual conversation. The ekphrasis tool was used for preprocessing the text. They fine-tuned the following models: the universal language model (ULM) [67], BERT model, OpenAI's Generative Pre-training (GPT) [123] model, DeepMoji [49] model, and a DeepMoji model trained with NTUA [18] embedding. The results show that the ULM model has the best performance among the other models, and the DeepMoji model trained with NTUA embedding came in second. However, ensembling these models obtained the highest result. They combined these models by taking the unweighted average of the posterior probabilities for these models, and the emotion class with the largest averaged probability was selected.

Ragheb et al. [124] proposed a deep learning model for emotion recognition in textual conversation. The three parts of the conversation were concatenated and inputted into the embedding layer. The output of the embedding layer is fed into three consecutive layers of Bi-LSTM trained by average stochastic gradient descent. Then, a self-attention mechanism followed by an average pooling was applied on the first and third parts of the conversation. The difference between the two pooled scores is taken as an input to a two DNN followed by softmax to obtain the emotion labels. The Wiktexit-103 dataset [94] was used for training the language model. The results show the low performance of recognizing the happy emotion label.

Ma et al. [91] proposed a deep learning model for emotion recognition in textual conversation. To overcome the out of vocabulary problem caused by using pretrained word embeddings, they replaced the emojis with a suitable emotion word. The embeddings are fed into a Bi-LSTM layer, while an attention mechanism increases the weights of the emotion words. The inner product is taken from the output of the Bi-LSTM and the attention weights and fed into another Bi-LSTM layer. Then, global max-pooling, global average pooling, and last tensor are used on the output of the Bi-LSTM layer. The pooling scores are fed into an LSTM layer and then a DNN with a softmax activation function. The results show the low performance of recognizing the happy emotion label.

Ge et al. [53] proposed a deep learning model for emotion recognition in textual conversation. Three pretrained embeddings, which are word2vec-twitter [58], GloVe, and ekphrasis

[17], were used. The embedding layer is fed into a Bi-LSTM followed by an attention layer and a CNN layer. The outputs of the Bi-LSTM and the CNN are concatenated, and global max-pooling is applied. The pooling scores are fed into a DNN with softmax activation function for classification. The results show that using pretrained embeddings improved the performance. Also, by combining the outputs of Bi-LSTM and CNN layers, the model was able to learn local features as well as long-term features.

Rathnayaka et al. [125] proposed a deep learning model for multi-label emotion detection in microblogs. They used ekphrasis tool for preprocessing. The pretrained word embedding GloVe was used. The embedding layer is fed into two Bi-GRU layers. Then, the embedding layer and the output of the first Bi-GRU layer are fed into the first attention layer. Also, the embedding layer, the output of the first Bi-GRU layer, and the output of the second Bi-GRU layer are fed into the second attention layer. Then, the two attention layers are concatenated and fed into a DNN with a sigmoid activation function to perform the classification. They achieved the state-of-the-art results with their model.

Seyeditabari et al. [135] formulated emotion recognition in text as a binary classification problem. Two word embedding models were used, which are ConceptNet Numberbatch [144] and fastText [97]. The embedding layer is fed into a Bi-GRU layer. Then, they used a concatenation of global max-pooling and average pooling layers. The pooling scores are fed into a DNN, and a sigmoid layer performs the classification. The results show that deep learning models can learn more informative features, which improve the performance significantly.

Shrivastava et al. [138] proposed a deep learning model for emotion recognition in multimedia text. The word2vec [95] model was used for constructing the words embeddings. The embedding layer is fed into the convolutional layers, followed by a max-pooling layer and then a DNN layer. The output of the DNN was then fed into an attention layer. The classification was performed by softmax. The results show that the precision of the emotion labels anger and fear is better than other emotion labels, while the recall and F1-score of the emotion label happiness are better than those of the other emotion labels.

4.5 Hybrid approaches

Seol et al. [134] proposed a hybrid of keyword-based and learning-based approaches. First, the system searches for emotion keywords in a sentence using the emotional keyword dictionary (EKD), which consists of words that express emotional meaning. If the system finds at least one emotional keyword, then the sentence is classified according to the EKD. However, if the input sentence does not contain any emotional keyword, then the knowledge-based artificial neural network (KBANN) classifier is used. The KBANN is a type of artificial neural network (ANN) that uses domain knowledge to initialize the network. Except for neutral emotions, each emotion is trained to be recognized by a separate KBANN. Moreover, Haggag [63] proposed a KBANN that is trained using an evolutionary algorithm. A structured knowledge base is created to store semantic and syntactic information for frame elements. Emotions are recognized via a matching process. Moreover, there are four methods to match a frame against a knowledge-based frame set, which are first matching, best matching, best opposite matching, and average matching. This approach allows for a trade-off between the performance and the strength of the matches found. The experimental results showed that the recognition accuracy of their proposed model is better than those of other existing emotion approaches, including keyword-spotting and the supervised machine learning models.

Gievska et al. [57] proposed a hybrid approach that uses both a lexicon and learning-based approaches. A lexicon of emotion words related to Ekman's six basic emotions was derived from the following: WordNet-Affect, AFINN, H4Lvd³⁰, and the NRC Word-Emotion Association Lexicon. A number of classification algorithms were tested, including naïve Bayes, SVM, and decision trees. The SVM provided the best results; thus, it was selected. The results showed that the lexicon approach limitation is improved with the help of the SVM classifier; it was able to recognize the implicit emotions in the sentences.

Shaheen et al. [136] proposed a framework that combines rule-based and learning-based approaches. Their emotion recognition system had two main phases. First, a set of the annotated reference called emotion recognition rules (ERRs), which are used to capture the emotional part of the sentence, is constructed. Second, the input sentence ERR is compared with the annotated ERRs using the KNN classifier. The KNN takes the input ERR and searches the annotated ERR set for a similar match. There are two similarity measures: semantic similarity and keyword similarity. The semantic similarity shows how close the two ERRs are in meaning, whereas the keyword similarity shows the number of matched words between the two ERRs. The input ERR will take the emotion label of the annotated ERR with the maximum semantic similarity. If there is a tie, then the keyword similarity is used. If the KNN classifier fails to find a match, which may occur when the training dataset is small, a PMI classifier is used. If it fails, a PMI with information retrieval (PMI-IR) is used. PMI-IR uses search engines (the authors used Google) to find a match. Two datasets were used: Aman's dataset and a dataset constructed from sentences are collected from Twitter. In one of their experiments, the authors used the second dataset for training and the first dataset for testing. The results demonstrated the strengths and robustness of their approach, where they trained using one dataset and tested using a completely different one.

Amelia and Maulidevi [8] proposed a hybrid method that combined the keyword-spotting technique and learning-based method to recognize the dominant emotion in short stories. The emotion words used in the keyword-spotting technique came from the NRC Emotion Lexicon. Although the NRC lexicon contains emotion words from 20 different languages, none of the languages corresponded to Bahasa Indonesia. Thus, the emotion words were translated from English into Indonesian using Google Translate³¹ and kamus.net.³² Then, they double-checked the translation to avoid auto-translation mistakes using Kamus Besar Bahasa Indonesia (KBBI)³³, which is a great dictionary of the Indonesian language of the language center. For the learning-based method, they used three algorithms: logistic regression, SVM, and naïve Bayes. Both methods were run separately, and each recognized one or more dominant emotion for each short story. Then, the hybrid method was used to select the most dominant emotion. However, if neither result of the two methods could be chosen, then the result obtained via the keyword-spotting technique was chosen. We believe that since no syntactic information and semantic information were used in the features, the keyword-spotting technique performed better than the learning-based methods.

Li et al. [88] proposed a hybrid neural network (HNN) composed of latent semantic machine (LSM), which uses the biterm topic model (BTM). Three experiments were performed. The first one evaluated the influence of the number of hidden neurons with one hidden layer. The results showed that the best numbers of hidden neurons on ISEAR and SemEval-2007 are 80 and 60, respectively. In the second experiment, the authors compared

³⁰ <http://www.wjh.harvard.edu/~inquirer/Home.html>.

³¹ <https://translate.google.com>.

³² <http://www.kamus.net>.

³³ <http://kbbi.web.id>.

the performance between their HNN and CNN, both with one hidden layer. Their HNN outperformed the CNN for both datasets. In the third experiment, they compared the performance of using two hidden layers. The HNN with two layers outperformed the CNNs with one and two hidden layers. However, the HNN with one hidden layer performed better than the HNN with two hidden layers.

Riahi and Safari [127] proposed a hybrid model, which consists of three submodels: a machine learning submodel, VSM submodel, and keyword-based submodel. Each submodel analyzes the input text from a different aspect and outputs an emotion label. If all submodels produce the same emotion label, then the label will be assigned to the input text; otherwise, the input text is left without an emotion label.

Herzig et al. [65] proposed an ensemble approach that combines the traditional representation of text as a BOW vector with a new representation that utilized pretrained word vectors, which are GloVe³⁴ and Word2Vec (GoogleNews). To obtain the document representation from the word embedding, they experimented with three methods: continuous bag-of-words (CBOW), TF-IDF weights, and classifier weights (CLASS). The experiments were performed on five datasets from different domains, and a (one vs. all) SVM classifier was used. The results showed that word vectors trained by GloVe achieved higher performance than Word2Vec-based vectors. The results showed that there is an advantage in combining traditional text representation, such as BOW, with embedded document representation.

Park et al. [113] proposed two models for multi-label emotion recognition in tweets. The first model was formulated as a linear regression with label distance as the regularization term. The second model was formulated as a logistic regression classifier chain. Classifier chain treats a multi-label problem as a sequence of binary classification problem while taking the prediction of the previous classifier as an additional input. For the features, the authors trained a CNN using another Twitter corpus distantly labeled with hashtags to obtain emotional word vectors. Additionally, they used two deep models to learn emoji vectors. In the first model, they used the pretrained deep learning network of Felbo et al. [49]. This network consists of Bi-LSTM with attention layer to extract features from the original competition datasets. For the second model, they collected 8.1 million tweets, which contained 34 different emojis relevant to the emotion labels. They then clustered these emojis into 11 clusters based on the distance on the correlation matrix of the hierarchical clustering from [49]. Next, they trained a one-layer Bi-LSTM classifier with 512 hidden units to predict the emoji cluster of each sample. They also included human-engineered features, such as the number of elongated words and the number of exclamation and question marks. The results showed that the regularized linear regression performed better than the classifier chain. However, the best result was achieved from the ensemble of both models. Their model achieved 3rd rank in the SemEval-2018 competition.

