Different Approach for Emotion Detection in Text:

A Systematic Review

Desi Dwi Purwantia), Ridi Ferdianab), and Teguh Bharata Adjic)

Department of Electrical and Information Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia

a) Corresponding author: [desi.dwi.p@mail.ugm.ac.id](mailto:desi.dwi.p@mail.ugm.ac.id)

b) [ridi@ugm.ac.id](mailto:ridi@ugm.ac.id)

c) [adji@ugm.ac.id](mailto:adji@ugm.ac.id)

**Abstract.**  Emotion detection in text has become a crucial area of research due to its wide-ranging applications in fields such as customer service, healthcare, marketing, etc. This article explores the significance of advancing research in text-based emotion detection, emphasizing its potential to revolutionize human-computer interaction and enhance user experience in digital environments. Text data sources that are widely available on the internet can be a source of data to find various information about feelings or emotions expressed through text. We discuss methodologies and challenges associated with accurately identifying emotions from textual data, highlighting the necessity for robust computational models capable of discerning subtle emotional nuances. This study aims to offer insights into recent advancements in text-based emotion detection while underscoring the necessity for continued research in this domain. Additionally, it proposes future research directions focused on enhancing the reliability and interpretability of text-based emotion detection systems to fully harness the potential of this innovative technology.

**Keywords:** emotion detection, text classification, emotion recognition, systematic literature review.

# INTRODUCTION

The ability to process emotions is important in a computer system. Recognizing user emotions can improve the human-computer interaction's quality [1]. Furthermore, the intelligence of computer systems can be increased by adding emotional factors in it. Before the computer can respond according to the emotions of the user, the detection process must first be detected from voice, facial expressions, gestures and text. Text is one of the media used to communicate and convey information. Not only contains information, text can also express emotions [2]. Text is also the main medium of communication using computers (Computer-Mediated Communication) such as email, blogs and social media.

Emotion detection from text finds applications across various fields such as sentiment analysis, verbal expression in human-computer interaction, consumer analysis, online learning and computer games. Some of these things show the importance of detecting emotions from texts. With the widespread use of digital communication platforms and the growing dependence on automated systems for analyzing sentiment, understanding and interpreting emotional nuances in textual data has become both a necessity and a challenge.

Many difficulties arise when attempting to discern emotions in text. Firstly, automated systems find it challenging to reliably detect and categorize human emotions due to the subjectivity and ambiguity of these feelings. The endeavor is made more difficult by linguistic variations, meanings that vary depending on context, and cultural variances. Furthermore, it can be difficult for traditional techniques to identify differences between comparable emotions and to capture subtle emotional cues. The fact that language is varied and always changing adds to these difficulties and might cause problems for emotion detection models regarding their ability to scale and and generalizability.

The majority of earlier research on emotion detection has concentrated on using deep learning and machine learning methods to categorize emotions in text. Sentiment analysis, feature engineering, and lexicon-based techniques have established the foundation for comprehending emotional content. These techniques, however, frequently rely on static models and predetermined emotion categories, which might not take into consideration the dynamic and context-sensitive character of human emotions. Although there has been progress in pre-trained language models and contextual embeddings, problems remain with data sparsity, interpretability of the models, and emotional granularity. Researchers have also emphasized the need for more flexible and all-encompassing frameworks that can better manage the subtle differences in emotional expression between settings and languages.

Various methods have been suggested to address the problem of how to identify emotions in textual data. There are 3 main approaches in detecting emotion from text, namely keyword-spotting, statistical, and ruled-based [3]. The keyword-spotting approach works by finding keywords that relate to a certain type of emotion. While the statistical approach involves large amounts of data (corpus) for training the emotion detection model. The data (corpus) is also used to create a number of rules in the ruled-based approach.

This paper’s contribution is to provide an overview of the existing approaches that have been used in detecting emotions from a text. This research article focuses on tackling the challenges and constraints involved in detecting emotions from text. It offers a thorough review of current methods and suggests new approaches to move the field forward. By analyzing the strengths and weaknesses of existing techniques, the study aims to highlight ways to improve them. Furthermore, it proposes a new framework designed to integrate recent advancements, enhancing the accuracy, flexibility, and interpretability of emotion detection systems. The goal is to advance the field with a more profound insight into how emotions are expressed and provide practical solutions to improve emotion detection across various textual scenarios.

