# Text-Based Emotion Detection and Applications: A Literature Review

**Conference Paper** · October 2022

DOI: 10.1109/ICCR56254.2022.9995902



3 authors:

Mashary N. Alrasheedy
University of Hail


Ravie Chandren Muniyandi
Universiti Kebangsaan Malaysia


Fariza Fauzi
Universiti Kebangsaan Malaysia

# Text-Based Emotion Detection and Applications: A Literature Review

1st Mashary N. Alrasheedy
*Faculty of Information Science and Technology*
*Universiti Kebangsaan Malaysia*
Bangi, Malaysia
P105399@siswa.ukm.edu.my

2nd Ravie Chandren Muniyandi
*Faculty of Information Science and Technology*
*Universiti Kebangsaan Malaysia*
Bangi, Malaysia
ravie@ukm.edu.my

3rd Fariza Fauzi
*Faculty of Information Science and Technology*
*Universiti Kebangsaan Malaysia*
Bangi, Malaysia
fariza.fauzi@ukm.edu.my

***Abstract*—Emotion Detection is the sentiment analysis process used to extract emotions from the text that best represent the author's mental state. In recent years, the emotion detection domain has become very popular due to its potential applications in artificial intelligence, human-computer interaction, psychology, and marketing. Despite its vast application, emotion detection is challenging in natural language processing. Detecting emotions from a text or image requires exhaustive knowledge and analysis. However, with machine learning, artificial intelligence, and data mining advancement, it has become possible to face this challenge. Furthermore, the huge amount of textual data available online through social media, blogs, news, and articles helped the cause. For emotion detection, most of the studies have relied on machine learning and deep learning models and achieved good results. However, the researchers face some challenges that need to be addressed, such as the inability to extract the semantic information, the feature extraction process being time-consuming and inefficient, difficulty in identifying different emotions from non-standard language, imbalanced datasets, etc. This article aims to explore the existing approaches, methods, and evaluation measures used for emotion detection. The significant contribution, the methodology applied, and the results obtained by different researchers to gain the best possible results are also highlighted. Finally, the article highlights the limitations and provides the future direction that can be useful for research in emotion detection from text.**

***Index Terms*—Emotion Detection, Sentiment Analysis, Natural Language Processing, Machine Learning, Social Media**

## I. Introduction

The artificial intelligence (AI) and machine learning (ML) based model has been applied successfully in various problems. Natural language processing ((NLP) is the branch of ML that deals with human language generation and interpretation. However, it achieved success in solving language problems. However, the tasks are challenging to master because of the ambiguities in natural languages [1]. There are several subdomains of NLP, such as classification, entity or information extraction, document similarity matching, summarizing, or natural language generation. Classification domain comprises of several different tasks, like sentiment analysis [2], emotion classification [3]–[5], fake news detection [6], [7], author profiling [8]–[11], bots and gender [12], hate speech detection [13], etc. Sentiment analysis and emotion detection are used interchangeably but differ from each other. Sentiment analysis is a problem in NLP used to determine whether a piece of text is positive, negative, or neutral in terms of its emotional undertone. The goal of emotion categorization is to determine a statement's emotion by considering its underlying semantics. Emotions include anger, sadness, worry, etc.

In emotion detection, the problem is identifying the human emotions from a piece of text that best represents the author's mental state. Emotion detection has gained significant attention from researchers due to its potential applications in marketing, medicine, E-learning, etc. There is a number of classes for emotion detection task, including anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, trust, etc. [3], [4]. Emotions have their own language, play a key role in communication, and allow understanding of the context of communication. It happens often; we need to understand the emotions to grasp the communication. Emotions can be classified generally as positive, negative, and neutral. Positive emotions are expressed as happiness, excitement, and joy, while negative emotions include fear, sadness, and unhappy. Neutral means there is no emotion in text [14]. In the last decade, social media has been considered a powerful source of communication. People around the globe share their feelings, arguments, and opinions on different topics. Among the social media platforms, Facebook, Twitter, YouTube, and Instagram are the most commonly used platforms where people spend their time posting their expressions and emotions. Social media allows people to express their thoughts and opinions on various topics [15].

978-1-6654-6122-1/22/$31.00 ©2022 IEEE

### A. Application of Emotion Detection

Sentiment analysis and emotion detection have a variety of potential applications in numerous fields. For example, E-commerce companies use their platform feedback or reviews about the product and services. This user rating and reviews help the service provider to improve the quality of services and products according to user demands. Furthermore, text analysis helped the vendor make better future decisions [16]. When we talk about social media, many business organizations use these platforms, including Facebook, YouTube, and Twitter, to launch their products. This allows the organization to collect user feedback, measure customer satisfaction, and track its competitors. The sentiment analysis helped the market understand customer perspectives to make product changes [17].

Twitter is very popular in the healthcare sector for gaining information from healthcare professionals. One of the best examples is the recent COVID-19 pandemic, where patients were isolated from their families. This can cause harm to their mental health and raise problems of depression and anxiety in patients. In medicine, emotion classification can determine the feelings and comfort level of the patient towards the treatment. Moreover, it helped the practitioners to judge the patient's mental health using posts, and online blogging [18], [19].

Emotion analysis also has a key role in the education sector for both students and teachers. Emotion analysis can adjust the learning techniques in conformity with the learner. The feedback from students allows the teachers to improve their teaching skills. Correspondingly, it assists the organization in making the correct decision. Moreover, educational institutes use social media platforms like Facebook and Twitter for marketing. This allows the students and guardians to gain the required information about the educational institutes [20].

### B. Emotion Models

Emotion is an English word that means physical disruption. Before the 19th century, desire, hunger, and love were considered mental situations. Whereas in the 19th century, emotions were measured as a psychological word. In psychology, the complex feelings that change a person's actions, thoughts, and behavior are referred to as emotions [21]. The emotions models are categorized as dimensional and categorical models.