Gee and Wang [54] proposed a model for multi-label emotion recognition in English tweets. The proposed model consists of five submodels, which are a Bi-LSTM, an LSTM with attention mechanism, a Bi-LSTM, a lexicon vector, and five layers of DNNs. Transfer learning was performed to learn the weights of the LSTM networks. The input to the first two submodels was word embeddings, while the input to the third and fourth submodels was a lexicon vector extracted by TweetToLexiconFeatureVector AffectiveTweets WEKA filter. The outputs of the four submodels were concatenated and fed into the fifth submodel. The

³⁴ <http://nlp.stanford.edu/data/glove.840B.300d.zip>.

model was trained incrementally for emotions within the same cluster formed by hierarchical clustering. Their model achieved 4th rank in the SemEval-2018 competition among the teams.

Kim et al. [82] proposed a model for multi-label emotion recognition in tweets. The proposed model used pretrained word embeddings. The embeddings were fed into three self-attention layers. The output of the self-attention layers was fed into a CNN, and then, max-pooling was performed. The output of the max-pooling was fed into a DNN for classification. They experimented with the impact of using emojis, self-attention layers, and lexicon features. The results showed that utilizing emojis, attention mechanism, and lexicon features improve the results. Their model achieved 5th rank for English and 1st rank for Spanish among the teams in the SemEval-2018 competition.

Rozental and Fleischer [130] proposed a model for multi-label emotion recognition in English tweets. Two preprocessing pipelines, which are simple and complex, were implemented. Both versions used the following steps: word tokenization using the CoreNLP,³⁵ [92] POS tagging using the Tweet NLP³⁶ tagger [111], replacing emojis with representative keywords, replacing URLs with a special keyword, removing duplications, and breaking hashtags into individual words. The complex preprocessing version included these additional steps: word lemmatization using CoreNLP, name entity recognition using CoreNLP, and replacing the entities with representative keywords, synonym replacement, and word replacement using Wikipedia dictionary. Two hundred million tweets were randomly sampled using the Twitter Firehose service. The authors cleaned the gathered tweets using the preprocessing pipelines (simple and complex). Then, they trained the embeddings using the Gensim³⁷ package [126]. They created four embeddings for the words and two embeddings for the POS tags. In addition to the deep features, they extracted lexicon features and semantic and syntactic features. The embeddings were fed into a bidirectional gated recurrent unit (Bi-GRU) with a CNN attention mechanism. The output was then fed into two fully connected neural nets. Their model achieved 6th rank in the SemEval-2018 competition.

De Bruyne et al. [35] formulated emotion recognition in tweets as a binary classification problem. Different syntactic, semantic, and stylistic features were used to represent the tweets. Additionally, different classifiers were tested, including SVM, linear SVM with stochastic gradient descent learning (SGD), logistic regression, and RF. The authors took the best performing classifier for each emotion label and combined them in a classifier chine, where the prediction from the previous model was passed to the next classifier as additional features. Their model achieved 11th rank in the SemEval-2018 competition.

Kravchenko and Pivovarova [85] proposed a model for multi-label emotion recognition in English tweets. The proposed model hybridized two types of features, which are lexicon features and word embeddings. The gradient boosting classifier was used for the classification. They concluded that the model performed better with word embeddings than lexicon features, and the best result was achieved from the combination of both. Their model achieved 15th rank in the SemEval-2018 competition.

³⁵ <https://stanfordnlp.github.io/CoreNLP>.

³⁶ <http://www.cs.cmu.edu/~ark/TweetNLP>.

³⁷ <https://radimrehurek.com/gensim/index.html>.

Badaro et al. [12] proposed a model for multi-label emotion recognition in Arabic tweets. Several features were tested including n -grams, affect lexicons, sentiment lexicon, and word embeddings from AraVec³⁸ [143] and FastText³⁹ [21]. AraVec embeddings outperformed the other features. The authors also tested several learning models including support vector classifier (SVC) with both penalties L1 and L2, ridge classification (RC), RF, and an ensemble of the three. Linear SVC with L1 outperformed the other learning models. Their model achieved 1st rank in the SemEval-2018 competition.

Agrawal and Suri et al. [3] proposed to combine lexical and deep learning features for emotion recognition in textual conversation. The aim was to build a model robust to emoticons, slang, abbreviations, spelling mistakes, and style of writing. They trained LightGBM [80] and logistic regression models. LightGBM performed better than logistic regression. They performed a hold-one-out experiment on the features. The results show that the maximum gain was from character n -grams.

4.6 Evaluation measures

The following section presents different evaluation measures used in related work. These include the multi-label accuracy (the Jaccard accuracy) (Eq. 1), accuracy (Eq. 2), F^{micro} (Eq. 5), and F^{macro} (Eq. 9).

$$\text{Jaccard accuracy} = \frac{1}{|S|} \sum_{s \in S} \frac{|G_s \cap P_s|}{|G_s \cup P_s|} \quad (1)$$

where G_s is the set of gold labels for sentence s , P_s is the set of predicted labels for sentence s , and S is the set of sentences.

$$\text{Accuracy} = \frac{\sum_{e \in E} \text{TP} + \sum_{e \in E} \text{TN}}{\sum_{e \in E} \text{TP} + \sum_{e \in E} \text{TN} + \sum_{e \in E} \text{FP} + \sum_{e \in E} \text{FN}} \quad (2)$$

where E is the set of emotion labels, TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

For the micro-averaged results, the TP, FP, and FN for each emotion label are summed; then, the average is taken. P^{micro} and R^{micro} are calculated as follows:

$$P^{\text{micro}} = \frac{\sum_{e \in E} \text{TP}}{\sum_{e \in E} \text{TP} + \sum_{e \in E} \text{FP}} \quad (3)$$

$$R^{\text{micro}} = \frac{\sum_{e \in E} \text{TP}}{\sum_{e \in E} \text{TP} + \sum_{e \in E} \text{FN}} \quad (4)$$

F^{micro} is the harmonic mean of P^{micro} and R^{micro} .

$$F^{\text{micro}} = 2 \cdot \frac{P^{\text{micro}} \times R^{\text{micro}}}{(\text{micro } P + R^{\text{micro}})} \quad (5)$$

F^{macro} computes the harmonic mean of precision and recalls independently for each emotion label e and then takes the average, hence treating all emotion labels equally.

³⁸ <https://github.com/bakrianoo/aravec>.

³⁹ <https://fasttext.cc/docs/en/crawl-vectors.html>.

Table 3 Summary of related work in emotion recognition in text

References	Language	Method	Dataset	Performance	
				Measure	%
Ma et al. [90]	English	Keyword-based	Text messages in chat system	–	–
Perikos and Hatzilygeroudis [116]	English	Keyword-based	Self-created	–	–
Shivhare et al. [137]	English	Keyword-based	Different blogs	Accuracy	79.57
Tao [151]	Chinese	Keyword-based	Speech corpus	–	–
Lee et al. [86]	Chinese	Rule-based	Sinica [30]	<i>F</i> -score	47.95
Udochukwu and He [155]	English	Rule-based	ISEAR	<i>F</i> -score	51.3
			SemEval-2007		57.3
			Alm		65.5
Alm et al. [4]	English	Learning-based	Alm	<i>F</i> -score	47
Aman and Szpakowicz [6]	English	Learning-based	Aman	Accuracy	73.89
Aman and Szpakowicz [7]	English	Learning-based	Aman	<i>F</i> -score	58.6
Anusha and Sandhya [9]	English	Learning-based	ISEAR	<i>F</i> -score	63.1
Bandhakavi et al. [15]	English	Learning-based	ISEAR	<i>F</i> -score	39.48
			SemEval-2007		40.99
			Tweets		64.24
			Blogs		59.66
Chaffar and Inkpen [28]	English	Learning-based	SemEval-2007	Accuracy	40.30
			Neviarouskaya et al. [107]		57.81
			Neviarouskaya et al. [108]		35.45
Danisman and Alpkocak [33]	English	Learning-based	SemEval-2007	<i>F</i> -score	32.2
Deborah et al. [37]	English	Learning-based	SemEval-2018	Jaccard accuracy	46.8
Desmet and Hoste [38]	English	Learning-based	Suicide database [117]	<i>F</i> -score	53.87
Douiji et al. [41]	English	Learning-based	YouTube comments	Accuracy	68.82
Gao et al. [52]	Chinese	Learning-based	Blog post from Weibo	<i>F</i> -score	71.67

Table 3 continued

References	Language	Method	Dataset	Performance	
				Measure	%
Ghazi et al. [55]	English	Learning-based	Aman	<i>F</i> -score	50.86
			Alm		48.8
Ghazi et al. [56]	English	Learning-based	Aman	<i>F</i> -score	65
Ho and Cao [66]	English	Learning-based	ISEAR	<i>F</i> -score	35.3
Jain et al. [76]	Bilingual (English and Hindi)	Learning-based	Election dataset	<i>F</i> -score	74
			Healthcare		62.2
			Sports		61.8
Kim et al. [81]	English	Learning-based	SemEval-2007	<i>F</i> -score	50.5
			ISEAR		37.2
			Alm		73.3
Muljono et al. [105]	Indonesian	Learning-based	Various web-sites	<i>F</i> -score	80.68
Mulki et al. [106]	Arabic	Learning-based	SemEval-2018	Jaccard accuracy	45.6
	English				48.1
	Spanish				43.8
Plaza-del-Arco et al. [119]	English	Learning-based	SemEval-2018	Jaccard accuracy	39.4
	Spanish				31.8
Singh et al. [140]	English	Learning-based	ISEAR	Accuracy	72.43
Thomas et al. [153]	English	Learning-based	ISEAR	Accuracy	76.96
Xu et al. [167]	Chinese	Learning-based	Sina Weibo posts	<i>F</i> -score	72.7
Xu et al. [168]	English	Learning-based	SemEval-2018	Jaccard accuracy	50.1
Yan and Turtle [169]	English	Learning-based	Tweets	<i>F</i> -score	57
Yuan and Purver [173]	Chinese	Learning-based	Sina Weibo posts	<i>F</i> -score	82.21
Zhang et al. [174]	Chinese	Learning-based	Sina Weibo posts	<i>F</i> -score	73.9
Abdullah and Shaikh [1]	Arabic	Deep learning	SemEval-2018	Jaccard accuracy	44.6
	English				47.1
Basile et al. [16]	English	Deep learning	SemEval-2019	<i>F</i> -score	77