# METHODS

Articles discussing the emotion detection system are dispersed across multiple journals. The following sections outline our approach to retrieving these articles, as well as our article selection criteria and filtering process.

## Research Questions

Given the wide reach and significant influence of previous research on emotion detection in text, a systematic literature review is necessary to address these research questions.

**RQ1** : What datasets are used for these methods?

**RQ2** : What methods are used for emotion detection in text and what are their effectiveness in detecting and classifying emotions in text?

**RQ3** : What metrics are used to evaluate these methods?

## Search Strategy

A thorough search was conducted for relevant research articles to address the research questions, utilizing searches based on titles, abstracts, and keywords. We combined identified search terms in different ways to locate relevant articles within our chosen database. Our search process followed procedural steps using ‘AND’ and ‘OR’ operators. We used the same search syntax and keywords across all databases.

(Emotion Detection **OR** Emotion Recognition **OR** Emotion Classification) **AND** (in Text **OR** Text Based **OR** from Text) **AND/OR** (Chatbot Application **OR** Academic Chatbot)

The search strategy was determined by the following decisions:

Search database : IEEE Explore, Science Direct, Springer Link, ACM Digital Directory.

Search items : Journal articles and conference papers

Search applied on : Full text

Publication period : 2019 to 2024.

## Inclusion and Exclusion Criteria

Based on the search strategy outlined earlier, a variety of document types were identified and reviewed. Non-relevant articles were excluded after analysing their abstracts, reducing the number of potential papers. During the process of selecting relevant articles, we consistently applied specific criteria for inclusion and exclusion.

The following are the inclusion criteria applied:

1. Articles written in English
2. Articles published between 2019 and 2024
3. Availability of the full text for review
4. Research focused on detecting emotions from text or textual data.

The following are the exclusion criteria applied:

1. Duplicate or redundant research.
2. Studies that do not address the research questions.
3. Research not primarily focused on emotion detection from textual data, including those focusing on other data forms such as images or audio.

## RESULT AND DISCUSSION

This section presents the results derived from the research process related to emotion detection in text. Only the most important articles were selected after careful consideration of these articles. A total of 20 articles constitutes the final dataset for the present study. The research findings are presented in relation to the research questions.

## RQ1: What datasets are used for these methods?

This section covers a discussion of the corpus and lexicon used for text-based emotion detection in related research.

1. Corpus

The methods of emotion detection depend on the availability of data or corpus. A corpus is a collection of text data that has been labelled with one or more types of emotions. The corpus is used as data to train the detection model on machine learning-based methods and also used to create a number of rules in the ruled-based approach. Meanwhile, in the lexicon-based method, detection is carried out based on a word dictionary or lexicon. Generally, the emotion lexicon only works well for a particular domain. Therefore, the emotional corpus can be used to develop a lexicon for a particular domain. Availability of corpus becomes very important in detecting emotion from text. In this research, an emotion labelled corpus is presented for evaluation of emotions in text. **TABLE 1** describes the different corpus available in the research.

**TABLE 1.** List of Corpus

|  |  |
| --- | --- |
| **Datasets** | Features |
| Aman [4] | Blog post |
| Amazon Alexa Reviews Emotions [5] | Alexa reviews |
| Alm [6] | 185 children’s stories |
| Cecilia Ovesdotter Alm's Affect data [7] | Stories |
| CrowdFlower [8] | Tweets |
| DailyDialog [9] | Dialogues |
| EmoBank [10] | Blogs, newspapers, news headlines, letters and travel guide |
| Emotion Lines | Facebook message chats and Television show |
| Emotion Stimulus [11] | Designed from FrameNets |
| Grounded emotions [12] | Tweets |
| MELD data [13] | Television show dialogues |
| SemEval-2007 [14] | News headlines |
| SemEval-2017 [15] | Social media and news |
| SemEval-2018 [16] | Tweets |
| SemEval-2019 [17] | Text dialogue between 2 individuals |
| SMILE dataset [18] | Tweets |
| The Valence and Arousal Dataset [19] | Facebook Post |
| WASSA-2017 Emotion Intensities [20] | Tweets |

1. Lexicons

Most of the emotion detection methods in the text use the emotion lexicon. The lexicon comprises a collection of words (dictionaries) and their corresponding emotional categories. Various emotional lexicons have been created and used to detect emotions, as outlined in **TABLE 2** below.