*1) Dimension Based Emotional Model:* This model characterizes emotions based on valence, provocation, and power. The valence indicates polarity, and provocation means the excitement of the feeling, while power signals the command over emotion [22]. The psychological states based on these parameters are illustrated in Fig. 1.

*2) Plutchik's Wheel of Emotion:* In the two-dimensional model, Fig. 2 presents Plutchik's wheel of emotion [23], which indicates the emotions in a wheel. The wheel is used for emotion in the concentric circle, the innermost circle as a variant of eight basic emotions: anger, anticipation, joy, trust, fear, surprise, sadness, and disgust. Furthermore, these eight basic emotions are an outermost area of the wheel and are combined with primary emotions.

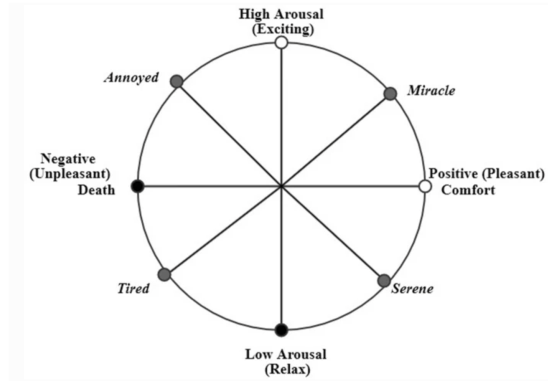


Fig. 1. The Dimension Emotion Model based on [22].

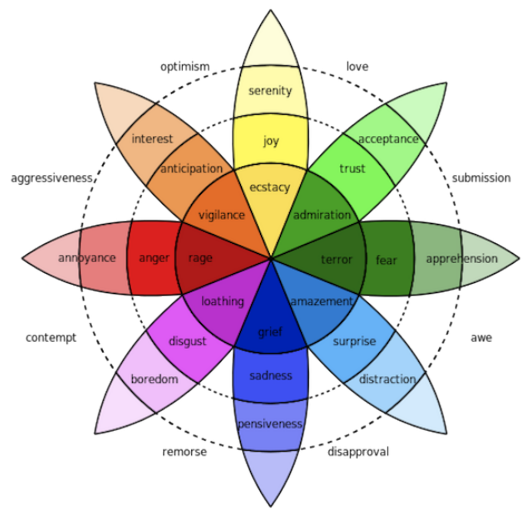


Fig. 2. The two dimensional Plutchik's Wheel of Emotions.

*3) Categorical Emotion Model:* In this type of model, the emotions are distinct as anger, joy, fear, and sadness. The model depends upon a particular category of four, six, and eight models. The four indicated the four basic emotions: positive, negative, relaxed, and exciting. At the same time, the six models included happiness, fear, surprise, anger, sadness, and disgust (a.k.a Paul Ekman's emotion model). The eight emotional models included surprise, anticipation, joy, sadness, anger, fear, trust, and disgust (a.k.a Plutchik's emotion model). Among these models, most of the researchers used Ekman and Plutchik [24], while the other models are Izard [25], [26], Shaver Emotion Model [27], and Tomkins (surprise, anguish, interest, fear, disgust, shame, anger, joy) [28]. Fig. 3 illustrates these emotion models.

This review paper aims to explore the concept of text-based emotion detection and methodologies used for emotion detection and highlights the challenges and future direction. The rest of the paper is organized as follows: section II describes the techniques and evaluation measures used for textual emotion detection. Section III explains the evaluation measures

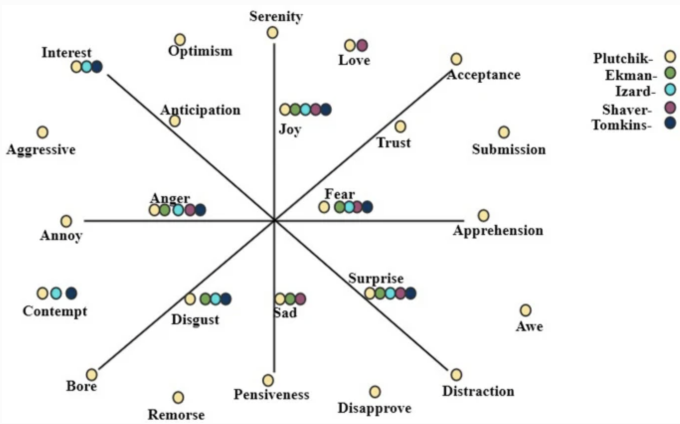


Fig. 3. The highlighted image of different categorical emotion model [22]

used for emotion detection. Section IV highlights researchers' challenges and limitations in emotion detection problems and discusses possible solutions. Section V concludes the paper.

## II. Approaches for Emotion Detection

Several methods are applied for emotion detection from text, including Lexicon, ML, deep learning, and hybrid-based techniques.

### *A. Lexicon-Based Emotion Detection Approaches*

The Lexicon-based approach used the keyword-built search method that searches for emotion keywords with some psychological state [29]. The most common techniques used for emotion detection are Word-Net-Affect [30] and National Research Council (NCR) Canada word-emotion lexicon [24]. The Word-Net-Affect is the extended form of WordNet with a list of effective words with emotion labels. At the same time, NRC Lexicon has 14000 words, where each word has a specific emotion and sentiment. These lexicons are categorized as each word that has a state of emotion. However, the main problem with this classification is that it ignores the intensity or concentration of emotions. It only measures the emotions but overlooks the supporting words that represent the intensity or level of emotions. Therefore these lexicons are less effective. To address this problem and maintain a more effective and informative lexicon, Li et al. [31] proposed an effective approach to gain the word-level emotion distribution that can assign the emotions with intensity by merging a dimensional dictionary called NRC-valence arousal dominance. Braun et al. [32] proposed a multilingual Emotional Football Corpus (MEmoFC). The corpus consists of football intelligence from English and German Web sites. The corpus has two metadata tables, one for explaining the match details and the other for the football abbreviations. Both are used to reference the outcome of football and the use of positive and negative emotions. The author also explains that corpus can be applied to various methods to influence the game outcomes.

