Cognitive Computation (2022) 14:110–129
https://doi.org/10.1007/s12559-021-09828-7

A Multitask Framework to Detect Depression, Sentiment and Multi-label Emotion from Suicide Notes

Soumitra Ghosh[1] · **Asif Ekbal**[1] · **Pushpak Bhattacharyya**[1]

Received: 16 June 2020 / Accepted: 12 January 2021 / Published online: 5 February 2021

Abstract
The significant rise in suicides is a major cause of concern in public health domain. Depression plays a major role in increasing suicide ideation among the individuals. Although most of the suicides can be avoided with prompt intercession and early diagnosis, it has been a serious challenge to detect the at-risk individuals. Our current work focuses on learning three closely related tasks, *viz.* depression detection, sentiment citation, and to investigate their impact in analysing the mental state of the victims. We extend the existing standard emotion annotated corpus of suicide notes in English, *CEASE*, with additional 2539 sentences collected from 120 new notes. We annotate the consolidated corpus with appropriate depression labels and multi-label emotion classes. We further leverage weak supervision to annotate the corpus with sentiment labels. We propose a deep multitask framework that features a knowledge module that uses SenticNet's *IsaCore* and *AffectiveSpace* vector-spaces to infuse external knowledge specific features into the learning process. The system models emotion recognition (the primary task), depression detection and sentiment classification (the secondary tasks) simultaneously. Experiments show that our proposed multitask system obtains the highest cross-validation MR of 56.47 %. Evaluation results show that all our multitask models perform better than their single-task variants indicating that the secondary tasks (depression detection and sentiment classification) improve the performance of the primary task (emotion recognition) when all tasks are learned jointly.

Keywords Fine-grained emotion recognition · Sentiment classification · Depression detection · Deep multitask learning · Suicide notes

Introduction

People who feel suicidal are swamped by a range of mixed emotions which make them believe death as the only way to end their suffering. Little do they realize that they can be helped and brought out from this temporary state of misery and avert this dreadful thing from happening. Suicide is a major public health concern all over the world as it accounts for second most deaths among young people[1] (age between 15-24 years) in the world and 10th most deaths in the USA (for all ages)[2]. Suicides alone contributed to $69 billion costs of injury to the US[3] in the year 2015. The numbers of cases of suicides are unsettling as around 8 lacs people happens to commit suicide in the world each year[4], out of which 48k is from the USA[5]. To curb this ever-increasing rate of suicide, the necessity for having effective suicide prevention efforts has become extremely essential at this moment.

There is no single reason attributing to the rise or fall of suicide rates. The problem is complex spanning a plethora of mixed emotions [51] felt by a person with suicidal thoughts. Most common factors [82] that contribute to the suicide ideation are:

– *Health factors:* substance abuse, underlying mental disorders, etc.
– *Environmental factors:* relationship issues, work stress, financial crisis, etc.
– *Historical factors:* history of suicides in the family, childhood abuse, etc.

✉ Soumitra Ghosh
ghosh.soumitra2@gmail.com

1 Department of Computer Science and Engineering, Indian Institute of Technology, Patna Patna, India

1 https://save.org/about-suicide/suicide-facts/
2 https://www.apa.org/monitor/2019/03/trends-suicide
3 https://afsp.org/suicide-statistics/
4 https://www.who.int/news-room/fact-sheets/detail/suicide
5 https://www.economist.com/graphic-detail/2020/01/30/americas-suicide-rate-has-increased-for-13-years-in-a-row

Approximately, 90% of people who die by suicides have a mental disorder at the time of their death[6]. More severe than a mental health condition is the chance of it getting unnoticed or untreated which elevates the risk of suicide ideation. A major cause of suicide is a mental illness, specifically and very commonly the depression. It is estimated that 264 million people across all age groups suffer from depression worldwide[7] and 17.6 million Americans each year[8]. A suicide note serves as a rich source of information which on proper analysis can provide a sound understanding of the mental state of the deceased person. This eventually helps experts and other personnel to identify at-risk subjects and provide proper treatment or assistance at the right time, thus, averting the horrendous act from happening.

Albeit firmly related, *sentiment* and *emotion* do not convey the similar thing [9]. Sentiment is frequently evaluated with a *positive* and *negative* score, called polarities. There is a third evaluation called *neutral* that implies an event that is neither positive nor negative. On the contrary, emotion goes beyond into varied digressions in mood of an individual. Therefore, a *positive* sentiment label can be related among various *emotion* classes like happiness, pride, confidence, thankfulness and so forth.

In this work, we put our efforts to investigate the role of depression and polarity information in assisting the recognition of possible emotion classes in suicide texts. Availability of real-life suicide notes is scarce and proper identification of emotions from such emotion-rich content is even more challenging. Multitask learning has proved to be essential when working with correlated tasks. Research consistently shows a strong link between suicide ideation and depression. To study this further, we learn the tasks of depression detection, sentiment classification and emotion recognition in a multitask setting. Such an approach provides the opportunity to improve over the existing single-task systems for the aforesaid tasks where availability of the data is limited. We consider the benchmark dataset, *CEASE* [28], which is an emotion annotated dataset of suicide notes in English, comprising of 2393 sentences (from 205 suicide notes), with single emotion-tag per sentence. To the best of our knowledge, *CEASE*[9] happens to be the only publicly available emotion-annotated corpus of suicide notes at this moment which was introduced by [28] to facilitate research and innovation in emotion recognition and suicide prevention. We extend this corpus by adding 2539 sentences collected from 120 new suicide notes[10]. We consider the same 15 emotion tagset[11] as used in [28] for labelling the instances with multi-label emotions. We adopt *weak supervision* to label each sentence with a sentiment polarity tag, i.e. positive, negative or neutral. We also mark instance with a tag signifying whether the instance shows hint(s) of depression or not (depressive or non-depressive).

We build several end-to-end deep neural-based multitask systems for sentence-level emotion recognition (primary task), depression and sentiment detection (auxiliary tasks) tasks. Each multitask system we build is an improvised and upgraded (performance-wise) version of its predecessor. We infuse external knowledge information to our system by leveraging the effectiveness of a couple of SenticNet's [11] tools. We evaluate our best performing multitask system by comparing to its per-task performance with the single-task models. We further compare our approach with a couple of systems built according to the state-of-the-art multitask architectures designed specifically for emotion and sentiment analysis tasks. Our best model attains a cross-validation Mean Recall (MR) of 56.47% on the emotion recognition task, 1.98% greater than the single-task equivalent model.

The main contributions of our proposed work are summarized below:

– We extend the benchmark dataset, *CEASE*, with additional 2539 sentences collected from 120 suicide notes.
– Each instance in the extended *CEASE* dataset is marked with multilabel emotion classes as opposed to the single-labelling scheme followed in the introductory work [28].
– We employ a weak supervision-based sentence labelling method to associate a sentiment label to each instance in the new dataset.
– Each sentence in the dataset is also marked with a tag signifying whether it shows any sign(s) or effect(s) of depression or not.
– We propose a multi-task learning framework aided by external knowledge information which attains the overall better performance for the emotion classification task on suicide notes.

The rest of the paper is structured as follows. Section 2 describes some of the previous works on this domain. Next, we discuss vividly the dataset creation and annotations in Section 3. In Section 4, we describe our proposed methodology for the multitask experiments. We discuss the

[6] https://chapterland.org/wp-content/uploads/sites/13/2017/11/US_FactsFigures_Flyer.pdf

[7] https://www.who.int/news-room/fact-sheets/detail/depression

[8] https://emedicine.medscape.com/article/805459-overview

[9] Resource available at: http://www.iitp.ac.in/~ai-nlp-ml/resources.html

[10] Resource to be made available at: https://www.iitp.ac.in/~ai-nlp-ml/resources.html

[11] Emotion labels: *forgiveness, happiness_peacefulness, love, pride, hopefulness, thankfulness, blame, anger, fear, abuse, sorrow, hopelessness, guilt, information, instructions*

experiments performed and their results with discussion in Section 5. Finally, we conclude our work and define the scope of future work in Section 6.

Related Work

The authors in [71] pioneered the study on real-life suicide notes dating back to 1950s, and it was observed that the people attempting suicide were not in a healthy state of mind. This work considered a collection of 66 suicide notes, 33 genuine and 33 fabricated ones, to identify specific features (using discourse analysis-based techniques) for distinguishing the genuine notes from the fake ones. Several other studies based on this resource used various other techniques like text characteristics (modal and auxiliaries) [23], language characteristics (Part-of-Speech variation, etc.) [30, 80] to tackle the same problem. An analysis of the content of some real suicide notes by [53] led to the findings that people with an unrest state of mind wrote conflicting responses frequently and in a disorganized way.

Early 1960 onward, various researchers [14, 83] started studying the content of suicide notes to investigate the various implied reasons for the suicides. The socioeconomic and psychological variables of suicides were studied in [72]. Topic-based analysis of real and simulated suicide notes was performed [68] on a collection of 286 suicide notes, collected from the Birmingham Coroner's Office. An additional 66 notes (33 simulate and 33 genuine notes) from Los Angeles were also part of the study. The study found various overlapping features as well as distinctive features related to each category of notes. The authors in [70] analysed some real suicide notes and observed several gender-based behavioural differences among the individuals who died in suicides.

A gender biasness was also observed [13] concerning the chances of suicide attempts. Women happen to attempt suicide more than men but men have a higher completion rate than that of women. Authors in [31] studied 224 suicide notes from 154 individuals and shared some valuable insights about the content of the notes and writing styles of the authors. Young people wrote lengthier notes with richer emotional content than notes written by elderly people. Young people wrote more of regrets and apologies while elderly people wrote mostly in a neutral tone (more information and instructions). Among the notes studied, most of them were from young unmarried females with no history of mental illness and also happened to be the first-time attempters. The feeling of guilt towards love ones is observed more in the notes of completers than attempters [39]. It was found that suicide is common among the young unmarried females with prior history of mental illness and with lack of basic education.

In a series of attempts [57, 58, 60] to check the effectiveness of machine-learning algorithms in distinguishing real from the simulated suicide notes, it was observed that machine learning methods outperformed human annotators in terms of classification accuracy. Because of the sensitiveness and stigma associated with any suicide act, availability of suicide notes in the public domain is quite limited. Within this context, Track 2 of the 2011 i2b2/VA/Cincinnati Medical Natural Language Processing Challenge was introduced by [59] to facilitate research and innovation in emotion recognition and suicide prevention. The task organizers made available a large emotion annotated corpus of sentences collected from 900 suicide notes in English. Each sentence in the corpus was labelled with multiple labels from a set of fine-grained emotion classes[12]. This dataset facilitated the development in many facets [22, 43, 64, 85, 87] from a classification perspective using the various NLP techniques [47].