**TABLE 2.** List of Lexicons

|  |  |
| --- | --- |
| **Lexicon** | **Descriptions** |
| AFINN [21] | Manually graded words for valence in integers between -5 and +5. |
| Bing Liu Lexicon [22] | List of sentiment (positive and negative opinion) words. |
| EmoSenticNet [23] | Emotion labels to Sentic Concept |
| NRC Hashtag Emotion Lexicon [24] [25] | Tweets containing hashtags with emotional terms. It is especially related to words like fear, anger, surprise, trust and happiness. |
| NRC Hashtag Sentiment Lexicon [26] | Tweets that include sentimental hashtags like #amazing. It is associated with negative or positive sentiments. |
| NRC Word-Emotion Association Lexicon [27] [28] | Use Amazon Truck Mechanic, Lexicon Annotated Manual. Consists of 8 emotions, includes 2 negative and positive sentiments. |
| Sentiment140 Lexicon [26] | Tweets containing emoji. |
| WordNet [29] | Online English lexicon database. It includes nouns, verbs, adverbs and adjectives in a set of synonyms. |

## RQ2: What methods are used for emotion detection in text and what are their effectiveness in detecting and classifying emotions in text?

This section highlights the various approaches to detecting emotion from texts and summarizes the strengths and weaknesses of each in depth.

1. Keyword-based Approach

This method is straightforward to execute and intuitive since it revolves around identifying specific words within text. The challenge of keyword pattern matching involves locating instances where a keyword from a specified set appears as a substring within a given string.

An emotion recognition method based on keywords involves taking a text document as input and producing an output representing an emotion category. This approach comprises five sequential steps. Initially, the text data is tokenized, followed by the identification and detection of emotion-related words from these tokens. The subsequent steps involve marking the input text, analysing the intensity of the identified emotional words, and checking for negation within sentences. Finally, the method determines the emotion class as the desired output.

1. Corpus-based Approach

Approaches based on corpuses use supervised learning to acquire information from sources such as annotated or semi-annotated emotional lexicons within a corpus. These methods involve extracting emotions based on predefined emotion theories, leveraging semantic and syntactic patterns in text, such as those found in Wikipedia. There is also a growing emphasis on lexicons driven by extensive research in sentiment analysis.

Del Arco et al. (2020) [29] used a multilingual twitter dataset. It includes 8409 datasets labelled in Spanish and 7303 datasets in English from tweets. They used machine learning to identify emotions then evaluated the methods using 10-fold cross validation. They achieved an accuracy of 55% for English and 64% for Spanish

1. Rule-based Approach

Rule-based classification categorizes emotions by applying a series of "if-then" rules. The "if" part is known as the "rule antecedent," while the "then" part is referred to as the "rule consequent." These emotion rules are derived using statistical, linguistic, and analytical principles. The most effective rules are then chosen and applied to the emotion dataset to identify emotion labels.

Seal et all’s (2020)[30] study concentrates on emotion detection using semantic rules and emotional keywords, particularly focusing on phrasal verbs. They assembled a phrasal verbs list and developed a database including synonyms for the verbs. By identifying and ranking keywords and phrasal verbs associated with particular emotions, they attained an accuracy rate of 65%. Nonetheless, their method did not address existing system issues, such as incomplete lists of emotional keywords and the neglect of contextual semantics.

1. Machine Learning Approach

Machine learning constructs computational models using sample data known as "training data" to make decisions or predictions. This approach can deduce decision rules for identifying emotions by analysing labelled training examples.

Badugu & Suhasini (2020)[31] used machine learning method to identify emotion in Twitter messages. Their result shows that their Naive Bayes model was more efficient than the K Nearest Neighbor. Nasir et al. (2020) [32] developed a system for detecting emotions from text using supervised learning algorithms. They used Multinomial Naive Bayes, KNN, DT, and SVM for the ISEAR dataset and achieved the highest accuracy rate with 64.08%.

Mozafari & Tahayori (2019)[33] introduced similarity techniques utilizing VSM and STASIS methods. When tested on the ISEAR21 dataset, the VSM method surpassed the Keyword and STASIS methods in terms of detection performance. Their experiment showed that their emotion detection achieved a precision of 0.53.

In order to increase emotion detection accuracy using a unique feature representation, Anzum & Gavrilova (2023) [34] presented an adjusted weight ensemble classifier using a genetic algorithm. Six emotions—sadness, anger, fear, love, joy, and surprise—were used to label each entry in the standard Twitter emotion detection dataset used to train and assess this classifier. With the best accuracy, recall, F1-score, and precision (96.49%), the suggested strategy outperformed existing conventional machine learning-based emotion recognition techniques.