### *B. Machine Learning-Based Emotion Detection Approaches*

Emotion detection from text is a classification problem that uses different domains of NLP and ML. ML can be classified as supervised, unsupervised, and reinforcement learning [33]. For emotion detection, mostly supervised and unsupervised ML methods are used for classification. Naïve Bayes (NB), support vector machine (SVM) and decision tree are the most important ML models. Hasan et al. [34] used the ML models, including Naïve Bayes, decision tree, and SVM for emotion detection from text data. The system was divided into two tasks: First, it collected the Twitter dataset using hashtags and trained the model. Second, it developed a two-stage EmoText Stream that was used to separate emotionless tweets and identify the emotions by using the trained model of step one. The SVM model achieved an accuracy of 90% for emotion detection. Suhasini and Badugu [35] used ML models to the Twitter dataset for emotion detection and applied Naïve Bayes and K-Nearest Neighbor (KNN) models. The Naïve Bayes achieved an accuracy of 72.59%, while KNN achieved 55.60% accuracy. Asghar et al. [36] also used multiple ML models over the ISEAR dataset to calculate the best models. The logistic regression (LR) model achieved a recall of 83% among all applied classifiers. Lee and Wang [37] proposed emotion detection from monolingual and bilingual text from English to Chinese and from Chinese to English. They applied a semi-supervised multi-view learning approach. They obtained 4195 posts from a famous Chinese social media platform, Weibo using Cohen Kappa Coefficient. Among 4000 posts, 2311 had the words for motions. They used code-switching text identification to separate Chinese from English. The model predicts emotions as happy, sad, fear, surprise, and anger with an F1 score of 0.48.

Allouch et al. [38] proposed an ML-based model that allowed disabled children to find the insult in different situations. The dataset has insulting and non-insulting sentences, a total of 1250 sentences. The ML models used for classification were the SVM, Tree Bagger, and Neural Network. The models achieved a recall of 80% and precision of 75%.

Wikarsa et al. [39] employed the NP methods for detecting emotions from Twitter using NB methods. It extracted 100 tweets using Twitter API. Tweets were preprocessed using stop words, removal of URLs, and conversion of emotions into text. The emotion was classified as Joy, Sadness, Fear, Disgust, Surprise, and Anger. For classifier performance measures, they used 10-fold cross-validation. The model achieved an accuracy of 83% for the Naïve Bayes classifier.

Amjad, Khan, and Chang [40] applied different ML classifiers for emotion detection and recognition on multiple datasets to compare different model performances. The experimented model included the SVM, Random Forest, Linear Regression, KNN, and multi-layer perception (MLP). The models were tested on BAUM, eNTERFACE05 for emotions including joy, surprise, anger, disgust, sadness, and fear. The results were compared using an accuracy measure. The SVM and MLP models have achieved the highest accuracy of 0.43 and 0.55

for joy while 0.32 and 0.48 for fear emotions, respectively.

Roman Urdu is widely used over social media for communication. As it is not considered a proper language, therefore, it has the limitation of dataset and emotions list. In order to detect emotion in Roman Urdu, Majeed, Mujtaba, and Beg [41] used machine learning models for emotion detection. Another challenge for the authors was translating the Roman Urdu words into English for emotion detection. The fear, love, anger, sadness, happy and neutral emotions were targeted for detection. The KNN, SVM, and random forest ML models were applied to the developed dataset. The SVM achieved the F1 score of 0.69, while KNN and random forest only achieved the F1 score of 0.59 and 0.63, respectively. However, the research opened the door for many researchers to work on emotion detection in low-resourced languages, like Roman Urdu.

Moreover, Ameer et al. [4] worked on a multi-label emotion classification problem on code-mixed (Roman Urdu and English) SMS messages. They applied classical ML–word n-grams (N = 1-3) and character n-grams (N = 3-10) (SVM, J48, Naive Bayes, bagging, decision tree, AdaBoost), deep learning (LSTM, Bi-LSTM, CNN, GRU, Bi-GRU, RNN, Bi-RNN), and transfer learning techniques (BERT, XLNet) on the newly developed dataset and compared the performance of all three approaches. Classical ML methods outperformed ML and deep learning models on proposed code-mixed (Roman Urdu and English) SMS messages. Word uni-gram yielded the best results (micro precision = 0.67, micro recall = 0.54, micro F1 = 0.67)

Mashal et al. [42] proposed a model that can extract fine-grained emotion intensities. The data was collected from 30 references and preprocessed by applying 70 references. After its construction, the feature vector can predict the four categories: Happy, sad, angry, and fear. The Linear Regression model was applied to predict the degree of emotions, but it has the issue of a limited number of categories to predict.

Nida et al. [43] developed an automatic emotion classifier for Twitter using emotion corpora. They extracted the features from the dataset individually and applied the WorldNet emotion dictionary for emotion synonyms. They found dependencies in a sentence and the context of the words. They trained the SVM classifier on the 6 Ekman categories of emotions. The model has achieved 60%, 63%, and 67.86% accuracy for three different emotion corpus they applied.

