

METHODOLOGY

Open Access



# Emotion AWARE: an artificial intelligence framework for adaptable, robust, explainable, and multi-granular emotion analysis

Gihan Gamage<sup>1</sup>, Daswin De Silva<sup>1\*</sup>, Nishan Mills<sup>1</sup>, Dammindala Alahakoon<sup>1</sup> and Milos Manic<sup>2</sup>

\*Correspondence:  
d.desilva@latrobe.edu.au

<sup>1</sup> Centre for Data Analytics and Cognition, La Trobe University, Melbourne, Australia

<sup>2</sup> Department of Computer Science, Virginia Commonwealth University, Richmond, USA

## Abstract

Emotions are fundamental to human behaviour. How we feel, individually and collectively, determines how humanity evolves and advances into our shared future. The rapid digitalisation of our personal, social and professional lives means we are frequently using digital media to express, understand and respond to emotions. Although recent developments in Artificial Intelligence (AI) are able to analyse sentiment and detect emotions, they are not effective at comprehending the complexity and ambiguity of digital emotion expressions in knowledge-focused activities of customers, people, and organizations. In this paper, we address this challenge by proposing a novel AI framework for the adaptable, robust, and explainable detection of multi-granular assembles of emotions. This framework consolidates lexicon generation and finetuned Large Language Model (LLM) approaches to formulate multi-granular assembles of two, eight and fourteen emotions. The framework is robust to ambiguous emotion expressions that are implied in conversation, adaptable to domain-specific emotion semantics, and the assembles are explainable using constituent terms and intensity. We conducted nine empirical studies using datasets representing diverse human emotion behaviours. The results of these studies comprehensively demonstrate and evaluate the core capabilities of the framework, and consistently outperforms state-of-the-art approaches in adaptable, robust, and explainable multi-granular emotion detection.

## Introduction

The rapid digitalisation of society has empowered knowledge-focussed human activities and communication to transpire on hyper-connected, digital platforms. This spectrum of intrapersonal, interpersonal, and group activities have led to the generation and management of high volumes of big social data that represents patterns of behaviour of individuals and organizations, and how they leverage insights drawn from that information for further engagement and collaborative activities [1]. Expressions of emotion are encapsulated in these digital platforms which is highly useful towards accurately modelling human behaviour [2]. The persistence of this textual digital record enables the use of computational approaches to process, analyse and synthesise emotion expressions. Computational approaches for emotion detection have been classified using several

schemes in existing literature. Acheampong et al. [3]. proposed three categories, rule-based, machine learning and hybrid methods. Alswaidan et al. [4] proposed a scheme of five categories, keyword-based, rule-based, classical learning, deep learning and hybrid. In reviewing these schemes, we have summarised into three technical categories, (1) heuristics (which includes keywords, rule-based, probabilistic and statistical), (2) Artificial Intelligence (AI) (consisting of classical learning, machine reasoning and deep learning) and (3) hybrids of the two. Despite the maturity of this topic in terms of classification schemes and the prevalence of many approaches across these three classes, the complexity and ambiguity of emotion expressions on digital platforms have not been fully addressed. We substantiated this challenge of complexity and ambiguity in terms of four capabilities, (1) output (granularity of emotion detection output), (2) domain specificity, (3) adaptability, and (4) explainability.

We conducted a systematic literature review of the state-of-the-art of recent emotion analysis and detection research published in the last five years, from 2018 to 2022. The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram for this review is reported in Supplementary Fig. 1 (Filename: emotionaware supp Fig. 1.docx). The review produced 83 articles that aligned with the selection criteria, which then we evaluated in terms of the four capabilities noted above. Supplementary Table 1 (Filename: EmotionAwareSuppTable1.xlsx) presents the results of this evaluation.

Based on the findings of the literature review and the subsequent evaluation against capabilities, we propose a novel framework for Emotion Assembles With Adaptability Robustness and Explainability (AWARE). This Emotion AWARE framework involves heuristics and AI techniques with lexicon generation and finetuned Large Language Models (LLM) into a hetero-hierarchical structure that receives text containing emotion expressions as input and produces as output an assemble of emotions with corresponding intensity values. Emotion assembles can be created at three levels of granularity, two, eight and fourteen. The framework is adaptable as the hetero-hierarchical structure can be revised and reintroduced to reflect a domain or topic of interest. The framework is robust in its ability to detect implied emotion expressions through the context of surrounding terms as well as scale the intensity values based on negations, intensifiers, and inhibitors. The framework is explainable in its identification of terms and phrases for each emotion expression, leading up to a collection of terms that can be used to profile and compare multiple assembles.

In comparison to related work on emotion detection, the Emotion AWARE framework is novel in its construction of emotion assembles with intensity values, and the explainability, adaptability and robustness of these emotion assembles. On approach, AWARE leverages prior knowledge of lexicons and learned knowledge of the finetuned language models, in contrast to the singular approaches adopted in related work, and it is the only approach evaluated on eight datasets (across studies). In terms of output, it produces multi-granular emotion assembles of 2,8, and 14 emotions with intensity scores, in contrast to the class-based output produced by other methods. In terms of valence and arousal, the proposed framework detects valence across a broad spectrum of 14 emotion categories, and each category is assigned a score from 0 to 1. This scoring reflects arousal levels and is determined while taking modifiers and negations into consideration. All

related methods in recent literature are limited to a specific domain or general application, whereas AWARE is intrinsically generic but can be adapted to a domain of interest. This feature is aptly demonstrated in the experimental results (study 5 and (6). Explainability, adaptability and modifier resolution are similarly more advanced than those reported in existing literature, mainly due to the effectiveness of the hybrid approach of prior knowledge from lexicons and learned knowledge from finetuned language models.

### Literature review

As noted above, we conducted a systematic literature review of the state-of-the-art research on emotion analysis published in the last 5 years, from 2018 to 2022. The PRISMA flow diagram and the evaluation of the selected work against the four capabilities are reported in Supplementary Fig. 1 (Filename: EmotionAwareSuppFig1.docx) and Supplementary Table 1 (filename: EmotionAwareSuppTable1.xlsx), respectively. Here, we delineate key findings in terms of the three categories, heuristics, AI and hybrids.

Heuristic approaches include keyword recognition, rule-based logical/grammatical affinities, statistical and probabilistic methods. These methods are grounded in emotional lexicons, corpora and dictionaries that represent prior knowledge of how emotion is expressed in that domain or discipline. The emotion lexicon is typically a list of synonyms and related words used for each emotion category, where each word may also be assigned a fixed intensity value. Besides a list, the lexicon can also be organised hierarchically in a tree structure or interlinked as a graph or map structure. Several emotion lexicons reported in the literature are, Plutchik's emotional terms [5], theWordNet-Affect [6], EmoSenticNet [7], DepecheMood [8], SentiWordNet dictionaries [9]. Keyword recognition methods [10] rely on locating keywords representing emotions in a given text and assigning an emotion label based on these keyword counts and other statistics. These methods can be used for explicit emotion detection. For example, "*their arrival made me happy*" explicitly expresses the emotion happiness/joy with the keyword "*happy*". But often emotions are not explicitly mentioned and can be negated or modified to give different or opposing interpretations than a keyword search method would suggest. In such cases more advance heuristics are required. Rule-based approaches incorporate text processing methods such as tokenization, part-of-speech tagging, and dependency parsing along with corpora and lexicons to find the most effective rules sets for emotion detection [11, 12]. Several other approaches use lexical affinity with the support of lexicons to capture contextual and semantic relatedness to generate probabilistic values for each emotion category [13]. Furthermore, some approaches utilize dimensionality reduction and categorical feature extraction methods such as Latent semantic analysis (LSA) [14], Probabilistic LSA [15] for improved emotion detection [16]. The use of lexicons enables domain adaptation in emotion detection as lexicons can be easily extended or altered to suit the target domain. Furthermore, these methods can be extended for emotion intensity calculation, negation and modifier detection as they can locate the keywords and evaluate the corresponding neighbourhood. However, a major drawback of all heuristic methods is that emotion expressions that are not specified in the lexicon and those that are implied or ambiguous are not detected. Due to these reasons, methods that are purely based on lexicons are not comparable to benchmark performance of AI based methods [3].

AI-based methods can be subdivided into two, conventional supervised learning methods situated in annotated datasets and the contemporary transfer learning methods that leverage pre-training contextual language models. The conventional methods require large, labelled datasets where each sentence, paragraph or segment in the corpus is pre-assigned an emotion category (or label), typically by a human expert. This annotated dataset is used to train a multiclass classification model using supervised learning algorithms. Emotion classification and intensity calculation using XGBoost [17, 18], Support Vector Machines (SVM) [19, 20], Naïve Bayes (NB) [21, 22], k-Nearest Neighbor (kNN) [22] and Decision Trees [23, 24] are some prominent techniques reported in related literature. More recently, deep learning algorithms such as Long Short Term Memory (LSTM) networks [25, 26], Gated Recurrent Units (GRU) [27, 28] and Deep Neural Networks (DNN) [29, 30] have also been used in the same supervised learning context but with increased performance. Collectively, all supervised learning methods have reported accuracies in the range of 65–80% on benchmark datasets [3]. However, supervised learning methods are impeded by two major limitations, the scarcity of large, domain independent labelled datasets and the challenge of ambiguous and implicit emotion expressions. More recent AI methods address these limitations by leveraging the semantic context of emotion expressions embedded in pre-trained language models. Unlike supervised methods, these methods can be fine-tuned with smaller labelled datasets using transfer learning. Emotion extraction using variations of BERT [31–33], GPT [34, 35], XLNet [36, 37] are such methods that leverage the contextual knowledge embedded in language models. These approaches report state-of-the-art accuracies for emotion detection from benchmark datasets in the range of 75–99% [38]. However, this strength is also a weakness due to the limited generalisability across new, unforeseen emotion expressions, as well as intensifiers, inhibitors, and negations of emotion expressions, lack of explainability and constrained domain adaptation. Collectively, these limitations question the practical value of the high accuracies reported in empirical evaluation [39].

Several hybrid methods also have been proposed in the recent literature combining heuristics with AI methods to improve accuracy and refine the emotion categories. Tzacheva et al. (2019) [20] proposed lexicon-based emotion annotation to train SVM classifiers, for emotion extraction in tweets. Wu and Chuang [40] utilized a rule-based approach to extract semantics related to emotions and combined it with lexicon ontology to extract emotions. In Salim et al. [41], authors presented self-supervised hybrid methodology for sentiment classification from unlabelled data that combines a machine learning classifier with a lexicon-based strategy. Li et al. [42] proposed a hybrid emotion detection system combining hand crafted rules and lexicon with machine learning based classifier to extract emotional levels in online blogs.

Collectively across all three categories, the practical value of these methods in the management of information and extraction of patterns of behaviour of individuals and organizations is vast. Large scale analyses of social media during elections [43, 44], patient-centred care for chronic illnesses such as Alzheimer's disease, cancer, and diabetes [45–47], real-time depression detection on social networks [48, 49], expressions of emotion and sentiment during the COVID-19 global pandemic [50–52], highlight the practical value in social and individual settings. In organisational settings, financial

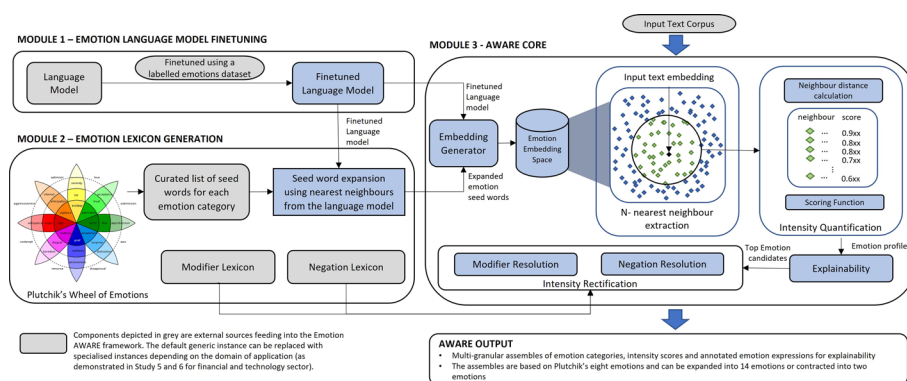
sentiment analysis [53], understanding consumer satisfactions [54], the role of social media in stock price moments [55], and the influence of review credibility and review usefulness [56] are pivotal studies that signify the continuing and incremental value of emotion analysis in digitalised content for all stakeholders.