Table 3 continued

References	Language	Method	Dataset	Performance	
				Measure	%
Baziotis et al. [18]	English	Deep learning	SemEval-2018	Jaccard accuracy	58.8
Du and Nie [42]	English	Deep learning	SemEval-2018	Jaccard accuracy	47.3
Ezen-Can and Can [48]	English	Deep learning	SemEval-2018	Jaccard accuracy	39.8
Ge et al. [53]	English	Deep learning	SemEval-2019	<i>F</i> -score	75.42
Li et al. [87]	English	Deep learning	SemEval-2018	Jaccard accuracy	40.7
Ma et al. [91]	English	Deep learning	SemEval-2019	<i>F</i> -score	75.57
Meisheri and Dey [93]	English	Deep learning	SemEval-2018	Jaccard accuracy	58.2
Ragheb et al. [124]	English	Deep learning	SemEval-2019	<i>F</i> -score	75.82
Rathnayaka et al. [125]	English	Deep learning	SemEval-2018	<i>F</i> -score	58.9
Seyeditabari et al. [135]	English	Deep learning	CrowdFlower	<i>F</i> -score	63.2
Shrivastava et al. [138]	English	Deep learning	TV transcript	<i>F</i> -score	72.48
Wang et al. [159]	Chinese	Deep learning	EACWT	<i>F</i> -score	60.05
			Ren_CECps		38.68
Xiao [166]	English	Deep learning	SemEval-2019	<i>F</i> -score	76.86
Agrawal and Suri et al. [3]	English	Hybrid	SemEval-2019	<i>F</i> -score	77.65
Amelia and Maulidevi [8]	Bahasa Indonesia	Hybrid	Short stories	Accuracy	65.71
Badaro et al. [12]	Arabic	Hybrid	SemEval-2018	Jaccard accuracy	48.9
De Bruyne et al. [35]	English	Hybrid	SemEval-2018	Jaccard accuracy	52
Gee and Wang [54]	English	Hybrid	SemEval-2018	Jaccard accuracy	57.4
Gievska et al. [57]	English	Hybrid	Mixture of ISEAR, SemEval-2007, Alm, and two other blog sources	<i>F</i> -score	84
Haggag [63]	English	Hybrid	Oxford Practice Grammar book and different electronic linguistic magazines	Accuracy	87

Table 3 continued

References	Language	Method	Dataset	Performance	
				Measure	%
Herzig et al. [65]	English	Hybrid	ISEAR	<i>F</i> -score	64.1
			SemEval-2007		52.5
			Alm		73.3
			Blog posts		66.3
			Twitter dialogs		52.2
Kim et al. [82]	English	Hybrid	SemEval-2018	Jaccard accuracy	57.4
	Spanish				46.9
Kravchenko and Pivovarova [85]	English	Hybrid	SemEval-2018	Jaccard accuracy	47.7
Li et al. [88]	English	Hybrid	ISEAR	<i>F</i> -score	51.75
			SemEval-2007		38.52
Park et al. [113]	English	Hybrid	SemEval-2018	Jaccard accuracy	57.6
Riahi and Safari [127]	English	Hybrid	ISEAR	<i>F</i> -score	68
Rozental and Fleischer [130]	English	Hybrid	SemEval-2018	Jaccard accuracy	56.6
Seol et al. [134]	English	Hybrid	Scripts of drama, novel, and public web diary	Accuracy	57.75
Shaheen et al. [136]	English	Hybrid	Aman	<i>F</i> -score	84

$$\text{precision}_e = \frac{\text{TP}_e}{\text{TP}_e + \text{FP}_e} \quad (6)$$

$$\text{recall}_e = \frac{\text{TP}_e}{\text{TP}_e + \text{FN}_e} \quad (7)$$

$$F_e = 2 \cdot \frac{\text{precision}_e}{\text{recall}_e} \quad (8)$$

$$F^{\text{macro}} = \frac{1}{|E|} \sum_{e \in E} F_e \quad (9)$$

4.7 Summary

The following section presents a summary of the reviewed state-of-the-art approaches, comparing them within tables. Table 3 reports the language of the text, the approach, the corpus used for testing, and the obtained result of the reviewed state-of-the-art approaches. Table 4 presents the features used by the classical learning-based approaches, the deep learning approaches, and the hybrid approaches. Table 5 presents the strengths and limitations of the reviewed state-of-the-art approaches.

Table 4 Features used for emotion recognition in text

Paper	Features
Alm et al. [4]	First sentence in story Direct speech (i.e., whole quote) in sentence Thematic story type Special punctuation (! and ?) All words written in uppercase Sentence length (in words) Ranges of story progress Percentage of adjectives, nouns, verbs, and adverbs Number of verbs in sentence Positive and negative word counts WordNet emotion words Interjections and affective words BOW content
Aman and Szpakowicz [6]	General inquirer features Emotion words Positive words Negative words Interjection words Pleasure words Pain words WordNet-Affect features Happiness words Sadness words Anger words Disgust words Surprise words Fear words Emoticons Special punctuation (! and ?)
Aman and Szpakowicz [7]	Unigrams Roget's thesaurus features WordNet-Affect features
Anusha and Sandhya [9]	TF-IDF
Bandhakavi et al. [15]	Emotion lexicon features Total emotion count Total emotion intensity Max emotion intensity Graded emotion count Graded emotion intensity

Table 4 continued

Paper	Features
	<i>n</i> -grams
	POS
	Contextual features
	Capitalized words
	Elongated words
	Punctuation
	Emoticons
	Negation
	Sentiment features
Chaffar and Inkpen [28]	BOW
	<i>n</i> -grams
	WordNet-Affect features
Danisman and Alpkocak [33]	TF-IDF
Deborah et al. [37]	BOW
Desmet and Hoste [38]	Lemmas
	Trigrams
	Subjectivity clues
	Lemmas + POS
	Pruned lemmas + POS
	WordNet features
	SentiWordNet features
Douiji et al. [41]	Normalized version of PMI (NPMI)
Gao et al. [52]	POS
	Dependency
	Emotional words
Ghazi et al. [55]	A combination of the emotional lists of words from Roget's Thesaurus [7] and WordNet-Affect
Ghazi et al. [56]	Emotion word features
	The emotion of a word according to WordNet-Affect
	The polarity of a word according to the prior-polarity lexicon
	The presence of a word in a small list of modifiers from the intensifier lexicon
	POS
	Sentence length (in words)
	Dependency tree
Ho and Cao [66]	HMM features
Jain et al. [76]	WordNet-Affect features
	SentiWordNet features
	Hindi WordNet-Affect ^a features
Kim et al. [81]	WordNet-Affect features
	ANEW [22] features
Muljono et al. [105]	TF-IDF
Mulki et al. [106]	TF-IDF

Table 4 continued

Paper	Features
Plaza-del-Arco et al. [119]	TF-IDF Lexicon features WordNet-Affect (English and Spanish) NRC Word-Emotion Association Lexicon (English and Spanish) Spanish Emotion Lexicon
Singh et al. [140]	Tokens
Thomas et al. [153]	n -grams ($n = 1, 2, 3$)
Xu et al. [167]	POS Repeated exclamation marks (!) and question marks (?) Negative words and adjacent adjectives (extract up to three words in length)
Xu et al. [168]	Linguistic features Lemma unigram Negation Named entities replacement (NER) Sentiment lexicon features From the sentiment lexicons: Bing Liu lexicon, General Inquirer lexicon IMDB, MPQA, NRC Emotion Sentiment lexicon, AFINN, and NRC Hashtag Sentiment Lexicon, and NRC Sentiment140 Lexicon, the following features were extracted: The ratio of positive words to all words The ratio of negative words to all words The maximum sentiment scores The minimum sentiment scores The sum of sentiment scores The sentiment score of the last word in a tweet Emotion lexicon features From the lexicons: NRC Hashtag Sentiment Lexicon, NRC Affect Intensity Lexicon, NRC Word-Emotion Association Lexicon, and ANEW Lexicon, the following features were extracted: The maximum scores The sum of scores The number of words exists in lexicons Domain-specific features Punctuation Bag-of-Hashtags Emoticon Intensity Words
Yan and Turtle [169]	n -grams
Yuan and Purver [173]	Character n -grams ($n = 1, 2, 3, 4, 5$) Word n -grams ($n = 1, 2, 3, 4, 5$)
Zhang et al. [174]	Tokens KTM

Table 4 continued

Paper	Features
Abdullah and Shaikh [1]	Word Embeddings AraVec pretrained embeddings on Arabic tweets Word2vecs trained on Google News collection
Basile et al. [16]	Bi-LSTM with deep self-attention mechanism USE BERT
Baziotis et al. [18]	Bi-LSTM with deep attention mechanism
Du and Nie [42]	GRU
Ezen-Can and Can [48]	GRU Bi-GRU
Ge et al. [53]	BiLSTM CNN
Li et al. [87]	LSTM
Ma et al. [91]	Bi-LSTM with max-pooling
Meisheri and Dey [93]	Bi-LSTM with attention layer Bi-LSTM with max-pooling
Ragheb et al. [124]	Bi-LSTM
Rathnayaka et al. [125]	Bi-GRU with pyramid attention network
Seyeditabari et al. [135]	Bi-GRU
Shrivastava et al. [138]	CNN
Wang et al. [159]	CNN
Xiao [166]	ULM BERT OpenAI's GPT DeepMoji DeepMoji trained with NTUA embedding
Agrawal and Suri et al. [3]	Term frequency (TF) of word n -grams ($n = 2, 3$) TF of character n -grams ($n = 2, 3$) Lexicon features VAD lexicon Emotion intensity using NRC Word-Emotion Association Lexicon Style features Number of interrogation marks Number of exclamation marks Number of uppercase letters The length of each turn in the conversation (in words) The length of each turn in the conversation (in characters) SSWE embeddings [150] Signals from adult and offensive classifiers [171]
Amelia and Maulidevi [8]	Words from short stories
Badaro et al. [12]	n -grams ($n = 1, 2, 3$) Affect lexicons

Table 4 continued

Paper	Features
De Bruyne et al. [35]	ArSEL [13]
	ArSenL [11]
	Word embeddings
	AraVec trained on three datasets (Wikipedia, Text data from Web and Twitter)
	FastText trained on Wikipedia
	Character n -grams ($n = 2, 3, 4, 5$)
	Word n -grams ($n = 1, 2$)
	Lexicon features
	TweetToLexiconFeatureVector
	TweetToSentiStrengthFeatureVector
	Warriner et al. [160]
	Style features
	Average word/sentence length
	Number of words and sentences
	Number of capitals
	Number of punctuation marks
	Number of nonstandard words
	Number of connectives
	Syntactic features
	POS n -grams
	POS frequencies
	POS first token
	Presence imp.
	Presence fut.
	Semantic features
	Synset depth
	TweetToEmbeddingsFeatureVector ^b
	Clusters
Gee and Wang [54]	LSTM
	Bi-LSTM
	TweetToLexiconFeatureVector
Gievska et al. [57]	Tokens
Haggag [63]	Semantic frames
Herzig et al. [65]	BOW
	CBOw
	TF-IDF weights
	Classifier weights (CLASS)
Kim et al. [82]	CNN
	NRC lexicon [103]