### *C. Deep Learning-Based Emotion Detection Approaches*

Deep learning is the branch of ML that uses the human learning process for learning purposes. The deep learning models have multiple layers of neurons. These neurons work together to extract hidden features from data that are important in emotion detection tasks to classify emotions from textual data correctly. Recently, deep learning has been applied in many domains to solve complex problems. The deep learning model artificial neural network (ANN), Convolution Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short Term Memory (LSTM) have gained great success due to their nonlinear problem-solving nature. These models have been applied in computer vision, image recognition, text classification, and other prediction problems [44]. Chatterjee et al. [45] developed a deep learning-based model called sentiment and semantic emotion detection (SSBED). The model used the two LSTM layers for feeding the sentiment and semantics. These illustrations are combined and forwarded to the main network for classification. The approach is based on the probability of multiple emotions in the sentences and uses semantics and sentiment for emotion classification, as underlying semantics are important in emotion classification, and some deep learning models lack there.

Krommyda et al. [46] used the Ekphrasis method, tokenization, spell correction, and normalization on a pre-trained word embedding model, GloVe. They applied the embedded layer for feature extraction and then passed them to the recurrent unit layer. The output of Bi-GRU was given to the attention layer, and output from the first and second Bi-GRU embedding layers was provided to DNN for classification. They used the ten emotions: anger, anticipation, disgust, joy, fear, surprise, sadness, and trust. Dashtipour [47] proposed a deep learning model that has been applied for sentiment investigation in the Persian language. Although identifying different emotions from non-standard language is challenging, the deep neural network models, including LSTM and CNN, has achieved great success and outperformed the existing ML models in term of accuracy and precision.

Abdullah et al. [48] used the CNN-LSTM-based deep learning model for emotion detection in Arabic tweets. They applied two different sets of experiments: Experiment 1 used the feed-forward network, and experiment 2 used the CNN-LSTM model. In experiment 1: They feed the model with 4900 input vectors using fully connected layers and three hidden layers. While in experiment 2, they used the input of 300 vectors into the CNN model. The model achieved an accuracy of 40% for experiment 1 and 605 for experiment 2. The models achieved reasonably good accuracy, but the problem is that systems were applied on a very small dataset.

Huang et al. [49] used a 4-part Episodal Memory Network (EMN) along self-attention model to identify the emotions from text data. They applied the SemEval dataset to four basic emotions: anger, fear, joy, and sadness. The EMN has achieved accuracy better than CNN, LSTM, and RNN models. The model achieved a precision of 56.8%, a recall of 63.1%, and an F1 score of 64.3%. The model was successful in detecting emotions, but the categories were limited.

Polignano et al. [50] proposed a model that used the Bi-LSTM, Self-Attention, and Convolution Neural Network (CNN) together for emotion detection. They found that word embedding is best for improving the performance of text-based emotion detection. Therefore, they compared three embedding systems: Google Word Embedding, GloVe Embedding, and FastText Embedding. They used the ISEAR, SemEval-2018, and SemEval-2019 datasets. The results showed that ISEAR has high performance, precision, and recall value. While GloVe embedding and FastText have emotion classes, including joy, guilt, anger, fear, shame, and disgust. On the other hand, SemEval-2018 and 2019 with FastText embedding gain higher precision, recall, and F1 measure. They concluded that FastText is an excellent embedding scheme for future research.

Ameer et al. [51] developed a multiple attention method that exposed the impact of each word on each emotion. For multi-label emotion classification, the researchers examined the application of LSTMs and the fine-tuning of Transformer Networks through TL, as well as a single-attention network and a multiple-attention network. The experimental findings demonstrated that pre-trained transformers used in transfer learning models, both with and without multiple attention processes, could outperform the state-of-the-art performance. On the difficult SemEval-2018 E-c: Detecting Emotions (multi-label classification) dataset for English, their top-performing RoBERTa-MA model outperformed the state-of-the-art. On the Ren-CECps dataset for Chinese, the XLNet-MA model fared better than the other suggested models.

Table II-C is a detailed analysis of the most recent literature presented in emotion detection. The table included the paper details, dataset information, method and algorithm applied for detection, the research objective, advantages and disadvantages of evaluation performance, and emotions detected that have been used in different emotion detection research models. More importantly, the table is arranged as the application domains. In our analysis, it has been discovered that social media is the prominent research area where research has been done for emotion detection. It is also observed that the emotion targeted by most of the researchers are happiness, anger, sadness, joy, fear, and disgust because these are most common emotion. Lastly, we have identified that ML and deep learning models have achieved good performances, whereas there is still room to improve.

## III. Evaluation Measures

The evaluation metrics are applied to measure the performance of the model. Several evaluation parameters can be used, including kappa Coefficient, Accuracy, Precision, Recall, Jaccard Accuracy, Pearson Correlation, and Chi-Square.

### A. Kappa Coefficient

The kappa Coefficient [62] is a statistical model used to calculate the inter-annotator reliability. It is used to assess the document to calculate the agreement between two annotators. The formula for Kappa is given below.

$$K = \frac{P_o - p_e}{1 - p_o} = 1 - \frac{1 - p_o}{1 - p_e} \tag{1}$$

### B. Jaccard Accuracy

The Jaccard Accuracy [63] is the size of the intersection divided by the size of two labeled sets. It is used to compare predictions for a sample with respect to a set of labels in the original value. It also ranges from 0 to 1. The formula for Jaccard Accuracy is given as:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{2}$$

### C. Precision

The precision [38] denoted by (P) is called the number of True positives (TP) divided by true positives and false positives (FP). The following formula describes the precision:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

### D. Recall

The Recall (Wang, 2015) is called as number of True Positive divided by a total number of true positive and false negative. Mathematically Recall is described as:

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

Where R is the value of Recall, TP is the number of True Positive, and FN is the false negative value.

### E. F-Score

The F-Score [64] is the measurement used to provide balance between precision and recall. The value of F-Score varies between 0 and 1. The value of F-Score can be calculate as follows:

$$F1 = 2 * \frac{P * R}{P + R} \tag{5}$$

Where P is precision and R is the value of recall.