In concluding the literature review, we elaborate on the four capabilities and their potency in addressing the challenges of the complexity and ambiguity of digital emotion expressions in knowledge-focused activities. The first capability is the output of the emotion detection approach. In most cases, this is limited to an emotion label without an intensity score for that emotion. This emotion label is also limited to a single granularity which cannot be further analysed in terms of its constituents. Most approaches assign a single emotion per atomic unit of text (sentence, paragraph or document), and overlook the presence of multiple emotions. The second capability, domain specificity relates to the generalisability of the approach across diverse domains. Most approaches are highly specific to the syntax or semantics of a given domain, such as emotions in short text like tweets [57, 58], emotions in poetry [59, 60], emotions in code switched text [61, 62], and consumer reviews [63, 64]. These are developed using supervised learning and then evaluated using labelled custom datasets which further limit generalisability and its application in diverse domains. Despite the custom datasets, some methods can be adapted (or retrained) for a new application, which is the third capability of adaptability. In recent work that is based on language models and annotated datasets, this capability is limited due to the large number of parameters and the opacity of transformer-based learning. They cannot be adapted without a significant volume of work on configuration and fine-tuning which is equivalent to developing an entirely new approach. The fourth capability is explainability of the detected emotion which is becoming more important given our increasing dependence on AI and automation. Explainability has been overlooked in most approaches, mainly due to design limitations that have focused on producing emotion labels of singular granularity. We do not consider accuracy as a core capability as it can be configured (or tweaked) in the design phase as an offset between the availability of annotated datasets for supervised learning and the need for generalisability across multiple domains. A high-quality human-annotated dataset can be leveraged by a supervised learning approach to produce highly accurate emotion classifications. In summary, the granularity of emotion detection output, domain specificity, adaptability and explainability are the formative capabilities of the proposed method for addressing the complexity and ambiguity of emotion expressions.

## Methods

As illustrated in Fig. 1, the emotion AWARE framework consists of three modules, Module 1—Emotion Language Model Finetuning, Module 2—Emotion Lexicon Generation and Module 3—AWARE Core. The components depicted in grey are external sources feeding into the Emotion AWARE framework, where the general instances we have used in this study can be replaced with specialised instances depending on the domain of application (this is demonstrated in Study 5 and 6 for the financial and technology sector).

Module 1 begins with a state-of-the-art language model, such as BERT [65] which has been effectively applied on diverse NLP tasks such as Reading Comprehension [66,



**Fig. 1** The modular composition of the emotion AWARE framework

[67] and Natural Language Inference [68, 69]. State-of-the-art language models are pre-trained on large volumes of unlabelled data to generate deep contextualised word representations by considering syntaxes and semantics [70]. In application, these pre-trained models are finetuned using labelled datasets through transfer learning techniques. For this framework, we selected the DistilBERT [71] base-case model with Huggingface [72] PyTorch implementation for the finetuning. As the finetuning dataset we selected Emotion dataset [73] due to its substantial size, granularity of emotions, and widespread acceptance in the research community. It contained 20,000 tweets based on six emotions joy (33.5%), sadness (29.2%), anger (13.5%), fear (12.1%), love (8.2%), and surprise (3.6%). For the finetuning, we combined train and validation sets, randomized and selected a subset of 5653 points where 1000 samples per each emotion except surprise which was 653 points. Finetuning settings were, a default token length of 128 enabled by both padding and truncation, and batch size of 64 with 8 epochs. At a learning rate of 0.00002 and weight decay of 0.01, the finetuning completed with an F1 score of 0.9394 for the test segment of the dataset. The finetuned language model is utilised by Module 2 for the expansion of a curated list of emotion seed words and in Module 3 for emotion embedding space generation. As noted earlier, DistilBERT can be replaced with any other language model that is closely aligned with the domain of interest.

Module 2 initiates with an emotion seed word list constructed and curated using a combination of automated and manual methods. In developing our emotion lexicon, we referenced Plutchik's model [74] which identifies eight primary emotion classes, each further divided into three subcategories, resulting in a comprehensive 24-class system. Initially, seed keywords for each of these classes were manually curated from an online thesaurus [75]. However, we encountered a scarcity of unique terms for certain emotions, which necessitated the merging of closely related categories joy and ecstasy, amazement and surprise, disgust and loathing, interest and vigilance, anger (rage, anger, annoyance), fear (terror, fear, apprehension). As a result, we consolidated the model into 14 broader emotion classes, each supported by 15–20 thesaurus-derived terms.

While manually curating seed terms yielded high-quality initial seeds, the number of words was insufficient for comprehensive lexicon construction. Therefore, we utilized the vocabulary of the finetuned DistilBERT model itself and extracted embeddings for each of our seed words and compared them with the raw embeddings from the model's

vocabulary terms to find contextually and emotionally similar words. However, due to the ambiguity of individual term embeddings, the relevance of these expanded terms was not highly consistent. To address this, we first clustered seed words into 4 (with  $k=4$  set via the Elbow method [76]) subgroups using the constrained k-means algorithm [77] and then used the average embedding of each subgroup for the expansion. This process extended each subgroup to include highest similar 25 terms from the model’s vocabulary, aiming for a total of 100 terms per each of the 14 emotion classes. Subsequent refinement involved removing duplicates and terms conflicting with Plutchik’s polar opposites to improve the lexicon coherence.

The resulting vocabulary size for each emotion class contained between 80 to 100 terms. To standardize the lexicon, we pruned it by considering the centrality of term embeddings where we compared each term’s embedding to the average category embedding and retained the 80 most pertinent terms per class. The final emotion lexicon comprised 1120 terms across the 14 classes. Table 1 depicts the alignment of the 2, 8 and 14 emotion classification schemes. The 8 classes of emotion contained 80 words per class with total of 640 terms. The version with two classes contained 480 words per category with total of 960 terms. Module 2 also contains externally sourced lexicons for modifiers (inhibitors and intensifiers) and negations, which was based on the valence detection work described in VADER [78]. VADER employs an advanced process that integrates human annotations, heuristic rules, and statistical modelling to determine the valence and polarity of the modifiers. Module 2 provides these two lexicons and the expanded emotion terms as output into Module 3.

Module 3 received the expanded emotion word terms and their corresponding embeddings to generate an emotion embedding space. In case lexicon is constructed from the scratch this step will be skipped as words are already tagged with embeddings during the expansion. For external lexicons each word will go through embedding extractor and tagged with the corresponding embedding. The high dimensional vectors of this emotion embedding space can be visualised using the t-SNE algorithm

**Table 1** Alignment of the 2, 8 and 14 emotion classification schemes

8 Emotions	14 Emotions	2 Emotions
Joy	Serenity	Positive
	Joy/Ecstasy	
Trust	Acceptance	
	Trust	
	Admire	
Anticipation	Interest /Vigilance	
	Anticipation	
Surprise	Amazement/Surprise	Negative
	Distraction	
Sad	Sad	
Disgust	Disgust/loathing	
	Boredom	
Anger	Anger	
Fear	Fear	

on a 2-D grid as shown in Fig. 2. Each point on this Fig. 2 corresponds to an emotion term, with clear separation between green and red, where green is for positive emotions, and red for negative, in 14 emotion categorization.

Next, the sample input text or an entire text corpus is received by Module 3. This input is pushed through the embedding generator and then projected on to the emotional embedding space. The  $n$  nearest neighbour extraction process identifies the closest emotion terms based on this projection. This process is depicted in Module 3 where the nearest neighbours are green dots and the blue are all other emotion embeddings. Based on these nearest neighbours, the Intensity Quantification component calculates the intensities of each of the relevant emotion classes. Here each neighbour will receive a score based on the proximity to the sample input. The terms are sorted and ranked based on similarity, then the terms are grouped based on emotion category and the summed scores for each category are normalised to create the emotion assemble of two, eight and fourteen emotions per input text. See Eq. 1.

$$\theta_e = \frac{\sum_{x \in A} (s_x)}{\frac{n \times (n+1)}{2}} \tag{1}$$

*Equation 1—Calculating Emotional Intensity*

where,  $\theta_e$  intensity of emotion e. n number of nearest neighbours. A subset of nearest neighbours with emotion e.  $S_x$  distance score of the neighbour x

The next phase in Module 3 is the Explainability component. Explainability in AI aims to understand and interpret output made by the model. In the context of emotion AWARE, this is achieved by identifying and extracting the words that have contributed significantly towards forming the emotion profile. Here term embeddings extracted from the input text vector representation are compared with the mean embedding of the entire text. These terms are ranked based on similarity and the top



**Fig. 2** The emotion embedding space generated by module 3 of the emotion AWARE framework



N terms are recorded for explainability and also sent across to the intensity rectification component.

The intensity rectification component consisted of two resolution processors for modifiers (intensifiers and inhibitors) and negations. The adjacent terms of the top N terms are passed through the corresponding lexicons to check for negative, intensive or inhibitive terms. Modifier resolution is completed prior to negations in order to detect intensified or inhibited negations. For detected intensifiers and inhibitors, the score of the top emotion in the profile is revised depending on the intensity of the modifier. Then emotion profile will be normalized so that the increment/decrement of top emotion will affect the other emotions in the profile. In case of negations, the emotion categories are revised based on Plutchik's polar opposites. See Eqs. 2 and 3.

$$\theta_{e^k}^* = \theta_{e^k} + b \times a \quad (2)$$

*Equation 2–Rectifying Emotional Intensity*

$$\theta_e^{normalized} = \frac{\theta_e}{\sum_{i \in E} \theta_i} \quad (3)$$

*Equation 3–Normalizing Emotional Intensity*

Here variables are as follows,  $\theta_{e^k}^*$  - Updated intensity of top keyword's emotion,  $\theta_{e^k}$  - Current intensity of top keyword's emotion,  $b$  - Modifier polarity (intensifier (+1) or inhibitor (-1)),  $a$  - Modifier valence,  $\theta_e^{normalized}$  - Normalized intensity of emotion  $e$ ,  $\theta_e$  - Intensity of emotion  $e$   $E$  - Set of all intensities in the emotion profile. Both modifier and negation lexicons as well as polarity and valences are based on prior work of VADER [78].

