Decision Support Systems 159 (2022) 113813

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

# Social influence-based contrast language analysis framework for clinical decision support systems

Xingwei Yang [a], Alexandra Joukova [b], Anteneh Ayanso [c,*], Morteza Zihayat [b]

[a] Smith School of Business, Queen's University, Goodes Hall, 143 Union St, Kingston, ON K7L 3N6, Canada
[b] Ted Rogers School of Management, Ryerson University, 55 Dundas St W, Toronto, ON M5G 2C3, Canada
[c] Goodman School of Business, Brock University, 1812 Sir Isaac Brock Way, St. Catharines, ON L2S 3A1, Canada

ARTICLE INFO

*Keywords:*
Contrast language analysis
Clinical Decision Support System (CDSS)
Depression detection
Social network

ABSTRACT

Depression is a leading mental health problem affecting 300 million people globally. Recent studies show that social networks provide a tremendous potential for mental health professionals as a source of supplemental information about their patients. This study presents a methodological framework for clinical decision support systems (CDSSs) through analysis of social network data to distinguish the language usage of individuals with early signs of depression (i.e., contrast language analysis). By analyzing the contrast language patterns of different user groups, we are able to uncover constructive and actionable insights into the pain points and characteristics of users with signs of depression as decision support mechanisms for clinicians during intervention, (early) diagnosis and treatment plans. First, we discover terms that represent contrasting language by analyzing the percentage difference of terms in two user groups, labeled as"depressed" and"non-depressed" for ease of reference. Second, by building topic models based on social network contents, the *topic-level contrast features* are discovered. Finally, we consider the structure of the social network to discover the *network-level contrast features*. To illustrate the effectiveness of the proposed framework, we present a case study on early depression detection using a real-world dataset. The proposed framework has methodological contributions in enhancing the features and functionalities of CDSS for clinicians. It also contributes to evidence-based health research through cost-effective data and analytical insights that can supplement or improve the traditional survey and time-consuming interview methods.

## 1. Introduction

Depression is the leading cause of disability, and the social and economic costs associated with it are growing rapidly. The National Institute of Mental Health estimates the projected global mental illness costs to be $6.0 trillion by 2030 [1]. Furthermore, depression can cause poor performance at work, school and social relationships. At worst, it can lead to suicide, which is the second leading cause of death among 15–29 year olds [2]. Given the scale of the problem, increasing attention has been directed towards clinical decision support systems (CDSSs) for diagnosis, early intervention and treatment of depression. However, the traditional CDSS environment is characterized by information primarily obtained through surveys and interviews that involve significant resources and one-on-one physical engagements. The quality of data in such settings is very low given that they are self-reported and often impacted by the mental status of the person at the time of reporting [3]. Therefore, it is extremely challenging for existing CDSSs to detect a person with a high risk of depression early enough, thus limiting the opportunity for proactive intervention.

With a growing deployment of electronic health records (EHRs), data-driven CDSSs have become a viable alternative to enhance decisions with respect to diagnosis, early intervention and treatment of health problems [4]. Data-driven CDSSs are empowered by a set of algorithms and analytical methods that can provide automated measurements and advanced decision-making capabilities in different scenarios [5–7]. Furthermore, the scope of functions provided by data-driven CDSSs depends not only on structured clinical datasets, but also the rapidly growing user-generated content (UGC) from patient interactions. In the context of mental health, since language is the most powerful reflection of mental mood and emotions, the UGC representing language features (e.g., sentiments, syntax, or semantic) are used as the most fundamental components of the methods proposed in prior studies

* Corresponding author.
*E-mail address:* aayanso@brocku.ca (A. Ayanso).

https://doi.org/10.1016/j.dss.2022.113813
Received 7 October 2021; Received in revised form 28 April 2022; Accepted 10 May 2022
Available online 20 May 2022

to uncover the *most informative words or topics* for specific purposes (e.g., depression detection) [8–10].

While content-based analysis provides the foundation for understanding mental mood and emotions, the underlying user interaction features (e.g., interactions among patients, and between patients and physicians) are mostly ignored. The social influence theory stipulates that an individual's attitudes, beliefs, and subsequent actions or behaviours are influenced by referent others and peers in their social environments [11]. In particular, online social interactions are interpreted as intentional social actions in the context of participation [12]. In such interaction forums, users write posts and comments with positive or negative emotions to express their feelings and thoughts. In addition, users tend to form norms and a collective sense of social cohesion within groups, as well as separation from others. These norms or conventions involve shared opinions, behaviours and languages, including the use of jargon or acronyms [13]. Therefore, such network-based social processes and interactions present tremendous potential for enhancing the knowledge base and scope of functions of CDSSs with respect to diagnosis, early detection or treatment plans [14–16]. Furthermore, while existing methods in this domain show strong performance accuracy, their downside is the lack of model interpret-ability on their recommendations. In other words, they are limited in explaining the differences in languages and contents used by depressed versus non-depressed users. The rising prevalence of depression, the barriers to obtaining high-quality and prompt data, as well as the lack of interpretable insights to understand differences and interactions in the social environment lead to a high risk of undiagnosed depressed patients. Therefore, there is an urgent need for a CDSS functionality to detect depression and offer justified recommendations during early intervention and treatments.

This study aims to contribute to enhanced scope of functionality and decision-making in CDSSs by proposing a social influence-based contrast language analysis framework that incorporates natural language processing (NLP) and machine learning (ML) methods to detect depression, and potentially offer proactive recommendations and guidance for early intervention and customized treatment plans. The proposed framework extracts a set of language patterns, called *social influence-based contrast language patterns*, in online social groups (e.g., patient status social networks) by considering the attitudes, topics and activities of patients. Social influence-based contrast language patterns refer to the language features (e.g., terms, topics, etc.) that differentiate depressed and non-depressed user groups in expressing their feelings and emotions as well as their influence on other users in the context of an online social environment. For example, more negative-tone words such as"sad","-bad", etc., might be used more by depressed users to describe their feelings, while a non-depressed user might use more positive words such as"happy","good", etc. We argue that the sentiment polarity of terms or topics of content posted by friends may have a positive or negative impact on the user's mental moods in the social network environment. In addition to enhancing decision-making in CDSSs from the methodological perspective, such patterns address the gap between model accuracy and interpret-ability. Lack of explainable recommendation is a major limitation of existing approaches. Thus, the features extracted in our framework can provide clinicians and mental health professionals more actionable insights into the differences in users' mental health states. Moreover, potential concerns regarding the privacy of individuals while using content-based data are addressed through privacy-preserving mechanisms during the different stages of our framework, including data collection, storage and processing [17]. Overall, our framework exhibits a good balance between model accuracy and interpret-ability while maintaining user anonymity through aggregating patterns of user data.

The rest of this paper is structured as follows. In Section 2, we provide a theoretical foundation and a review of related studies. In Section 3, we present our framework, describe the social influence-based contrast language features discovery processes and how our framework can be integrated into a CDSS. In Section 4, we evaluate the effectiveness of our framework through an experiment that involves social influence and early depression detection based on real data collected from Facebook. In Section 5, we provide discussions and implications of our results. Finally, in Section 6, we conclude our paper and outline limitations and future research directions.

## 2. Literature review

### 2.1. Online social platforms, social influence and mental health

Studying mental health using information from online social platforms has received significant attention recently. Social network applications, virtual communities and online social forums have been very instrumental in providing supplemental information for clinical practice to understand sentiments and moods of users with mental health issues [15,18] and detect depression [9,14,16]. Previous work shows that personal information, especially status updates which describe users' current feelings, experience and thoughts can reveal vital health related information [19]. In particular, Moreno et al. [20] study a group of college students, and apply Diagnostic and Statistical Manual (DSM) criteria to one-year status updates from each Facebook profile to determine the prevalence of displayed depression symptoms. The study confirms that college students commonly discuss their depressive symptoms publicly on Facebook. Li et al. [21] analyze users' Sina Weibo status updates, a famous microblogging service in China, to monitor users' stigmatizing attitudes towards depression and show how depression stigma is exhibited on social media. Chen et al. [15] use Twitter users' status updates to explore the role of basic emotions and how they change over time. They detect eight emotion features from each tweet and show that emotion-related expressions from online postings can provide insights into an individual's psychological states and identify users with or at risk of depression. Li and Du [22] focus on the micro-blogging data in this environment as unique and convenient, and propose a two-part framework to investigate the persuasiveness of micro-blogs. Their findings show that the attitudes of opinion leaders in the community can impact the persuasiveness of their messages in the social environment.

According to the social influence theory [11], an individual's attitudes, behaviours or actions are influenced by referent others through different mechanisms (i.e., compliance, identification, and internalization). For example, in the context of social learning [23], social influence occurs when an individual's behavior is influenced by communications with trusted friends [24]. Furthermore, several researchers have studied the impact of social influence on participation in online social networks. For example, Wang et al. [25] examine social influence in the context of knowledge management systems use. Cheung et al. [26] also examine social influence and online social networks in the context of students' use of Facebook. Bagozzi et al. [13] study social influence variables (e.g., subjective norms, group norms, social identity) in conjunction with individual attributes to understand the bases of participation in online social interactions. More directly relevant to this study, Yan and Tan [27] find that different forms of social support exchanged via discussion board posts have different influences on patients' health conditions. The study shows that informational support is the most prevalent type in the online healthcare community; and emotional support plays the most significant role in helping patients develop healthier states. In the context of online smoking cessation communities, Wang et al. [28] leveraged ML techniques on content shared by users to build more effective and personally tailored cessation treatment recommendations. By extracting domain-specific features, author-based features and thread-based features, the researchers showed an improved performance in discerning important user characteristics (i.e., smoking status) to enhance decision support for tailored intervention.

### 2.2. Analysis of user-generated content online

Most of the prior research focuses on studying the language usages in UGC. More specifically, n-grams and Latent Dirichlet Allocation (LDA) topic modeling are the most common methods. LDA is the most popular topic modeling analysis technique that aims to find topics a document belongs to, the proportion of topics in the entire document, and the probability that a particular word will be included in a topic based on the words in the document [29]. However, most of the existing research mainly concentrates on finding the most predictive features for detecting depression or identifying individuals with risk of depression through the analysis of the postings of users and extracting language features [8,9,14]. Resnik et al. [30] examine the differences in language use between depressed and non-depressed users and demonstrate that supervised topic models can perform better than unsupervised LDA for detecting depression. They generate topics with tweets from users using one LDA model. Shen and Rudzicz [31] use n-grams language modeling, LDA, emotional norms and vector embeddings to classify anxiety-related posts from personal narratives in Reddit. Saha et al. [10] build a framework using psycholinguistic processes and LDA to extract languages and topics from Live Journal where users can create and maintain online personal blogs. They use these features to classify mental health-related communities with interest in depression-related topics. They show that topics and linguistic styles they identified have a strong predictive power to classify communities interested in depression-related topics. Schwartz et al. [14] predict and characterize Facebook users' degree of depression based on their language use in their status updates using n-grams and LDA topic analysis with linear regression method. By analyzing the language contents, their model also captures seasonal within-person changes for depression detection. Tadesse et al. [8] employs NLP techniques and ML approaches to detect depression-related posts from Reddit. Their research focuses on finding the best predictive power of depression by exploring the different combination of uni-grams, bi-grams, topics generated by LDA and linguistics features such as Linguistic Inquiry and Word Count (LIWC) [32]. Despite the effectiveness of the previous methods, they mainly focused on UGC as an independent source of data from social interactions among users.

### 2.3. Contrast language in different user groups

Different user groups use different language styles when they express themselves. Several studies have shown that depressed users tend to have more negative language styles compared to non-depressed users. Segrin [38] analyzes social skills deficit associated with depression. Their study shows that more negative self-disclosures and negative verbal contents are exhibited for depressed people, especially when they are interacting with a friend. Slonim [39] also confirms that depressed individuals' verbal contents are more self-focused which include more first-person pronouns such as"I″. Their languages involve more negative sentiments with emotionally negative words such as"sad","unhappy" etc. Rude et al. [40] examine language differences in written essays by currently-depressed, formerly-depressed, and never-depressed college students using text analytics. They explain that currently-depressed students use more negatively valenced words and are more likely to use the word"I″ than did never-depressed students. Formerly-depressed students do not differ from never-depressed students in using negative words. However, formerly-depressed students indicate more frequent use of the word"I″ than never-depressed students in the final portion of the essays. Nguyen et al. [37] explore the differences between online depression communities and other communities in terms of moods, languages styles, and content topics using LDA extracted from online posts. Their study aims to find predictors of depression, consisting of the aspects that differentiate mental health communities from others.Ríssola et al. [41] consider various language dimensions, including vocabulary, psychometric attributes and emotional indicators. They find that positive instances of mental disorders show significant differences from control individuals in how they write and express emotions in social media. Other studies show that there is a perception difference between depressed and non-depressed users and an individual's state of mind will influence his/her perception of the world and the people within it [42]. What online social platform users post online will impact their friends who read their posts. A bad mood can be transferred from one person to another through social interactions [43]. This social interaction therefore leads to the consideration of social influence-based features to study language impact and differences among different user groups. Research such as Nguyen et al. [37] and Ríssola et al. [41] conduct content analysis, but social influence-based analysis is not examined. Yang et al. [9] use linguistic features and consider two approaches: intention modeling and shortest path to investigate the social influence-based features and the social impact of depressed and non-depressed friends in a user's Facebook network. However, topic-level features are not considered. Our paper considers both linguistics and topic-level features and extends Yang et al. [9]'s social influence-based features by obtaining network-based contrast features based on similarity of influential contents from depressed and non-depressed user groups. In Table 1, we illustrate the main differences and contributions of recent studies against our proposed framework in this paper. As shown, this study is the first leverages UGC data while considering the influence of users in generated contents and proposes features that differentiates depressed and non-depressed users.