### F. Accuracy

The accuracy is the measurement of the classification of models. The accuracy is calculated by using the following formula:

$$Accuracy = \frac{\text{Number of correct Predication}}{\text{Total Number of predication}}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

### G. Pearson Correlation

Pearson Correlation [65] is a statistical model that indicates the relationship or an association. It is between two variables and divided by the product of their standard deviation. The following equation is used to calculate the Pearson Correlation:

$$r = \frac{\sum \left(x_i - \bar{X}\right)\left(y_i - \bar{Y}\right)}{\sqrt{\sum \left(x_i - \bar{X}\right)^2 \left(y_i - \bar{Y}\right)^2}} \tag{7}$$

Where r is the value of Pearson correlation and $x_i$ is the value of x is the mean value of x. The $Y_i$ is the value of y variable and y is the mean value of y-variable.

## IV. Challenges and Improvements

We have explored the literature and identified the major challenges and limitation of the studies for emotion detection.

TABLE I
STUDIES ON EMOTION DETECTION

| Paper | Dataset (open source) | Model | Objective | Advantages | Disadvantages | Evaluation Measure | Emotions Detected |
|---|---|---|---|---|---|---|---|
| Joan [52] | SemEval-2019 | ML models Chi-square POS tagger, SVM | Identifying the Human Emotion from text data | Enhance Performance Solved Semantic Extraction Problem | Ignore intensity of emotions. | Accuracy 72.43% | Joy, Sadness, Fear, Surprise, Disgust |
| Choudrie and Dwivedi [53] | CrowdFlower | Deep learning-based model | Identify emotions of people during COVID-19 | Performance | Limited dataset | Accuracy 80.20% | Joy, Fear, Anxiety, Happiness, Sadness |
| Allouch et al. [38] | Self-Created | Naive Bayes SVM Decision Tree | Insult detection for children | SVM highest accuracy | Ignored Cintextual Information | Accuracy 76.2% | Insulting, Non Insulting, Not Clear |
| Suhasini and Srinivasu [35] | Twitter Sentiment | ML Models NB KNN | Detect emotion based on tweets | NB has gain higher accuracy from KNN | No Contextual Information applied | Accuracy 72.5% | Happy, Normal, Unhappy |
| De Bruyne et al. [54] | SemEval-2018 | ML Models SVM Logistic Regression RF | Multi-level emotion classification for tweets | Very large list of features applied | Unbalanced data | Jaccard Accuracy 52% | anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust |
| Baziotis et al. [55] | SemEval-2018 | Deep Learning Bi-LSTM | Emotion Detection in Tweets | Spelling Correction Performance | Problem for out of vocabulary words | Accuracy 58.40% | anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust |
| Ezen-Can and Ethem [56] | SemEval-2018 | Deep Learning Bi-GRU | Various classification for Tweet post | emoji embedding | No application of affected lexicons | Jaccard Accuracy 40.00% | anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust |
| Rathnayaka et al. [57] | SemEval-2018 | Deep Learning Bi-GRU Attention Network | Emotion detection in microblogs | Multiple emotion in a single text | Low recognition for some classes | (F-Score) 58.00% | anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust |
| Krommyda [58] | Tweets Annotated to emotions | Deep Learning ML LSTM SVM | Emotion Detected in online Social Media Chat | Fully annotated dataset created for Tweets | Trust Emotion have Low Accuracy | Accuracy 91.9% | anger, anticipation, joy, trust, fear, surprise, sadness and disgust |
| Kratzwald et al. [59] | Year, Twitter | ML Models Random Forest SVM Bi-LSTM Deep Learning Model | Emotion Detected in decision support system | Achieved performance over conventional models | Small dataset | F1 Score 23.2% | joy, anger, sadness, trust, surprise, anticipation |
| Kolekar [60] | Customized Dataset from shopping portal | ML with Bags of Words | Identify sentiments in product review | Addressed polarity shift problem | Used small and customised dataset | | negative, positive, neutral |
| Shrivastava et al. [61] | TV Transcript | Deep Learning CNN Max Pooling | Emotion detection in multimedia text | Context Features Applications | Overlap between classes like anger and disgust | F1-Score 72.48% | anger, disgust, fear, happiness, sadness, surprise |

| Paper | Dataset (open source) | Model | Objective | Advantages | Disadvantages | Evaluation Measure | Emotions Detected |
|---|---|---|---|---|---|---|---|
| Ameer et al. [4] | CM-MEC-21 | ML, deep learning, transfer learning | Emotion Classification | Multi-label Emotion Classification in SMS Messages | Poor performance of ML and transfer learning | Micro F1 67% | anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust |
| Ameer et al. [51] | SemEval-2018 and Ren-CECps | transfer learning with multiple attention mechanism | Emotion Classification | Multi-label Emotion Classification in Tweets | not able to model the relationships between phrases and classes | Micro F1 62.4% | anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust |

## A. Challenges

**Unable to Extract the Semantic Information:** Sometimes, when we write a sentence, there are different negation signs used for emotional representation. However, detecting the emotion from that text may cause ambiguity because words have different contexts with different emotions. Therefore, these issues need to be identified to gain the correct emotion detection from text data.

**Feature Extraction Process is Time-Consuming and Inefficient:** As we have seen, ML and deep learning models have achieved great success in emotion detection on textual data. However, most ML models need extracted features for efficient emotion detection. However, the manual feature extraction process is a time-consuming, challenging, and error-prone task. Sometimes mislabeling emotions can occur during the labeling process as it is a very brainstorming task. Therefore, for the ML model to work efficiently, there should be an automated system for feature extraction.