Table 4 continued

Paper	Features
Kravchenko and Pivovarova [85]	Lexicon features DepecheMood [145] Vader (Valence Aware Dictionary and sEntiment Reasoner) ^c [72] Word embeddings GloVe trained on Common Crawl GloVe trained on Twitter Word2vecs trained on Google News collection
Li et al. [88]	TF
Park et al. [113]	Emotional Word Vectors (300 dimensions) learned by CNN Emoji Sentence Representations DeepMoji representations (2304 dimensions) learned by Bi-LSTM with attention layer Emoji cluster representations (512 dimensions) Tweet-specific features Number of words in uppercase Number of positive and negative emoticons Sum of emoji valence score ^d Number of elongated words The number of exclamation and question marks
Riahi and Safari [127]	n -grams ($n = 1, 2$) POS Frequency of emotion words in WordNet-Affect ANEW [22]
Rozental and Fleischer [130]	Bi-GRU with a CNN as an attention mechanism NRC Affect Intensity lexicon Sentence emotions score Semantic and syntactic features: Number of magnifier and diminisher words Logarithm of length of sentences Existence of elongated words Fully capitalized words The symbols # and @ appearing in the sentence Predictions of external packages: Vader TextBlob ^e

Table 4 continued

Paper	Features
Seol et al. [134]	Hierarchical features defined by a set of rules
Shaheen et al. [136]	Emotion recognition rules (ERRs) WordNet-Affect features SentiWordNet features ConceptNet features

^a<http://www.cfilt.iitb.ac.in/~wordnet/wn.old>^b<https://affectivetweets.cms.waikato.ac.nz>^c<https://github.com/cjhutto/vaderSentiment>^d<https://github.com/words/emoji-emotion>^e<https://github.com/slوريا/textblob>**Table 5** Strengths and limitations of the reviewed papers

Paper	Strengths	Limitations
Ma et al. [90]	Simple approach	The use of a weak strategy to manage negation
Perikos and Hatzilygeroudis [116]	The use of a dependency graph to specify the emotional word strength	The patterns could be extended to cover complex sentences
Shivhare et al. [137]	Simple approach	No use of semantic information
Tao [151]	Simple approach	Unable to recognize implicit emotions
Lee et al. [86]	The creation of 15 linguistic rules	The performance results were low
Udochukwu and He [155]	The use of lexicon and syntactic information	Sensitive to the text quality
Alm et al. [4]	The use of syntactic, stylistic, and lexicon features	No use of semantic features
Aman and Szpakowicz [6]	The use of emoticons information	No use of syntactic or semantic features
Aman and Szpakowicz [7]	A lexicon of emotion-related words was built automatically	No use of syntactic or semantic features
Anusha and Sandhya [9]	Use of semantic information	The use of one feature; more features must be explored
Bandhakavi et al. [15]	The use of lexicon, stylistic, and syntactic features The use of emoticon and hashtag information	No use of semantic features
Chaffar and Inkpen [28]	Several classification algorithms were explored	No use of syntactic or semantic features
Danisman and Alpkocak [33]	Several classification algorithms were explored	No use of syntactic or semantic features The performance results were low

Table 5 continued

Paper	Strengths	Limitations
Deborah et al. [37]	Simple learning-based system	<p>The informal style of writing in tweets was not taken into consideration when preprocessing</p> <p>The lack of used features</p> <p>No use of emojis and emoticons</p> <p>No use of affect lexicons</p> <p>No use of syntactic and semantic features</p> <p>No use of automatically extracted features</p>
Desmet and Hoste [38]	<p>The use of syntactic and lexicon features</p> <p>The correction of spelling errors</p> <p>Able to recognize 15 emotions</p>	<p>The class size impacted the performance</p> <p>No use of semantic features</p>
Douiji et al. [41]	<p>Easily configured to add new concepts</p> <p>The use of semantic and syntactic relations</p>	The semantic rules were insufficient and must be extended
Gao et al. [52]	The use of lexicon, syntactic, and semantic features	<p>No consideration of the polysemy in the construction process of the emotional lexicon</p> <p>No use of some complicated linguistic patterns</p>
Ghazi et al. [55]	Reduced impact of the class size on the performance	No use of syntactic or semantic features
Ghazi et al. [56]	The use of different types of features	Only sentences with one or more emotion words were used
Ho and Cao [66]	Took into account linguistic information and psychological characteristic of emotion	The performance results were low
Jain et al. [76]	Emotion recognition in multilingual text	No use of stylistic, syntactic, and semantic features
Kim et al. [81]	Estimated categorical model and dimensional model for emotion recognition	No use of syntactic, or semantic features
Muljono et al. [105]	Several classification algorithms were explored	No use of syntactic or semantic features

Table 5 continued

Paper	Strengths	Limitations
Mulki et al. [106]	Used emoji information	The performance results were low No use of semantic, syntactic or stylistic features
Plaza-del-Arco et al. [119]	The informal style of writing in tweets was taken into consideration when preprocessing	No use of emojis No use of syntactic and semantic features
Singh et al. [140]	The semantics and statistics information were considered for feature selection	Did not explore other types of features
Thomas et al. [153]	Simple approach	No use of syntactic or semantic features
Xu et al. [167]	Able to recognize 19 emotions The use of stylistic and syntactic features	No use of semantic features Complex feature space The informal style of writing in Chinese blog was not taken into consideration when preprocessing
Xu et al. [168]	The use of emojis, emoticons, and hashtags information Different types of features were explored Different classification algorithms were explored	No use of semantic features
Yan and Turtle [169]	Able to recognize 28 emotions Took into consideration Twitter style of writing in preprocessing Included emoji and emoticon information	No use of syntactic or semantic features
Yuan and Purver [173]	Took into consideration the informal style of writing in Chinese blog in preprocessing	No use of stylistic, syntactic, or semantic features
Zhang et al. [174]	Took into consideration Sina Weibo style of writing	Time-consuming
Abdullah and Shaikh [1]	The binary formulation of a multi-label multiclass classification problem allowed each DNN to learn the optimized weights for each emotion separately	No use of emojis, emoticons, and hashtag information The performance results were low

Table 5 continued

Paper	Strengths	Limitations
Basile et al. [16]	The use of different deep submodels	The <i>F</i> -score of emotion label happy is lower than sad and angry
Baziotis et al. [18]	The vocabulary and expressions were preserved by taking into consideration Twitter style of writing The vocabulary size was reduced without losing any words by performing word normalization and segmentation Spell correction was performed	Out-of-vocabulary words problem
Du and Nie [42]	The use of semantic features	Did not use emojis, emoticons, and hashtag information The performance results were low
Ezen-Can and Can [48]	The use of binary models allowed for a better handle on unbalanced data The use of pretrained weights The use of emoji embeddings The use of a greedy approach for dividing hashtags corresponding words	No use of affect lexicons
Ge et al. [53]	The ability to learn local features and long-term features	The <i>F</i> -score of emotion label happy is lower than sad and angry
Li et al. [87]	Emoticons and hashtags information were preserved The use of semantic features	The performance results were low The classification of the tweet with three emotions regardless of how low the calculated scores were
Ma et al. [91]	Emojis information were preserved	The <i>F</i> -score of emotion label happy is lower than sad and angry
Meisheri and Dey [93]	Overcame the out-of-vocabulary words problem by using a robust representation that utilized several pretrained emojis, words and characters	The # symbol was removed from hashtags, and no further preprocessing, such as segmentation, was performed No use of emoticon information

Table 5 continued

Paper	Strengths	Limitations
Ragheb et al. [124]	Easily configured to add new concepts	The <i>F</i> -score of emotion label happy is lower than sad and angry
	Easily applied to other emotional and sentiment classification	
Rathnayaka et al. [125]	Learning semantic and context features	The low recognition of underrepresented classes
Seyeditabari et al. [135]	Learning semantic and context features	The informal style of writing in tweets was not taken into consideration when preprocessing
Shrivastava et al. [138]	Learning semantic and context features	Some classes get confused with others such as anger with disgust and happiness with surprise
Wang et al. [159]	The use of semantic features	Only explored one language model, more language models must be explored
Xiao [166]	The informal style of writing in tweets was taken into consideration when preprocessing	The <i>F</i> -score of emotion label happy is lower than sad and angry
	Emojis and hashtags information were preserved	
	The use of semantic features	
Agrawal and Suri et al. [3]	The model is robust to spelling mistakes, slangs, abbreviations, and emoticons	The <i>F</i> -score of emotion label happy is lower than sad and angry
Amelia and Maulidevi [8]	Several classification algorithms were explored	No use of syntactic or semantic features
Badaro et al. [12]	The use of emoji information	The test of each feature separately; no hybridization of different types of features
	The testing of several human-engineered features, affect lexicons and pretrained embedding	Sparsity caused by dialectal Arabic
	The out-of-vocabulary words problem was managed by taking the average of the embeddings of the words that had the smallest Levenshtein distance to the target word	

Table 5 continued

Paper	Strengths	Limitations
De Bruyne et al. [35]	The use of wide range of features A customized model for each emotion label	The unbalanced dataset impacted the performance
Gee and Wang [54]	The data preprocessing was tailored to Twitter Improved performance of multi-label multiclass emotion recognition by training correlated emotion labels together	Out-of-vocabulary words problem No use of syntactic and stylistic features
Gievska et al. [57]	Explored several classification algorithms	No use of syntactic or semantic features
Haggag [63]	Used semantic features The use of evolutionary algorithm to train the KBANN	The performance was sensitive to the sample size
Herzig et al. [65]	The hybridization of human-engineered features and automatically extracted features	Only explored two pretrained word embeddings
Kim et al. [82]	The use of emoji and hashtag information The use of semantic features	No use of syntactic and stylistic features
Kravchenko and Pivovarova [85]	The use of hashtag information	No use of emoji and emoticon information No use of syntactic and stylistic features
Li et al. [88]	Semantic features were learned by hybridizing neural network with LSM	The performance results were low
Park et al. [113]	The hybridization of human-engineered features and automatically extracted features Robust representation of emojis	No use of emoticon information
Riahi and Safari [127]	The hybridization of three different approaches where each one analyzed the data from a different aspect	High dimensionality No use of semantic features

Table 5 continued

Paper	Strengths	Limitations
Rozental and Fleischer [130]	<p>The hybridization of human-engineered features and automatically extracted features</p> <p>The use of affect lexicon features</p> <p>The use of emojis in calculating the sentence emotions score feature</p> <p>The embeddings were trained on two hundred million tweets</p>	Low-performance result on macro-average F1
Seol et al. [134]	The use of domain knowledge and semantic features	Highly susceptible to misclassification
Shaheen et al. [136]	<p>Used syntactic and semantic features</p> <p>Achieved high performance result</p>	Required a large dataset for training

5 Discussion

Different approaches have been examined to address emotion recognition in text. The keyword-based approaches are the main approach for explicit emotion recognition. However, these approaches do not always succeed in recognizing explicit emotions. If a sentence expresses an emotion but does not include any word from the emotion keyword set, then the emotion will not be recognized. Even when a sentence includes an emotion keyword, it is not guaranteed to express the same emotion because the word meaning can change according to the context. The main approaches for implicit emotion recognition are rule-based approaches, learning-based approaches, deep learning approaches, and hybrid approaches. Rule-based approaches are affected by the quality of the text. Thus, if the text is written in an informal style and contains several grammatical mistakes, then this approach may not be able to recognize the implicit emotion correctly. Additionally, an implicit emotion can only be recognized if there is a rule that represents it in the rule set. Classical learning-based approaches need efficient features to be able to recognize implicit emotion. Human-engineered features do not cover all the cases of how emotions are expressed. Therefore, many implicit emotions are mislabeled or missed, and only those that the learning approach is trained to recognize are successfully recognized. Deep learning offers high-quality features and eliminates the need for feature engineering, which is one of the most time-consuming parts of machine learning practice. However, deep learning requires a large quantity of training data. Implementing a hybrid approach can improve the results because it takes advantage of the approaches integrated into it. However, the disadvantages and limitations of these approaches can also be inherited.

