IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 21 December 2023, accepted 9 January 2024, date of publication 19 January 2024, date of current version 7 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3356357

SURVEY

Challenges and Opportunities of Text-Based Emotion Detection: A Survey

ABDULLAH AL MARUF[1], FAHIMA KHANAM[1], MD. MAHMUDUL HAQUE[1], ZAKARIA MASUD JIYAD[1], M. F. MRIDHA[2], (Senior Member, IEEE), AND ZEYAR AUNG[3,4], (Senior Member, IEEE)

[1]Department of Computer Science and Engineering, Bangladesh University of Business and Technology (BUBT), Dhaka 1216, Bangladesh
[2]Department of Computer Science, American International University-Bangladesh (AIUB), Dhaka 1229, Bangladesh
[3]Center for Secure Cyber-Physical Systems (C2PS), Khalifa University, Abu Dhabi, United Arab Emirates
[4]Department of Computer Science, Khalifa University, Abu Dhabi, United Arab Emirates

Corresponding author: Zeyar Aung (zeyar.aung@ku.ac.ae)

This work was supported in part by Khalifa University, Abu Dhabi, United Arab Emirates.

ABSTRACT Emotion detection has become an intriguing issue for researchers owing to its psychological, social, and commercial significance. People express their feelings directly or indirectly through facial expressions, language, writing, or behavior. An emotion detection tool is a critical and practical way to recognize and categorize moods in various applications. Artificial intelligence (AI) is often used to identify emotions. Machine learning and deep learning algorithms produce high-quality solutions for diagnosing emotional diseases among social media users. Numerous studies and survey articles have been published on emotion detection using textual data. However, most of these studies did not comprehensively address the emerging architectures and performance analyses in emotion detection. This study provides an extensive survey of state-of-the-art systems, techniques, and datasets for textual emotion recognition. Another goal of this study is to emphasize the limitations and provide up-and-coming research directions to fill these gaps in this rapidly evolving field. This survey paper investigated the concepts and performance of different categories of textual emotion detection models, approaches, and methodologies.

INDEX TERMS Text-based emotion detection, datasets, machine learning models, performance metrics, challenges and emerging trends.

I. INTRODUCTION

In today's fast-paced society, mental health disorders, such as stress, anxiety, and depression, have emerged as significant challenges. Almost 264 million people suffer from depression, which is the most crucial illness worldwide. It is alarming that mental illnesses such as depression and others isolate people, and depressed people stay reserved from ordinary life. Social networks have become the most exciting way for users to communicate and share their thoughts as well as unhappiness with their circle of friends and other people. Persons with depression use social media to connect with others, share their emotional experiences, and support each other. Just as psychologists predicted human behavior long ago through experience, experimentation, and observation, scientists in other fields could predict results based on their findings. Machine learning (ML) techniques detect emotions regarding current trends, themes, and discussions on the Web on people's psyches by determining their emotional states [1].

In the field of research, multiple methods for identifying emotions in the textual data have been devised using various machine learning algorithms and classifiers such as decision trees, support vector machines, naive Bayes, logistic regression, and k-nearest neighbor [2]. Applying machine learning algorithms to a large volume of textual data poses a challenge. For example, Twitter (now "X") has an average of 6,000 tweets per second, leading to a staggering 350,000 tweets per minute, with daily totals reaching 500 million and an annual total of approximately 200 billion tweets [3]. Scalability is an important issue in dealing with such big datasets.

The associate editor coordinating the review of this manuscript and approving it for publication was M. Shamim Kaiser.

Changes in emotional expression are reflected in the language used. Schwarz-Friesel [4] conducted a study on the relationship of language and emotion from the cognitive linguistic perspective. Speech emotion recognition (SER) research has aroused great interest in dynamic computing and human-computer interaction (HCI). One of the objectives of SER is to improve the interaction between humans and machines. For example, the SER system has been used to identify and categorize six primary emotions in the Sepedi language (one of South Africa's official languages): sadness, disgust, anger, fear, neutrality, and joy [5].

Emotional well-being is a primary factor in creating a smart social environment. Stress, depression, and frustration are related to emotional issues. Several researchers have investigated this topic [6]. Mental stability is a significant issue among tech people. In [7], the author used the ML technique to determine stress in the tech industry. They conducted surveys and gathered responses to questionnaires in order to compile their dataset.

Nowadays, physiological signals are also used to detect human emotions. Jang et al. [8] studied physiological signals induced by basic emotions. Recently, various algorithms have been used to perform emotion recognition based on physiological signals.

Owing to its broad applicability across numerous fields and the skills and abilities of those who use it, emotion detection has become increasingly significant. This technology has advanced past its original limitations and is now an important tool in various industries, including marketing, customer service, healthcare, and education. Its ability to improve interactions between people, decision-making, and user experience highlights its significance in tackling difficulties and opportunities worldwide. There are two types of emotional expressions: (i) affective items (words are determined by context and do not have direct assignation) and (ii) emotive things (assigning emotional categories of happiness, sadness, love, etc.).

Various media, such as speech, text, signals, audio, and video, can be used for detecting human emotions. In this survey, we focused only on text-based emotion detection (TBED). This topic has received considerable attention. Several surveys and review articles have focused on text-based emotion detection [9], [10]. Some researchers have dived into deep learning (DL)-based TBED [11] while some have attended to transformer-based TBED [12].

Some scholars have directed their concentration towards the analysis of sentiments that are target-dependent and aspect-based [13], [14]. Zhang et al. [14] analyzed the tasks of aspect-based TBED and presented methods and challenges. However, in their research, they did not discuss all of the components of TBED. Abudalfa and Ahmed [13] targeted a micro-blog analysis of social media in their survey paper. They explored techniques and gaps.

Researchers have identified the classifiers, algorithms, and lexicons related to TBED. Some researchers investigated datasets that are available for emotion detection. A few scholars have examined the essential components of TBED. However, many of them did not present any challenges or future research directions. As a result, each of the existing survey and review articles has some limitations. We studied these limitations and aimed to address these gaps. An overview of the existing TBED review and survey investigations conducted by various researchers is presented in Table 1.

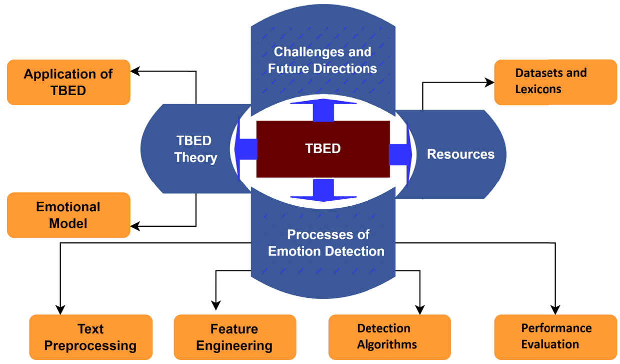

FIGURE 1. Overall structure of this survey.

The main contributions of this survey study are as follows:

1) This paper focuses on emotional models and their architecture, including the emotional types of a model, as well as the pros and cons.
2) This study aims to investigate various machine learning (ML) and deep learning (DL) techniques for text-based emotion detection (TBED), describe the basic architecture, and demonstrate a comparative investigation of different strategies.
3) The study extensively covers different text preprocessing techniques on how text can be made ready for the subsequent learning task using various tools, packages, and methods.
4) The survey presents feature extraction methods and their models, architectures, types, advantages, and disadvantages.
5) In this study, we detailed the resources (datasets and lexicons) used for textual emotion detection. Additionally, the datasets were divided into various emotion detection study domains.
6) The study further describes the difficulties faced by current emotion detection systems, presents emerging trends and explores their prospects.
7) We kept a special focus on other languages in addition to English. We discussed their datasets, language models, and other special language-specific considerations.
8) We briefly explored domain-specific and aspect-based sentiment analysis (ABSA) challenges and solving techniques.

This paper delves into the application and emotional model descriptions in Section II. This is followed by a comprehensive presentation of TBED resources, including datasets and lexicons, in Section III. The processes of

TABLE 1. Summary of the existing survey/review papers on TBED. (Features of surveys/reviews are F1: Applications, F2: Emotional Models, F3: Text Preprocessing, F4: Feature Extraction, F5: Approach, F6: Classifiers, F7: Performance, F8: Evaluation Metrics, F9: Datasets, and F10: Future Directions.)

Ref.	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Comments
[12]	✗	✓	✗	✓	✗	✓	✓	✗	✓	✓	The author focused mainly on the transformer-based model, its weaknesses, and its strengths for TBED in this survey.
[9]	✗	✓	✗	✗	✓	✓	✓	✗	✓	✓	It examined the concept, critical methodologies, futuristic approaches, datasets used, strengths and shortcomings of TBED systems. The authors did not specifically address the performance or classifier. They somewhat explained this in related work. There is no comparative analysis seen between performance and the classifier.
[15]	✗	✓	✗	✓	✓	✓	✗	✗	✓	✓	This survey covered progress in emotion detection investigation, such as various emotion models, detection algorithms, associated datasets, features, limitations, and future methods in text and speech-based emotions.
[16]	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	It focused on identifying both hidden and explicit emotions in writing. It highlighted cutting-edge methodologies, their properties with benefits and drawbacks, various datasets, and various lexicons available for text-based emotion recognition. Unfortunately, the authors did not provide any comparative analysis of the outcomes.
[17]	✗	✓	✗	✓	✓	✗	✗	✗	✓	✗	It examined the work completed in TBED. The authors contended that many approaches, procedures, and models created to identify emotion in the text are insufficient for various reasons.
[18]	✓	✓	✗	✗	✓	✗	✗	✗	✓	✓	The survey looked at the data and emotions corpora. It just evaluated the results of numerous datasets. A classifier's or model's performance was not compared to others.
[19]	✗	✓	✗	✓	✓	✗	✓	✗	✓	✗	It gave a general review of emotional theories, lexicons, and datasets. It was also justified to use emotion mining-related polarity classification algorithms.
[20]	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	It examined techniques, classifiers, application fields, and challenges.
[21]	✗	✓	✓	✗	✓	✗	✗	✓	✓	✗	It reviewed and categorizes the most recent and noteworthy emotion detection investigations. The dataset and evaluation procedure are also covered.
[22]	✗	✓	✗	✗	✓	✗	✗	✗	✗	✗	The survey work highlighted recent research on TBED. The detection technique and emotional theory were studied, and the ideal strategy was determined.
[23]	✗	✓	✓	✓	✓	✗	✓	✓	✓	✗	The mechanism for sentiment analysis and emotion recognition was discovered. The dataset and domain were also discussed, and some work on an algorithm to categorize emotions was done.
[24]	✗	✓	✗	✓	✓	✗	✗	✗	✓	✓	It examined the extraordinary challenges and future research directions in deep text emotion recognition. The methodologies of DL for text emotion recognition were the primary emphasis.
[25]	✗	✗	✓	✗	✓	✗	✗	✗	✗	✗	It was mainly concerned with emotion detection work based on ML.
[26]	✗	✗	✗	✗	✓	✗	✗	✗	✓	✗	It reviewed the ways to detect emotions in Bangla (Bengali) text.
[27]	✗	✓	✗	✗	✓	✗	✗	✓	✓	✗	The authors explored the techniques and emotional models, resources such as corpora and lexicons. The classifier's performance was not precisely mentioned.
[28]	✗	✓	✗	✗	✓	✗	✗	✗	✓	✓	The authors aimed to evaluate how different methodologies have been utilized to construct emotion dimension datasets, optimize affective computing systems, and undertake review analysis in this survey.
[29]	✗	✗	✓	✗	✓	✗	✗	✗	✓	✓	The authors surveyed emotion detection systems based on text. In this research, they analyzed methodologies and preprocessing.
[30]	✗	✗	✗	✗	✓	✓	✗	✗	✓	✓	This survey described the DL classifier, architecture, and model for emotion detection. Strengths and weaknesses of the model and architecture were also explored, as well as difficulties and future directions.

emotion detection are described in Section IV, where we present the text preprocessing techniques, feature engineering approaches, detection algorithms, and evaluation metrics. Finally, Section V discusses the challenges of TBED and explores their future directions. The overall structure of the survey is illustrated in Figure 1.

II. EMOTION DETECTION: BACKGROUND

In this section, we discuss the uses of TBED. In addition, we describe the emotional models and present short comparisons of all models.

A. APPLICATIONS OF TBED

Because emotions are a fundamental part of human nature, emotion analysis has received considerable attention in psychology, neurology, and behavioral science. Figure 2 depicts some of the applications of emotion detection.

Politicians may understand the public's concerns regarding security issues based on the emotions displayed in social media debates. Marketers can develop sales agendas to advertise their goods and services [3], [31]. One of the important social media is Twitter [32] (now re-branded as "X"). As reported in [1], the authors concentrated on Twitter data to examine historical and present feeds to extract emotions. In Twitter posts, hashtags may convey various semantic payloads [33], [34]. Scholars obtained Arabic texts from Twitter's social networking platform, and human annotations assigned appropriate emotions to each text [35]. Understanding patients' emotional states can be gained by gaining insight into their emotional conditions by detecting subtle emotions in online health groups [36]. In [37], the authors detected emotions in customer service emails. Another piece of work [38] forecasts and presents the patterns of Facebook post reactions. In [39], the scholars utilized the Facebook reactions for supervision emotion detection. The language used in the YouTube comments was used to classify emotions in [40]. In [3], the author proposed a model where emotions were detected to understand the people's sentiments towards national parties.

Chatbots can perform various functions including recommendation systems, information providers, instructors, advisors, discussion partners, and entertainers. Using a model developed by researchers, chatbots can automatically assess human participants based on their distinct emotional features, extract emotions from conversations, and detect changes over time [41]. Because people now connect through a variety of text-based technologies, the ability to extract emotions from text has gained significant attention. The authors [42] used a dataset called "Emotion Categorization" with six different emotional categories. Short messages are known as "microblogs" and frequently feature new vocabulary, loud noise, and abbreviations.

Social behavior analysis is becoming increasingly common when studying social robots [43]. A social robot was developed using multi-layer perception (MLP) to improve machine-human interaction [44]. SNS phishing is one type of social engineering assault. A solution to counteract such attacks was proposed in the form of an Intelligent Security Chatbot Assistant (ISCA). This chatbot was crafted using Google Dialogflow on Telegram and was powered by an AWS server [45].

Emotion detection can also be employed in the review systems. People express their thoughts about the products and films. Business Intelligence (BI) encompasses a range of techniques, principles, and practices for transforming data into insightful information to streamline business operations. User satisfaction with a product, system, or service is increasingly being recognized as a crucial element of successful design and is essential to the overall competitiveness of a product or service [46]. The authors of [47] researched ways to capture and recover user input data from online customer evaluations to facilitate product design. Another study suggested using genetic programming to adapt domains for product evaluations [48].

With the popularity of online social media, many efforts have been made to identify individuals exhibiting depressive symptoms and utilize these platforms for early diagnosis, therapy, and prevention [49]. However, depression is one of the main factors contributing to mental illness, which has been linked to a higher chance of dying young. Suicidal ideation is greatly exacerbated by depression, and daily functioning is severely hampered [50]. Moreover, anxiety and depression are intimately related to behavioral mental illness, and anxious depression is a person's mental state [51].

We graphically represent different applications of TBED in Figure 2.

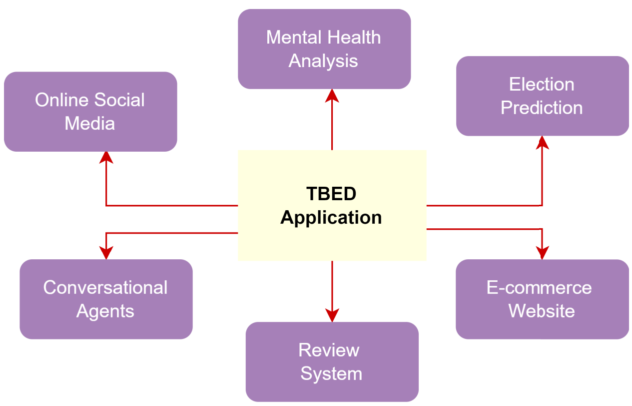

FIGURE 2. Various applications of text-based emotion detection (TBED). (Note: they are by no means exhaustive.)

B. EMOTION MODELS

Psychology researchers have identified three primary emotion modeling techniques [52], [53]. These are described below and summarized in Table 2.

1) **Discrete Model:** This method is based on the concept that there are only a few fundamental and widely recognized emotions [53]. Paul Ekman's model [54] is the most widely used in emotion recognition research, and includes six fundamental emotions: happiness, sorrow, anger, fear, surprise, and disgust.

2) **Dimensional Emotion Model:** Dimensional emotion models specify feelings based on a few dimensions defined by specific factors. Most dimensional emotion models include two or three dimensions [55], [56].
   - **Valence:** This calculates the degree of emotion as positive or negative [53].
   - **Arousal:** It measures whether a feeling is aroused or apathetic [53].
   - **Power:** It denotes the level of power [53].

   The paradigm categorizes emotions into arousal and valance. Arousal is divided into action potentials and disbands. On the other hand, valence differentiates between unpleasantness and pleasure. According to the circumplex effect, emotions are linked rather than separated. Plutchik's 2-dimensional emotional wheel [57] has parallel mean arousal and a perpendicular valence axis. Homocentric circles indicate emotions on the wheel. The core of the emotion wheel consists of basic emotions followed by primary emotions. The outer rim displays a variation of fundamental emotions, and the arrangement of emotions on the wheel highlights their interrelatedness and similarity based on their position. Russell and Mehrabian [58] presented a 3-dimensional model of arousal, pleasure or valence, and dominance.
3) **Componential Emotion Model:** This model is an updated version of the dimensional model. For an event, emotions are experienced by the individuals according to it. The outcome is determined by an individual's prior experience, expectations, and action options. According to "appraisal theories" [59], a person's experience of emotions is contingent upon their appraisal of entities that directly impact them. The ultimate determination of this appraisal depends on an individual's goals, prior experiences, and possibilities for action. Language may use appraisal terms to describe the circumstances that lead to feelings. The strongest predictor of an emotional result is an independent evaluation of the former object, incident, or situation [60]. On the other hand, emotion is categorized by Parrott [61] into three grades: primary, secondary, and third level.

III. RESOURCES

This section describes the dataset used for text-based emotion classification and its features and size. Available lexicons are also displayed. We provide download links for datasets and lexicons so that future researchers can readily find them.

A. DATASETS

- **Alm:** The corpus includes tales from Grimm, H.C. Andersen, B. Potter, and roughly 185 other children's stories. The assessment was done at the sentence level using one of the following categories: neutral, disgust-anger, sadness, fear, joy, pleasant surprise, and unpleasant surprise. The writers then annotated the same tales in a group. Every analyst receives independent training to eliminate potential prejudice. The first writer of the work selects one of its chosen emotion tags if there is a dispute between observers [62].
- **Aman:** This dataset extracts blog articles from this corpus with word embeddings for Ekman's six fundamental emotions. For the class of pleasant feelings, the terms happy, delighted, and enjoy were chosen as the word embeddings. Eight different emotion categories were used to mark the text: no feelings and mixed emotions. Four annotators individually analyzed the corpus for interpretation [63].
- **ISEAR:** About 3000 students took part in the test and reported on circumstances in which they felt their respective emotional states: happiness, fear, anger, sorrow, disgust, humiliation, and guilt. A considerable number of analysts worldwide contributed to it [64].
- **SemEval (2007):** This corpus comprises newspaper headlines. Headlines have a framework that enables sentence-level annotations. Anger, disgust, fear, excitement, grief, and astonishment are highlighted in at least one of the headlines [65].
- **SemEval (2018):** Tweets comprise this corpus. Every tweet reflects 11 predetermined emotions: anger, joy, disgust, fear, love, optimism, pessimism, sorrow, surprise, and trust or neutral. In addition, different training, testing, and experiment datasets are offered for tweets in different languages [66].
- **SemEval [(2019):** This corpus comprises two-person textual conversations. The first person initiates the discussion, which becomes a part of the second person before returning to the first person. Each dialogue was classified as happy, angry, unhappy, or other. The final round of discussion serves as the foundation for categorizing emotional categories. In addition, they offer specific training, trials, and testing samples [67].
- **Neviarouskaya:** This corpus uses the ten descriptors of Izard [68]: anger, guilt, panic, disguised, unhappiness, shock, neutral, joy, shame, and interest. The three corpora handle the marking process [63].
  - **Dataset 1:** It comprises 1000 sentences with annotations taken from stories in 13 distinct categories and arranged by theme [69].
  - **Dataset 2:** It contains 700 phrases with annotations that the dataset took from blog posts that resemble diaries [70].
- **EMOBANK:** The Valence-Arousal-Dominance (VAD) emotion representation concept was used to categorize approximately 10,000 words dimensionally. It was created in 2017 by Buechel and Hahn [71]. These phrases were from various sources. Some facts have been divided into categories based on Ekman's primary model, which qualifies for double expressive methods [72].
- **Valence and Arousal:** This data collection was conducted in [73]. It contains labels for two different grades,

TABLE 2. Summary of the three most popular emotion modeling methods.

Item	Discrete	Dimensional	Componential
Primitive ontology	Fundamental emotions	2 or 3 dimensions	Variables in appraisal
Number and variety of emotions possible	A small set of fundamental emotions	A wide range of emotions	An extensive range of emotions
The level of elaboration of the evaluation process	Low	Low	High
Amount of information on effective dynamics	Low/qualitative	Medium/concentration on arousal	A few/qualitative
Advantage	Because of its simplicity, it is widely used	Complex and mixed emotions, as well as a large range of emotion classifications, are properly covered	Concentrate on the irregularity of various feeling states as a result of varied assessment categorizes
Disadvantage	Only use limited emotions, as well as the difficulty of dealing with complicated and conflicted emotions	Some fundamental emotions are incompatible with spatial space, while others exist outside of it	There are no established standard evaluation criteria

which are not related to each other: Valence, sometimes known as sentiment, is a nine-point scale that ranges from 1 (highly negative) to 5 (neutral) and 5 to 9 (very positive), which measures the polarity of an article's dynamic content (very positive).

- **MELD:** A multimodal corpus, MELD, contains data from text, video, and audio sources. It contains over 1400 talks and over 13,000 sentences from the Friends TV series. The dataset was developed in 2018 [74].
- **CBET:** From the tweets, a dataset was created in 2018 [75]. The nine emotions the author sought in the tweets were joy, anger, sorrow, love, contempt, surprise, fear, guilt, and thankfulness. He utilized hashtags to find those tweets. There are two parts in CBET. Since not all emotional combinations occur evenly and regularly, the dataset contains 4,303 tweets that are imbalanced.
- **GoEmotions:** 58 thousand comments were included in the dataset. The dataset was then labeled with 27 emotional classes. This corpus was created in 2020 [76].
- **Amazon:** The dataset of product reviews was created in 2007 [77]. Two thousand sentences were classified into books, DVDs, electronics, and kitchen goods. Each part contains an equal number of classes to balance the dataset. Each sentence provides detailed information on the review.
- **IMDB:** This is the polarity dataset's unlabeled variant. The categorization of binary sentiments in film reviews is made more accessible using the IMDB dataset. The dataset was divided equally across the training and testing datasets, with 25,000 sentences. Its creation began in 2011. The dataset consists of 27,886 raw and unannotated HTML files.
- **MR Dataset:** The dataset is also based on movie reviews and is divided into positive and negative classes among 10,662 sentences [78].

A summary of the datasets' features, sizes, and download links is provided in Table 3.

1) DATASETS FOR OTHER LANGUAGES

We explored the well-known datasets in other languages that are used for emotion detection. Here, we list the datasets from other languages. A summary of these datasets is provided in Table 4.