**Algorithm 1** EDGstar\_Pathfinding

---

```

input:
  input_text, nearest neighbours list
  current_emotion_profile <- Output from nearest neighbor algorithm
output:
  rectified emotion profile
  emotional keywords.

model_output <- Output after passing text through finetuned model
initialize emotion_keyword_list, top_emotional_keywords
initialize final_emotion_profile = current_emotion_profile
input_terms_list = split input text to words
input_text_mean_emb = extract_embedding (model_output, input_text)

foreach w ∈ input_terms_list do
  w_emb = extract_embedding (model_output, w)
  w_dis = cosine_distance (w_emb, input_text_mean_emb)
  append [w, w_emb, w_dis] to emotion_keyword_list
end

top_emotion_keywords = sort emotion_keyword_list on w_dis and pick top n
best_emotion_keyword = select best candidate for intensity resolution

best_kwd_emo = map_keyword_to_emotion (best_emotion_keyword, nearest_neighbors_list)
best_kwd_neighborhood = extract_kwd_neighborhood (text, best_emotion_keyword)

Bool modifier = check_modifier (best_kwd_neighborhood)
If modifier then rectified_emotion_profile = modify_emotion (current_emotion_profile, best_kwd_emo)
rectified_emotion_profile = normalize (rectified_emotion_profile)
Bool negation = check_negation (best_kwd_neighborhood)
If negation then rectified_emotion_profile = map_opposite_emotion (current_emotion_profile)

return rectified_emotion_profile, top_emotional_keywords

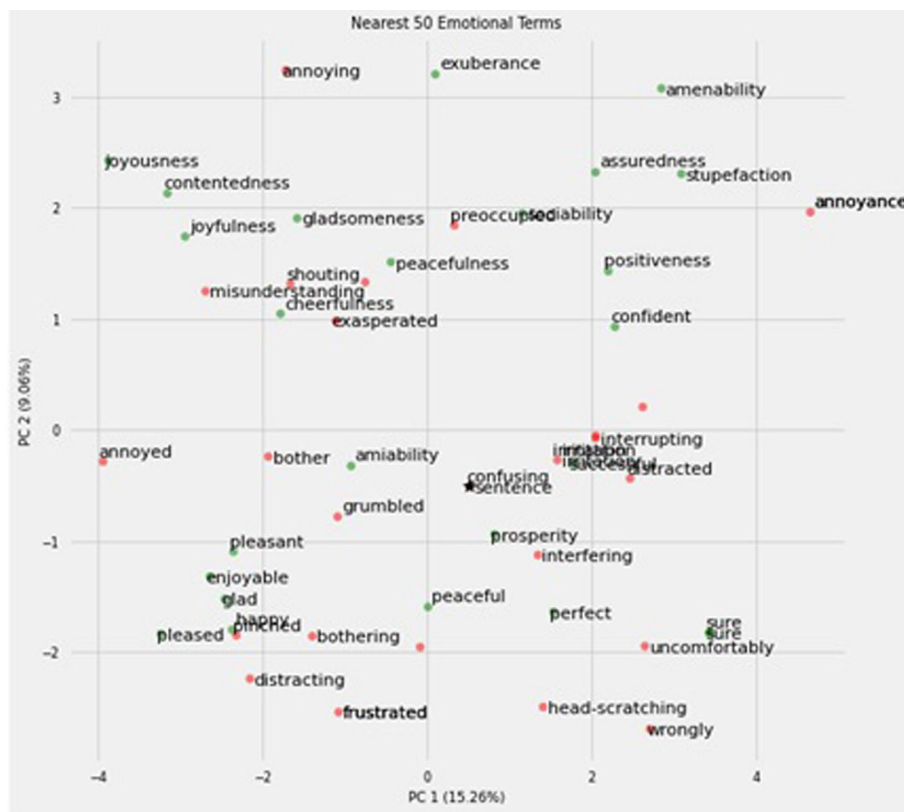
```

Algorithm 1 further describes the explainability component and insensitivity rectification. This algorithm takes nearest neighbours list and current emotion profile as inputs and generates as output, a rectified emotion profile with emotion keywords for explainability.

Figure 3 illustrates an instance of how AWARE constructs an emotion assemble for a given input text, each row of Fig. 3 depicts in the input text and relevant components of the output. The neighbourhood size is 50 and the input text is “*The movie had a great start, but the ending was awful*”. Given the emotional ambiguity of this input, the ‘Emotion Assemble’ presents similar intensity scores for polar emotions, ‘disgust’ and ‘joy’. This is also visible in the neighbour count vector. The explainable emotion terms are ‘awful’ and ‘great’, which provides a rationale for the polarity of the emotion assemble.

## Results

We designed nine studies that demonstrate the capabilities of the framework for the elicitation of multi-granular adaptable, robust, and explainable emotion assembles (Table 2). Each study is composed of a set of experiments where the datasets are drawn from a state-of-the-art collection that represent realistic conversations and content on digital media (Table 3). The results generated from this combination of



**Fig. 3** An emotion assemble generated by the emotion AWARE framework for mixed polarity sample text

nine studies across eight datasets confirms and validates the effectiveness of the proposed framework in the detection and analysis of emotions expressed in digital medium. The same configurations were used for all experiments, such as the fine-tuned language model, modifier and negation lexicons, scoring and explainability modules. Emotion lexicons/embedding spaces were based on the corresponding 2, 8 and 14 classes.

**Study 1: Elicitation of two-emotion assembles (positive and negative) using ISEAR and twitter sentiment datasets**

This study demonstrates the generation of two-emotion assembles of positive and negative emotions, the accuracy of which is then validated with existing methods for the same binary classification. We used two datasets Twitter Sentiment and ISEAR, in which we aggregated sad, anger, fear, disgust as negative and joy as positive. The two-emotion assembles were evaluated with three other methods reported in the literature, they are (1) linear keyword matching using Plutchik’s emotion terms list [], (2) stemmed keyword matching [10] with negation, inhibitor, intensifier detection components and (3) SentiWordNet 3.0 [87]. The evaluation was conducted across four metrics, accuracy, precision, recall and F1-score. Table 4 presents the results, where Emotion AWARE surpasses all three methods.

**Table 2** Nine studies evaluating and demonstrating capabilities of the proposed Emotion AWARE framework

Emotion AWARE capability	Study ID	Description
Elicitation of emotion assembles	Study 1	Objective: Elicitation of two-emotion assembles—positive and negative Datasets: ISEAR[79] and twitter sentiment [80] datasets
	Study 2	Objective: elicitation of four emotion assembles—anger, fear, sadness, joy Datasets: semeval 2007 (“Affective Text”) <sup>55</sup> , ISEAR and Fairy Tales [81]
	Study 3	Objective: elicitation of four emotion assembles—disgust, surprise, trust, anticipation Datasets: goemotions [82] and SemEval-2018[83]
	Study 4	Objective: elicitation of 2, 8 and 14 emotion assembles in increasing granularity Datasets: demonstrated on sample sentences of GoEmotions
Adaptability	Study 5	Objective: emotion AWARE adapted for the finance sector Dataset: PhraseBank [84]
	Study 6	Objective: emotion AWARE adapted for the technology sector Dataset: Senti4SD [85]
	Study 7	Objective: robustness of emotion AWARE across intensifiers and inhibitors Dataset: manually curated dataset based on fairy tales [81]
Explainability	Study 8	Objective: robustness of emotion AWARE in negation detection Dataset: manually curated dataset based on Fairy Tales
	Study 9	Objective: explaining emotion assembles using constituent intensity scores and terms of emotional significance Dataset: demonstrated on fairy tales dataset

### Study 2: Elicitation of four emotion assembles (anger, fear, sadness, joy) using SemEval 2007 (“Affective Text”), ISEAR and fairy tales

As noted prior, the proposed framework is capable of detecting all emotions in Plutchik’s wheel of emotions [88]. However, only a handful of related work have proposed techniques to detect all eight emotions. Therefore, we split the eight emotions into two subsets (common and rare) in order to ensure that Emotion AWARE can be evaluated with state-of-the-art approaches in extant literature. Study 2 evaluates the common subset anger, fear, sadness, joy, while study 3 evaluates the rare subset, disgust, surprise, anticipation, and trust. In study 2, we compared AWARE with rule-based, hybrid as well as machine learning techniques. Rule-based includes emotional linear keyword matching, stemmed keyword matching as well as the more advanced rule-based methods that consider contextuality and affinity-based methods CLSA, CPLSA, DIM. Here, CLSA and CPLSA are categorical classifications based on LSA and PLSA. Additionally, we also compared with context-based emotion vector construction methods [89], namely context-based Wiki, context-based Guten, context-based W-G. For machine learning methods, we finetuned DistilBERT [71] model on Emotion [90] dataset. Collectively, study 2 compares Emotion AWARE with ten similar techniques proposed in recent literature, using SemEval 2007, ISEAR and Fairy Tales datasets. For this, we incorporated the experiments included in previous work [16, 89]. As presented in Table 5, AWARE outperforms all methods for most combinations of dataset and emotions.

**Table 3** Description of datasets used in the experiments, with percentage distribution of each emotion

Dataset name	Description
ISEAR [79]	7665 sentences by 3000 respondents in 37 countries on all 5 continents describing situations where they experienced seven major emotions classified emotions are joy (14.27%), sadness (14.3%), fear (14.28%), anger (14.3%), guilt (14.26%), disgust (14.3%), and shame (14.3%) emotions
Twitter sentiment [80]	3534 tweets extracted from figure eight's data for everyone platform classified emotions are positive (31.21%),negative (28.32%) and neutral (40.46%). Test set is used
SemEval 2007 ("affective text") [86]	1250 sentences from tweets, news headlines, Google News, and other major newspapers classified emotions are Ekman's six basic emotions, fear (15.52%), anger (7.28%), joy (35.28%), sadness (21.2%), disgust (3.36%), and surprise (17.36%)
Fairy tales [81]	All sentences with affective high agreement from 176 fairy tales(1207 datapoints) by B. Potter, H.C. Andersen, and Grimm's classified emotions are angry-disgusted (18.06%), fearful (13.75%), happy (36.87%), sad (21.87%), and surprised (9.44%)
GoEmotions [82]	58,000 comments extracted from popular English-language subreddits on the reddit social media platform classified into admiration (9.44%), amusement (5.34%), anger (3.61%), annoyance (5.7%), approval (6.79%), caring (2.53%), confusion (3.08%), curiosity (5.02%), desire (1.48%), disappointment (2.92%), disapproval (4.76%), disgust (1.87%), embarrassment (0.69%), excitement (1.94%), fear (1.41%), gratitude (6.21%), grief (0.18%), joy (3.29%), love (4.75%), nervousness (0.38%), optimism (3.64%), pride (0.26%), realization (2.55%), relief (0.34%), remorse (1.23%), sadness (2.99%), surprise (2.45%), neutral (32.75%), some comments contained more than 1 emotion
SemEval-2018 [83]	10,983 manually curated English tweets from Affect in Tweets which included five subtasks on inferring the affectual state of a person from their tweet classified into 11 different emotions anger (36.1%), anticipation (13.9%), disgust (36.6%), fear (16.8%), joy (39.3%), love (12.3%), optimism (31.3%), pessimism (11.6%), sadness (29.4%), surprise (5.2%), trust (5.0%). Some comments contained more than 1 emotion
Financial phrasebank [84]	4840 sentences from financial news on the LexisNexis database, classified emotions are positive (28.2%), negative (12.5%), and neutral (59.4%)
Senti4SD dataset [85]	4424 posts extracted from stack overflow ( <a href="http://www.stackoverflow.com">www.stackoverflow.com</a> ), classified emotions are positive (34.61%), negative (27.15%), and neutral (38.23%)

### Study 3: Elicitation of four emotion assembles (disgust, surprise, trust, anticipation) using GoEmotions and SemEval-2018

For the rare emotions of disgust, trust, anticipation, and surprise, we used GoEmotions and SemEval-2018 datasets and compared with stemmed keyword matching and DistilBERT model finetuned with the Emotions dataset. Table 6 presents the results where AWARE outperforms all other methods across the four emotions.

### Study 4: Elicitation of 2, 8 and 14 emotion assembles in increasing granularity

This study demonstrates Emotion AWARE's ability to generate emotion assembles at diverse levels of granularity. Table 7 presents these granular emotion assembles for the same text. Only the emotions with non-zero scores are shown in this table. For instance, row 2 depicts a positive score in the two-emotion assemble, anticipation and trust as the detected emotions in the eight emotions assemble, and in the 14 emotions assemble, trust is further split into trust, acceptance, and admiration alongside the corresponding intensity scores.

