Journal of Affective Disorders 355 (2024) 40–49

Contents lists available at ScienceDirect

Journal of Affective Disorders

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

Research paper

Identifying depression-related topics in smartphone-collected free-response speech recordings using an automatic speech recognition system and a deep learning topic model

Yuezhou Zhang [a,*], Amos A. Folarin [a,b,c,d], Judith Dineley [a,e], Pauline Conde [a], Valeria de Angel [a], Shaoxiong Sun [a,f], Yatharth Ranjan [a], Zulqarnain Rashid [a], Callum Stewart [a], Petroula Laiou [a], Heet Sankesara [a], Linglong Qian [a], Faith Matcham [a,g], Katie White [a], Carolin Oetzmann [a], Femke Lamers [h], Sara Siddi [i], Sara Simblett [a], Björn W. Schuller [e,j], Srinivasan Vairavan [k], Til Wykes [a,c], Josep Maria Haro [i], Brenda W.J.H. Penninx [h], Vaibhav A. Narayan [k,l], Matthew Hotopf [a,c], Richard J.B. Dobson [a,b,c,d,1], Nicholas Cummins [a,*,1], RADAR-CNS consortium[2]

[a] Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[b] University College London, London, UK
[c] South London and Maudsley NHS Foundation Trust, London, UK
[d] Health Data Research UK London, University College London, London, UK
[e] University of Augsburg, Augsburg, Germany
[f] Department of Computer Science, University of Sheffield, Sheffield, UK
[g] School of Psychology, University of Sussex, Falmer, East Sussex, UK
[h] Department of Psychiatry, Amsterdam Public Health Research Institute and Amsterdam Neuroscience, Amsterdam University Medical Centre, Vrije Universiteit and GGZ InGeest, Amsterdam, the Netherlands
[i] Parc Sanitari Sant Joan de Déu, Fundació Sant Joan de Déu, CIBERSAM, Universitat de Barcelona, Barcelona, Spain
[j] GLAM - Group on Language, Audio, & Music, Imperial College London, London, UK
[k] Janssen Research and Development LLC, Titusville, NJ, USA
[l] Davos Alzheimer's Collaborative, Geneva, Switzerland.

ARTICLE INFO

Keywords:
Depression
Speech
Automatic speech recognition
Topic modeling
Smartphone

ABSTRACT

Background: Prior research has associated spoken language use with depression, yet studies often involve small or non-clinical samples and face challenges in the manual transcription of speech. This paper aimed to automatically identify depression-related topics in speech recordings collected from clinical samples.

Methods: The data included 3919 English free-response speech recordings collected via smartphones from 265 participants with a depression history. We transcribed speech recordings via automatic speech recognition (Whisper tool, OpenAI) and identified principal topics from transcriptions using a deep learning topic model (BERTopic). To identify depression risk topics and understand the context, we compared participants' depression severity and behavioral (extracted from wearable devices) and linguistic (extracted from transcribed texts) characteristics across identified topics.

Results: From the 29 topics identified, we identified 6 risk topics for depression: 'No Expectations', 'Sleep', 'Mental Therapy', 'Haircut', 'Studying', and 'Coursework'. Participants mentioning depression risk topics exhibited higher sleep variability, later sleep onset, and fewer daily steps and used fewer words, more negative language, and fewer leisure-related words in their speech recordings.

Limitations: Our findings were derived from a depressed cohort with a specific speech task, potentially limiting the generalizability to non-clinical populations or other speech tasks. Additionally, some topics had small sample sizes, necessitating further validation in larger datasets.

* Corresponding authors.
E-mail addresses: yuezhou.zhang@kcl.ac.uk (Y. Zhang), nick.cummins@kcl.ac.uk (N. Cummins).

[1] These authors contributed equally.
[2] www.radar-cns.org

https://doi.org/10.1016/j.jad.2024.03.106
Received 26 September 2023; Received in revised form 18 March 2024; Accepted 22 March 2024
Available online 27 March 2024

Conclusion: This study demonstrates that specific speech topics can indicate depression severity. The employed data-driven workflow provides a practical approach for analyzing large-scale speech data collected from real-world settings.

1. Introduction

Depression, a prevalent mental health disorder, imposes substantial economic and societal burdens (Hawton et al., 2013; Lenox-Smith et al., 2013; Lerner et al., 2004; Lewinsohn et al., 2000). Current clinical depression diagnostics, relying on subjective recall (questionnaires and interviews) and the expertise of clinicians (Althubaiti, 2016; Devaux and Sassi, 2016), often lead to inadequate and delayed treatment (Kessler et al., 2005). Consequently, there is a need for more objective and effective early-stage depression detection methods (De Angel et al., 2022b).

Natural language use can reflect an individual's personality, emotions, and mental health status, including depression (Losada and Crestani, 2016). Prior research has explored the relationship between language use and depression (Pennebaker et al., 2003; Tølbøll, 2019), analyzing written formats, like essays (Holmes et al., 2007; Rude et al., 2004), diaries (Baikie et al., 2006; Rodriguez et al., 2010), and unstructured surveys (Cook et al., 2016), as well as oral formats such as clinical interview (Li et al., 2023), therapy session speech tasks (Sonnenschein et al., 2018; Zimmermann et al., 2017), and responses to oral questions (Fast and Funder, 2010; Zimmermann et al., 2013). These studies consistently link increased use of negative words and first-person singular pronouns with worsening depression (Tølbøll, 2019). However, the traditional collection of language samples is time-consuming and expensive, resulting in limited sample sizes (Losada and Crestani, 2016).

To address this limitation, recent studies have applied natural language processing (NLP) technologies to analyze a vast volume of social media posts for depression prediction (Nanomi Arachchige et al., 2021; Skaik and Inkpen, 2020; Tadesse et al., 2019). Yet, these social media studies mostly use non-clinical depression labels (e.g., keywords and tags), overlook non-social media users' language data, and lack contextual information, constraining their generalizability and interpretability.

Recent mobile health studies, incorporating mobile and Internet-of-Things technologies, can concurrently track individuals' spoken language recordings, depression questionnaires, and wearable-measured behaviors (Dineley et al., 2021; Matton et al., 2019). This approach provides a unique opportunity to link spoken language use with depression in real-world settings, contextualized by behaviors (De Angel et al., 2022b; Rohani et al., 2018). However, a major challenge in these studies is the time-consuming and costly manual transcription of speech recordings. Advances in automatic speech recognition (ASR) systems, such as OpenAI's Whisper (Radford et al., 2022), offer near-human-level accuracy for large-scale transcription of speech-to-text, enabling the subsequent NLP analysis.

Topic modeling, a popular NLP method (Wallach, 2006), can be utilized to identify specific depression-related topics within transcribed speech texts and alongside depression questionnaires, thereby revealing intuitive and interpretive language markers of depression. Notably, several topics such as feelings of depression, loneliness, hostility, somatic complaints, and medical references have been found to be significantly associated with depression (Eichstaedt et al., 2018). The present study explored depression risk topics in mobile health speech recordings and extracted contextual information (behaviors and linguistic characteristics) to enhance the interpretation of the emergence of these risk topics. Conducting a secondary analysis of free-response speech recordings gathered from two mobile health studies, this study aimed to (i) develop a data-driven workflow for topic identification from speech recordings collected in daily-life settings; (ii) identify and contextualize risk topics for depression; and (iii) investigate the association between topic changes and fluctuations in depression severity.

2. Methods

2.1. Remote Assessment of Disease and Relapse Major Depressive Disorder (RADAR-MDD) dataset

The RADAR-MDD study is a large observational investigation assessing the utility of remote technologies for monitoring depression (Matcham et al., 2019). The study recruited 623 participants with a depression history from three sites in the United Kingdom, Spain, and the Netherlands, and followed them for up to over 2 years (Matcham et al., 2019). Enrollment began in November 2017, ended in June 2020, and data collection finished in April 2021 (Matcham et al., 2022). The study used RADAR-base (Ranjan et al., 2019), an open-source platform, to remotely gather participants' active (questionnaires and speech tasks) and passive (smartphones and Fitbit) data streams.