- **AETD:** This dataset consists of tweets in the Egyptian Arabic dialect. The tweets were classified as belonging to anger, fear, happiness, love, sadness, surprise, sympathy, or none [79].
- **IAEDS:** The Iraqi Arabic Emotion Data set and Facebook posts written in the Iraqi dialect comprise this corpus. It is broken down into six datasets, each of which contains examples of Ekman's fundamental emotions [80].
- **DINA:** The following emotions are annotated in the tweet dataset: surprise, fear, anger, disgust, sadness, and happiness. For every class, a set of seeds was used to query Twitter to collect tweets [81].
- **LAMA-DINA:** Dataset for Arabic Dialects (AD) and Modern Standard Arabic (MSA) emotions; tweets were labeled with Plutchik emotions [82].
- **AraEmoCorpus:** Arabic tweets tagged with the following emotion categories can be found in the corpus: surprise, joy, sadness, fear, anger, and disgust [83].
- **AEELex:** Following the collection of tweets from users residing in Mecca and Riyadh, the Arabic-English emotion lexicon was categorized using Plutchik's eight fundamental emotion categories [84].
- **BanglaSenti:** The "BanglaSenti" dataset, which contains 43,000 labeled Bangla (Bengali) words, is an essential resource for sentiment analysis and Bangla Natural Language Processing (BNLP) applications [85].
- **BEmoC:** The BEmoC dataset contains 7,000 labeled texts across six emotional categories, assessed with a high Cohen's score of 0.969. The dataset highlights happiness, with fewer instances in the surprise category [86].
- **Part-of-Speech (POS) Tagset:** Data were created and used to train taggers in the Bangla and Hindi languages

TABLE 3. Emotion detection datasets: Sizes, sources, download links, and emotional models.

Dataset	Source	Size	Download Link	Emotion Model
Alm	Stories of 185 children	15,302	1	Dimensional
Aman	Blog post	4,266	2	Discrete
ISEAR (2017)	Worldwide survey on causes and consequences	7,666	3	Dimensional
SemEval (2007)	Headlines	1,250	4	Discrete
SemEval (2018)	Blogs, Tweets	22,761	5	Discrete
SemEval (2019)	Text conversation between two people	38,424	6	Discrete
EMOBANK	Blogs, newspaper headlines, letters, and travel guides	10,062	7	Dimensional
Valence and arousal	Facebook posts	3,120	8	Dimensional
MELD	Dialogues	15,141	9	Discrete
CBET	Tweets	76,860	10	Componential
GoEmotions	Blogs (Reddit)	58,009	11	Discrete, Componential
Amazon	Product Review	8,000	-	Discrete
Amazon Alexa	Customer Review	3,000	12	Discrete
Emotion Lines	FB message discussions and TV shows	1,000	13	Discrete
CrowdFlower	Tweets	40,001	14	Discrete
ISEAR (2018)	Cross-cultural studies	40,785	15	Discrete

during the first annotation phase. The Bangla tagger was trained on 1,457 sentences containing approximately 20,000 words, whereas the Hindi tagger was trained on 2366 sentences, which contained about 45,000 words. The original form of the obtained dataset contained approximately 3,000 sentences and 42,000 words. This dataset was annotated with a broad category of 32 tagsets [87].

- **EmoNoBa:** This dataset is explicitly created for fine-grained emotion detection in the Bangla language. It consists of 22,698 multi-label emotions expressed in noisy Bangla text. These are public comments on 12 distinct topics from three social media platforms [88].
- **BEmoC:** BEmoC, a new Bangla emotion classification corpus, includes 6,243 preprocessed and annotated texts from various sources, with Facebook contributing the most (2,796). The data, organized into training (4,994), validation (624), and test (625) sets, help in the assessment of emotion classification models [89].
- **BMDb:** The dataset for Bangla movie reviews was compiled from various sources, including Facebook groups and pages related to movie reviews, Twitter posts by well-known reviewers, and the Bangla Movie Database (BMDb). The dataset consisted of 800 movie reviews in Bangla. To ensure unbiased polarity, each review was manually annotated by two independent annotators [90].
- **ABSA:** The ABSA dataset comprises comments about the Bangladeshi cricket and encompasses 2,979 rows and five columns, with data sourced and annotated from BBC Bangla. The dataset comprises a comment column and target column that classifies comments as positive, negative, or neutral. This dataset was used to train the sentiment analysis model [91].
- **BRBT:** The BRBT dataset contains 9,337 Bangla sentiment analysis samples featuring both native Bangla and Romanized Bangla texts, the latter written in English. This inclusion facilitates easy typing without a Bangla keyboard and reflects the widespread use of Romanized Bangla in various communication contexts, including government messages. The dataset is currently private for security and ongoing enhancements [92].
- **EmoInHind:** This dataset includes 15 different emotion labels for the dialogue's utterances: anticipation, confidence, hope, anger, sadness, joy, compassion, fear, disgust, annoyance, gratitude, impressed, cautious, surprised, and guilty. Another label, neutral, has been added to the list of annotated emotions [93].
- **BHAAV:** The dataset includes 20,304 sentences from 230 well-known Hindi short stories that span 18 commonly used genres, including historical, mystery, and patriotic, among others [94].
- **Emo-Dis-HI:** It comprises 2,668 sentences taken from news articles about disasters, each of which is labeled with one of the following emotion categories: no emotion, fear or anxiety, joy, disgust, anger, surprise, sadness, sympathy/pensiveness, optimism, etc. [95].
- **AT-ODTSA:** It is a collection of Arabic tweets for open-domain-targeted sentiment analysis. A total of 3,000 samples constituted the dataset. They focused on tourism, media, arts, and sports primarily for the domain base dataset [96].

[1] http://people.rc.rit.edu/ coagla/affectdata/index.html
[2] http://saimacs.github.io/
[3] http://www.affectivesciences.org
[4] http://web.eecs.umich.edu/ mihalcea/affectivetext/#resources
[5] https://competitions.codalab.org/competitions/17751
[6] https://www.linkedin.com/groups/12133338/
[7] http://saifmohammad.com/WebPages/EmotionIntensity-SharedTask.html
[8] http://wwbp.org/downloads/public_data/datasetfb-valence-arousalanon.csv
[9] https://github.com/SenticNet/MELD
[10] https://github.com/chenyangh/CBET-dataset
[11] https://www.tensorflow.org/datasets/catalog/goemotions
[12] https://www.kaggle.com/datasets/sid321axn/amazon-alexa-reviews
[13] https://sites.google.com/view/emotionx2019/datasets
[14] ttps://www.crowdflower.com/wpcontent/uploads/2016/07/text_emotion.csv
[15] https://www.kaggle.com/datasets/faisalsanto007/isear-dataset

- **Icon Dataset:** This dataset was sourced from the icon shared task. It is composed of code-mixed tweets in Hindi and English. A total of 5,525 test samples and 12,936 training samples were provided. There were three feelings in the dataset: neutral, negative, and positive [97].
- **Emotion Dataset:** Six emotions were identified from the dataset of tweets. There were 151,311 messages distributed as follows: 26,364 tweets that expressed happiness, 21,024 for sadness, 29,306 for anger, 19,138 for fear, 35,797 for disgust, and 19,682 for surprise. This was a Hindi-English code mixed dataset [98].
- **Sentiment Dataset:** Also, the dataset is a mixed Hindi-English code dataset. Three categories are used to categorize tweet sentiment: negative, neutral, and positive. There are 3,879 tweets in total on it [99].
- **DravidianCodeMix:** The dataset, which was created from comments on social media, is multilingual and manually annotated for three Dravidian languages that lack resources. More than 60,000 YouTube comments were annotated in the dataset for sentiment analysis and offensive language detection. Approximately 44,000 comments in Tamil-English, 7000 comments in Kannada-English, and 20,000 comments in Malayalam-English make up the dataset [100].
- **BE-CM Dataset:** This dataset includes user reviews for various Play Store apps that are exclusive to Bangla-speaking users. They gathered high-quality code-mixed Bangla-English data from the Google Play Store using a web scraping tool. Approximately, 970,852 samples were taken in total [101].
- **Punjabi-English:** This dataset was compiled from social media remarks and unofficial celebrity interview transcripts. To achieve better code-mixing, the dataset was reduced from a total of over 10,000 sentences to 4,812. The prepared dataset is accessible through an open repository[16] [102].

B. LEXICONS

Lexicons are very important when emotions are detected based on a keyword or lexical approach. In this section, we describe some lexicons that are frequently used to detect emotions.

- **NRC Word-Emotion Association Lexicon:** This lexicon has been manually annotated using Mechanical Turk from Amazon. Eight different feelings were used[17] [103].
- **NRC-10 Expanded:** It provides word alliance for emotions that correspond to the NRC Word-Emotion Association Lexicon's Twitter extension[18] [104].

TABLE 4. Emotion detection datasets from other languages.

Dataset	Size	Language
AETD	10,065 tweets	Egyptian Arabic
IAEDS	1,365 Facebook posts	Iraqi Arabic
DINA	3,000 tweets	Arabic
LAMA-DINA	9,064 tweets	Arabic
AraEmoCorpus	5.5 million tweets	Arabic
AEELex	35,383 tweets	Arabic
BanglaSenti	43,000 words	Bangla
BEmoC	7,000 texts	Bangla
Part-of-Speech (POS) Tagse	20,000 words	Bangla
EmoNoBa	22,698 text	Bangla
BEmoC	6,243 texts	Bangla
BMDb	800 movie reviews	Bangla
BERT	9,337 texts	Bangla
EmoInHind	44,247 utterances	Hindi
BHAAV	20,304	Hindi
Emo-Dis-HI	2,668	Hindi
Icon Dataset	18,461 Samples	Code Mixed Hindi-English
Emotion Dataset	151,311 messages	Code Mixed Hindi-English
DravidianCodeMix	60,000 Youtube comments	Code Mixed Dravidian-English
Punjabi-English	4,812 Samples	Code Mixed Punjabi-English

- **Sentiment140 Lexicon:** It aggregates the positive and negative word scores supplied by the Sentiment140 Lexicon to determine positive and negative factors[19] [105].
- **NRC Hashtag Sentiment Lexicon:** It uses the positive and negative word scores supplied by the NRC Hashtag Sentiment Lexicon to construct positive and negative variables[19] [105].
- **NRC Hashtag Emotion Lexicon:** This vocabulary was produced automatically using tweets that contained emotion-related hashtags, such as happy. Feelings of rage, contempt, fear, grief, anticipation, surprise, delight, and trust are connected to the words in this passage[20] [106].
- **NRC Valence, Arousal, and Dominance Lexicon:** The valence, arousal, and dominance ratings of over 20,000 words were included in this lexicon. The results range from 0 to 1[21].
- **NRC Affect Intensity Lexicon:** For the emotions of wrath, fear, sorrow, and joy, this vocabulary offers real-valued intensity scores[22] [107].
- **AFINN:** This lexicon comprises words that have been carefully assessed for polarity with a number between minus five (negative) and five (positive)[23] [108].
- **Bing Liu Lexicon:** This vocabulary includes a list of words that express both favorable and unfavorable sentiments[24] [109].

[16] https://github.com/gptansh/Bilingual Sentiment Analysis Pun-En
[17] http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm
[18] https://www.cs.waikato.ac.nz/ml/sa/lex.html#emolextwitter
[19] http://saifmohammad.com/WebPages/lexicons.html#NRCTwitter
[20] http://saifmohammad.com/WebPages/lexicons.html#HashEmo
[21] http://saifmohammad.com/WebPages/nrc-vad.html
[22] http://www.saifmohammad.com/WebPages/AffectIntensity.htm
[23] https://github.com/fnielsen/afinn
[24] https://www.cs.uic.edu/ liub/FBS/sentiment-analysis.html#lexicon

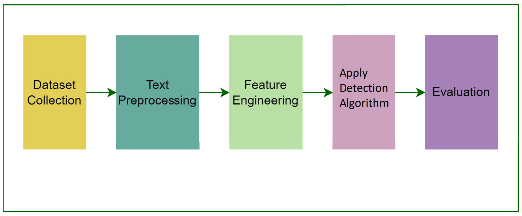

FIGURE 3. Basic steps of textual emotion detection process.

- **MPQA:** This vocabulary contains almost 8,000 subjective single-word cues, each of which is rated as either positive or negative[25] [110].
- **SenticNet:** This is a list of words and their corresponding feelings[26] [111].
- **SentiWordNet:** This lexicon gives every synset of WordNet one of three emotion scores: positivity, negativity, or objectivity[27] [112].
- **WordNet-Affect:** This vocabulary is a useful addition to WordNet. Semantic labels tag a subset of WordNet synsets, which are words that reflect emotions directly or indirectly[28] [113].
- **WordNet:** It is an online dictionary of English words. It organizes words into groups of equivalents called synsets, including verbs, nouns, adjectives, and adverbs[29] [114].
- **EmoSenticNet:** It provides labels for Emotions to Sentic Concept[30] [115].
- **ConceptNet:** ConceptNet is a knowledge graph that uses labeled edges to connect natural language words and sentences. It derives contents from several sources, including expert-created documents, crowdsourcing, and purpose-driven games. ConceptNet is the most recent form of CRIS/most OMCSNet.Version 2.0 has 1.6 million assertions connecting 300,000 nodes[31] [116].

IV. PROCESSES OF EMOTION DETECTION

Emotion detection occurs at various stages. These stages are essential for the creation of better systems. The general steps of the TBED are shown in Figure 3.

A. TEXT PREPROCESSING

The initial phase of the emotion detection process involves text preprocessing. Raw data may contain several unwanted features. The data cannot be trained without cleaning such artifacts, and the desired accuracy cannot be achieved. Text cleaning, tokenization, and normalization procedures are discussed in this section. The most crucial stage in text classification is data preprocessing [117]. To illustrate the idea of the text preprocessing steps, we graphically present them in Figure 4.

1) DATA CLEANING

Certain superfluous items in the raw data may impact the performance of any system. The data-cleaning steps are as follows:

- **Punctuation Removal:** Punctuation is also removed during the cleaning stage. A standard library (e.g., string module in Python) can be used to remove these as it has a predefined collection of punctuation marks.
- **Lower-case Conversion:** The common aspect of preprocessing is converting all characters into the same case. Most of the time, it is preferable to convert them into a lowercase.
- **Stop Word Removal:** Unicode and unwanted strings are regarded as leftovers from the crawling operation, which clutter the data and are useless for text processing. In addition, practically all Twitter (now "X") users include URLs for additional information, user-mentioned tags, and hashtag symbols to relate their tweets to a specific topic or to convey their feelings. These provide supplementary information helpful to humans but are considered "noise" and do not supply relevant data to machines. Researchers have proposed many methods to manage this additional information supplied by users, such as replacing URLs and removing user mentions [118].
- **Replacing Emote and Emoji:** Social media users express their feelings and ideas using a variety of emoticons and emojis. To identify tweets appropriately, it is crucial to record this helpful information. A limited number of tokenizers can record several expressions and feelings and substitute them with their corresponding meanings [119].
- **Slang and Acronym Substitution:** In some social media, such as Twitter (now "X"), there are character limitations. That drives users to write online using acronyms, abbreviations, and slang. An acronym, such as MIA, which means unaccounted for, is a condensed term. On the other hand, slang is a colloquial language often confined to a certain setting or group of individuals and is used to communicate ideas or meanings. To improve performance without losing any information, it is necessary to address this informal word style by substituting it with its true meaning [120].
- **Removing Repeating Characters:** Social media users occasionally purposefully employ prolonged terms, such as "marufffffffff, collllllll," when they purposefully type or add additional letters repeating more frequently. Therefore, classification algorithms should not consider these as separate words. Handling and transforming these terms into their base words are crucial [121].

[25] http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/
[26] http://sentic.net
[27] http://sentiwordnet.isti.cnr.it
[28] http://wndomains.fbk.eu/wnaffect.html
[29] https://wordnet.princeton.edu
[30] https://www.gelbukh.com/emosenticnet/
[31] https://conceptnet.io/

2) TOKENIZATION

Tokenization involves fragmenting a sentence into its constituent words or tokens. Sentences can be transformed into tokens, whereas records can be labeled as sentences. A text sequence is separated into letters, signs, phrases, or characters during tokenization [122]. Its primary objective is to distinguish the characters within a statement. Tokenization is typically the initial step in emotion detection work [123].

3) NORMALIZATION

- **Stemming:** This is also known as the standardization phase because it involves stemming or reducing words from their root or base form. For example, the words 'emotional,' 'emotions,' and 'emotion' all stem from the base word 'emotion.' Stemming can negatively impact the meaning of words by reducing them to their root form, which can sometimes result in non-standard or meaningless English terms. Stemming reduces wording to the stems and removes the suffix of the word according to certain grammatical rules.
- **Lemmatization:** It ensures that the meaning of a term is preserved. Lemmatization exploits predefined terminology to shorten terms by recording the context of words and checking them against definitions.

4) DATA AUGMENTATION

Deep learning requires enormous amounts of data for the training and testing phases to perform tasks efficiently [124], [125]. This can play a crucial role in text categorization and emotion detection [126]. Back translation, word embeddings, thesaurus, substitution of synonyms, text creation, etc., are methods for data augmentation. As an illustration, a text classifier combining semi-supervised learning and data augmentation based on a deep neural network was constructed using free-text data from a survey focused on sleep-related concerns [127]. According to several studies, data augmentation enhances the pre-training performance of the model, ultimately improving the results. One popular approach for improving data is the use of "generative adversarial networks" (GAN). This novel architecture pits neural networks against each other to produce fresh synthesized data samples that resemble accurate data. For instance, in [128], because of the unbalanced dataset, data augmentation was used to detect emotions in Spanish.

B. FEATURE ENGINEERING

Feature extraction from textual data is vital for ML in identifying and interpreting data as humans do. It involves transforming unprocessed input into numerical data that machines can read. A "feature vector" is another name for extracting features from raw data. Without a large corpus, it is difficult to extract solid word representations from English because of the difficulty in conveying thoughts, feelings, and intentions. Academics now have access to massive amounts of data from social media networks. Categorizing

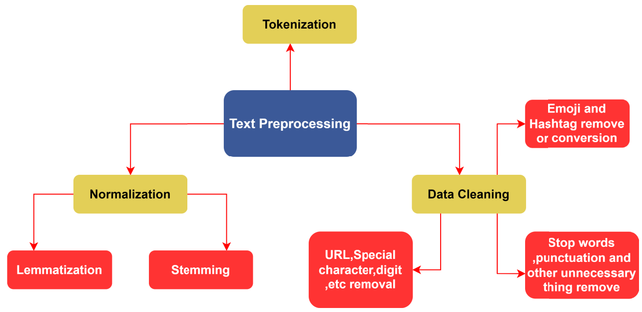

FIGURE 4. Text preprocessing techniques. (Note: Data augmentation is omitted from this diagram because it could be optional.)

this enormous amount of social media data is challenging. To simplify the annotation process, researchers first looked for signs of mood and emotion within the text's content, such as emojis and hashtags [129], [130], [131]. Figure 5 depicts the feature extraction strategies. A summary of these feature extraction models is presented in Table 6.

1) CLASSICAL FEATURE EXTRACTION METHODS

The following methods were used to convert textual data into vectors containing the document's word count. Classical models have two types of word representation systems.

- **One-hot Encoding:** This is a popular technique for text representation. The height of the lexicon corresponds to the total number of phrases it contains. Each expression in the dictionary represents an individual binary variable, including 1 or 0, meaning that every word comprises ones and zeros. The index of the linked term is indicated by 1, while all others are marked with 0.
- **Bag-of-words (BOW):** This is a more advanced type of one-hot encoding. It includes all sentence's single-hot word representations. A technique for collecting characteristics from text sources is called the "bag of words." These characteristics can be used to train machine learning systems. It develops vocabulary from every singular phrase found in the training corpus.
- **Part of Speech (PoS) Tagging:** The word type "PoS" serves as an example of its use and helps in the natural language processing (NLP) task. It comprises adjectives, adverbs, verbs, and nouns.
- **Word n-gram:** The "n-grams" characteristic is a group of n identical words or letters that are thought to be particularly helpful for classifying text. Consider the following sentence. An example of n-grams is presented in Table 5.
- **TF-IDF:** TF-IDF is the combined form of term frequency and inverse document frequency. The TF-IDF vectorizer is a prime example of an acceptable input format because it uses raw integers to determine if words are present in a text [132], [133]. The term frequency refers to how often a particular word (term) appears in a document. On the other hand, inverse document frequency considers the count of the documents in

TABLE 5. Example of n-grams.

N-grams type	Value of n	Output
Unigram	1	"Emotion," "detection," "from," "text"
Bigram	2	"Emotion detection," "detection from," "from text"
Trigram	3	"Emotion detection from," "detection from text"

the corpus in which that particular term occurs. The calculation of the TF-IDF score is given in the equation below.

$$TF\text{-}IDF(t, d, D) = TF(t, d) \times \log(\frac{D}{df_t}) \qquad (1)$$

Here, t is a particular term, D is the set of documents, and df_t indicates how many documents include the term t.

- **Count Vectorizer:** The count vectorizer uses a BOW approach to circumvent textual structures. Instead, information is extracted using only word counts. Therefore, every text is first transformed into a vector form. Vector input determines the frequency of a specific word appearing in the material [134].

2) WORD EMBEDDING METHODS

The classical methods have limitations, and the researchers chose to learn the distributed word representation in the lower range [135]. Input representations significantly affect the performance of ML models [136]. Deep learning (DL) is replacing traditional feature learning approaches. Word embedding is a patterning method of learning that links a vocabulary word to an N-dimensional vector. In other words, it is a trained representation that enables words with similar meanings to share similar representations. Embedding methods preserve word order and capture the meanings of expressions. They have proven to be useful in many NLP applications.

- **Word2vec:** Mikolov et al. [137] created the Word2vec word representation model. This model includes two hidden units in an artificial neural network to generate a vector for each word. The semantic and grammatical information is meant to be contained in the Word2vec skip-gram and Continuous Bag of Words models. As a result, it is advisable to use a large corpus to represent words better. Based on this context, Word2vec carries out prediction using one of two neural network models, namely, continuous bag of words (CBOW) and skip-gram.
- **FastText:** FastText was introduced by Bojanowski et al. [138] and is the foundation of its CBOW. Compared to previous algorithms, FastText reduces training time while maintaining performance. Instead of using whole words as input, FastText breaks down words into n-grams so that a neural network that can recognize word semantics and understand the relationships between letters can process them. FastText produces improved outcomes with better word representations, particularly for unusual words.
- **Context2vec:** Melamud et al. [139] introduced this model in 2016 based on the CBOW model of Word2vec. However, they replaced the transitional term encoding in a fixed window with a more powerful bidirectional long short-term memory (bi-LSTM) neural network. Using a sizable text corpus, a neural network that incorporates targeted phrases and sentence context in a lower weight was trained. It is then modified to reflect the relationships between the target and the complete sentential environment.
- **CoVe:** The model was built using Context2vec. This method was developed by McCann et al. [140]. They started with GloVe word vectors. Bi-LSTM with two layers was trained for sequential translation. The outcome of the serial encoder is given the name CoVe, integrated with GloVe vectors, and used in a more detailed form later on via transfer learning. Instead of utilizing Word2vec or other methods, CoVe was constructed by applying machine translation.
- **ELMo:** This model uses a bi-LSTM technique, meaning that the words are used in both directions before and after. Peters et al. [141] introduced this approach, which is a contextual embedding method that calculates the words around it. This character-based language model can conceive of the preceding or next words and those not found in the dictionary. It improves approaches to challenging NLP issues, such as sentiment analysis and textual context question answering. It also examines the semantic and grammatical characteristics of words and how they change in different linguistic contexts, such as the development of lexical ambiguity.