**Table 4** Comparison of results with 95% CI for two-emotion assembles using ISEAR and Twitter

Dataset	Approach	Positive			Negative			Accuracy
		Prec.	Rec.	F1	Prec.	Rec.	F1	
ISEAR aggregated	Linear keyword matching	0.402 ± 0.029	0.938 ± 0.143	0.562 ± 0.029	0.350 ± 0.014	0.383 ± 0.015	0.365 ± 0.014	0.391 ± 0.019
	Stemmed keyword matching	0.422 ± 0.03	0.954 ± 0.012	0.585 ± 0.029	0.339 ± 0.014	0.427 ± 0.015	0.378 ± 0.014	0.405 ± 0.13
	SentiWordNet	0.657 ± 0.028	0.918 ± 0.016	0.766 ± 0.025	0.739 ± 0.012	0.36 ± 0.014	0.485 ± 0.016	0.673 ± 0.124
	Emotion AWARE	0.836 ± 0.022	0.956 ± 0.012	0.892 ± 0.018	0.845 ± 0.011	0.562 ± 0.015	0.675 ± 0.014	0.837 ± 0.009
Twitter sentiment	Linear keyword matching	0.335 ± 0.029	0.854 ± 0.021	0.482 ± 0.03	0.407 ± 0.03	0.873 ± 0.021	0.555 ± 0.031	0.373 ± 0.021
	Stemmed keyword matching	0.326 ± 0.027	0.853 ± 0.021	0.471 ± 0.029	0.451 ± 0.031	0.851 ± 0.022	0.589 ± 0.03	0.391 ± 0.021
	SentiWordNet	0.568 ± 0.029	0.774 ± 0.025	0.605 ± 0.03	0.639 ± 0.03	0.729 ± 0.028	0.681 ± 0.029	0.605 ± 0.022
	Emotion AWARE	0.715 ± 0.026	0.9 ± 0.018	0.797 ± 0.024	0.928 ± 0.016	0.782 ± 0.026	0.849 ± 0.022	0.827 ± 0.016

**Study 5: Emotion AWARE adapted for the finance sector using the PhraseBank dataset**

Domain adaptability is a core capability of Emotion AWARE. In study 5 and 6, we demonstrate this capability for the financial and technology sector. For the financial sector, we used the PhraseBank dataset which contains financial statements classified for positive and negative emotions. Emotion AWARE was adapted to this domain by simply expanding the vocabulary with 20 words each for positive and negative classes using the L&M financial emotion lexicon [91]. Following the domain adaptation, two-emotion assembles were generated and compared with the stemmed keyword matching technique, finetuned DistilBERT with the Emotion dataset and SentiWordNet. Emotion AWARE is used with both the default vocabulary and the extended vocabulary using L&M. Table 8 summarizes the results, notably AWARE surpasses all methods across all metrics.

**Study 6: Emotion AWARE adapted for the technology sector using Senti4SD8 dataset**

Study 6 is the domain adaptation for the technology sector, where we used Senti4SD dataset which contains conversations from the stackoverflow community classified by emotion. Similar to study 5, we evaluated the proposed approach with default vocabulary as well as extended vocabulary along with stemmed keyword matching, SentiWordNet, and finetuned DistilBERT. Here both positive and negative classes were extended with 20 words extracted using Emotion AWARE running on the training set. As shown in Table 9, Emotion AWARE outperforms all other methods in this adaptability task.

**Table 5** Comparison of F1 score with 95% CI for four emotion assemblies (anger, fear, sadness, joy)

Method	Emotion	dataset	Linear Keyword matching	Stemmed keyword matching	CLSA	CPLSA	CNMF	DIM	Context-based Wiki	Context-based Guten	Context-based W-G	DistilBERT (emotions)	Emotion AWARE
Anger	SemEval		0.07 ± 0.052	0.119 ± 0.067	0.112 ± 0.065	0.244 ± 0.088	0.278 ± 0.092	0.175 ± 0.078	-	-	-	0.25 ± 0.089	0.371 ± 0.099
	ISEAR		0.274 ± 0.025	0.192 ± 0.022	0.631 ± 0.028	0.456 ± 0.031	0.579 ± 0.029	0.286 ± 0.026	0.628 ± 0.028	0.582 ± 0.029	0.668 ± 0.028	0.494 ± 0.03	0.571 ± 0.029
	Fairy Tales		0.2 ± 0.052	0.193 ± 0.052	0.51 ± 0.065	0.313 ± 0.061	0.65 ± 0.062	0.392 ± 0.064	-	-	-	0.575 ± 0.065	0.708 ± 0.059
Fear	SemEval		0.252 ± 0.061	0.291 ± 0.064	0.511 ± 0.07	0 ± 0	0.618 ± 0.068	0.404 ± 0.069	-	-	-	0.292 ± 0.064	0.488 ± 0.07
	ISEAR		0.393 ± 0.028	0.451 ± 0.029	0.071 ± 0.015	0 ± 0	0.056 ± 0.014	0.351 ± 0.027	0.592 ± 0.028	0.536 ± 0.031	0.574 ± 0.029	0.673 ± 0.027	0.602 ± 0.029
	Fairy Tales		0.385 ± 0.075	0.363 ± 0.073	0.64 ± 0.074	0 ± 0	0.741 ± 0.067	0.255 ± 0.066	-	-	-	0.508 ± 0.076	0.696 ± 0.071
Joy	SemEval		0.203 ± 0.038	0.258 ± 0.041	0.402 ± 0.046	0.254 ± 0.041	0.648 ± 0.045	0.71 ± 0.042	-	-	-	0.698 ± 0.043	0.772 ± 0.039
	ISEAR		0.365 ± 0.028	0.379 ± 0.029	0.103 ± 0.018	0.34 ± 0.027	0.01 ± 0.006	0.515 ± 0.031	0.564 ± 0.03	0.574 ± 0.029	0.542 ± 0.029	0.673 ± 0.027	0.738 ± 0.025
	Fairy Tales		0.632 ± 0.045	0.621 ± 0.045	0.727 ± 0.041	0.436 ± 0.047	0.781 ± 0.038	0.789 ± 0.038	-	-	-	0.772 ± 0.039	0.741 ± 0.04
Sadness	SemEval		0.226 ± 0.05	0.28 ± 0.054	0.337 ± 0.057	0.378 ± 0.058	0.475 ± 0.06	0.253 ± 0.052	-	-	-	0.281 ± 0.054	0.554 ± 0.06
	ISEAR		0.383 ± 0.029	0.36 ± 0.028	0.106 ± 0.018	0.282 ± 0.027	0.017 ± 0.008	0.337 ± 0.028	0.408 ± 0.029	0.253 ± 0.026	0.296 ± 0.027	0.557 ± 0.029	0.611 ± 0.029
	Fairy Tales		0.462 ± 0.06	0.449 ± 0.06	0.642 ± 0.058	0.37 ± 0.058	0.76 ± 0.052	0.24 ± 0.052	-	-	-	0.655 ± 0.057	0.741 ± 0.053
Weighted Average of F1 Score	SemEval		0.206 ± 0.025	0.258 ± 0.027	0.379 ± 0.03	0.236 ± 0.026	0.562 ± 0.031	0.479 ± 0.032	-	-	-	0.466 ± 0.031	0.621 ± 0.03
	ISEAR		0.353 ± 0.014	0.345 ± 0.014	0.227 ± 0.011	0.269 ± 0.013	0.165 ± 0.010	0.372 ± 0.015	0.548 ± 0.016	0.486 ± 0.017	0.52 ± 0.017	0.599 ± 0.015	0.63 ± 0.014
	Fairy Tales		0.467 ± 0.03	0.454 ± 0.03	0.65 ± 0.029	0.329 ± 0.028	0.745 ± 0.026	0.496 ± 0.03	-	-	-	0.664 ± 0.028	0.728 ± 0.027

**Table 6** Comparison of F1 scores with 95% CI for four emotion assembles (disgust, surprise, trust, anticipation)

Dataset	Method	disgust	surprise	trust/approval	anticipation/desire
GoEmotions	Stemmed keyword matching	0.377 ± 0.029	0.234 ± 0.022	0.076 ± 0.008	0.048 ± 0.014
	DistilBERT (Emotions)	0.45 ± 0.03	0.5 ± 0.026	0.36 ± 0.015	0.49 ± 0.033
	Emotion AWARE	0.648 ± 0.028	0.655 ± 0.025	0.707 ± 0.014	0.491 ± 0.033
SemEval 18	Stemmed keyword matching	0.25 ± 0.013	0.084 ± 0.023	0.147 ± 0.03	0.068 ± 0.013
	DistilBERT (Emotions)	0.515 ± 0.015	0.185 ± 0.032	0.175 ± 0.032	0.215 ± 0.021
	Emotion AWARE	0.826 ± 0.012	0.515 ± 0.041	0.439 ± 0.042	0.289 ± 0.023

**Table 7** Demonstrating the Elicitation of 2, 8 and 14 emotion assembles in increasing granularity

ID	Sentence	Ground truth label	Two Emotion assemble	8 Emotion assemble (main emotions shown)	14 Emotion profile (Main emotions shown)
1	"That's a maryland fan comment. You should be ashamed of yourself."	Embarrassment(negative)	Negative: 0.937	disgust: 0.558 anger: 0.179 sadness: 0.169	disgust_loathing': 0.299 'boredom': 0.259 anger: 0.179 sadness: 0.169
2	"Our father will protect us < 3"	Caring(positive)	Positive: 0.977	trust: 0.529,  'anticipation': 0.268	Trust': 0.371 'acceptance': 0.115 'admire': 0.043 'interest_vigilance': 0.207 'anticipation': 0.061
3	"But I'm sort of confused as to how the pic was taken, sorry if you said it but I'm a bit stupid and tired"	Confusion(negative)	Negative: 0.797	surprise: 0.347,  'disgust': 0.211	'distraction': 0.292 amazement_surprise: 0.055, 'boredom': 0.151 disgust_loathing': 0.06
4	"Best bedtime christmas story ever."	Admiration(positive)	Positive: 0.998	'sadness': 0.186 'trust': 0.376  'joy': 0.337 Surprise': 0.285	'sadness': 0.186 'admire': 0.308 'acceptance': 0.068 'joy_ecstasy': 0.337 'amazement_surprise': 0.285

**Study 7: Robustness of Emotion AWARE across intensifiers and inhibitors**

Intensifiers and inhibitors are subjectively used in emotion expressions, which means an emotion detection method must be robust to intensifiers and inhibitors, specifically in digitalised emotion expressions where physical cues unavailable. To demonstrate this robustness property of Emotion AWARE, we created a new dataset because state-of-the-art datasets used in related work are limited in their inclusion of varying intensifiers and inhibitors. For constructing this manually curated dataset, we selected a random subset of 80 sentences from the Fairy Tales dataset and introduced intensifiers and inhibitors to each sentence to generate additional 160 sentences.



