Information & Management 57 (2020) 103349

Contents lists available at ScienceDirect

Information & Management

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

What reveals about depression level? The role of multimodal features at the level of interview questions

Guohou Shan[a], Lina Zhou[b,*], Zhang Dongsong[b]

[a] Department of Information Systems University of Maryland, Baltimore County 1000 Hilltop Circle, Baltimore, MD 21250, United States
[b] Department of Business Information Systems and Operations Management, University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC, 28223, United States

ARTICLE INFO

Keywords:
Depression detection
Interview question
Multimodal feature
Sensitivity analysis
Question category

ABSTRACT

Early depression detection can enable timely intervention. Automatic depression detection has relied on features extracted from individual-level data, which may be too coarse to support effective detection. Existing detection models have largely overlooked interview questions commonly used in clinical depression assessment. This research proposes a two-layered multi-modal model for depression detection, which not only extracts features from responses at a level of individual interview questions, but also identifies semantic categories of those questions. The evaluation results demonstrate that the proposed model outperforms the state-of-the-art methods for depression detection. The research findings have broad and cross-disciplinary implications.

1. Introduction

Depression is a common and serious mental disorder that severely affects how people feel, think, and handle daily activities [1]. According to Anxiety and Depression Association of America[1], major depressive disorder affects 16.1 million American adults and it is the leading cause of disability for people with ages between 15 and 44 in the U.S. Depression is dangerous, and can result in severe consequences if not detected and treated in a timely manner. According to WHO, 90 % of people who die from suicide have depression [2]. Depression also imposes a significant economic burden on patients and hospitals. The total economic burden of major depression disorder was estimated to be $210.5 billion per year [3].

Early detection of depression is the key to the prevention of its adverse consequences [4]. The predominant methods for depression assessment used in current clinical practice include interviews and specially designed questionnaires such as Patient Health Questionnaire-9 (PHQ-9) [5]. Fully structured interviews are neither theoretically adequate nor easily accessible for obtaining psycho-diagnostic information [6]. Pre-defined questionnaires (e.g., PHQ-9) suffer from similar problems. In addition, the modes of questionnaire administration can have serious biasing effects on data quality [7]. Semi-structured interviews allow clinicians to probe for more details about specific areas pertaining to an individual's depression state, but they rely heavily on the experience and skills of clinicians or non-clinicians who are specially trained on how to observe, interpret, and assess patients' verbal and nonverbal expressions. Such expertise may not be easily accessible and scalable. Automatic depression detection provides a promising solution to the above-mentioned problems.

Despite recent advances of automatic depression detection owing to the advanced information technologies, such as social media (e.g., tweets), wearable sensors, and/or smartphones [8–10], existing methods for depression assessment have several limitations. For instance, sensor- or smartphone-based depression detection methods track and analyze a user's daily behavior (e.g., interactions with one's smartphone), which introduces privacy concerns [11]. Social media content such as tweets and blogs may not be rich enough to capture manifestations of depression [12]. More importantly, the automated depression detection models overlook the mainstream method used in clinical depression screening—interviews, and consequently may miss important signals of depression and induce significant resistance to adoption.

Given the unique and complementary strengths of automated depression detection and clinical depression assessment, combining the two has potential to bring out the best in both. Despite a lack of well-defined and universally accepted characteristics of individuals with depression, it has been widely recognized that depressed people tend to speak anxiously in short phrases and monotonously, avoid eye contact, and be less engaged in verbal communication than those without depression [13]. The behavioral characteristics of individuals with

* Corresponding author.
E-mail addresses: gshan2@umbc.edu (S. Guohou), lzhou8@uncc.edu (Z. Lina), dzhang15@uncc.edu (Z. Dongsong).
[1] https://adaa.org/about-adaa/press-room/facts-statistics

https://doi.org/10.1016/j.im.2020.103349
Received 18 January 2019; Received in revised form 24 July 2020; Accepted 25 July 2020
Available online 31 July 2020
0378-7206/

depression [14,15] can be grouped into verbal (e.g., linguistic style), visual (e.g., facial expressions and body gestures), and vocal (e.g., sound pitch and vocal frequency) categories. These behaviors can manifest themselves in multimodal communication, including vocal, visual, and verbal modalities. However, most of the prior models (e.g. [16–19],) have focused on extracting and using features from a single modality (e.g., audio or text). This study aims to improve the effectiveness of depression detection by combining a wide spectrum of multimodal features of an individual's behavior.

Despite that a few recent studies have explored multimodal data for depression detection (e.g., [20,21]), they developed machine learning models for depression detection by extracting features at an individual level. None of them has considered extracting features at a level of interview questions. To make a clinical or psychiatric interview effective, it is essential for a psychoanalyst to establish connection (e.g., empathy and therapeutic connection) with a patient [22]. In a semi-structured psychiatric interview, questions are contextually adapted to a patient's narrative in order to ask for more details or further examples [6]. As a result, not all questions in an interview are equally important to the detection of the interviewee's depression state. An interviewee may consciously or unconsciously reveal signs of depression through vocal, visual, and/or verbal characteristics of his/her responses to some interview questions but not others. Therefore, analyzing interviewees' responses at a question level provides an opportunity to identify finer grained behavioral characteristics of depression.