Contractor and DeChurch [44] provide an integrated understanding of networks (i.e., who to influence whom) as well as the mechanisms of influence (i.e., how to influence) by proposing a Structured-Influence-Process (SIP) framework. They show that the SIP framework bridges the network view of social influence structures (i.e., the science of networks, opinion leaders and peers) with the psychological view of social influence processes (i.e., core social motives) to understand both aspects of social influence. They conclude that the SIP framework of social influence will play a positive role in transforming scientific information into actionable solutions that advance the public good in communities. In this broader context of social influence, our framework can be extended to other problem settings beyond depression detection. More importantly, our comprehensive framework enhances prior methods as it can thoroughly analyze term-level, topic-level and network-based contrast features to find social influence-based contrast patterns for depressed and non-depressed users, and interpret the differences between them. This can enhance the knowledge base as well as the functionality of CDSSs for psychologists and clinicians to provide more targeted diagnosis and treatment plans.

## 3. Social influence-based contrast language analysis framework

*Social influence-based contrast language analysis framework* is our proposed analytics approach to extract language features (e.g., terms, topics, etc.) that differentiate depressed and non-depressed user groups in expressing their feelings and emotions as well as their influence on other users in an online social environment. It can provide deeper insights into the process of discriminating depressed from non-depressed users in terms of content they generate in different online social platforms that contain patient network data (e.g., patient discussion boards, social media, online health forums, etc.). The discovery of such features can strengthen the functionality and capability of CDSSs for such purposes as depression detection, customized treatment plan recommendation and early prevention of depression severity. Fig. 1 shows a conceptual representation of our proposed social influence-based contrast language analysis framework integrated into a data-driven CDSS [45] to improve healthcare delivery by enhancing clinical decisions with targeted medical knowledge, patient information, social interactions, and other clinical and sensory information collected from hospitals, clinics and their patients. The available data can be programmed into a data-driven techniques interface for a CDSS. Data integration and pre-processing are very important components in CDSS

**Table 1**
Recent studies on users' language features using online social platforms.

| Author | Title | What? | Linguistic Features | Topic Modeling | Network Structure | Social Influence-based Contrast Analysis |
|---|---|---|---|---|---|---|
| Chen et al. [33] | A linguistic signaling model of social support exchange in online health communities | Investigate how linguistic signals in online posts, including sentiment valence, linguistic style matching, readability and etc. affect the amount of support received from online health communities. | ✓ | ✘ | ✘ | ✘ |
| Yang et al. [9] | A big data analytics framework for detecting user-level depression from social networks | Detect depression using linguistics features LIWC, intentional modeling, personality trait, and social network structural features from Facebook. | ✓ | ✘ | ✓ | ✘ |
| Feldhege et al. [34] | Who says what? content and participation characteristics in an online depression community | Investigate topics discussed in an online depression community and how they are related to participation styles using LDA topic model. | ✘ | ✓ | ✘ | ✘ |
| Tadesse et al. [8] | Detection of depression-related posts in reddit social media forum | Detect depression-related posts from Reddit using n-grams and LDA topic models. | ✓ | ✓ | ✘ | ✘ |
| Grover et al. [35] | "Technology enabled Health" – Insights from Twitter Analytics with a Socio-technical Perspective | Find top technologies in health domain using hashtag analysis, word analysis and co-occurrence of words within the Tweets. | ✓ | ✘ | ✘ | ✘ |
| Li et al. [21] | Detecting depression stigma on social media: A linguistic analysis | Design more effective stigma reduction strategies by using linguistics analysis methods from a Chinese social media (Weibo). | ✓ | ✘ | ✘ | ✘ |
| Islam et al. [16] | Depression detection from social network data using machine learning techniques | Investigate emotional features, linguistics styles from Facebook users comments to measure and detect depression. | ✓ | ✘ | ✘ | ✘ |
| Nguyen et al. [36] | Kernel-based Features for predicting Population Health Indices from Geocoded Social Media Data | Propose to use kernel-based features based on the distributions of textual features including language features LIWC and LDA topic model from tweets. | ✓ | ✓ | ✘ | ✘ |
| Shen and Rudzicz [31] | Detecting anxiety through reddit | Use n-grams, LDA, emotional norms and vector embeddings to classify anxiety-related posts from personal narratives in Reddit. | ✓ | ✓ | ✘ | ✘ |
| Saha et al. [10] | A framework for classifying online mental health-related communities with an interest in depression | Propose a framework to classify mental health-related communities with interest in depression related topics using psycholinguistic processes and LDA topic model to extract languages and topics from online blogs Live Journal. | ✓ | ✓ | ✘ | ✘ |
| Nguyen et al. [37] | Affective and content analysis of online depression communities | Use moods, psycholinguistic processes and content topics extracted from online communities to differences between depression and control communities who discussed mental issues. | ✓ | ✓ | ✘ | ✘ |

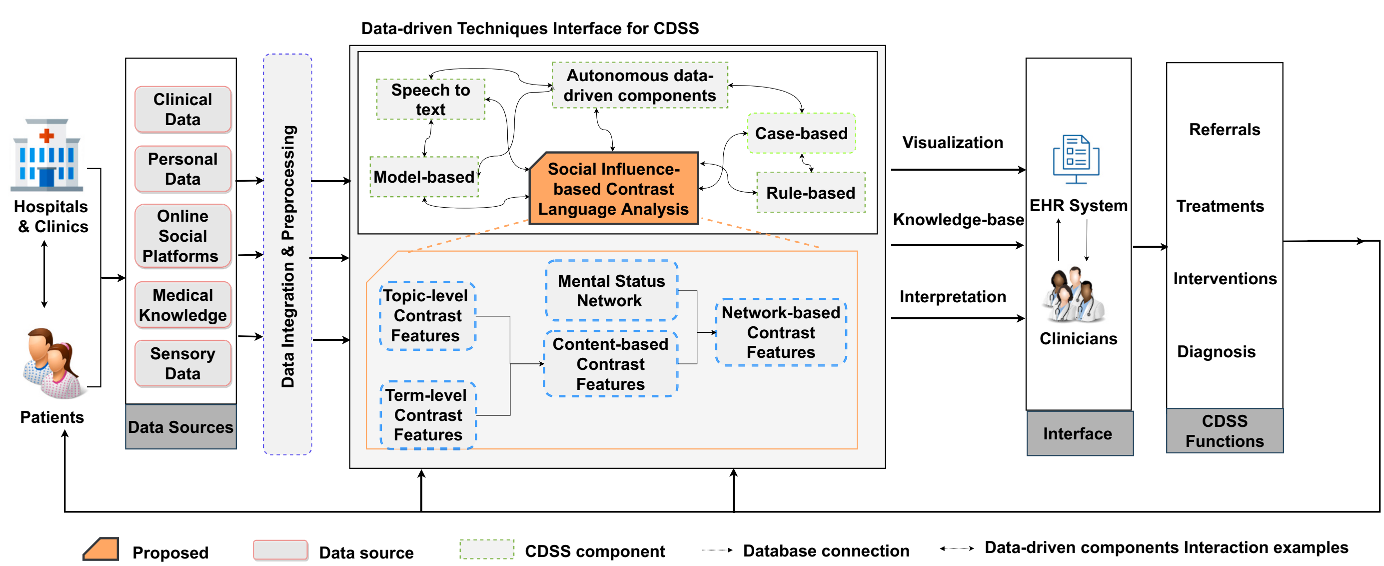



**Fig. 1.** Social influence-based contrast language analysis framework integrated with data-driven CDSS.

to maintain data quality and provide a reliable source of information [6]. The pre-processed information can be fed into different components that enable a CDSS to apply programmed or data-driven rules in order to generate output visualizations and interpretations of results. This can generate EHR for further analysis or serve as a knowledge base for clinicians in the form of recommendations, referrals, or decisions related to early interventions to avoid severe consequences of mental health disorder, high quality and interpret-able diagnosis to detect early signs and potential causes of depression, and personalized treatment plans for those who are already affected. All actions and decisions can then be readjusted by the output from the interface engine when new information is collected.

Our research is based on data collected from online social platforms, but the input data can be extended from many other sources and may have interactions with other data-driven components in CDSS as shown in Fig. 1. After extensive data preparation and cleaning procedures, we focus on extracting social influence-based contrast features obtained from the network structure and the content-based contrast features among depressed and non-depressed users. We obtain network-based contrast features to uncover influential contents (e.g., terms-level and topics-level contrast contents) of depressed and non-depressed users considering the network structure. To achieve this, first we need to extract content-based contrast features. We propose to use a novel procedure to discover topics that differentiate depressed from non-depressed users, called *topic-level contrast features*. We also propose a new measure to identify the most discriminant word features for depressed and non-depressed users, called *term-level contrast features*. It is important to note that the content-based contrast features are inputs of the network-based contrast features extraction.

### 3.1. Data preparation

We use the datasets used in Kosinski et al. [46] for illustration and evaluation of our proposed framework. The datasets contain responses from users filling in psychological research questionnaires and Facebook status updates. We use Facebook social network data only as a demonstration of user study. Our framework can be applied in other general network-based online and organizational discussion forums The CES-D scale is one of the most common approaches used by mental health professionals to diagnose depression. The CES-D scale is a 20-questions self-report scale that can measure a respondent's depression state [47]. CES-D includes questions related to how often a user encounters depressive associated symptoms such as feeling sad and lonely, and have a poor appetite and restless sleep. The participants can choose from 4 options with different scale of frequency ranging from *Rarely or None of the time (score = 0)* to *Most or Almost all the time (score = 3)* based on their mental mood in the past. The scale of the questionnaire ranges from 0 to 60. The higher the total score, the more likely the participant is diagnosed as depressed. We obtain Facebook users' responses to CES-D questionnaire from the dataset that measures a respondent's depression state. We calculate the score for Facebook users and create a binary label, namely, depressed and non-depressed, to represent two main user groups based on a threshold. We use a score of 25 to be our threshold of a participant to be considered as depressed because research suggests that the CES-D threshold of 25 and above provides adequate sensitivity and specificity for detecting major depressive disorder [48]. As a result, users whose score is above the threshold are labeled as depressed, and users whose score is below the threshold are labeled as non-depressed. For all the labeled users, we obtain their Facebook status updates and aggregate all the status updates per user using anonymous ID. During the data integration and prepossessing stage, we first aggregate all the status updates per user using anonymous ID. To maintain consistent contents, we run spell checker using SymSpell module [49]. We stripe HTML tags and remove special characters, and expand contracted words. We perform word lemmatization to bring all words to their respective bases. We applied stop words using python NLTK package. We keep personal pronouns as previous literature suggests that there is a positive association between mental wellness and the use of first person singular pronoun "I", and a negative association with the use of second and third person pronouns [39,50]. However, common words with low interpretation such as 'de', 'en', 'que', 'la', 'www', 'http' are eliminated.

### 3.2. Interactions with other data-driven components in a CDSS

Besides taking pre-processed inputs from online social platforms, our framework can interact with other data-driven components in CDSS. For example, it can interact with text data generated through speech recognition that can convert speech or sound from interviews into text data [51] where the text can be used as input in our framework to generate features. It can also interact with a case-based reasoning component in CDSS and combine knowledge discovery and case-based mechanisms to capture contextual information and generate classifications, or case-retrievals [52]. For example, after we find contrast features from different user groups, we can use the case-based engine to match new patients with the existing ones, and use the information to recommend solutions to the current problem or revise the proposed solutions for new cases. On the other hand, different cases found through the case-based engine can be fed into our framework to extract contrast features among different cases for better comparative representations. Our framework can be also integrated with a rule-based component in CDSS [53]. The rule-based engine can detect logical elements from our framework such as features and composite logic, which can be interpreted as rule functions in the rule engine to select or generate new features based on rules. The rule engine can suggest logical rules such as clinical constraints and criteria for our framework to generate contrast features that may lead to further actions by clinicians or hospitals. Furthermore, model-based components can interact with our framework. The features generated by our framework can be applied in the model-based engine which uses advanced predictive or deep learning models to capture complex relationships between input features and predictions [54]. We will present a case study of this interaction in our experiments. Finally, autonomous decision components can be integrated with our framework or other data-driven components to support such decision scenarios in CDSSs.

### 3.3. Social influence-based contrast language features discovery

In this section, we present how our framework extracts social influence-based contrast language features: term-level, topic-level and network-based contrast features. Term-level and topic-level contrast features are content-based contrast language features that differentiate depressed and non-depressed users based on different influential contents. The network-based contrast feature is a social influence-based contrast language feature that considers network structure from online social platforms and content-based contrast language features in the network. Below, we first present the proposed content-based contrast features and then illustrate how they are leveraged to extract network-based contrast features. The contrast language features are discovered from depressed and non-depressed user group to show the discrimination between these user groups. The contrast language features are very explainable and efficient on a two-classes comparison. The clinician can take advantage of the contrast languages depressed and non-depressed user group use to get more actionable insight and support targeted analysis for early depression diagnosis, intervention and design treatment plans based on different users' language behaviours.

#### 3.3.1. Content-based contrast features discovery

Our proposed framework incorporates content analysis as foundations to obtain social influence-based contrast language features to support decision making. Below, we propose two novel procedures to uncover influential content-based contrast features: *term-level* and *topic-level contrast features* to discover terms and topics that differentiate depressed from non-depressed users part of which are used to generate contents vectors to obtain the network-based contrast features.

*3.3.1.1. Term-level contrast features discovery.* Research shows that the content of a document can be illustrated by the frequency and proportion of specific words appearing in it [55]. Thus, we extract term-level contrast features where the term is an n-gram (e.g., uni-gram, bi-gram, tri-gram, etc.) extracted from users online contents. Moreover, we propose an effective approach to visualize such features for further analysis by clinicians. Fig. 2 outlines the proposed procedure.