In [26] the role of various biometric applications (speech, handwriting, face, etc.) in detecting depression, Parkinson's and Alzheimer's disease was studied. The study [40] showed conclusive shred of evidences of depression being a major reason behind a suicide or at least a primary factor behind increasing the risk of suicide. Independent researchers have shown direct pieces of evidence of association between depression and disability, depression and suicide ideation and depression and suicide ideation in people suffering from Multiple Sclerosis.

In [45] a cross-sectional study was performed to investigate the relationship between all three. The authors in [79] observed that depressive symptoms correlate positively with suicide ideation, whereas depressive symptoms and suicide ideation both have negative correlate with positive mental health. Childhood trauma and depression increase the risk of suicide ideation by five-folds and six-folds, respectively [27].

A psychometric-based (using interRAI ChYMH-S data) decision support system has been reported in [76] to detect children and youth who are at high risk of causing self-harm or engage in suicidal acts. A comprehensive survey presented in [37] shows the various detection methods for suicide ideation spanning several clinical methods, feature engineering based machine learning approaches or deep learning based methods. To detect suicidal ideation and provide early intervention to at-risk individuals, many researchers conducted psychological and clinical studies [81] using the classified responses of questionnaires [21, 73] and analysis of social media data [46] based on feature

[12] *abuse, anger, blame, fear, forgiveness, guilt, happiness_peacefulness, hopefulness, hopelessness, love, pride, sorrow, thankfulness, information and instructions*

engineering [52, 69], sentiment analysis [63, 89], and deep learning [6, 38, 36] techniques.

The study in [50] showed the feasibility and effectiveness of available mobile applications meant for suicide prevention. Authors assessed 69 apps adhering to six evidence-based clinical guideline recommendations[13] out of which only 5 apps incorporated all the six strategies. The 7% apps provided non-existent or wrong helpline numbers, whereas most of the apps provided emergency contact information and/or a crisis helpline number. A novel semi-supervised machine learning model has been proposed in [33] for reasoning on a knowledge base using a combination of random projection scaling and support vector machine. Results show commendable improvement in emotion and sentiment detection tasks paving the way for exploration of semi-supervised based approaches to big social data analytics. In study [67] on internet-based interventions (IBIs) for depression, it was observed that suicidality is viewed more as a criterion of exclusion than as a treatable condition.

Understanding the role that emotions play in human behaviour can lead to the development of better, competent and flexible Artificial Intelligence (AI)-based systems [2, 3]. The authors in their work [86] addressed the problem of ambiguity in contextual polarity using opinion-level features and a probabilistic Bayesian model and achieved significant improvements in word polarity disambiguation across the four domains. The deep ensemble multitask system has been presented in [3] for detecting emotion, sentiment and intensity values for several problems and domains. This study has motivated us in building one of our baselines in this work. A deep stacked model has been presented in [2] to the predict intensities of emotions and sentiments. Relation network based attention mechanism has been introduced in [35] to enhance the performance of suicidal ideation and for other mental disorders detection. The study in [78] addressed the computation of emotions through a theoretical framework and broadened the scope for exploration of the biaseness of emotion towards the factors associated with emotional processing in both normal cognitive functions and pathological dysfunctions in affective disorders. Because of the high emotive content in suicide notes and varying types of emotions that cannot be restricted to Ekman's [24] or Plutchik's [61] basic emotions, we adopt the 15 fine-grained emotion tags as introduced in the 2011 i2b2/VA/Cincinnati Medical Natural Language Processing Challenge, Track 2 Shared Task for sentiment analysis in suicide notes [59]. This Shared Task focused on the development of automatic systems that identify, at the sentence level, affective text of 15 specific emotions from suicide notes. We develop several deep multitask classifiers for depression detection, sentiment classification and emotion recognition tasks. We draw performance comparison among the models developed for single-task and multitask systems, and also with some of the state-of-the-art systems.

[13] mood and suicidal thought tracking, safety plan development, the recommendation of activities to deter suicidal thoughts, information and education, access to support networks, and access to emergency counselling

Popular Emotion Models in Computational Analysis

Several researchers and psychologists introduced various models of emotion mostly pertaining to the studies related to psychology. Primarily, they can be broadly divided into two categories: *dimensional models of emotion*, involving emotion continuum with several dimensions, and *categorical models of emotion* (involving distinct categories of emotion). The study presented in [42] provides a systematic presentation of a diverse range of available emotion annotated corpora following various annotation schema, domains and file formats. We briefly discuss some of the popular models of emotions in this section.

Russell's Circumplex Model of Emotions

Introduced by [65], this particular model of emotion has found applications not only in psycholinguistic studies but also in computational linguistics. This model arranges various states of emotion across 2 dimensions in a circular space. The two dimensions are *Valence*, signifying the degree of pleasantness or unpleasantness of an emotion, and *Arousal*, signifying the degree of calmness or excitement felt in presence of a particular emotion. The model enables continuous measurements of valence (often referred as polarity) and arousal (often referred as intensity) unlike mapping instances to discrete emotional categories (as in categorical models of emotions).

Valence–Arousal–Dominance (VAD) Space

Categorical models [24, 61] may not be sufficient to realize the range of complex emotions that humans feels and expresses in daily basis. Dimensional emotion models like VAD model [66] helps to represent these various emotions systematically. This model maps emotional states as real-valued vector in an orthogonal VAD space accounting for fine-grained emotions compared to categorical models. The framework presented in [54] shows the prediction of fine-grained dimensional VAD scores from a model trained on discrete emotion labels. A multitask approach for emotion prediction across VAD dimensions is presented in [8] that outperforms the single-task learning method across diverse languages.

Ekman's Basic Emotions

The six basic emotions of Ekman [25] comprising of anger, disgust, fear, joy, sadness and surprise could accommodate the discrete categories as opposed to continuous nature of Russell's circumplex model. Among these basic emotions, anger-disgust and fear-surprise are believed to be closely overlapping emotions. Ekman's basic emotions are well acknowledged in computational analysis and are extensively used [4, 16] in the field of emotion analysis specifically for the emotion recognition tasks. The neural network architecture presented in [5] considered both syntactic/semantic and emotional information contained in the training data and evaluated their approach on several datasets that followed Ekman's notion of emotion.

Plutchik's Wheel of Emotions

Robert Plutchik introduced an emotional model, popularly known as *Plutchik's wheel of emotions* [61], where various states of emotions are arranged in a wheel-like structure. The model consists of eight primary bi-polar emotions *(joy–sadness; anger–fear; trust–disgust; surprise–anticipation)* and eight derived emotions (from the basic ones). The emotions are arranged in the wheel based on their level of intensities, with most intense emotions concentrated near the centre of the wheel and least intense emotions near the circumference of the wheel. The authors [1] introduced a large annotated dataset based on Plutchik's fine-grained emotions and developed deep-learning models on the same. In the work by [90], Plutchik's 8 primary emotions are used in a multilabel setting along with their intensities.

[77] revisited the Hourglass of Emotions which is an emotion categorization model and a reinterpretation of Plutchik's Wheel of Emotions. The Hourglass of Emotions is mainly optimized for sentiment analysis. [84] has reviewed and discussed various existing emotion classification models and also proposed several approaches that have potential to enhance the existing emotion studies. Comparatively, categorical models of emotion have found more applications in computational analysis than the dimensional models. However, these models are insufficient to capture subtle emotional states like *abuse, blame, guilt, hopefulness, hopelessness, thankfulness, etc.,* which exist and are considerably relevant in the context of suicide notes. This motivates us to work with the CEASE dataset that deals with a rich set of fine-grained emotions that suits our requirement.

Corpus Creation

The *CEASE* dataset [28] consists of 2393 English sentences collected from the 205 real-life suicide notes. We extend this benchmark corpus by adding 2539 instances from 120 new real-life suicide notes. To minimize the variations in the quality, nature and type of data, we try to adhere to the similar data collection, cleaning and annotation procedures as mentioned in the introductory work [28] while extending the *CEASE* corpus with new instances.

Data Collection

We mainly try to collect notes with a substantial amount of contents in them, ignoring the small notes. To attain this, we use various search criteria[14] while scavenging for notes using a web browser. We also look at the resultant images for the same search criteria. Instances of images showing lengthy writings are chosen and further searched for their credible source(s) to fetch the note's content. On availability of multiple sources for the same note, we consider the more credible or popular source to minimize the chance of the note being a fabricated one. The majority of sources of our collected notes are news bulletins, e-newspapers, blogs, etc.[15]. Handwritten notes found in an image format are manually transcripted to its plain-text form. We use a third party OCR software[16] to convert non-handwritten notes (typed notes) to its corresponding plain text.

Data Annotation

To maintain the anonymity of names of persons, places, organization details, personal details, etc., mentioned in any suicide note, we replace such instances with relevant tag(s) from the following tags: *NAME, ADDRESS, ORGANIZATION, TIME, DATE, SECTION_NOT_CLEAR, DELETED, SIGNATURE*. All plain texts are sentence

[14] 'suicide notes', 'suicide note + < YEAR >', 'social media + suicide note', etc.

[15] List of some sources of collected notes:
https://www.denverpost.com/2007/12/07/mall-gunmans-suicide-note-i-just-snapped/
https://www.huffpost.com/entry/bill-zeller-dead-princeto_n_805689
https://twitter.com/harrisongolden/status/880816422145970177
https://www.mirror.co.uk/sport/football/news/gary-speeds-widow-speaks-suicide-13256742
http://www.homelandnewsng.com/other-news/4518-18-year-old-girl-kills-self-over-man
https://www.writechoice.co.in/writechoice/geetika-sharma-suicide-note/
https://en.wikipedia.org/wiki/File:Snead-Ocey_note.jpg
https://www.storypick.com/doctor-suicide-gay-husband-dowry/

[16] https://ocr.space/

Table 1 Annotation samples

Sentence	Is_Dep	Sent	Emo1	Emo2	Emo3
But now, its pretty empty.	depressive	negative	hopelessness	information	
They will destroy everything that is mine.	depressive	negative	hopelessness	fear	sorrow
you're my best friend and you're such an amazing person and you're going to be so successful and i love you.	non-depressive	positive	pride	hopefulness	love
I never thought I would end up letting all of you down.	depressive	negative	sorrow	guilt	
It's been long and brutal and full of a lot of tears and angst and it's finally coming to a close	depressive	negative	sorrow	fear	guilt
Thank you for being the best. parents.	non-depressive	positive	thankfulness		

tokenized using NLTK sentence tokenizer[17] and shuffled to prevent any chance of establishment of a full note from its sentences. Each instance of the newly collected data is annotated by 3 annotators for depression, sentiment and emotion classes. The newly collected 2539 instances are merged and shuffled with 2393 instances from the *CEASE* dataset to result in a dataset of 4932 instances and annotated for depression, sentiment and emotion labels.