Speech recordings involved two tasks: (i) a scripted task, reading a script from Aesop's fable The North Wind and The Sun (Association, 1999), and (ii) a free-response task, allowing participants to describe what they were looking forward to in the upcoming week (Cummins et al., 2015; Mundt et al., 2007). Details of the RADAR-MDD dataset and speech data collection have been reported in our previous publications (Cummins et al., 2023; Dineley et al., 2021; Matcham et al., 2019; Matcham et al., 2022). This study focused only on English free-response speech recordings from the UK site due to our research aims and the performance variation of Whisper in language translation.

2.2. Remote Assessment of Treatment Prognosis in Depression (RAPID) dataset

The RAPID study investigated the feasibility of remote technologies during the treatment of depression (De Angel et al., 2023; de Angel et al., 2022a). From June 2020 to June 2021, 66 adults in London, UK seeking depression treatments were recruited and followed for 7 months (De Angel et al., 2023; de Angel et al., 2022a). Since the RAPID used the same speech collection protocol as the RADAR-MDD, the free-response speech recordings collected in the RAPID were used for validation.

2.3. Ethics considerations

The RADAR-MDD and RAPID studies were both conducted per the Declaration of Helsinki and Good Clinical Practice, adhering to principles outlined in the National Health Service (NHS) Research Governance Framework for Health and Social Care (2nd edition). The ethical approval of the RADAR-MDD of the UK site had been obtained in London from the Camberwell St Giles Research Ethics Committee (REC reference: 17/LO/1154). The RAPID study was reviewed by the London Westminster Research Ethics Committee and received approval from the Health Research Authority (reference number 20/LO/0091). All participants of both studies signed informed consent.

2.4. Patient involvement

The RADAR-MDD protocol was co-developed with a patient advisory board that shared their opinions on several user-facing aspects of the study including the choice and frequency of survey measures and speech tasks, the usability of the study app, participant-facing documents, selection of optimal participation incentives, selection, and deployment of wearable device as well as the data analysis plan.

2.5. Analytical workflow

Fig. 1 shows the data analytical workflow in this study. We first utilized Whisper (Radford et al., 2022) to automatically transcribe the speech recordings into texts. Then, we applied the BERTopic model (Grootendorst, 2022) to identify primary topics in transcribed texts. To identify depression risk topics and understand topic contexts, we compared the depression symptom severity and behavioral/linguistic characteristics across identified topics.

2.6. Speech-to-text transcription

Whisper, an OpenAI automatic speech recognition system (ASR) trained on 680,000 h of web-collected audio data, outperformed the state-of-the-art speech recognition systems on public speech datasets covering various domains, tasks, and languages (Radford et al., 2022). We have previously used Whisper to transcribe smartphone-collected speech samples (Dineley et al., 2023). In this paper, we utilized the Whisper Medium model (https://github.com/openai/whisper) to transcribe our free-response speech recordings into texts. After transcription, we manually reviewed the texts and removed abnormalities that were caused by technical or operational issues (e.g., empty recordings).

2.7. Topic modeling

BERTopic (Grootendorst, 2022), a deep learning-based topic model, was chosen to identify principal topics in the transcribed texts due to its contextual understanding and efficiency with short texts (Baird et al., 2022). Then, we assigned an appropriate title to represent each of the identified topics by reviewing keywords generated by BERTopic.

2.8. Comparisons between topics

We summarized and compared depression symptom severity and behavioral/linguistic features across identified topics. Differences between topics were assessed for significance using Kruskal-Wallis tests (McKight and Najab, 2010). These are briefly described below.

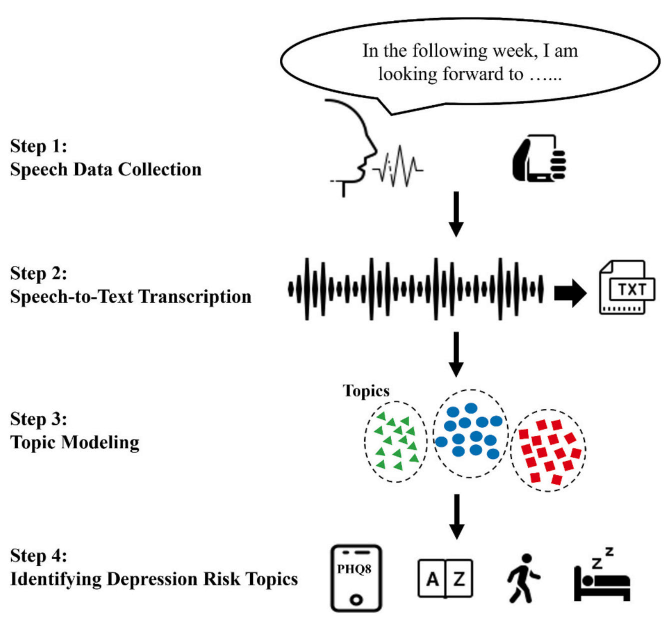

Fig. 1. The data analytical workflow. It contains 4 steps: (1) free-response speech recording collection, (2) speech-to-text transcription via Whisper, (3) topic modeling using BERTopic, (4) comparisons of PHQ8 scores and behavioral/linguistic characteristics across identified topics to identify risk topics for depression.

2.8.1. Depression symptom severity

The participant's depression symptom severity was measured biweekly using the 8-item Patient Health Questionnaire (PHQ-8) (Kroenke et al., 2009) concurrently with the speech task in the RADAR-MDD study. The PHQ-8 score ranges from 0 to 24, indicating increasing depression severity. A PHQ-8 score $\geq$ 10 is the recommended threshold for depression screening (Kroenke et al., 2009). Therefore, the topics mentioned by participants with a median PHQ-8 score $\geq$ 10 were considered as the risk topics for depression; other topics were regarded as non-risk topics.

2.8.2. Behavioral features

Participants were instructed to wear a Fitbit Charge 2/3 wristband during the follow-up period (Matcham et al., 2019). Fitbit's effectiveness in identifying sleep periods and measuring steps has been previously validated (De Zambotti et al., 2018; Feehan et al., 2018; Haghayegh et al., 2019; Paul et al., 2015). Therefore, to understand the potential reasons for topic emergence, from 1-week Fitbit data preceding the speech task, we extracted the following three behavioral features: *Sleep Variability* (standard deviation of sleep duration over a period), *Sleep Onset* (mean sleep onset time over a period), and *Daily Steps* (mean steps per day), as they have been reported to be significantly associated with depression severity (McKercher et al., 2009; Zhang et al., 2021).

2.8.3. Linguistic features

The inclusion of linguistic features is to compare the linguistic differences between risk and non-risk topics and understand the linguistic context of topics. We extracted 20 linguistic features commonly used in previous depression-language studies (Al-Mosaiwi and Johnstone, 2018; Bernard et al., 2016; Landoni et al., 2023; Rathner et al., 2018) using the Linguistic Inquiry and Word Count tool (LIWC-22) (Boyd et al., 2022). These features encompass four perspectives of linguistic characteristics, including summary variables (e.g., word count), personal pronouns, emotional words, and lifestyle, detailed in Supplementary Table 4.

2.9. Topic shifts and changes in depression severity over time

This study also aimed to examine potential associations between speech topic shifts and changes in depression severity. There are four types of topic shifts between two consecutive speech tasks: Risk to Risk, Risk to Non-Risk, Non-Risk to Risk, and Non-Risk to Non-Risk. We utilized the Kruskal-Wallis test (McKight and Najab, 2010) to assess whether differences in PHQ-8 scores between the two speech tasks are significantly different across these four topic shift patterns.

3. Results

3.1. Cohort characteristics

According to our data inclusion criteria, a total of 3919 transcribed speech texts from 265 participants were analyzed in this study. The cohort had a median (IQR) age of 46.00 (31.00, 58.00) years and was predominantly female (78.5 %) and White (88.3 %). Detailed sociodemographics are provided in Supplementary Table 1.

3.2. Topic modeling results

We identified 29 topics in the transcribed texts using the BERTopic model (Table 1). For clear and logical presentation, these topics were manually categorized based on their actual meanings.

Table 1
A summary of 29 topics identified in the free-response speech tasks of the RADAR-MDD dataset.