**Table 8** Comparison of results with 95% CI adapted for the finance sector using the PhraseBank dataset

Approach	Positive			Negative			Weighted average of F1
	Prec.	Rec.	F1	Prec.	Rec.	F1	
Stemmed keyword matching	0.405 ± 0.026	0.649 ± 0.025	0.499 ± 0.027	0.228 ± 0.033	0.523 ± 0.04	0.318 ± 0.037	0.443 ± 0.022
SentiWordNet	0.307 ± 0.024	0.511 ± 0.027	0.383 ± 0.026	0.833 ± 0.03	0.698 ± 0.037	0.76 ± 0.034	0.498 ± 0.022
DistilBERT (emotions)	0.831 ± 0.02	0.937 ± 0.013	0.871 ± 0.018	0.813 ± 0.031	0.59 ± 0.039	0.69 ± 0.037	0.815 ± 0.017
Emotion AWARE (default vocab)	0.922 ± 0.014	0.853 ± 0.019	0.886 ± 0.017	0.699 ± 0.037	0.829 ± 0.03	0.759 ± 0.034	0.847 ± 0.016
Emotion AWARE (LnM + default vocab)	0.903 ± 0.016	0.879 ± 0.017	0.891 ± 0.017	0.765 ± 0.034	0.805 ± 0.032	0.785 ± 0.033	0.858 ± 0.015

**Table 9** Comparison of results with 95% CI when adapted for the technology sector using Senti4SD8 dataset

Approach	Positive			Negative			Weighted Average of F1
	Prec.	Rec.	F1	Prec.	Rec.	F1	
Stemmed keyword matching	0.973 ± 0.008	0.672 ± 0.024	0.795 ± 0.02	0.397 ± 0.028	0.922 ± 0.015	0.555 ± 0.028	0.69 ± 0.017
SentiWordNet	0.722 ± 0.022	0.772 ± 0.021	0.746 ± 0.022	0.64 ± 0.027	0.848 ± 0.02	0.73 ± 0.025	0.739 ± 0.016
DistilBERT (emotions)	0.896 ± 0.015	0.945 ± 0.011	0.92 ± 0.014	0.925 ± 0.015	0.861 ± 0.02	0.892 ± 0.018	0.907 ± 0.011
Emotion AWARE (default vocab)	0.954 ± 0.01	0.868 ± 0.017	0.909 ± 0.014	0.816 ± 0.022	0.933 ± 0.014	0.871 ± 0.019	0.892 ± 0.012
Emotion AWARE (extended vocab)	0.945 ± 0.011	0.911 ± 0.014	0.928 ± 0.013	0.883 ± 0.018	0.927 ± 0.015	0.904 ± 0.017	0.917 ± 0.01

Table 10 demonstrates the evaluation of a single sentence using known intensifiers and inhibitors and their corresponding impact on the emotion score and emotion category. Here the valence and intensity of modifiers is derived from prior work of VADER [78]. In case of incrementing or decrementing modifier, current top emotion's score will be increased or decreased with a factor of corresponding modifier intensity as explained in the Eq. 2. Then the emotion profile will be normalized according to the Eq. 3. For this experiment we used a sample sentence from SemEval-2018 dataset. As depicted in Table 9, the base sentence “work was good for the first half” is classified as joy\_ecstasy with an intensity score of 0.339 and admire with a score of 0.229. In the subsequent rows, we added intensifiers and inhibitors with varied valence that modifies the emotion expressed in the sentence. In descending order of Table 10, the intensity score of the top emotion of the base sentence (joy\_ecstasy) decreases. This illustrates that AWARE has correctly identified all modifiers and attributed emotion labels and varied intensity scores accordingly.

**Table 10** Demonstrating the variation of emotion intensity score based on intensifiers and inhibitors

Sentence (modifier is bolted)	Modifier valence (increment / decrement)	Modifier intensity	Emotion profiles (only top emotions are showed)	Original sentence top emotion's score (joy_ecstasy)
Work was good for the first half	–	–	'Admire': 0.229, 'joy_ecstasy': 0.339	0.339
Work was incredibly good for the first half	Increment (+)	0.9	'Amazement_surprise': 0.197, 'joy_ecstasy': 0.444	0.444
Work was very good for the first half	Increment (+)	0.8	'Admire': 0.216, 'joy_ecstasy': 0.426	0.426
Work was quite good for the first half	Increment (+)	0.6	'Admire': 0.194, 'joy_ecstasy': 0.425	0.425
Work was considerably good for the first half	Increment (+)	0.6	'Admire': 0.232, 'joy_ecstasy': 0.409	0.409
Work was almost good for the first half	Decrement (–)	0.2	'Admire': 0.202, 'joy_ecstasy': 0.316	0.316
Work was kinda good for the first half	Decrement (–)	0.3	'Joy_ecstasy': 0.244, 'admire': 0.255	0.244
Work was sort of good for the first half	Decrement (–)	0.4	'joy_ecstasy': 0.206, 'acceptance': 0.214, 'admire': 0.262	0.206

**Table 11** Demonstrating robustness across intensifiers and inhibitors emotion of a sentence

Version	Original	Intensified	Inhibited
Sentence	She is happy	She is extremely happy	She is partly happy
Most significant emotion	Joy	Joy	Joy
Most significant emotion score	0.54(x)	0.68(> x)	0.41(< x)

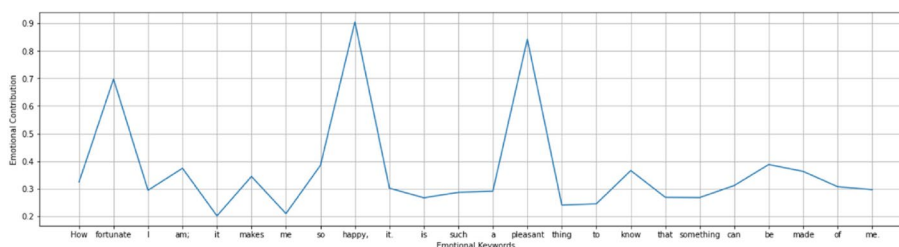
The manually curated dataset was used to evaluate Emotion AWARE, SentiWordNet, and stemmed keyword matching. Even though these approaches construct multi-facet emotion profiles, for this experiment we have only considered the most significant emotion as it is the most impacted from such modifications. For instance, if the most significant emotion in the original sentence is joy and has score of  $x$ , it is expected that intensified sentence score of joy be  $> x$  where inhibited sentence score of joy be  $< x$  (Table 11). Thus, we considered the most significant emotion score of the original sentence in inhibited and intensified cases to determine if this approach has correctly identified the modifiers. As the dataset consisted of 80 sentences, we calculated the mean of the most significant emotion score as the evaluation metric. Here DistilBERT (Emotion) is not included as it provides only labels (Table 12).

**Table 12** Performance of inhibitor and intensifier detection

Approach	Mean of most significant emotion's score		
	Original sentences	Intensified sentences	Inhibited sentences
Emotion AWARE	0.561	0.907	0.400
Stemmed keyword matching	0.188	0.237	0.103
SentiWordNet	0.122	0.107	0.107

**Table 13** Results for robustness of emotion AWARE in negation detection

Approach	F1 Score	
	Original sentences	Negated sentences
Improved keyword matching	0.435	0.462
SentiWordNet	0.799	0.297
DistilBERT (Emotions)	0.805	0.056
Emotion AWARE	0.818	0.841



**Fig. 4** Demonstrating explainability of emotion assembles using constituent intensity

As seen in the mean emotion scores, Emotion AWARE has increased from 0.346 intensified case and decreased from 0.161 in inhibited case. This shows that AWARE has correctly modified the emotions compared to corresponding original sentences. Stemmed keyword matching has incorporated the modifiers to some extent but it’s bottlenecked with limitations of modifier capturing. When considering SentiWordNet, none of the modifiers were detected, where it has mitigated the scores even in intensified sentences.

**Study 8: Robustness of emotion AWARE in negation detection**

Similar to Study 7, we randomly selected 80 sentences from the Fairy Tales dataset and manually negated to create a new dataset of 80 negated sentences. Here we used negation terms such as ‘no’, ‘not’, ‘never’ to reverse the emotions. We used this dataset to evaluate robustness of Emotion AWARE with that of stemmed keyword matching, SentiWordNet and DistilBERT finetuned on Emotion dataset. Table 13 presents mean F1 scores of emotion detection for original and negated sentences in this dataset. It is interesting to note that although SentiWordNet and DistilBERT show comparable accuracies to AWARE for the original sentences, they perform poorly for the negated sentences, unlike Emotion AWARE which scores 0.841 F1 score. We hypothesize that this observed behaviour is likely a result of the model’s tendency to prioritize emotion-specific terms while disregarding the presence of negating words within the sentences. The datasets used in Study 7 and 8 will be made publicly available as a secondary outcome of this work. This dataset consisted of 320 sentences 80 per original, negated, intensified, and inhibited and optimal for modifier evaluation.

**Study 9: Explainability of emotion assembles using constituent intensity scores and terms of emotional significance**

Study 9 evaluates explainability of the emotion assembles generated by the Emotion AWARE framework, using both intensity scores and terms that contribute to the detection of an emotion. Figure 4 illustrates this capability for a sample sentence randomly selected from the Fairy Tales dataset, *“How fortunate I am; it makes me so happy, it is such a pleasant thing to know that something can be made of me”*. The framework generates intensity peaks for the terms *“fortunate”, “happy”* and *“pleasant”*, which distinguishes the contributing terms and their significance in the emotion assemble. These intensities are based on the *w\_dist* in the Algorithm 1. scores as explained as *w\_dist* in the algorithm 1.

Table 14 presents a further demonstration of explainability with emotion keyword extraction. Here the positive, negative samples are randomly selected from the Fairy Tales dataset. We combined some samples to create a mixed sample. The colour scheme depicts emotion significance, where shades of green are for positive emotions and shades of red are for negative emotions. The intensity scores are depicted on the right side of the image, which further improves the explainability of the emotion assemble.

The following table (Table 15) summarizes the emotion keyword results for the entire fairy tales dataset. Here for each sample in the dataset, top emotion and top keyword is extracted. The table contain each of the emotion category fear, anger, joy, surprise and sadness along with the 10 most frequent keywords per category. These keywords reflect the corresponding emotions which further validates AWARE.

**Table 14** Contributing terms and corresponding intensity scores for emotion explainability

Example	Intensity scores		
	Term	Score	
<b>Positive sample</b>			
The son said: “Most gracious father, I will show her to you in the form of a beautiful flower,” and he thrust his hand into his pocket and brought forth the pink, and placed it on the royal table, and it was so beautiful that the king had never seen one to equal it	Beautiful	0.74	
	Beautiful	0.73	
	Gracious	0.73	
	Royal	0.66	
	Equal	0.55	
<b>Negative sample</b>	Term	Score	
The king ordered the man to be brought before him, and threatened with angry words that unless he could before the morrow point out the thief, he himself should be looked upon as guilty and executed	Angry	0.78	
	Executed	0.75	
	Threatened	0.74	
	Thief	0.71	
<b>Mixed sample</b>	Guilty	0.68	
	Term	Score	
	“It was saying, ‘You are so beautiful, I like you very much. ‘Tweet, tweet,’ sang the bird, as he flew out into the green woods, and Tiny felt very sad. The little prince was at first quite frightened at the bird. It was like a giant, compared to such a delicate little creature as himself. But when he saw Tiny, he was delighted, and thought her the prettiest little maiden he had ever seen	Delighted	0.90
	Frightened	0.88	
	Delicate	0.85	
Beautiful	0.81		
Sad	0.76		
Prettiest	0.71		

**Table 15** 10 most frequent keywords per emotion category in fairy tales dataset

Emotion	Keywords with highest significance per category
Fear	'frightened', 'terrified', 'afraid', 'uneasy', 'alarmed', 'fear', 'cried', 'trembling', 'anxious', 'trembled'
Anger	'angry', 'stupid', 'angrily', 'cried', 'growled', 'annoyed', 'nasty', 'rage', 'fury', 'refused'
Joy	'happy', 'pleased', 'joy', 'glad', 'merry', 'delighted', 'beautiful', 'rejoiced', 'happily', 'good'
Surprise	'exclaimed', 'astonished', 'surprised', 'heavens', 'startled', 'shocked', 'bewildered', 'yelping', 'sudden', 'interrupting'
Sadness	'wept', 'sad', 'grieved', 'cry', 'sorrowfully', 'poor', 'unhappy', 'mournfully', 'troubled', 'tears'

## Discussion

The study of emotion has a vibrant history, beginning with the evolutionary context where Charles Darwin [92] posited that emotions are an expressive behaviour that has evolved to increase our chances of survival, right up to Barrett [93] constructivist view where an emotion is constructed by cognitively classifying an affect based on past knowledge of that emotion. A multitude of studies have been conducted on the types of emotions, using methods such as philosophical postulations, factor analytic studies, similarity scaling studies, child development studies, cross cultural studies and facial expression studies. Based on studies of facial expression, Ekman [94, 95] proposed six basic emotions; anger, disgust, fear, happiness, sadness and surprise. This was followed by Plutchik's [74] eight primary emotions interlinked by polarity; joy and sadness, trust and disgust, surprise and anticipation, anger and fear. Plutchik also proposed the wheel of emotions, a three-dimensional circumplex that illustrates degrees of similarity/polarity between emotions [74]. The wheel is split into eight sectors for eight primary emotions, layers within each sector signify varying intensities (for instance with joy, intense joy being ecstasy and less intense being serenity) and gaps between sectors represent the mix of two primary emotions. The more recent digitalisation of emotion expressions has led to new challenges in complexity and ambiguity due to the absence of physical cues and observer inference Table 15.