* 1. Deep Learning Approach

Deep learning is a branch of machine learning, involves neural networks. It utilizes algorithms similar to those in machine learning but operates with multiple layers. These layers are known as artificial neural networks. These algorithms mimic the human brain, where neural networks are interconnected, enabling deep learning to address complex problems through its layered processes and algorithms.

Park et al. (2020)[35] developed a model for emotion detection using CNN with a dataset of 144,701 tweets and ROC story data. Their findings showed the highest accuracy of 73.3% for detecting the joy emotion, while the anger emotion had the lowest Kappa score of 0.216 and an accuracy of 36.7%.

In order to integrate the discovered correlations into a cohesive framework for identifying various emotions, Zhang et al. (2020) [36] created a factor graph model. An F1 score of 62.7 was attained by their multilabel learning technique.

Malte & Ratadiya (2019)[37] introduced a generative deep learning approach using a two-way transformer-based BERT architecture. Their tests on English and Hindi datasets yielded F1 scores of 0.5520 for English and 0.4521 for Hindi. Waleed et al. (2019)[38] built a model utilizing deep transfer learning for language modelling, achieving a micro-F1 score of 0.7582.

1. Hybrid Approach

This method combines elements of both keyword-based and learning-based approaches. By incorporating knowledge-rich linguistic information from dictionaries and synonyms, it aims to improve accuracy through training. This approach seeks to strike a balance between the resource-intensive nature of information retrieval tasks and minimizing associated challenges.

Mahima et al. (2021)[39] introduced a hybrid approach that combines rules, sentiments, and context to clarify word meanings. They utilize sentence transformers to detect emotions through NLP, BERT, sentence embeddings and similarity techniques, aiming to address these limitations effectively. Their research focuses on accurately identifying Ekman’s emotions and neutral emotions, tagging multiple emotions based on specific contexts. This hybrid methodology has demonstrated significant superiority over current methods in detecting multiple emotions.

**TABLE 3** provides a summary of the previous research with an emphasis on the approach applied, the main contributions and limitations.