C. EMOTION DETECTION ALGORITHMS

After finishing text preprocessing and generating feature vectors, the next important step is detecting (classifying) emotions. Different categories of text-based emotion detection algorithms are described in this section.

1) KEYWORD-BASED APPROACH

This is the easiest and most effective way of detecting emotions in a piece of writing from a sentence. According to this approach, a text reflecting one or more emotions is considered as the input and divided into tokens. Subsequently, word counting is used to determine which emotional words are repeated. Finally, the output, which contains the expressive class, is provided for this input. Keyword matching occurs between a predetermined keyword list and the emotional words in the text. The effectiveness of the emotion keywords is then assessed. Unfortunately, this approach can fail when feelings are weakly correlated with terms. The steps involved in using this strategy are shown in Figure 6.

2) LEXICAL-BASED APPROACH

The likelihood or weight of an unbroken word for a certain emotion is determined using the lexical/corpus-based

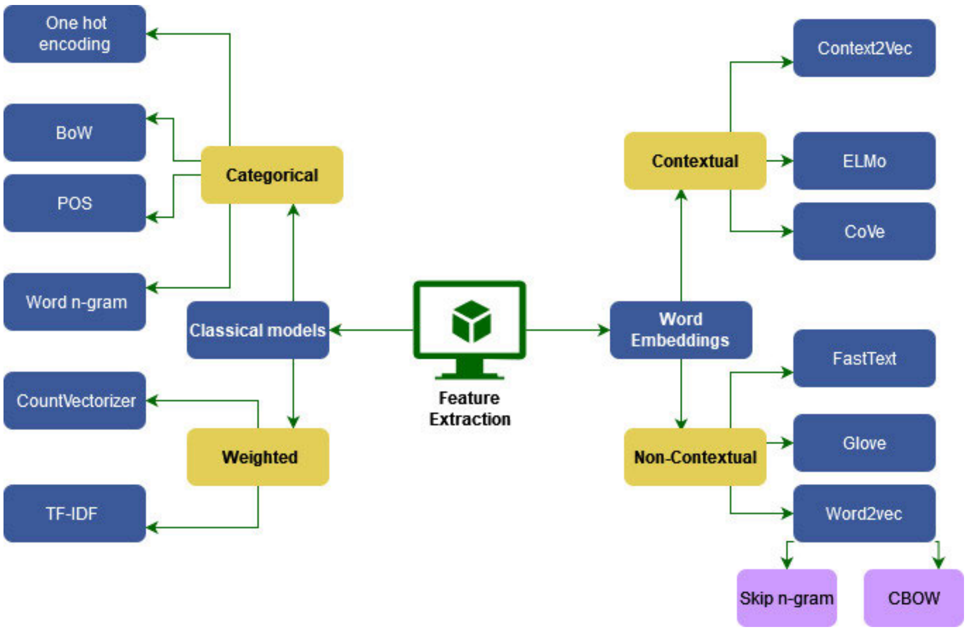

FIGURE 5. Overview of feature extraction techniques.

TABLE 6. Summary of classical and word embedding models for feature extraction.

Model	Architecture	Pros	Cons	Type
One Hot Encoding, PoS, Word n-gram, BOW	-	Easy to calculate; Compatible with unknown words, and using a fundamental metric to extract phrases.	Misses out on semantic and grammatical information; Cannot capture the feeling of words; Common words impact the outcomes.	Count based
TF-IDF, Count vectorizer	-	Easy to calculate; Basic measure for extracting descriptive words.	Semantic and grammatical information is not captured; Unable to express feelings in words terminology.	Count based
Word2vec	Log Bilinear	Records the text's semantics and syntax; Trained using a large corpus (pre-trained).	Contextual information is not captured; OOV (out-of-vocabulary) terms are not captured; Learning requires a large corpus.	Prediction based
GloVe	Log Bilinear	Uncovers sub-linear correlations; Induces interaction using vectors in the vector space; Less weight will not impact how well training goes for typical word pairings like stop words.	Contextual information is not captured; Large memory is needed for storage; OOV words are not captured; Learning requires a large corpus (pre-trained).	Count Based
FastText	Log Bilinear	Works well with uncommon words; Takes care of OOV terms.	Contextual information is not captured; Much memory is consumed for storage; Computationally more expensive than GloVe and Word2vec.	Prediction based
CoVe, ELMo, Context2vec	Bi-LSTM	Addresses the issue of contextual information; Improves efficiency.	Extensive computation required; For LSTM and feed-forward layers, different word embeddings are used.	Prediction based

method, which is an enlarged keyword-based technique. It can be divided into two categories: ontology-based emotions and consequences based on language. Branchlet emotions are at the bottom of the emotion ontology with low weights, whereas fundamental emotions are at the top with high weights [142]. In Section III-B, we have covered some lexicons. Researchers typically use these lexicons to detect emotions following the workflow outlined in Figure 7. Finally, negation is performed if needed. Sentence level emotion word classifier algorithm is listed as follows:

The studies based on lexical approaches are summarized in Table 7.

3) MACHINE LEARNING APPROACH

There are some constraints in both linguistic and keyword-based approaches to overcome the fact that this approach pursues data on its own. It also seeks a relationship between the specified text and the related out-turn of emotion by constructing a predictive model rather than the requisite direct link to emotion from the inserted text. The steps of

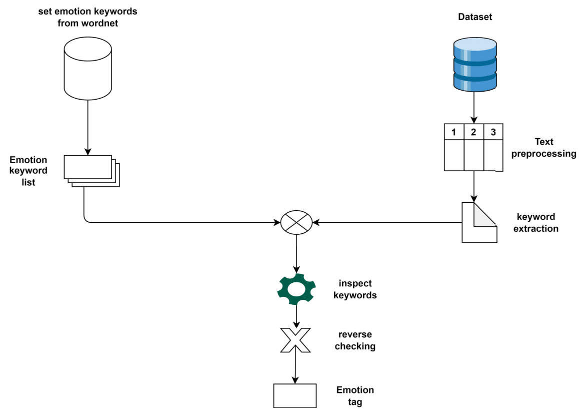

FIGURE 6. Keyword approach workflow.

while *(Sentence in the dataset)* **do**
  Do preprocessing;
  if *(Word is found in the lexicon)* **then**
    Perform scoring of emotion words calculation and aggregation;
    Select the emotion category of emotion words with the greatest score;
  end
  if *(Sentence has a negation)* **then**
    Perform classification using a negation classifier;
  end
end

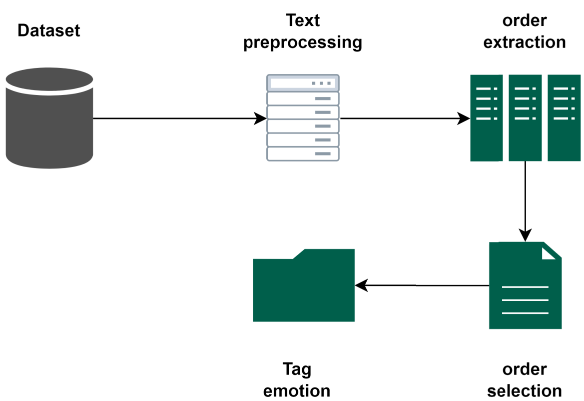

FIGURE 7. Lexical-based approach workflow.

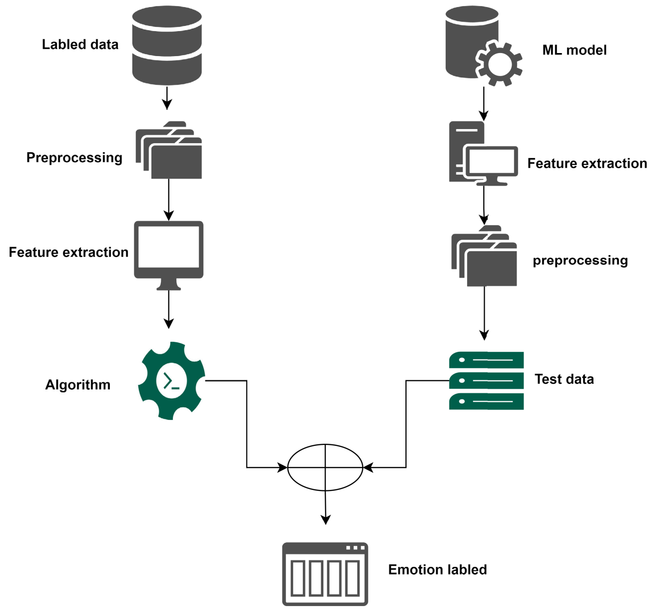

FIGURE 8. Machine learning approach workflow.

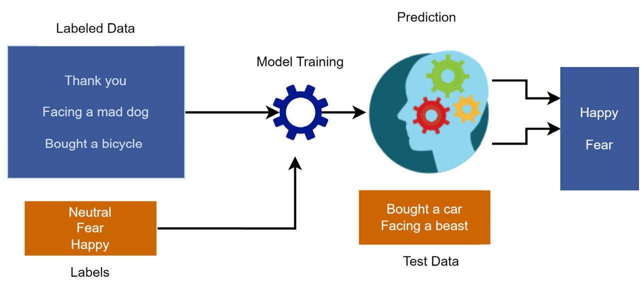

FIGURE 9. Process of supervised learning (classification).

the machine learning (ML) approach are demonstrated in Figure 8. They can generally be divided into two types.

- **Supervised learning:** This method is based on a dataset tagged with class labels. These data are used for training the emotion classifier. The data are trained explored, and the model is constructed. The rest of the dataset is classified into emotion categories based on the trained learning model. The training and the testing phases are separated in a simple ML setup. In the training phase, samples from the training data are utilized as input, with features learned by the algorithm used to create a learning model [155]. The process of predicting the test data is carried out using the learning model. The emotion classes (labels) that have been originally tagged to the test instances are compared against the predicted labels. The learning process is illustrated in Figure 9.
- **Unsupervised learning:** This is an ML method in which the models are built from a training dataset without class labels. However, models uncover hidden patterns and insights into the presented data. It is analogous to learning new things in the human brain. Unsupervised learning is a type of ML in which models are trained on unlabeled datasets and then run unsupervised. This technique presupposes ML algorithms with flaws where a large dataset is required for accurate training processing. Unlike supervised learning, unsupervised learning includes input data but no corresponding output data; therefore, it cannot be used immediately for regression or classification. Although some studies have been conducted on the application of unsupervised learning in TBED [156], [157], we will not focus on them in this survey.

In this survey, we focus on supervised learning (particularly classification) models used in TBED. The classifier models can be divided into four broad categories.

TABLE 7. An overview of the studies using the lexical approach.

Ref.	Language	Dataset	Text Preprocessing	Model	Result	Evaluation Metric
[143]	Vietnamese	VnEmoLex	Standardizing words, Conversion of emoticons into emoji	BERTology models	XLM-R: 67.03%, 88.29%, 68.29%, 88.08%	Accuracy, Macro and Weighted F-score
[144]	English	Blogs of 135	Cleaning and Conversion	Emotion Ontology	79.57%, 88.08%	Accuracy, Precision
[145]	English	myPersonality	-	NRC lexicon	-	Pearson Chi
[146]	English	ISEAR, SemEval and Twitter	-	Optimization framework	61%, 90%, 63%	Precision, Recall, F-score
[147]	English	Twitter, news, and blogs	-	GPELs, DSELs	52.84%, 56.05%	F-score
[148]	English	-	Tokenization	KA, KNA	80%	Accuracy
[149]	English	Manually created corpus	-	Sentence emotion extractor	93%, 85%, 89%	Precision, Recall, F-score
[150]	English	Aman's, Alm's	-	SVM	-	Precision, Recall, F-score
[151]	English	Site of ictclas.org news	Segmentation	POS	63.27%	Pearson correlation coefficient average
[152]	English	Task organizer	Tokenization, Lowercasing, PoS	SVM	49%, 64%, 52%	Macro and Micro F-score, Accuracy
[153]	Egyptian, Gulf Arabic	Egyptian Company named RDI	-	SATALex	69.6%, 80.3%, 73.5%, 86.8%	Precision, Recall, F-score and Accuracy
[154]	Arabic	Arabic senti-lexicon, MASC	Tokenisation, normalization, stop word removal and stemming	NB, KNN, SVM, LR and NN	-	Macro-F1

(a) Conventional machine learning,
(b) Ensemble learning,
(c) Deep learning, and
(d) Transformer (distinct sub-category of deep learning)

Figure 10 overviews the taxonomy (categorization) of machine learning classifier models used in TBED.

a: CONVENTIONAL MACHINE LEARNING

Among the many conventional ML algorithms, those most frequently employed for classifying emotions using textual data are described below.

- **Decision Tree (DT):** DT is a graph that shows choices and outcomes in the shape of a tree. The nodes in the graph indicate an event, whereas the edges reflect the criteria or rules for decision making. The DT classifier sets a threshold for classification accuracy [158].
- **Random Forest (RF):** The RF technique comprises several DTs, allowing each to forecast the class's outcome independently. The entire Random Forest process determines which class receives the most "votes." This method has several benefits, including the ability to control fitting problems and the reduction of prediction mistakes that may arise when using a single individual tree alone [159].
- **Support vector Machine (SVM):** SVM is another of the most commonly used modern ML methods. The SVM in machine learning is a supervised learning model with corresponding learning algorithms that examine the data used for regression and classification analyses. A technique for automatically detecting emotions was implemented using an SVM in [160].
- **k-Neatest Neighbors (KNN):** KNN is a technique for categorizing an item based on training data nearest to the item. Problems involving classification and regression can be resolved using the KNN technique. Although simple to use and comprehend, a major drawback of this method is that it becomes noticeably slower as the amount of data used increases [161].
- **Naive Bayes (NB):** The NB classification method, which is based on the Bayes Theorem, assumes that predictors are independent of one another. NB mostly focuses on the text categorization sector. NB has various benefits, including ease, speed, and accuracy. Depending on the occurrence probability, it is mostly utilized for clustering and classification purposes [162], [163].
- **Logistic Regression (LR):** To predict the probability of the target variable, supervised learning uses a classification process known as LR. LR is used to divide the data into two categories. One is binary, and the other is a multi-class LR. This method can solve classification problems using linear regression [164], [165].

A summary of the studies on TBED based on conventional ML is presented in Table 8.

b: ENSEMBLE LEARNING

Data scientists across many industries use ensemble learning, which is a complex machine learning algorithm. The strength of ensemble learning methods is their capacity to combine forecasts from various machine learning models. Combination mechanisms include maximum voting, averaging, and weighted averaging. Although individual baseline classifiers offer low accuracy and tend to overfit, an ensemble of

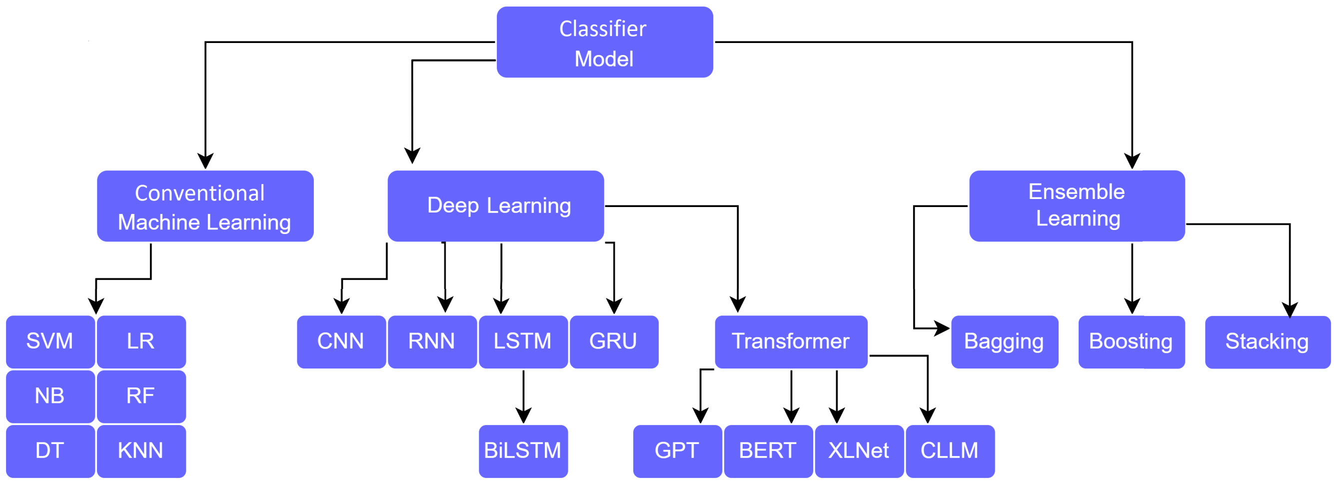

FIGURE 10. Taxonomy of machine learning classifier models used in TBED.

these base classifiers can reduce overfitting and improve accuracy [190]. In addition, ensemble learning offers good generalization and robustness [191]. Ensemble models, including Random Forest, AdaBoost, and Gradient Boosting, have been applied to detect emotions [192].

Ensemble learning techniques can generally be categorized into three types: bagging, boosting, and stacking, as illustrated in Figure 11. Bagging merges the conclusions of numerous models to produce a more generalized outcome. In the bagging procedure, subsets provide a clear image of the distribution [193]. The boosting technique boosts overall performance by combining several weak learners to create a strong learner [194].

Maruf et al. [190] used bagging and boosting ensemble models. Initially, the authors used the baseline classifier and found that the accuracy was low and overfitting occurred. Overfitting was reduced when the ensemble method was used. They used several baseline classifiers as the base estimators. The authors also showed that the performance of the ensemble methods mostly depends on the parameters.

While bagging and boosting use homogeneous base learners, stacking often utilizes heterogeneous base learners, trains them in parallel, and combines them by employing a meta-learner to produce a final prediction based on different base estimator outputs. An example of stacking different classifiers for TBED can be found in [191].

A combination of LSTM and the Ada boosting ensemble method showed the best result in detecting offensive text from Bangla text [195]. Using the ensemble technique improves the system performance by 44% [196]. Instead of choosing one classifier and decreasing the accuracy of a given language corpus, an ensemble learning technique has the benefit of pooling the effects of all the classifiers.

According to experiments in [197], within-corpus testing increased the accuracy of the Urdu corpus by 13%, the German corpus by 8%, the Italian corpus by 11%, and the English corpus by 5%. When using Urdu data for training and testing and German data for comparison, cross-corpus trials

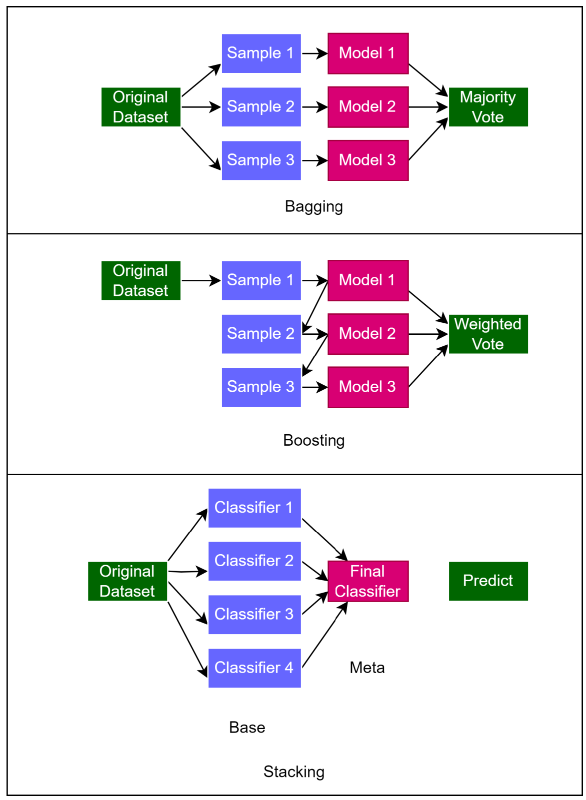

FIGURE 11. Architecture of bagging, boosting and stacking ensemble learning.

showed 2% and 15% improvement, respectively. Accuracy increased by 7% when testing with German data and training with Italian data, 3% when testing with Urdu data and training with Italian data, and 5% when testing with Urdu data and training with English data.

Table 9 summarizes the ensemble models used in TBED.

c: DEEP LEARNING

Deep learning (DL) is a distinct ML subcategory. It has become widely used, outperforming conventional machine

TABLE 8. Overview of TBED papers using conventional machine learning approach.

Ref.	Language	Dataset	Text Preprocessing	Model	Result	Evaluation Metric
[166]	Bangla	ABSA Dataset	Normalization, tokenization, and stemming	BTSC algorithm and Baseline ML	82.21%	Accuracy
[167]	Arabic	SemEval-2018	Removing stop words, repeating chars, English characters, mentions, punctuation marks, Arabic diacritics	DT, KNN NB, SVM	DT, KNN-74%, SVM-63%	Accuracy
[168]	Bangla	Movie and short film scripts	Remove punctuation, stop words and punctuation	DT, KNN, RF, MNB, SVM LR, Stochastic gradient descent	SVM-81.68%, 92.24%, 86.64%, 85.59%	Precision, recall, F-score accuracy
[169]	Hindi	Social media websites	Remove stop words and special symbols	RF, SVM	RF-88.77%, 87.97%, 88.53%, 91.07%	Precision, Recall, F-score Accuracy
[170]	Chinese	Chinese and RenCECps corpus	-	EDL	-	Chi-square
[171]	English	Twitter	Cleaning, Stemming	SMO, NB	NB- (44.2%, 45.1%, 44%)	Precision, Recall, F-score
[172]	English	Twitter	-	SVM, RF, NB, KNN	SVM-81%	Accuracy
[173]	Chinese	Sina News Tencent news	-	LDM, SVM, NB, LR	63.5%, 63.4%	Precision, Recall, F-score
[174]	English	BBC, Myspace, Twitter, Youtube Runner	-	TME, WME	TME-85.01%, 31%	Precision, Accuracy
[175]	Chinese	weibo.com	-	Baseline, Multi-view learning	Multi-view-48.6%	F-score
[176]	English	Grimms', H.C. Andersen's, B. Potter's stories	-	NB	63%	10-fold validation (Accuracy)
[177]	English	3600 samples	-	SVM, DT, NB, KNN	All above 80%	Precision, Recall F-score, Accuracy
[178]	English	Twitter	Stop words removal, Cleaning	NB, SVM	NB-75.7% 75.26%, 75.34%	Precision, Recall, F-score
[179]	English	WASAA	Lowercasing, Tokenization, Stop words removal, Stemming	-	-	Precision, Recall, Accuracy
[180]	English	Twitter	Stemming, Stop word removal, Tokenization	NB, SVM	NB-92.5%	Accuracy
[181]	English	YouTube comments	Stemming	SVM	92.75%, 68.82%	Precision, Accuracy
[182]	English	Twitter message	Cleaning	SVM, KNN	KNN-90 %	Kappa Coefficient
[183]	English	Twitter message	Cleaning	SVM, KNN, DT, NB	KNN-Above 90 %	Precision, Recall, F-score
[184]	English	105 Tweets	Lowercasing, Cleaning	NB, SVM	SVM-77.55%	Accuracy
[185]	English	Newspaper headline	Tokenization, Stop words Removal Stemming	NB, SVM, DT, RF	SVM-55.91 %	Accuracy
[186]	English	Blogs of135	Cleaning, Conversion	SVM, NB	73.89%	Accuracy
[187]	English	Twitter	Cleaning	NB, SVM, DT	DT-0.1%, 89.9%, 90%	Precision, Recall, F-score
[188]	English	Autism Center	-	SVM, NB, DT	SVM-80%, 75%	Precision, Recall
[189]	English	ISEAR	Tokenization	SVM	80%	F-score

learning techniques. The primary objective of the DL method is feature learning. It is a multi-layer nonlinear processing method for learning complex feature representations based on the original feature input. DL can design new classifiers or generate tools using automated learning's feature representation and accomplish domain-oriented classification or other activities when paired with specialized domain tasks. In deep learning, a system learns to complete classification tasks from text. DL algorithms learn high-level features from the data. The approach to solving this

TABLE 9. Overview of TBED papers using ensemble learning approach.