This study shows that Chinese is the most dominant language after English in terms of emotion recognition in text. Additionally, there is newly published research from other languages, including Arabic, Hindi, Indonesia, and Spanish. For emotion recognition in English, some researchers used one or more existing emotion-annotated corpora—Alm, Aman, ISEAR, SemEval-2007, SemEval-2018, and SemEval-2019—to measure the performance of their models, while others used their own created corpora, such as Ma et al. [90], Shivhare et al. [137], Perikos and Hatzilygeroudis [116], Bandhakavi et al. [15], Douiji et al. [41], Yan and

Turtle [169], and Haggag [63], to measure the performance of their models. For emotion recognition in Chinese, most of the work except for that of Lee et al. [86] and Wang et al. [159] evaluated the performance of their approach using a self-built corpus but used the same resource for text, Sina Weibo, which is a Chinese microblogging Web site. The researchers who created their own corpus had the opportunity to test their model on recognizing more emotions. However, as the number of emotions increases, the difficulties also increase, which results in a performance reduction.

This study shows that the most used learning-based method is SVM. In comparison with other methods, SVM almost obtained the best results. (It achieved the second best result in Ghazi et al. [56].) The most important part of any learning-based approach is the features. The success of the approach depends on whether the correct set of features is selected. Representation learning gained attention due to the success of word embeddings by Mikolov et al. [95,96]. This study shows that utilizing word embeddings and deep neural networks enhances the performance. It also shows that hybrid approaches managed to reach good results in the annotated dataset. Herzig et al. [65] tested the performance of their model, which combines the traditional representation of text and word embeddings on Alm dataset, ISEAR dataset, and SemEval-2007 dataset. Their system obtained one of the best results for the ISEAR dataset and Alm dataset, but it did not achieve good results for the SemEval-2007 dataset. The ISEAR dataset offers over 1000 instances per class, which the SemEval-2007 dataset does not have. To overcome the dataset size limitation, many participants in the SemEval-2018 competition used transfer learning to pretrain the weights of the deep neural networks.

Although English corpora exist, some of them are not large enough to train deep learning approaches. Moreover, the accuracy of recognizing an emotion is reliant on how well-balanced the dataset is. In the case of the Alm dataset, only 9.86% of the sentences are labeled as expressing a positive emotion, and the result of recognizing this class is the worst. In SemEval-2019, the low performance of recognizing the emotion label happy is a common problem among the participants in this competition. The opposite is true for Yuan and Purver [173], where the highest accuracy for recognizing an emotion was obtained for the happy emotion label, as the size of the annotated sentences for this class is the largest. In SemEval-2018, the highest performance result was 58.8, which is not that high, and looking at the dataset, it is very imbalanced. Almahdawi and Teahan [5] tested the effect of downsizing the classes to the class with the smallest size, and the accuracy significantly improved.

Emotion recognition in text has made some progress in the last few years. Looking at SemEval-2018 and SemEval-2019 participants, it is apparent that deep learning approaches are dominating the emotion recognition field. More language models were created [27,39,67,123]. Strong deep models were built along with creative deep attention mechanisms. Nevertheless, more research must be done to overcome the following challenges:

- **The difficulties of recognizing implicit emotions** Emotions are complex; humans have problems expressing and understanding them. Recognizing emotions in text increases the difficulty of understanding emotion due to the lack of visible facial expressions, body gestures, and voice. Automating emotion recognition is a difficult task. A machine needs to deal with the complexity of linguistics and the context of written text.
- **The quality of the datasets** The available datasets are not large enough to cope with the new trends, especially deep learning. Moreover, all of the datasets except for ISEAR are imbalanced. This shifts the focus from the task of emotion recognition to how to deal with problems caused by under represented classes. Thus, high-quality data must be created to improve emotion recognition models.

- **The limited resources in languages other than English** Emotion recognition in other languages is not as advanced as emotion recognition in English. Thus, resources, including high-quality data and lexicons, must be created for other natural languages.

In the future, we predict that using pretrained word embeddings in emotion recognition in text would be a standard practice, and new pretrained models will be developed. Further, transfer learning would play more important role, especially with the lack of large datasets to train deep learning models. Lastly, transformer [157] models would dominate deep learning models. The transformer model is gaining popularity and has already been used in Open AI's GPT-2, which is a successor to the GPT language model. Moreover, Dai et al. [32] proposed a new improvement to the transformer model that enables learning dependency beyond a fixed length.

6 Conclusion

In this paper, we surveyed existing approaches for both explicit and implicit emotion recognition in text. The keyword-based approaches are mostly used for explicit emotion recognition. However, they fail to fully recognize implicit emotion in text due to the lack of linguistic information. The main approaches for implicit emotion recognition include rule-based approaches, classical learning-based approaches, deep learning approaches, and hybrid approaches. Rule-based approaches can only recognize implicit emotion that are already represented in their rule sets. Classical learning-based approaches can recognize implicit emotions, given that the classifier has been already trained on such types of emotions. However, they do not require large training datasets to achieve a reasonable performance. Deep learning-based approaches can outperform the other approaches, given a very large quantity of training data. Hybrid approaches generally inherit the advantages of the approaches integrated into them, in addition to their disadvantages and limitations. Although most of the best results are obtained by learning-based approaches, deep learning approaches, and hybrid approaches, there are other approaches that performed rather well and must be further investigated. These include the compression-based approach and constraint optimization approach.

The results of this work show that POS tagging, parsing, and other simple NLP tasks can highly impact the performance of emotion recognition systems. This study also identified the sets of features used by the best performing approaches. It highlighted the features automatically extracted by deep learning models, which can capture explicit and implicit information. Combining handcrafted features and word embedding for classical machine learning or deep learning approaches represents a promising research avenue.

Acknowledgements This work was supported by the Research Center of the College of Computer and Information Sciences, King Saud University. We are grateful for this support. We also would like to thank the anonymous reviewers for their valuable and insightful comments.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Human and animal rights This article does not contain any studies with human or animal subjects performed by any of the authors.

Informed consent Informed consent was not required, as no human or animals were involved.

References

1. Abdullah M, Shaikh S (2018) TeamUNCC at SemEval-2018 task 1: emotion detection in English and Arabic tweets using deep learning. In: Proceedings of the 12th international workshop on semantic evaluation. Association for Computational Linguistics, pp 350–357
2. Agrawal A, An A (2012) Unsupervised emotion detection from text using semantic and syntactic relations. In: Proceedings of the 2012 IEEE/WIC/ACM international joint conferences on web intelligence and intelligent agent technology. IEEE Computer Society, Washington, DC, WI-IAT '12, pp 346–353
3. Agrawal P, Suri A (2019) NELEC at SemEval-2019 task 3: think twice before going deep. In: May J, Shutova E, Herbelot A, Zhu XZ, Apidianaki M, Mohammad SM (eds) Proceedings of the 13th international workshop on semantic evaluation. Association for Computational Linguistics, Minneapolis, pp 266–271
4. Alm CO, Roth D, Sproat R (2005) Emotions from text: machine learning for text-based emotion prediction. In: Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, Stroudsburg, PA, HLT '05, pp 579–586
5. Almahdawi A, Teahan WJ (2017) Emotion recognition in text using PPM. In: Bramer M, Petridis M (eds) Artificial intelligence XXXIV, vol 10630. Lecture notes in computer science. Springer, Cham, pp 149–155
6. Aman S, Szpakowicz S (2007) Identifying expressions of emotion in text. In: Proceedings of the 10th international conference on text, speech and dialogue, TSD'07. Springer, Berlin, pp 196–205
7. Aman S, Szpakowicz S (2008) Using roget's thesaurus for fine-grained emotion recognition. In: Proceedings of the 3rd international joint conference on natural language processing (IJCNLP), pp 312–318
8. Amelia W, Maulidevi NU (2016) Dominant emotion recognition in short story using keyword spotting technique and learning-based method. In: 2016 International conference on advanced informatics: concepts, theory and application (ICAICTA), pp 1–6
9. Anusha V, Sandhya B (2015) A learning based emotion classifier with semantic text processing. In: El-Alfy MES, Thampi MS, Takagi H, Piramuthu S, Hanne T (eds) Advances in intelligent informatics. Springer, Cham, pp 371–382
10. Baccianella S, Esuli A, Sebastiani F (2010) Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the seventh conference on international language resources and evaluation (LREC'10). European Language Resources Association (ELRA), Valletta, vol 25, pp 2200–2204
11. Badaro G, Baly R, Hajj H, Habash N, El-Hajj W (2014) A large scale arabic sentiment lexicon for arabic opinion mining. In: Proceedings of the EMNLP 2014 workshop on arabic natural language processing (ANLP). Association for Computational Linguistics, pp 165–173
12. Badaro G, El Jundi O, Khaddaj A, Maarouf A, Kain R, Hajj H, El-Hajj W (2018) EMA at SemEval-2018 task 1: emotion mining for arabic. In: Proceedings of the 12th international workshop on semantic evaluation. Association for Computational Linguistics, pp 236–244
13. Badaro G, Jundi H, Hajj H, El-Hajj W, Habash N (2018) Arsel: a large scale arabic sentiment and emotion lexicon. In: The 3rd workshop on open-source arabic corpora and processing tools (OSACT3) co-located with LREC 2018
14. Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. [arXiv:1409.0473](https://arxiv.org/abs/1409.0473)
15. Bandhakavi A, Wiratunga N, Padmanabhan D, Massie S (2017) Lexicon based feature extraction for emotion text classification. Pattern Recognit Lett 93:133–142
16. Basile A, Franco-Salvador M, Pawar N, Štajner S, Chinea Rios M, Benajiba Y (2019) SymantoResearch at SemEval-2019 task 3: combined neural models for emotion classification in human-chatbot conversations. In: May J, Shutova E, Herbelot A, Zhu XZ, Apidianaki M, Mohammad SM (eds) Proceedings of the 13th international workshop on semantic evaluation. Association for Computational Linguistics, Minneapolis, pp 330–334
17. Baziotis C, Pelekis N, Doukeridis C (2017) Datastories at semeval-2017 task 4: deep LSTM with attention for message-level and topic-based sentiment analysis. In: Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017). Association for Computational Linguistics, pp 747–754
18. Baziotis C, Nikolaos A, Chronopoulou A, Kolovou A, Paraskevopoulos G, Ellinas N, Narayanan S, Potamianos A (2018) NTUA-SLP at SemEval-2018 task 1: predicting affective content in tweets with deep attentive rnns and transfer learning. In: Proceedings of The 12th international workshop on semantic evaluation. Association for Computational Linguistics, pp 245–255