*3.3.1.1.1. Extracting term-level contrast features.* We calculate the frequency of each n-gram and the relative term frequency of the n-gram for each group. The frequency for a specific term appears in generated contents can be used as information to grasp the ideas of the contents. If there is a difference in the term frequency of a specific n-gram in the depressed and non-depressed user group generated contents, then the term can indicate a distinct feature for distinguishing depressed and non-depressed users. This distinct feature discovered through term frequency shows the discrimination language used between these user groups and will provide informative and explainable insights for clinicians to find most significant features from different user groups and make corresponding diagnosis, intervention and treatment plan. Given generated contents (e.g., update statuses) of all users in both user groups (e.g., depressed and non-depressed), we first convert the contents into n-grams (e.g., uni-gram). We compute the frequency of each n-gram for each group. Then, we filter out n-grams whose frequency is more than a threshold (80% in our experiments as it reveals the more meaningful results) of the group size (e.g., number of records in the group). Furthermore, we remove low frequency n-grams as both low frequent and high frequent terms will not present discriminant information. Given the final list of n-grams for each user group, we define and calculate the relative term frequency of n-grams for each group as follows:

$$TF_g(t) = \frac{count_g(t)}{\sum_j count_g(j)} \tag{1}$$

where $count_g(t)$ is the weighted frequency of term $t$ in group $g$ (depressed $= d$, non-depressed $= nd$).

The above computation gives us a relative frequency of term $t$ in each group (i.e., $TF_d(t)$, $TF_{nd}(t)$). To extract the most discriminant terms, we rank them based on their ratio difference to compare the proportions of terms appearing in the depressed and non-depressed group. The ratio differences of term $t$ for depressed and non-depressed user groups are calculated as follows:

$$\Lambda_d(t) = TF_d(t) - TF_{nd}(t) \tag{2}$$

$$\Lambda_{nd}(t) = TF_{nd}(t) - TF_d(t) \tag{3}$$

where $\Lambda_d(t)$ represents ratio difference of a discriminant language term $t$ for depressed group and $\Lambda_{nd}(t)$ represents ratio difference of a discriminant language term $t$ for non-depressed group.

Given the the ratio differences of all terms for each user group, the terms are then sorted in the order of their statistical significance. We measure the significance using *p*-values proposed by [56]. A log-likelihood ratio statistic is calculated using chi-square distribution to assess the size and significance of a word's frequency of use in the two texts. We determine key terms in a corpus which differentiate one document from another and compare two groups to find statistically significant key differences between them. The lower the p-values (e.g., the larger the value of statistic) the the more significantly discriminant the key term between the two groups is. Given the sorted terms based on their p-values, we consider *top-k* terms from each group to be the term-level contrast features of the group. Given update statuses of user groups in our processed dataset, Table 2 shows that uni-grams are sorted in order of p-values. For k = 5, top-5 term-level contrast features are shown in this table. The words "im" (stands for"I am"), "text,""me","daughter" appeared relatively more frequent in the depressed group and represent its contrast term features. Conversely, the right side of Table 2 shows the term frequencies had been rearranged based on the term frequency of non-depressed group minus the term frequency of depressed group. We define these words that appeared relatively frequent in the non-depressed group as term-level contrast features of the depressed group.

*3.3.1.1.2. Term-level contrast features visualization.* A list of term-level contrast features might not provide enough insights or decision support for end-users. Therefore, we perform a cluster analysis to compare n-grams used by the depressed and non-depressed groups. To do so, top-ranked terms are transformed into vector space using a recent text embedding approach called Sentence-BERT (SBERT) [57] to compute meaningful sentence embeddings. This model transforms sentences or words combinations with similar meanings to a vector space in which their vectors are close to each other. SBERT model was pre-trained with SNLI [58] and the Multi-Genre NLI [59] datasets. All words from the input data go into the embedding layer where they get an embedding representation. The application of the SBERT results in fixed-size embeddings (768 dimensions) for each type of n-gram where the size is too big to visualize. Therefore, the principal component analysis (PCA) is employed, and the embedded vectors are reduced to the first two PCA elements. The two new components represent the two main dimensions of variations. Scatter plots are then used to visualize the results, with different colours representing the selected depressed and non-depressed n-grams. PCA has been applied to different embedding models including BERT embeddings to compute reduced dimensionality embeddings and generate visualizations [60,61]. Note that, although we use PCA to reduce dimension, the dimension reduction can be equally extended to other methods such as TSNE [62] and UMAP [63]. The visualization allows determining whether the two-component representation of n-grams results in clusters, as well as the relative positions of n-grams and the distances between them. We will discuss our findings in the experiment section.

*3.3.1.2. Topic-level contrast features discovery.* Another critical component of a language is the topics discussed by users in their online social platforms. In addition to term-level contrast features, we argue that the topic-level contrast features can provide a deeper insight into a user's language in online social platforms.Topic modeling in online social platforms has been conducted in the past for depression behavior analysis [8,30,37]. A straightforward approach to use topic modeling for depression analysis is to build the topic models on the content and generate the topic distribution vectors for depressed and non-depressed user groups. Then, given a new user, the similarity of his topic vector to the depressed and non-depressed vectors is calculated to recognize if the user is prone to depression. However, this approach is generalized and only a vector generated by the general topic model is used to represent each group. Furthermore, this approach is not able to indicate the most discriminant topics between user groups effectively. We present a novel approach to extract contrast topics presented by depressed and non-depressed user groups. Such topics reveal the main differences between the two groups in terms of the topics of discussions in a online social platform. This contrast topics discovered from each user group help to identify the distinct contents discussed from different user group which further provide a deeper understanding of mental health and interpretable knowledge to clinicians to detect early signs and potential causes of depression, apply appropriate early intervention for users at risk and treatment plans for those already affected. Fig. 3 illustrates the procedure to discover topic-level contrast features. We first build our topics using LDA [64] to identify the topics that appear in the two groups of users. LDA represents the status updates as a mixture of topics, and each topic is represented as a mixture of words with certain probabilities. We split the input corpus of status updates into depressed and non-depressed datasets. Then, we run LDA to build the topic model for each dataset. One of the main input parameters of LDA is the number of topics. To find the optimum number of topics, we use the well-known measure of *coherence*. The coherence of a topic model measures the degree of similarity exhibited among words in a topic. The coherence score helps distinguish between topics that are semantically interpretable [65]. We use *Mallet* by McCallum [66] to run LDA for a varying number of topics, and measure the coherence scores for all the topic models built using LDA. The model with the highest coherence score

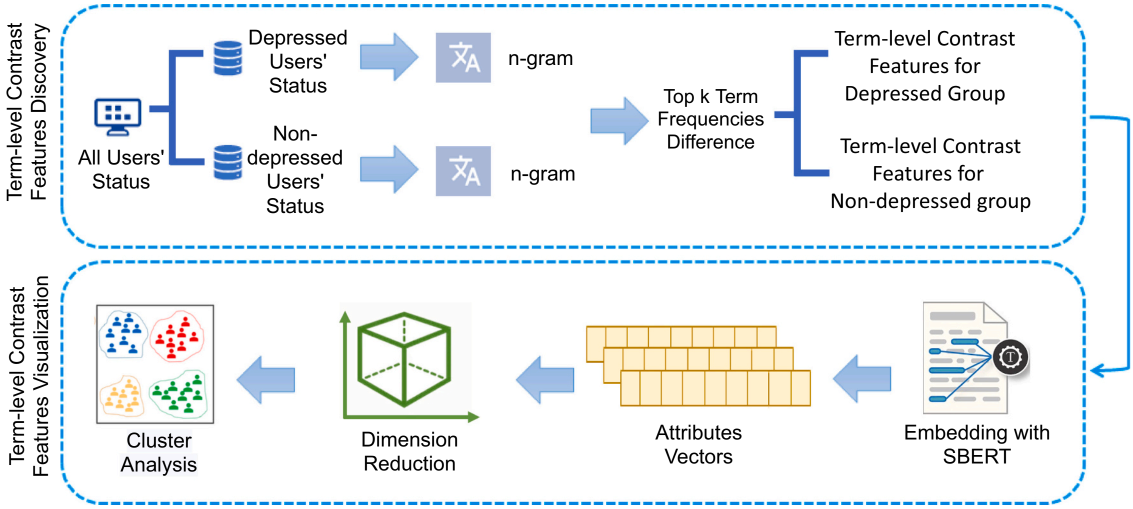



**Fig. 2.** Term-based contrast features extraction process.

**Table 2**
Contrast n-grams.

| | Term-level contrast features for Depressed Group | | | | | | Term-level contrast features for Non-depressed Group | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Term Frequency Ratio (TF) | | | | | | Term Frequency Ratio (TF) | | | |
| Uni-gram | Depressed | Non-depressed | $\Lambda_d(t)$ | P-value Rank | | Uni-gram | Depressed | Non-depressed | $\Lambda_{nd}(t)$ | P-value Rank |
| Im | 0.15% | 0.07% | 0.08% | 1 | | forward | 0.04% | 0.08% | 0.04% | 1 |
| Text | 0.11% | 0.06% | 0.05% | 2 | | snow | 0.08% | 0.14% | 0.06% | 2 |
| | 0.12% | 0.08% | 0.04% | 3 | | class | 0.12% | 0.18% | 0.06% | 3 |
| Me | 2.34% | 2.16% | 0.18% | 4 | | week | 0.24% | 0.34% | 0.10% | 4 |
| Daughter | 0.05% | 0.03% | 0.02% | 5 | | today | 0.57% | 0.73% | 0.16% | 5 |

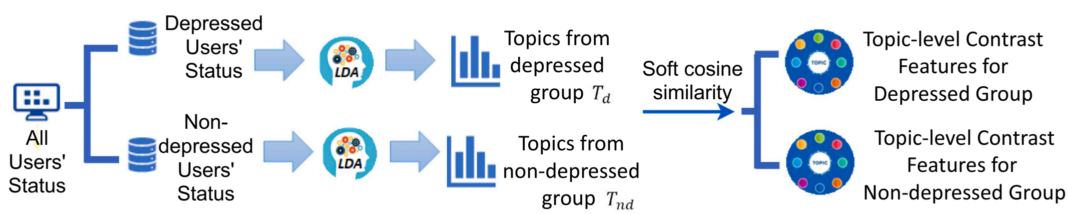



**Fig. 3.** Topic-level contrast features discovery procedure.

before a major drop in the coherence is chosen as the best model, and the corresponding number of topics is selected to be the optimum number of topics.

Let $T_d = \{t_d^1, t_d^2, \dots, t_d^m\}$ be the topics of depressed user group and $T_{nd} = \{t_{nd}^1, t_{nd}^2, \dots, t_{nd}^n\}$ be the topics of non-depressed user group, where $t_g^i$ is a topic vector representing $i^{th}$ topic for user group $g$. To find the topic-level contrast features, we calculate the similarity of topics between the two models. We use *soft cosine similarity* to compute the similarity between each depressed and non-depressed topic pair. Note that there is a difference between traditional cosine similarity and soft cosine similarity. Traditional cosine similarity characterizes "hello" and "hi" as zero similarity because there are no common words, while soft cosine similarity would consider "hello" and "hi" as similar since they are synonymous. Soft cosine similarity generalizes the concept of cosine similarity and considers the semantic meaning of words, and words with similar meanings are treated as similar. To find topic-level contrast features, we run LDA independently on each user group dataset because the topic models' vocabulary is not the same. Thus, cosine similarity is not applicable here. Soft cosine similarity is chosen to measure the similarity between two collections of words. Each topic is represented as a vector with words, and a matrix is used to calculate the similarity between the features [67]. Assume $t_d^k$ and $t_{nd}^l$ are topic vectors extracted from depressed and non-depressed contents, respectively. Given two topic vectors $t_d^k$ and $t_{nd}^l$, the soft cosine similarity is calculated as follows:

$$Soft-Cosine\left(t_d^k, t_{nd}^l\right) = \frac{\sum \sum_{i,j=1}^{V} s_{ij} t_d^k(i) t_{nd}^l(j)}{\sqrt{\sum \sum_{i,j=1}^{V} s_{ij} t_d^k(i) t_d^k(j)} \sqrt{\sum \sum_{i,j=1}^{V} s_{ij} t_{nd}^l(i) t_{nd}^l(j)}} \tag{4}$$

where $i$ and $j$ represent the $i^{th}$ and $j^{th}$ word in a topics and $s_{ij}$ is a matrix based on Levenshtein distance [67].

Assume that $t_d^k$ and $t_{nd}^l$ are two topic vectors extracted from depressed and non-depressed groups, respectively. The soft similarity is computed through the matrix $s_{ij}$ which indicates the similarity of words between these two vectors of topics. Each topic is a vector of words, each word in topic $t_d^k$ and each word in topic $t_{nd}^l$ are compared to each other using a levenshtein distance. Levenshtein distance is used as a metric to quantify the similarity between words [68]. It is defined as the minimum number of edits needed to transform one word $a$ into another word $b$ (e.g., insert, delete or replace a character in word $a$). For example, to calculate the levenshtein distance between two words peel (e.g., word $a$) and hello (e.g., word $b$), we first replace"p" with"h" for word peel (peel→ heel), then we delete"e" (heel → hel), we next insert an"o" (hel → helo), and a"l" (helo → hello). As a result, the distance is four as four operations have been applied. Then, the levenshtein distance for word $a$ and $b$ can be calculated by using a matrix $s_{ij}$ that computes the similarity of these two words.