Ref.	Language	Dataset	Text Preprocessing	Model	Result	Evaluation Metric
[198]	Arabic	SemEval2018 -Ar-Ec	Removing English characters, numbers, stop words, repeating chars, punctuation marks, and Arabic diacritics, normalization and emojis replacement	Ensemble(Bi-LSTM, Bi-GRU, MARBERT)	54%, 63%, 55%, 52%, 70%	Jaccard Score, Precision, Recall, F-score (micro macro)
[199]	Arabic	OSACT-2022	Eliminated non-Arabic letters, punctuation marks, digits, removal of unnecessary characters and normalized	Ensemble (LightGBM, MARBERT, QARiB, AraBERT-B-T)	87.044%	Macro-F1
[200]	Bangla	From 20 participants	Noise reduction, audio transcoding, stereo to monaural, frequency downsampling	Ensemble (MLP, KNN RF, DT, SVM)	84%	Accuracy
[201]	Bangla, Hindi	Twitter	Remove irrelevant characters	Heterogeneous ensemble model	62.63%	Accuracy
[192]	English	-	-	RF, Adaboost Gradient Boosting	-	-
[202]	English	ISEAR, OANC, CrowdFlower	Noise elimination, Cleaning, Tokenization, Lemmatization	Ensemble (KNN, MLP, DT)	82%, 92%, 81%, 88.49%	Precision, Recall, F-score, Accuracy
[196]	English	IMDB, Sentiment140 SemEval-2017	Cleaning	RF, Adaboost, Gradient Boosting	Stacking-86%, 90%, 88%, 87%	Precision, Recall F-score, Accuracy
[191]	English	6328 sentences	cleaning	Stacking	85.49%, 91.21%	Accuracy, AUC
[190]	English	Social Media	Cleaning, Segmentation, Tokenization, Lemmatization	Bagging, Boosting	93%, 94%, 94%, 90.45%	Precision, Recall, F-score, Accuracy
[198]	English	SemEval	Cleaning, Replacing emoji	Combined MARBERT transformer, Bi-LSTM, and Bi-GRU	63.4%, 55% 52.7%, 54%	Precision, Recall, F-score, Accuracy

problem is end-to-end. DL has been used for large datasets and accurate systems. The most popular DL techniques are based on artificial neural networks (ANN). DL is defined as a model (e.g., a neural network) with several layers taught layer-by-layer. The approach to the deep learning workflow is shown in Figure 12.

The deep learning classifiers used primarily in TBED are briefly described below.

- **Convolutional Neural Network (CNN):** Convolutional layers provide outputs by combining numerous filters after filtering the incoming data to create features. The quality of the layered features is decreased by pooling or down-sampling to increase the CNN's tolerance for noise and distortion. CNNs are often used in classifiers and NLP tasks [203].
- **Recurrent Neural Network (RNN):** The core purpose of an RNN is to process successive inputs using memory obtained through directed cycles. Unlike conventional neural networks, RNNs can recall and apply prior knowledge to source states in the input sequence. Early data points in a series are assigned higher weights in RNN, which employs a specific form of neural network design. This technique correctly categorizes text and sequence data [204].

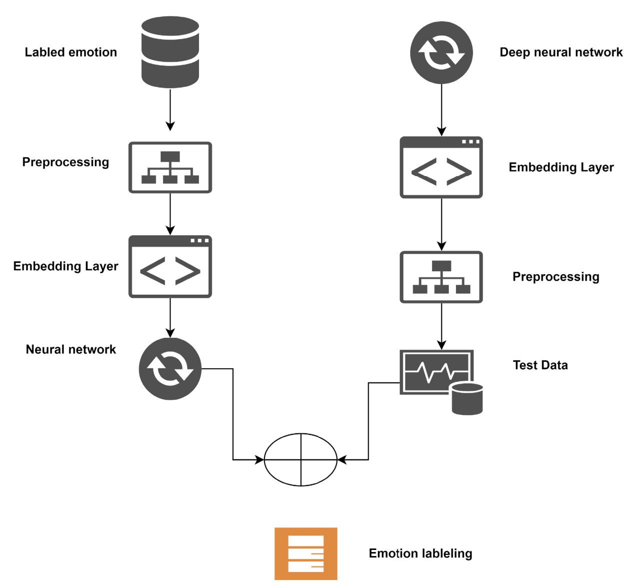

FIGURE 12. Deep learning approach workflow.

Long Short-Term Memory (LSTM): LSTM, a subtype of RNN, excels at addressing problems involving long-term dependencies more effectively than conventional RNN. This is achieved by incorporating specialized mechanisms that facilitate the retention of information over extended sequences, thereby overcoming the limitations associated with vanishing gradients in traditional RNNs. RNN and LSTM both have a chain-like design; however, LSTM utilizes several gates to limit the amount of information allowed through every node tightly. LSTM and its variants have been demonstrated to be highly beneficial in TBED [205].

Gated Recurrent Unit (GRU): An easier design than LSTM is the GRU architecture. While LSTMs have relatively complicated internal memory, a GRU only has two gates. A typical challenge with an RNN is the hidden gradient problem, which is resolved using a GRU [204]. The model uses the update gate to determine the amount of historical data from earlier steps that should be included in subsequent stages. The reset gate determines the quantity of deleted data.

A summary of the studies on TBED based on DL is presented in Table 10.

d: TRANSFORMER

The transformer model breaks new ground in the field of NLP research at a breathtaking rate by offering a significant answer to persistent issues with sequential manipulations. (Although transformers can be viewed as a sub-category of deep learning, we present it here as a separate sub-section because of its high importance.) The encoder and decoder blocks constitute the model, and a softmax activation function is used to normalize the output probabilities. The input of the model is a series of data. The input words are embedded and then sent through positional encoders, which provide word vectors based on where they are in a phrase, thereby extracting the context of the input words. The transformer model, which was created for machine translation, is now utilized for language modeling, making it relevant for additional NLP applications, including text classification, document summarization, and question answering.

- **GPT:** The following highlights the advantages of the GPT model. When GPT is used, lexical robustness is enhanced. The pre-trained GPT model can be fine-tuned without model change to carry out additional tasks. For various domain-specific language modeling tasks, the GPT model outperforms other models trained on domain-specific datasets and generates SOTA outcomes. Fine-tuning is not necessary for GPT-2 and GPT-3. The resource-intensive nature of the GPT model makes the pre-training stage expensive, which is a weaknesses of the model. Additionally, the model frequently struggles with its inability to model dependencies that are larger than specified fixed lengths. The GPT, GPT-2, and GPT-3 models use transformer architecture to anticipate the following tokens or sentences in a sequence. While GPT-3 makes a huge step forward in terms of scale and performance by adding alternately dense and sparse attention patterns in its transformer layers, GPT-2 enhances the original GPT with varied model sizes and normalization layers.
- **BERT:** By receiving training in Next Sentence Prediction (NSP) and Masked Language Modeling (MLM), BERT can comprehend language. To learn the bi-directional contexts of the sentences, BERT used MLM as a blinder. Consequently, it uses certain random sentences as input, masks some of the words within, and then reconstructs the masked words from the surrounding texts at the output. It achieves NSP because of its capacity to simultaneously input two sentences and detect whether the second sentence follows the first one. This capability enables the model to preserve text associations across large distances. The BERT-base and BERT-large models are the two available variations in BERT. The BERT-base model consists of 12 layered transformer encoder blocks, each of which has 768 hidden layers and 12 head self-attention layers, producing approximately 110 million parameters. In contrast, BERT-large consists of 24-layered transformer encoder blocks with 24-head self-attention layers in each block, producing a total of approximately 340 million parameters.
- **Cross-lingual Language Models:** The BERT model requires monolingual data to achieve promising outcomes in NLP tasks. By pre-training cross-lingual models for NLP, the Cross-Lingual Language Models (XLMs) enhance BERT for classification and translation tasks [225]. This model extends the MLM provided in BERT using the model translation idea proposed by Lample et al. [226]. Training both the MLM and TLM while switching between them is required. The corpora contain identical sentences written in two different languages. One sentence from each language corpus is simultaneously supplied to the input of the model. To predict masked words, the model explores texts in both languages and each individual language to gain context.
- **XLNet:** The PLM implementation in the model provides the XLNet model with the capacity to extract contextual information, and the model is known to outperform BERT in a wider range of language modeling applications. Most notably, it eliminates the fixed-length limitation of BERT and can be trained beforehand to reach the SOTA. However, these benefits are accompanied by some computing challenges, making XLNet a computationally intensive system.
- **AraBERT:** The pre-trained contextualized text representation model and BERT-based AraBERT model are pre-trained for Arabic. The prototype was trained using four more manually crawled news websites, the 1.5 billion-word Arabic Corpus, the OSIAN Corpus, Assafir news articles, and Arabic Wiki dumps. AraBERTv1, AraBERTv02, and AraBERTv2 are some of the available versions. The difference lies in the use of pre-segmented text in AraBERTv1, where prefixes and

TABLE 10. Overview of TBED papers using deep learning approach.

Ref.	Language	Dataset	Text Preprocessing	Model	Result	Evaluation Metric
[206]	Arabic	Twitter	Remove all non-Arabic words, Remove hashtags, URLs, replies, emojis, digits, punctuation, non-meaningful stop -words, Eliminate repetitive words	Bi-LSTM	78%, 83%, 81%, 83%	Precision, Recall, F-score, Accuracy
[207]	Urdu	Urdu Nastalique Emotions Dataset	Normalization, Tokenization	MLP, LSTM, Bi-LSTM	82%, 88%, 85%, 85%	Precision, Recall, F-score, Accuracy
[208]	Bangla	Youtube, Twitter, and Reddit	Stop word removal, Tokenization, slang expansion	BiGRU	86.52	Accuracy
[93]	Hindi	EmoInHindi	-	C-Attention-Trans, C-Fourier-Trans	65.98%, 66.52%, 66.24%, 67.14%	Precision, Recall F-score, Accuracy
[209]	English	MOUD	-	RNN, CNN	CNN-61.25%	Accuracy
[210]	English	Poetry text	Stop word removal, Tokenization, Lowercasing	CNN, LSTM, Bi-LSTM Attention based C-BiLSTM,	Attention based C-BiLSTM-88%	Precision, Recall, F-score
[211]	English	SemiEval	Normalization, Punctuation removal	NB, SVM, MLP	NB-99.90%	Accuracy
[212]	English	Shared task public data	Spelling correction, Special character removal	LSTM	81.61%, 74.7%	Recall, F-score
[213]	English	Twitter	Stemming, White space removal	LSTM	71.89%	F-score
[214]	English	SemEval	Tokenization, Normalization, Lowercasing	Attention-base RNN	59%	Accuracy
[215]	Persian	Bijan kha's	Segmentation, Removal of non-Arabic characters and punctuation	GRU, SVM, NB	GRU-97%	Accuracy
[216]	English	Isear, SemEval	-	SVM, RF, NB, LSTM, RNN	All performed better with fastText word embeddings	Precision, Recall, F-score
[217]	English	Twitter	Punctuation removal, Lowercasing	LSTM	98.86%, 99.04%	Recall, F-score
[218]	English	-	Tokenization, Lemmatization, Stop word removal, Lowercasing	CNN	72.43%, 72.54%, 72.48%	Precision, Recall, F-score
[219]	English	SemEval 2018	Tokenizing, Normalization, Spelling correction, Segmentation	GRU	58.9%, 70.1%, 55.50%	Micro and Macro F-score
[220]	English	IMDB movie review data	Removing non-ASCII characters, Cleaning, Lowercasing	CNN, RNN, LSTM, SVM	CNN-89.3%, 87.5%, 88.7%	Accuracy, Recall, F-score
[221]	English	NLPCC	Cleaning, Removal of numbers and special characters	GRU, BiGRU, CNN, LSTM, Bi-LSTM	Bi-GRU-85.35% 57.83%, 68.94%	Precision, Recall, F-score
[222]	English	Online website	-	-	-	Sensitivity, Specificity, Precision, Accuracy
[223]	English	Arabic tweets	Cleaning, Tokenizing, Stemming	CNN, LSTM	Both works best	Accuracy
[224]	English	Sina Weibo	-	CNN, SVM, RNN, LSTM	CNN-97%	Accuracy

suffixes are divided using arasaSegmenter, an Arabic-specific text segmenter. AraBERT is assessed using three tasks: sentiment analysis, NER, and QA. AraBERT is composed of 12 heads, 768 hidden layers, and 12 layers [227].

- **ArabicBERT:** ArabicBERT uses a corpus larger than the previous AraBERT. A recent version of Arabic Wikipedia, OSCAR, and other Arabic resources were used to pre-train the models. ArabicBERT comes in four sizes: small, medium, base, and large, based on the size of the architecture. There are 256 hidden layers spread over four layers in mini ArabicBERT with 11 million parameters. On the other hand, in addition to 512 and 1024 hidden layers, the medium and the large models have 8 and 24 layers, respectively. There are 12 layers, 768 hidden layers, and 110 million parameters in the base model [228].

- **AraELECTRA:** The pre-trained ELECTRA is consists of two modules: a discriminator and a generator. A discriminator is typically used and refined for downstream tasks. The same text corpus used for AraBERT was employed to train Arabic-specific ELECTRA. Even so, AraELECTRA outperforms AraBERT, which has twice as many parameters, and is considerably more economical in terms of computation. It yields comparable outcomes. Consequently, this may be the best choice in situations where resources are scarce [229].
- **Arabic ALBERT:** ALBERT is available in Arabic through Arabic ALBERT2. This model was trained using the Arabic Wikipedia and the Arabic version of the OSCAR corpus. There are three primary versions of this model based on the number of parameters. Using parameter reduction techniques, such as factorized embedding parameterization and cross-layer parameter sharing, the model reduces the number of parameters and accelerates the training process. MSA is primarily used to train Arabic language models for emotion analysis of Arabic tweets.
- **MARBERT:** MARBERT has received training to improve its ability to deal with dialectal Arabic. The authors added a set of one billion randomly gathered Arabic tweets to the training set. Approximately half of the final training dataset's textual content, which is about 128 GB, was made up of tweets. Among other NLP tasks, this model has been assessed on sentiment analysis, named entity recognition, topic categorization, dialect identification, and social meaning prediction. MARBERT comprises 24 layers, with a total parameter of 60M and 2048 hidden layers [230].
- **HindiBERT:** Hindi and English codes were mixed into the Roman text to create the HingBERT model. This is an enhanced iteration of the L3Cube-HingCorpus Multilingual BERT (mBERT) model. 1.04 billion tokens (HingCorpus's currency) were made using 52.93 million phrases from Twitter [231].
- **HingRoBERTa:** An English RoBERTa model optimized for Roman Hindi-English code mixed text is called HingRoBERTa. Once again, the L3Cube-HingCorpus Roman version was employed for fine-tuning.
- **BanglaBERT:** The model is derived from the base ELECTRA model, which is a 12-layer transformer encoder with the batch size of 256 for 2.5M steps on a v3-8 TPU instance on GCP. It has 768 embeddings, 768 hidden units, 12 attention heads, 3,072 feed-forward units, generator-to-discriminator ratio of 1:3, and 110 million parameters. The model was optimized using a linear warmup of 10,000 steps and Adam optimizer and 2e-4 learning rate [232].

A summary of the studies on TBED based on transformers is presented in Table 11.

4) HYBRID APPROACH

Sometimes, standalone approaches cannot provide satisfactory outcomes. A hybrid system uses a combination of two or more different approaches to achieve a better in performance in emotion detection. A hybrid strategy can be used when an unlabeled dataset is discovered. The dataset is labeled using an unsupervised method, and a supervised approach is subsequently utilized. Lexicon-based and machine-learning systems can be used for this purpose.

A summary of the studies based on the hybrid approach is presented in Table 12.

D. COMPARISON OF THE CLASSIFIERS

In the previous sections, we presented the approach-wise performance of different classifiers and models. According to our thorough examination of numerous models and classifiers, SVM classifier is the most popular classifier in the conventional ML approach. According to our analysis, the SVM classifier performed better on average than the NB and KNN classifiers, with an average performance of 74.13% compared to 73.06%. Furthermore, we found that, with an average accuracy of 76.85%, the DT classifier performed well. The RF classifier exhibited the best performance, with an average accuracy of 83.9%.

However, our survey found that DL-based classifiers, such as CNN, GRU, Bi-GRU, LSTM, and Bi-LSTM, showed better accuracy than most conventional ML classifiers. According to our research, DL-based classifiers may perform better in a variety of applications than the conventional ML classifiers. For example, the average accuracy of the LSTM and GRU classifiers was 85%, whereas that of Bi-GRU was 85.76%. The average accuracies displayed by the Bi-LSTM and CNN classifiers were 78.8% and 80.63%, respectively. Xu et al. [224] achieved 97% accuracy by applying the CNN. On the other hand, Sadegh et al. [215] also obtained 97% accuracy using GRU.

A summary of the indicative average accuracy results by different classifiers is represented in Figure 13.

E. PERFORMANCE EVALUATION

1) CONFUSION MATRIX

For a classification system such as emotion detection, the target emotion is identified as positive, and the others are regarded as negative. For example, if we are currently focusing on "happiness," a data instance (text) with the class label of "happiness" is identified as positive. On the other hand, a data instance (text) with any of the other class labels (such as "sadness," "anger," etc.) is treated as

TABLE 11. Overview of TBED papers based on transformer model.

Ref.	Language	Dataset	Model	Result	Evaluation Metric
[233]	Arabic	Tweets, News paragraphs	AraBERT	-	F-score (macro, micro)
[234]	Bangla	UBMEC	m-BERT	71.03%	Weighted F1
[98]	Hindi-English	Twitter	BERT, RoBERTa, ALBERT	71.43%	Accuracy
[235]	English	GoEmotions, Vent	BERT	91.2%	Accuracy
[236]	English	-	Cross-lingual, BERT	90%	Jaccard score
[237]	English	conversation benchmark	BERT	Improve 5-10%	Accuracy
[238]	English	SemEval-2019 Task 3 and SMP2020-EWECT	BERT, BLS	IBERT+CFEBLS 88.87% 73.32%	Accuracy, F-score
[239]	English	SemEval-2016, EmoBank, GoEmotions, ISEAR	RoBERTa	Not Mentioned	F-score
[240]	English	ISEAR	BERT, RoBERTa DistilBERT, and XLNet	RoBERTa 74.31%	Accuracy

TABLE 12. Overview of the studies using the hybrid approach.

Ref.	Language	Dataset	Text Preprocessing	Model	Result	Evaluation Metric
[241]	Arabic	AraPersonality	Normalization, Redundancy removal, Stemming, Emoji and emoticon conversion	AraBERT	82.73%	Macro-F1
[242]	Bangla	Bangla dataset	Remove Bangla stop-words, letters, digits, duplicate information, links, or URLs, numerals, punctuation marks, emojis, noisy data, unwanted value	Bi-LSTM	85%	Accuracy
[1]	English	Twitter	Cleaning	K-means, NB, SVM	SVM-77%, 79%, 82%	Precision, Recall, F-score
[2]	English	Twitter	Cleaning	SVM	68.3%, 67.4%	Precision, F-score
[243]	English	Newspaper Website, Blogs	Segmentation, Tokenization Stop word removal, Stemming	NB, SVM	-	-
[244]	English	Facebook	Text Segmentation Handling Emoticon Stop word removal Stemming	NB	78%	Accuracy
[245]	English	YouTube video (Comment)	-	NER, PoS, UNSET	-	-
[246]	English	Twitter, Facebook, News Agencies	Cleaning noises, Tokenization, Stemming	SVM, LR, NB, DT (J48)	-	Precision, Recall, F-score
[247]	English	Sports, Health Care, Election	-	NB, SVM	Both work better for different datasets	Precision, Recall Accuracy, F-score
[248]	English	Twitter, Post of Blogs, ISEAR, SemEval	Lemmatization	Ensemble Transformer	61.4%	Avg F-score
[249]	English	ISEAR, SemEval, Alms	-	NB+lexicon	65.5%	F-score

negative. By testing a classifier on a test dataset, we obtain the following four basic units from the cells of a confusion matrix [250] as shown in Figure 14.

- True Positive (TP): Both the actual and predicted classes are positive.
- True Negative (TN): Both the actual and predicted classes are negative.
- False Positive (FP): The predicted class is positive, but the actual class is negative.
- False Negative (FN): The predicted class is negative, but the actual class is positive.

2) EVALUATION METRICS

The performance of the emotion detection model can be evaluated using various metrics based on a confusion matrix. These metrics measure the outcomes of the model and its effectiveness. The following are the four most commonly used criteria in TBED. The values range between 0 and 1 for all four metrics. A higher value indicates a better performance.

- **Precision:** This measures the capacity of a classifier to correctly recognize positive emotions. Low precision means that many negatives are incorrectly reported as

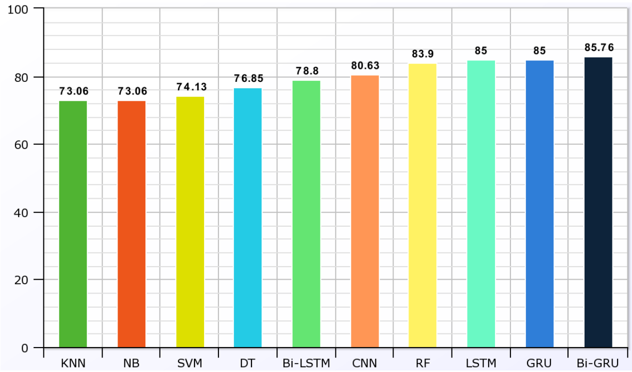

FIGURE 13. Average accuracy of classifiers. (Note: Since the accuracy values reported in various papers were achieved on different datasets using different experimental protocols, they should be treated as merely indicative.).

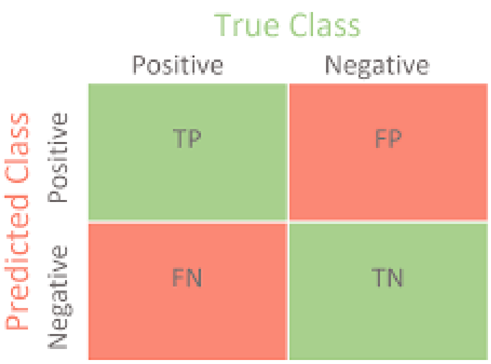

FIGURE 14. Confusion matrix.

positive.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

- **Recall:** This refers to the proportion of positive emotions identified among the available emotions. This measures the completeness of the results. A low recall indicates that many instances of positive emotions remained undetected.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

- **F-score:** This is also known as the F1-score, F-value, or F-measure. This is the harmonic mean of precision and recall, which are weighted equally.

$$F\text{-}score = 2\times\frac{Precision\times Recall}{Precision + Recall} \tag{4}$$

Accuracy: The accuracy of the model is used to determine its overall efficacy. In emotion detection investigations, precision signifies repeatability, whereas accuracy denotes authenticity.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

The prevalence of the metrics used by researchers to evaluate their TBED systems is shown in Figure 15.