19. Biagioni R (2016) Senticnet. In: The SenticNet sentiment lexicon: exploring semantic richness in multi-word concepts. Springer, Cham, pp 17–31
20. Binali H, Potdar V (2012) Emotion detection state of the art. In: Proceedings of the CUBE international information technology conference, CUBE'12. ACM, New York, pp 501–507
21. Bojanowski P, Grave E, Joulin A, Mikolov T (2017) Enriching word vectors with subword information. *Trans Assoc Comput Linguist* 5:135–146
22. Bradley MM, Lang PJ (1999) Affective norms for English words (ANEW): stimuli, instruction manual, and affective ratings. Tech. rep., Center for Research in Psychophysiology, University of Florida, Gainesville
23. Bravo-Marquez F, Frank E, Mohammad SM, Pfahringer B (2016) Determining word-emotion associations from tweets by multi-label classification. In: 2016 IEEE/WIC/ACM international conference on web intelligence, WI 2016. IEEE Computer Society, pp 536–539
24. Cambria E, Livingstone A, Hussain A (2012) The hourglass of emotions. In: Esposito A, Esposito AM, Vinciarelli A, Hoffmann R, Müller VC (eds) Cognitive behavioural systems. Springer, Berlin, pp 144–157
25. Canales L, Martínez-Barco P (2014) Emotion detection from text: a survey. In: Processing in the 5th information systems research working days (JISIC 2014), pp 37–43
26. Carlson A, Cumby C, Rosen J, Roth D (1999) The SNoW learning architecture. Tech. rep., Technical report UIUCDCS
27. Cer D, Yang Y, Kong S, Hua NH, Limtiaco N, St John R, Constant N, Guajardo-Cespedes M, Yuan S, Tar C, Sung Y, Strophe B, Kurzweil R (2018) Universal sentence encoder. CoRR abs/1803.11175
28. Chaffar S, Inkpen D (2011) Using a heterogeneous dataset for emotion analysis in text. In: Butz C, Lingras P (eds) Proceedings of the 24th Canadian conference on advances in artificial intelligence, Canadian AI'11. Springer, Berlin, pp 62–67
29. Chatterjee A, Narahari KN, Joshi M, Agrawal P (2019) Semeval-2019 task 3: emocontext: contextual emotion detection in text. In: Proceedings of the 13th international workshop on semantic evaluation (SemEval-2019), Minneapolis
30. Chen KJ, Huang CR, Chang LP, Hsu HL (1996) Sinica corpus: design methodology for balanced corpora. In: Proceedings of the 11th Pacific Asia conference on language, information and computation. Kyung Hee University, pp 167–176
31. Cohen WW (1995) Fast effective rule induction. In: Prieditis A, Russell S (eds) Machine learning proceedings 1995. Morgan Kaufmann, San Francisco, pp 115–123
32. Dai Z, Yang Z, Yang Y, Carbonell JG, Le QVL, Salakhutdinov R (2019) Transformer-xl: attentive language models beyond a fixed-length context. [arXiv:1901.02860](https://arxiv.org/abs/1901.02860)
33. Danisman T, Alpkocak A (2008) Feeler: emotion classification of text using vector space model. In: AISB 2008 convention communication, interaction and social intelligence, Aberdeen, vol 2, pp 53–60
34. Darwin C (1872) The expression of the emotions in man and animals. John Murray, London
35. De Bruyne L, De Clercq O, Hoste V (2018) LT3 at SemEval-2018 task 1: a classifier chain to detect emotions in tweets. In: Proceedings of the 12th international workshop on semantic evaluation. Association for Computational Linguistics, pp 123–127
36. Deborah SA, Milton R, Hannah S (2016) A survey of emotion analysis. *Middle East J Sci Res* 24:32–38
37. Deborah SA, Rajalakshmi S, Rajendram SM, Mirnalinee TT (2018) SSN MLRG1 at SemEval-2018 task 1: Emotion and sentiment intensity detection using rule based feature selection. In: Proceedings of the 12th international workshop on semantic evaluation. Association for Computational Linguistics, pp 324–328
38. Desmet B, Hoste VH (2013) Emotion detection in suicide notes. *Expert Syst Appl* 40(16):6351–6358
39. Devlin J, Chang M, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, NAACL-HLT 2019, vol 1, pp 4171–4186
40. Dong Z, Dong Q (1999) Hownet knowledge database
41. Douji Y, Mousannif H, Al Moatassime H (2016) Using youtube comments for text-based emotion recognition. *Procedia Comput Sci* 83:292–299
42. Du P, Nie JY (2018) Mutux at SemEval-2018 task 1: exploring impacts of context information on emotion detection. In: Proceedings of the 12th international workshop on semantic evaluation. Association for Computational Linguistics, pp 345–349
43. Eisner B, Rocktäschel T, Augenstein I, Bosnjak M, Riedel S (2016) emoji2vec: learning emoji representations from their description. In: Proceedings of the fourth international workshop on natural language processing for social media. Association for Computational Linguistics, pp 48–54

44. Ekman P (1999) Basic emotions. In: Dalgleish T, Power M (eds) *The handbook of cognition and emotion*. Wiley, New York, pp 45–60
45. Ellsworth PC (2013) Appraisal theory: old and new questions. *Emotion Rev* 5(2):125–131
46. Esuli A, Sebastiani F (2005) Determining the semantic orientation of terms through gloss classification. In: *Proceedings of the 14th ACM international conference on information and knowledge management, CIKM'05*. ACM, New York, pp 617–624
47. Esuli A, Sebastiani F (2006) Sentiwordnet: a publicly available lexical resource for opinion mining. In: *Proceedings of the 5th conference on language resources and evaluation (LREC'06)*. European Language Resources Association (ELRA), Genoa, pp 417–422
48. Ezen-Can A, Can EF (2018) RNN for affects at SemEval-2018 task 1: formulating affect identification as a binary classification problem. In: *Proceedings of the 12th international workshop on semantic evaluation*. Association for Computational Linguistics, pp 162–166
49. Felbo B, Mislove A, Søgaard A, Rahwan I, Lehmann S (2017) Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In: *Proceedings of the 2017 conference on empirical methods in natural language processing*. Association for Computational Linguistics, pp 1615–1625
50. Fellbaum C (1998) *WordNet: an electronic lexical database*. Language, speech, and communication. MIT Press, Cambridge
51. Frijda NH (1993) Moods, emotion episodes and emotions. In: Lewis M, Haviland JM (eds) *Handbook of emotions*. Guilford Press, New York, pp 381–403
52. Gao K, Xu H, Wang J (2014) Emotion classification based on structured information. In: *2014 International conference on multisensor fusion and information integration for intelligent systems (MFI)*, pp 1–6
53. Ge S, Qi T, Wu C, Huang Y (2019) THU_NGN at SemEval-2019 task 3: dialog emotion classification using attentional LSTM-CNN. In: May J, Shutova E, Herbelot A, Zhu XZ, Apidianaki M, Mohammad SM (eds) *Proceedings of the 13th international workshop on semantic evaluation*. Association for Computational Linguistics, Minneapolis, pp 340–344
54. Gee G, Wang E (2018) psyML at SemEval-2018 task 1: transfer learning for sentiment and emotion analysis. In: *Proceedings of the 12th international workshop on semantic evaluation*. Association for Computational Linguistics, pp 369–376
55. Ghazi D, Inkpen D, Szpakowicz S (2010) Hierarchical versus flat classification of emotions in text. In: *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, CAAGET'10*. Association for Computational Linguistics, Stroudsburg, pp 140–146
56. Ghazi D, Inkpen D, Szpakowicz S (2014) Prior and contextual emotion of words in sentential context. *Comput Speech Lang* 28(1):76–92
57. Gievska S, Koroveshevski K, Chavdarova T (2014) A hybrid approach for emotion detection in support of affective interaction. In: *2014 IEEE international conference on data mining workshop (ICDMW)*, pp 352–359
58. Godin F, Vandersmissen B, De Neve W, Van de Walle R (2015) Multimedia lab @ ACL WNUT NER shared task: Named entity recognition for twitter microposts using distributed word representations. In: *Proceedings of the workshop on noisy user-generated text*. Association for Computational Linguistics, Beijing, pp 146–153
59. Goodfellow I, Bengio Y, Courville A (2016) *Deep learning*. MIT Press, Cambridge
60. Grandjean D, Sander N, Scherer KR (2008) Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Conscious Cognit* 17(2):484–495
61. Graves A, Schmidhuber J (2005) Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural Netw* 18(5):602–610
62. Gunes H, Pantic M (2010) Automatic, dimensional and continuous emotion recognition. *Int J Synth Emot (IJSE)* 1(1):68–99
63. Haggag MH (2014) Frame semantics evolutionary model for emotion detection. *Comput Inf Sci* 7(1):136–161
64. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) *The weka data mining software: an update*. *SIGKDD Explor Newsl* 11(1):10–18
65. Herzig J, Shmueli-Scheuer M, Konopnicki D (2017) Emotion detection from text via ensemble classification using word embeddings. In: *Proceedings of the ACM SIGIR international conference on theory of information retrieval, ICTIR'17*. ACM, New York, pp 269–272
66. Ho DT, Cao TH (2012) A high-order hidden markov model for emotion detection from textual data. In: *Proceedings of the 12th Pacific rim conference on knowledge management and acquisition for intelligent systems, PKAW'12*. Springer, Berlin, pp 94–105