Annotation for Depression

Annotators are provided with ample materials[18] [7, 34] related to the study of depression to get acquainted with the problem in hand. Instances are marked with the label depressive if it contains any recognized symptom(s) or sign(s) of depression. Instances citing any fact, event or situation that may be a consequence of depression is also marked to be depressive. Rest all the instances are marked as non-depressive. Some common symptoms of depression as recognized by many clinical studies are:

- Feeling of sadness, hopelessness, guilt and helplessness.
- Lack of interest in daily activities or hobbies.
- Appetite and sleep changes and other health-related problems.
- Trouble in concentrating and frequent mood changes.
- Pessimistic thoughts leading to thoughts of suicide or suicide attempts.

Inter-rater Agreement To know the consistency among the ratings of the annotators, we use Fleiss-Kappa [74] measure to calculate the inter-rater agreement for the depression annotation task. The kappa, κ, can be defined as,

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

where, $1 - \bar{P}_e$ is the *attainable* degree of agreement above chance and $\bar{P} - \bar{P}_e$ is the *actually attained* degree of agreement above chance. We attain an agreement score of 0.7855 that shows that the annotations are of good quality.

Multi-label Emotion Annotation

One of the limitations of CEASE dataset is that instances are labelled with single emotion only. This does not allow to capture other secondary emotions for the sentences that also carry vital emotional traits. We take this opportunity to address this limitation by assigning each instance in our dataset with multiple labels (at most 3 emotion classes, *Emo*1, *Emo*2 and *Emo*3) from the following emotion labels:

- **Positive emotions:** *Forgiveness, Happiness_Peacefulness, Hopefulness, Love, Pride, Thankfulness*
- **Negative emotions:** *Abuse, Anger, Blame, Fear, Guilt, Hopelessness, Sorrow*
- **Neutral emotions:** *Information and Instructions*

The first emotion (*Emo*1) is marked with the most prominent emotion (primary emotion) and then the secondary emotions (if present) are marked in *Emo*2 and *Emo*3.

For the emotion annotation task, where multiple raters are involved in tagging more than one label for a particular instance, popular agreement metrics such as Cohen's Kappa coefficient [19], Fleiss-Kappa [74], etc., cannot be used for this scenario. Another statistical measure is Krippendorff's α coefficient [44] which accommodates multiple annotators (3 in our case), missing data and multilevel agreement as well. In simple terms, Krippendorff's α coefficient can be realized by the following expression:

$$\alpha = 1 - \frac{D_o}{D_e} \quad (1)$$

[17] https://www.nltk.org/api/nltk.tokenize.html

[18] Some https://www.nimh.nih.gov/health/publications/depression/index.shtml

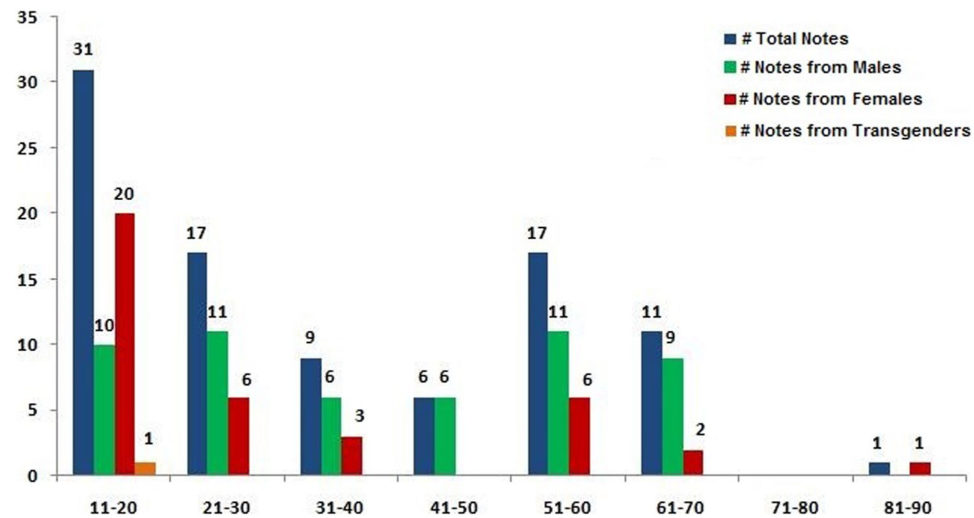

Fig. 1 Gender-wise distribution of notes over various age-intervals

where, D_o is the disagreement observed and D_e is the disagreement expected by chance.

We attain a sentence inter-annotation agreement score of 0.59 which is significantly higher than what was attained in a similar dataset (now discontinued) [59] depicting that the annotations are of good quality.

Some annotated sample instances are shown in Table 1.

Weak Supervision-based Sentiment Annotation

Each emotion class in the fine-grained emotion tagset as introduced by [59] has a strong association with at most one of the sentiment polarities (positive, negative or neutral). There is minimal or no overlap among the sentiment polarities for any particular emotion class. Instances belonging to emotion classes (Abuse, Anger, Blame, Fear, Guilt, Hopelessness and Sorrow) bearing negative sentiment are labelled as *Negative*. Similarly, positive emotion-oriented instances (Forgiveness, happiness_Peacefulness, Hopefulness, Love, Pride, Thankfulness) are marked as *Positive*. Other than these subjective classes, there are two objective classes, *Information* and *Instructions*, that carry no specific emotion in themselves. We label them as *neutral* instances. The first annotated emotion *(Emo1)* being the primary emotion for that instance is considered while deciding on the corresponding sentiment label for that instance (refer to Table 1).

Such an application of weak supervision-based annotation approach saves the effort of performing annotations from scratch and also preserves the natural relationship between various emotions and their associated polarity labels. The incurred cost due to the generation of some noisy labels is very minimal. A somewhat similar approach is used in [87] to associate sentences from suicide notes into distinct sentiment polarities based on their emotion labels. We obtain a skewed distribution of 18%, 31% and 51% of positive, negative and neutral instances, respectively.

Statistical Analysis

Out of total 120 new notes collected, we found the names of the authors of 101 notes from the various sources of the notes or the notes themselves. The collected notes are from the persons of various countries like *USA, India, England, Japan, Zimbabwe, etc.*[19]. Fifteen persons posted their notes online via personal blogs, mail, microblogs. Gender information was found for 104 notes, whereas age of the deceased persons could be established for 92 notes.

Figure 1 shows the gender distribution over various age intervals of the deceased persons whose notes we collected. We observe that for most of the intervals, a number of notes from males are higher than females. Also, the availability of notes is high for the age interval 11-30 and 51-70. We had observed similar distribution in [28]. Among the 92 notes whose gender information is found, 53 are males, 38 are females and 1 transgender.

Proposed Methodology

Multitask approaches have proved to be helpful when learning the correlated tasks [3, 49, 62]. We build an end-to-end multitask deep learning system to detect depression, sentiment and emotion from the suicide sentences, learning all the tasks jointly. We also leverage SenticNet's[20] IsaCore [12] and AffectiveSpace [10] vector-spaces to infuse common and common-sense knowledge to further enhance the performance of our multitask architecture. The overall

[19] Full list of countries: Australia, Canada, China, England, India, Ireland, Japan, Pakistan, Philippines, South Korea, USA and Zimbabwe

[20] https://sentic.net/downloads/

architecture of the proposed approach is shown in Fig. 4. The various components of the architecture are discussed in the following sections.

Word Embedding Representation

Pre-trained word embeddings help to capture both the syntactic and semantic information of the words. The input to the models is the embedded representation of the words in a sentence. The embeddings are fetched from the pre-trained GloVe[21] [56] word embedding which is trained on *Common Crawl* (840 billion tokens) corpus. Inputs sequences are made to be of the same length by padding arrays of zeros. We prepare an *embedding matrix* (E) that contain the embedding vectors (x_{ij}) for the words (w_{ij}) in our training set. We load this embedding matrix into a Keras[22] embedding layer and set the trainable attribute as False to limit the weights/embedding vectors from getting updated during training. The output of the embedding layer is passed to the word sequence encoder to capture the contextual information from the sentence.

GRU-based Word Sequence Encoder

The word vectors from the embedding layer is passed through a Bidirectional GRU [17] layer that captures contextual information from both the directions, forward ($\overrightarrow{GRU}$) and backward ($\overleftarrow{GRU}$) timesteps, and produce hidden representation (h_{it}) of each word (w_{it}) in the sentence (s_i) as shown below:

$$\begin{aligned}\overrightarrow{h_{it}} &= \overrightarrow{GRU}(x_{it}),\\ \overleftarrow{h_{it}} &= \overleftarrow{GRU}(x_{it}),\\ h_{it} &= [\overrightarrow{h_{it}}, \overleftarrow{h_{it}}]\end{aligned}$$

where, i is the sentence number and t refers to the t_{th} word in the sentence. $\overrightarrow{h_{it}}$ and $\overleftarrow{h_{it}}$ are the forward and backward hidden representations of x_{it}, respectively, which are summed up to produce h_{it}.

Attention Layer

To focus on the words that contribute the most in the sentence meaning, we leverage attention mechanism [88] after the word encoding layer in our models that outputs an attended sentence vector. Specifically,

$$\begin{aligned}u_{it} &= tanh(W_w h_{it} + b_w)\\ \alpha_{it} &= \frac{exp(u_{it}{}^T u_w)}{\sum_t exp(u_{it}{}^T u_w)}\\ s_i &= \sum_t (\alpha_{it} * h_{it})\end{aligned}$$

u^{it} is the hidden representation of h^{it} (word representation from word encoding layer). u^w is the context vector and α^{it} is the attention weight for a word. s^i is the output sentence vector from the attention layer.

Loss Functions

We consider *Categorical Cross-entropy* as the cost function (L_{CE}) for the *depression* and *sentiment* tasks as they are two-class classification problems. The loss function can be realized by the following:

$$L_{CE}(a,p) = -\frac{1}{X}\sum_{j=0}^{X}\sum_{i=0}^{Y}(a_{ij} * log(p_{ij}))$$

Multitask Framework for Depression, Sentiment and Emotion detection a is the actual value and p is the predicted value. The double summations from $j = 0\ to\ X$ and $i = 0\ to\ Y$ are over the number of instances (X) and number of classes (Y), respectively. The neural network returns a vector of Y probabilities depicting the chances of belonging of an instance to the respective classes.

For the multi-label *emotion* task, where each instance may belong to more than one class, we consider Binary Cross-entropy to evaluate the loss (L_{BE}) as it allows the model to decide for each class whether a particular instance belongs to that class or not.

$$L_{BE}(a,p) = -\frac{1}{X}\sum_{j=0}^{X}(a_j * log(p_j) + (1-a_j) * log(1-p_j))$$

Multitask Framework

Multitask learning [15] improves feature learning by exploiting the underlying commonality among the correlated tasks and also reduces the computation time involved in testing. More often than not, a multitask system involving a set of tasks produces better output than the task-specific single task systems. We propose a single-level cascaded multitask framework for solving our three-task problem.