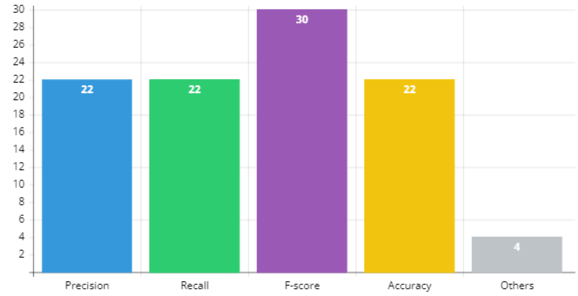

FIGURE 15. Prevalence of evaluation metrics in the evaluation process of TBED in the papers surveyed in this study. (Note: many papers use more than one metric.).

When there are N emotion class labels, there are N confusion matrices, one for each class label. N sets of the evaluation metrics mentioned above can be computed. Typically, researchers report the average value of each evaluation metric for all the class labels.

Apart from the above four most common criteria, the other often-used measures include the Jaccard score, sensitivity, specificity, kappa coefficient, and Pearson correlation coefficient.

3) COMMON EXPERIMENTAL SETUPS

The following are the most common experimental setups in classification [251], including the TBED experiments.

- **Training and Testing:** The dataset is divided into a larger chunk for training and a smaller chunk for testing. Typically, splits, such as 90:10, 80:20, and 70:30, are used. Evaluation metrics are reported for the testing set. In this setup, one could fine-tune the classifier model multiple times to achieve better results on the testing set.
- **Training, Validation, and Testing:** The dataset is divided into three parts. The training and validation sets can be used many times (most commonly for classifier fine-tuning) to obtain an optimal/final classifier model. The testing set is held out during the training+validation process. It is used only once at the end to evaluate the performance of the final classifier model. Splits, such as 80:10:10, 70:15:15, and 60:20:20, are commonly used. This setup is considered more rigorous/objective than the above "training and testing only" approach.
- **k-Fold Cross-Validation:** The dataset is divided into k (k=10, 5 or any integer number) chunks. One is for testing, whereas the remaining ones are used for training. The average score of each evaluation metric is used to evaluate the performance after the k tests. Although this setup is more time-consuming (especially for resource-intensive methods, such as deep learning), the evaluation scores yielded by cross-validation are considered more robust than the above two.

V. CHALLENGES AND EMERGING TRENDS

Some limitations were realized after reviewing the literature on TBED. This section addresses the challenges associated

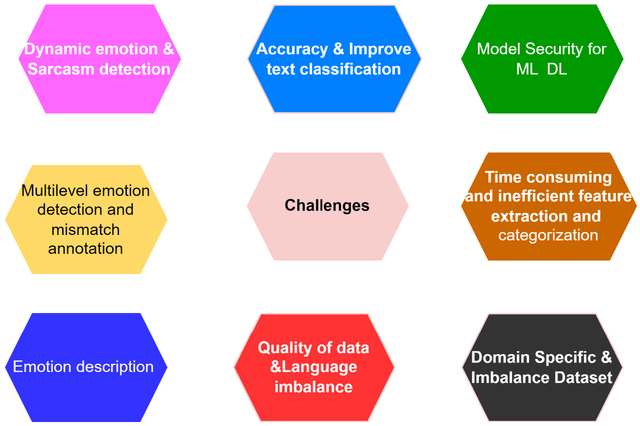

FIGURE 16. Challenges in TBED.

with TBED. This section also discusses the emerging trends to address these challenges.

A. CHALLENGES

The challenges faced in TBED are listed together in Figure 16.

- **Difficult to identify emotion:** Identifying concealed emotions is difficult; yet another challenge is identifying emotions in writing when no emotive words or expressions are used. There are numerous possible meanings for a text's words and sentences. Numerous emotions and perspectives may be contained in a single sentence. For example, people can use a correction for both love and happiness. A sentence may have a different meaning from another perspective. Therefore, emotion recognition is difficult to achieve. This issue must be addressed to improve the performance or accuracy of automated emotion recognition systems.
- **Multi-label emotion recognition:** Existing emotion classification methods divide emotions into many categories, including Ekman's six basic emotions, which seem simplistic and ignore the diversity of emotions. Truth be said, human emotion is nuanced. As a result, numerous connections exist among various emotional classifications, and no clear boundary exists, and it can be challenging to provide an exact term for an emotional aspect. Furthermore, emotion-based feelings can be personal. Humans may feel diverse emotions in response to the same linguistic statement depending on their own experience. Because of the hazy emotional borders and subjective feelings of humans, TBED becomes challenging. For more complex explanations of emotions, many academics use a variety of emotional categories simultaneously. Multiple fundamental emotions are combined to form multi-label emotions, which can more fully explain complex feelings and overcome the limitations mentioned above.
- **Annotating system mismatch:** Existing TBED tasks use discrete, dimensional, and componential emotion models extensively. However, no common annotation standard exists, resulting in incompatibility between the different datasets. For example, if one dataset has Ekman's six basic emotions, another contains four, whereas the other includes eight. In addition, cross-corpus emotion resources are frequently incompatible owing to the various emotion labels used. Although several emotional corpora have been produced, incompatible annotations may explain why insufficient training data exist.
- **Imbalance datasets:** Numerous studies have been conducted to train TBED models by reducing misclassification faults. To enhance the prediction performance, further investigation must be performed to ensure the innate characteristics of emotion datasets. Most studies assume that the amount of text labeled with each emotion is approximately equal and distributed over the corpus of data. However, this is not always true in reality. Most currently used emotional datasets are generated and labeled manually. Therefore, the data distribution for every emotion is constantly unbalanced. For instance, a dataset comprising 10% love, 40% surprise, and 40% happiness is regarded as an unbalanced dataset. Unbalanced datasets can lead to incorrect predictions.
- **Domain Adaptation and no universal dataset:** There are some datasets, and the maximum is imbalanced and performs a specific domain of emotion detection. However, no datasets can be used for multiple domains. For example, suppose a dataset is used for a product review. Furthermore, the dataset cannot be used to detect the emotions of chatbot applications. Therefore, it is necessary to create different datasets for different domains. However, this is a significant challenge for TBED. Emotion detection using other languages poses a significant challenge. There are approximately 7,111 languages worldwide [252]. Among these, only a few languages have established the TBED datasets. A majority of the TBED datasets are in English. Thus, detecting emotions in different languages is a tricky business.
- **Detecting emotion from substandard languages:** Users of online social platforms convey their feelings through sarcasm, irony, comedy, or other means. As opposed to formal language, slang, spelling mistakes, hashtags, emoticons, acronyms, etc., social media texts use informal language. As a result, interpreting such informal writings for automatic TBED systems becomes difficult.
- **Inefficient and time-consuming extraction and categorization of features:** Effective feature extraction is necessary for most machine learning methods to successfully recognize emotions. In contrast, manually extracting features requires time and is prone to mistakes. In addition, the manual classification process may also be complicated because mislabeled sentiments may appear. Therefore, poor feature extraction and

labeling can seriously impair the precision of text-based emotion recognition.

- **Generating dynamic emotions:** Tracking contextual information and examining the discourse's overall tone are the main focuses of the methodologies mentioned earlier. These methods are mostly concerned with inertia at the emotional level. They presume that individuals' emotional states are stable and that the background depicts a constant conversational environment. However, this presumption ignores the fundamental dynamic emotional shifts experienced by each side during a debate. The emotions of the interlocutors are separate, even though they influence each other during the conversation. The stimulus causes the feelings to fluctuate as conversation progresses. As a counterbalance, dynamic emotion analytics offers a variety of applications, such as emotional chatbot companionship and healthcare robot conversations. It is beneficial for better dynamic management and counseling during the talk if doctors and patients actively monitor their emotions.
- **Detecting sarcasm:** Sarcasm is the use of language to convey the antithesis of what is intended to be said, typically to offend, irritate, or make someone laugh. For example, it is a sarcasm when one refers to a disorganized group by saying, "They're fully on top of things." Consequently, distinguishing emotions from sarcasms is a complicated process.
- **Detecting emotion from low resources languages:** In our survey, we comparatively discussed the different aspects of TBED. After a full observation, we found that low-resource languages, such Bangla, Arabic and Hindi, impose some specific challenges compared to English languages as discussed below.
  1) **Orthographic ambiguity:** The Arabic language suffers from orthographic ambiguity, which causes words and characters to have different spellings depending on the context; conversely, emotions in brief texts, such as tweets, can have unclear meanings. Furthermore, the interpretation of Arabic words with emotional connotations varies depending on the context in which they occur. Comparably, it is far more difficult to identify implicit emotions in the emotional text that does not contain explicit emotional phrases than it is to identify explicit emotions in direct, explicit text. Moreover, Arabic has a rich morphology, meaning that a single verb can have thousands of different forms. The same problem was observed in Bangla and Hindi.
  2) **Diversity of dialects:** Arabic is spoken in many different regions of the country, giving rise to various dialects. Since every dialect may have unique idioms, expressions, and emotional language variations, it is difficult to create a model that works well everywhere in the world where Arabic is spoken. Cultural factors frequently impact how people express their emotions, so what one culture views as an expression of a given emotion may not be the same in another. It is extremely difficult to create models that are sensitive to cultural differences and sufficiently flexible to represent a wide range of emotional states. Hindi and Bengali cultures express emotions in different ways. Accurately detecting emotions requires an understanding of and accommodation for these cultural variations. Hindi and Bangla both have regional dialects with distinct linguistic traits. Complexity is added when a model is modified to account for the variety of dialects in these languages.
  3) **Code-switching:** A common practice among Arabic speakers is code-switching, which involves the incorporation of English or French vocabulary into Arabic writing. Because of this, emotion detection models have become more complex as they have to take into account different languages and their corresponding emotional expressions. Similarly, code-switching is a common linguistic occurrence in communities that speak Bangla and Hindi, which complicates emotion detection algorithms. Bangla or Hindi expressions frequently incorporate English or other regional languages smoothly. Emotion detection models face difficulties as a result of this practice because they have to deal with the subtle differences between different languages' emotional expressions. Due to its rich cultural past and frequent code-switching with English and other languages, Bangla makes it difficult for models to identify subtle emotional differences between the two languages. Similarly, speakers of Hindi often mix in English or other regional languages, adding to the ever-changing linguistic landscape. Deep comprehension of Bangla, Hindi, and the inserted languages is necessary for emotion detection models functioning in this context to be able to recognize and interpret emotional states.
  4) **Lack of resources:** The scarcity of resources in languages such as Hindi, Bangla, and Arabic makes emotion detection extremely difficult. Bangla, Arabic, and Hindi have limited resources for data cleaning compared to English, which has an abundance of predefined libraries and tools. It is difficult to effectively clean and prepare textual data for further analysis during the preprocessing stage due to the lack of standardized libraries designed for these languages. Furthermore, there are not many transformer models and lexicons that are unique to Bangla and Hindi. Although there are many transformer architectures and lexicon resources available for English, Bangla and Hindi may not have as many or as well-developed ones. Owing to their scarcity, models may not be able to take advantage of the most advanced transformer architectures or extensive lexicons, which are essential for precise emotion detection and

language analysis. As previously mentioned, there are some transformer-based methods in Arabic, but they still require improvement. Moreover, one of the major challenges in training robust models is the creation of large annotated datasets. An abundance of annotated datasets available for English makes it easier to train emotion detection models. In contrast, the ability to create high-performing models is limited by the scarcity of large annotated datasets available for Hindi, Bangla, and Arabic. The development of efficient emotion detection models for Bangla and Hindi is hampered by the lack of annotated data, which is crucial for teaching models the nuances of emotional expressions in these languages.

- **Multilingual emotion detection:** We discovered that multilingual models mostly represent Western cultural values and fail to accurately capture cultural variances related to emotion. Multilingual language models' emotion embeddings are fixed in English, and text completions produced in response to prompts in languages other than English do not reflect the emotional inclinations of the users' expected cultural contexts. For example, GPT-4 responds as an American who speaks Japanese fluently but is not familiar with Japanese culture or values when prompted in Japanese. The authors of an investigation [253] advised against relying solely on the emotional representations that language models (LMs) learned for use in downstream applications. There are risks associated with using machine translation for label transfer or multilingual LMs in a zero-shot setting for languages that have not been seen before. This is because the multilingual representations of emotion that these models learn do not always accurately reflect the ways in which the corresponding cultures express emotion.
- **Translation accuracy:** A more thorough understanding of how emotions are expressed can be achieved through the analysis of emotions across linguistic boundaries made possible by translation. Researchers employed machine translation to convert the sentence for multilingual and cross-cultural emotion detection. The model's performance can be improved by translating the code-switched text into the training language of the model. Even if they are not native speakers of every language used, translators can still provide light on the content of the code-switched text. This makes it easier to comprehend the data more thoroughly. Translating an idiom's various shades of meaning can be difficult. This is a difficult problem in translation. The explanation for this is that translating idioms frequently entails understanding meta-linguistic details, such as social and cultural norms. Conversational idioms are widely used in tweets due to their informal nature. The manual analysis has revealed that many idioms were translated literally, which not only affected the source text's ability to retain sentiment but also frequently resulted in the nonsensical target text. The same problem exists in other code-mixed or low-resource datasets [254].
- **Cross-domain emotion detection:** The basic tenet of the cross-domain emotion analysis method is that given enough words that convey a range of emotions from multiple domains, emotions contained in current comment information can be swiftly and accurately identified. Traditional techniques for analyzing emotions, however, frequently overlook the domain-dependent aspects of emotive words and may even purposefully select domain-independent elements (e.g., emoji). There is a growing need for cross-domain emotion analysis research due to the emergence of emotional corpus resources in various domains and the growing demand for practical application.
- **Challenges of Aspect-based Sentiment Analysis (ABSA):** A sizable multi-dialect and multi-domain benchmark dataset for Arabic, Hindi, Bangla, and other languages covered by ABSA is needed, along with comprehensive details on the dialects and domains covered, in order to assess the generalizability of the suggested models for these languages. covering topics that are either unexplored yet (health care, entertainment events, travel, and transportation services) or understudied (airline, telecoms, education, etc.).It has not yet been addressed how to handle various language text variations, such as Arabizi for Arabic, Assamese for Bangla, and Devanagari script for Hindi. Scholars ought to experiment with alternative neural network architectures, such as recurrent neural networks, GCNs, and capsule neural networks. It has been demonstrated that contextualized language models, such as BERT, perform better than other embeddings. Further research that compares different contextualized language models is advised. For example, for all ABSA tasks, research could compare the performance of Electra, ELMO, GPT, and BERT at different levels of granularity, such as character, sub-word, word, and sentence. While attention mechanisms and transformer-based pre-trained language models both produce encouraging results, combining the two approaches could boost performance even further.
  Promising results have been shown for English when using DL models to solve ABSA tasks collaboratively or with related tasks, such as sentiment term extraction via multitask learning [255], [256]. Therefore, researching this method for Arabic, Bangla, Hindi, and other languages may be beneficial. It is advised to create new methods for automatically or semi-automatically creating domain-dependent sentiment lexicons because of ABSA domain dependency and the benefits the sentiment lexicon provides for performance. This can be achieved by creating novel methods or

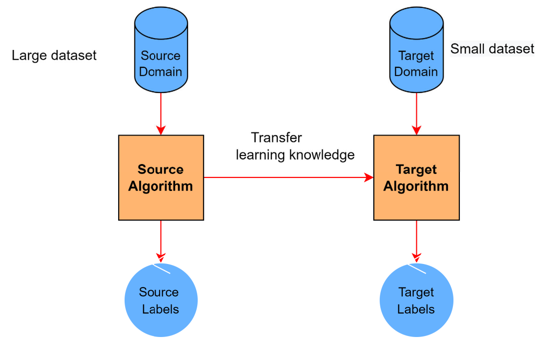

FIGURE 17. General procedure of transfer learning workflow.

modifying existing methods that have been developed for English [257], [258].

B. EMERGING TRENDS

This study focused on how the body of literature contributes to AI-based textual emotion recognition. Consequently, this section covers several emerging and potentially promising future research directions in AI for TBED. In the previous sub-section, we have discussed the common difficulties faced in a TBED system. We discovered significant issues such as imbalanced datasets, poor text classification results, the need for performance generalization, and non-standard language. We will now investigate the up-and-coming techniques that become increasingly popular in addressing these issues.

1) TRANSFER LEARNING

Transfer learning techniques are employed with assistance from the source domain to enhance the overall learning of a targeted field [259]. A source is a collection of data with sufficient data samples and many labels, and is generally of good quality. In [95], the authors suggested employing cross-lingual embeddings for transfer learning to recognize 758 emotions. Through the standard realm of English and Hindi, their deep transfer learning system, which uses CNN and Bi-LSTM, effectively transfers vital knowledge that has been recorded. Researchers such as [260], [261], and [262] suggested employing a transfer learning strategy to tackle the lack of data, as it can be used for adapting the domain. A simple overview of the transfer learning architecture is presented in Figure 17.

2) MULTI-TASK LEARNING

Multitask learning involves learning numerous tasks from multiple domains simultaneously, with no distinction between the source and goal. With a series of learning n tasks $(T_1, T_2, T_3, \ldots, T_n)$ together, it entails learning each activity simultaneously. By doing so, each scenario's information can be more effectively transferred and a rich composite feature vector can be developed from all of the many scenarios within the same domain. Using common knowledge, the learner optimizes the learning/performance across all n tasks [263]. Multitask learning can be viewed as a distinct form of transfer learning. It can improve the generalization performance of the TBED model by capitalizing on information, such as network parameters, that are shared among tasks [264].

3) GRAPH-BASED NETWORKS

Over the past few years, graph neural networks (GNN) have attracted the attention of numerous academics. Applications of NLP have employed GNNs in several ways. Text categorization is an example that uses GNN for the purpose of neural machine translation and relational reasoning [265], [266], [267]. In [266], a method of classifying texts was suggested, in which a single text graph was built for the corpus based on word co-occurrence and the document-word relationship in the GNN.

A crucial and common issue in TBED is its robustness for text classification. The Graph Convolution Network (GCN) is a prominent sub-type of GNN. It is a multi-layer neural network that operates directly on a graph and produces embedding vectors for nodes based on the properties of the regions around them. As a result, GCN can produce robust text classification results and develop predictive document annotators using a variety of cutting-edge text embedding and classification techniques [268], [269].

4) REPRESENTATION LEARNING

During the encoding stage, feature vectors are translated into a dimension space of any higher or lower order, allowing the original feature vector to be recovered in the subsequent decoding stage with the least amount of reconstruction error. In noisy situations, autoencoders can enhance system performance and handle missing data. These can also be trained using unsupervised learning. For example, in [270], robust feature representations have been created for emotion recognition using autoencoders. In this study, the denoising autoencoder receives the normalized static feature set as the input. Two hidden representations were mapped to this input: one for neutral information and the other for emotional information extraction. The model parameters were trained to reduce the squared difference in input error between the original and the reconstructed inputs. The outcomes showed a significant performance improvement over static features. Because of the outstanding findings published in [271], the field of text production has paid considerable attention to variational autoencoders (VAE) [272], [273]. Semi-supervised learning employing VAE can also enhance discriminative outcomes when unlabeled data are used [274].

5) ADVERSARIAL MACHINE LEARNING (AML)

Many machine learning and deep learning algorithms are highly efficient for various NLP tasks. Numerous publications state that these models are becoming more susceptible to hostile scenarios [275]. The best machine learning practices, reliable statistics, and computer security are combined in a relatively new field of research called AML. This refers to attacking systems based on machine learning and fortifying

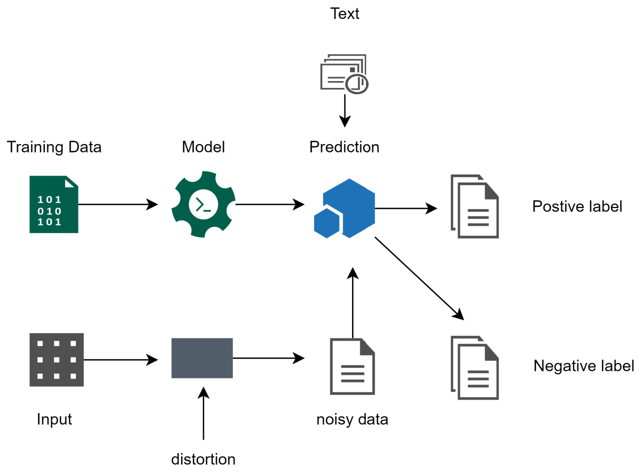

FIGURE 18. Adversarial machine learning (AML) workflow.

them against attacks. AML modifies inputs by generating fresh ones. The target model malfunctions because of minor perturbations and produces false results. Utilizing the flaws or weaknesses in the machine learning model is the objective. As a result, the usefulness of model may be significantly reduced. In [276], the authors examined hostile attacks on deep learning NLP models. They investigated the properties of adversarial machine learning, notably in text creation and analysis. The study [277] discussed the most significant contributions to the field of AML, including GAN algorithms, models, attack types, and countermeasures. Recently, the specific issue of adversarial attacks on TBED was addressed in [278]. Figure 18 depicts the AML workflow.

6) ADVANCES IN TACKLING CLASS IMBALANCE

Imbalanced datasets pose a significant issue in TBED. This is one of the major causes of inaccurate classification outputs. To balance the performance of TBED closely, the case can be primarily fixed by adding data (e.g., SMOTE) or removing data (e.g., NCL) [279]. The Synthetic Minority Oversampling Technique (SMOTE) extends sample data from the minority class. On the other hand, the Neighborhood Cleaning Rule (NCL) methodology properly balances the dataset by removing outliers and redundancies from the majority class. Heuristic algorithms include random oversampling. Random repetition of minority target instances is used to the balance class spreading [280]. Random undersampling is one of the simplest yet best simple undersampling algorithms [281]. It is also a heuristic approach that balances the target distributions by deleting instances randomly from the majority class. The general oversampling and undersampling processes can be visualized in Figure 19.

Recently, there has emerged a promising trend of using GAN-based techniques such as Text-GAN, SeqGAN, and Wasserstein GAN (WGAN) to oversample textual datasets [282], [283], [284]. GAN-based methods typically outperform traditional oversampling and undersampling methods.

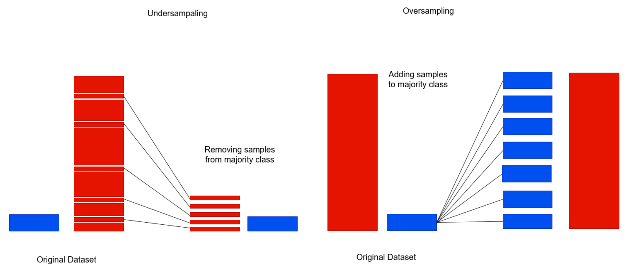

FIGURE 19. Visualization of oversampling and undersampling.

7) LANGUAGE AUGMENTATION

Code mixing refers to the use of multiple languages in a single text. The use of code-mixed data, especially English combined with a regional language, on social media platforms is evolving to a greater extent. The implicit language information in the code-mixed text is not utilized by the DL models that are currently in use. Word-level interleaving and post-sentence placement of language information can be used for language augmentation. All characters are changed to lowercase during the preprocessing phase, and the corresponding language tags are changed to uppercase. Two distinct approaches (the interleaved word-language approach and the adjacent sentence-language approach) can be used to train the model after each word has been given the proper language tag. The interleaved word-language approach involves appending language tags to each word as it is spoken. The Adjacent Sentence-Language Approach involves appending language tags to every word at the conclusion of the sentence. This approach showed a massive improvement in transformer models in experiments [285].

8) EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI)

For several strong reasons, XAI can be used in emotion detection. Because emotion detection models, particularly those built on intricate DL techniques, can be difficult to understand, XAI approaches offer transparency by revealing how decisions are made. In addition to increasing user trust, this transparency makes it easier for stakeholders to interpret the model and recognize the important characteristics and trends that affect emotion prediction. Additionally, XAI tools are very helpful in debugging models by identifying and fixing problems or biases. XAI creates a dynamic relationship between users and the emotion detection system by promoting user feedback, aligning model decisions with human intuition, and offering concise explanations for emotion predictions. There are many frameworks for this approach, such as Shapley additive explanation (SHAP) [286] and Local Interpretable Model-agnostic Explanation (LIME) [287]. Velampalli et al. [288] implemented the SHAP base XAI framework to determine if there were any biases. This analysis helped them to find that the models for sentence embedding were effective in capturing semantic similarities. When the models saw entirely new sentences, the method performed admirably. XAI techniques are used nowadays to make the result most trustworthy [289]. Li et al. [290]

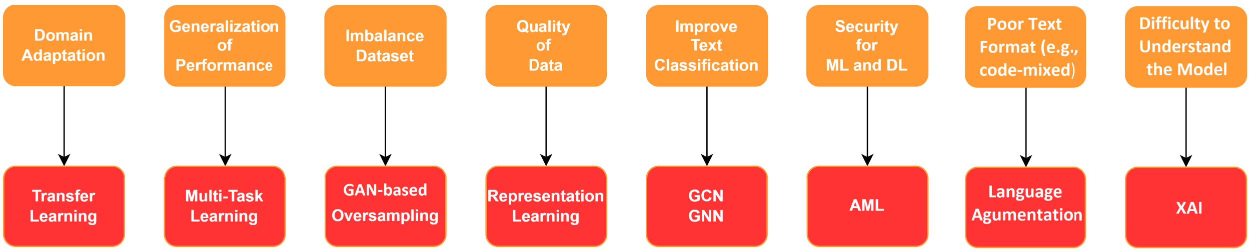

FIGURE 20. Emerging techniques towards addressing challenges in TBED.

identified words that elicited strong emotions and their causes. Additionally, they used visualization, which enhances one's understanding of the explanation.

In this sub-section, we have discussed the emerging trends in TBED. In order to elucidate which technique is suited for which specific problem, we graphically represented those emerging techniques together with the corresponding issues that they can be used to address in TBED in Figure 20.

VI. CONCLUSION

Emotion detection has a significant impact on studying human-computer interaction. Researchers have explored various factors that affect the detection of emotions, including emotional models and their advantages and disadvantages. Different methods have been identified for text-based emotion detection, with ML and DL being the most popular and effective. Feature engineering methods such as traditional and word embedding have also been explored, with word embedding outperforming traditional methods.

Text preparation stages are necessary to achieve improved accuracy in text-based emotion detection. The preprocessing stages of the text were outlined in this survey, including essential steps such as text cleaning and normalization. This study has examined publicly available datasets and lexicons. Text-based emotion detection has a range of applications and domains, but despite many studies being conducted, there is still no optimal solution. Our study highlights the challenges faced in this field and provides guidance for future development.

An extensive summary of the state of text-based emotion detection is presented in this survey article. This will be helpful for researchers who wish to understand the latest methodologies, classifier models, resources, and limitations in this field and help them in their future research and development efforts.

REFERENCES

[1] S. A. Salam and R. Gupta, "Emotion detection and recognition from text using machine learning," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 6, pp. 341–345, Jun. 2018.

[2] A. D. Souza and R. D. Souza, "Affective interaction based hybrid approach for emotion detection using machine learning," in *Proc. Int. Conf. Smart Syst. Inventive Technol. (ICSSIT)*, Nov. 2019, pp. 337–342.

[3] S. Kuamri and C. N. Babu, "Real time analysis of social media data to understand people emotions towards national parties," in *Proc. 8th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2017, pp. 1–6.

[4] M. Schwarz-Friesel, "Language and emotion: The cognitive linguistic perspective," in *Emotion in Language: Theory—Research—Application*. Amsterdam, The Netherlands: John Benjamins, 2015, pp. 157–173.

[5] P. J. Manamela, M. J. Manamela, T. I. Modipa, T. J. Sefara, and T. B. Mokgonyane, "The automatic recognition of sepedi speech emotions based on machine learning algorithms," in *Proc. Int. Conf. Adv. Big Data, Comput. Data Commun. Syst. (icABCD)*, Aug. 2018, pp. 1–7.

[6] A. Zajenkowska, D. Jasielska, and J. Melonowska, "Stress and sensitivity to frustration predicting depression among young adults in Poland and Korea—Psychological and philosophical explanations," *Current Psychol.*, vol. 38, no. 3, pp. 769–774, Jun. 2019.

[7] V. U. Kumar, A. Savithri, M. J. Bhavani, A. M. Priya, K. V. S. B. Jahnavi, and N. D. N. Lakshmi, "Finding psychological instability using machine learning," in *Proc. 7th Int. Conf. Smart Struct. Syst. (ICSSS)*, Jul. 2020, pp. 1–4.

[8] E.-H. Jang, B.-J. Park, S.-H. Kim, M.-A. Chung, M.-S. Park, and J.-H. Sohn, "Classification of human emotions from physiological signals using machine learning algorithms," in *Proc. 6th Int. Conf. Adv. Comput.-Human Interact. (ACHI)*, 2013, pp. 395–400.

[9] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, "Text-based emotion detection: Advances, challenges, and opportunities," *Eng. Rep.*, vol. 2, no. 7, Jul. 2020.

[10] S. Zad, M. Heidari, H. James Jr., and O. Uzuner, "Emotion detection of textual data: An interdisciplinary survey," in *Proc. 2021 IEEE World AI IoT Congr. (AIIoT)*, 2021, pp. 255–261.

[11] B. Kratzwald, S. Ilić, M. Kraus, S. Feuerriegel, and H. Prendinger, "Deep learning for affective computing: Text-based emotion recognition in decision support," *Decis. Support Syst.*, vol. 115, pp. 24–35, Nov. 2018.

[12] F. A. Acheampong, H. Nunoo-Mensah, and W. Chen, "Transformer models for text-based emotion detection: A review of BERT-based approaches," *Artif. Intell. Rev.*, vol. 54, no. 8, pp. 5789–5829, Dec. 2021.

[13] S. Abudalfa and M. Ahmed, "Survey on target dependent sentiment analysis of micro-blogs in social media," in *Proc. 9th IEEE-GCC Conf. Exhib. (GCCCE)*, May 2017, p. 1.

[14] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "A survey on aspect-based sentiment analysis: Tasks, methods, and challenges," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 11, pp. 11019–11038, Nov. 2023.

[15] K. Sailunaz, M. Dhaliwal, J. Rokne, and R. Alhajj, "Emotion detection from text and speech: A survey," *Social Netw. Anal. Mining*, vol. 8, no. 1, pp. 1–26, Dec. 2018.

[16] N. Alswaidan and M. E. B. Menai, "A survey of state-of-the-art approaches for emotion recognition in text," *Knowl. Inf. Syst.*, vol. 62, pp. 2937–2987, Aug. 2020.

[17] A. Seyeditabari, N. Tabari, and W. Zadrozny, "Emotion detection in text: A review," 2018, *arXiv:1806.00674*.

[18] L. A. M. Oberländer and R. Klinger, "An analysis of annotated corpora for emotion classification in text," in *Proc. 27th Int. Conf. Comput. Linguistics (COLING)*, 2018, pp. 2104–2119.

[19] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current state of text sentiment analysis from opinion to emotion mining," *ACM Comput. Surv.*, vol. 50, no. 2, pp. 1–33, Mar. 2018.

[20] S. Goyal and N. Tiwari, "Emotion recognition: A literature survey," *Int. J. Eng. Res. Technol.*, vol. 4, no. 9, pp. 1502–1524, 2017.

[21] S. Al-Saqqa, H. Abdel-Nabi, and A. Awajan, "A survey of textual emotion detection," in *Proc. 8th Int. Conf. Comput. Sci. Inf. Technol. (CSIT)*, Jul. 2018, pp. 136–142.

[22] L. Canales and P. Martínez-Barco, "Emotion detection from text: A survey," in *Proc. Workshop Natural Lang. Process. 5th Inf. Syst. Res. Work. Days (JISIC)*, 2014, pp. 37–43.

[23] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social Netw. Anal. Mining*, vol. 11, no. 1, pp. 1–19, Dec. 2021.
[24] J. Deng and F. Ren, "A survey of textual emotion recognition and its challenges," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 49–67, Jan. 2023.
[25] U. Jain and A. Sandhu, "A review on the emotion detection from text using machine learning techniques," *Emotion*, vol. 5, no. 4, pp. 2645–2650, 2009.
[26] T. Rabeya, S. Ferdous, H. S. Ali, and N. R. Chakraborty, "A survey on emotion detection: A lexicon based backtracking approach for detecting emotion from Bengali text," in *Proc. 20th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2017, pp. 1–7.
[27] A. R. Murthy and K. M. Anil Kumar, "A review of different approaches for detecting emotion from text," *IOP Conf. Series: Mater. Sci. Eng.*, vol. 1110, no. 1, Mar. 2021, Art. no. 012009.
[28] K. Hulliyah, N. S. A. A. Bakar, and A. R. Ismail, "Emotion recognition and brain mapping for sentiment analysis: A review," in *Proc. 2nd Int. Conf. Informat. Comput. (ICIC)*, Nov. 2017, pp. 1–5.
[29] D. D. Gosai, H. J. Gohil, and H. S. Jayswal, "A review on an emotion detection and recognization from text using natural language processing," *Int. J. Appl. Eng. Res.*, vol. 13, no. 9, pp. 6745–6750, 2018.
[30] L. Cao, S. Peng, P. Yin, Y. Zhou, A. Yang, and X. Li, "A survey of emotion analysis in text based on deep learning," in *Proc. IEEE 8th Int. Conf. Smart City Informatization (iSCI)*, Dec. 2020, pp. 81–88.
[31] W. Chung and D. Zeng, "Dissecting emotion and user influence in social media communities: An interaction modeling approach," *Inf. Manage.*, vol. 57, no. 1, Jan. 2020, Art. no. 103108.
[32] Y. Bao, C. Quan, L. Wang, and F. Ren, "The role of pre-processing in Twitter sentiment analysis," in *Proc. Int. Conf. Intell. Comput. (ICIC)*, 2014, pp. 615–624.
[33] F. Kunneman, C. Liebrecht, and A. van den Bosch, "The (un)predictability of emotional hashtags in Twitter," in *Proc. 5th Workshop Lang. Anal. Social Media (LASM)*, 2014, pp. 26–34.
[34] S. M. Mohammad and S. Kiritchenko, "Using hashtags to capture fine emotion categories from tweets," *Comput. Intell.*, vol. 31, no. 2, pp. 301–326, May 2015.
[35] S. Cassab and M.-B. Kurdy, "Ontology-based emotion detection in Arabic social media," *Int. J. Eng. Res. Technol.*, vol. 9, no. 8, pp. 1991–2013, 2020.
[36] S. Kusal, S. Patil, K. Kotecha, R. Aluvalu, and V. Varadarajan, "AI based emotion detection for textual big data: Techniques and contribution," *Big Data Cogn. Comput.*, vol. 5, no. 3, p. 43, Sep. 2021.
[37] N. Gupta, M. Gilbert, and G. D. Fabbrizio, "Emotion detection in email customer care," *Comput. Intell.*, vol. 29, no. 3, pp. 489–505, Aug. 2013.
[38] B. T. Raad, B. Philipp, H. Patrick, and M. Christoph, "ASEDS: Towards automatic social emotion detection system using Facebook reactions," in *Proc. IEEE 20th Int. Conf. High Perform. Comput. Commun., IEEE 16th Int. Conf. Smart City, IEEE 4th Int. Conf. Data Sci. Syst. (HPCC/SmartCity/DSS)*, Jun. 2018, pp. 860–866.
[39] C. Pool and M. Nissim, "Distant supervision for emotion detection using Facebook reactions," 2016, *arXiv:1611.02988*.
[40] P. Sarakit, T. Theeramunkong, C. Haruechaiyasak, and M. Okumura, "Classifying emotion in Thai Youtube comments," in *Proc. 6th Int. Conf. Inf. Commun. Technol. Embedded Syst. (IC-ICTES)*, Mar. 2015, pp. 1–5.
[41] A. Adikari, D. De Silva, D. Alahakoon, and X. Yu, "A cognitive model for emotion awareness in industrial chatbots," in *Proc. IEEE 17th Int. Conf. Ind. Informat. (INDIN)*, vol. 1, Jul. 2019, pp. 183–186.
[42] J. Yin, Z. Chen, K. Zhou, and C. Yu, "A deep learning based chatbot for campus psychological therapy," 2019, *arXiv:1910.06707*.
[43] M.-C. Lee, S.-Y. Chiang, S.-C. Yeh, and T.-F. Wen, "Study on emotion recognition and companion chatbot using deep neural network," *Multimedia Tools Appl.*, vol. 79, nos. 27–28, pp. 19629–19657, Jul. 2020.
[44] W. Graterol, J. Diaz-Amado, Y. Cardinale, I. Dongo, E. Lopes-Silva, and C. Santos-Libarino, "Emotion detection for social robots based on NLP transformers and an emotion ontology," *Sensors*, vol. 21, no. 4, p. 1322, Feb. 2021.
[45] J. Yoo and Y. Cho, "ICSA: Intelligent chatbot security assistant using text-CNN and multi-phase real-time defense against SNS phishing attacks," *Expert Syst. Appl.*, vol. 207, Nov. 2022, Art. no. 117893.
[46] S. M. C. Loureiro, J. Guerreiro, and I. Tussyadiah, "Artificial intelligence in business: State of the art and future research agenda," *J. Bus. Res.*, vol. 129, pp. 911–926, May 2021.
[47] B. Yang, Y. Liu, Y. Liang, and M. Tang, "Exploiting user experience from online customer reviews for product design," *Int. J. Inf. Manage.*, vol. 46, pp. 173–186, Jun. 2019.
[48] I. Chaturvedi, E. Cambria, S. Cavallari, and R. E. Welsch, "Genetic programming for domain adaptation in product reviews," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2020, pp. 1–8.
[49] S. Park, S. W. Lee, J. Kwak, M. Cha, and B. Jeong, "Activities on Facebook reveal the depressive state of users," *J. Med. Internet Res.*, vol. 15, no. 10, p. e2718, Oct. 2013.
[50] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," in *Proc. Int. Conf. Intell. Sustain. Syst. (ICISS)*, Dec. 2017, pp. 858–862.
[51] A. Kumar, A. Sharma, and A. Arora, "Anxious depression prediction in real-time social data," 2019, *arXiv:1903.10222*.
[52] D. Grandjean, D. Sander, and K. R. Scherer, "Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization," *Consciousness Cognition*, vol. 17, no. 2, pp. 484–495, Jun. 2008.
[53] H. Gunes and M. Pantic, "Automatic, dimensional and continuous emotion recognition," *Int. J. Synth. Emotions*, vol. 1, no. 1, pp. 68–99, Jan. 2010.
[54] P. Ekman, "Basic emotions," in *Handbook of Cognition and Emotion*, vol. 98. Hoboken, NJ, USA: Wiley, 1999, pp. 45–60.
[55] H. Binali and V. Potdar, "Emotion detection state of the art," in *Proc. CUBE Int. Inf. Technol. Conf.*, Sep. 2012, pp. 501–507.
[56] X. Jin and Z. Wang, "An emotion space model for recognition of emotions in spoken Chinese," in *Proc. Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, 2005, pp. 397–402.
[57] R. Plutchik, "A general psychoevolutionary theory of emotion, emotion research, theory, and experience," in *Theories of Emotion*, vol. 1, R. Plutchik and H. Kellerman, Eds. New York, NY, USA: Academic Press, 1980.
[58] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *J. Res. Pers.*, vol. 11, no. 3, pp. 273–294, 1977.
[59] T. Dalgleish and M. Power, *Handbook of Cognition and Emotion*. Hoboken, NJ, USA: Wiley, 2000.
[60] A. Balahur, J. M. Hermida, and A. Montoyo, "Building and exploiting EmotiNet, a knowledge base for emotion detection based on the appraisal theory model," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 88–101, Jan. 2012.
[61] W. G. Parrott, *Emotions Social Psychology: Essential Readings*. New York, NY, USA: Psychology Press, 2001.
[62] P. Agrawal and A. Suri, "NELEC at SemEval-2019 task 3: Think twice before going deep," 2019, *arXiv:1904.03223*.
[63] S. Chaffar and D. Inkpen, "Using a heterogeneous dataset for emotion analysis in text," in *Proc. Can. Conf. Artif. Intell. (CANAI)*, 2011, pp. 62–67.
[64] K. R. Scherer and H. G. Wallbott, "Evidence for universality and cultural variation of differential emotion response patterning," *J. Personality Social Psychol.*, vol. 66, no. 2, pp. 310–328, 1994.
[65] C. Strapparava and R. Mihalcea, "SemEval-2007 task 14: Affective text," in *Proc. 4th Int. Workshop Semantic Eval. (SemEval)*, 2007, pp. 70–74.
[66] S. Mohammad, F. Bravo-Marquez, M. Salameh, and S. Kiritchenko, "Semeval-2018 task 1: Affect in tweets," in *Proc. 12th Int. Workshop Semantic Eval. (SemEval)*, 2018, pp. 1–17.
[67] A. Chatterjee, K. N. Narahari, M. Joshi, and P. Agrawal, "SemEval-2019 task 3: EmoContext contextual emotion detection in text," in *Proc. 13th Int. Workshop Semantic Eval. (SemEval)*, 2019, pp. 39–48.
[68] C. E. Izard, *The Face Emotion*. New York, NY, USA: Appleton-Century-Crofts, 1971.
[69] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "@AM: Textual attitude analysis model," in *Proc. NAACL HLT Workshop Comput. Approaches Anal. Gener. Emotion Text*, 2010, pp. 80–88.
[70] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "Compositionality principle in recognition of fine-grained emotions from text," in *Proc. Int. AAAI Conf. Web Social Media (ICWSM)*, 2009, vol. 3, no. 1, pp. 278–281.
[71] S. Buechel and U. Hahn, "EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics*, vol. 2, 2017, pp. 578–585.
[72] S. Buechel and U. Hahn, "Readers vs. writers vs. texts: Coping with different perspectives of text understanding in emotion annotation," in *Proc. 11th Linguistic Annotation Workshop*, 2017, pp. 1–12.

[73] D. Preotiuc-Pietro, H. A. Schwartz, G. Park, J. Eichstaedt, M. Kern, L. Ungar, and E. Shulman, "Modelling valence and arousal in Facebook posts," in *Proc. 7th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal. (WASSA)*, 2016, pp. 9–15.

[74] S. Poria, D. Hazarika, N. Majumder, G. Naik, E. Cambria, and R. Mihalcea, "MELD: A multimodal multi-party dataset for emotion recognition in conversations," 2018, *arXiv:1810.02508*.

[75] A. G. Shahraki, "Emotion mining from text," M.S. thesis, Dept. Comput. Sci., Univ. Alberta, Alberta, WC, Canada, 2015.

[76] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "GoEmotions: A dataset of fine-grained emotions," 2020, *arXiv:2005.00547*.

[77] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in *Proc. 45th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2007, pp. 440–447.

[78] I. N. Khasanah, "Sentiment classification using fastText embedding and deep learning model," *Proc. Comput. Sci.*, vol. 189, pp. 343–350, Jan. 2021.

[79] A. Al-Khatib and S. R. El-Beltagy, "Emotional tone detection in Arabic tweets," in *Proc. 18th Int. Conf. Comput. Linguistics Intell. Text Process. (CICLing)*, 2018, pp. 105–114.

[80] W. J. Teahan, "A new Arabic dataset for emotion recognition," in *Proc. Intell. Comput. Conf.*, vol. 2, 2019, pp. 200–216.

[81] M. Abdul-Mageed, H. AlHuzli, and M. D. D. Elhija, "DINA: A multi-dialect dataset for Arabic emotion analysis," in *Proc. 2nd Workshop Arabic Corpora Processing Tools*, 2016, pp. 29–43.

[82] H. Alhuzali, M. Abdul-Mageed, and L. Ungar, "Enabling deep learning of emotion with first-person seed expressions," in *Proc. 2nd Workshop Comput. Model. People's Opinions, Personality, Emotions Social Media*, 2018, pp. 25–35.

[83] A. Al-Laith and M. Alenezi, "Monitoring people's emotions and symptoms from Arabic tweets during the COVID-19 pandemic," *Information*, vol. 12, no. 2, p. 86, 2021.

[84] K. A. Shakil, K. Tabassum, F. S. Alqahtani, and M. A. Wani, "Analyzing user digital emotions from a holy versus non-pilgrimage city in Saudi Arabia on Twitter platform," *Appl. Sci.*, vol. 11, no. 15, p. 6846, Jul. 2021.

[85] H. Ali, Md. F. Hossain, S. B. Shuvo, and A. Al Marouf, "BanglaSenti: A dataset of Bangla words for sentiment analysis," in *Proc. 11th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2020, pp. 1–4.

[86] T. Ahmed, S. F. Mukta, T. Al Mahmud, S. A. Hasan, and M. G. Hussain, "Bangla text emotion classification using LR, MNB and MLP with TF-IDF & CountVectorizer," in *Proc. 26th Int. Comput. Sci. Eng. Conf. (ICSEC)*, Dec. 2022, pp. 275–280.

[87] M. A. Rahman and M. H. Seddiqui, "Comparison of classical machine learning approaches on Bangla textual emotion analysis," 2019, *arXiv:1907.07826*.

[88] K. I. Islam, S. Kar, M. S. Islam, and M. R. Amin, "SentNoB: A dataset for analysing sentiment on noisy Bangla texts," in *Proc. Findings Assoc. Comput. Linguistics (AACL/IJCNLP)*, 2021, pp. 128–134.

[89] A. Das, O. Sharif, M. M. Hoque, and I. H. Sarker, "Emotion classification in a resource constrained language using transformer-based approach," 2021, *arXiv:2104.08613*.

[90] N. Banik and M. H. H. Rahman, "Evaluation of Naïve Bayes and support vector machines on Bangla textual movie reviews," in *Proc. Int. Conf. Bangla Speech Language Process. (ICBSLP)*, 2018, pp. 1–6.

[91] S. A. Mahtab, N. Islam, and M. M. Rahaman, "Sentiment analysis on Bangladesh cricket with support vector machine," in *Proc. Int. Conf. Bangla Speech Lang. Process. (ICBSLP)*, Sep. 2018, pp. 1–4.

[92] A. Hassan, M. R. Amin, A. K. A. Azad, and N. Mohammed, "Sentiment analysis on Bangla and Romanized Bangla text using deep recurrent models," in *Proc. Int. Workshop Comput. Intell. (IWCI)*, Dec. 2016, pp. 51–56.

[93] G. V. Singh, P. Priya, M. Firdaus, A. Ekbal, and P. Bhattacharyya, "EmoInHindi: A multi-label emotion and intensity annotated dataset in Hindi for emotion recognition in dialogues," 2022, *arXiv:2205.13908*.

[94] Y. Kumar, D. Mahata, S. Aggarwal, A. Chugh, R. Maheshwari, and R. R. Shah, "BHAAV—A text corpus for emotion analysis from Hindi stories," 2019, *arXiv:1910.04073*.

[95] Z. Ahmad, R. Jindal, A. Ekbal, and P. Bhattachharyya, "Borrow from rich cousin: Transfer learning for emotion detection using cross lingual embedding," *Expert Syst. Appl.*, vol. 139, Jan. 2020, Art. no. 112851.

[96] S. Sahmoud, S. Abudalfa, and W. Elmasry, "AT-ODTSA: A dataset of Arabic tweets for open domain targeted sentiment analysis," *Int. J. Comput. Digit. Syst.*, vol. 11, no. 1, pp. 1299–1307, 2022.