Topic	Number of Recordings	Number of Participants	Keywords
Nothing to Expect			
No Expectations	302	101	nothing to look forward to
No Expectations due to Covid	65	39	lockdown, restrictions, nothing to look forward to
Social Networks and Activities			
Family	289	123	daughter, son, parents
Friend	171	94	friends, meeting, seeing
Festival	154	101	Christmas, new year, Easter
Celebration	123	82	birthday, celebrating, party
Conversation	26	23	phone, skype, chat
Entertainment and Hobbies			
Art Activity	97	56	rehearsal, pantomime, theatre
Holiday	61	44	holiday, half term, away
Traveling	39	35	scenery, train, trip
Weekend	116	70	weekend, day off, relaxing
Gardening	86	55	gardening, planting, plants
Hobby	81	45	yoga, baking, pottery
Study and Work			
Coursework	72	38	university, course, exams
Studying	27	16	book, reading, writing
Online Meeting	63	36	zoom, quiz, meeting
Working	46	37	job, work, project
Sports			
Fitness	45	34	gym, exercise, fitness
Walking	63	42	walk, walking, going
Outdoor Activity	37	22	ride, climbing, cycling
Swimming	50	29	swimming, pool, sea
Health			
Hospital	98	60	hospital, operation, pain
Mental Therapy	28	17	mental, therapy, NHS
Sleep	36	27	sleep, tired, rest
Other Themes			
Pet	73	49	dog, puppy, cat
Weather	86	54	sunshine, rain, warm
House	90	62	house, decorating, bedroom
Covid-19	132	82	Covid-19, virus, vaccine
Haircut	23	21	haircut, hair, cut

3.2.1. *Nothing to expect*

This category encompasses two topics, namely 'No Expectations' and 'No Expectations due to Covid'. Both topics revealed participants had nothing to look forward to in the coming week. Notably, the 'No Expectations' topic accounted for a considerable proportion, encompassing 302 recordings from 101 participants. The latter topic, 'No Expectations due to Covid', showed that the lack of expectations was attributed to COVID-19 restrictions, as exemplified by "Because of the lockdown and social isolation, I have nothing to look forward to over the next seven days."

3.2.2. *Social networks and activities*

Some participants' expectations for the next week were related to social networks and activities, including topics of 'Family' (keywords: daughter, son, parents), 'Friend' (keywords: friends, meeting, seeing), 'Festival' (keywords: Christmas, new year, Easter), 'Celebration' (keywords: birthday, celebrating, party), and 'Conversation' (keywords: phone, skype, chat).

3.2.3. *Entertainment and hobbies*

The forthcoming entertainment plans of participants were manifest in topics of 'Art Activity' (keywords: rehearsal, pantomime, theatre), 'Holiday' (keywords: holiday, half term, away), 'Traveling' (keywords: scenery, train, trip), and 'Weekend' (keywords: weekend, day off, relaxing). Additionally, the topics of 'Gardening' (keywords: gardening, planting, plants) and 'Hobby' (keywords: yoga, baking, pottery) reflected participants' desires to engage in hobbies during the next week.

3.2.4. *Study and work*

Participants' expectations regarding academic and work-related activities for the upcoming week were reflected in topics of 'Coursework' (keywords: university, course, exams), 'Studying' (keywords: book, reading, writing), 'Online Meeting' (keywords: zoom, quiz, meeting), and 'Working' (keywords: job, work, project).

3.2.5. *Sports*

Four distinct topics were associated with participants' sport-related plans for the coming week, including 'Fitness' (keywords: gym, exercise, fitness), 'Walking' (keywords: walk, walking, going), 'Outdoor Activity' (keywords: ride, climbing, cycling), and 'Swimming' (keywords: swimming, pool, sea).

3.2.6. *Health*

The topics of 'Hospital' (keywords: hospital, operation, pain), 'Mental Therapy' (keywords: mental, therapy, NHS), and 'Sleep' (keywords: sleep, tired, rest) reflect the health conditions of participants.

3.2.7. *Other themes*

Additional topics included subjects such as 'Pet' (keywords: dog, puppy, cat), 'Weather' (keywords: sunshine, rain, warm), 'House' (keywords: room, decorating, bedroom), 'Covid-19' (keywords: Covid-19, virus, vaccine), and 'Haircut' (keywords: haircut, hair, cut).

3.3. *Comparison between topics*

3.3.1. *Depression symptom severity*

We observed significant variability in depression scores (PHQ-8) across these 29 topics (Kruskal-Wallis test: $P < .001$) (displayed in Fig. 2 and Supplementary Table 2). There were six topics with a median PHQ-8 score ≥ 10 (depression screening threshold), including 'No Expectations' (13.00 [8.00, 18.00]), 'Sleep' (13.00 [7.75, 15.25]), 'Mental Therapy' (12.50 [5.00, 15.75]), 'Haircut' (11.00 [7.50, 15.50]), 'Studying' (11.00 [7.50, 14.50]), and 'Coursework' (10.50 [6.75, 14.00]). These six topics were regarded as risk topics for depression. Other topics with median PHQ-8 scores ranging from 5 to 9 were regarded as non-risk topics for depression such as 'Art Activity' (6.00 [3.00, 12.00]), 'Gardening' (6.00 [3.00, 11.00]), 'Holiday' (5.00 [2.00, 9.00]), and 'Outdoor Activity' (5.00 [2.00, 8.00]).

3.3.2. *Behavioral characteristics*

We found that behavioral features, *Sleep Variability*, *Sleep Onset*, and *Daily Steps*, were significantly different across topics (Kruskal-Wallis test: $P < .001$) (Fig. 3 and Supplementary Table 3).

Participants who mentioned risk topics, except for the 'Studying' topic, had higher sleep variability over the preceding week compared to those discussing non-risk topics. Participants discussing the 'Sleep' topic had the highest sleep variability (1.65 [1.33, 1.78] hours), while participants who mentioned the 'Outdoor Activity' topic had the lowest sleep variability (0.67 [0.37, 0.80] hours) (Fig. 3a). For sleep onset time,

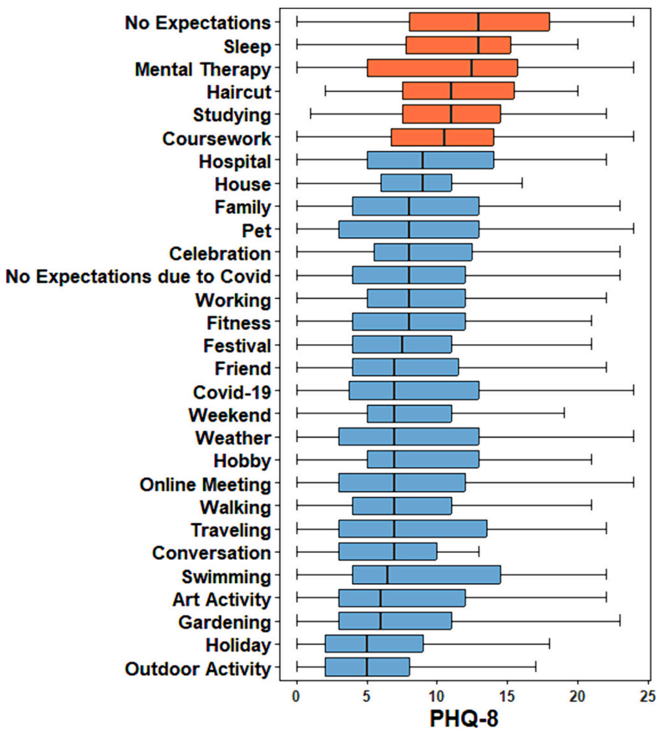

Fig. 2. Boxplots of PHQ-8 scores for 29 topics identified in RADAR-MDD's free-response speech recordings. Note, orange represents risk topics (median PHQ-8 $\geq$ 10) and blue represents non-risk topics (median PHQ-8 $<$ 10). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

participants who mentioned 'Sleep' (01:16 [23:37, 03:19]), 'Coursework' (00:40 [23:18, 01:54]), and 'No Expectations' (00:28 [23:22, 02:03]) topics slept later than participants mentioning the non-risk topics (23:53 [22:58, 00:59]) (Fig. 3b).

Regarding daily activities, participants mentioning risk topics for depression (except for the 'Haircut' topic) had fewer daily steps than participants who talked about non-risk topics. Specifically, participants discussing 'Coursework' (3202.57 [1741.29, 6569.86] steps), 'No Expectations' (3923.00 [1156.50, 6589.07] steps), and 'Sleep' (3899.71 [3036.14, 6951.71] steps) had relatively low daily step count, whereas participants mentioning the 'Holiday' topic had the most daily steps (7747.14 [5269.71, 11,147.29] steps) (Fig. 3c).