Problem Definition

We formally define our problem as follows:

[21] GloVe: http://nlp.stanford.edu/data/wordvecs/glove.42B.300d.zip

[22] a high-level neural networks API: https://keras.io/

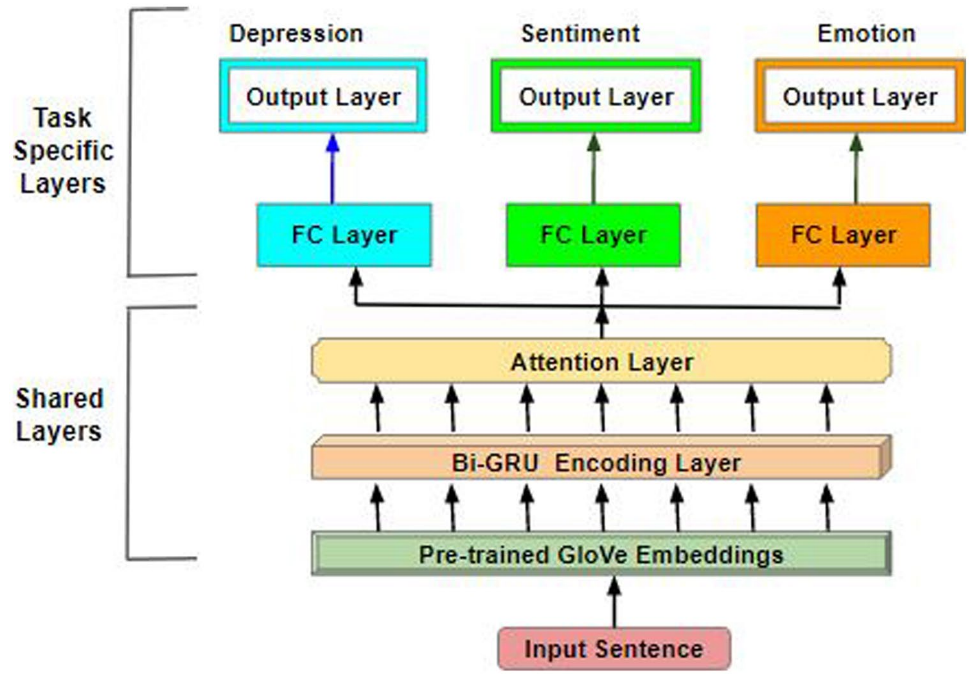

Fig. 2 Basic multitask architecture

Let $(i_a, d_a, s_a, e_a)_{a=1}^{I}$ be a set of I sentences where d_a, s_a and e_a represents the corresponding depression, sentiment and emotion labels for i_a^{th} sentence. Here, $i_a \epsilon\ I$, $d_p \epsilon\ D$ (depression classes), $s_p \epsilon\ S$ (sentiment classes) and $e_p \epsilon\ E$ (emotion classes). The goal of our multitask learning framework is to fit a function that maps an unknown instance i_u (from related domain) to its appropriate depression label d_p, sentiment label s_p and emotion label e_p.

Basic Multitask System (BMS)

Figure 2 shows an end-to-end basic multitask model that takes as input a text instance and produces 3 outputs: depressive/non-depressive label, sentiment class and all possible emotion labels. Our multitask model consists of a combination of shared layers (sharing of information among all the tasks) and some task-specific layers (specific to each task). The input sentence is passed through the embedding layer which has been pre-assigned with embeddings from GloVe. The embedded representation passes through a BiGRU-based word encoder (128 units) that learns the contextual information of the sentence and further gets attended upon by an attention layer that focuses on the relevant parts of the sentence. Till now, the layers are being shared among all three tasks which allow sharing of information among them. The output from the attention layer is passed to three separate task-specific fully connected layers (representing the three distinct tasks) of 100 neurons each and are followed by their respective output layers.

Calculation of Loss: Our overall loss function involving two cross-entropy losses and a binary cross-entropy loss is defined as follows:

$$L = p * L^{D}{}_{CE} + q * L^{S}{}_{CE} + r * L^{E}{}_{BE} \tag{2}$$

We calculate a weighted sum of the various losses from the individual tasks to produce the overall loss. Here, p, q and r are the constants in the range 0 to 1 that sets the loss weights defining the per-task's loss contribution to the overall loss.

Cascaded Multitask System (CMS)

We build a single-level cascaded multitask system (refer to Fig. 3) on top of the previous architecture where a later stage of the model depends on the outputs of an earlier stage. The task-specific learned features from the depression and sentiment tasks (colour coded blue and green dotted lines, respectively) are concatenated with the output from the attention layer. The concatenated rich-feature vector is passed through two successive task-specific fully connected layers and an output layer to get the final emotion predictions. The first feature layer (300 neurons) is intended to capture the increased dimensionality resulted due to the concatenation of the other task features and the second layer (100 neurons) focuses on finding depression and sentiment aware emotion-specific features. We also try multi-level cascading as in [20] by using depression specific features to determine sentiment specific features and then use these depression and sentiment specific features in determining

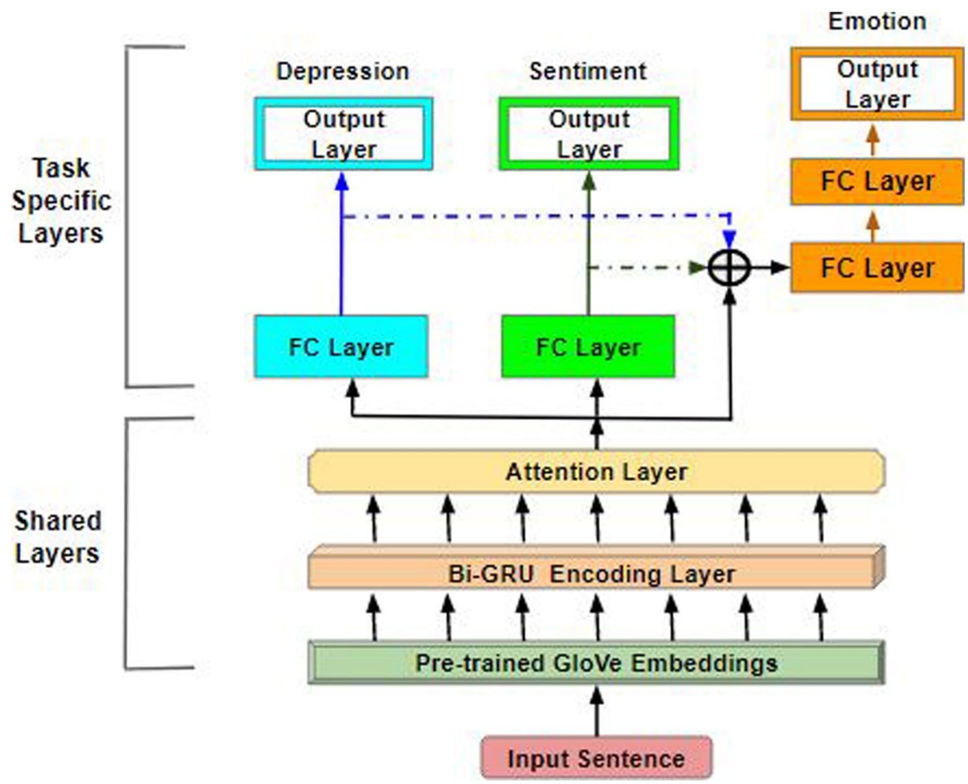

Fig. 3 Cascaded multitask architecture

emotion specific features. It is observed from the results that the single-level cascading produced better results than the multi-level variant, hence we consider the single-level cascaded multitask architecture to build our proposed approach.

Calculation of Loss The calculation of overall loss term is slightly different from that in equation 2 as the individual loss terms of the tasks are not independent here (a later stage's loss relies on the output of an earlier stage). A unified loss function to train our cascaded multitask architecture is defined as follows:

$$L = p * L^{D}{}_{CE} + q * L^{S}{}_{CE} + r * (L^{E}{}_{BE} | L^{D}{}_{CE}, L^{S}{}_{CE}) \tag{3}$$

Eqn.3 indicates that the loss concerning the emotion recognition task, L^E_{BE}, is dependent on both L^D_{CE} and L^S_{CE}.

Cascaded Multitask System with External Knowledge Infusion (CMSEKI)

To enhance the performance of our cascaded multitask system, we incorporate SenticNet's external knowledge tools to the existing architecture that introduces common and common-sense knowledge information to the learning process. Exploiting common-sense knowledge helps to tackle the challenge of aspect-based sentiment analysis as well as targeted sentiment analysis [48].

IsaCore IsaCore [12] is an open-domain resource that helps in inferring general conceptual and affective information and finds applications in many affect identification and opinion mining tasks. It introduces a vector space where both semantic and sentiment polarity are preserved based on the relationships of instances ('birthday party' and 'china') and concepts ('special occasion' and 'country') and affective labels.

AffectiveSpace 2 Although similar in structure, unlike IsaCore which performs aspect extraction by exploiting semantics, Affective Space 2 [10] is concerned with inferring polarity at the concept-level using affective information (sentics). It is a vector space of affective common-sense knowledge (trivial knowledge that would not normally be found on the Web) leveraging multiple relationships (e.g. LocatedAt, IsUsedFor, Arises, etc.).

We prepare two distinct matrices, C and S, that contains the vectors (c_{ij}) and ($_{ij}$) for the words (w_{ij}) in our training set fetched from *IsaCore* [12] and *AffectiveSpace* [10] vector spaces. We load these matrices into two separate Keras embedding layers, C_{EL} and S_{EL}. The same input that is passed to the GloVe embedding layer is also passed to C_{EL} and S_{EL}. Embedded outputs from C_{EL} and S_{EL} are passed through separate Bi-GRU layers followed by the respective attention layers. These results in two distinct representations, C' and S' for the same input according to the different vector spaces involved. Finally, we concatenate C' and S' with G'