For each pair of topics between depressed and non-depressed groups, we calculate the soft-cosine similarity score. Given the similarity scores, a possible approach is to choose the topics with the lowest score as topic-level contrast features. However, this approach fails as the most dissimilar topics are not necessarily in the same context and do not provide clinicians with valuable knowledge to compare depressed and non-depressed topics of discussion. Therefore, we first discover topics of depressed and non-depressed groups in the similar context, and then the topics that are the most discriminant topics in the same context are considered as topic-level contrast features for each group. We compute a $\mu_{td^m}$ to represent the soft cosine similarity value for topic $t_d^m$ and its most similar topic in group $nd$ (non-depressed). For topic $t_d^m$, we compute all the soft cosine similarity values between topic $t_d^m$ and each non-depressed topic and find a non-depressed topic that is the most similar to $t_d^m$. $\mu_{td^m}$ is the cosine similarity values that show the highest values between $t_d^m$ and the found non-depressed topic. Similarly, we compute a $\mu_{tnd^n}$ to represent the soft cosine similarity value for topic $t_{nd}^n$ and its most similar topic in group $d$ (depressed). Given topics in each group, the calculations are done as follows. we first find the most similar topics to each topic in the other group as follows:

$$\forall t_d^m \in T_d, \mu_{t_d^m} = max\left(Soft - Cosine\left(t_d^m, t_{nd}^n\right)\right), \forall t_{nd}^n \in T_{nd} \tag{5}$$

$$\forall t_{nd}^n \in T_{nd}, \mu_{t_{nd}^n} = max\left(Soft - Cosine\left(t_{nd}^n, t_d^m\right)\right), \forall t_d^m \in T_d \tag{6}$$

Next, the topic-level contrast features (i.e., $\tau$) are selected by finding topics with the minimum $\mu$ values in each group. In Eq.7, $\tau_d$ represents topics with the minimum $\mu_{td^m}$ values for depressed group, and in Eq. 8, $\tau_{nd}$ represents topics with the minimum $\mu_{tnd^n}$ values for non-depressed group:

$$\tau_d := \left\{\forall t_d^m \in T_d, k \text{ topics with smallest } \mu_{t_d^m}\right\} \tag{7}$$

$$\tau_{nd} := \left\{\forall t_{nd}^n \in T_{nd}, k \text{ topics with smallest } \mu_{t_{nd}^n}\right\} \tag{8}$$

Fig. 4 shows an example of how we discover topic-level contrast features. For brevity, we present four topic vectors for depressed users and three topics for non-depressed user group. The top-3 words per topic are also presented. Given the topic models, we first calculate the maximum cosine value based on all the calculated soft cosine similarities. Then, top-k (k = 3) topics are selected per group as contrast features.

Note that, Rayson and Garside [56] discovers keywords in the corpora which differentiate one corpus from another by frequency, while our method uses frequency ratio for term-level features, and also explores how topics-level features differentiate one corpus from another. Moreover, Rayson and Garside [56] calculates log-likelihood values and only finds one set of keywords. In contrast, our method finds two sets of keywords for depressed and non-depressed groups separately for both term-level and topic-level features.

### 3.3.2. Network-based contrast features

Several studies show that an individual's mental states are influenced directly or indirectly through family and friends or colleagues [69]. Social influence is primarily captured through the lens of social network analysis where nodes (users, entities) are influenced by others for various reasons and social contexts (e.g., family, friends, colleagues, social groups, etc.). Tang et al. [70] examine how to differentiate the social influences from different angles (topics) and quantify the strength of those social influences on real large networks such as Facebook or research collaboration networks. Extending the focus on macro-level models such as degree distributions, diameter, clustering coefficient, etc., they argue that a user's influence on others not only depends on their own topic distribution, but also the kinds of social relationships they have with others. Consequently, they present a method that leverages both the local attributes (topic distribution) and the global structure (network information) for social influence analysis. Grounded on the theory of factor graph [71], their method primarily leverages an affinity propagation at the topic level for social influence identification where the observation data are cohesive on both local attributes and relationships. In a Twitter setting, Zhang et al. [72] also study how users' behaviours are influenced by friends in their ego network. Examining retweet behaviours in a large microblogging network, they formulate social influence locality in the social network by distinguishing local patterns (pairwise influence and structure influence) from global patterns (e.g., influence from an opinion leader and influence by a hot topic). Their formalization of social influence locality is based on instantiation functions on pairwise influence and structural diversity, as well as the assumption that users' behaviours are mainly influenced by close friends in their ego networks. In this section, we aim at taking general network structure into account to discover the social influence-level contrast features represented by ***network-based contrast features***. Below, we first define the network structure used in this study.

**Definition 1.** A Mental Status Network $MSN = (U, C)$ is a graph where the nodes (i.e., U) are users and the connections represent how each user is connected to others whose depression status are determined through CES-D scale survey.

Given an MSN, Fig. 5 shows the process of discovering network-based contrast features (i.e., NBC). Given content-based contrast features (term-level features and topic-level contrast features) which are discussed in Section 3.3.1 for depressed and non-depressed groups respectively, we infer the content features vector for each user with respect to each model. As result, we have two content vectors, denoted by $V_{u,\ d}$ and $V_{u,\ nd}$, for each user $u$ in which the elements are the content features from UGC with respect to each user group $g$ (depressed $= d$, non-depressed $= nd$). We argue that the contents that depressed or non-depressed users discussed in a MSN have a social influence on other users via direct and non-direct connections. The closer the connection, the more social influence the contents would have on the target user, and the more similar the contents of two nearby users, the more likely one user gets influenced by the other. To reflect such impacts, we first compute a social influence score for each user, representing the social influence of each neighbor on the user. The social influence scores and the content vectors are then used to calculate cosine similarities of a user's influential contents with those neighbors belonging to depressed and non-depressed user groups in MSN. As the result of the similarity analysis, two contrast features are extracted that represent the degrees of differences between influential contents a user discusses and those discussed by depressed and non-depressed user groups. Below, we first present the proposed procedure with an example for a particular user, thus the same procedure can be applied to calculate the features for all users in a MSN.

Fig. 6 shows a partial network. Assume that we would like to calculate the network-based contrast features for Lucy. Lucy has direct social connections with Bob and Alice, second level social connections with Jim, and a third level social connection Ava. Given the types of the social connections, we define the distance between Lucy and her first neighbors (e.g., Bob and Alice) to be 1, the distance between Lucy and her second level neighbors (e.g., Jim) to be 2, the distance between Lucy and her third level social connection (e.g., Ava) to be 3, and so on. The closer the friend is, the more social influence the friend might have on Lucy. Therefore, the activities of Bob and Alice might have a higher impact on Lucy compared to Jim. Most likely, Ava has the least social influence as she is a third level social connection. This means that we

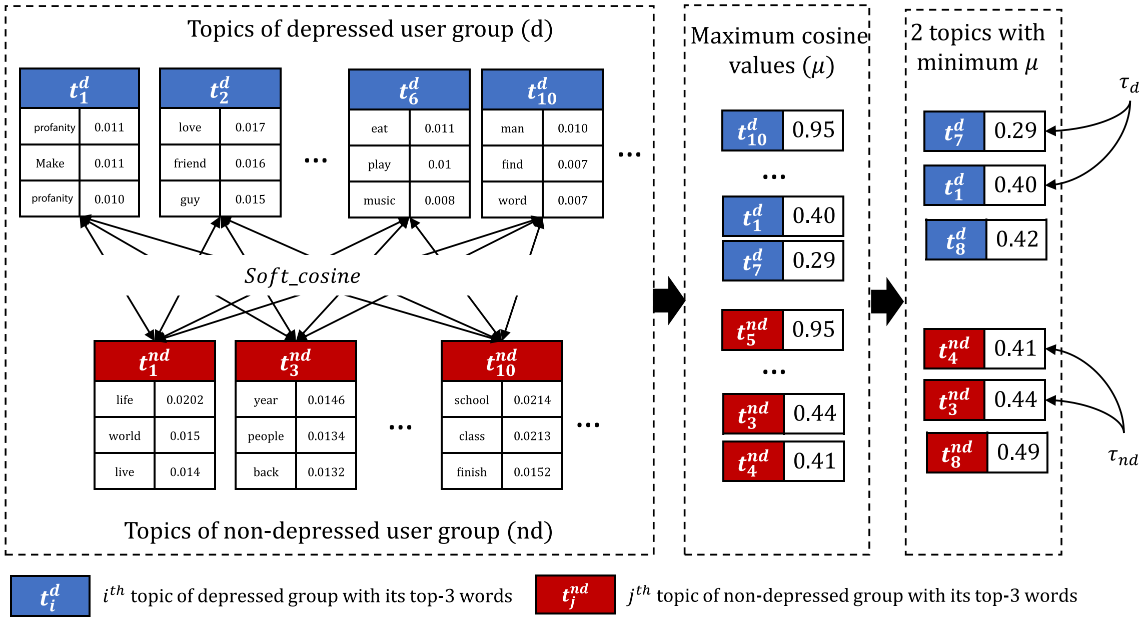



**Fig. 4.** An example of topic-level contrast features discovery.

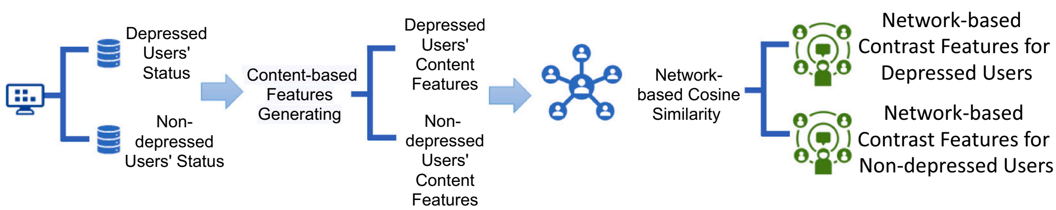



**Fig. 5.** Network-based contrast features discovery.

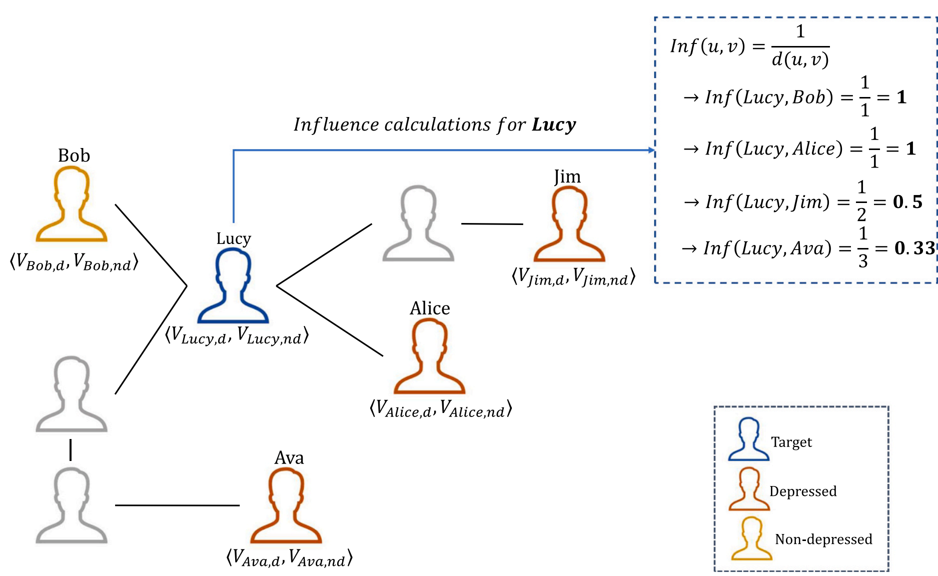



**Fig. 6.** An example of network-based contrast features discovery. $\langle V_{usr,\ d}, V_{usr,\ nd} \rangle$ represent contents features vectors of user *usr* generated using depressed and non-depressed contents features, respectively.

should assign a higher weight to Bob and Alice compared to Jim and Ava to reflect their social influence using a reciprocal of the distance between Lucy and her neighbors. To calculate the impact of neighbor $v$ on user $u$, we define *social influence* score as follows:

$$Inf(u,v) = \frac{1}{d(u,v)} \tag{9}$$

where $d(u,v)$ is the number of steps to reach user $v$ from user $u$ in the network.

Note that, since there are different paths to reach a user with confirmed mental status (i.e., depressed or non-depressed user), we consider the shortest path to calculate the distance. In Fig. 6, we calculate the social influence scores of Bob, Alice, Jim, and Ava on Lucy.

Now that we have the social influence scores from user's friends, which form a influence matrix for the user (e.g., Lucy) shown in Fig. 6. Then, we calculate the cosine similarity of content vectors of users. Let $V_{u,\ d}$ and $V_{v,\ d}$ be the content vectors generated by content features discovery from depressed group (i.e.,term-level and topic-level contrast features) for users $u$ and $v$. We calculate the similarity of the contents for the depressed group and non-depressed group:

$$\delta_d(u,v) = \frac{V_{u,d}\cdot V_{v,d}}{\sqrt{V_{u,d}\cdot V_{u,d}}\sqrt{V_{v,d}\cdot V_{v,d}}} \tag{10}$$

$$\delta_{nd}(u,v) = \frac{V_{u,nd}\cdot V_{v,nd}}{\sqrt{V_{u,nd}\cdot V_{u,nd}}\sqrt{V_{v,nd}\cdot V_{v,nd}}} \tag{11}$$

$\delta_d$ and $\delta_{nd}$ are calculated for all the users with the confirmed mental status. Then, we calculate the network-based contrast features by combining the similarity and social influence scores:

$$NBC_d(u,\Phi_u) = \frac{\sum_{v\in\Phi_u} \lambda_v Inf(u,v)\delta_d(u,v)}{\sum_{v\in\Phi_u} Inf(u,v)\delta_d(u,v)} \tag{12}$$

$$NBC_{nd}(u,\Phi_u) = \frac{\sum_{v\in\Phi_u} \beta_v Inf(u,v)\delta_{nd}(u,v)}{\sum_{v\in\Phi_u} Inf(u,v)\delta_{nd}(u,v)} \tag{13}$$

where $\Phi_u$ is the neighbors of user $u$ and $\lambda_v = 1$ if neighbor $v$ is a depressed user, otherwise $\lambda_v = 0$; And $\beta_v = 1$ if neighbor $v$ is a non-depressed user, otherwise $\beta_v = 0$.

$NBC_d(u,\Phi_u)$ and $NBC_{nd}(u,\Phi_u)$ are network-based contrast features that examine impact of network-based contrast contents from depressed and non-depressed users in the mental status network. The higher the values of $NBC_d(u,\Phi_u)$ and $NBC_{nd}(u,\Phi_u)$, the more influence the depressed and non-depressed friends in MSN has on the target user. Algorithm 1 illustrates the pseudo code of the proposed method.