67. Howard J, Ruder S (2018) Universal language model fine-tuning for text classification. In: Proceedings of the 56th annual meeting of the association for computational linguistics, vol 1. Association for Computational Linguistics, Melbourne, pp 328–339
68. Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the 10th ACM SIGKDD international conference on knowledge discovery and data mining, KDD'04. ACM, New York, pp 168–177
69. Huang CR, Chen Y, Lee SYM (2010) Textual emotion processing from event analysis. In: Proceedings of the joint conference on Chinese language processing, Beijing
70. Hudlicka E (2015) Computational analytical framework for affective modeling: towards guidelines for designing computational models of emotions. In: Vallverdú J (ed) Handbook of research on synthesizing human emotion in intelligent systems and robotics. IGI Global, Hershey, pp 1–62
71. Hume D (2012) Emotion and moods. In: Robbins SP, Judge TA (eds) Organizational behaviour. Pearson, New York, pp 258–297
72. Hutto C, Gilbert E (2014) Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the 8th international conference on weblogs and social media, ICWSM 2014, pp 216–225
73. Izard CE (1971) The face of emotion. Century psychology series. Appleton-Century-Crofts
74. Izard CE (1977) Human emotions. Plenum Press, New York
75. Jain U, Sandhu A (2015) A review on the emotion detection from text using machine learning techniques. *Int J Curr Eng Technol* 5(4):2645–2650
76. Jain VK, Kumar S, Fernandes SL (2017) Extraction of emotions from multilingual text using intelligent text processing and computational linguistics. *J Comput Sci* 21:316–326
77. Jarmasz M, Szpakowicz S (2001) The design and implementation of an electronic lexical knowledge base. In: Stroulia E, Matwin S (eds) Advances in artificial intelligence. Lecture notes in artificial intelligence, vol 2056. Springer, Berlin, pp 325–334
78. Jin X, Wang Z (2005) An emotion space model for recognition of emotions in spoken Chinese. In: Proceedings of the first international conference on affective computing and intelligent interaction, ACHI'05. Springer, Berlin, pp 397–402
79. Kao ECC, Liu CC, Yang TH, Hsieh CT, Soo VW (2009) Towards text-based emotion detection—a survey and possible improvements. In: Proceedings of the 2009 international conference on information management and engineering, ICIME'09. IEEE Computer Society, Washington, pp 70–74
80. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY (2017) Lightgbm: a highly efficient gradient boosting decision tree. In: Advances in neural information processing systems. Curran Associates, Inc., pp 3146–3154
81. Kim SM, Valitutti A, Calvo RA (2010) Evaluation of unsupervised emotion models to textual affect recognition. In: Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, CAAGET'10. Association for Computational Linguistics, Stroudsburg, pp 62–70
82. Kim Y, Lee H, Jung K (2018) AttnConvnet at SemEval-2018 task 1: attention-based convolutional neural networks for multi-label emotion classification. In: Proceedings of the 12th international workshop on semantic evaluation. Association for Computational Linguistics, pp 141–145
83. Kiritchenko S, Zhu X, Mohammad SM (2014) Sentiment analysis of short informal texts. *J Artif Intell Res (JAIR)* 50(1):723–762
84. Kleinginna PR, Kleinginna AM (1981) A categorized list of emotion definitions, with suggestions for a consensual definition. *Motiv Emotion* 5(4):345–379
85. Kravchenko D, Pivovarova L (2018) DL Team at SemEval-2018 task 1: tweet affect detection using sentiment lexicons and embeddings. In: Proceedings of the 12th international workshop on semantic evaluation. Association for Computational Linguistics, pp 172–176
86. Lee SYM, Chen Y, Huang CR (2010) A text-driven rule-based system for emotion cause detection. In: Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, CAAGET'10. Association for Computational Linguistics, Stroudsburg, pp 45–53
87. Li M, Dong Z, Fan Z, Meng K, Cao J, Ding G, Liu Y, Shan J, Li B (2018) ISCLAB at SemEval-2018 task 1: Uir-miner for affect in tweets. In: Proceedings of the 12th international workshop on semantic evaluation. Association for Computational Linguistics, pp 286–290
88. Li X, Pang J, Mo B, Rao Y (2016) Hybrid neural networks for social emotion detection over short text. In: 2016 International joint conference on neural networks (IJCNN), pp 537–544
89. Liu H, Singh P (2004) Conceptnet—a practical commonsense reasoning tool-kit. *BT Technol J* 22(4):211–226

90. Ma C, Prendinger H, Ishizuka M (2005) Emotion estimation and reasoning based on affective textual interaction. In: Tao J, Tieniu T, Picard RW (eds) *Affective computing and intelligent interaction*. Springer, Berlin, pp 622–628
91. Ma L, Zhang L, Ye W, Hu W (2019) PKUSE at SemEval-2019 task 3: emotion detection with emotion-oriented neural attention network. In: May J, Shutova E, Herbelot A, Zhu XZ, Apidianaki M, Mohammad SM (eds) *Proceedings of the 13th international workshop on semantic evaluation*. Association for Computational Linguistics, Minneapolis, pp 287–291
92. Manning C, Surdeanu M, Bauer J, Finkel J, Bethard S, McClosky D (2014) The stanford corenlp natural language processing toolkit. In: *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*. Association for Computational Linguistics, pp 55–60
93. Meisheri H, Dey L (2018) TCS research at SemEval-2018 task 1: learning robust representations using multi-attention architecture. In: *Proceedings of the 12th international workshop on semantic evaluation*. Association for Computational Linguistics, pp 291–299
94. Merity S, Xiong C, Bradbury J, Socher R (2017) Pointer sentinel mixture models. In: *5th International conference on learning representations*, ICLR 2017
95. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*
96. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: *Proceedings of the 26th international conference on neural information processing systems*, Vol 2, NIPS'13. Curran Associates Inc., USA, pp 3111–3119
97. Mikolov T, Grave E, Bojanowski P, Puhrsch C, Joulin A (2018) Advances in pre-training distributed word representations. In: *Proceedings of the eleventh international conference on language resources and evaluation (LREC-2018)*. European Languages Resources Association (ELRA), Miyazaki
98. Mohammad SM (2012) #Emotional tweets. In: *Proceedings of the first joint conference on lexical and computational semantics—volume 1: proceedings of the main conference and the shared task, and volume 2: proceedings of the sixth international workshop on semantic evaluation, SemEval'12*. Association for Computational Linguistics, Stroudsburg, pp 246–255
99. Mohammad SM (2018) Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In: *Proceedings of the annual conference of the association for computational linguistics (ACL)*, pp 174–184
100. Mohammad SM (2018) Word affect intensities. In: *Proceedings of the 11th edition of the language resources and evaluation conference (LREC-2018)*, Miyazaki
101. Mohammad SM, Kiritchenko S (2015) Using hashtags to capture fine emotion categories from tweets. *Comput Intell* 31(2):301–326
102. Mohammad SM, Turney PD (2010) Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon. In: *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, CAAGET'10*. Association for Computational Linguistics, Stroudsburg, pp 26–34
103. Mohammad SM, Turney PD (2013) Crowdsourcing a word-emotion association lexicon. *Comput Intell* 29(3):436–465
104. Mohammad SM, Bravo-Marquez F, Salameh M, Kiritchenko S (2018) SemEval-2018 Task 1: affect in tweets. In: *Proceedings of international workshop on semantic evaluation (SemEval-2018)*, New Orleans
105. Muljono, Winarsih NAS, Supriyanto C (2016) Evaluation of classification methods for Indonesian text emotion detection. In: *2016 International seminar on application for technology of information and communication (ISemantic)*, pp 130–133
106. Mulki H, Bechikh Ali C, Haddad H, Babaoglu I (2018) Tw-StAR at SemEval-2018 task 1: preprocessing impact on multi-label emotion classification. In: *Proceedings of the 12th international workshop on semantic evaluation*. Association for Computational Linguistics, pp 167–171
107. Neviarouskaya A, Prendinger H, Ishizuka M (2009) Compositionality principle in recognition of fine-grained emotions from text. In: *Proceedings of the third international ICWSM conference*, pp 278–281
108. Neviarouskaya A, Prendinger H, Ishizuka M (2010) AM: textual attitude analysis model. In: *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, CAAGET'10*. Association for Computational Linguistics, Stroudsburg, pp 80–88
109. Nielsen FÅ (2011) A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In: *Proceedings of the ESWC2011 workshop on 'Making Sense of Microposts': big things come in small packages*. Heraklion, Crete, pp 93–98
110. Ortony A, Clore GL, Collins A (1990) *The cognitive structure of emotions*. Cambridge University Press, Cambridge
111. Owoputi O, O'Connor B, Dyer C, Gimpel K, Schneider N, Smith NA (2013) Improved part-of-speech tagging for online conversational text with word clusters. In: *Proceedings of the 2013 conference of the*

- North American chapter of the association for computational linguistics: human language technologies. Association for Computational Linguistics, pp 380–390
112. Panksepp J (2004) Affective neuroscience: the foundations of human and animal emotions. Series in affective science. Oxford University Press, Oxford
 113. Park JH, Xu P, Fung P (2018) PlusEmo2Vec at SemEval-2018 task 1: exploiting emotion knowledge from emoji and #hashtags. In: Proceedings of the 12th international workshop on semantic evaluation. Association for Computational Linguistics, pp 264–272
 114. Parrott WG (ed) (2001) Emotions in social psychology: essential readings. Key readings in social psychology. Psychology Press, New York
 115. Pennington J, Socher R, Manning C (2014) Glove: global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). Association for Computational Linguistics, pp 1532–1543
 116. Perikos I, Hatzilygeroudis I (2013) Recognizing emotion presence in natural language sentences. In: Iliadis L, Papadopoulos H, Jayne C (eds) Engineering applications of neural networks. Springer, Berlin, pp 30–39
 117. Pestian J, Nasrallah H, Matykiewicz P, Bennett A, Leenaars A (2010) Suicide note classification using natural language processing: a content analysis. *Biomed Inform Insights* 3:19–28
 118. Picard RW (1997) Affective computing. MIT Press, Cambridge
 119. Plaza-del Arco FM, Jiménez-Zafra SM, Martín M, Ureña-López LA (2018) SINAI at SemEval-2018 task 1: emotion recognition in tweets. In: Proceedings of the 12th international workshop on semantic evaluation. Association for Computational Linguistics, pp 128–132
 120. Plutchik R (2001) The nature of emotions. *Am Sci* 89(4):344–350
 121. Posner J, Russell JA, Peterson BS (2005) The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev Psychopathol* 17(3):715–734
 122. Quan C, Ren F (2010) A blog emotion corpus for emotional expression analysis in chinese. *Comput Speech Lang* 24(4):726–749
 123. Radford A, Narasimhan K, Salimans T, Sutskever I (2018) Improving language understanding by generative pre-training. Tech. rep., Preprint. https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf
 124. Ragheb W, Azé J, Bringay S, Servajean M (2019) LIRMM-advance at SemEval-2019 task 3: attentive conversation modeling for emotion detection and classification. In: May J, Shutova E, Herbelot A, Zhu XZ, Apidianaki M, Mohammad SM (eds) Proceedings of the 13th international workshop on semantic evaluation. Association for Computational Linguistics, Minneapolis, pp 251–255
 125. Rathnayaka P, Abeysinghe S, Samarajeewa C, Manchanayake I, Walpola MJ, Nawaratne R, Bandaragoda T, Alahakoon D (2019) Gated recurrent neural network approach for multilabel emotion detection in microblogs. arXiv preprint [arXiv:1907.07653](https://arxiv.org/abs/1907.07653)
 126. Řehůřek R, Sojka P (2010) Software framework for topic modelling with large corpora. In: Proceedings of the LREC 2010 workshop on new challenges for nlp frameworks. European Language Resources Association (ELRA), Valletta, pp 45–50
 127. Riahi N, Safari P (2016) Implicit emotion detection from text with information fusion. *J Adv Comput Res* 7(2):85–99
 128. Roseman JJ (1991) Appraisal determinants of discrete emotions. *Cognit Emotion* 5(3):161–200
 129. Rosenthal S, Farra N, Nakov P (2017) SemEval-2017 task 4: sentiment analysis in twitter. In: Proceedings of the 11th international workshop on semantic evaluation. Association for Computational Linguistics, Vancouver, SemEval'17
 130. Rozental A, Fleischer D (2018) Amobee at SemEval-2018 task 1: Gru neural network with a cnn attention mechanism for sentiment classification. In: Proceedings of the 12th international workshop on semantic evaluation. Association for Computational Linguistics, pp 218–225
 131. Russell JA (1980) A circumplex model of affect. *J Personal Soc Psychol* 39(6):1161–1178
 132. Scherer KR (2005) Appraisal theory. In: Dalgleish T, Power MJ (eds) Handbook of cognition and emotion. Wiley, New York, pp 637–663
 133. Scherer KR, Wallbott HG (1994) Evidence for universality and cultural variation of differential emotion response patterning. *J Personal Soc Psychol* 66(2):310–328
 134. Seol YS, Kim DJ, Kim HW (2008) Emotion recognition from text using knowledge-based ann. In: Proceedings of the 32nd international technical conference on circuits/systems, computers and communications (ITC-CSCC 2008), pp 1569–1572
 135. Seyeditabari A, Tabari N, Gholizadeh S, Zadrozny W (2019) Emotion detection in text: focusing on latent representation. arXiv preprint [arXiv:1907.09369](https://arxiv.org/abs/1907.09369)