3.3.3. Linguistic characteristics

We observed four notable differences (Kruskal-Wallis test: $P < .001$) in *Word Count*, *Negation*, *Leisure*, and *Negative Emotion* between risk and non-risk topics for depression (Fig. 4). Further details of comparisons are provided in Supplementary Table 5.

In the speech tasks, participants discussing 'Mental Therapy' (40.00 [28.50, 66.50] words) and 'Holiday' (36.00 [16.00, 71.00] words) topics spoke a greater number of words, while participants mentioning 'Studying' (13.00 [10.50, 21.50] words) and 'No Expectations' (13.00 [10.00, 23.00] words) topics spoke less (Fig. 4a). Participants expressing the 'No Expectations' topic used the highest percentage of negation terms in their speech tasks (90.1 %) compared to other participants (Fig. 4b). Also, participants used more leisure-related words when they were talking about topics of 'Art Activity' (68.0 %), 'Gardening' (77.9 %), and 'Holiday' (55.7 %), whereas participants who mentioned 'No Expectations' (2.6 %) and 'Haircut' (4.3 %) topics used fewer leisure-related words (Fig. 4c). Furthermore, speech contents of the 'Sleep' and 'Mental Therapy' topics contained more negative emotion words, while speech texts of the 'Studying' and 'Art Activity topics' contained fewer negative emotion words (Fig. 4d).

3.4. Topic shifts and changes in depression severity over time

We observed that the PHQ-8 difference of the Risk to Non-Risk shift was significant (Kruskal-Wallis test: $P < .001$), indicating if the speech topic changed from the risk topic to the non-risk topic, the corresponding PHQ-8 score decreased by 1.1 points on average (Fig. 5a). We also observed that the PHQ-8 scores of two consecutive risk topics (Risk to Risk) were higher than those with only one risk topic in two consecutive speech tasks (Fig. 5b). Additionally, after/before a risk topic, the PHQ-8 was still higher even if the corresponding speech topic was non-risk, compared to those of two consecutive non-risk topics (7.0 [4.0, 12.0]) (Kruskal-Wallis test: $P < .001$) (Fig. 5b).

3.5. Validation on the RAPID dataset

A total of 356 transcribed speech texts from 57 participants were available in the RAPID dataset. We used the established BERTopic model to classify these speech texts into topics (Supplementary Table 6). Note, different from the RADAR-MDD, the RAPID study utilized the 9-item Patient Health Questionnaire (PHQ-9) (Kroenke et al., 2001) for measuring participants' depression severity. Due to the small sample size of the dataset, we only compared PHQ-9 scores across topics comprising a minimum of 10 transcribed speech texts. We found that participants who mentioned 'No Expectations' and 'Sleep' topics had significantly higher median PHQ-9 scores (17.00 [15.00, 19.00] and 18.00 [15.00, 19.00]) compared to those discussing other topics (Kruskal-Wallis test: $P = .006$).

3.6. Results on data before the COVID-19 pandemic

Prior to the first reported COVID-19 case in the UK (31 January 2020), 827 transcribed speech texts from 183 participants were collected. The PHQ-8 scores of topics on the data before the COVID-19 pandemic are summarized in Supplementary Table 7. We found that participants who mentioned 'No Expectations', 'Mental Therapy', and 'Coursework' topics had relatively higher PHQ-8 scores (12.00 [5.50, 17.50], 11.00 [4.00, 15.00] and 12.50 [8.25, 14.75]). However, the sample sizes of transcribed speech texts related to other depression risk topics, including 'Sleep', 'Haircut', and 'Studying' topics, were insufficient for comparison.

4. Discussion

Utilizing automatic speech recognition, topic modeling, and mobile technologies, the present study proposed a data-driven approach for analyzing large-scale spoken language data. To the best of our knowledge, it is the first study to explore depression-related language topics in real-world speech recordings collected from clinical populations in their daily life. Here, we discuss our key findings alongside prior research.

4.1. Risk topics for depression

A notable finding is a considerable proportion of participants with high depression severity expressed having nothing to look forward to in the upcoming week in their free-response speech tasks, which is consistent in both datasets. This could be attributed to hopelessness being a key facet of depression (Abramson et al., 1989; Marchetti, 2019). The incapacity to hope is frequently found in first-person reports of depressed people (Ratcliffe, 2013). Prior literature described such sentiments of depression as "all sense of hope had vanished", "a paralysis of hope", and "no words to explain the depths of my despair" (Brampton, 2018; Schoeneman et al., 2004). Moreover, previous quantitative studies reported a significant correlation between hopelessness (measured by questionnaires) and depression (Hedayati and Khazaei, 2014; Schrank et al., 2008). Thus, the expression of hopelessness for the future may serve as an indicative marker of depression.

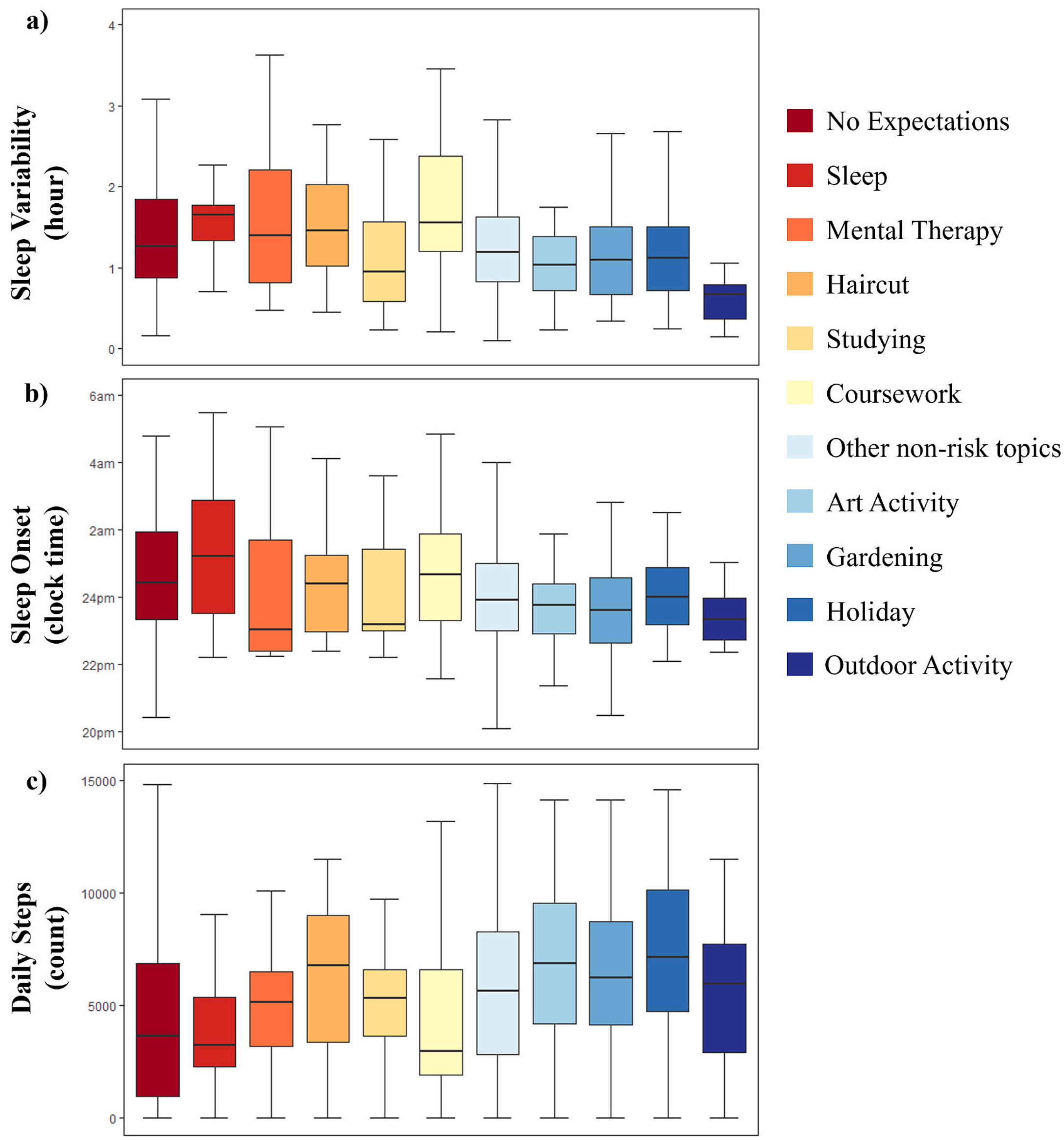

Fig. 3. Boxplots of behavioral features for 6 identified risk topics, 4 non-risk topics with the lowest median PHQ-8 scores ('Art Activity', 'Gardening', 'Holiday', and 'Outdoor Activity'), and other non-risk topics (grouped as a single category).