**TABLE 3**. Summary of Previous Works on Emotion Detection in Text

|  |  |  |  |
| --- | --- | --- | --- |
| **Proposed Work** | **Approaches** | Contributions | Limitations |
| Jasvitha et al. (2024)[40] | Machine Learning | The MLP achieved the highest performance with an average F-measure of 1.0 across all classes. On the other hand, Decision Tree, SVM and Random Forest classifiers showed nearly identical performance, each reporting a 0.99 F-measure. | Neural networks are exceptional at learning intricate patterns but often demand significant computational resources. |
| Khurdula et al. (2024)[41] | Deep Learning | BERT performs slightly worse than BiLSTM in recognizing certain emotions such as sadness. It demonstrates reduced accuracy for emotions like love (0.84) and surprise (0.83), which aligns with BiLSTM's performance and highlights the issue of class imbalance. | The fine-tuned Text-Bison-001 model's performance cannot be directly compared to Bi-LSTM or BERT because it produces its own labels specifically suited to the text. Instead, its metrics need to be quantified through manual validation. |
| Anzum & Gavrilova (2023) [34] | Machine Learning | Used a genetic algorithm to build the representation of the input, sentiment, incorporating stylistic, and linguistic features derived from tweets. This proposed method achieved exceptional precision, recall, F1-score and accuracy (96.49%). | Did not capture a broad spectrum of emotions and has not been tested on different user groups. |
| Machová et al. (2023)[42] | Machine Learning | Experimented with lexicon-based approaches, Naïve Bayes, SVM and neural networks. Their neural network model particularly effective in classification of multi-class emotion and achieved a 0.95 F1-score for detecting sadness. | In certain instances, the text may only weakly convey the emotion, or multiple emotions may be present simultaneously. |
| Yang & Zhang (2023) [43] | Machine Learning | The MLI-ED model enhances micro-F1 from 0.444 to 0.603 and average precision from 0.702 to 0.719, surpassing the DATN model. Compared to TeCAP, the MLI-ED model boosts macro-F1 from 0.438 to 0.484 and increases average precision from 0.579 to 0.719. Additionally, compared to LR-GCN, the proposed model lowers Hamming loss from 0.178 to 0.150 and improves accuracy. | Did not incorporate enough emotion information and lacks effective fusion of text features. |
| Cahyani et al. (2022)[44] | Deep Learning | BERT+CNN achieved accuracy values of 86.47% on commuter line data, 87.23% on transjakarta data, and 86.18% on combined commuter line and transjakarta data. Word2Vec+CNN achieved the second-highest accuracy, with GloVe+CNN showing the lowest accuracy. | dataset used in the research might have imbalanced data across different emotion classes, which can negatively impact the model's performance. |
| Izadkhah, Habib (2022) [45] | Deep Learning | Employed FastText and GloVe for encoding textual data into numerical representations to identify syntactic, semantic, and word similarities. The attention property was also used to increase accuracy. | The addition of layers is restricted, resulting in reduced accuracy (degradation issue). |
| Mahima et al. (2021) [39] | Hybrid | Achieved the highest classification accuracy of 57.447% on the EmoDB and GoEmotions datasets using Custom VSM with Sentiment Analyzer for multi-emotion classification. | Not include all possible emotions, particularly those not covered by Ekman's emotions. |
| Acheampong et al. (2021) [46] | Deep Learning | Evaluates how well a combination of XLNet and RoBERTa identifies emotions within the ISEAR dataset and achieved the highest F1-score of 0.75. | Not include all possible emotions, particularly those not covered by ISEAR dataset. |
| Del Arco et al. (2020)[29] | Corpus-based | Using machine learning and a multilingual dataset for emotion detection, It obtained 0.64 accuracy for Spanish and 0.55 accuracy for English. | The recommendation system failed to identify emotions of anger, disgust, fear, and sadness. |
| Seal et al. (2020)[30] | Rule-based | Identify emotions using the ISEAR dataset by paying special attention to phrasal verbs. | Contextual and word meanings for the lexicon is still lacking. |
| Badugu & Suhasini (2020) [31] | Machine Learning | Proving the efficiency of K Nearest Neighbor and Naive Bayes then obtaining accuracy results of 55.50% and 72.06%, respectively. | Restricted extraction of pertinent information within sentences. |
| Ab. Nasir et al. (2020)[32] | Machine Learning | Four different machine learning algorithms were evaluated in terms of performance. With the greatest mean accuracy score of 64.08%, the Multinomial Naïve Bayes classifier surpassed the SVM (15.37%), Decision Tree classifiers (52.36%), and k-NN (47.95%). | Only analyzed Ekman’s six basic emotions, potentially missing out on recognizing more complex or mixed emotional expressions present in texts. |
| Park et al. (2020) [35] | Deep Learning | Used a CNN to develop a model of embedded emotions, achieving a peak accuracy of 73.3%, a low accuracy of 36.7%, and the lowest 0.216 Kappa score. | Inability to manage sentence-level expressions that negate the overall sentiment. |
| Zhang et al. (2020)[36] | Deep Learning | Utilizing factorial graph modeling to identify multiple emotions, implementing an approach to multi-label learning and acquiring contextual features, resulting 62.7 of F1-score. | Used a small-scale dataset and lack of retrospective checks in the annotation process. |
| Mozafari & Tahayori (2019) [33] | Machine Learning | VSM, STASIS and Keyword Base approaches were used to find different emotion types and achieved an accuracy of 0.53. | Weakness of qualitative information extraction. |
| Malte & Ratadiya (2019) [37] | Deep Learning | Build BERT (bidirectional transformer) to test both English and Hindi text, the F1 score obtained the accuracy of 0.4521 (Hindi) and 0.5520 (English). | Insufficient handling of slang text in languages ​​other than English, which affects the accuracy of word representations. |
| Ragheb et al. (2019)[38] | Deep Learning | Using SemEval2019 dataset, they created a model using deep transfer learning, turn-based conversation model and self-attention mechanism for emotions classifying. The result attained a 0.7582 F1 score. | Can’t find the emotions – “happy” |
| Shah et al. (2019)[23] | Machine Learning | Created a model employing lexical analysis with WordNet-Affect and EmoSenticNet, integrated with supervised classifiers, to identify emotions in tweets. Their SVM classifier achieved an accuracy of 89.28% in the Anger class. | Used a limited dataset and encountering language ambiguity issues concerning texts that depict multiple emotions simultaneously. |
| Tzacheva et al. (2019)[47] | Hybrid | SVM and the NRC emotion lexicon were used on WEKA, Spark software, yielding accuracy rates of 84.92 percent and 88.01 percent in order to identify useful emotional patterns in tweets. | The limited number of emotional classifications makes generalizing inappropriate. |

## RQ3: What metrics are used to evaluate these methods?

Performance metrics were used to statistically evaluate the effectiveness of the propose methods. Some of the metrics used are as follows.