Emotion AWARE addresses this complexity and ambiguity of emotion detection through its four capabilities of multi-granular emotion assembles, adaptability, robustness and explainability. Unlike related work in emotion detection, the proposed framework generates emotion assembles based on prior knowledge of heuristics and learned knowledge of the finetuned language models. Drawing upon the literature review, we conducted a capability comparison of Emotion AWARE against the most effective and relevant studies as tabulated in Table 16. Following this capability comparison, we developed empirical evidence through the experimental evaluation of Emotion AWARE across nine studies that are based on state-of-the-art datasets containing diverse human emotion expressions. Studies 1–4 evaluate the detection of a spectrum of emotion assembles, starting with binary (or sentiment), the four common emotions from Plutchik's wheel of emotion (anger, fear, sadness, joy), the four rare emotions (disgust, surprise, trust, anticipation), and the increasing granularity of emotions from 2, 4 to 14 categories. 2, Emotion AWARE outperforms a finetuned DistilBERT, highlighting the importance of prior knowledge contained in lexicons. Adaptability of the framework is demonstrated in Study 5 and 6 where AWARE was adapted for the finance and

**Table 16** Comparison of Emotion AWARE with related work in emotion detection

Author	Approach	Datasets	Output	Domain	Explainability	Adaptability	Modifier resolution and negation
Mukherjee et al. [98]	Machine learning algorithms with TD-IDF	Amazon reviews	Sentiment labels (positive, negative)	Cell phone reviews	No	No	Negation prediction
Adikari et al [10]	Lexicon based	Online mental health forums	emotion profiles (8 classes)	Healthcare	Limited to emotion keywords in the lexicon	No	Resolution limited to variations in the lexicon
Sharma and Sharma [99]	Deep learning and SVM	Twitter	Top sentiment label (positive, negative and neutral)	Airlines	No	No	No
Adoma et al. [100]	Pretrained language Models (BERT, RoBERTa, DistilBERT, and XLNet)	ISEAR	Top emotion label (7 classes)	General	No	No	No
Shrivastava et al [101]	GNN (convolutional neural networks) and Word2Vec	TV shows transcripts	Top emotion label (7 classes)	TV shows	No	No	No
Tzacheva et al [102]	Hybrid (lexicon with SVM)	Tweets	Top emotion label (8 classes)	General	No	Challenging as large lexicon extensions are required	No
Huang et al [103]	Pretrained language Models (BERT)	Friends, EmotionPush	Top emotion label (8 classes)	General (emotions in conversations)	No	No	No
Rabeya et al. [104]	Lexicon based	Social media and news	Sentiment (happiness and sadness)	General	Emotion keywords (limited lexicon)	No	No
Emotion AWARE framework	lexicon and finetuned language models	ISEAR, Twitter SemEval, Senti4SD, Goemotions, Fairy Tales, Affective Text, PhraseBank	Multifaceted emotion profiles with intensities granularity levels of 2,8,14 classes	General and specific (adaptable to domain specific emotion extraction)	Emotion keyword extraction (capture wide range of related words with embeddings)	Can adapt lexicons with small extensions of the lexicon. lexicons, embeddings are customizable	Negation, intensifier and inhibitor resolution followed by intensity rectification

technology domains. In Study 5, AWARE demonstrates a 6% improvement in F1-score with an extended vocabulary compared to finetuned DistilBERT. Most related work in recent literature forego domain adaptability, where the challenges include frequency and scarcity as well as changing emotion polarity across domains. For example, “unpredictable” is frequently used as a positive emotion expression in film reviews (e.g., “The plot of this movie is fun and unpredictable”), whereas it is a negative expression in financial markets or human resource management (e.g., “the impact on share market indices is unpredictable” or “the employee response to governance is unpredictable”) [96]. Language model-based approaches have limited adaptability across domains due to the scale of training data required for finetuning while lexicon-based approaches require large hand-crafted, domain-specific lexicons [97]. Emotion AWARE is able to overcome both limitations by leveraging a short list of domain specific terms with the usage of embeddings, which introduces context through meaning and emotion instead of exact matching. Robustness of the framework is demonstrated in Studies 7 and 8 where implied emotions and the presence of intensifiers, inhibitors and negations are detected and assigned intensity values relative to other emotions expressed in the same text.

Also, in Study 8 which demonstrates robustness of negation detection, DistilBERT and SentiWordNet perform poorly in comparison to Emotion AWARE due to its exclusive focus on learned knowledge of emotion expressions. For instance, DistilBERT can accurately identify emotions of sentences “*I am truly glad to hear it!*”(joy) and “*I am truly sad to hear it!*”(sad) but incorrectly detect the emotion as joy in the negated version “*I am truly not glad to hear it!*”. This highlights the significance of incorporating a heuristic approach to manage negations in Emotion AWARE, enhancing the accuracy of emotion detection. Finally, study 9 demonstrates the explainability capability where contributing terms and corresponding intensity scores of emotion assembles effectively unpack and rationalise the detected emotions.

The practical implications of this framework are broad. The robust, domain adaptable and explainable detection of emotion expressions has wide application value as we increasingly express emotions using digital media. For instance, in a long-term healthcare setting of multiple stakeholders (such as cancer care involving a clinician, patient, and social worker), this framework can be adapted to suit the vocabulary of each stakeholder and the generated emotion assembles can be explained using the constituent terms, which yields further capabilities of converging or diverging the emotion profiles of all stakeholders for decision value and consensus building among human behaviours in such complex settings.

## Conclusion

The exponential transition of knowledge-focussed human activities and communication into digital spaces and physical hybrids has necessitated the manifestation, communication and persistence of our expressions of emotions on digital media. The proposed Emotion AWARE framework enables the objective and unambiguous detection of such emotions, with adaptability, robustness and explainability, for the subsequent generation and management of information that represents patterns of behaviour of individuals and organizations. The results from eight experimental studies confirm its practical value and contribution towards the comprehension of such expressions and behaviour

of individuals and organizations. As future work, we intend to address the limitations of Emotion AWARE in complex settings where emotion is implied using either highly technical, jargonistic or informal emoji-based expressions, and figurative expressions of emotion such as the detection of metaphors and similes. We will also work on the integration of detected emotions along with other dimensions and modalities of information into the decision-making activities of individuals and organizations.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40537-024-00953-2>.

Supplementary File 1.

### Acknowledgements

Not applicable

### Author contributions

All authors contributed to the ideation and design of the proposed framework. G.G., D.S and N.M developed the AI framework. G.G, D.A and M.M conducted experiments and evaluations across the eight studies. G.G., D.S and N.M wrote the initial draft of the manuscript, followed by first round of reviews and revisions by D.A and MM. All authors reviewed and finalised the manuscript.

### Funding

Not applicable

### Availability of data and materials

The full code repository and related content of the Emotion AWARE framework will be publicly available on GitHub, following the peer review process. The datasets that support the findings of this study are publicly available in the following links, ISEAR—<https://www.unige.ch/cisa/research/materials-and-online-research/research-material/>. Twitter Sentiment—<https://www.kaggle.com/c/tweet-sentiment-extraction/data>. SemEval 2007 ("Affective Text")—[https://web.eecs.umich.edu/~mihalcea/affective\\_text/#datasets](https://web.eecs.umich.edu/~mihalcea/affective_text/#datasets). Fairy Tales—<http://people.rc.rit.edu/~coagla/affectdata/index.html>. GoEmotions <https://www.kaggle.com/datasets/debarshichanda/goemotions>. SemEval-2018—[https://competitions.codalab.org/competitions/17751#learn\\_the\\_details-datasets](https://competitions.codalab.org/competitions/17751#learn_the_details-datasets). Financial PhraseBank—<https://github.com/ProsusAI/finBE>. RT. Senti4SD Dataset—<https://github.com/collab-uniba/Senti4SD>. As noted above, all data used in the experiments are open source and publicly accessible from the original data owners and corresponding hyperlinks.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 5 May 2023 Accepted: 28 June 2024

Published online: 10 July 2024

## References

1. Olshannikova E, Olsson T, Huhtamäki J, Kärkkäinen H. Conceptualizing big social data. *J Big Data*. 2017;4:3. <https://doi.org/10.1186/s40537-017-0063-x>.
2. K. A, P. D, Sam Abraham S, V. L. L, P. Gangan M., Readers' affect: predicting and understanding readers' emotions with deep learning. *J Big Data*. 2022;9:82. <https://doi.org/10.1186/s40537-022-00614-2>.
3. Acheampong FA, Wenyu C, Nunoo-Mensah H. Text-based emotion detection: advances, challenges, and opportunities. *Eng Reports*. 2020. <https://doi.org/10.1002/eng2.12189>.
4. Alswaidan N, Menai MEB. A survey of state-of-the-art approaches for emotion recognition in text. *Knowl Inf Syst*. 2020;62:2937–87. <https://doi.org/10.1007/s10115-020-01449-0>.
5. Plutchik R. The nature of emotions. *Am Sci*. 2001;89:344. <https://doi.org/10.1511/2001.4.344>.
6. Strapparava C, Valitutti A (2004) WordNet Affect: an Affective Extension of WordNet. In: Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04). European Language Resources Association (ELRA), Lisbon, Portugal
7. Poria S, Gelbukh A, Cambria E, Hussain A, Huang G-B. EMOsenticspace: a novel framework for affective common-sense reasoning. *Knowl-Based Syst*. 2014;69:108–23. <https://doi.org/10.1016/j.knosys.2014.06.011>.