Participants expressing the expectation of sleep had relatively high depression severity, which is also consistent in both datasets. The potential reason is that participants looking forward to sleep may have suffered from poor sleep or been exhausted before their speech tasks. It is known that sleep disturbances are key manifestations of depression (Alvaro et al., 2013; Smith-Nielsen et al., 2018). Further, fatigue is also a somatic symptom of depression (Demyttenaere et al., 2005). Therefore, mentioning the expectation of sleep may be also an indicator of depression.

Many of our participants who mentioned studying, reading, examinations, and coursework had relatively high depression symptom severity. Several studies reported that workload (Erschens et al., 2019; Kitzrow, 2003; Sprung and Rogers, 2021), exams (Hunt and Eisenberg, 2010; Scholz et al., 2016), and concerns about academic performance (Ishii et al., 2018; Stallman, 2010) are risk factors for stress, anxiety, and depression among college students. Since the 'Mental Therapy' topic is directly related to mental disorders, participants who mentioned this topic were likely experiencing mental health problems. Intriguingly, 21 participants who mentioned the 'Haircut' topic also exhibited relatively high depression severity. Previous research indicates that barbers and hairstylists can be trusted and respected conversationalists for individuals with mental illness to seek help and social support (Cowen et al., 1979; Mbilishaka, 2018; Shabazz, 2016). Nonetheless, due to the absence of detailed reasons for mentioning the 'Haircut' topic and limited sample size, this specific link needs future investigation and validation on large datasets.

4.2. Non-risk topics for depression

Participants discussing topics related to art activities, gardening, holidays, and outdoor activities had the lowest depression severity in our cohort. Engaging in art activities, such as dance, singing, and theatrical performances, has been reported to have beneficial effects on individuals' mental health (Curtis et al., 2018; Fujiwara et al., 2014; Zarobe and Bungay, 2017). Moreover, since exposure to nature can promote feelings of connectedness and relaxation (Pretty, 2004), participating in gardening activities has been linked to reduced depression and anxiety symptoms, as well as increased positive emotions (Adevi and Mårtensson, 2013; Gonzalez et al., 2010). Similarly, holiday trips have been found to contribute to lessening loneliness

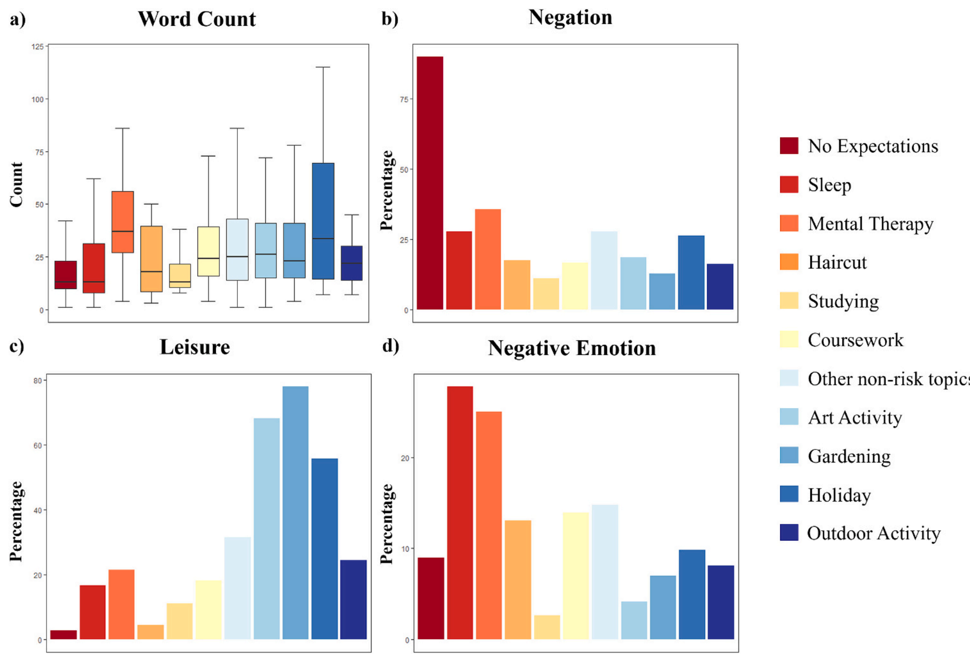

Fig. 4. Comparison of linguistic features of Word Count, Negation, Leisure, and Negative Emotion across 6 risk topics, 4 non-risk topics with the lowest median PHQ-8 scores ('Art Activity', 'Gardening', 'Holiday', and 'Outdoor Activity'), and other non-risk topics (grouped as one category).

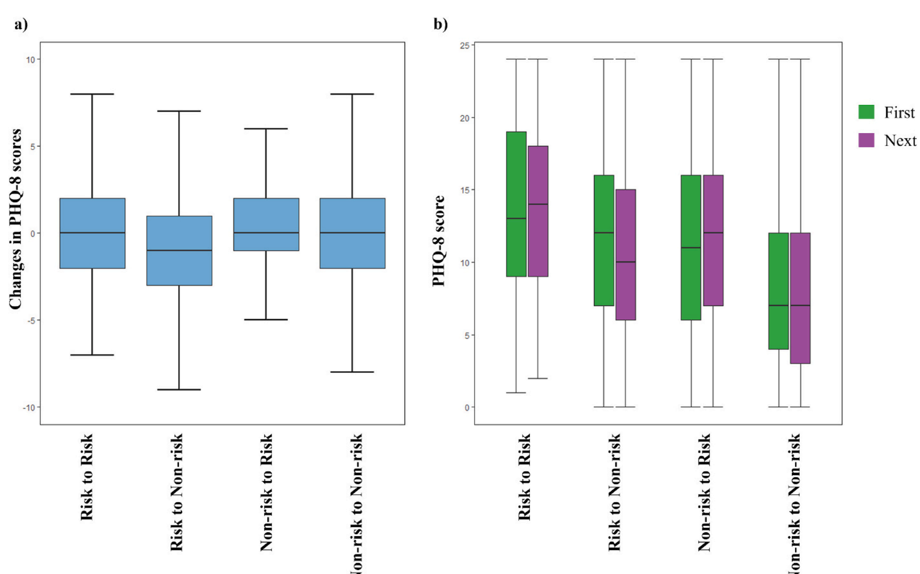

Fig. 5. The correlation between topic shifts and changes in depression severity. a) The boxplots of changes in PHQ-8 scores between two consecutive speech tasks among the 4 topic shifts. b) boxplots of PHQ-8 scores corresponding to two consecutive speech tasks among the 4 topic shifts. Note, there are 4 topic shifts between two consecutive speech tasks: Risk to Risk, Risk to Non-Risk, Non-Risk to Risk, and Non-Risk to Non-Risk.

(Pagan, 2020), enhancing life satisfaction (Pagán, 2015), and alleviating depression (Filep and Bereded-Samuel, 2012). Furthermore, engaging in outdoor activities like walking, running, and cycling, has demonstrated a notable association with a reduced risk of depression (Matias et al., 2022).