### 3.4. Downstream tasks: CDSS functionality and enhanced decision-making

Our comprehensive analysis framework offers tremendous potential for a CDSS environment to enhance clinical decision making for three main downstream tasks, early depression detection, early intervention, and personalized treatment plans as outlined below: **1) Early depression detection**: Studies show that users with signs of depression have different language styles, written contents and word usage [37,41]. By analyzing social influence-based contrast language features and better understanding of the differences between depressed and non-depressed user groups, a new set of features can be captured to build a high-quality decision support tool that improves the effectiveness of diagnosis, and more importantly detects early depression signs. **2) Early intervention**: Early intervention plans are one of the most effective processes to avoid severe consequences of mental health disorders. Considering the low quality of data provided by questionnaires, early intervention is extremely challenging as clinicians should spend a lot of time to understand an individual's symptoms and characteristics. Contrast language features profile a comparative representation of language style, words used and contents of depressed versus non-depressed users. Thus, if an individual develops language features similar to depressed users, this individual might be at risk of depression. Our framework can enhance the features and functionalities as well as effectiveness of a CDSS by identifying the behaviours and language patterns that can signal early intervention and behavior adjustment for high-risk people. **3) Personalized treatment plan:** Psychotherapy is one common depression treatment. This involves lengthy sessions between individuals and psychiatrists to learn about issues and suggest plans or medications for patients to transition into better mental states. Contrast language features can be a rich source of clinical decision support materials to develop customized treatment plans for depressed users, informed by the attitudes, topics and activities from this user group.

## 4. Exploratory analysis of social influence-based contrast features

In this section, we present the results of a user study we conducted to evaluate our social influence-based contrast language features discovery framework. We use 197,230 Facebook status updates from 1047 Facebook users and their associated CES-D questionnaires scores as a demonstration for our framework. Our framework can be applied in any network that involves online social interactions. The data contains all the status updates of 569 depressed and 478 non-depressed users who are labeled based on the threshold of 25 on CES-D scale. We investigate

**Algorithm 1** Network-based Contrast Features

**Input: Mental Status Network, Users' Content Vectors**
**Output: Network-based Contrast Features for all users** ($NBC_d$, $NBC_{nd}$)

$NBC_d \leftarrow \emptyset$ ▷ NBC feature based on the depressed group
$NBC_{nd} \leftarrow \emptyset$ ▷ NBC feature based on the non-depressed group
**for** $\forall\ u \in$ Mental Status Network **do**
  $\Delta_{u,d} \leftarrow \emptyset$ ▷ Depressed contents similarity of user $u$ and his neighbors
  $\Delta_{u,nd} \leftarrow \emptyset$ ▷ Non-depressed contents similarity of user u and his neighbors
  $Inf_u \leftarrow \emptyset$ ▷ Influence scores for user $u$
  **for** $\forall v \in \Phi_u$ **do**
    Calculate $\delta_d(u,v)$, and add $\delta_d(u,v)$ to $\Delta_{u,d}$
    Calculate $\delta_{nd}(u,v)$ and add $\delta_{nd}(u,v)$ to $\Delta_{u,nd}$
    Calculate $Inf(u,v)$ and add $Inf(u,v)$ to $Inf_u$
  **end for**
  Calculate $NBC_d(u,\Phi_u)$ and add to $NBC_d$
  Calculate $NBC_{nd}(u,\Phi_u)$ and add to $NBC_{nd}$
**end for**
**return** $NBC_d, NBC_{nd}$

**Table 3**
The values of network-based contrast features for depressed and non-depressed users example.

| Users | Depression Status | $NBC_d$ features based on depressed group | $NBC_{nd}$ features based on non-depressed group |
|---|---|---|---|
| $u_1$ | Depressed | 0.759 | 0.098 |
| $u_2$ | Depressed | 0.680 | 0.257 |
| $u_3$ | Non-depressed | 0.172 | 0.776 |
| $u_4$ | Non-depressed | 0.186 | 0.776 |

the interpretability of the proposed social influence-based contrast features that involve term-level, topic-level and network-based contrast features. We show the step-by-step analysis of extracting social influence-based contrast features and what they look like in practice. The privacy of our data is well protected following the proposed mechanisms in [17].

### 4.1. Exploratory analysis of network-based contrast features

We use the content-based contrast features for both depressed and non-depressed group and the MSN structure to obtain network-based contrast features. Table 3 presents an example of network-based contrast features from the depressed and non-depressed group in a MSN. According to our labeled data, $u_1$ and $u_2$ are depressed users. Both users' $NBC_d(u, \Phi_u)$ are higher than $NBC_{nd}(u, \Phi_u)$. This indicates that for depressed user group, the content-level[1] impact from their depressed friends is larger than non-depressed friends in their MSN. The contents that $u_1$ and $u_2$ talk about are more similar to their depressed friends and less similar to their non-depressed friends. Likewise, for non-depressed users $u_3$ and $u_4$, both of their $NBC_{nd}(u, \Phi_u)$ is larger than $NBC_d(u, \Phi_u)$. This shows that $u_3$ and $u_4$ are more likely to be affected by their non-depressed friends than depressed friends. The contents that $u_3$ and $u_4$ talk about are more likely similar to their non-depressed friends than depressed friends. This illustrate that depressed and non-depressed friends in the MSN have contrast impact on its users. We plot the values of network-based contrast features from depressed and non-depressed users in Fig. 7 with depressed user and non-depressed user network-based contrast features on the horizontal and the vertical axis, respectively. The plot shows a downward trend: as the value of depressed user network-based contrast features increases, the value of non-depressed user network-based contrast features decreases. This shows depressed and non-depressed user groups in the network have a different impact on their friends. The friends who are impacted more by the depressed user group tend to cover contents more similar to the depressed user group and vice versa. This figure asserts the fact that the distance of users in a mental status social network can be investigated to measure the mental health of a user.

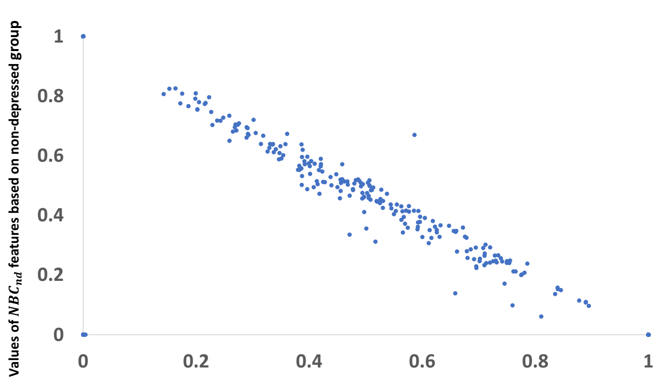



**Fig. 7.** The values of network-based contrast features from depressed and non-depressed groups.

### 4.2. Exploratory analysis of content-based contrast features

#### 4.2.1. Term-level contrast features

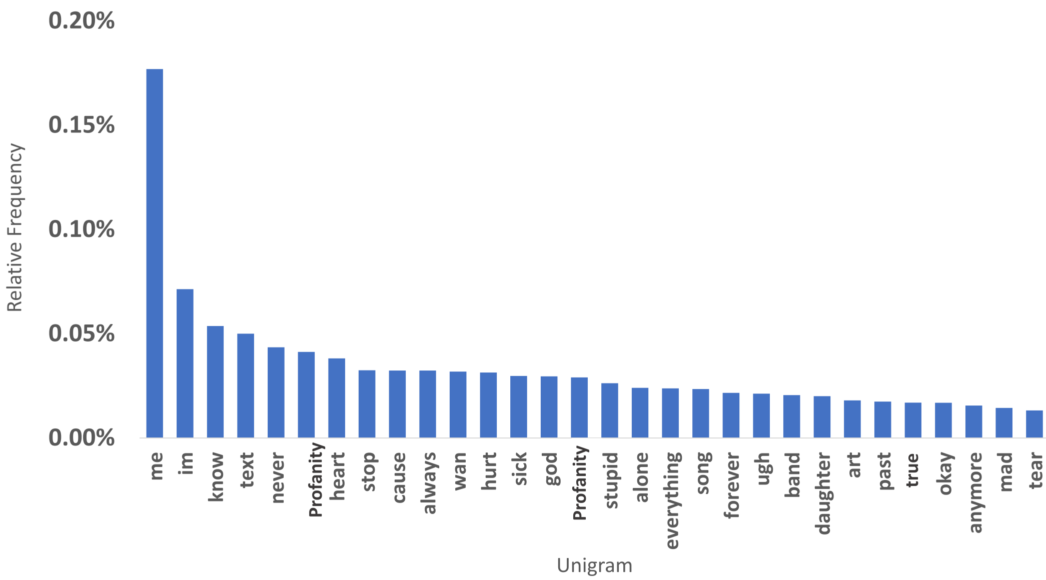
Following the proposed procedure in Section 3.3.1, we select statistically significant 30 term-level contrast features for uni-grams from each depressed and non-depressed group based on *p*-values rankings during the contrast feature discovery. This can be extended to bi-grams, tri-gram and etc. The frequency distributions of all uni-grams is also statistically significant, and categorized as depressed and non-depressed groups. The outputs are shown in Fig. 8. The higher the term frequency difference, the more discriminant the terms are. According to Fig. 8 (A), the depressed user group often uses first person pronouns "I'm" and"me" and terms with negative sentiments such as "alone", "hurt", "sick", and profanity[2] which is consistent with findings in previous literature [39]. On the other hand, Fig. 8(b) shows the distribution of terms for the non-depressed user group in which they are more likely to use terms referred to events such as"test","class","snow" and"work"; and time, which includes"forward ","week","today","day","weekend","hour","last","morning" and etc.

The terms representing the contrast features that we extract from update status texts are visualized using the proposed approach in Section 3.3.1. Fig. 9 shows the visualization. In this figure, the depressed term-level contrast features are presented by red triangles, while non-depressed term-level contrast features are presented by blue circles. As shown in Fig. 9, the majority of term-level contrast features for the depressed user group are located on the left side of Fig. 9, while the term-level contrast features for the non-depressed group are mostly located on the right.

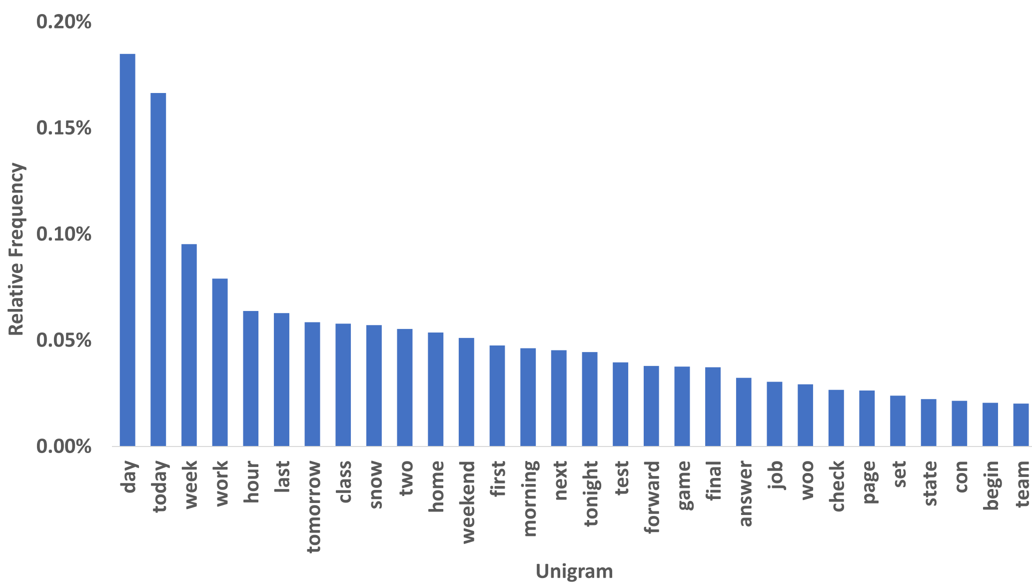
Note that, given the context aware embedding approach we use to represent the terms in embedding space, terms that are close to each other are usually used together in similar contexts; thus, we categorize these words into specific groups through cluster analysis as seen in Fig. 9. Our clustering shows that there are clear term-level contrast feature differences between depressed and non-depressed user groups. Generally speaking, depressed and non-depressed user groups tend to express themselves with different language in term-level. The upper left side of Fig. 9 shows the words related to negation, such as"cant","never","stop", and negative feelings, such as"sick","alone","mad","hurt","tear" are used frequently by depressed user group. The upper right side of Fig. 9 shows non-depressed user group tends to use words related to the moment in time, such as"weekend","tomorrow","tonight","morning","hour","day", "today","last" and place or event such as"job","work","game","class" and"home". In the middle of Fig. 9, we find depressed user group tends to talk more about existence in indefinite time or amount such as"anymore","always","past","god". In the bottom of Fig. 9, first person pronoun such as"im" and"me" are found to appear frequently in depressed user group. Depressed user group also uses words"true","okay","know" to show they agree or compromise with something. In the right bottom side, we can see that verbs, numbers or forward looking words such as"test","set","answer","fist","next","begin","forward", etc. are used

[1] In our study, we infer topic distribution vectors given on topic models $T_d$ and $T_{nd}$. We focus on using each user $u's$ topic distribution with respect to each topic model as inputs for content-level feature vectors $V_{u,\ d}$ and $V_{u,\ nd}$ to obtain network-based contrast features for each user. Content feature vectors can also be extended use individual's term-level contrast features as influential contents.

[2] Due to the sensitivity of keywords from the depressed group, we have avoided listing them here.




(a) Term-level contrast features of depressed group




(b) Term-level contrast features of non-depressed group

**Fig. 8.** Term-level contrast feature distributions based on term frequency difference ratio.

frequently by non-depressed users.

#### 4.2.2. Topic-level contrast features

By applying the procedure outlined in Section 3.3.1, we identify topic-level contrast features for depressed and non-depressed user groups separately. The first step of the proposed model is to build topic models for each user group corpus. To find the best number of topics in each corpus, we run LDA models for each group dataset by varying number of topics and calculating coherence score for the topic models built by LDA. Fig. 10 shows the coherence scores for topics models with different number of topics. We notice that topic models with number of topics more than 15 do not reveal a better score. According to Fig. 10(a), the optimum number of topics is determined to be 10 where the coherence value is 0.28. Furthermore, we choose 10 as the number of topics for non-depressed user group as it yields the best coherence score before a significant drop as shown in 10(b). As the first exploratory analysis, we focus on top-10 words in each topic's distribution. These words are effective representation of a topic. Table 4 presents the common words among the topics of the two user groups and the words unique to depressed and non-depressed group topics. The depressed topics contain multiple negative words in terms of sentiment, including swear words[3] as well as emotional words such as"ill","fear", and"tired". The non-depressed group terms do not contain as many words as depressed-group with negative sentiment. Given the best topic model for each group, we then use soft-cosine similarity to find the similarities between depressed and non-depressed group topics. We discover three topics from each group that are the most discriminant topics in the same context to be considered as topic-level contrast features for depressed

[3] Due to the sensitivity of keywords from the depressed group, we have avoided listing them here.