[97] B. G. Patra, D. Das, and A. Das, "Sentiment analysis of code-mixed Indian languages: An overview of SAIL_code-mixed shared task @ICON-2017," 2018, *arXiv:1803.06745*.

[98] A. Wadhawan and A. Aggarwal, "Towards emotion recognition in hindi-english code-mixed data: A transformer based approach," 2021, *arXiv:2102.09943*.

[99] A. Prabhu, A. Joshi, M. Shrivastava, and V. Varma, "Towards sub-word level compositions for sentiment analysis of Hindi–English code mixed text," 2016, *arXiv:1611.00472*.

[100] B. R. Chakravarthi, R. Priyadharshini, V. Muralidaran, N. Jose, S. Suryawanshi, E. Sherly, and J. P. McCrae, "DravidianCodeMix: Sentiment analysis and offensive language identification dataset for dravidian languages in code-mixed text," *Lang. Resour. Eval.*, vol. 56, no. 3, pp. 765–806, Sep. 2022.

[101] M. Tareq, M. F. Islam, S. Deb, S. Rahman, and A. A. Mahmud, "Data-augmentation for Bangla–English code-mixed sentiment analysis: Enhancing cross linguistic contextual understanding," *IEEE Access*, vol. 11, pp. 51657–51671, 2023.

[102] K. Yadav, A. Lamba, D. Gupta, A. Gupta, P. Karmakar, and S. Saini, "Bilingual sentiment analysis for a code-mixed Punjabi English social media text," in *Proc. 5th Int. Conf. Comput., Commun. Secur. (ICCCS)*, Oct. 2020, pp. 1–5.

[103] S. Mohammad and P. Turney, "Emotions evoked by common words and phrases: Using mechanical Turk to create an emotion lexicon," in *Proc. NAACL HLT Workshop Comput. Approaches Anal. Gener. Emotion Text*, 2010, pp. 26–34.

[104] F. Bravo-Marquez, E. Frank, S. M. Mohammad, and B. Pfahringer, "Determining word-emotion associations from tweets by multi-label classification," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Oct. 2016, pp. 536–539.

[105] S. Kiritchenko, X. Zhu, and S. M. Mohammad, "Sentiment analysis of short informal texts," *J. Artif. Intell. Res.*, vol. 50, pp. 723–762, May 2014.

[106] S. Mohammad, "Emotional tweets," in *Proc. 1st Joint Conf. Lexical Comput. Semantics (SEM), Proc. 6th Int. Workshop Semantic Eval. (SemEval)*, vols. 1–2, 2012, pp. 246–255.

[107] S. Mohammad and S. Kiritchenko, "WikiArt emotions: An annotated dataset of emotions evoked by art," in *Proc. 11th Int. Conf. Lang. Resour. Eval. (LREC)*, 2018, pp. 1125–1238.

[108] F. Å. Nielsen, "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs," 2011, *arXiv:1103.2903*.

[109] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, Aug. 2004, pp. 168–177.

[110] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in *Proc. Human Lang. Technol. Conf. (HLT) Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2005, pp. 347–354.

[111] R. Biagioni, *The SenticNet Sentiment Lexicon: Exploring Semantic Richness in Multi-Word Concepts*, vol. 4. Cham, Switzerland: Springer, 2016.

[112] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *Proc. 7th Int. Conf. Lang. Resour. Eval. (LREC)*, 2010, pp. 2200–2204.

[113] C. Strapparava and A. Valitutti, "WordNet affect: An affective extension of WordNet," in *Proc. 7th Int. Conf. Lang. Resour. Eval. (LREC)*, vol. 4, 2004, pp. 1083–1086.

[114] G. A. Miller, *WordNet: An Electronic Lexical Database*. Cambridge, MA, USA: MIT Press, 1998.

[115] F. M. Shah, A. S. Reyadh, A. I. Shaafi, S. Ahmed, and F. T. Sithil, "Emotion detection from tweets using AIT-2018 dataset," in *Proc. 5th Int. Conf. Adv. Electr. Eng. (ICAEE)*, Sep. 2019, pp. 575–580.

[116] H. Liu and P. Singh, "ConceptNet—A practical commonsense reasoning tool-kit," *BT Technol. J.*, vol. 22, no. 4, pp. 211–226, Oct. 2004.

[117] V. Kalra and R. Aggarwal, "Importance of text data preprocessing & implementation in RapidMiner," in *Proc. 1st Int. Conf. Inf. Technol. Knowl. Manage.*, Jan. 2018, pp. 71–75.

[118] F. H. Khan, S. Bashir, and U. Qamar, "TOM: Twitter opinion mining framework using hybrid classification scheme," *Decis. Support Syst.*, vol. 57, pp. 245–257, Jan. 2014.

[119] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and N. A. Smith, "Part-of-speech tagging for Twitter: Annotation, features, and experiments," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2010, pp. 42–47.
[120] E. Kouloumpis, T. Wilson, and J. Moore, "Twitter sentiment analysis: The good the bad and the OMG!" in *Proc. 2011 Int. AAAI Conf. Web Social Media (ICWSM)*, 2011, vol. 5, no. 1, pp. 538–541.
[121] A. Balahur, "Sentiment analysis in social media texts," in *Proc. 4th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal. (WASSA)*, 2013, pp. 120–128.
[122] J. A. Balazs and J. D. Velásquez, "Opinion mining and information fusion: A survey," *Inf. Fusion*, vol. 27, pp. 95–110, Jan. 2016.
[123] A. Giachanou, J. Gonzalo, I. Mele, and F. Crestani, "Sentiment propagation for predicting reputation polarity," in *Proc. Eur. Conf. Inf. Retr. (ECIR)*, 2017, pp. 226–238.
[124] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proc. 7th Int. Conf. Document Anal. Recognit., Proceedings.*, vol. 3, Aug. 2003, pp. 958–963.
[125] A. J. Keya, M. A. H. Wadud, M. F. Mridha, M. Alatiyyah, and M. A. Hamid, "AugFake-BERT: Handling imbalance through augmentation of fake news using BERT to enhance the performance of fake news classification," *Appl. Sci.*, vol. 12, no. 17, p. 8398, Aug. 2022.
[126] M. Bayer, M.-A. Kaufhold, B. Buchhold, M. Keller, J. Dallmeyer, and C. Reuter, "Data augmentation in natural language processing: A novel text generation approach for long and short text classifiers," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 1, pp. 135–150, Jan. 2023.
[127] H. Shim, S. Luca, D. Lowet, and B. Vanrumste, "Data augmentation and semi-supervised learning for deep neural networks-based text classifier," in *Proc. 35th Annu. ACM Symp. Appl. Comput.*, Mar. 2020, pp. 1119–1126.
[128] H. Luo, "Emotion detection for Spanish with data augmentation and transformer-based models," in *Proc. Iberian Lang. Eval. Forum (IberLEF)@SEPLN*, 2021, pp. 35–42.
[129] K. Kowsari, J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Information*, vol. 10, no. 4, p. 150, Apr. 2019.
[130] J. Suttles and N. Ide, "Distant supervision for emotion classification with discrete binary values," in *Proc. 14th Int. Conf. Comput. Linguistics Intell. Text Process. (CICLing)*, 2013, pp. 121–136.
[131] W. Wang, L. Chen, K. Thirunarayan, and A. P. Sheth, "Harnessing Twitter 'big data' for automatic emotion identification," in *Proc. Int. Conf. Privacy, Secur., Risk Trust (PASSAT), Int. Conf. Social Comput. (SocialCom)*, Sep. 2012, pp. 587–592.
[132] F. Pedregosa. (2013). *Sklearn TfidfVectorizer*. [Online]. Available: https://scikit-learn.org/stable/modules.feature_extraction.text .TfidfVectorizer.html
[133] M. F. Mridha, Md. S. Rahman, and A. Q. Ohi, "Human abnormality detection based on Bengali text," in *Proc. IEEE Region 10 Symp. (TENSYMP)*, Jun. 2020, pp. 1102–1105.
[134] A. Dadhich and B. Thankachan, "Sentiment analysis of Amazon product reviews using hybrid rule-based approach," in *Smart Systems: Innovations in Computing*. Singapore: Springer, 2022, pp. 173–193.
[135] T. Bolukbasi, K.-W. Chang, J. Y. Zou, V. Saligrama, and A. T. Kalai, "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 4356–4364.
[136] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[137] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 3111–3119.
[138] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Trans. Assoc. Comput. Linguistics*, vol. 5, pp. 135–146, Dec. 2017.
[139] O. Melamud, J. Goldberger, and I. Dagan, "context2vec: Learning generic context embedding with bidirectional LSTM," in *Proc. 20th SIGNLL Conf. Comput. Natural Lang. Learn.*, 2016, pp. 51–61.
[140] B. McCann, J. Bradbury, C. Xiong, and R. Socher, "Learned in translation: Contextualized word vectors," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 6297–6308.
[141] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," 2018, *arXiv:1802.05365*.
[142] G. R. Reddy, M. V. S. Reddy, M. Stanlywit, and S. Khaleel, "Emotion detection from text and analysis of future work: A survey," *Rivista Italiana di Filosofia Analitica Junior*, vol. 14, no. 1, pp. 59–73, 2023.
[143] A. L. Doan and S. T. Luu, "Improving sentiment analysis by emotion lexicon approach on Vietnamese texts," in *Proc. Int. Conf. Asian Lang. Process. (IALP)*, Oct. 2022, pp. 39–44.
[144] S. N. Shivhare, S. Garg, and A. Mishra, "EmotionFinder: Detecting emotion from blogs and textual documents," in *Proc. Int. Conf. Comput., Commun. Autom.*, May 2015, pp. 52–57.
[145] G. Farnadi, G. Sitaraman, M. Rohani, M. Kosinski, D. Stillwell, M.-F. Moens, S. Davalos, and M. De Cock, "How are you doing? Emotions and personality in Facebook," in *Proc. 2nd Workshop Emotions Personality Personalized Services (EMPIRE), 22nd Conf. User Modelling, Adaptation Personalization (UMAP)*, 2014, pp. 45–56.
[146] Y. Wang and A. Pal, "Detecting emotions in social media: A constrained optimization approach," in *Proc. 24th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2015, pp. 996–1002.
[147] A. Bandhakavi, N. Wiratunga, S. Massie, and D. Padmanabhan, "Lexicon generation for emotion detection from text," *IEEE Intell. Syst.*, vol. 32, no. 1, pp. 102–108, Jan. 2017.
[148] R. Rahman, "Detecting emotion from text and emoticon," *London J. Res. Comput. Sci. Technol.*, vol. 7, no. 3, pp. 9–13, 2017.
[149] I. Perikos and I. Hatzilygeroudis, "Recognizing emotion presence in natural language sentences," in *Proc. Int. Conf. Eng. Appl. Neural Netw. (EANN)*, 2013, pp. 30–39.
[150] D. Ghazi, D. Inkpen, and S. Szpakowicz, "Hierarchical versus flat classification of emotions in text," in *Proc. NAACL HLT Workshop Comput. Approaches Anal. Gener. Emotion Text*, 2010, pp. 140–146.
[151] J. Lei, Y. Rao, Q. Li, X. Quan, and L. Wenyin, "Towards building a social emotion detection system for online news," *Future Gener. Comput. Syst.*, vol. 37, pp. 438–448, Jul. 2014.
[152] L. De Bruyne, O. De Clercq, and V. Hoste, "LT3 at SemEval-2018 task 1: A classifier chain to detect emotions in tweets," in *Proc. 12th Int. Workshop Semantic Eval. (SemEval)*, 2018, pp. 123–127.
[153] A. Shoukry and A. Rafea, "SATALex: Telecom domain-specific sentiment lexicons for Egyptian and Gulf Arabic dialects," in *Proc. 15th Int. Conf. Web Inf. Syst. Technol. (WEBIST)*, 2019, pp. 169–176.
[154] T. Al-Moslmi, M. Albared, A. Al-Shabi, N. Omar, and S. Abdullah, "Arabic senti-lexicon: Constructing publicly available language resources for Arabic sentiment analysis," *J. Inf. Sci.*, vol. 44, no. 3, pp. 345–362, Jun. 2018.
[155] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 44–53, Jan. 2018.
[156] C. C. Sujadi, Y. Sibaroni, and A. F. Ihsan, "Analysis content type and emotion of the presidential election users tweets using agglomerative hierarchical clustering," *Sinkron, Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 3, pp. 1230–1237, Jul. 2023.
[157] J. Fang, D. K. Chiu, and K. K. Ho, "Exploring cryptocurrency sentiments with clustering text mining on social media," in *Intelligent Analytics With Advanced Multi-Industry Applications*. Hershey, PA, USA: IGI Global, 2021, pp. 157–171.
[158] M. A. Siddiqui, S. A. Ali, and N. G. Haider, "Reduced feature set for emotion based spoken utterances of normal and special children using multivariate analysis and decision trees," *Eng., Technol. Appl. Sci. Res.*, vol. 8, no. 4, pp. 3228–3233, Aug. 2018.
[159] O. Zhurakovskaya, L. Steinkamp, K. M. Tymann, and C. Gips, "EmoDex: An emotion detection tool composed of established techniques," in *Proc. Conf. Lernen, Wissen, Daten, Analysen (LWDA)*, 2020, pp. 105–116.
[160] K. Mohan Kudiri, A. Md Said, and M. Y. Nayan, "Emotion detection using relative amplitude-based features through speech," in *Proc. Int. Conf. Comput. Inf. Sci. (ICCIS)*, vol. 1, Jun. 2012, pp. 522–525.
[161] A. Yudhana, A. Muslim, D. E. Wati, I. Puspitasari, A. Azhari, and M. M. Mardhia, "Human emotion recognition based on EEG signal using fast Fourier transform and K-nearest neighbor," *Adv. Sci., Technol. Eng. Syst. J.*, vol. 5, no. 6, pp. 1082–1088, 2020.
[162] M. Batta, "Machine learning algorithms—A review," *Int. J. Sci. Res.*, vol. 9, no. 1, pp. 381–386, 2020.
[163] N. M. A. Lestari, I. K. G. D. Putra, and A. K. A. Cahyawan, "Personality types classification for Indonesian text in partners searching website using Naïve Bayes methods," *Int. J. Comput. Sci. Issues*, vol. 10, no. 1, p. 1, 2013.

[164] M. A. H. Rafath, F. T. Z. Mim, and M. S. Rahman, "An analytical study on music listener emotion through logistic regression," *World Academics J. Eng. Sci.*, vol. 8, no. 3, pp. 15–20, 2021.

[165] M. C. Guerrero, J. S. Parada, and H. E. Espitia, "EEG signal analysis using classification techniques: Logistic regression, artificial neural networks, support vector machines, and convolutional neural networks," *Heliyon*, vol. 7, no. 6, Jun. 2021, Art. no. e07258.

[166] N. R. Bhowmik, M. Arifuzzaman, M. R. H. Mondal, and M. S. Islam, "Bangla text sentiment analysis using supervised machine learning with extended lexicon dictionary," *Natural Lang. Process. Res.*, vol. 1, nos. 3–4, p. 34, 2021.

[167] F. Aljwari, "Emotion detection in Arabic text using machine learning methods," *Int. J. Inf. Syst. Comput. Sci.*, vol. 6, no. 3, pp. 175–185, 2022.

[168] M. Hassan, S. Shakil, N. N. Moon, M. M. Islam, R. A. Hossain, A. Mariam, and F. N. Nur, "Sentiment analysis on Bangla conversation using machine learning approach," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 12, no. 5, p. 5562, Oct. 2022.

[169] T. Jain, P. Garg, R. Gupta, P. Goyal, and N. Chalil, "Sentiment analysis of movie reviews in Hindi language using machine learning," in *Proc. IEEE Int. Conf. Data Sci. Inf. Syst. (ICDSIS)*, Jul. 2022, pp. 1–5.

[170] D. Zhou, X. Zhang, Y. Zhou, Q. Zhao, and X. Geng, "Emotion distribution learning from texts," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2016, pp. 638–647.

[171] O. Rabie and C. Sturm, "Feel the heat: Emotion detection in Arabic social media content," in *Proc. Int. Conf. Data Mining, Internet Comput. Big Data (BigData)*, 2014, pp. 37–49.

[172] L. Sana, K. Nasir, A. Urooj, Z. Ishaq, and I. A. Hameed, "BERS: Bussiness-related emotion recognition system in Urdu language using machine learning," in *Proc. 5th Int. Conf. Behav., Econ., Socio-Cultural Comput. (BESC)*, Nov. 2018, pp. 238–242.

[173] X. Quan, Q. Wang, Y. Zhang, L. Si, and L. Wenyin, "Latent discriminative models for social emotion detection with emotional dependency," *ACM Trans. Inf. Syst.*, vol. 34, no. 1, pp. 1–19, Oct. 2015.

[174] Y. Rao, H. Xie, J. Li, F. Jin, F. L. Wang, and Q. Li, "Social emotion classification of short text via topic-level maximum entropy model," *Inf. Manage.*, vol. 53, no. 8, pp. 978–986, Dec. 2016.

[175] S. Y. M. Lee and Z. Wang, "Multi-view learning for emotion detection in code-switching texts," in *Proc. Int. Conf. Asian Lang. Process. (IALP)*, Oct. 2015, pp. 90–93.

[176] C. O. Alm, D. Roth, and R. Sproat, "Emotions from text: Machine learning for text-based emotion prediction," in *Proc. Conf. Human Lang. Technol. Empirical Methods Natural Lang. Process.*, 2005, pp. 579–586.

[177] B. Sun, Q. Ma, S. Zhang, K. Liu, and Y. Liu, "iSelf: Towards cold-start emotion labeling using transfer learning with smartphones," *ACM Trans. Sensor Netw.*, vol. 13, no. 4, pp. 1–22, 2017.

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

[179] E. Junianto and R. Rachman, "Implementation of text mining model to emotions detection on social media comments using particle swarm optimization and Naïve Bayes classifier," in *Proc. 7th Int. Conf. Cyber IT Service Manage. (CITSM)*, vol. 7, Nov. 2019, pp. 1–6.

[180] R. Rana and V. Kolhe, "Analysis of students emotion for Twitter data using Naïve Bayes and non linear support vector machine approachs," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 3, no. 5, pp. 2321–8169, 2015.

[181] D. Yasmina, M. Hajar, and A. M. Hassan, "Using Youtube comments for text-based emotion recognition," *Proc. Comput. Sci.*, vol. 83, pp. 292–299, Jan. 2016.

[182] M. Hasan, E. Agu, and E. Rundensteiner, "Using hashtags as labels for supervised learning of emotions in Twitter messages," in *Proc. ACM SIGKDD Workshop Health Informat. (HI-KDD)*, 2014, vol. 34, no. 74, p. 100.

[183] M. Hasan, E. Rundensteiner, and E. Agu, "EMOTEX: Detecting emotions in Twitter messages," in *Proc. ASE BIGDATA/SOCIALCOM/CYBERSECURITY Conf.*, 2014, pp. 1–10.

[184] L. Wikarsa and S. N. Thahir, "A text mining application of emotion classifications of Twitter's users using Naïve Bayes method," in *Proc. 1st Int. Conf. Wireless Telematics (ICWT)*, Nov. 2015, pp. 1–6.

[185] A. M. G. de Almeida, S. Barbon, and E. C. Paraiso, "Multi-class emotions classification by sentic levels as features in sentiment analysis," in *Proc. 5th Brazilian Conf. Intell. Syst. (BRACIS)*, Brazil, Oct. 2016, pp. 486–491.

[186] S. Aman and S. Szpakowicz, "Identifying expressions of emotion in text," in *Proc. Int. Conf. Text, Speech Dialogue (TSD)*, 2007, pp. 196–205.

[187] M. Hasan, E. Rundensteiner, and E. Agu, "Automatic emotion detection in text streams by analyzing Twitter data," *Int. J. Data Sci. Anal.*, vol. 7, no. 1, pp. 35–51, Feb. 2019.

[188] M. Allouch, A. Azaria, R. Azoulay, E. Ben-Izchak, M. Zwilling, and D. A. Zachor, "Automatic detection of insulting sentences in conversation," in *Proc. IEEE Int. Conf. Sci. Electr. Eng. Isr. (ICSEE)*, Dec. 2018, pp. 1–4.

[189] L. Singh, S. Singh, and N. Aggarwal, "Two-stage text feature selection method for human emotion recognition," in *Proc. 2nd Int. Conf. Commun., Comput. Netw. (ICCCN)*, 2019, pp. 531–538.

[190] A. A. Maruf, Z. M. Ziyad, M. M. Haque, and F. Khanam, "Emotion detection from text and sentiment analysis of Ukraine Russia war using machine learning technique," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 12, pp. 868–882, 2022.

[191] W. Bian, C. Wang, Z. Ye, and L. Yan, "Emotional text analysis based on ensemble learning of three different classification algorithms," in *Proc. 10th IEEE Int. Conf. Intell. Data Acq. Adv. Comput. Syst. Technol. Appl.*, vol. 2, 2019, pp. 938–941.

[192] S. A. Deborah, S. Rajalakshmi, S. M. Rajendram, and T. T. Mirnalinee, "Contextual emotion detection in text using ensemble learning," in *Proc. Int. Conf. Emerg. Current Trends Comput. Expert Technol. (COMET)*, 2019, pp. 1179–1186.

[193] D. P. Gaikwad and R. C. Thool, "Intrusion detection system using bagging ensemble method of machine learning," in *Proc. Int. Conf. Comput. Commun. Control Autom.*, Feb. 2015, pp. 291–295.

[194] A. Mosavi, F. S. Hosseini, B. Choubin, M. Goodarzi, A. A. Dineva, and E. R. Sardooi, "Ensemble boosting and bagging based machine learning models for groundwater potential prediction," *Water Resour. Manage.*, vol. 35, no. 1, pp. 23–37, Jan. 2021.

[195] M. A. H. Wadud, M. M. Kabir, M. F. Mridha, M. A. Ali, M. A. Hamid, and M. M. Monowar, "How can we manage offensive text in social media—A text classification approach using LSTM-BOOST," *Int. J. Inf. Manage. Data Insights*, vol. 2, no. 2, Nov. 2022, Art. no. 100095.

[196] S. G. Tesfagergish, J. Kapočiūtė-Dzikienė, and R. Damaševičius, "Zero-shot emotion detection for semi-supervised sentiment analysis using sentence transformers and ensemble learning," *Appl. Sci.*, vol. 12, no. 17, p. 8662, Aug. 2022.

[197] W. Zehra, A. R. Javed, Z. Jalil, H. U. Khan, and T. R. Gadekallu, "Cross corpus multi-lingual speech emotion recognition using ensemble learning," *Complex Intell. Syst.*, vol. 7, no. 4, pp. 1845–1854, Aug. 2021.

[198] A. Mansy, S. Rady, and T. Gharib, "An ensemble deep learning approach for emotion detection in Arabic tweets," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 4, pp. 980–990, 2022.

[199] A. Mostafa, O. Mohamed, and A. Ashraf, "GOF at Arabic hate speech 2022: Breaking the loss function convention for data-imbalanced Arabic offensive text detection," in *Proc. 5th Workshop Open-Source Arabic Corpora Processing Tools Shared Tasks Qur'an QA Fine-Grained Hate Speech Detection*, 2022, pp. 167–175.

[200] R. D. G. Ayon, M. S. Rabbi, U. Habiba, and M. Hasana, "Bangla speech emotion detection using machine learning ensemble methods," *Adv. Sci., Technol. Eng. Syst. J.*, vol. 7, no. 6, pp. 70–76, Nov. 2022.