8. Staiano J, Guerini M (2014) DepecheMood: A lexicon for emotion analysis from crowd-annotated news. 52nd Annu Meet Assoc Comput Linguist ACL 2014 - Proc Conf 2:427–433. <https://doi.org/10.3115/v1/p14-2070>
9. Esuli A, Sebastiani F (2006) SentiWordNet: a publicly available lexical resource for opinion mining. european language resources association (ELRA)
10. Adikari A, Gamage G, de Silva D, Mills N, Wong SMJ, Alahakoon D. A self structuring artificial intelligence framework for deep emotions modeling and analysis on the social web. *Futur Gener Comput Syst*. 2021;116:302–15. <https://doi.org/10.1016/j.future.2020.10.028>.
11. Udochukwu O, He Y. A rule-based approach to implicit emotion detection in text natural language processing and information systems. Berlin: Springer; 2015. [https://doi.org/10.1007/978-3-319-19581-0\\_17](https://doi.org/10.1007/978-3-319-19581-0_17).
12. Seal D, Roy UK, Basak R. Sentence-level emotion detection from text based on semantic rules advances in intelligent systems and computing. Berlin: Springer; 2020. [https://doi.org/10.1007/978-981-13-7166-0\\_42](https://doi.org/10.1007/978-981-13-7166-0_42).
13. Agrawal A, An A (2012) Unsupervised emotion detection from text using semantic and syntactic relations. In: Proceedings - 2012 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2012. IEEE. 346–353. <https://doi.org/10.1109/WI-IAT.2012.170>
14. Latent Semantic Analysis. [https://en.wikipedia.org/wiki/Latent\\_semantic\\_analysis](https://en.wikipedia.org/wiki/Latent_semantic_analysis). Accessed 12 Dec 2022
15. Probabilistic Latent Semantic Analysis. [https://en.wikipedia.org/wiki/Probabilistic\\_latent\\_semantic\\_analysis](https://en.wikipedia.org/wiki/Probabilistic_latent_semantic_analysis). Accessed 12 Dec 2022
16. Mac Kim S, Valitutti A, Calvo RA. Evaluation of unsupervised emotion models to textual affect recognition workshop on computational approaches to analysis and generation of emotion in text association for computational linguistics. Los Angeles: CA; 2010.
17. Hama Aziz RH, Dimillier N. SentiXGboost: enhanced sentiment analysis in social media posts with ensemble XGBoost classifier. *J Chinese Inst Eng*. 2021;44:562–72. <https://doi.org/10.1080/02533839.2021.1933598>.
18. Winata GI, Madotto A, Lin Z, Shin J, Xu Y, Xu P, Fung P (2019) CAiRE HKUST at SemEval-2019 task 3: Hierarchical attention for dialogue emotion classification. In: NAACL HLT 2019 - International Workshop on Semantic Evaluation, SemEval 2019, Proceedings of the 13th Workshop. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 142–147. <https://doi.org/10.18653/v1/S19-2021>
19. Nida H, Mahira K, Mudasar M, Mudasar Ahmed M, Mohsin M. Automatic emotion classifier advances in intelligent systems and computing. Berlin: Springer; 2019. [https://doi.org/10.1007/978-981-13-1708-8\\_52](https://doi.org/10.1007/978-981-13-1708-8_52).
20. Tzacheva A, Ranganathan J, Mylavarapu SY. Actionable pattern discovery for tweet emotions. Berlin: Springer; 2020.
21. Suhasini M, Srinivasu B. Emotion detection framework for twitter data using supervised classifiers advances in intelligent systems and computing. Berlin: Springer; 2020.
22. Jain VK, Kumar S, Fernandes SL. Extraction of emotions from multilingual text using intelligent text processing and computational linguistics. *J Comput Sci*. 2017;21:316–26. <https://doi.org/10.1016/j.jocs.2017.01.010>.
23. Ghanbari-Adivi F, Mosleh M. Text emotion detection in social networks using a novel ensemble classifier based on Parzen Tree Estimator (TPE). *Neural Comput Appl*. 2019;31:8971–83. <https://doi.org/10.1007/s00521-019-04230-9>.
24. Hasan M, Rundensteiner E, Agu E. Automatic emotion detection in text streams by analyzing Twitter data. *Int J Data Sci Anal*. 2019;7:35–51. <https://doi.org/10.1007/s41060-018-0096-z>.
25. Chatterjee A, Narahari KN, Joshi M, Agrawal P (2019) SemEval-2019 task 3: EmoContext contextual emotion detection in text. In: NAACL HLT 2019 - International Workshop on Semantic Evaluation, SemEval 2019, Proceedings of the 13th Workshop. Association for Computational Linguistics, Minneapolis, Minnesota, USA, pp 39–48. <https://doi.org/10.18653/v1/S19-2005>
26. Vijayvergia A, Kumar K. Selective shallow models strength integration for emotion detection using GloVe and LSTM. *Multimed Tools Appl*. 2021;80:28349–63. <https://doi.org/10.1007/s11042-021-10997-8>.
27. Du P, Nie J-Y (2018) Mutux at SemEval-2018 Task 1: Exploring Impacts of Context Information On Emotion Detection. In: Proceedings of The 12th International Workshop on Semantic Evaluation. Association for Computational Linguistics, New Orleans, Louisiana, pp 345–349. <https://doi.org/10.18653/v1/S18-1052>
28. Rozental A, Fleischer D (2018) Amobee at SemEval-2018 Task 1: GRU Neural Network with a CNN Attention Mechanism for Sentiment Classification. In: Proceedings of The 12th International Workshop on Semantic Evaluation. Association for Computational Linguistics, New Orleans, Louisiana, pp 218–225. <https://doi.org/10.18653/v1/S18-1033>
29. Wang Y, Feng S, Wang D, Yu G, Zhang Y. Multi-label chinese microblog emotion classification via convolutional neural network lecture notes in computer science. Berlin: Springer; 2016.
30. Park JH, Xu P, Fung P (2018) PlusEmo2Vec at SemEval-2018 Task 1: Exploiting emotion knowledge from emoji and #hashtags. 264–272. <https://doi.org/10.18653/v1/S18-1039>
31. Cambria E, Li Y, Xing FZ, Poria S, Kwok K (2020) SenticNet 6: Ensemble Application of Symbolic and Subsymbolic AI for Sentiment Analysis. In: International Conference on Information and Knowledge Management, Proceedings. ACM, New York, NY, USA, pp 105–114. <https://doi.org/10.1145/3340531.3412003>
32. Zeng Z, Zhang S, Ren L, Lin H, Yang L (2021) Senti-BSAS: A BERT-based Classification Model with Sentiment Calculating for Happiness Research. In: ACM International Conference Proceeding Series. ACM, New York, NY, USA, pp 272–277. <https://doi.org/10.1145/3467707.3467748>
33. Li X, Gao W, Feng S, Wang D, Joty S (2021) Span-Level Emotion Cause Analysis by BERT-based Graph Attention Network. In: International Conference on Information and Knowledge Management, Proceedings. ACM, New York, NY, USA, pp 3221–3226. <https://doi.org/10.1145/3459637.3482185>
34. Xiao J (2019) Figure eight at SemEval-2019 task 3: Ensemble of transfer learning methods for contextual emotion detection. In: NAACL HLT 2019 - International Workshop on Semantic Evaluation, SemEval 2019, Proceedings of the 13th Workshop. Association for Computational Linguistics, Minneapolis, Minnesota, USA, pp 220–224. <https://doi.org/10.18653/v1/S19-2036>
35. Fadel A, Al-Ayyoub M, Cambria E (2021) JUSTers at SemEval-2020 Task 4: Evaluating Transformer Models against Commonsense Validation and Explanation. In: Proceedings of the Fourteenth Workshop on Semantic Evaluation.

- International Committee for Computational Linguistics, Barcelona (online), pp 535–542. <https://doi.org/10.18653/v1/2020.semeval-1.66>
36. Shen W, Chen J, Quan X, Xie Z. DialogXL: all-in-one xlnet for multi-party conversation emotion recognition. *AAAI*. 2020;35(15):13789–97. <https://doi.org/10.1609/aaai.v35i15.17625>.
  37. Alduailej A, Alothaim A. AraXLNet: pre-trained language model for sentiment analysis of Arabic. *J Big Data*. 2022;9:72. <https://doi.org/10.1186/s40537-022-00625-z>.
  38. Acheampong FA, Nunoo-Mensah H, Chen W. Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artif Intell Rev*. 2021;54:5789–829. <https://doi.org/10.1007/s10462-021-09958-2>.
  39. Dunietz J, Burnham G, Bharadwaj A, Rambow O, Chu-Carroll J, Ferrucci D (2020) To Test Machine Comprehension, Start by Defining Comprehension. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 7839–7859. <https://doi.org/10.18653/v1/2020.acl-main.701>
  40. Wu C-H, Chuang Z-J, Lin Y-C. Emotion recognition from text using semantic labels and separable mixture models. *ACM Trans Asian Lang Inf Process*. 2006;5:165–83. <https://doi.org/10.1145/1165255.1165259>.
  41. Sazzed S, Jayarathna S. SSentiA: a self-supervised sentiment analyzer for classification from unlabeled data. *Mach Learn with Appl*. 2021;4: 100026. <https://doi.org/10.1016/j.mlwa.2021.100026>.
  42. Li TMH, Chau M, Wong PWC, Yip PSF. A hybrid system for online detection of emotional distress proceedings of the pacific Asia conference on intelligence and security informatics. Berlin: Springer; 2012. [https://doi.org/10.1007/978-3-642-30428-6\\_6](https://doi.org/10.1007/978-3-642-30428-6_6).
  43. Ali RH, Pinto G, Lawrie E, Linstead EJ. A large-scale sentiment analysis of tweets pertaining to the 2020 US presidential election. *J Big Data*. 2022;9:79. <https://doi.org/10.1186/s40537-022-00633-z>.
  44. Budiharto W, Meiliana M. Prediction and analysis of Indonesia presidential election from twitter using sentiment analysis. *J Big Data*. 2018;5:51. <https://doi.org/10.1186/s40537-018-0164-1>.
  45. De Silva D, Ranasinghe W, Bandaragoda T, Adikari A, Mills N, Iddamalagoda L, Alahakoon D, Lawrentschuk N, Persad R, Osipov E, Gray R, Bolton D. Machine learning to support social media empowered patients in cancer care and cancer treatment decisions. *PLoS ONE*. 2018;13: e0205855. <https://doi.org/10.1371/journal.pone.0205855>.
  46. Saffar AH, Mann TK, Ofoghi B. Textual emotion detection in health: advances and applications. *J Biomed Inform*. 2023;137: 104258. <https://doi.org/10.1016/j.jbi.2022.104258>.
  47. Ranasinghe S, Gamage G, Moraliyage H, Mills N, McCaffrey N, Bucholc J, Lane K, Cahill A, White V, De Silva D. An artificial intelligence framework for the detection of emotion transitions in telehealth services. In 2022 15th international conference on human system interaction (HSI). Piscataway: IEEE; 2022. <https://doi.org/10.1109/HSI55341.2022.9869503>.
  48. Nijhawan T, Attigeri G, Ananthakrishna T. Stress detection using natural language processing and machine learning over social interactions. *J Big Data*. 2022;9:33. <https://doi.org/10.1186/s40537-022-00575-6>.
  49. Angskun J, Tipprasert S, Angskun T. Big data analytics on social networks for real-time depression detection. *J Big Data*. 2022;9:69. <https://doi.org/10.1186/s40537-022-00622-2>.
  50. Corti L, Zanetti M, Tricella G, Bonati M. Social media analysis of Twitter tweets related to ASD in 2019–2020, with particular attention to COVID-19: topic modelling and sentiment analysis. *J Big Data*. 2022;9:113. <https://doi.org/10.1186/s40537-022-00666-4>.
  51. Adikari A, Nawaratne R, De Silva D, Ranasinghe S, Alahakoon O, Alahakoon D. Emotions of COVID-19: content analysis of self-reported information using artificial intelligence. *J Med Internet Res*. 2021;23: e27341. <https://doi.org/10.2196/27341>.
  52. Prabagar K, Srikanthabala K, Loganathan N, De Silva D, Gamage G, Rathnayaka P, Perera AS, Alahakoon D. Investigating COVID-19 vaccine messaging in online social networks using artificial intelligence In 2022 15th international conference on human system interaction (HSI). Piscataway: IEEE; 2022. <https://doi.org/10.1109/HSI55341.2022.9869484>.
  53. Sohangir S, Wang D, Pomeranets A, Khoshgoftaar TM. Big data: deep learning for financial sentiment analysis. *J Big Data*. 2018;5:3. <https://doi.org/10.1186/s40537-017-0111-6>.
  54. Kumar S, Zymbler M. A machine learning approach to analyze customer satisfaction from airline tweets. *J Big Data*. 2019;6:62. <https://doi.org/10.1186/s40537-019-0224-1>.
  55. Smith S, O'Hare A. Comparing traditional news and social media with stock price movements; which comes first, the news or the price change? *J Big Data*. 2022;9:47. <https://doi.org/10.1186/s40537-022-00591-6>.
  56. Filieri R, Acikgoz F, Ndou V, Dwivedi Y. Is tripadvisor still relevant? The influence of review credibility, review usefulness, and ease of use on consumers' continuance intention. *Int J Contemp Hosp Manag*. 2021;33:199–223. <https://doi.org/10.1108/IJCHM-05-2020-0402>.
  57. Abdullah M, AlMasawa M, Makki I, Alsolmi M, Mahrous S. Emotions extraction from Arabic tweets. *Int J Comput Appl*. 2020;42:661–75. <https://doi.org/10.1080/1206212X.2018.1482395>.
  58. Sintsova V, Pu P. Dystemo. *ACM Trans Intell Syst Technol*. 2016;8:1–22. <https://doi.org/10.1145/2912147>.
  59. Khattak A, Asghar MZ, Khalid HA, Ahmad H. Emotion classification in poetry text using deep neural network. *Multimed Tools Appl*. 2022;81:26223–44. <https://doi.org/10.1007/s11042-022-12902-3>.
  60. Ahmad S, Asghar MZ, Alotaibi FM, Khan S. Classification of poetry text into the emotional states using deep learning technique. *IEEE Access*. 2020;8:73865–78. <https://doi.org/10.1109/ACCESS.2020.2987842>.
  61. Ilyas A, Shahzad K, Malik MK. Emotion detection in code-mixed roman urdu - english text. *ACM Trans Asian Low-Resource Lang Inf Process*. 2022. <https://doi.org/10.1145/3552515>.
  62. Lee S, Wang Z (2015) Emotion in Code-switching Texts: Corpus Construction and Analysis. In: Proceedings of the Eighth SIGHAN Workshop on Chinese Language Processing. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 91–99. <https://doi.org/10.18653/v1/W15-3116>
  63. Martin L, Pu P. Prediction of helpful reviews using emotions extraction. *Proc AAAI Conf Artif Intell*. 2014. <https://doi.org/10.1609/aaai.v28i1.8937>.
  64. Cao J, Li J, Yin M, Wang Y. Online reviews sentiment analysis and product feature improvement with deep learning. *ACM Trans Asian Low-Resource Lang Inf Process*. 2022. <https://doi.org/10.1145/3522575>.