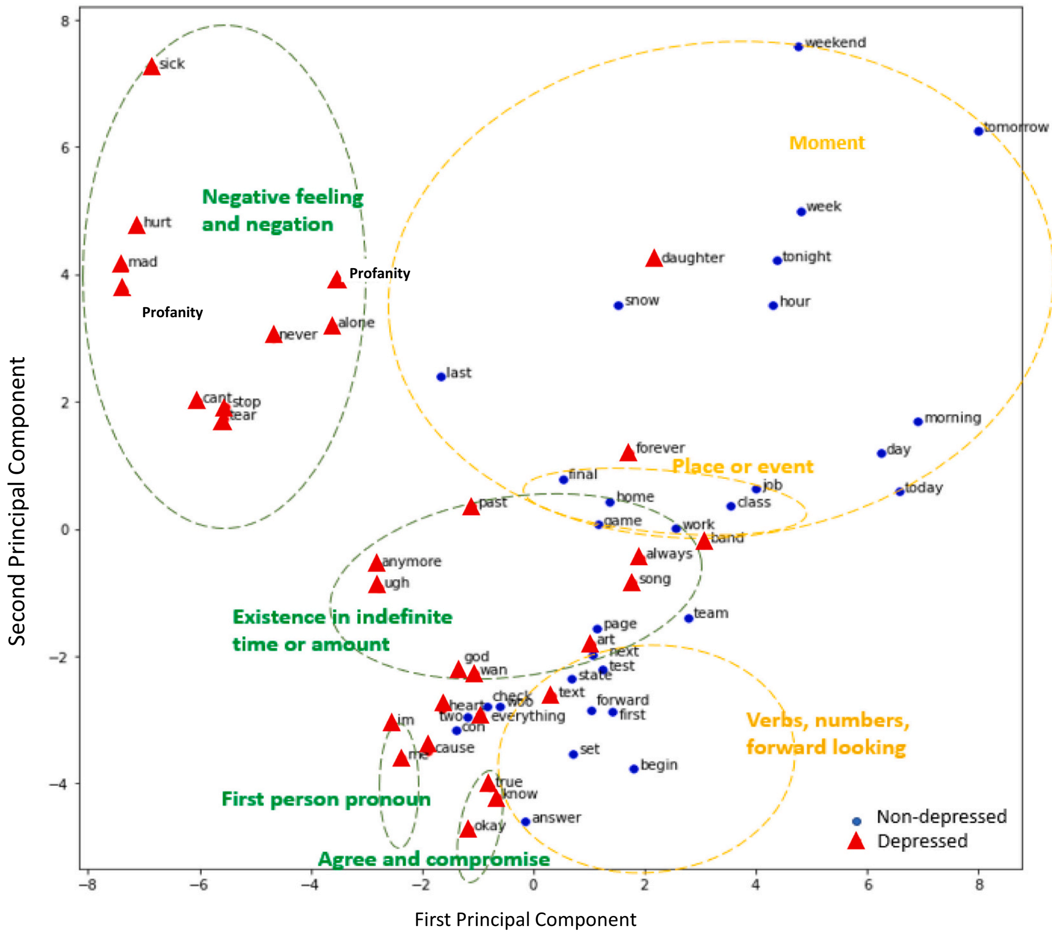



**Fig. 9.** Cluster analysis of term-level contrast features.

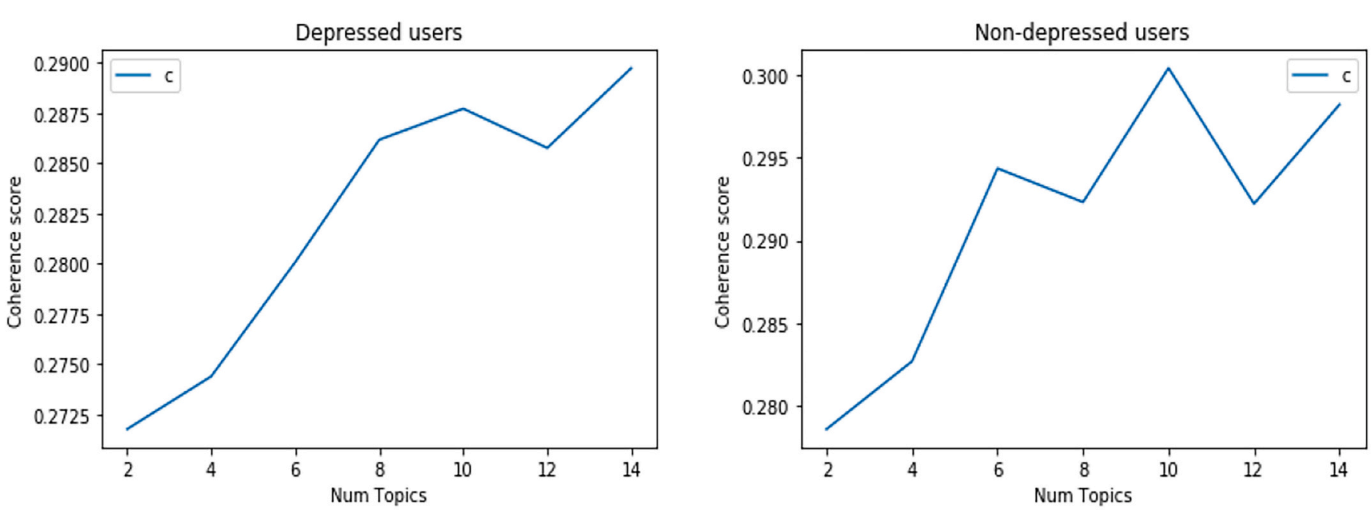



**Fig. 10.** Coherence scores of topic models for depressed and non-depressed groups.

and non-depressed group, respectively.

Fig. 11 shows the top-10 words in the topic-level contrast features for each user group, respectively. Interestingly, a significant contrast is clear among topics of the two user groups. The tone, sentiment and context of the topics are different between users with signs of depression versus users without the signs. The selected topic-level contrast features for depressed group contain multiple terms with negative sentiments, such as 'hate', and 'fear' while those of non-depressed group do not contain such terms. Note that, given a new user update status, we can calculate the probability of each of these topics using LDA. The probabilities represent the importance of the topic in a user's update status. Then, they can be used as indicators of depression signs.

**Table 4**
Topics' terms comparison of depressed and non-depressed group.

| Depressed vs. Non-depressed topics words | | |
|---|---|---|
| Common words | Unique depressed words | Unique non-depressed words |
| Bore, back, post, homework, guy, leave, play, night, give, happy, thing, work, home, love, week, status, mom, friend, life, man, day, put, tomorrow, class, world, read, live, time, hate, feel, make, watch, sleep, find, today, school, good, music, awesome, year, finish, eat, people, word | Birthday, call, dream, eye, family, fear, great, head, heart, ill, money, movie, power, snow, sound, text, tired | Bad, bed, big, bit, check, drink, end, final, fun, game, hour, job, kid, listen, mind, pay, start, test, tonight, turn, write |

### 4.3. User study: early depression detection in mental status networks

The proposed contrast feature discovery framework can be the foundation for many different tasks in mental health domain such as (early) diagnosis, treatment and intervention. We design an early depression diagnosis user study to investigate the effectiveness of the proposed framework using ML algorithms. ML methods have been used in CDSS to gain knowledge from data to support decisions [5,6,73]. The proposed features can be considered as input features for any supervised ML algorithm to diagnose early depression. This approach is a common practice for detecting depression or early depression. However, the majority of existing studies focus on the features extracted directly from the data and do not consider contrast features which we believe are of paramount importance in diagnosing (early) depression. To do so, we use four different supervised ML algorithms: 1) Logistic Regression (**LR**), 2) K-Nearest Neighbors (**KNN**), 3) Neural Network (**NN**), 4) Classification and Regression Trees (**CART**). Note that, other supervised learning algorithms that take similar features as input can be used here. We follow the literature in selecting these four methods [8,9,16]. We run experiments with social influence-based contrast language features extracted from our proposed contrast language framework with the components of term-level contrast features, topic-level contrast features and network-based contrast features to predict early depression with the four different ML methods. We train our model by randomly selecting 80% of data and test on 20% of data. We compare the performance of the models using the following performance measures:

1. **Accuracy** $= \frac{TP+TN}{TP+TN+FP+FN}$
2. **Precision** $= \frac{TP}{TP+FP}$: ratio of number of depressed users correctly classified to depression class and the total number of depressed users.
3. **Recall** $= \frac{TP}{TP+FN}$: rate of depressed users correctly classified to be in depression class
4. **F_Score** $= 2 \times \frac{Precision \cdot Recall}{Precision + Recall}$: harmonic average of precision and recall ranging from 0 to 1; F_Score = 1 presents the perfect precision and recall while F_Score = 0 represent the worst.

where TP = True Positive; FP = False Positive; TN = True Negative; FN = False Negative.

The performance measures are based on early depression detection. This means we only account for the predictions when the date that a user took the depression questionnaires is after the last day that he or she posted a status update. Fig. 12 shows the performance of different algorithms in terms of the performance metrics. In this experiment, CART achieves the highest accuracy, precision and F_Score while KNN yields the best recall and CART's is the second best. According to the figure, CART's performance is dominant with an accuracy of 69.05%, precision of 76.19%, recall of 66.67%, and F_Score of 71.11%. Since CART achieves the best performance, we run another experiment using CART to explore the early depression prediction power of different feature subsets. The details of feature subsets are presented in Table 5. One of the most common types of features being used in the past for depression detection is LIWC features [8,9,36]. LIWC is a text analysis computer program that counts words and calculate the percentage of words in each LIWC's linguistics, and the psychological and topical categories for a given text [74]. It is commonly used as linguistic features for depression detection [8,9,15]. Given the status update text corpus, we run

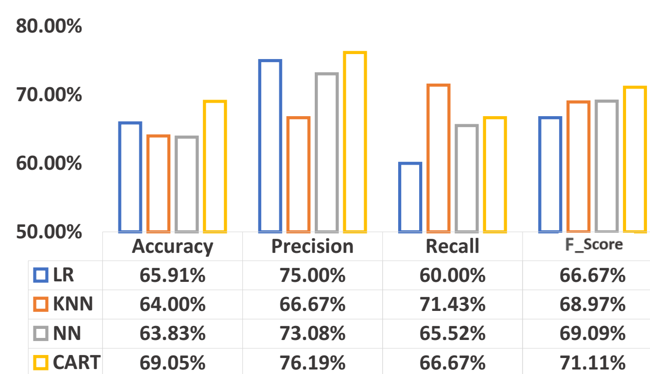



**Fig. 12.** Performance of different ML methods with all proposed social influence-based contrast features.


**Fig. 11.** Depressed and non-depressed discriminate topics.

**Table 5**
Feature sets.

| # | Feature Sets |
|---|---|
| SBERT | Embedding features extracted by SBERT from raw text |
| LIWC | LIWC Features |
| GTF | LIWC Features and General Topics Features |
| RFE Features | Selected features through Recursive Feature Elimination (RFE) feature selection algorithm |
| FS1 | LIWC Features and Social Influence-based Contrast Features: Topic-level Contrast Features |
| FS2 | LIWC Features and Social Influence-based Contrast Features: Network-based Contrast Features |
| FS3 | LIWC Features and Social Influence-based Contrast Features: Topic-level Contrast Features, Network-based Contrast Features |
| FS4 | LIWC Features and Social Influence-based Contrast Features: Term-level Contrast Features, Topic-level Contrast Features |
| FS5 | LIWC Features and Social Influence-based Contrast Features: Term-level Contrast Features, Topic-level Contrast Features, Network-based Contrast Features |
| FS6 | Social Influence-based Contrast Features: Term-level Contrast Features, Topic-level Contrast Features, Network-based Contrast Features |
| FS7 | Social Influence-based Contrast Features: Topic-level Contrast Features, Network-based Contrast Features |
| FS8 | Social Influence-based Contrast Features: Term-level Contrast Features, Topic-level Contrast Features |

LIWC tool [32] and report the percentages of the words in each LIWC category out of all of the words for each user used in the combined status updates. We use LIWC features (LIWC) as one of our baselines to benchmark with the depression detection using proposed contrast languages features. Several studies adopt topic modeling for depression detection [8,30]. To show the power of the proposed contrast features compared to the existing approach of using topic modeling for depression detection, we implement the method in [8], which uses LIWC and a general LDA topic model to obtain generalized topics for depressed and non-depressed user groups. We present this method as General Topic Feature (GTF) model in our experiment and compare its performance with the proposed contrast topic features. In order to show the prediction power of our selected contrast features compared to non-selected features, we use Recursive Feature Elimination (RFE) feature selection algorithms from Python *scikit-learn* [75] to select most informative features from the non-selected contrast term-level and topic-level features. Recursive Feature Elimination is a common feature selection algorithm that selects features by removing the least important features We present this baseline as RFE features in our experiment and compare its performance with the proposed contrast features that we selected by choosing the most discriminant contrast features. We designed and developed an SBERT-based method [57] for depression detection as another baseline. The intuition is based on the fact that users in the same group (e.g., depressed or non-depressed) usually publish content with similar semantic and context. Therefore, we used SBERT to generate embedding vectors representing contents (e.g., sentences) of each user's raw status updates. The values of features in the embedding space are being fed to the classifiers to train the depression detection model. Fig. 13 shows accuracy, precision, recall and F_Score with different feature sets compared to baselines SBERT, LIWC, GTF and RFE. The features sets with proposed contrast features significantly improve the detection power compared to these baselines. To be more specific, the models based on FS1-FS5, with LIWC and different combinations of term-level, topic-level and network-based contrast features all outperform baseline LIWC and SBERT. The model with LIWC and with term-level, topic-level and network-based contrast features FS5, has the best accuracy and precision, and the model with term-level and topic-level contrast features FS4 shows the best recall among all features with LIWC. Compared to GTF, the performance of FS1 shows that the proposed topic-level contrast features improve detection accuracy, precision, recall and F_Score of the general topic features. The model based on FS8, with topic-level and term-level contrast features shows that the selected contrast features based on our feature selection criteria (e.g., find the most discriminant features) perform better than a general feature selection method. The model based on FS7, with topic-level and network-based contrast features, generates the best accuracy, precision and F_Score while the model based on FS8 shows the best recall.The model based on FS6, with all proposed contrast features, show the second best performance for all metrics among all feature sets. The performance of F7 in F_Score, accuracy and precision and performance of F8 in recall is a little bit higher than FS6 which uses all contrast features. This suggests that using all contrast features does not necessarily improve the result. In summary, the models with only feature sets from the proposed contrast language feature set FS6-FS8 perform the best among all feature sets and all outperform SBERT.