1. Kappa Coefficient [36]

If the rows and columns of the contingency table represent the same categories, the degree of agreement measures the association between the two variables. Therefore, the contingency table will be symmetrical because both variables classify categories similarly. The Kappa coefficient is employed to quantify this level of agreement. Generally, the Kappa coefficient can be utilized for:

1. Assessing the two raters' degree of agreement when categorizing objects into groups.
2. Evaluating agreement between new alternative methods and established methods.

The Kappa coefficient is computed using the **EQUATION (1)**:

(1)

Here, represents the observed agreement between raters, while denotes the theoretical probability of chance agreement. These probabilities are calculated based on the observed data, reflecting the likelihood that each rater would randomly select each category. The value ranges from 0 to 1.

1. Jaccard Accuracy [37], [46]

Jaccard Accuracy is a technique employed to assess the similarity between two items. The Jaccard coefficient quantifies similarity within a defined set by determining the ratio of the intersection to the union of the samples.

Note that by design . If intersection B is empty, then. Jaccard Accuracy is calculated using the **EQUATION (2)**:

(2)

1. Precision, Recall, F-Score Accuracy

Precision () is the ratio of True Positives () to the total number of data points predicted as positive. [34], [35], [42], [43], [44], [45], [46], [47] calculate the precision using the **EQUATION (3)**:

(3)

Recall is the ratio of True Positive () to the total number of data points that are actually positive. The recall is calculated using the **EQUATION (4)**:

(4)

Where is recall, is true positive and is false negative.

The F1-Score represents the harmonic mean of precision and recall. A perfect F1-Score is 1.0, while the worst score is 0. A high F1-Score indicates that our classification model achieves good precision and recall. [35], [37], [38] and [39] , [41], [43], [44], [45], [46], [47] calculate the F1 score is calculated using the **EQUATION (5)**:

(5)

Where represents precision and represents recall.

Classification models are measured using accuracy metrics. [30], [33], [35], [42], [43], [45] are calculated using the **EQUATION (6, 7)**:

(6)

(7)

Where is true positive, is true negative, is false positive, and is false negative.

1. 10-Fold Cross Validation [30]

Cross-validation is a statistical method employed to evaluate model performance by partitioning data into training and validation (or test) subsets. A commonly used technique is 10-fold cross-validation, where data is split into 10 equal parts (folds). In iterative cycles, the model is trained on 9 folds and validated on the last fold, making sure the validation set is used exactly once for each fold. When compared to conventional cross-validation, leave-one-out cross-validation, and bootstrap approaches, this approach typically produces less biased accuracy estimates.

# FUTURE WORKS

While the previous studies has made significant progress in the field of emotion detection in text, several areas remain ripe for further exploration and improvement. Future research can build upon our findings in the following ways:

1. **Contextual Understanding**: Develop techniques to better understand the context surrounding text-based conversations, including dialogue history, user intent, and situational context. Context-aware models can improve the accuracy of emotion detection by considering the broader context in which emotions are expressed.
2. **Fine-grained Emotion Recognition**: Investigate methods for fine-grained emotion detection that distinguish between subtle variations of emotions (e.g., distinguishing between sadness and disappointment). This requires more nuanced models that can capture and differentiate between closely related emotional states.
3. **Real-time Emotion Detection**: Explore real-time emotion detection methods that can dynamically adjust responses based on changes in the state of user's emotional during the conversation. This can enhance the responsiveness and effectiveness of chat bot interactions.

By pursuing these future research directions, researchers can advance the field of emotion detection in text, leading to more accurate, interpretable, and contextually aware systems that enhance human-computer interaction and support a wide range of applications in both academic and commercial domains.