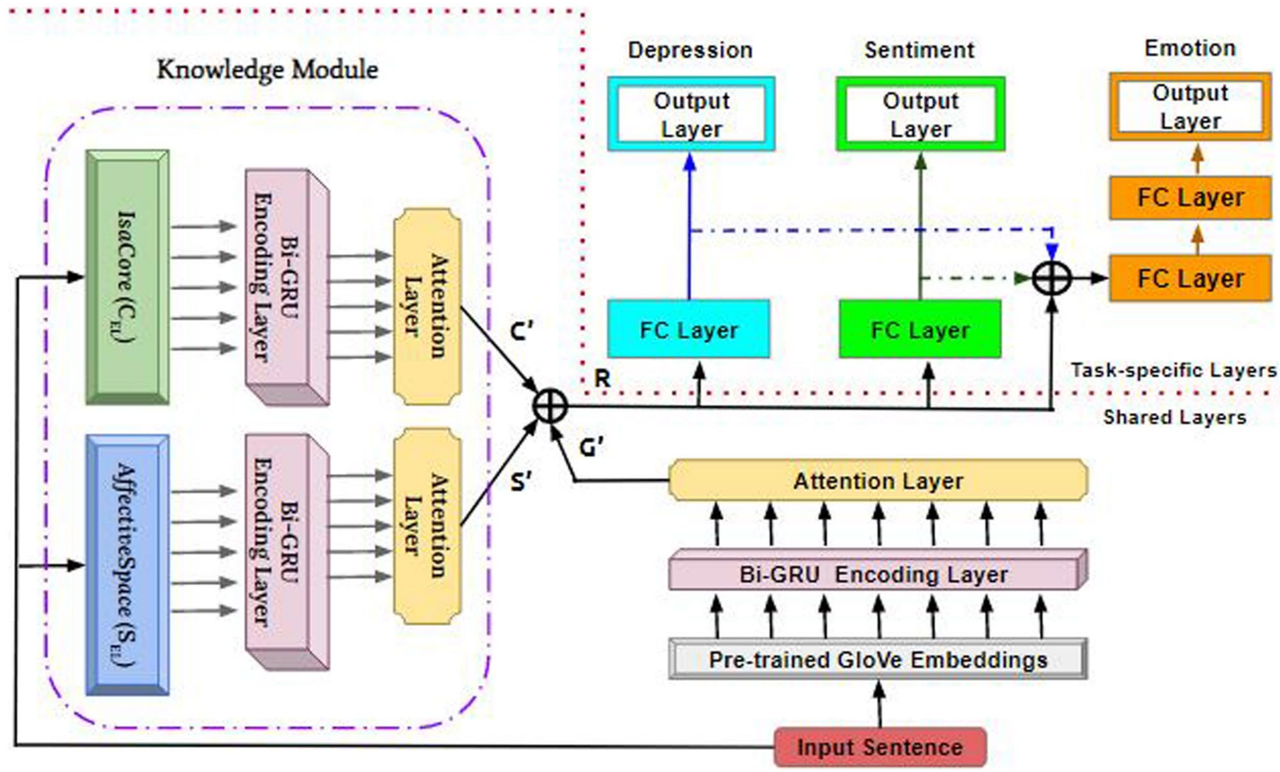

Fig. 4 Multitask model incorporated with external knowledge module. The red dotted line marks the separation of the shared layers from the task-specific layers. **R** is the concatenated output vector from all the three representations (**G'**, **C'**and **S'**)

(the output from the attention layer as in Fig. 3) to produce R, which is passed to the task-specific layers. The overall architecture of the proposed approach is shown in Fig. 4. The calculation of loss is realized by equation 3.

Table 2 shows some domain-specific word pair correlations based on the considered vector spaces to realize how a pair of word correlates differently in different representation schema. Figures 5, 6 and 7 show the arrangement of 100 most frequent words[23] in each of the three vector spaces, *IsaCore*, *AffectiveSpace* and *GloVe*, considered in our experiments.

Along with the multitask models, we build three independent single-task models, STL_Dep, STL_Sent and STL_Emo, for the depression, sentiment and emotion tasks, respectively. The architecture is similar to our BMS approach, except that there are no task-specific layers in the single-task models. Instead, we employ a single fully connected layer after the attention layer followed by the output layer.

[23] not necessarily same 100 words are considered for all 3 spaces, as vectors are fetched as on availability in the respective vector spaces

Model parameters

The output layers for depression detection and sentiment classification tasks use *Softmax* activation with 2 and 3 neurons, respectively. For the emotion recognition task, there are 15 neurons in the output layer and use *Sigmoid* activation. We use *Tanh* activation in the GRU units and also employ a *dropout* [75] and *recurrentdropout* of 25% each in the GRU units to prevent overfitting. We apply dropouts [75] of 25% after the attention layer and also after every linear layer to reduce the impact of overfitting. We perform

Table 2 Cosine similarity scores of word pairs based on various vector spaces. While finding correlations using *IsaCore* which consists of 'concepts' (not necessarily of single word always), we consider: 'happy family' in place of 'happy'; 'family out' in place of 'family'; 'love story' in place of 'love'; 'forgiveness' in place of 'forgive'; 'death family' in place of 'death'

Word Pairs	IsaCore	Affective Space	Glove
(suicide–happy)	0.07	0.21	0.20
(family–love)	0.04	0.39	0.44
(suicide–forgive)	0.38	0.10	0.27
(understand–forgive)	0.90	0.48	0.52
(suicide–death)	0.58	0.27	0.68

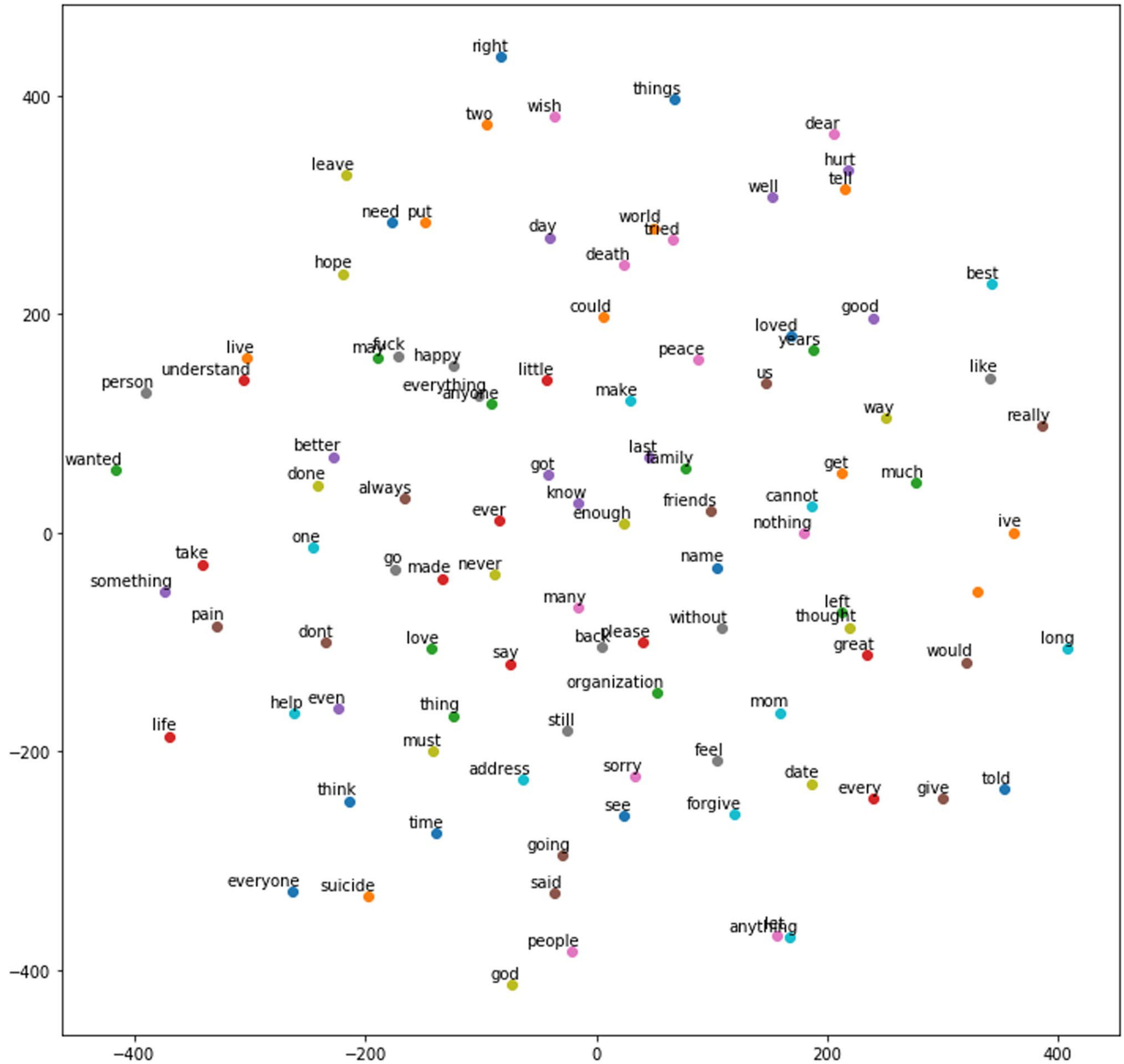

Fig. 5 t-SNE representation of 100 most frequent words from our dataset in IsaCore vector space

backpropagation during training using the *Adam* [41] optimizer to minimize our loss functions for the multitask architectures.

Experiments, Results and Analysis

In this section, we present the experimental setup, report the evaluation results along with the necessary analysis.

Experimental Setup

We use the python-based libraries Keras[24] [18] and Scikit-learn[25] [55] at various stages of our implementations. We compute the accuracies for the depression and sentiment classification tasks and calculate the MR for the multilabel emotion recognition task to evaluate our models. MR is calculated by calculating the recall for each sample and

[24] https://keras.io/

[25] https://scikit-learn.org/stable/

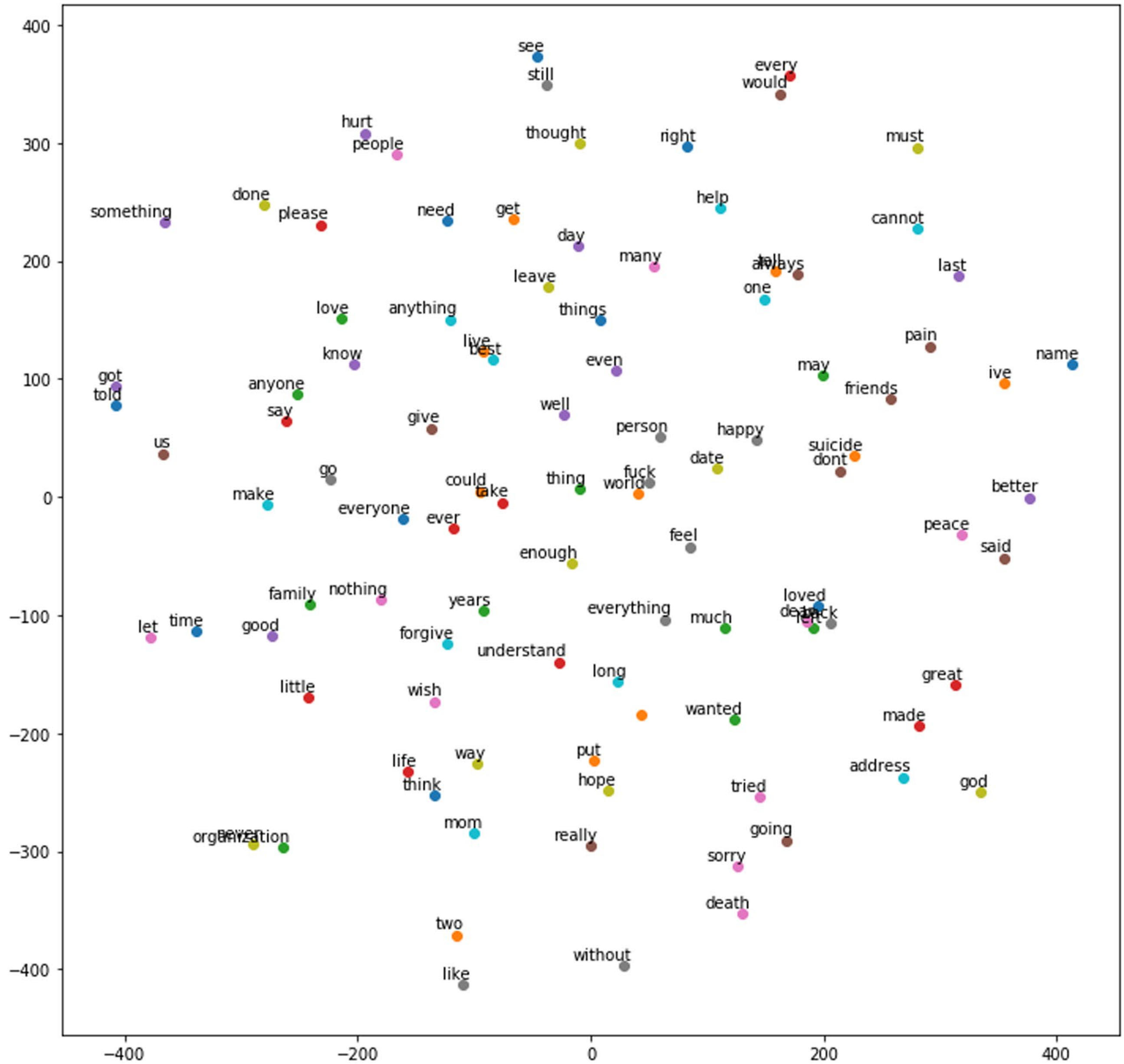

Fig. 6 t-SNE representation of 100 most frequent words from our dataset in AffectiveSpace vector space