In the next experiment, we identify the importance of social influence-based contrast language features and their prediction power in detecting early depression. Fig. 14 shows top-20 features. We see that all topic-level contrast features from the depressed and non-depressed user groups are the most important features in detecting early depression. They are followed by term-level contrast features such as non-depressed group contrast features"week","last","answer" and depressed group term-level contrast features"sick","heart","alone". The network-based contrast features from the non-depressed group is also one of the most critical drivers to depression early detection.

## 5. Discussions and implications

### 5.1. Research contributions

Language is the most powerful reflection of mental states and emotions. Online forums and social interactions have been the most common phenomena on the Internet in various social settings to share feelings and thoughts with friends, family and online social groups. The online UGC have become a rich source of information to improve the current approaches to (early) depression diagnosis and treatment strategies. Existing studies in (early) depression detection using ML techniques mainly focus on extracting data-driven features to build a predictive model. However, there is a lack of model interpretability that can be used to provide clinicians with appropriate and justified decision support during the diagnosis, early intervention and treatments. Social influence theory shows that the way an individual behaves, thinks, or acts is influenced by her social environment. Prior studies fail to explain how depressed users communicate differently in online social platforms, in terms of the social influence-based language styles and contents involved in their social activities and interactions. We demonstrated a social influence-based contrast language analysis framework that identifies discriminant language features between depressed and non-depressed users in online social forums (e.g., patient status social networks) to enhance decision making in CDSS. Unlike existing approaches, our framework is designed to address the gap between model accuracy and their interpretability. We show such features can offer more actionable information to understand the differences in users' mental health state, identify depressive behaviours and language at an early stage and design customized treatment plans. For clinicians and mental health professionals. Our framework advances current knowledge and evidence-based healthcare practice by proposing social influence-based contrast language features from influential content-based and network-based data for depression analysis that aims to contribute to enhanced decision-making in CDSSs. We learn from the social network structure and discover the social influence-based contrast features using network-based contrast features that consider influential contents (topic-level and term-level contrast features) from depressed and non-depressed user groups. During the influential contents generation, we first analyze the percentage difference of terms in depressed and non-depressed user groups to discover term-level contrast features. Then, we identify topic-level contrast features by first building topic models based on social network contents, then conducting a deep similarity analysis of topics.

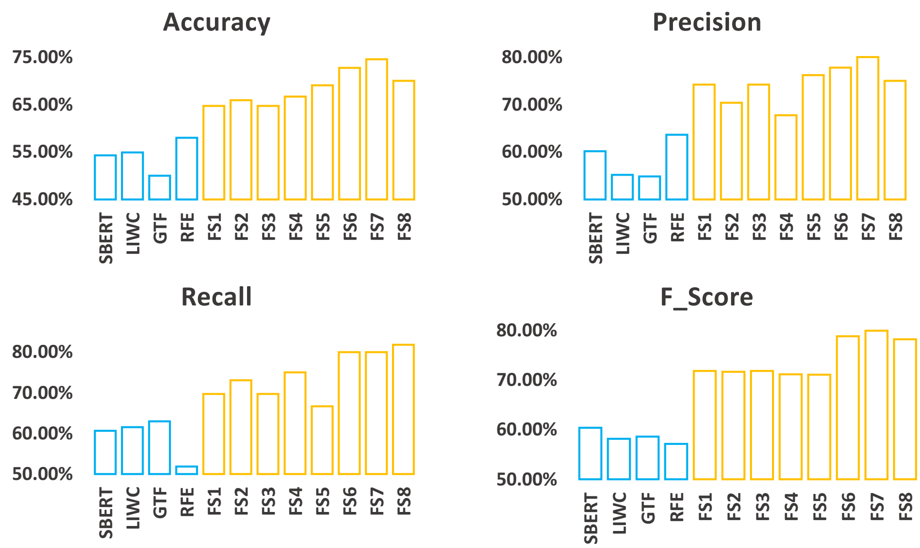



**Fig. 13.** Prediction power of different feature sets with classification and regression trees (CART).

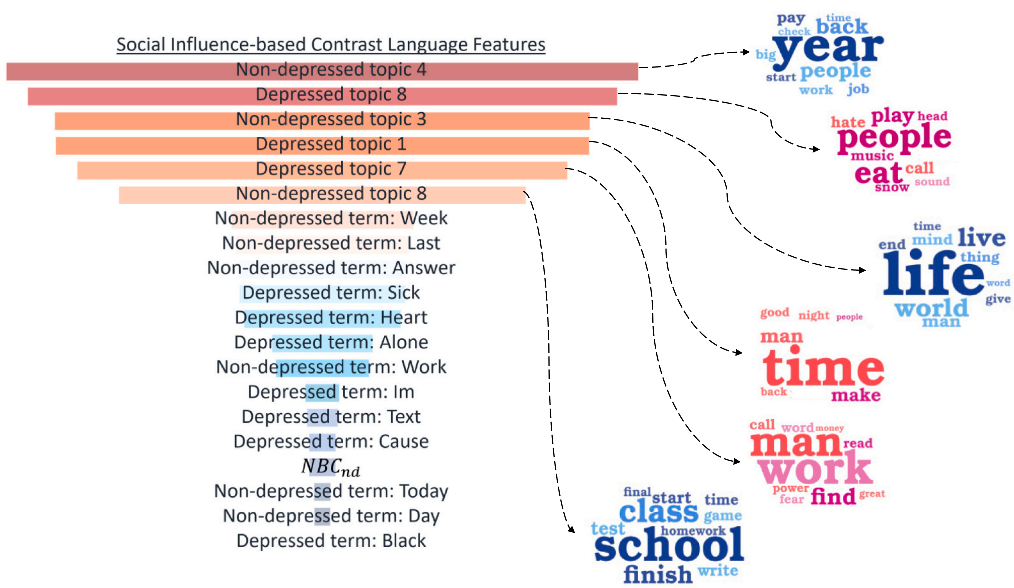



**Fig. 14.** Feature importance and representation of topic-level features.

Our framework provides a deeper understanding of mental health and interpretable knowledge to clinicians to support targeted analysis to a patient, detect early signs and potential causes of depression, apply appropriate early intervention for users at risk and treatment plans for those already affected.

### 5.2. Implications for practice

Our framework can be used to develop predictive models based on mental status network or other similar networks that involve online social interactions. These models can be also used in other domains and problem settings involving more than one class of users to support decisions. More importantly, our framework can be applied in different scenarios to enhance decision-making in clinical environment. In particular, it benefits clinicians and health professionals in three main aspects. First, it can help mental health professionals to understand their patients better and capture distinct patterns between user groups to improve the effectiveness of (early) depression diagnosis. Second, the treatments for depressed users usually involve long interviews to learn about patient issues. However, clinicians mostly ask about how patients feel at a given moment, which can be influenced by the current mental state of the user and potentially misleading. Although we used the social media platforms to demonstrate our framework, it can be adapted to domain-specific data generated by CDSSs. In a CDSS setting, content can be internally captured through patient interactions and feedback channels. Current CDSSs have collaborative features that allow not only patient-to-patient interactions, but also direct engagements between clinicians and patients. These include functions such as customized patient interaction, documentation, diagnostic support, contextual information, alerts and reminders, among others. By incorporating content generated through such channels and contrast features of a patient during a treatment, clinicians can gain a deeper insight into her emotions during different time periods, which is usually challenging to discover. Our framework also provides different topics to lead the discussions during meetings with a patient more effectively. Third, besides the benefits for depressed individuals, the framework is also useful for non-depressed users who are at a risk of developing depression symptoms. It can generate a comparative representation of language style, words used and contents of depressed users versus non-depressed users. This can enrich a CDSS to detect if a non-depressed user uses similar language patterns as depressed users.

In addition to its use in the context of depression, our framework can be used in other related domains with multiple groups of individuals where differentiating the social influence-based language patterns and styles are of importance. For example, it can be used to discriminate depression communities from the other online subgroups, such as bipolar disorder, self-harm, grief, bereavement, and suicide from online conversations [76]. Contrast Topics and psycholinguistic features can be extracted from each sub-group, which can help to understand different mental moods better and develop methods for treatment and prevention from worsening mental disorder symptoms. Other than the health care context, our framework can be used to examine how customers' different emotions expressed through online postings from social platforms influence institutional investors' investment decisions [77]. Another example is our framework can be used to analyze bilateral reviews collected from both sellers and buyers from e-commerce platforms and predict transaction results [78]. Based on the social influence-based contrast languages used by buyers and sellers regarding different products, the platform can understand their customers' needs better and present buyers with offerings that are more likely to lead to a transaction between a buyer and a seller.

## 6. Conclusion

This study presents a data-driven analytics framework that incorporates NLP and ML methods for social influence-based contrast language features analysis, including the extraction of content-based and network-based contrast features, in a CDSS environment. By analyzing different language habits used by depressed and non-depressed groups, we are able to derive meaningful and actionable insights and analyze different behaviours of individuals from the two groups. This makes online social platforms a tremendous potential and a key source of information for mental health professionals and clinicians to provide a better social and decision support in terms of (early) depression diagnosis, supporting depressed users with more customized treatment as well as helping prevent non-depressed individuals who are at high risk of developing depression symptoms with early intervention. We designed a early depression detection user study to evaluate the effectiveness of our framework using different ML algorithms. We show that the proposed social influence-based contrast language framework effectively detects depression and significantly outperforms the baselines. We acknowledge that there are limitations in this study that can be addressed in future research. Our study focuses on finding discriminant language features between depressed and non-depressed individuals. This framework can be used by mental health professionals to offer customized treatment plans for depressed individuals by monitoring and comparing their language differences from non-depressed online platform users. However, for depressed users, different intensity levels may exist, and they may use different languages when expressing themselves. Thus, one possible future direction is to study different intensity levels of depression and how language use is different at each stage of depression and their potential insights for the customization of treatment plans. Furthermore, we plan to build a case-based recommendation system incorporating discriminant treatment information from a different intensity level of depressed users. It can take language information from various sources and recommend medical charts that health practitioners could easily access and offer corresponding treatments based on previous patient cases. This requires collaborations with the medical professionals for more focused field studies on how to design treatment plans.

## Author statement

Xingwei Yang: Writing- Original draft preparation, Methodology, Visualization.

Alexandra Joukova: Investigation, Data curation, Development.

Anteneh Ayanso: Writing- Reviewing and Editing, Supervision.

Morteza Zihayat: Supervision, Conceptualization, Methodology, Software.