65. Devlin J, Chang MW, Lee K, Toutanova K (2018) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL HLT 2019 - 2019 Conf North Am Chapter Assoc Comput Linguist Hum Lang Technol - Proc Conf 1:4171–4186. <https://doi.org/10.18653/v1/N19-1423>
66. Ramnath S, Nema P, Sahni D, Khapra MM (2020) Towards Interpreting BERT for Reading Comprehension Based QA. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, Stroudsburg, PA, USA, pp 3236–3242. <https://doi.org/10.18653/v1/2020.emnlp-main.261>
67. Joshi M, Chen D, Liu Y, Weld DS, Zettlemoyer L, Levy O. SpanBERT: improving pre-training by representing and predicting spans. *Trans Assoc Comput Linguist.* 2020;8:64–77. [https://doi.org/10.1162/tacl\\_a\\_00300](https://doi.org/10.1162/tacl_a_00300).
68. Lan Z, Chen M, Goodman S, Gimpel K, Sharma P, Soricut R (2019) ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. In: International Conference on Learning Representations. <https://doi.org/10.48550/arXiv.1909.11942>
69. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V (2019) RoBERTa: A Robustly Optimized BERT Pretraining Approach. <https://doi.org/10.48550/arXiv.1907.11692>
70. Qiu X, Sun T, Xu Y, Shao Y, Dai N, Huang X. Pre-trained models for natural language processing: a survey. *Sci China technol sci.* 2020. <https://doi.org/10.1007/s11431-020-1647-3>.
71. Sanh V, Debut L, Chaumond J, Wolf T (2019) DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. <https://doi.org/10.48550/arXiv.1910.01108>
72. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, Cistac P, Rault T, Louf R, Funtowicz M, Davison J, Shleifer S, von Platen P, Ma C, Jernite Y, Plu J, Xu C, Scao T Le, Gugger S, Drame M, Lhoest Q, Rush AM (2019) HuggingFace's Transformers: State-of-the-art Natural Language Processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. <https://doi.org/10.18653/v1/2020.emnlp-demos.6>
73. Saravia E, Liu H-CT, Huang Y-H, Wu J, Chen Y-S (2018) CARER: Contextualized Affect Representations for Emotion Recognition. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 3687–3697. <https://doi.org/10.18653/v1/D18-1404>
74. Plutchik, R. (1991) *The Emotions*. University Press of America.
75. Thesaurus. <https://www.merriam-webster.com/thesaurus/>. Accessed 12 Dec 2022
76. Syakur MA, Khotimah BK, Rochman EMS, Satoto BD. Integration K-means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conf Ser Mater Sci Eng.* 2018;336: 012017. <https://doi.org/10.1088/1757-899X/336/1/012017>.
77. Bradley PSKP, Bennett. Constrained k-means clustering. *Microsoft Res.* 2000;20:1–8.
78. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. *Proc Int AAAI Conf Web Soc Media.* 2014;8:216–25. <https://doi.org/10.1609/icwsm.v8i1.14550>.
79. Scherer KR, Wallbott HG. Evidence for universality and cultural variation of differential emotion response patterning. *J Pers Soc Psychol.* 1994;66:310–28. <https://doi.org/10.1037/0022-3514.66.2.310>.
80. (2020) Twitter Sentiment Dataset | Kaggle. <https://www.kaggle.com/datasets/saurabhshahane/twitter-sentiment-dataset>. Accessed 12 Dec 2022
81. Alm CO (2009) *Affect in Text and Speech*. VDM Verlag Dr. Müller (January 9, 2009)
82. Demszky D, Movshovitz-Attias D, Ko J, Cowen A, Nemade G, Ravi S (2020) GoEmotions: A Dataset of Fine-Grained Emotions. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp 4040–4054. <https://doi.org/10.18653/v1/2020.acl-main.372>
83. Mohammad S, Bravo-Marquez F, Salameh M, Kiritchenko S (2018) SemEval-2018 Task 1: Affect in Tweets. In: Proceedings of The 12th International Workshop on Semantic Evaluation. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 1–17. <https://doi.org/10.18653/v1/S18-1001>
84. Malo P, Sinha A, Korhonen P, Wallenius J, Takala P. Good debt or bad debt: detecting semantic orientations in economic texts. *J Assoc Inf Sci Technol.* 2014;65:782–96. <https://doi.org/10.1002/asi.23062>.
85. Calefato F, Lanubile F, Maiorano F, Novielli N (2018) Sentiment polarity detection for software development. In: Proceedings of the 40th International Conference on Software Engineering. ACM, New York, NY, USA, pp 128–128. <https://doi.org/10.1145/3180155.3182519>
86. Strapparava C, Mihalcea R (2007) SemEval-2007 Task 14: Affective Text. In: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007). Association for Computational Linguistics, Prague, Czech Republic, pp 70–74. <https://doi.org/10.3115/1621474.1621487>
87. Baccianella S, Esuli A, Sebastiani F (2010) SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). European Language Resources Association (ELRA), Valletta, Malta
88. Plutchik R. The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *Am Sci.* 2001. <https://doi.org/10.1511/2001.28.344>.
89. Agrawal A, An A (2012) Unsupervised emotion detection from text using semantic and syntactic relations. *Proc - 2012 IEEE/WIC/ACM Int Conf Web Intell WI 2012* 346–353. <https://doi.org/10.1109/WI-IAT.2012.170>
90. Emotion dataset. <https://github.com/huggingface/datasets/tree/master/datasets/emotion>. Accessed 12 Dec 2022
91. Loughran T, McDONALD B. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *J Finance.* 2011;66:35–65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>.
92. Darwin C. *The expression of the emotions in man and animals*. London: John Murray; 2004.
93. Barrett LF. The theory of constructed emotion: an active inference account of interoception and categorization. *Soc Cogn Affect Neurosci.* 2017;12:1–23. <https://doi.org/10.1093/scan/nsw154>.
94. Ekman P. Emotions revealed. *Bmj.* 2004;328:0405184. <https://doi.org/10.1136/sbmj.0405184>.
95. Ekman P. Are there basic emotions? *Psychol Rev.* 1992;99:550–3. <https://doi.org/10.1037/0033-295X.99.3.550>.

96. Wu F, Huang Y, Yan J (2017) Active Sentiment Domain Adaptation. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Stroudsburg, PA, USA, pp 1701–1711. <https://doi.org/10.18653/v1/P17-1156>
97. De Silva D, Mills N, El-Ayoubi M, Manic M, Alahakoon D. Chatgpt and generative ai guidelines for addressing academic integrity and augmenting pre-existing chatbots. In: 2023 IEEE International Conference on Industrial Technology (ICIT) 2023 Apr 4 (p. 1–6). IEEE.
98. Mukherjee P, Badr Y, Doppalapudi S, Srinivasan SM, Sangwan RS, Sharma R. Effect of negation in sentences on sentiment analysis and polarity detection. *Procedia Comput Sci.* 2021;185:370–9. <https://doi.org/10.1016/j.procs.2021.05.038>.
99. Kumar Y, Saini S, Sharma H, Payal R, Mishra A (2022) Feedback Investigation on Twitter Dataset Using Classification Approaches. In: Proceedings of International Conference on Recent Trends in Computing. pp 251–262. [https://doi.org/10.1007/978-981-16-7118-0\\_22](https://doi.org/10.1007/978-981-16-7118-0_22)
100. Adoma AF, Henry N-M, Chen W (2020) Comparative Analyses of Bert, Roberta, Distilbert, and Xlnet for Text-Based Emotion Recognition. In: 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). IEEE, pp 117–121
101. Shrivastava K, Kumar S, Jain DK. An effective approach for emotion detection in multimedia text data using sequence based convolutional neural network. *Multimed Tools Appl.* 2019;78:29607–39. <https://doi.org/10.1007/s11042-019-07813-9>.
102. Tzacheva A, Ranganathan J, Mylavarapu SY. Actionable pattern discovery for tweet emotions advances in intelligent systems and computing. Berlin: Springer; 2020. [https://doi.org/10.1007/978-3-030-20454-9\\_5](https://doi.org/10.1007/978-3-030-20454-9_5).
103. Huang Y-H, Lee S-R, Ma M-Y, Chen Y-H, Yu Y-W, Chen Y-S (2019) EmotionX-IDEA: Emotion BERT -- an Affective Model for Conversation. <https://doi.org/10.48550/arXiv.1908.06264>
104. Rabeya T, Ferdous S, Ali HS, Chakraborty NR (2017) A survey on emotion detection: A lexicon based backtracking approach for detecting emotion from Bengali text. In: 2017 20th International Conference of Computer and Information Technology (ICIT). IEEE, pp 1–7. <https://doi.org/10.1109/ICCITECHN.2017.8281855>

### Publisher's Note

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