**Emotions Classification with Their Intensities Level:** The written text has words that can indicate the emotion in a text. However, limited words can be associated with the degree of emotion or sentiments. While detecting the emotion, the intensity of words or sentences can help identify the intensity in textual data. So, these words help in identifying the correct emotion with their intensity level.

**Identifying Different Emotions from Non-standard Language:** Informal words, slang, misspelled words, hashtags, abbreviations, etc., are widely used on social media platforms for emotional expression as these words are non-standard words. Therefore, it is very challenging for models to perceive such words for emotion detection.

**Existence Models Failed to Perform:** Most text-based emotion detection models used ML and deep learning models for classification. However, the problem with these models is that they require annotated dataset. Preparing such a dataset requires time, resources, and human efficiency. Also, applying deep learning models need to process knowledge. While on the other hand, these models require a large amount of data to improve their performance.

**Imbalanced Datasets:** Very few datasets are available online and can be used for research as most online datasets have imbalanced instances. So, these datasets cannot help ML and deep learning to achieve the required accuracy goal.

## B. Possible Improvements

This study focused on the detection of emotion from text data. So, some future directions can help the researchers improve existing technologies' performance. In this section, we proposed some future directions to the above-explained challenges.

**Domain Adaptation and Transfer Learning:** As we have explained earlier, the existing dataset has limited labeled data and has domain dependency and imbalanced datasets. The solution to these problems is domain adaptation and transfer learning. The deep learning models are trained in one domain and tested in another using the domain adaptation methodology. At the same time, semi-supervised learning is another solution to the problem.

**Ensemble Model Models:** The deep learning models are very impressive for emotion detection but still suffer to perform well. The Ensemble model is the solution for deep learning model problems. Both models combine together to improve performance and accuracy. The Ensemble model combines the prediction from the neural network for better performance. Likewise, the performance can be enhanced by applying the model to large and domain-related datasets.

**Pre-trained Word Embedding:** As explained above, using informal language, including sarcasm, irony, hashtags, and misspelled words, is another challenge for the model to perform well. Applying pre-trained word embedding can counter the problem.

**Graph Neural Network:** Another problem was the meaning of the text being classified. In this scenario, the Graph Neural Network can be applied to understand the text semantically to enhance the results of the text classification problem.

# V. Conclusion

This review article provides significant insight into textual emotion detection and shed some light on existing techniques and evaluation measures used to detect emotion from a piece of text. The study highlighted the challenges and limitations of the literature on emotion detection tasks. Moreover, it provided possible solutions to tackle these challenges to improve the performance of emotion detection systems. The study concluded that ML and deep learning models had gained great accuracy over conventional rule-based approaches.

## ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support of the Laboratory of Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Malaysia and University of Hail, Kingdom of Saudi Arabia.

## REFERENCES


[1] K. Chowdhary, "Natural language processing," *Fundamentals of artificial intelligence*, pp. 603–649, 2020.

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

[3] I. Ameer, N. Ashraf, G. Sidorov, and H. Gómez Adorno, "Multi-label emotion classification using content-based features in twitter," *Computación y Sistemas*, vol. 24, no. 3, pp. 1159–1164, 2020.

[4] I. Ameer, G. Sidorov, H. Gomez Adorno, and R. Nawab, "Multi-label emotion classification on code-mixed text: Data and methods," *IEEE Access*, vol. 10, pp. 8779–8789, 01 2022.

[5] S. Mohammad, F. Bravo-Marquez, M. Salameh, and S. Kiritchenko, "Semeval-2018 task 1: Affect in tweets," in *Proceedings of the 12th international workshop on semantic evaluation*, 2018, pp. 1–17.

[6] I. Ameer, C. P. Capetillo, H. Gómez-Adorno, and G. Sidorov, "Automatic fake news detection in urdu language using transformers," in *CEUR Workshop Proceedings*, vol. 3159. CEUR-WS, 2021, pp. 1127–1134.

[7] M. Arif, A. L. Tonja, I. Ameer, O. Kolesnikova, A. Gelbukh, G. Sidorov, and A. G. M. Meque, "Cic at checkthat! 2022: multi-class and cross-lingual fake news detection," *Working Notes of CLEF*, 2022.

[8] I. Pervaz, I. Ameer, A. Sittar, and R. M. A. Nawab, "Identification of author personality traits using stylistic features: Notebook for pan at clef 2015." in *CLEF (Working Notes)*. Citeseer, 2015, pp. 1–7.

[9] A. Sittar and I. Ameer, "Multi-lingual author profiling using stylistic features." in *FIRE (Working Notes)*, 2018, pp. 240–246.

[10] I. Ameer, G. Sidorov, and R. M. A. Nawab, "Author profiling for age and gender using combinations of features of various types," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 5, pp. 4833–4843, 2019.

[11] I. Ameer and G. Sidorov, "Author profiling using texts in social networks," in *Handbook of Research on Natural Language Processing and Smart Service Systems*. IGI Global, 2021, pp. 245–265.

[12] M. H. F. Siddiqui, I. Ameer, A. F. Gelbukh, and G. Sidorov, "Bots and gender profiling on twitter." in *CLEF (Working Notes)*, 2019.

[13] I. Ameer, M. H. F. Siddiqui, G. Sidorov, and A. Gelbukh, "Cic at semeval-2019 task 5: Simple yet very efficient approach to hate speech detection, aggressive behavior detection, and target classification in twitter," in *Proceedings of the 13th International Workshop on Semantic Evaluation*, 2019, pp. 382–386.

[14] S. Volet, C. Seghezzi, and S. Ritchie, "Positive emotions in student-led collaborative science activities: Relating types and sources of emotions to engagement in learning," *Studies in higher education*, vol. 44, no. 10, pp. 1734–1746, 2019.