## References

[1] NIMH, National institute of mental health, statistics. http://tinyurl.com/452x2rv, 2011 accessed: January, 2020.
[2] WHO, World health organization, report of depression, 2020. https://www.who.int/news-room/fact-sheets/detail/depression. accessed: January, 2021.
[3] A. Althubaiti, Information bias in health research: definition, pitfalls, and adjustment methods, J. Multidiscip. Healthc. 9 (2016) 211.
[4] M.Z. Hydari, R. Telang, W.M. Marella, Saving patient Ryan—can advanced electronic medical records make patient care safer? Manag. Sci. 65 (2019) 2041–2059.
[5] S. Piri, D. Delen, T. Liu, H.M. Zolbanin, A data analytics approach to building a clinical decision support system for diabetic retinopathy: developing and deploying a model ensemble, Decis. Support. Syst. 101 (2017) 12–27.
[6] S. Piri, Missing care: a framework to address the issue of frequent missing values; the case of a clinical decision support system for parkinson's disease, Decis. Support. Syst. 136 (2020), 113339.
[7] Y. Li, A. Vo, M. Randhawa, G. Fick, Designing utilization-based spatial healthcare accessibility decision support systems: a case of a regional health plan, Decis. Support. Syst. 99 (2017) 51–63.
[8] M.M. Tadesse, H. Lin, B. Xu, L. Yang, Detection of depression-related posts in reddit social media forum, IEEE Access 7 (2019) 44883–44893.
[9] X. Yang, R. McEwen, L.R. Ong, M. Zihayat, A big data analytics framework for detecting user-level depression from social networks, Int. J. Inf. Manag. 54 (2020), 102141.
[10] B. Saha, T. Nguyen, D. Phung, S. Venkatesh, A framework for classifying online mental health-related communities with an interest in depression, IEEE J. Biomed. and Health Inform. 20 (2016) 1008–1015.
[11] H.C. Kelman, Compliance, identification, and internalization three processes of attitude change, J. Confl. Resolut. 2 (1958) 51–60.
[12] R.P. Bagozzi, U.M. Dholakia, Intentional social action in virtual communities, J. Interact. Mark. 16 (2002) 2–21.
[13] R.P. Bagozzi, U.M. Dholakia, A. Mookerjee, Individual and group bases of social influence in online environments, Media Psychol. 8 (2006) 95–126.
[14] H.A. Schwartz, J. Eichstaedt, M. Kern, G. Park, M. Sap, D. Stillwell, M. Kosinski, L. Ungar, Towards assessing changes in degree of depression through facebook, in, in: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 2014, pp. 118–125.
[15] X. Chen, M.D. Sykora, T.W. Jackson, S. Elayan, What about mood swings: Identifying depression on twitter with temporal measures of emotions, in: Proceedings of the The Web Conference 2018, 2018, pp. 1653–1660.
[16] M.R. Islam, M.A. Kabir, A. Ahmed, A.R.M. Kamal, H. Wang, A. Ulhaq, Depression detection from social network data using machine learning techniques, Health Inform. Sci. and Syst. 6 (2018) 8.
[17] A. Mehmood, I. Natgunanathan, S. Guo, Protection of big data privacy, IEEE Access 4 (2016) 1821–1834.
[18] B. Kratzwald, S. Ilić, M. Kraus, S. Feuerriegel, H. Prendinger, Deep learning for affective computing: text-based emotion recognition in decision support, Decis. Support. Syst. 115 (2018) 24–35.
[19] R. Thackeray, B.L. Neiger, A.K. Smith, S.B. Van Wagenen, Adoption and use of social media among public health departments, BioMed central, Public Health 12 (2012) 1–6.
[20] M.A. Moreno, L.A. Jelenchick, K.G. Egan, E. Cox, H. Young, K.E. Gannon, Feeling bad on facebook: depression disclosures by college students on a social networking site, Depress. Anxiety 28 (2011) 447–455.
[21] A. Li, D. Jiao, T. Zhu, Detecting depression stigma on social media: a linguistic analysis, J. Affect. Disord. 232 (2018) 358–362.
[22] F. Li, T.C. Du, Listen to me—evaluating the influence of micro-blogs, Decis. Support. Syst. 62 (2014) 119–130.
[23] A. Bandura, D.C. McClelland. Social learning theory (vol.1), Englewood Cliffs Prentice Hall, 1977.
[24] L. Qin, Y. Kim, J. Hsu, X. Tan, The effects of social influence on user acceptance of online social networks, Int. J. Hum.-Comp. Interact. 27 (2011) 885–899.
[25] Y. Wang, D.B. Meister, P.H. Gray, Social influence and knowledge management systems use: evidence from panel data, Manag. Inf. Syst. Q. (2013) 299–313.
[26] C.M. Cheung, P.-Y. Chiu, M.K. Lee, Online social networks: why do students use facebook? Comput. Hum. Behav. 27 (2011) 1337–1343.
[27] L. Yan, Y. Tan, Feeling blue? Go online: an empirical study of social support among patients, Inf. Syst. Res. 25 (2014) 690–709.
[28] X. Wang, K. Zhao, S. Cha, M.S. Amato, A.M. Cohn, J.L. Pearson, G. D. Papandonatos, A.L. Graham, Mining user-generated content in an online smoking cessation community to identify smoking status: a machine learning approach, Decis. Support. Syst. 116 (2019) 26–34.
[29] Y.W. Teh, A hierarchical bayesian language model based on pitman-yor processes, in, in: Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, 2006, pp. 985–992.

[30] P. Resnik, W. Armstrong, L. Claudino, T. Nguyen, V.-A. Nguyen, J. Boyd-Graber, Beyond LDA: Exploring supervised topic modeling for depression-related language in twitter, in: Proceedings of 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 2015, pp. 99–107.
[31] J.H. Shen, F. Rudzicz, Detecting anxiety through reddit, in, in: Proceedings of the 4th Workshop on Computational Linguistics and Clinical Psychology, 2017, pp. 58–65.
[32] Linguistic Inquiry and World Count, Liwc2015:how it works, 2020. http://liwc.wpengine.com/how-it-works/. accessed: January, 2021.
[33] L. Chen, A. Baird, D. Straub, A linguistic signaling model of social support exchange in online health communities, Decis. Support. Syst. 130 (2020), 113233.
[34] J. Feldhege, M. Moessner, S. Bauer, Who says what? Content and participation characteristics in an online depression community, J. Affect. Disord. 263 (2020) 521–527.
[35] P. Grover, A.K. Kar, G. Davies, "Technology enabled health"–insights from twitter analytics with a sociotechnical perspective, Int. J. Inf. Manag. 43 (2018) 85–97.
[36] T. Nguyen, M.E. Larsen, B. O'Dea, D.T. Nguyen, J. Yearwood, D. Phung, S. Venkatesh, H. Christensen, Kernelbased features for predicting population health indices from geocoded social media data, Decis. Support. Syst. 102 (2017) 22–31.
[37] T. Nguyen, D. Phung, B. Dao, S. Venkatesh, M. Berk, Affective and content analysis of online depression communities, IEEE Trans. Affect. Comput. 5 (2014) 217–226.
[38] C. Segrin, Social skills deficits associated with depression, Clin. Psychol. Rev. 20 (2000) 379–403.
[39] T. Slonim, Verbal Behavior in Individuals with Generalized Anxiety Disorder and Depressive Disorders, City University of New York, 2014.
[40] S. Rude, E.-M. Gortner, J. Pennebaker, Language use of depressed and depression-vulnerable college students, Cognit. Emot. 18 (2004) 1121–1133.
[41] E.A. Ríssola, M. Aliannejadi, F. Crestani, Beyond modelling: understanding mental disorders in online social media, Eur. Conf. on Inform. Retrieval, Springer (2020) 296–310.
[42] M. Park, D. McDonald, M. Cha, Perception differences between the depressed and non-depressed users in twitter, in: Proceedings of the International Association for the Advancement of Artificial Intelligence Conference on Web and Social Media volume 7, 2013, pp. 476–485.
[43] R. Neumann, F. Strack, "Mood contagion": the automatic transfer of mood between persons, J. Pers. Soc. Psychol. 79 (2000) 211.
[44] N.S. Contractor, L.A. DeChurch, Integrating social networks and human social motives to achieve social influence at scale, Proc. Natl. Acad. Sci. 111 (2014) 13650–13657.
[45] R.T. Sutton, D. Pincock, D.C. Baumgart, D.C. Sadowski, R.N. Fedorak, K.I. Kroeker, An overview of clinical decision support systems: benefits, risks, and strategies for success, Nat. Partner J. Digital Med. 3 (2020) 1–10.
[46] M. Kosinski, S.C. Matz, S.D. Gosling, V. Popov, D. Stillwell, Facebook as a research tool for the social sciences: opportunities, challenges, ethical considerations, and practical guidelines, Am. Psychol. 70 (2015) 543.
[47] L.S. Radloff, The ces-d scale: a self-report depression scale for research in the general population, Appl. Psychol. Meas. 1 (1977) 385–401.
[48] L.J. Julian, S.E. Gregorich, C. Tonner, J. Yazdany, L. Trupin, L.A. Criswell, E. Yelin, P.P. Katz, Using the center for epidemiologic studies depression scale to screen for depression in systemic lupus erythematosus, Arthritis Care Res. 63 (2011) 884–890.
[49] Wolf Garbe, Symspell 6.7, 2020. https://github.com/wolfgarbe/symspell. accessed: January, 2020.
[50] J.D. Bernard, J.L. Baddeley, B.F. Rodriguez, P.A. Burke, Depression, language, and affect: an examination of the influence of baseline depression and affect induction on language, J. Lang. Soc. Psychol. 35 (2016) 317–326.
[51] M. Johnson, S. Lapkin, V. Long, P. Sanchez, H. Suominen, J. Basilakis, L. Dawson, A systematic review of speech recognition technology in health care, BioMed. Central Med. Inform. Decision Making 14 (2014) 1–14.
[52] Z.Y. Zhuang, C.L. Wilkin, A. Ceglowski, A framework for an intelligent decision support system: a case in pathology test ordering, Decis. Support. Syst. 55 (2013) 476–487.
[53] J. Lee, J. Kim, I. Cho, Y. Kim, Integration of workflow and rule engines for clinical decision support services, in: Medinfo 2010, IOS Press, 2010, pp. 811–815.
[54] J.-H. Park, J.-H. Baek, S.J. Sym, K.Y. Lee, Y. Lee, A data-driven approach to a chemotherapy recommendation model based on deep learning for patients with colorectal cancer in Korea, BioMed Central Medical Informatics and Decision Making 20 (2020) 1–12.
[55] S. Morinaga, K. Yamanishi, K. Tateishi, T. Fukushima, Mining product reputations on the web, in, in: Proceedings of the Association for Computing Machinery Special Interest Group on Knowledge Discovery and Data Mining Conference on Knowledge Discovery and Data Mining, 2002, pp. 341–349.
[56] P. Rayson, R. Garside, Comparing corpora using frequency profiling, in, Proceedings of the Workshop on Comparing Corpora (2000) 1–6.
[57] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2019, pp. 671–688.
[58] S. Bowman, G. Angeli, C. Potts, C.D. Manning, A large annotated corpus for learning natural language inference, in, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015, pp. 632–642.
[59] A. Williams, N. Nangia, S. Bowman, A broad-coverage challenge corpus for sentence understanding through inference, in, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018, pp. 1112–1122.
[60] Y. Si, J. Wang, H. Xu, K. Roberts, Enhancing clinical concept extraction with contextual embeddings, J. Am. Med. Inform. Assoc. 26 (2019) 1297–1304.
[61] Y. Liang, R. Cao, J. Zheng, J. Ren, L. Gao, Learning to remove: towards isotropic pre-trained bert embedding, Int. Conf. on Artificial Neural Networks, Springer (2021) 448–459.
[62] L. Van der Maaten, G. Hinton, Visualizing data using t-sne, J. Machine Learning Research 9 (2008).
[63] L. McInnes, J. Healy, N. Saul, L. Großberger, Umap: uniform manifold approximation and projection, J. Open Source Softw. 3 (2018).
[64] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
[65] K. Stevens, P. Kegelmeyer, D. Andrzejewski, D. Buttler, Exploring topic coherence over many models and many topics, in, in: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2012, pp. 952–961.
[66] A.K. McCallum, Mallet: A Machine Learning for Language Toolkit. Http://mallet.cs.umass.edu, 2002.
[67] G. Navarro, A guided tour to approximate string matching, Association for Computing Machinery computing Surveys 33 (2001) 31–88.
[68] D. Jan, L. Zeevaert, Receptive Multilingualism: Linguistic Analyses, Language Policies and Didactic Concepts volume 6, John Benjamins Publishing, 2007.
[69] Government of Canada, About mental illness, 2017. https://www.canada.ca/en/public-health/services/about-mental-illness.html. accessed: January, 2020.
[70] J. Tang, J. Sun, C. Wang, Z. Yang, Social influence analysis in large-scale networks, in: Proceedings of the 15th Association for Computing Machinery Special Interest Group on Knowledge Discovery and Data Mining International Conference on Knowledge Discovery and Data Mining, 2009, pp. 807–816.
[71] F.R. Kschischang, B.J. Frey, H.-A. Loeliger, Factor graphs and the sum-product algorithm, IEEE on Inform. Theory 47 (2001) 498–519.
[72] J. Zhang, B. Liu, J. Tang, T. Chen, J. Li, Social influence locality for modeling retweeting behaviors, in: Proceedings of International Joint Conference on Artificial Intelligence volume 13, 2013, pp. 2761–2767.
[73] T.N. Poly, M.M. Islam, M.S. Muhtar, H.-C. Yang, P.A.A. Nguyen, Y.-C.J. Li, Machine learning approach to reduce alert fatigue using a disease medication–related clinical decision support system: model development and validation, J. Med. Internet Res. Med. Inform. 8 (2020), e19489.
[74] Y.R. Tausczik, J.W. Pennebaker, The psychological meaning of words: Liwc and computerized text analysis methods, J. Lang. Soc. Psychol. 29 (2010) 24–54.
[75] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[76] T. Nguyen, B. O'Dea, M. Larsen, D. Phung, S. Venkatesh, H. Christensen, Using linguistic and topic analysis to classify sub-groups of online depression communities, Multimed. Tools Appl. 76 (2017) 10653–10676.
[77] H. Nguyen, R. Calantone, R. Krishnan, Influence of social media emotional word of mouth on institutional investors' decisions and firm value, Manag. Sci. 66 (2020) 887–910.
[78] J. Chen, Y. Yang, H. Liu, Mining bilateral reviews for online transaction prediction: a relational topic modeling approach, Inf. Syst. Res. 32 (2) (2021) 541–560.

**Xingwei Yang** is a PhD candidate in Management Analytics at the Smith School of Business at Queen's University. She received her B.Sc. in Mathematics from University of British Columbia. Her research interests are in the area of machine learning, Big Data and financial mathematics.

**Alexandra Joukova** received master's degree in Data Science Program at Ryerson University. Her main research interests are text mining and big data.

**Anteneh Ayanso** is Professor of Information Systems and founding director of the Centre for Business Analytics at the Goodman School of Business, Brock University, Canada. His research interests focus primarily on data management and information retrieval, Big Data analytics, electronic commerce, and electronic government. His articles are published in leading journals such as *Decision Sciences, Decision Support Systems, European Journal of Operational Research, Journal of Business Research, International Journal of Electronic Commerce, Government Information Quarterly*, among others. His research has been funded by government grants, including NSERC Discovery Research Grant, SSHRC Insight Development Grant, NSERC Engage Grant, and Ontario Centres of Excellence (OCE). He is currently serving as Senior Editor at *Information Technology & People*, Associate Editor at *Decision Support Systems*, and *International Journal of Business and Data Analytics*; and review board member at *Journal of Database Management*, and *International Journal of Convergence Computing*.

**Morteza Zihayat** is an Associate Professor at the School of Information Technology Management of Ryerson University from 2016 and IBM Faculty Fellow from 2018. Before joining ITM, he was a Postdoctoral Fellow at University of Toronto and a research fellow in the IBM Cloud Analytics. His research concerns Big Data analytics and machine learning. He obtained his PhD from York University in 2016.He has published in various well-respected journals and conferences in data mining, machine learning and information systems such as *International Journal of Information Management (IJIM), Decision Support Systems (DSS), IEEE TKDE, Information Processing and Management, Information Sciences, ACM SIGKDD, SIGIR, ECIR, PKDD, SDM SIAM, EDBT.*