# CONCLUSIONS

In this paper, we have conducted a systematic review of the various proposed solutions for emotion detection in texts with a focus on the data set used, the approach applied and the evaluation metrics. This review paper summarizes a collection of corpus and lexicon sources used as datasets in previous studies. The methods that have been used in the previous work are keyword-based approach, rules-based approach, machine learning approach, deep learning approach and hybrid approach. Furthermore, the technical evaluation process is carried out using several metrics, such as Kappa Coefficient, Jaccard Accuracy, Precision, Recall, F-Score Accuracy and 10-Fold Cross Validation. Significant contributions and limitations to previous studies are also discussed to compare the performance of these methods. Several studies to detect emotions in texts have been carried out quite well, but there are very few emotions labelled data sources other than English. The availability of data sources in other languages such as Bahasa (Indonesian) can greatly encourage research in the area of emotion detection to balance work done in multiple languages.

# REFERENCES

1. Lopatovska, I. and I. Arapakis, *Theories, methods and current research on emotions in library and information science, information retrieval and human–computer interaction.* Information Processing & Management, 2011. **47**(4): p. 575-592.

2. Mohammad, S.M. and S. Kiritchenko, *Using hashtags to capture fine emotion categories from tweets.* Computational Intelligence, 2015. **31**(2): p. 301-326.

3. Neviarouskaya, A., H. Prendinger, and M. Ishizuka, *Affect analysis model: novel rule-based approach to affect sensing from text.* Natural Language Engineering, 2011. **17**(1): p. 95-135.

4. Aman, S. and S. Szpakowicz. *Identifying expressions of emotion in text*. in *International Conference on Text, Speech and Dialogue*. 2007. Springer.

5. Garcia, D. and F. Schweitzer. *Emotions in product reviews--empirics and models*. in *IEEE International Conference on and IEEE International Confernece on Social Computing (SocialCom) Privacy, Security, Risk and Trust (PASSAT)*. 2011. IEEE.

6. Alm, C. *Emotions from text: machine learning for text-based emotion prediction*. in *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*. 2005.

7. Cecilia, E. and A. Ovesdotter, *Affect in text and speech.* ProQuest, Citeseer, 2008.

8. Bostan, L.A.M. and R. Klinger, *An analysis of annotated corpora for emotion classification in text.* Proceedings of the 27th International Conference on Computational Linguistics, 2018.

9. Li, Y., et al., *Dailydialog: A manually labelled multi-turn dialogue dataset.* arXiv preprint arXiv:1710.03957, 2017.

10. Buechel, S. and U. Hahn. *Readers vs. writers vs. texts: Coping with different perspectives of text understanding in emotion annotation*. in *Proceedings of the 11th Linguistic Annotation Workshop*. 2017.

11. Ghazi, D., D. Inkpen, and S. Szpakowicz. *Detecting emotion stimuli in emotion-bearing sentences*. in *Computational Linguistics and Intelligent Text Processing*. 2015. Springer.

12. Liu, V., C. Banea, and R. Mihalcea. *Grounded emotions*. in *Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. 2017. IEEE.

13. Poria, S., et al., *Meld: A multimodal multi-party dataset for emotion recognition in conversations.* arXiv preprint arXiv:1810.02508, 2018.

14. Carlo Strapparava, R.M. *SemEval-2007 Task 14: Affective Text*. in *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007*. 2007.

15. Rosenthal, S., N. Farra, and P. Nakov. *SemEval-2017 Task 4: Sentiment Analysis in Twitter*. in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. 2019.

16. Mohammad, S., et al. *SemEval-2018 Task 1: Affect in Tweets*. in *Proceedings of the 12th International Workshop on Semantic Evaluation*. 2018.

17. Ankush Chatterjee, K.N.N., Meghana Joshi and Puneet Agrawal. *SemEval-2019 Task 3: EmoContext*

*Contextual Emotion Detection in Text*. in *Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019)*. 2019.

18. Bo Wang, M.L., Arkaitz Zubiaga, Rob Procter, Eric Jensen. *SMILE: Twitter Emotion Classification using Domain Adaptation*. in *Proceedings of the 4th Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2016)*. 2016.

19. Preoţiuc-Pietro, D., et al. *Modelling Valence and Arousal in Facebook posts*. in *Proceedings of NAACL-HLT 2016*. 2016.

20. Saif M. Mohammad, F.B.-M. *WASSA-2017 Shared Task on Emotion Intensity*. in *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. 2017.

21. Nielsen, F.A. *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 (2011)*. 2011.

22. Hu, M. and B. Liu, *Mining and summarizing customer reviews*, in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. 2004, Association for Computing Machinery: Seattle, WA, USA. p. 168–177.