[15] C. Malighetti, S. Sciara, A. Chirico, and G. Riva, "Emotional expression of# body on instagram," *Social Media+ Society*, vol. 6, no. 2, p. 2056305120924771, 2020.

[16] S. Al Ajrawi, A. Agrawal, H. Mangal, K. Putluri, B. Reid, G. Hanna, and M. Sarkar, "Evaluating business yelp's star ratings using sentiment analysis," *Materials Today: Proceedings*, 2021.

[17] Z. Drus and H. Khalid, "Sentiment analysis in social media and its application: Systematic literature review," *Procedia Computer Science*, vol. 161, pp. 707–714, 2019.

[18] K. Garcia and L. Berton, "Topic detection and sentiment analysis in twitter content related to covid-19 from brazil and the usa," *Applied soft computing*, vol. 101, p. 107057, 2021.

[19] I. Ameer, M. Arif, G. Sidorov, H. Gòmez-Adorno, and A. Gelbukh, "Mental illness classification on social media texts using deep learning and transfer learning," *arXiv preprint arXiv:2207.01012*, 2022.

[20] P. Bhardwaj, P. Gupta, H. Panwar, M. K. Siddiqui, R. Morales-Menendez, and A. Bhaik, "Application of deep learning on student engagement in e-learning environments," *Computers & Electrical Engineering*, vol. 93, p. 107277, 2021.

[21] T. Dixon, ""emotion": The history of a keyword in crisis," *Emotion Review*, vol. 4, no. 4, pp. 338–344, 2012.

[22] I. Bakker, T. Van Der Voordt, P. Vink, and J. De Boon, "Pleasure, arousal, dominance: Mehrabian and russell revisited," *Current Psychology*, vol. 33, no. 3, pp. 405–421, 2014.

[23] R. Plutchik, "A general psychoevolutionary theory of emotion," in *Theories of emotion*. Elsevier, 1980, pp. 3–33.

[24] S. Mohammad and P. Turney, "Nrc emotion lexicon, national researc h council of canada," Technical Report, Tech. Rep., 2013.

[25] W. S. Ark, D. C. Dryer, and D. J. Lu, "The emotion mouse." in *HCI (1)*, 1999, pp. 818–823.

[26] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[27] K. W. Fischer, P. R. Shaver, and P. Carnochan, "How emotions develop and how they organise development," *Cognition and emotion*, vol. 4, no. 2, pp. 81–127, 1990.

[28] S. Tomkins, *Affect imagery consciousness: Volume I: The positive affects*. Springer publishing company, 1962.

[29] S. Zad, M. Heidari, H. James Jr, and O. Uzuner, "Emotion detection of textual data: An interdisciplinary survey," in *2021 IEEE World AI IoT Congress (AIIoT)*. IEEE, 2021, pp. 0255–0261.

[30] R. C. Balabantaray, M. Mohammad, and N. Sharma, "Multi-class twitter emotion classification: A new approach," *International Journal of Applied Information Systems*, vol. 4, no. 1, pp. 48–53, 2012.

[31] Z. Li, H. Xie, G. Cheng, and Q. Li, "Word-level emotion distribution with two schemas for short text emotion classification," *Knowledge-Based Systems*, vol. 227, p. 107163, 2021.

[32] N. Braun, C. van der Lee, L. Gatti, M. Goudbeek, and E. Krahmer, "Memofc: introducing the multilingual emotional football corpus," *Language resources and evaluation*, vol. 55, no. 2, pp. 389–430, 2021.

[33] W. Khan, S. Ali, U. K. Muhammad, M. Jawad, M. Ali, and R. Nawaz, "Adadiffgrad: An adaptive batch size implementation technique for diffgrad optimization method." in *2020 14th International Conference on Innovations in Information Technology (IIT)*. IEEE, 2020, pp. 209–214.

[34] M. Hasan, E. Rundensteiner, and E. Agu, "Automatic emotion detection in text streams by analyzing twitter data," *International Journal of Data Science and Analytics*, vol. 7, no. 1, pp. 35–51, 2019.

[35] M. Suhasini and B. Srinivasu, "Emotion detection framework for twitter data using supervised classifiers," in *Data Engineering and Communication Technology*. Springer, 2020, pp. 565–576.

[36] M. Z. Asghar, F. Subhan, M. Imran, F. M. Kundi, S. Shamshirband, A. Mosavi, P. Csiba, and A. R. Várkonyi-Kóczy, "Performance evaluation of supervised machine learning techniques for efficient detection of emotions from online content," *arXiv preprint arXiv:1908.01587*, 2019.

[37] S. Y. M. Lee and Z. Wang, "Multi-view learning for emotion detection in code-switching texts," in *2015 International Conference on Asian Language Processing (IALP)*. IEEE, 2015, pp. 90–93.

[38] M. Allouch, A. Azaria, R. Azoulay, E. Ben-Izchak, M. Zwilling, and D. A. Zachor, "Automatic detection of insulting sentences in conversation," in *2018 IEEE International Conference on the Science of Electrical Engineering in Israel (ICSEE)*. IEEE, 2018, pp. 1–4.

[39] L. Wikarsa and S. N. Thahir, "A text mining application of emotion classifications of twitter's users using naive bayes method," in *2015 1st International Conference on Wireless and Telematics (ICWT)*. IEEE, 2015, pp. 1–6.

[40] A. Amjad, L. Khan, and H.-T. Chang, "Semi-natural and spontaneous speech recognition using deep neural networks with hybrid features unification," *Processes*, vol. 9, no. 12, p. 2286, 2021.

[41] A. Majeed, H. Mujtaba, and M. O. Beg, "Emotion detection in roman urdu text using machine learning," in *Proceedings of the 35th IEEE/ACM international conference on automated software engineering workshops*, 2020, pp. 125–130.