averaging the per-sample recall scores over all samples. The per-sample recall (R) is computed as follows:

$$R = \frac{|True \cap Pred|}{|True|} \tag{4}$$

Here, *Pred* is the set of predicted categories and *True* is the set of actual labels for any particular instance. *R* signifies the fraction of the true labels that are correctly predicted. We also calculate the samples-averaged F1 (used specifically for multilabel tasks) and micro-averaged F1 (performance of the model on the overall dataset [32]) for the emotion task. We perform 10-fold cross-validation on our dataset for the experiments and report the averaged scores (over 10-folds) for the metrics. In the multitasking experiments, we weigh the losses[26] produced from the depression, sentiment and emotion tasks by order of 0.3. 0.3 and 1, respectively. We employ the Grid Search method to find the optimal loss weight constants. Table 3 shows the details of various hyper-parameters for training deep learning systems.

Results and Discussion

Table 4 shows the average cross-validation accuracies, samples-averaged and micro-averaged F1 for our various

[26] using *loss_weights* parameter of keras *compile* function

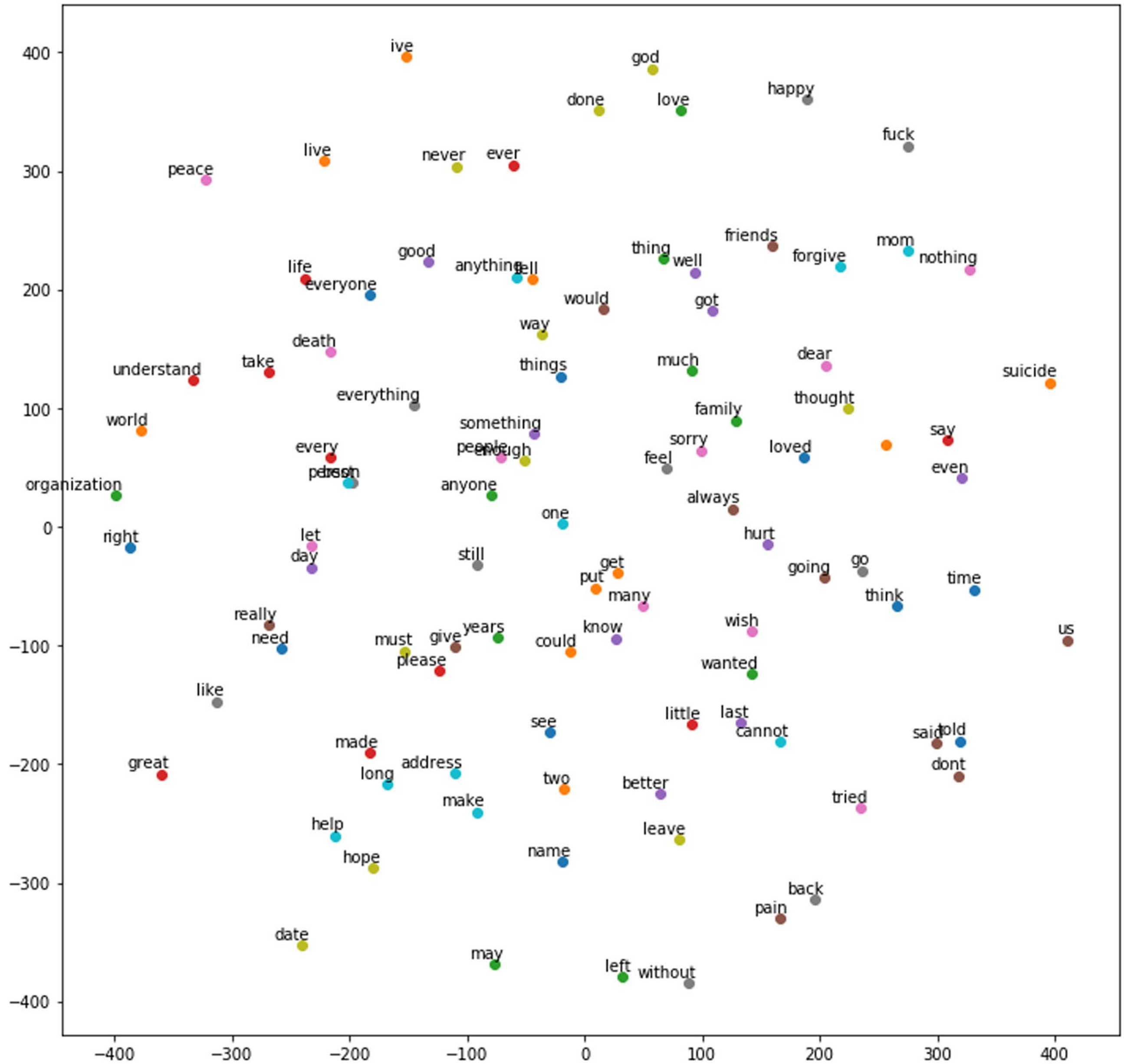

Fig. 7 t-SNE representation of 100 most frequent words from our dataset in GloVe vector space

models over the various tasks. Both the *BMS* and the *CMS* systems outperform the single-task classifiers for the sentiment and emotion tasks by significant margins. Furthermore, *CMS* system scores at par as the single-task classifier for the depression task. The cascading of outputs from the depression and sentiment specific layer to build emotion-specific features in the CMS model has shown significant improvement of 1.22 % for the depression tasks from the *BMS* system.

To get an estimate of the contribution of the individual auxiliary tasks for the, we train two additional multitask models: A). depression with emotion and B). sentiment with emotion, and observe their performance on the dataset. Comparing the obtained results with the single-task models, there is an improvement of 0.24 and 0.53 points on the emotion task from Model A and Model B, respectively. Performance fall of 0.22% is noted on the depression task from Model A and improvement of 0.44 point on the sentiment task from Model B. These results indicate that knowledge of sentiment is more effective than the depression task in improving the performance of the emotion detection task.

We build the CMSEKI model on top of CMS by including an *external knowledge module* guided by SenticNet's *IsaCore*

Table 3 Details of various hyper-parameters related to our experiments

Parameters	Details
Shared Layers	*All models:* 1 GRU layer (256 neurons) and 1 attention layer *Single-task models:* 1 fully connected layer (100 neurons) each *Multitask CMSEKI model:* 2 additional pairs of (1 GRU layer (256) + 1 attention layer) specific to IsaCore and AffectiveSpace vector spaces
Task-Specific layers	*All Multitask models:* 1 fully connected layer (100) each for depression and sentiment tasks and 2 fully connected layers (100) for emotion task
Output layer(s)	*Single-task models:* 1 output layer each (3, 2 and 15 neurons for depression, sentiment and emotion task, respectively) *Multitask models:* 3 output layers each representing depression (2 neurons), sentiment (3 neurons) and emotion task (15 neurons)
Hidden Activations	*ReLU* [29] for fully connected layers
Output Activations	*Softmax* for depression and sentiment tasks; *Sigmoid* for emotion task
Batch size	32
Epochs	20
Dropout [75]	25%
Loss	*Categorical Cross-entropy* for depression and sentiment tasks; *Binary Cross-entropy* for emotion task;
Loss Weights	[0.3, 0.3, 1]
Optimizer	*Adam* [41]

and *AffectiveSpace* vector spaces. This has enabled the system to achieve the best-attained MR of 56.47 % for the primary task of multilabel emotion recognition. Our proposed knowledge-infused multitask system can achieve significant improvements in terms of samples-F1 and micro-F1 than the single-task and other multitasks models for the emotion task. It can be observed from Table 4 that all the multitask models perform better than the single-task model for the sentiment task with the best-attained accuracy of 65.33 % by our proposed approach (CMSEKI). The CMSEKI system performs better than the single-task sentiment and the best performing multitask baseline with improvements of 1.24 and 0.16 points, respectively. The proposed best performing system (CMSEKI) is observed to be statistically significant[27] over the next best performing model (BMS) when tested against null hypothesis with p-value 0.035.

Table 5 shows some of the correct classifications from the CMSEKI model as opposed to the part/full misclassifications by the single-task models. The first sentence in the table is misclassified by the single-task depression classifier but correctly predicted by the single-task sentiment classifier. For the emotion task, the single-task system performs fairly well by classifying it as *instructions* though it is not the actual label because the sentence has a similar underlying tone as in most instruction type instances. The multitask system, CMSEKI captures the actual meaning and correctly predicts for all the three tasks. Similarly, in the second sentence as well the multitask predictions are better than the single-task ones. If we look into the third, fourth and fifth sentences in the table, it is observed that the single-task model correctly predicts the sentiment and depression label but could predict only one label correctly (information) for the emotion task. On the other hand, the multi-task model correctly predicts the desired pair of emotions (hopefulness and information). Through quantitative and qualitative analysis of the predicted instances has helped us to understand that the correct classifications are biased more towards classes having a quantifiable amount of instances (*information, instructions, sorrow, hopefulness*) than the under-represented classes (*abuse, pride, fear*). The model also performs well for the explicit instances belonging to any emotion classes (*forgiveness, thankfulness, love*) irrespective of the number of training instances in these classes. For example:

- I **forgive** him because he is born of the same mother & father as i am. Predicted: *Forgiveness*
- **Thank** you Lord for being here for me. Predicted: *Thankfulness*

In the above couple of sentences, the presence of explicit words (**forgive, thank**) depicting a particular emotion

[27] We perform *Student's t test* for assessing the statistical significance

Table 4 Scores from 10-fold cross-validation are shown below. MR: mean recall. Values in italics depict improvement in scores of the MTL models over the STL variants. Values in bold are the maximum scores attained

Tasks	Depression	Sentiment	Emotion		
Models	Accuracy (%)		MR (%)	samples-F1 (%)	micro-F1 (%)
Single-task baselines					
STL_Dep	75.34	-	-	-	-
STL_Sent	-	64.09	-	-	-
STL_Emo	-	-	54.49	50.8	50.8
Multitask baselines					
BMS	74.12	65.17	55.42	51.0	51.1
CMS	75.34	65.11	55.06	50.8	50.7
Proposed multitask approach					
CMSEKI	74.35	65.33	56.47	51.8	51.9

makes the classification task less challenging than in case of implicit instances (discussed in the next section) for our model.