23. Shah, F.M., et al. *Emotion Detection from Tweets using AIT-2018 Dataset*. in *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*. 2019.

24. Mohammad, S.M., et al., *Sentiment, emotion, purpose, and style in electoral tweets.* Information Processing & Management, 2015. **51**(4): p. 480-499.

25. Mohammad, S. and S. Kiritchenko, *Using Hashtags to Capture Fine Emotion Categories from Tweets.* Computational Intelligence, 2014. **31**.

26. Svetlana Kiritchenko, X.Z., Saif M. Mohammad, *Sentiment Analysis of Short Informal Texts.* Journal of Artificial Intelligence Research 2014: p. 723-762.

27. Mohammad, S. and P. Turney. *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*. 2010.

28. Mohammad, S. and P. Turney, *Crowdsourcing a Word-Emotion Association Lexicon.* Computational Intelligence, 2013. **29**.

29. Frohmann, M., et al. *Predicting the Price of Bitcoin Using Sentiment-Enriched Time Series Forecasting*. Big Data and Cognitive Computing, 2023. **7**, DOI: 10.3390/bdcc7030137.

30. Seal, D., U.K. Roy, and R. Basak. *Sentence-Level Emotion Detection from Text Based on Semantic Rules*. in *Information and Communication Technology for Sustainable Development*. 2020. Singapore: Springer Singapore.

31. Suhasini, M. and B. Srinivasu. *Emotion Detection Framework for Twitter Data Using Supervised Classifiers*. in *Data Engineering and Communication Technology*. 2020. Singapore: Springer Nature Singapore.

32. Ab. Nasir, A.F., et al., *Text-based emotion prediction system using machine learning approach.* IOP Conference Series: Materials Science and Engineering, 2020. **769**(1): p. 012022.

33. Mozafari, F. and H. Tahayori. *Emotion Detection by Using Similarity Techniques*. in *2019 7th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*. 2019.

34. Anzum, F. and M.L. Gavrilova, *Emotion Detection From Micro-Blogs Using Novel Input Representation.* IEEE Access, 2023. **11**: p. 19512-19522.

35. Park, S.H., 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)*. 2020.

36. Zhang, X., et al., *Emotion Detection in Online Social Networks: A Multilabel Learning Approach.* IEEE Internet of Things Journal, 2020. **7**(9): p. 8133-8143.

37. Malte, A. and P. Ratadiya. *Multilingual Cyber Abuse Detection using Advanced Transformer Architecture*. in *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*. 2019.

38. Waleed Ragheb, J.A., Sandra Bringay, Maximilien Servajean1, *Attention-based Modeling for Emotion Detection and Classification in Textual Conversations.* 2019.

39. Mahima, M.A., et al. *A Text-Based Hybrid Approach for Multiple Emotion Detection Using Contextual and Semantic Analysis*. in *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*. 2021.

40. Jasvitha, B.D., et al. *Emotion Detection from Text Using ML Framework*. in *2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE)*. 2024.

41. Khurdula, H.V., A. Pagutharivu, and J.S. Yoo. *The Future of Feelings: Leveraging Bi-LSTM, BERT with Attention, Palm V2 & Gemini Pro for Advanced Text-Based Emotion Detection*. in *SoutheastCon 2024*. 2024.

42. Machová, K., et al., *Detection of emotion by text analysis using machine learning.* Frontiers in Psychology, , 2023. **14**(1664-1078 (Print)).

43. Yang, H. and J. Zhang. *Multi-label Text Emotion Detection Based on Multi-level Information*. in *2023 9th International Conference on Computer and Communications (ICCC)*. 2023.

44. Cahyani, D.E., et al. *Emotion Detection in Text Using Convolutional Neural Network*. in *2022 International Conference on Electrical and Information Technology (IEIT)*. 2022.

45. Izadkhah, H. *Detection of multiple emotions in texts using a new deep convolutional neural network*. in *2022 9th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*. 2022.

46. Acheampong, F.A., H. Nunoo-Mensah, and W. Chen. *Recognizing Emotions from Texts Using an Ensemble of Transformer-Based Language Models*. in *2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. 2021.

47. Tzacheva, A., J. Ranganathan, and S.Y. Mylavarapu. *Actionable Pattern Discovery for Tweet Emotions*. in *Advances in Artificial Intelligence, Software and Systems Engineering*. 2020. Cham: Springer International Publishing.