136. Shaheen S, El-Hajj W, Hajj H, Elbassuoni S (2014) Emotion recognition from text based on automatically generated rules. In: 2014 IEEE international conference on data mining workshop (ICDMW), pp 383–392
137. Shivhare SN, Garg S, Mishra A (2015) Emotionfinder: detecting emotion from blogs and textual documents. In: International conference on computing, communication & automation (ICCCA), pp 52–57
138. Shrivastava K, Kumar S, Jain DK (2019) An effective approach for emotion detection in multimedia text data using sequence based convolutional neural network. *Multimed Tools Appl* 78:29607–29639
139. Sidorov G, Miranda-Jiménez S, Viveros-Jiménez F, Gelbukh A, Castro-Sánchez N, Velásquez F, Díaz-Rangel I, Suárez-Guerra S, Treviño A, Gordon J (2013) Empirical study of machine learning based approach for opinion mining in tweets. In: Batyrshin I, González Mendoza M (eds) *Advances in artificial intelligence*. Springer, Berlin, pp 1–14
140. Singh L, Singh S, Aggarwal N (2019) Two-stage text feature selection method for human emotion recognition. In: Krishna CR, Dutta M, Kumar R (eds) *Proceedings of 2nd international conference on communication, computing and networking, lecture notes in networks and systems*, vol 46. Springer, Singapore, pp 531–538
141. Smith CA, Ellsworth PC (1985) Patterns of cognitive appraisal in emotion. *J Pers Soc Psychol* 48(4):813–838
142. Smith CA, Lazarus RS (1993) Appraisal components, core relational themes, and the emotions. *Cognit Emotion* 7(3–4):233–269
143. Soliman AB, Eissa K, El-Beltagy SR (2017) Aravec: a set of Arabic word embedding models for use in Arabic NLP. *Procedia Comput Sci* 117:256–265
144. Speer R, Chin J, Havasi C (2017) Conceptnet 5.5: an open multilingual graph of general knowledge. In: *Proceedings of the thirty-first AAAI conference on artificial intelligence, AAAI'17*. AAAI Press, pp 4444–4451
145. Staiano J, Guerini M (2014) Depechemood: a lexicon for emotion analysis from crowd-annotated news. In: *Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: short papers)*, pp 427–433
146. Steunebrink BR, Dastani M, Meyer JJC (2009) The OCC model revisited. In: Reichardt D (ed) *Proceedings of the 4th workshop on emotion and computing—current research and future impact*, Paderborn, pp 40–47
147. Stone PJ, Dunphy DC, Smith MS, Ogilvie DM (1966) *The general inquirer: a computer approach to content analysis*. MIT Press, Cambridge
148. Strapparava C, Mihalcea R (2007) Semeval-2007 task 14: affective text. In: *Proceedings of the 4th international workshop on semantic evaluations, SemEval'07*. Association for Computational Linguistics, Stroudsburg, pp 70–74
149. Strapparava C, Valitutti A (2004) Wordnet-affect: an affective extension of wordnet. In: *Proceedings of the 4th international conference on language resources and evaluation (LREC-2004)*, pp 1083–1086
150. Tang D, Wei F, Yang N, Zhou M, Liu T, Qin B (2014) Learning sentiment-specific word embedding for twitter sentiment classification. In: *Proceedings of the 52nd annual meeting of the association for computational linguistics*, vol 1. Association for Computational Linguistics, Baltimore, pp 1555–1565
151. Tao J (2004) Context based emotion detection from text input. In: *Proceedings of the 8th international conference on spoken language processing (ICSLP)*, pp 1337–1340
152. Thelwall M, Buckley K, Paltoglou G (2012) Sentiment strength detection for the social web. *J Assoc Inf Sci Technol (JASIST)* 63(1):163–173
153. Thomas B, Vinod P, Dhanya KA (2014) Multiclass emotion extraction from sentences. *Int J Sci Eng Res (IJSER)* 5(2):12–15
154. Tomkins SS (1991) *Affect imagery consciousness: volume III: the negative affects: anger and fear*. Springer, Berlin
155. Udochukwu O, He Y (2015) A rule-based approach to implicit emotion detection in text. In: Biemann C, Handschuh S, Freitas A, Mezzanin F, Métais E (eds) *Natural language processing and information systems. Lecture notes in computer science*. Springer, Cham, pp 197–203
156. van der Goot R, van Noord G (2017) Monoise: modeling noise using a modular normalization system. *Comput Linguist Neth J* 7:129–144
157. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: *Proceedings of the 31st international conference on neural information processing systems, NIPS'17*. Curran Associates Inc, New York, pp 6000–6010
158. Wang M, Liu M, Feng S, Wang D, Zhang Y (2014) A novel calibrated label ranking based method for multiple emotions detection in chinese microblogs. In: Zong C, Nie JY, Zhao D, Feng Y (eds) *Natural language processing and chinese computing*. Springer, Berlin, pp 238–250

159. Wang Y, Feng S, Wang D, Yu G, Zhang Y (2016) Multi-label chinese microblog emotion classification via convolutional neural network. In: Li F, Shim K, Zheng K, Liu G (eds) *Web technologies and applications: APWeb 2016*, vol 9931. Lecture notes in computer science. Springer, Cham, pp 567–580
160. Warriner AB, Kuperman V, Brysbaert M (2013) Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behav Res Methods* 45(4):1191–1207
161. Watson D, Tellegen A (1985) Toward a consensual structure of mood. *Psychol Bull* 98(2):219–235
162. Watson D, Tellegen A (1999) Issues in dimensional structure of affect—effects of descriptors, measurement error, and response formats: comment on russell and carroll (1999). *Psychol Bull* 125:601–610
163. Weiss HM, Cropanzano R (1996) Affective events theory: a theoretical discussion of the structure, cause and consequences of affective experiences at work. In: Staw BM, Cummings LL (eds) *Research in organizational behavior: an annual series of analytical essays and critical reviews*, vol 18. JAI Press Inc, Stamford, pp 1–74
164. Wilson T, Wiebe J, Hoffmann P (2005) Recognizing contextual polarity in phrase-level sentiment analysis. In: *Proceedings of the conference on human language technology and empirical methods in natural language processing, HLT'05*. Association for Computational Linguistics, Stroudsburg, pp 347–354
165. Wundt WM (1904) *Principles of physiological psychology*. Swan Sonnenschein & Co., London
166. Xiao J (2019) Figure eight at SemEval-2019 task 3: ensemble of transfer learning methods for contextual emotion detection. In: May J, Shutova E, Herbelot A, Zhu XZ, Apidianaki M, Mohammad SM (eds) *Proceedings of the 13th international workshop on semantic evaluation*. Association for Computational Linguistics, Minneapolis, pp 220–224
167. Xu H, Yang W, Wang J (2015) Hierarchical emotion classification and emotion component analysis on chinese micro-blog posts. *Expert Syst Appl* 42(22):8745–8752
168. Xu H, Lan M, Wu Y (2018) ECNU at SemEval-2018 task 1: emotion intensity prediction using effective features and machine learning models. In: *Proceedings of the 12th international workshop on semantic evaluation*. Association for Computational Linguistics, pp 231–235
169. Yan JLS, Turtle HR (2016) Exploring fine-grained emotion detection in tweets. In: *Proceedings of the North American chapter of the association for computational linguistics: human language technologies (NAACL-HLT)*. San Diego, pp 73–80
170. Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E (2016) Hierarchical attention networks for document classification. In: *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*. Association for Computational Linguistics, San Diego, pp 1480–1489
171. Yenala H, Jhanwar A, Chinnakotla MK, Goyal J (2018) Deep learning for detecting inappropriate content in text. *Int J Data Sci Anal* 6(4):273–286
172. Yu C, Aoki PM, Woodruff A (2004) Detecting user engagement in everyday conversations. In: *Proceedings of 8th international conference on spoken language processing (ICSLP)*, pp 1329–1332
173. Yuan Z, Purver M (2015) Predicting emotion labels for chinese microblog texts. In: Gaber MM, Cosea M, Wiratunga N, Goker A (eds) *Advances in social media analysis*. Springer, Cham, pp 129–149
174. Zhang F, Xu H, Wang J, Sun X, Deng J (2016) Grasp the implicit features: hierarchical emotion classification based on topic model and SVM. In: *2016 International joint conference on neural networks (IJCNN)*, pp 3592–3599

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Nourah Alswaidan received a master's degree in computer science from King Saud University, Saudi Arabia, in 2014. She is currently a Ph.D. candidate in the Department of Computer Science at King Saud University. Her main interests include natural language processing, machine learning, and meta-heuristics.



Mohamed El Bachir Menai received a Ph.D. degree in computer science from Mentouri University of Constantine, Algeria, and University of Paris VIII, France, in 2005. He received also a “Habilitation Universitaire” in computer science from Mentouri University of Constantine, in 2007. (It is the highest academic qualification in Algeria, France, and Germany.) He is currently a professor in the department of computer science at King Saud University. His main interests include machine learning, natural language processing, satisfiability problems, evolutionary computing, and swarm intelligence.