Comparison with State-of-the-art Systems

In this work we consider a new dataset with the challenging tasks of depression detection, sentiment identification and emotion recognition. Though direct comparison with the state-of-the-art-systems is not possible, we try to establish the effectiveness of our work through some relative analysis of the similar tasks. Authors in [59] introduced a large dataset of 900 suicide notes with multi-label sentence annotations for emotion in the i2b2 shared task for emotion classification in suicide notes. The mean micro-F1 score attained in this shared task is 0.522 which is very close to 51.9 % attained by our proposed approach (CMSEKI). Moreover, we build our baselines on top of some popular state-of-the-art approaches and necessarily modify the architectures to suit our problem. The reported results show our proposed model outperforms them all on the sentiment and the emotion tasks.

Error Analysis

It is observed that most instances are predicted as belonging to the *information* class. Even in the case of multilabel predictions, the class *information* is frequently predicted along with other emotion of a particular instance. Table 6 shows some sample predictions where the class *information* is predicted by the CMSEKI system. The first two sentences in the table are misclassified for both the depression and sentiment tasks, whereas the 2nd sentence is partly classified for the emotion task. The 3rd sentence in the table

Table 5 Sample predictions where CMSEKI outperformed the single-task models in terms of correct classifications for all the tasks. MTO: multitask output; STO: single-task output

Sentence	Model	Is_Dep	Sent	Emo1	Emo2
never forget to treat yourself with kindness.	MTO	non-depressive	neutral	information	
	STO	depressive	neutral	instructions	
I wish I could have been apart of your growing up but instead I'll have to watch from Heaven.	MTO	non-depressive	neutral	information	hopefulness
	STO	non-depressive	positive	information	
i need justice then only my soul will rest in peace.	MTO	non-depressive	neutral	hopefulness	information
	STO	non-depressive	neutral	information	
I still love writing, english, history (India, world, art, science and maths history) and they are capable of exciting me In the darkest time And, I am the athiest.	MTO	non-depressive	positive	love	information
	STO	non-depressive	neutral	love	
for the past six or seven years i have been enduring the slings and arrows of outrageous misfortune until i feel like a worn out pin cushion.	MTO	depressive	negative	sorrow	information
	STO	depressive	negative	information	

Table 6 Qualitative analysis of some misclassified instances by the CMSEKI system. *Actual* shows the true labels and *Pred* shows the model predictions

Sentence	Labels	Is_Dep	Sent	Emo1	Emo2
some people do not talk to me.	Actual	depressive	negative	sorrow	
	Pred	non-depressive	neutral	information	
Why does fag have to mean anything bad, but i am gay not a fag and just a person like anyone else.	Actual	depressive	negative	sorrow	information
	Pred	non-depressive	neutral	information	
Sorry I could not face everything any more.	Actual	depressive	negative	hopelessness	sorrow
	Pred	non-depressive	negative	guilt	information
But the darkness always returned and built up something like a tolerance, because programming has become less and less of a refuge.	Actual	depressive	negative	hopelessness	sorrow
	Pred	non-depressive	negative	information	sorrow

is correctly classified as bearing *negative* polarity, but the corresponding emotion predicted, though belonging to a negative class (guilt), is not proper. This particular sentence has a sense of feeling pained or something unbearable for a continuous period of time which our multitask model is not able to capture, thus, wrongly predicting it as being *non-depressive*.

Our proposed system also finds difficulty in labelling implicit instances. This can be understood from the last sentence in Table 6 where our model partially predicted the correct emotion label (sorrow) but fails to predict *hopelessness* which is suggested but not directly expressed in the sentence itself.

Conclusion

In this paper, we have presented a multitask framework for depression, sentiment and multilabel emotion recognition. We have developed three multitask frameworks starting from an adaptation of a state-of-the-art approach to a more complex cascaded version of the same and finally an external knowledge-infused network for enhanced performance on the primary task of multilabel emotion recognition. Through our experiments, we investigate whether depression and sentiment specific features aid in learning emotion features when learned upon jointly. We compare the performance of our multitask models with the single-task counterparts representing the distinct tasks. We draw a comparison among the multitask models as well to show the performance change with changes in model complexities.

The proposed knowledge-infused model shows superior performance than the rest of the systems with the best attained MR score of 56.47 % for the emotion recognition and accuracy of 65.33 % for the sentiment classification. Our cascaded multitask system scored top among all the multitask systems for the depression task with 75.34 % accuracy, but could not beat the best score (from STL_DEP model) by a narrow margin of 0.18 %. Comparisons suggest that our proposed *external knowledge-infused* model performs significantly better than the single-task variants and the best performing state-of-the-art multitask system with improvements of 1.98 and 1.41 points on the emotion task, respectively. Overall results for our primary task of emotion recognition suggest that depression and sentiment information help in improving the predictive performance of the emotion recognition task in a multitasking scenario. Through this work, we learn that emotion data from suicide notes remain a challenge and have huge scope for multi-faceted research from multiple perspectives such as psychological analysis, computational psycholinguistic analysis, etc.

In this work, we only considered unigrams from our dataset to get the vector representations from IsaCore and AffectiveSpace. These vector spaces offer a fair share of vectors of popular bigrams as well which we would like to explore in our future work. We also plan to extend this work by studying other relevant contributing factors to suicide ideation such as anxiety, PTSD, depression, etc., from social media data and investigate the power of machine learning-based computational methods in addressing these concerns.

Acknowledgements The authors gratefully acknowledge the support from the project titled 'Development of CDAC Digital Forensic Centre with AI based Knowledge Support Tools', supported by Ministry of Electronics and Information Technology (MeitY), Government of India and Government of Bihar (project number P-264). The authors would also like to thank Suman Sekhar and Saroj Jha for their valuable efforts in labelling the sentences. Asif Ekbal acknowledges the Young Faculty Research Fellowship (YFRF), supported by Visvesvaraya PhD scheme for Electronics and IT, Ministry of Electronics and Information Technology (MeitY), Government of India, being implemented by Digital India Corporation (formerly Media Lab Asia.

Funding This study was funded as part of the project titled 'Development of CDAC Digital Forensic Centre with AI based Knowledge Support Tools' supported by Ministry of Electronics and Information Technology (MeitY), Government of India and Government of Bihar (project number P-264).

Compliance with Ethical Standards

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

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Abdul-Mageed M, Ungar. Emonet L. Fine-grained emotion detection with gated recurrent neural networks. In Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers) 2017. p. 718–728.
2. Akhtar MS, Ekbal A, Cambria E. How intense are you? predicting intensities of emotions and sentiments using stacked ensemble. IEEE Comput Intell Mag. 2020;15(1):64–75.
3. Akhtar MS, Ghosal D, Ekbal A, Bhattacharyya P, Kurohashi S. A multi-task ensemble framework for emotion, sentiment and intensity prediction. arXiv preprint 2018. arXiv:1808.01216
4. Bălan O, Moise G, Petrescu L, Moldoveanu A, Leordeanu M, Moldoveanu F. Emotion classification based on biophysical signals and machine learning techniques. Symmetry. 2020;12(1):21.
5. Batbaatar E, Li M, Ryu KH. Semantic-emotion neural network for emotion recognition from text. IEEE Access. 2019;7:111866–78.
6. Benton A, Mitchell M, Hovy D. Multi-task learning for mental health using social media text. arXiv preprint 2017. arXiv:1712.03538
7. Bucci W, Freedman N. The language of depression. Bull Men Clin. 1981;45(4):334.
8. Buechel S, Hahn U. Word emotion induction for multiple languages as a deep multi-task learning problem. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers) 2018. p. 1907–1918.
9. Cambria E. Affective computing and sentiment analysis. IEEE Intell Syst. 2016;31(2):102–7.
10. Cambria E, Fu J, Bisio F, Poria S. Affectivespace 2: Enabling affective intuition for concept-level sentiment analysis. In B. Bonet and S. Koenig, editors, Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. AAAI Press, Austin, Texas, USA 2015. p. 508–514.
11. Cambria E, Li Y, Xing FZ, Poria S, Kwok K. Senticnet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management 2020. p. 105–114.
12. Cambria E, Song Y, Wang H, Howard N. Semantic multidimensional scaling for open-domain sentiment analysis. IEEE Intell Syst. 2012;29(2):44–51.
13. Canetto SS, Lester D. The epidemiology of women's suicidal behavior. Springer Publishing Co, 1995.
14. Capstick A. Recognition of emotional disturbance and the prevention of suicide. British Medical Journal. 1960;1(5180):1179.
15. Caruana R. Multitask learning. Machine learning. 1997;28(1):41–75.
16. Chen SY, Hsu CC, Kuo CC, Ku LW, et al. Emotionlines: An emotion corpus of multi-party conversations. arXiv preprint 2018. arXiv:1802.08379
17. Cho K, Van Merriënboer B, Bahdanau D, Bengio Y. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint 2014. arXiv:1409.1259
18. Chollet F, et al. Keras: Deep learning library for theano and tensorflow. https://keras.io/k 7(8):T1, 2015.
19. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20(1):37–46.
20. Dai J, He K, Sun J. Instance-aware semantic segmentation via multi-task network cascades. In Proceedings of the IEEE Conf Comput Vis Pattern Recognit 2016. p. 3150–3158.
21. Delgado-Gomez D, Blasco-Fontecilla H, Alegria AA, Legido-Gil T, Artes-Rodriguez A, Baca-Garcia E. Improving the accuracy of suicide attempter classification. Artif Intell Med. 2011;52(3):165–8.
22. Desmet B, Hoste V. Emotion detection in suicide notes. Expert Systems with Applications. 2013;40(16):6351–8.
23. Edelman AM, Renshaw SL. Genuine versus simulated suicide notes: an issue revisited through discourse analysis. Suicide and Life-Threatening Behavior. 1982;12(2):103–13.
24. Ekman P. An argument for basic emotions. Cognit Emot. 1992;6(3–4):169–200.
25. Ekman P, Sorenson ER, Friesen WV. Pan-cultural elements in facial displays of emotion. Science. 1969;164(3875):86–88.
26. Faundez-Zanuy M, Hussain A, Mekyska J, Sesa-Nogueras E, Monte-Moreno E, Esposito A, Chetouani M, Garre-Olmo J, Abel A, Smekal Z, et al. Biometric applications related to human beings: there is life beyond security. Cognit Comput. 2013;5(1):136–51.
27. Gawęda Ł, Pionke R, Krężołek M, Frydecka D, Nelson B, Cechnicki A. The interplay between childhood trauma, cognitive biases, psychotic-like experiences and depression and their additive impact on predicting lifetime suicidal behavior in young adults. Psychol Med. 2020;50(1):116–24.
28. Ghosh S, Ekbal A, Bhattacharyya P. Cease, a corpus of emotion annotated suicide notes in english. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, editors, Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France,European Language Resources Association, 2020. p. 1618–1626.
29. Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. In Proceedings of the fourteenth international conference on artificial intelligence and statistics 2011. p. 315–323.
30. GOTTSCHALK L, GLESER G. An analysis of the verbal content of suicide notes. Br J Med Psychol. 33:195204, 1960.
31. Ho T, Yip PS, Chiu C, Halliday P. Suicide notes: what do they tell us? Acta Psychiatrica Scandinavica. 1998;98(6):467–73.
32. Holub M. Review: Natural language processing for online applications: Text retrieval, extraction and categorization, by peter jackson and isabelle moulinier. Prague Bull. Math. Linguistics. 2005;83:83–4.
33. Hussain A, Cambria E. Semi-supervised learning for big social data analysis. Neurocomputing. 2018;275:1662–1673.
34. Ingram RE, Miranda J, Segal ZV. Cognitive vulnerability to depression. Guilford Press; 1998.
35. Ji S, Li X, Huang Z, Cambria E. Suicidal ideation and mental disorder detection with attentive relation networks. arXiv preprint 2020. arXiv:2004.07601
36. Ji S, Long G, Pan S, Zhu T, Jiang J, Wang S. Detecting suicidal ideation with data protection in online communities. In International Conference on Database Systems for Advanced Applications. Springer, 2019. p. 225–229.