[42] S. X. Mashal and K. Asnani, "Emotion intensity detection for social media data," in *2017 International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2017, pp. 155–158.

[43] H. Nida, K. Mahira, M. Mudasir, M. Mudasir Ahmed, and M. Mohsin, "Automatic emotion classifier," in *Progress in Advanced Computing and Intelligent Engineering*. Springer, 2019, pp. 565–572.

[44] F. Murat, O. Yildirim, M. Talo, U. B. Baloglu, Y. Demir, and U. R. Acharya, "Application of deep learning techniques for heartbeats detection using ecg signals-analysis and review," *Computers in biology and medicine*, vol. 120, p. 103726, 2020.
[45] A. Chatterjee, U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, and P. Agrawal, "Understanding emotions in text using deep learning and big data," *Computers in Human Behavior*, vol. 93, pp. 309–317, 2019.
[46] M. Krommyda, A. Rigos, K. Bouklas, and A. Amditis, "An experimental analysis of data annotation methodologies for emotion detection in short text posted on social media," in *Informatics*, vol. 8, no. 1. MDPI, 2021, p. 19.
[47] K. Dashtipour, M. Gogate, J. Li, F. Jiang, B. Kong, and A. Hussain, "A hybrid persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks," *Neurocomputing*, vol. 380, pp. 1–10, 2020.
[48] M. Abdullah, M. Hadzikadicy, and S. Shaikhz, "Sedat: sentiment and emotion detection in arabic text using cnn-lstm deep learning," in *2018 17th IEEE international conference on machine learning and applications (ICMLA)*. IEEE, 2018, pp. 835–840.
[49] J. Huang, Z. Lin, and X. Liu, "Episodic memory network with self-attention for emotion detection," in *International Conference on Database Systems for Advanced Applications*. Springer, 2019, pp. 220–224.
[50] M. Polignano, P. Basile, M. De Gemmis, G. Semeraro, and V. Basile, "Alberto: Italian bert language understanding model for nlp challenging tasks based on tweets," in *6th Italian Conference on Computational Linguistics, CLiC-it 2019*, vol. 2481. CEUR, 2019, pp. 1–6.
[51] I. Ameer, N. Bölücü, M. H. F. Siddiqui, B. Can, G. Sidorov, and A. Gelbukh, "Multi-label emotion classification in texts using transfer learning," *Expert Systems with Applications*, p. 118534, 2022.
[52] J. Xiao, "Figure eight at SemEval-2019 task 3: Ensemble of transfer learning methods for contextual emotion detection," in *Proceedings of the 13th International Workshop on Semantic Evaluation*. Minneapolis, Minnesota, USA: Association for Computational Linguistics, Jun. 2019, pp. 220–224. [Online]. Available: https://aclanthology.org/S19-2036
[53] J. Choudrie and Y. K. Dwivedi, "Investigating factors influencing adoption of broadband in the household." *Journal of Computer Information Systems*, vol. 46, no. 4, 2006.
[54] L. De Bruyne, O. De Clercq, and V. Hoste, "Lt3 at semeval-2018 task 1: A classifier chain to detect emotions in tweets," in *Proceedings of The 12th International Workshop on Semantic Evaluation*, 2018, pp. 123–127.
[55] C. Baziotis, N. Pelekis, and C. Doulkeridis, "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)*, 2017, pp. 747–754.
[56] A. Ezen-Can and E. F. Can, "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*, 2018, pp. 162–166.
[57] P. Rathnayaka, S. Abeysinghe, C. Samarajeewa, I. Manchanayake, M. J. Walpola, R. Nawaratne, T. Bandaragoda, and D. Alahakoon, "Gated recurrent neural network approach for multilabel emotion detection in microblogs," *arXiv preprint arXiv:1907.07653*, 2019.
[58] M. Krommyda, A. Rigos, K. Bouklas, and A. Amditis, "An experimental analysis of data annotation methodologies for emotion detection in short text posted on social media," *Informatics*, vol. 8, no. 1, 2021. [Online]. Available: https://www.mdpi.com/2227-9709/8/1/19
[59] B. Kratzwald, S. Ilić, M. Kraus, S. Feuerriegel, and H. Prendinger, "Deep learning for affective computing: Text-based emotion recognition in decision support," *Decision Support Systems*, vol. 115, pp. 24–35, 2018.
[60] N. V. Kolekar, G. Rao, S. Dey, M. Mane, V. Jadhav, and S. Patil, "Sentiment analysis and classification using lexicon-based approach and addressing polarity shift problem," *Journal of Theoretical and Applied Information Technology*, vol. 90, no. 1, p. 118, 2016.
[61] K. Shrivastava, S. Kumar, and D. K. Jain, "An effective approach for emotion detection in multimedia text data using sequence based convolutional neural network," *Multimedia tools and applications*, vol. 78, no. 20, pp. 29 607–29 639, 2019.
[62] S.-H. Park, B.-C. Bae, and Y.-G. Cheong, "Emotion recognition from text stories using an emotion embedding model," in *2020 IEEE international conference on big data and smart computing (BigComp)*. IEEE, 2020, pp. 579–583.
[63] X. Zhang, W. Li, H. Ying, F. Li, S. Tang, and S. Lu, "Emotion detection in online social networks: a multilabel learning approach," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8133–8143, 2020.
[64] D. G. Haryadi, J. A. Orr, K. Kuck, S. McJames, and D. R. Westenskow, "Partial co2 rebreathing indirect fick technique for non-invasive measurement of cardiac output." *Journal of clinical monitoring and computing*, vol. 16, no. 5-6, pp. 361–74, 2000.
[65] C. Strapparava and A. Valitutti, "Wordnet-affect: an affective extension of wordnet," in *Proceedings of LREC*, vol. 4, 2004, pp. 1083–1086.