4.2. Topics and behavioral characteristics

To the best of our knowledge, this study is the first to link spoken language topics with matching behavioral information as measured by wearable devices. Passive behavioral features extracted from Fitbit recordings could help understand the potential emergence reason for some topics and the depressive status of participants when conducting speech tasks. Using this information, we were able to observe that participants expressing the 'Sleep' topic had the highest sleep variability and the latest sleep onset time during the week before their speech tasks, indicating that they were more likely to be experiencing unstable sleep or insomnia (Zhang et al., 2021). Also, participants who mentioned 'Coursework' had late and disrupted sleep during the past week, likely due to academic workload and study pressure. In contrast, participants who expressed non-risk topics slept relatively better, indicating that their mental health was likely in better status. Regarding daily activities, participants expressing 'No Expectations', 'Sleep', and 'Coursework' topics had fewer daily step counts than those mentioning other topics which may be caused by a lack of motivation, fatigue, and pressure of academic workload, respectively. The low activity levels of these participants indicated that they may have had relatively high severity of depression symptoms before speech tasks (McKercher et al., 2009).

4.4. Topics and linguistic characteristics

The linguistic characteristics of the transcribed speech texts could provide additional insights into the participants' mental status (e.g., emotion) during the speech task. We observed several significant differences in linguistic features across topics, which provide support to our findings. In one of our previous studies, we found participants with higher depression severity tended to speak less in their speech tasks (Cummins et al., 2023). This may explain why the speech texts of some risk topics ('No Expectations', 'Sleep', and 'Studying') contain fewer words than those of non-risk topics. Furthermore, prior research shows that engagement in leisure activities can be a protective factor against depression (Bone et al., 2022; Ponde and Santana, 2000). This aligns with our observation that participants discussing non-risk topics used more leisure-related words than those mentioning risk topics. Additionally, previous studies have linked elevated use of negative emotion words with higher depression severity (Al-Mosaiwi and Johnstone, 2018; Landoni et al., 2023; Yang et al., 2023). In our data, participants who mentioned 'Sleep' and 'Mental Therapy' topics used more negative emotional words than other topics, potentially indicating their depressed states. Moreover, participants who mentioned the 'No Expectations' topic were more likely to use negated words (such as nothing, don't, and can't) when expressing that they had nothing to look forward to in the next week.

4.5. Topic shifts and changes in depression severity over time

To the best of our knowledge, there is currently a lack of longitudinal research tracking how the changes in spoken language topics relate to fluctuations in depression severity over time. Our findings indicate that risk topics may reflect the depressive status of a long period and frequently discussing them may reflect severe depression severity, which highlights the importance of longitudinally monitoring language.

4.6. Limitations

Our findings should be interpreted in the context of certain limitations. First, our participants had a history of major depression (RADAR-MDD) or were undergoing depression treatment (RAPID), along with a predominantly white and female demographic, which may limit the generalizability of our findings to non-depressed or more diverse populations. Second, our findings were based on a specific speech task, describing one's expectations for the upcoming week, which may limit our generalizability to other free-response speech tasks. Third, some topics' sample sizes were small and required validation in larger corpora. Fourth, although we obtained some consistent results on the pre-COVID data, the majority of speech recordings were collected during the COVID-19 pandemic, which could have influenced the findings. Therefore, our findings need to be validated on post-COVID datasets. Fifth, the transcription of spoken language was performed with a pre-trained deep learning model that may contain errors. Depression can affect the acoustic properties of speech (Cummins et al., 2015), potentially impacting the accuracy of the transcripts, which needs further investigation. Sixth, the co-occurrence of depression with anxiety may have influenced some identified risk topics (such as 'Studying' and 'Coursework' topics). Future research should concurrently collect anxiety questionnaires for controlling participants' anxiety levels to obtain more specific depression-related risk topics.

5. Conclusions

Our findings demonstrate specific speech topics may indicate depression severity, though further validation is needed. The presented data-driven workflow provides a practical approach for analyzing large-scale speech data collected from real-world settings for digital health research.

Code availability

The complete code used for the analysis can be made available through reasonable requests to the RADAR-CNS consortium. Please email the corresponding author for details.

CRediT authorship contribution statement

Yuezhou Zhang: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Amos A. Folarin:** Writing – review & editing, Software, Project administration, Data curation, Conceptualization. **Judith Dineley:** Writing – review & editing, Data curation, Conceptualization. **Pauline Conde:** Writing – review & editing, Software, Data curation. **Valeria de Angel:** Writing – review & editing, Data curation. **Shaoxiong Sun:** Writing – review & editing, Methodology. **Yatharth Ranjan:** Writing – review & editing, Software. **Zulqarnain Rashid:** Writing – review & editing, Software. **Callum Stewart:** Writing – review & editing, Software. **Petroula Laiou:** Writing – review & editing, Methodology. **Heet Sankesara:** Writing – review & editing, Software. **Linglong Qian:** Writing – review & editing, Methodology. **Faith Matcham:** Writing – review & editing, Project administration, Data curation. **Katie White:** Writing – review & editing, Data curation. **Carolin Oetzmann:** Writing – review & editing, Data curation. **Femke Lamers:** Writing – review & editing, Project administration, Data curation. **Sara Siddi:** Writing – review & editing, Project administration, Data curation. **Sara Simblett:** Writing – review & editing. **Björn W. Schuller:** Writing – review & editing, Project administration, Funding acquisition. **Srinivasan Vairavan:** Writing – review & editing, Project administration. **Til Wykes:** Writing – review & editing, Project administration, Funding acquisition. **Josep Maria Haro:** Writing – review & editing, Project administration, Funding acquisition. **Brenda W.J.H. Penninx:** Writing – review & editing. **Vaibhav A. Narayan:** Writing – review & editing, Project administration, Funding acquisition. **Matthew Hotopf:** Writing – review & editing, Project administration, Funding acquisition. **Richard J.B. Dobson:** Writing – review & editing, Software, Project administration, Methodology, Funding acquisition. **Nicholas Cummins:** Writing – review & editing, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

S.V. and V.A.N. are employees of Janssen Research and Development LLC. M.H. is the principal investigator of the Remote Assessment of Disease and Relapse–Central Nervous System project, a private public precompetitive consortium that receives funding from Janssen, UCB, Lundbeck, MSD, and Biogen.

Data availability

The datasets used for the present study can be made available through reasonable requests to the RADAR-CNS consortium. Please email the corresponding author for details.

Acknowledgments

The RADAR-CNS project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115902. This Joint Undertaking receives support from the European

Union's Horizon 2020 research and innovation programme and EFPIA, www.imi.europa.eu. This communication reflects the views of the RADAR-CNS consortium and neither IMI nor the European Union and EFPIA are liable for any use that may be made of the information contained herein. The funding body has not been involved in the design of the study, the collection or analysis of data, or the interpretation of data.

This paper represents independent research part-funded by the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author (s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

We thank all the members of the RADAR-CNS patient advisory board for their contribution to the device selection procedures, and their invaluable advice throughout the study protocol design.

This research was reviewed by a team with experience of mental health problems and their careers who have been specially trained to advise on research proposals and documentation through the Feasibility and Acceptability Support Team for Researchers (FAST-R): a free, confidential service in England provided by the National Institute for Health Research Maudsley Biomedical Research Centre via King's College London and South London and Maudsley NHS Foundation Trust.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2024.03.106.