37. Ji S, Pan S, Li X, Cambria E, Long G, Huang Z. Suicidal ideation detection: A review of machine learning methods and applications. arXiv preprint 2019. arXiv:1910.12611
38. Ji S, Yu CP, Fung S, Pan S, Long G. Supervised learning for suicidal ideation detection in online user content. Complex. 2018;6157249(1–6157249):10.
39. Joiner TE, Pettit JW, Walker RL, Voelz ZR, Cruz J, Rudd MD, Lester D. Perceived burdensomeness and suicidality: Two studies on the suicide notes of those attempting and those completing suicide. J Soc Clin Psychol. 2002;21(5):531–45.
40. Joo J, Hwang S, Gallo J. Death ideation and suicidal ideation in a community sample who do not meet criteria for major depression. Crisis. 2016;37(2):161–5.
41. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint 2014. arXiv:1412.6980
42. Klinger R, et al. An analysis of annotated corpora for emotion classification in text. In Proceedings of the 27th International Conference on Computational Linguistics 2018. p. 2104–2119.
43. Kovačević A, Dehghan A, Keane JA, Nenadic G. Topic categorisation of statements in suicide notes with integrated rules and machine learning. Biomedical informatics insights, 5:BII–S8978, 2012.
44. Krippendorff K. Content Analysis: An Introduction to Its Methodology. SAGE Publications, 2013.
45. Lewis V, Williams K, Koko C, Woolmore J, Jones C, Powell T. Disability, depression and suicide ideation in people with multiple sclerosis. J Affect Disord. 2017;208:662–9.
46. Liu G, Wang C, Peng K, Huang H, Li Y, Cheng W. Socinf: Membership inference attacks on social media health data with machine learning. IEEE Transactions on Computational Social Systems. 2019;6(5):907–21.
47. Lopez-Castroman J, Moulahi B, Azé J, Bringay S, Deninotti J, Guillaume S, Baca-García E. Mining social networks to improve suicide prevention: A scoping review. J Neurosci Res. 2019;98:616–25.
48. Ma Y, Peng H, Cambria E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In S. A. McIlraith and K. Q. Weinberger, editors, Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, AAAI Press, 2018. p. 5876–5883.
49. Majumder N, Poria S, Peng H, Chhaya N, Cambria E, Gelbukh A. Sentiment and sarcasm classification with multitask learning. IEEE Intell Syst. 2019;34(3):38–433.
50. Martinengo L, Van Galen L, Lum E, Kowalski M, Subramaniam M, Car J. Suicide prevention and depression apps suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. BMC medicine. 2019;17(1):1–12.
51. O'Connor RC, Nock MK. The psychology of suicidal behaviour. The Lancet Psychiatry. 2014;1(1):73–85.
52. O'dea B, Wan S, Batterham PJ, Calear AL, Paris C, Christensen H. Detecting suicidality on twitter. Internet Interventions. 2015;2(2):183–188.
53. Osgood CE, Walker EG. Motivation and language behavior: a content analysis of suicide notes. J Abnorm Soc Psychol. 1959;59(1):58.
54. Park S, Kim J, Jeon J, Park H, Oh A. Toward dimensional emotion detection from categorical emotion annotations. arXiv preprint 2019. arXiv:1808.01216
55. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. Scikit-learn: Machine learning in python. J Mach Learn Res, 201112(Oct):2825–2830, .
56. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) 2014. p. 1532–1543.
57. Pestian J, Nasrallah H, Matykiewicz P, Bennett A, Leenaars A. Suicide note classification using natural language processing: A content analysis. Biomedical informatics insights, 3:BII–S4706, 2010.
58. Pestian JP, Matykiewicz P, Grupp-Phelan J. Using natural language processing to classify suicide notes. In Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing. Association for Computational Linguistics, 2008. p. 96–97.
59. Pestian JP, Matykiewicz P, Linn-Gust M, South B, Uzuner O, Wiebe J, Cohen KB, Hurdle J, Brew. C Sentiment analysis of suicide notes: A shared task. Biomedical informatics insights 2012. 5:BII–S9042.
60. Pestian J, Matykiewicz P. Classification of suicide notes using natural language processing. Proceedings of ACL Bio NLP 2008, 967.
61. 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;89(4):344–50.
62. Qureshi SA, Saha S, Hasanuzzaman M, Dias G. Multitask representation learning for multimodal estimation of depression level. IEEE Intell Syst. 2019;34(5):45–52.
63. Ren F, Kang X, Quan C. Examining accumulated emotional traits in suicide blogs with an emotion topic model. IEEE J Biomed Health Inform. 2015;20(5):1384–96.
64. Roberts K, Harabagiu SM. statistical and similarity methods for classifying emotion in suicide notes. Biomedical informatics insights 2012. 5:BII–S8958.
65. Russell JA. A circumplex model of affect. J Pers Soc Psychol. 1980;39(6):1161.
66. Russell JA, Mehrabian A. Evidence for a three-factor theory of emotions. Journal of research in Personality. 1977;11(3):273–94.
67. Sander L, Gerhardinger K, Bailey E, Robinson J, Lin J, Cuijpers P, Mühlmann C. Suicide risk management in research on internet-based interventions for depression: A synthesis of the current state and recommendations for future research. J Affect Disord. 2020;263:676–83.
68. Shapero JJ. The language of suicide notes. PhD thesis, University of Birmingham, 2011.
69. Shing HC, Nair S, Zirikly A, Friedenberg M, Daumé H III, Resnik P. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic 2018. p. 25–36.
70. Shneidman ES. Suicide notes reconsidered. Psychiatry. 1973;36(4):379–94.
71. Shneidman ES, Farberow NL. Clues to suicide. Public health reports. 1956;71(2):109.
72. Shneidman ES, Farberow NL. A socio-psychological investigation of suicide. In Perspectives in personality research. Springer, 1960. p. 270–293.
73. Simon GE, Coleman KJ, Rossom RC, Beck A, Oliver M, Johnson E, Whiteside U, Operskalski B, Penfold RB, Shortreed SM, et al. Risk of suicide attempt and suicide death following completion of the patient health questionnaire depression module in community practice. J Clin Psychol. 2016;77(2):221.
74. Spitzer RL, Cohen J, Fleiss JL, Endicott J. Quantification of agreement in psychiatric diagnosis: A new approach. Archives of General Psychiatry. 1967;17(1):83–7.
75. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15(1):1929–58.

76. Stewart SL, Celebre A, Hirdes JP, Poss JW. Risk of suicide and self-harm in kids: The development of an algorithm to identify high-risk individuals within the childrens mental health system. Child Psychiatry Hum Dev 2020. p. 1–12.
77. Susanto Y, Livingstone AG, Ng BC, Cambria E. The hourglass model revisited. IEEE Intell Syst. 2020;35(5):96–102.
78. Tam DN. Computation in emotional processing: quantitative confirmation of proportionality hypothesis for angry unhappy emotional intensity to perceived loss. Cogn Comput. 2011;3(2):394–415.
79. Teismann T, Forkmann T, Brailovskaia J, Siegmann P, Glaesmer H, Margraf J. Positive mental health moderates the association between depression and suicide ideation: A longitudinal study. Int J Clin Health Psychol. 2018;18(1):1–7.
80. Tuckman J, Kleiner RJ, Lavell M. Emotional content of suicide notes. American Journal of Psychiatry. 1959;116(1):59–63.
81. Venek V, Scherer S, Morency L-P, Pestian J, et al. Adolescent suicidal risk assessment in clinician-patient interaction. IEEE Trans Affect Comput. 2017;8(2):204–15.
82. Vioulès MJ, Moulahi B, Azé J, Bringay S. Detection of suicide-related posts in twitter data streams. IBM J Res Dev. 2018;62(1):7–1.
83. Wagner F. Suicide notes. Danish Medical Journal. 1960;7:62–4.
84. Wang Z, Ho SB, Cambria E. A review of emotion sensing: Categorization models and algorithms. Multimedia Tools and Applications 2020. p. 1–30.
85. Wicentowski R, Sydes MR. Emotion detection in suicide notes using maximum entropy classification. Biomedical informatics insights 2012. 5:BII–S8972.
86. Xia Y, Cambria E, Hussain A, Zhao H. Word polarity disambiguation using bayesian model and opinion-level features. Cogn Comput. 2015;7(3):369–80.
87. Yang H, Willis A, De Roeck A, Nuseibeh B. A hybrid model for automatic emotion recognition in suicide notes. Biomedical informatics insights 2012. 5:BII–S8948.
88. Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. Hierarchical attention networks for document classification. In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies 2016. p. 1480–1489.
89. Yue L, Chen W, Li X, Zuo W, Yin M. A survey of sentiment analysis in social media. Knowl Inf Syst 2018. p. 1–47.
90. Zhou D, Zhang X, Zhou Y, Zhao Q, Geng X. Emotion distribution learning from texts. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing 2016. p. 638–647.