[201] K. Sarkar, "Heterogeneous classifier ensemble for sentiment analysis of Bengali and Hindi tweets," *Sādhanā*, vol. 45, no. 1, p. 196, Dec. 2020.

[202] F. Ghanbari-Adivi and M. Mosleh, "Text emotion detection in social networks using a novel ensemble classifier based on Parzen tree estimator (TPE)," *Neural Comput. Appl.*, vol. 31, no. 12, pp. 8971–8983, Dec. 2019.

[203] K. Yang, Y. Gao, L. Liang, S. Bian, L. Chen, and B. Zheng, "CrowdTC: Crowd-powered learning for text classification," *ACM Trans. Knowl. Discovery from Data*, vol. 16, no. 1, pp. 1–23, Feb. 2022.

[204] M. H. K. Vardag, A. Saeed, U. Hayat, M. F. Ullah, and N. Hussain, "Contextual Urdu text emotion detection corpus and experiments using deep learning approaches," *ADCAIJ, Adv. Distrib. Comput. Artif. Intell. J.*, vol. 11, no. 4, pp. 489–505, Jun. 2023.

[205] P. Chandra, M. T. Ahammed, S. Ghosh, R. H. Emon, M. Billah, M. I. Ahamad, and P. Balaji, "Contextual emotion detection in text using deep learning and big data," in *Proc. 2nd Int. Conf. Comput. Sci., Eng. Appl. (ICCSEA)*, Sep. 2022, pp. 1–5.

[206] A. M. Almars, "Attention-based bi-LSTM model for Arabic depression classification," *Comput., Mater. Continua*, vol. 71, no. 2, pp. 3091–3106, 2022.

[207] M. F. Bashir, A. R. Javed, M. U. Arshad, T. R. Gadekallu, W. Shahzad, and M. O. Beg, "Context-aware emotion detection from low-resource Urdu language using deep neural network," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 5, pp. 1–30, May 2023.
[208] R. Rahman, S. A. Hasan, and F. A. Rubel, "Identifying sentiment and recognizing emotion from social media data in Bangla language," in *Proc. 12th Int. Conf. Electr. Comput. Eng. (ICECE)*, Dec. 2022, pp. 36–39.
[209] S. Poria, I. Chaturvedi, E. Cambria, and A. Hussain, "Convolutional MKL based multimodal emotion recognition and sentiment analysis," in *Proc. IEEE 16th Int. Conf. Data Mining (ICDM)*, Dec. 2016, pp. 439–448.
[210] S. Ahmad, M. Z. Asghar, F. M. Alotaibi, and S. Khan, "Classification of poetry text into the emotional states using deep learning technique," *IEEE Access*, vol. 8, pp. 73865–73878, 2020.
[211] M. Baali and N. Ghneim, "Emotion analysis of Arabic tweets using deep learning approach," *J. Big Data*, vol. 6, no. 1, pp. 1–12, Dec. 2019.
[212] H. Al-Omari, M. A. Abdullah, and S. Shaikh, "EmoDet2: Emotion detection in English textual dialogue using BERT and BiLSTM models," in *Proc. 11th Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2020, pp. 226–232.
[213] U. Rashid, M. W. Iqbal, M. A. Skiandar, M. Q. Raiz, M. R. Naqvi, and S. K. Shahzad, "Emotion detection of contextual text using deep learning," in *Proc. 4th Int. Symp. Multidisciplinary Stud. Innov. Technol. (ISMSIT)*, Oct. 2020, pp. 1–5.
[214] M. Jabreel and A. Moreno, "A deep learning-based approach for multi-label emotion classification in tweets," *Appl. Sci.*, vol. 9, no. 6, p. 1123, Mar. 2019.
[215] S. S. Sadeghi, H. Khotanlou, and M. R. Mahand, "Automatic Persian text emotion detection using cognitive linguistic and deep learning," *J. AI Data Mining*, vol. 9, no. 2, pp. 169–179, 2021.
[216] M. Polignano, P. Basile, M. de Gemmis, and G. Semeraro, "A comparison of word-embeddings in emotion detection from text using BiLSTM, CNN and self-attention," in *Proc. Adjunct Publication 27th Conf. User Model., Adaptation Personalization (UMAP)*, Jun. 2019, pp. 63–68.
[217] D. Haryadi and G. Putra, "Emotion detection in text using nested long short-term memory," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 6, pp. 351–357, 2019.
[218] 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 Appl.*, vol. 78, no. 20, pp. 29607–29639, Oct. 2019.
[219] 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," 2019, *arXiv:1907.07653*.
[220] F. Xia and Z. Zhang, "Study of text emotion analysis based on deep learning," in *Proc. 13th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, May 2018, pp. 2716–2720.
[221] T. Liu, Y. Du, and Q. Zhou, "Text emotion recognition using GRU neural network with attention mechanism and emoticon emotions," in *Proc. 2nd Int. Conf. Robot., Intell. Control Artif. Intell.*, Oct. 2020, pp. 278–282.
[222] N. Kanger and G. Bathla, "Recognizing emotion in text using neural network and fuzzy logic," *Indian J. Sci. Technol.*, vol. 10, no. 12, pp. 1–6, May 2017.
[223] M. Abdullah, M. Hadzikadicy, and S. Shaikhz, "SEDAT: Sentiment and emotion detection in Arabic text using CNN-LSTM deep learning," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2018, pp. 835–840.
[224] D. Xu, Z. Tian, R. Lai, X. Kong, Z. Tan, and W. Shi, "Deep learning based emotion analysis of microblog texts," *Inf. Fusion*, vol. 64, pp. 1–11, Dec. 2020.
[225] G. Lample and A. Conneau, "Cross-lingual language model pretraining," 2019, *arXiv:1901.07291*.
[226] G. Lample, M. Ott, A. Conneau, L. Denoyer, and M. Ranzato, "Phrase-based & neural unsupervised machine translation," 2018, *arXiv:1804.07755*.
[227] W. Antoun, F. Baly, and H. Hajj, "AraBERT: Transformer-based model for Arabic language understanding," 2020, *arXiv:2003.00104*.
[228] A. Safaya, M. Abdullatif, and D. Yuret, "KUISAIL at SemEval-2020 task 12: BERT-CNN for offensive speech identification in social media," in *Proc. 14th Workshop Semantic Eval. (SemEval)*, 2020, pp. 2054–2059.
[229] W. Antoun, F. Baly, and H. Hajj, "AraELECTRA: Pre-training text discriminators for Arabic language understanding," 2020, *arXiv:2012.15516*.
[230] M. Abdul-Mageed, A. Elmadany, and E. M. B. Nagoudi, "ARBERT & MARBERT: Deep bidirectional transformers for Arabic," 2020, *arXiv:2101.01785*.
[231] R. Nayak and R. Joshi, "L3Cube-HingCorpus and HingBERT: A code mixed hindi-english dataset and BERT language models," 2022, *arXiv:2204.08398*.
[232] A. Bhattacharjee, T. Hasan, W. U. Ahmad, K. Samin, M. S. Islam, A. Iqbal, M. S. Rahman, and R. Shahriyar, "BanglaBERT: Language model pretraining and benchmarks for low-resource language understanding evaluation in Bangla," 2021, *arXiv:2101.00204*.
[233] M. Hasanain, F. Alam, H. Mubarak, S. Abdaljalil, W. Zaghouani, P. Nakov, G. D. S. Martino, and A. A. Freihat, "ArAIEval shared task: Persuasion techniques and disinformation detection in Arabic text," 2023, *arXiv:2311.03179*.
[234] M. S. U. Sourav, H. Wang, M. S. Mahmud, and H. Zheng, "Transformer-based text classification on unified Bangla multi-class emotion corpus," 2022, *arXiv:2210.06405*.
[235] N. Alvarez-Gonzalez, A. Kaltenbrunner, and V. Gómez, "Uncovering the limits of text-based emotion detection," 2021, *arXiv:2109.01900*.
[236] S. Hassan, S. Shaar, and K. Darwish, "Cross-lingual emotion detection," 2021, *arXiv:2106.06017*.
[237] H. Yang and J. Shen, "Emotion dynamics modeling via BERT," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2021, pp. 1–8.
[238] S. Peng, R. Zeng, H. Liu, G. Chen, R. Wu, A. Yang, and S. Yu, "Emotion classification of text based on BERT and broad learning system," in *Proc. 5th Int. Joint Conf. Web Big Data (APWeb-WAIM)*, 2021, pp. 382–396.
[239] S. Park, J. Kim, S. Ye, J. Jeon, H. Y. Park, and A. Oh, "Dimensional emotion detection from categorical emotion," 2019, *arXiv:1911.02499*.
[240] A. F. Adoma, N.-M. Henry, and W. Chen, "Comparative analyses of BERT, RoBERTa, DistilBERT, and XLNet for text-based emotion recognition," in *Proc. 17th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. (ICCWAMTIP)*, Dec. 2020, pp. 117–121.
[241] H. El-Zayady, M. S. Mohamed, K. Badran, and G. Salama, "A hybrid approach based on personality traits for hate speech detection in Arabic social media," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 13, no. 2, p. 1979, Apr. 2023.
[242] M. T. Ahmed, A. S. Urmi, M. Rahman, A. Z. M. T. Islam, D. Das, and M. G. Rashed, "Cyberbullying detection based on hybrid ensemble method using deep learning technique in Bangla dataset," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 9, 2023.
[243] S. Grover and A. Verma, "Design for emotion detection of Punjabi text using hybrid approach," in *Proc. Int. Conf. Inventive Comput. Technol. (ICICT)*, vol. 2, Aug. 2016, pp. 1–6.
[244] S. Azmin and K. Dhar, "Emotion detection from Bangla text corpus using Naïve Bayes classifier," in *Proc. 4th Int. Conf. Electr. Inf. Commun. Technol. (EICT)*, 2019, pp. 1–5.
[245] Y. Douiji and H. Mousanif, "I-CARE: Intelligent context aware system for recognizing emotions from text," in *Proc. 10th Int. Conf. Intell. Systems: Theories Appl. (SITA)*, Oct. 2015, pp. 1–5.
[246] D. Chatzakou, A. Vakali, and K. Kafetsios, "Detecting variation of emotions in online activities," *Expert Syst. Appl.*, vol. 89, pp. 318–332, Dec. 2017.
[247] V. K. Jain, S. Kumar, and S. L. Fernandes, "Extraction of emotions from multilingual text using intelligent text processing and computational linguistics," *J. Comput. Sci.*, vol. 21, pp. 316–326, Jul. 2017.
[248] J. Herzig, M. Shmueli-Scheuer, and D. Konopnicki, "Emotion detection from text via ensemble classification using word embeddings," in *Proc. ACM SIGIR Int. Conf. Theory Inf. Retr.*, Oct. 2017, pp. 269–272.
[249] O. Udochukwu and Y. He, "A rule-based approach to implicit emotion detection in text," in *Proc. Int. Conf. Appl. Natural Lang. Inf. Syst. (NLDB)*, 2015, pp. 197–203.
[250] D. Krstinić, M. Braović, L. Šerić, and D. Božić-Štulić, "Multi-label classifier performance evaluation with confusion matrix," in *Proc. Comput. Sci. Inf. Technol.*, vol. 1, 2020, pp. 1–14.
[251] S. Marsland, *Machine Learning: An Algorithmic Perspective*, 2nd ed. London, U.K.: Chapman & Hall, 2014.
[252] B. Rahman and H. U. Rashid, "Teaching Arabic to non-native speakers: A foresight vision," *Perdana, Int. J. Academic Res.*, vol. 7, no. 1, pp. 67–85, 2020.
[253] S. Havaldar, S. Rai, B. Singhal, L. Liu, S. C. Guntuku, and L. Ungar, "Multilingual language models are not multicultural: A case study in emotion," 2023, *arXiv:2307.01370*.

[254] H. Saadany, C. Orasan, R. C. Quintana, F. do Carmo, and L. Zilio, "Challenges in translation of emotions in multilingual user-generated content: Twitter as a case study," 2021, *arXiv:2106.10719*.

[255] C. Yao, X. Song, X. Zhang, W. Zhao, and A. Feng, "Multitask learning for aspect-based sentiment classification," *Scientific Program.*, vol. 2021, pp. 1–9, Nov. 2021.

[256] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "An interactive multi-task learning network for end-to-end aspect-based sentiment analysis," 2019, *arXiv:1906.06906*.

[257] J. Bross and H. Ehrig, "Generating a context-aware sentiment lexicon for aspect-based product review mining," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol.*, vol. 1, Aug. 2010, pp. 435–439.

[258] P. Michelle, F. Z. Ruskanda, and A. Purwarianti, "Development of domain-specific lexicon for aspect-based sentiment analysis," in *Proc. 7th Int. Conf. Advance Informatics: Concepts, Theory Appl. (ICAICTA)*, Sep. 2020, pp. 1–6.

[259] S. Kusal, S. Patil, J. Choudrie, K. Kotecha, D. Vora, and I. Pappas, "A review on text-based emotion detection - techniques, applications, datasets, and future directions," 2022, *arXiv:2205.03235*.

[260] E. Omara, M. Mosa, and N. Ismail, "Emotion analysis in Arabic language applying transfer learning," in *Proc. 15th Int. Comput. Eng. Conf. (ICENCO)*, Dec. 2019, pp. 204–209.

[261] D. Hazarika, S. Poria, R. Zimmermann, and R. Mihalcea, "Conversational transfer learning for emotion recognition," *Inf. Fusion*, vol. 65, pp. 1–12, Jan. 2021.

[262] K. Yasaswini, K. Puranik, A. Hande, R. Priyadharshini, S. Thavareesan, and B. R. Chakravarthi, "IIITT@ DravidianLangTech-EACL2021: Transfer learning for offensive language detection in Dravidian languages," in *Proc. 1st Workshop Speech Lang. Technol. Dravidian Lang. (DravidianLangTech)@ECAL*, 2021, pp. 187–194.

[263] Y. Zhang and Q. Yang, "An overview of multi-task learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 30–43, Jan. 2018.

[264] X. Liang, Y. Zou, X. Zhuang, J. Yang, T. Niu, and R. Xu, "MMATERIC: Multi-task learning and multi-fusion for AudioText emotion recognition in conversation," *Electronics*, vol. 12, no. 7, p. 1534, Mar. 2023.

[265] L. Huang, D. Ma, S. Li, X. Zhang, and H. Wang, "Text level graph neural network for text classification," 2019, *arXiv:1910.02356*.

[266] L. Yao, C. Mao, and Y. Luo, "Graph convolutional networks for text classification," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 7370–7377.

[267] P. Battaglia, R. Pascanu, M. Lai, D. J. Rezende, and K. Kavukcuoglu, "Interaction networks for learning about objects, relations and physics," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 4509–4517.

[268] M. Firdaus, G. V. Singh, A. Ekbal, and P. Bhattacharyya, "Affect-GCN: A multimodal graph convolutional network for multi-emotion with intensity recognition and sentiment analysis in dialogues," *Multimedia Tools Appl.*, vol. 82, no. 28, pp. 43251–43272, Nov. 2023.

[269] W. Nie, R. Chang, M. Ren, Y. Su, and A. Liu, "I-GCN: Incremental graph convolution network for conversation emotion detection," *IEEE Trans. Multimedia*, vol. 24, pp. 4471–4481, 2022.

[270] R. Xia and Y. Liu, "Using denoising autoencoder for emotion recognition," in *Proc. 14th Annu. Conf. Int. Speech Commun. Assoc. (Interspeech)*, Aug. 2013, pp. 2886–2889.

[271] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[272] S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," 2015, *arXiv:1511.06349*.

[273] D. P. Kingma, S. Mohamed, D. J. Rezende, and M. Welling, "Semi-supervised learning with deep generative models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 3581–3589.

[274] X. Fu, Y. Wei, F. Xu, T. Wang, Y. Lu, J. Li, and J. Z. Huang, "Semi-supervised aspect-level sentiment classification model based on variational autoencoder," *Knowl.-Based Syst.*, vol. 171, pp. 81–92, May 2019.

[275] Y. Vorobeychik, M. Kantarcioglu, R. Brachman, P. Stone, and F. Rossi, *Adversarial Machine Learning*. San Rafael, CA, USA: Morgan & Claypool, 2018.

[276] W. E. Zhang, Q. Z. Sheng, A. Alhazmi, and C. L. Li, "Adversarial attacks on deep-learning models in natural language processing: A survey," *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 3, pp. 1–41, 2020.

[277] Y. Deldjoo, T. D. Noia, and F. A. Merra, "A survey on adversarial recommender systems: From attack/defense strategies to generative adversarial networks," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–38, Mar. 2022.

[278] A. Bajaj and D. K. Vishwakarma, "Evading text based emotion detection mechanism via adversarial attacks," *Neurocomputing*, vol. 558, Nov. 2023, Art. no. 126787.

[279] N. Junsomboon and T. Phienthrakul, "Combining over-sampling and under-sampling techniques for imbalance dataset," in *Proc. 9th Int. Conf. Mach. Learn. Comput.*, Feb. 2017, pp. 243–247.

[280] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine learning with oversampling and undersampling techniques: Overview study and experimental results," in *Proc. 11th Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2020, pp. 243–248.

[281] A. Estabrooks and N. Japkowicz, "A mixture-of-experts framework for learning from imbalanced data sets," in *Proc. Int. Symp. Intell. Data Anal. (IDA)*, 2001, pp. 34–43.

[282] A. Anand, K. Gorde, J. R. A. Moniz, N. Park, T. Chakraborty, and B.-T. Chu, "Phishing URL detection with oversampling based on text generative adversarial networks," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 1168–1177.

[283] Y. Luo, H. Feng, X. Weng, K. Huang, and H. Zheng, "A novel oversampling method based on SeqGAN for imbalanced text classification," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 2891–2894.

[284] V. W. de Vargas, J. A. S. Aranda, R. dos Santos Costa, P. R. da Silva Pereira, and J. L. V. Barbosa, "Imbalanced data preprocessing techniques for machine learning: A systematic mapping study," *Knowl. Inf. Syst.*, vol. 65, no. 1, pp. 31–57, Jan. 2023.

[285] G. Takawane, A. Phaltankar, V. Patwardhan, A. Patil, R. Joshi, and M. S. Takalikar, "Language augmentation approach for code-mixed text classification," *Natural Lang. Process. J.*, vol. 5, Dec. 2023, Art. no. 100042.

[286] W. Zhao, T. Joshi, V. N. Nair, and A. Sudjianto, "SHAP values for explaining CNN-based text classification models," 2020, *arXiv:2008.11825*.

[287] D. Mardaoui and D. Garreau, "An analysis of lime for text data," in *Proc. Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2021, pp. 3493–3501.

[288] S. Velampalli, C. Muniyappa, and A. Saxena, "Performance evaluation of sentiment analysis on text and emoji data using end-to-end, transfer learning, distributed and explainable AI models," *J. Adv. Inf. Technol.*, vol. 13, no. 2, pp. 167–172, 2022.

[289] K. R. Chowdhury, A. Sil, and S. R. Shukla, "Explaining a black-box sentiment analysis model with local interpretable model diagnostics explanation (LIME)," in *Proc. 5th Int. Conf. Adv. Comput. Data Sci. (ICACDS)*, 2021, pp. 90–101.

[290] Y. Li, J. Chan, G. Peko, and D. Sundaram, "An explanation framework and method for AI-based text emotion analysis and visualisation," *Decis. Support Syst.*, vol. 178, Mar. 2024, Art. no. 114121.

ABDULLAH AL MARUF received the B.Sc. degree from the Bangladesh University of Business and Technology (BUBT) in 2022. This milestone set the stage for his subsequent successes. He began his journey with a strong foundation. He stepped into a software engineer position after completing his academic studies with ease, where he developed his algorithmic thinking and problem-solving abilities. His quest eventually took him to the exciting field of machine learning, driven by his unrelenting curiosity. He accepted the post of Machine Learning Engineer, a position that used his existing experiences while opening him new opportunities for advancement, with a thirst for innovation. His research interests include bioinformatics, image processing, human psychology in AI, and natural language processing (NLP). Due to this intense interest, he has published several noteworthy research articles in esteemed journals and at important international conferences.

FAHIMA KHANAM received the B.Sc. degree in computer science and engineering from the Military Institute of Science and Technology (MIST) and the M.Sc. degree in information and communication technology from the Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh. She is currently a Lecturer with the Department of Computer Science and Engineering, Bangladesh University of Business and Technology (BUBT), Dhaka. Her research interests include machine learning, data mining, bioinformatics, network security, and natural language processing (NLP).

MD. MAHMUDUL HAQUE received the B.Sc. degree from the Bangladesh University of Business and Technology (BUBT). This major achievement shaped a path at the nexus of computer science, artificial intelligence (AI), and natural language processing (NLP), establishing the groundwork for a series of subsequent successes. He started his journey with a solid academic foundation. He showed an unshakable dedication to exploring the limits of knowledge and solving practical problems throughout his academic career. As his interest deepened, he dipped headfirst into the complex worlds of human emotion and psychology, AI technology, and the nuanced workings of natural language processing. His pursuit of knowledge and exploration took off as a result of his intrinsic desire to understand the workings of the human mind and his unwavering commitment to encouraging well-being. His research has appeared in major journals and conferences across the world. His research outputs are evidence not only of his technical prowess but also of his unwavering dedication to illuminating the complex interactions between AI and mental health.

ZAKARIA MASUD JIYAD received the bachelor's degree in computer science from the Bangladesh University of Business and Technology. His academic foundation provided him with a solid grounding in computer science. He is currently a Researcher at the intersection of computer science, artificial intelligence (AI), and mental health advocacy. His journey in academia has been marked by a relentless pursuit of innovative solutions to real-world challenges. As his interest grew, he diligently immersed himself in the realms of human psychology, AI technologies, and natural language processing. Armed with a deep-seated passion for unraveling the complexities of the human mind and a commitment to fostering well-being, his research trajectory took flight. His work has been distinguished by several important contributions, with his research publications having a big influence on prestigious journals and conferences worldwide. His insightful work demonstrates his technical prowess and his constant commitment to illuminating the complex interactions between AI and mental health. As he continues to explore new horizons and forge innovative pathways, he emerges as a trailblazer in the dynamic landscape of AI-driven research.

M. F. MRIDHA (Senior Member, IEEE) received the Ph.D. degree in AI/ML from Jahangirnagar University, in 2017. In June 2007, he joined the Department of Computer Science and Engineering, Stamford University Bangladesh, as a Lecturer, where he was promoted to a Senior Lecturer and an Assistant Professor, in October 2010 and October 2011, respectively. In May 2012, he joined the University of Asia Pacific (UAP), as an Assistant Professor. From 2012 to 2019, he was a Faculty Member of the CSE Department, UAP, and a Graduate Coordinator. He is currently working as an Associate Professor with the Department of Computer Science, American International University-Bangladesh (AIUB), Bangladesh. For more than ten years, he has been with the bachelor's and master's students, as a supervisor of their thesis work. His research experience, within both academia and industry, has resulted in more than 80 journals and conference publications. His research interests include artificial intelligence (AI), machine learning, deep learning, big data analysis, and natural language processing (NLP). He has served as a program committee member at several international conferences and workshops. He also served as an associate editor for several journals.

ZEYAR AUNG (Senior Member, IEEE) received the Ph.D. degree in computer science from the National University of Singapore, in 2006. From 2006 to 2010, he was a Research Fellow with the Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore. In 2010, he joined the Masdar Institute, which later became a part of Khalifa University, United Arab Emirates, as an Assistant Professor. He is currently an Associate Professor with the Department of Computer Science, Khalifa University. His previous research interests were in bioinformatics and chemoinformatics. His current research interests include data analytics, machine learning, and their applications in various domains, such as cyber security, social media, financial systems, renewable energy, and environmental science.