References

Abramson, L.Y., Metalsky, G.I., Alloy, L.B., 1989. Hopelessness depression: a theory-based subtype of depression. Psychol. Rev. 96, 358.
Adevi, A.A., Mårtensson, F., 2013. Stress rehabilitation through garden therapy: the garden as a place in the recovery from stress. Urban For. Urban Green. 12, 230–237.
Al-Mosaiwi, M., Johnstone, T., 2018. In an absolute state: elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. Clin. Psychol. Sci. 6, 529–542.
Althubaiti, A., 2016. Information bias in health research: definition, pitfalls, and adjustment methods. J. Multidiscip. Healthc. 211–217.
Alvaro, P.K., Roberts, R.M., Harris, J.K., 2013. A systematic review assessing bidirectionality between sleep disturbances, anxiety, and depression. Sleep 36, 1059–1068.
Association, I.P, 1999. Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet. Cambridge University Press.
Baikie, K.A., Wilhelm, K., Johnson, B., Boskovic, M., Wedgwood, L., Finch, A., et al., 2006. Expressive writing for high-risk drug dependent patients in a primary care clinic: a pilot study. Harm Reduct. J. 3, 1–8.
Baird, A., Xia, Y., Cheng, Y., 2022. Consumer perceptions of telehealth for mental health or substance abuse: a Twitter-based topic modeling analysis. JAMIA open 5, ooac028.
Bernard, J.D., Baddeley, J.L., Rodriguez, B.F., Burke, P.A., 2016. Depression, language, and affect: an examination of the influence of baseline depression and affect induction on language. J. Lang. Soc. Psychol. 35, 317–326.
Bone, J.K., Bu, F., Fluharty, M.E., Paul, E., Sonke, J.K., Fancourt, D., 2022. Engagement in leisure activities and depression in older adults in the United States: longitudinal evidence from the health and retirement study. Soc. Sci. Med. 294, 114703.
Boyd, R.L., Ashokkumar, A., Seraj, S., Pennebaker, J.W., 2022. The Development and Psychometric Properties of LIWC-22. University of Texas at Austin, Austin, TX, pp. 1–47.
Brampton, S., 2018. Shoot the damn dog: a memoir of depression. Bloomsbury publishing.
Cook, B.L., Progovac, A.M., Chen, P., Mullin, B., Hou, S., Baca-Garcia, E., 2016. Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. In: Computational and mathematical methods in medicine 2016.
Cowen, E.L., Gesten, E.L., Boike, M., Norton, P., Wilson, A.B., DeStefano, M.A., 1979. Hairdressers as caregivers. I. A descriptive profile of interpersonal help-giving involvements. Am. J. Community Psychol. 7, 633.
Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., Quatieri, T.F., 2015. A review of depression and suicide risk assessment using speech analysis. Speech Comm. 71, 10–49.
Cummins, N., Dineley, J., Conde, P., Matcham, F., Siddi, S., Lamers, F., et al., 2023. Multilingual markers of depression in remotely collected speech samples: a preliminary analysis. J. Affect. Disord. 341, 128–136.
Curtis, A., Gibson, L., O'Brien, M., Roe, B., 2018. Systematic review of the impact of arts for health activities on health, wellbeing and quality of life of older people living in care homes. Dementia 17, 645–669.
De Angel, V., Lewis, S., Munir, S., Matcham, F., Dobson, R., Hotopf, M., 2022a. Using digital health tools for the Remote Assessment of Treatment Prognosis in Depression (RAPID): a study protocol for a feasibility study. BMJ Open 12, e059258.
De Angel, V., Lewis, S., White, K., Oetzmann, C., Leightley, D., Oprea, E., et al., 2022b. Digital health tools for the passive monitoring of depression: a systematic review of methods. NPJ digital medicine 5, 3.
De Angel, V., Adeleye, F., Zhang, Y., Cummins, N., Munir, S., Lewis, S., et al., 2023. The feasibility of implementing remote measurement technologies in psychological treatment for depression: mixed methods study on engagement. JMIR Mental Health 10, e42866.
De Zambotti, M., Goldstone, A., Claudatos, S., Colrain, I.M., Baker, F.C., 2018. A validation study of Fitbit charge 2TM compared with polysomnography in adults. Chronobiol. Int. 35, 465–476.
Demyttenaere, K., De Fruyt, J., Stahl, S.M., 2005. The many faces of fatigue in major depressive disorder. Int. J. Neuropsychopharmacol. 8, 93–105.
Devaux, M., Sassi, F., 2016. Social disparities in hazardous alcohol use: self-report bias may lead to incorrect estimates. The European Journal of Public Health 26, 129–134.
Dineley, J., Lavelle, G., Leightley, D., Matcham, F., Siddi, S., Peñarrubia-María, M.T., et al., 2021. Remote smartphone-based speech collection: acceptance and barriers in individuals with major depressive disorder. arXiv preprint arXiv:2104.08600.
Dineley, J., Carr, E., Matcham, F., Downs, J., Dobson, R., Quatieri, T.F., et al., 2023. Towards robust paralinguistic assessment for real-world mobile health (mHealth) monitoring: an initial study of reverberation effects on speech. arXiv preprint arXiv:2305.12514.
Eichstaedt, J.C., Smith, R.J., Merchant, R.M., Ungar, L.H., Crutchley, P., Preoţiuc-Pietro, D., et al., 2018. Facebook language predicts depression in medical records. Proc. Natl. Acad. Sci. 115, 11203–11208.
Erschens, R., Keifenheim, K.E., Herrmann-Werner, A., Loda, T., Schwille-Kiuntke, J., Bugaj, T.J., et al., 2019. Professional burnout among medical students: systematic literature review and meta-analysis. Med. Teach. 41, 172–183.
Fast, L.A., Funder, D.C., 2010. Gender differences in the correlates of self-referent word use: authority, entitlement, and depressive symptoms. J. Pers. 78, 313–338.
Feehan, L.M., Geldman, J., Sayre, E.C., Park, C., Ezzat, A.M., Yoo, J.Y., et al., 2018. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth 6, e10527.
Filep, S., Bereded-Samuel, E., 2012. Holidays against depression? An Ethiopian Australian initiative. Curr. Issue Tour. 15, 281–285.
Fujiwara, D., Kudrna, L., Dolan, P., 2014. Quantifying and Valuing the Wellbeing Impacts of Culture and Sport. Department for Culture Media and Sport Research Paper.
Gonzalez, M.T., Hartig, T., Patil, G.G., Martinsen, E.W., Kirkevold, M., 2010. Therapeutic horticulture in clinical depression: a prospective study of active components. J. Adv. Nurs. 66, 2002–2013.
Grootendorst, M., 2022. BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.
Haghayegh, S., Khoshnevis, S., Smolensky, M.H., Diller, K.R., Castriotta, R.J., 2019. Accuracy of wristband Fitbit models in assessing sleep: systematic review and meta-analysis. J. Med. Internet Res. 21, e16273.
Hawton, K., i Comabella, C.C., Haw, C., Saunders, K., 2013. Risk factors for suicide in individuals with depression: a systematic review. J. Affect. Disord. 147, 17–28.
Hedayati, M.M., Khazaei, M.M., 2014. An investigation of the relationship between depression, meaning in life and adult hope. Procedia Soc. Behav. Sci. 114, 598–601.
Holmes, D., Alpers, G.W., Ismailji, T., Classen, C., Wales, T., Cheasty, V., et al., 2007. Cognitive and emotional processing in narratives of women abused by intimate partners. Violence Against Women 13, 1192–1205.
Hunt, J., Eisenberg, D., 2010. Mental health problems and help-seeking behavior among college students. J. Adolesc. Health 46, 3–10.
Ishii, T., Tachikawa, H., Shiratori, Y., Hori, T., Aiba, M., Kuga, K., et al., 2018. What kinds of factors affect the academic outcomes of university students with mental disorders? A retrospective study based on medical records. Asian J. Psychiatr. 32, 67–72.
Kessler, R.C., Chiu, W.T., Demler, O., Walters, E.E., 2005. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch. Gen. Psychiatry 62, 617–627.
Kitzrow, M.A., 2003. The mental health needs of today's college students: challenges and recommendations. J. Stud. Aff. Res. Pract. 41, 167–181.
Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16, 606–613.
Kroenke, K., Strine, T.W., Spitzer, R.L., Williams, J.B., Berry, J.T., Mokdad, A.H., 2009. The PHQ-8 as a measure of current depression in the general population. J. Affect. Disord. 114, 163–173.
Landoni, M., Silverio, S.A., Ciuffo, G., Daccò, M., Petrovic, M., Di Blasio, P., et al., 2023. Linguistic features of postpartum depression using linguistic inquiry and word count text analysis. J. Neonatal Nurs. 29, 127–134.
Lenox-Smith, A., Macdonald, M.T., Reed, C., Tylee, A., Peveler, R., Quail, D., et al., 2013. Quality of life in depressed patients in UK primary care: the FINDER study. Neurology and therapy 2, 25–42.
Lerner, D., Adler, D.A., Chang, H., Berndt, E.R., Irish, J.T., Lapitsky, L., et al., 2004. The clinical and occupational correlates of work productivity loss among employed patients with depression. Journal of occupational and environmental medicine/ American College of Occupational and Environmental Medicine 46, S46.
Lewinsohn, P.M., Solomon, A., Seeley, J.R., Zeiss, A., 2000. Clinical implications of" subthreshold" depressive symptoms. J. Abnorm. Psychol. 109, 345.
Li, T.M.H., Chen, J., Law, F.O.C., Li, C.-T., Chan, N.Y., Chan, J.W.Y., et al., 2023. Detection of suicidal ideation in clinical interviews for depression using natural

language processing and machine learning: cross-sectional study. JMIR Med. Inform. 11, e50221.
Losada, D.E., Crestani, F., 2016. A test collection for research on depression and language use, Experimental IR Meets Multilinguality, Multimodality, and Interaction: 7th International Conference of the CLEF Association, CLEF 2016, Évora, Portugal, September 5-8, 2016, Proceedings 7. Springer, pp. 28–39.
Marchetti, I., 2019. Hopelessness: a network analysis. Cogn. Ther. Res. 43, 611–619.
Matcham, F., di San, Barattieri, Pietro, C., Bulgari, V., De Girolamo, G., Dobson, R., Eriksson, H., et al., 2019. Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): a multi-centre prospective cohort study protocol. BMC Psychiatry 19, 1–11.
Matcham, F., Leightley, D., Siddi, S., Lamers, F., White, K.M., Annas, P., et al., 2022. Remote Assessment of Disease and Relapse in Major Depressive Disorder (RADAR-MDD): recruitment, retention, and data availability in a longitudinal remote measurement study. BMC Psychiatry 22, 136.
Matias, T.S., Lopes, M.V.V., da Costa, B.G.G., Silva, K.S., Schuch, F.B., 2022. Relationship between types of physical activity and depression among 88,522 adults. J. Affect. Disord. 297, 415–420.
Matton, K., McInnis, M.G., Provost, E.M., 2019. Into the Wild: Transitioning from Recognizing Mood in Clinical Interactions to Personal Conversations for Individuals with Bipolar Disorder, Interspeech.
Mbilishaka, A., 2018. PsychoHairapy: using hair as an entry point into Black women's spiritual and mental health. Meridians 16, 382–392.
McKercher, C.M., Schmidt, M.D., Sanderson, K.A., Patton, G.C., Dwyer, T., Venn, A.J., 2009. Physical activity and depression in young adults. Am. J. Prev. Med. 36, 161–164.
McKight, P.E., Najab, J., 2010. Kruskal-wallis test. The corsini encyclopedia of psychology 1.
Mundt, J.C., Snyder, P.J., Cannizzaro, M.S., Chappie, K., Geralts, D.S., 2007. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. J. Neurolinguistics 20, 50–64.
Nanomi Arachchige, I.A., Sandanapitchai, P., Weerasinghe, R., 2021. Investigating machine learning & natural language processing techniques applied for predicting depression disorder from online support forums: a systematic literature review. Information 12, 444.
Pagán, R., 2015. The contribution of holiday trips to life satisfaction: the case of people with disabilities. Curr. Issue Tour. 18, 524–538.
Pagan, R., 2020. How important are holiday trips in preventing loneliness? Evidence for people without and with self-reported moderate and severe disabilities. Curr. Issue Tour. 23, 1394–1406.
Paul, S.S., Tiedemann, A., Hassett, L.M., Ramsay, E., Kirkham, C., Chagpar, S., et al., 2015. Validity of the Fitbit activity tracker for measuring steps in community-dwelling older adults. BMJ Open Sport—Exercise Medicine 1.
Pennebaker, J.W., Mehl, M.R., Niederhoffer, K.G., 2003. Psychological aspects of natural language use: our words, our selves. Annu. Rev. Psychol. 54, 547–577.
Ponde, M.P., Santana, V.S., 2000. Participation in leisure activities: is it a protective factor for women's mental health? J. Leis. Res. 32, 457–472.
Pretty, J., 2004. How nature contributes to mental and physical health. Spiritual. Health Int. 5, 68–78.
Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C., Sutskever, I., 2022. Robust speech recognition via large-scale weak supervision. arXiv preprint arXiv: 2212.04356.
Ranjan, Y., Rashid, Z., Stewart, C., Conde, P., Begale, M., Verbeeck, D., et al., 2019. RADAR-base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. JMIR Mhealth Uhealth 7, e11734.
Ratcliffe, M., 2013. What is it to lose hope? Phenomenol. Cogn. Sci. 12, 597–614.
Rathner, E.-M., Djamali, J., Terhorst, Y., Schuller, B., Cummins, N., Salamon, G., et al., 2018. How Did you Like 2017? Detection of Language Markers of Depression and Narcissism in Personal Narratives.
Rodriguez, A.J., Holleran, S.E., Mehl, M.R., 2010. Reading between the lines: the lay assessment of subclinical depression from written self-descriptions. J. Pers. 78, 575–598.
Rohani, D.A., Faurholt-Jepsen, M., Kessing, L.V., Bardram, J.E., 2018. Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: systematic review. JMIR Mhealth Uhealth 6, e9691.
Rude, S., Gortner, E.-M., Pennebaker, J., 2004. Language use of depressed and depression-vulnerable college students. Cognit. Emot. 18, 1121–1133.
Schoeneman, T.J., Schoeneman, K.A., Stallings, S., 2004. "The black struggle": metaphors of depression in Styron's darkness visible. J. Soc. Clin. Psychol. 23, 325–346.
Scholz, M., Neumann, C., Ropohl, A., Paulsen, F., Burger, P., 2016. Risk factors for mental disorders develop early in German students of dentistry. Annals of Anatomy-Anatomischer Anzeiger 208, 204–207.
Schrank, B., Stanghellini, G., Slade, M., 2008. Hope in psychiatry: a review of the literature. Acta Psychiatr. Scand. 118, 421–433.
Shabazz, D.L., 2016. Barbershops as cultural forums for African American males. J. Black Stud. 47, 295–312.
Skaik, R., Inkpen, D., 2020. Using social media for mental health surveillance: a review. ACM Computing Surveys (CSUR) 53, 1–31.
Smith-Nielsen, J., Matthey, S., Lange, T., Væver, M.S., 2018. Validation of the Edinburgh postnatal depression scale against both DSM-5 and ICD-10 diagnostic criteria for depression. BMC Psychiatry 18, 1–12.
Sonnenschein, A.R., Hofmann, S.G., Ziegelmayer, T., Lutz, W., 2018. Linguistic analysis of patients with mood and anxiety disorders during cognitive behavioral therapy. Cogn. Behav. Ther. 47, 315–327.
Sprung, J.M., Rogers, A., 2021. Work-life balance as a predictor of college student anxiety and depression. J. Am. Coll. Health 69, 775–782.
Stallman, H.M., 2010. Psychological distress in university students: a comparison with general population data. Aust. Psychol. 45, 249–257.
Tadesse, M.M., Lin, H., Xu, B., Yang, L., 2019. Detection of depression-related posts in reddit social media forum. IEEE Access 7, 44883–44893.
Tølbøll, K.B., 2019. Linguistic features in depression: a meta-analysis. Journal of Language Works-Sprogvidenskabeligt Studentertidsskrift 4, 39–59.
Wallach, H.M., 2006. Topic modeling: beyond bag-of-words. In: Proceedings of the 23rd international conference on Machine learning, pp. 977–984.
Yang, C., Zhang, X., Chen, Y., Li, Y., Yu, S., Zhao, B., et al., 2023. Emotion-dependent language featuring depression. J. Behav. Ther. Exp. Psychiatry 81, 101883.
Zarobe, L., Bungay, H., 2017. The role of arts activities in developing resilience and mental wellbeing in children and young people a rapid review of the literature. Perspect. Public Health 137, 337–347.
Zhang, Y., Folarin, A.A., Sun, S., Cummins, N., Bendayan, R., Ranjan, Y., et al., 2021. Relationship between major depression symptom severity and sleep collected using a wristband wearable device: multicenter longitudinal observational study. JMIR Mhealth Uhealth 9, e24604.
Zimmermann, J., Wolf, M., Bock, A., Peham, D., Benecke, C., 2013. The way we refer to ourselves reflects how we relate to others: associations between first-person pronoun use and interpersonal problems. J. Res. Pers. 47, 218–225.
Zimmermann, J., Brockmeyer, T., Hunn, M., Schauenburg, H., Wolf, M., 2017. First-person pronoun use in spoken language as a predictor of future depressive symptoms: preliminary evidence from a clinical sample of depressed patients. Clin. Psychol. Psychother. 24, 384–391.