This research integrates psychiatric interviews and machine learning models for detection of depression. We propose a method that extracts multimodal features (i.e., vocal, visual, and verbal features) from an interviewee's responses to each interview question, which helps differentiate important questions from unimportant ones. In addition, we segment interview questions into nine categories based on their semantic concentration to examine the efficacy of different types of interview questions for depression detection. Drawing on the semantic categorization of interview questions, we build a novel two-layered machine learning model for depression detection. The first layer of the model detects depression at an interview question level, and the second layer detects at an individual level. The empirical evaluation results demonstrate that the proposed method outperforms the state-of-the-art methods for depression detection. The findings also offer practical implications for enabling early detection of depression and improving the design of interview questions for depression assessment.

The rest of this paper is organized as follows. We first introduce background and review related work on depression detection. We then describe interview questions, multimodal features, question categorization, and the proposed automated depression detection model. Next, we present the experimental design and results. Finally, we discuss research contributions, practical implications, and limitations of this study.

2. Background and related work

In this section, we first categorize depression behavior, and then introduce depression assessment questionnaires, which play a key role in clinical interviews for psychological assessment. Finally, we summarize studies on automatic depression detection.

2.1. Depression behavior

Research has associated depressed individuals with various psychological characteristics. For instance, adults with chronic depression are found to have a higher level of neuroticism, hopelessness, aggression, risk aversion, and rumination, but a lower level of extraversion, agreeableness, and conscientiousness, compared with non-depressed adults [23]. Some studies also show that failure of emotion regulation and immature personality contribute to high depression risks [24,25]. These psychological characteristics of individuals who experience depression have their behavioral manifestations.

Individuals with depression behave differently from those without it while communicating with others [13]. There is a host of studies on general depression behaviors (e.g. [23,26,27],). Based on communication modality, we classify depression behaviors into verbal, visual, and vocal categories.

- Verbal behavior, which is directly related to spoken or written language [28]. For example, people with depression tend to engage less in verbal communication and speak in short phrases [13].
- Visual behavior, which includes body postures, gestures, facial expression, and eye movement, etc. [28]. People with depression tend to avoid eye contact [29], generally have fewer animated facial expressions [29,30], and are more likely to hold their heads in a downward position and engage in self-touching (e.g., rubbing, scratching) than non-depressed counterparts [31,32].
- Vocal behavior includes voice or paralinguistic behaviors that cover all features of speech not directly related to verbal messages [28]. Those features consist of articulatory signals such as speech speed, pitch, volume, and silence; acoustic signals that result from the sound production process taking place prior to speech articulation; and vocalizations such as variations in pitch, volume, and sounds that are used as speech fillers [33]. Vocal features such as fundamental speech frequency and pause duration were found to be highly correlated with the Hamilton Rating Scale for Depression (HAMD)-17 depression score [34]. Compared with individuals without depression, the rate of change in voice frequency of individuals with depression is lower, and the pauses between an interviewer's questions and an interviewee's responses are longer [35]. People with depression tend to speak monotonously [13]. In addition, depressed individuals consistently demonstrate prosodic speech abnormalities, such as reduced variation in loudness, repetitious pitch inflections and stress patterns, monotonous pitch, and loudness [36].

The above-mentioned behavioral characteristics of individuals with depression can inform the design of automatic depression detection methods.

2.2. Depression assessment questionnaires

Based on the mode of administration, we classify depression assessment questionnaires into two categories: clinician-administration and self-administration, as shown in Table 1.

HAMD [37] and Montgomery-Åsberg Depression Rating Scale (MARSD) [38] require clinical professionals to administer an interview.

Table 1
Depression Assessment Questionnaires.

Type	Questionnaires	Acronyms	# of Items	Complete Time (min)
clinician-administration	Hamilton Rating Scale for Depression [37]	HAMD	21	20–30
	Montgomery-Åsberg Depression Rating Scale [38]	MARSD	10	20–30
self-administration	Beck Depression Inventory [39]	BDI	21	5–10
	Quick Inventory of Depressive Symptomatology [40]	QIDS	16	5–10
	Patient Health Questionnaire [41]	PHQ-9	9	< 5

HAMD aims to determine a psychiatric patient's levels of depression before, during, and after treatment, and MARSD is used to measure the severity of depressive episodes of patients with mood disorders in clinical assessment. In contrast, Beck Depression Index (BDI), Quick Inventory of Depressive Symptomatology (QIDS), and PHQ-9 are self-administered questionnaires. BDI is used for individuals aged 13 or above [39]; QIDS allows patients to observe trends of depressive symptoms over time [40]; and PHQ-9 is a self-screening and diagnostic tool for mental health disorders, including depression, anxiety, alcohol, eating, and somatoform disorders [41].

The clinician-administered questionnaires through interviews are considered more effective than self-administered ones for two main reasons. First, an interviewee in the former has an opportunity to provide detailed and unstructured responses to each question, which enables a clinician to make a more accurate assessment. Second, a clinician is able to personalize questions based on the interviewee's earlier responses and his/her own professional experience. Interviews can be either structured or semi-structured. Structured interviews involve the exactly same predetermined questions in a fixed order. Semi-structured interviews contain both general questions that apply to all individuals and specific questions personalized for individual interviewees based on their responses to earlier questions.

2.3. Automatic depression detection

Automatic depression detection requires training and testing machine learning models. The model training requires data collected from both individuals with depression and those without it. One of the key tasks involved in training models is to select predictive features. Based on the number of communication modalities (i.e., verbal (T), visual (V), and vocal (A)) from which predictive features are selected, we classify depression detection methods into three broad categories—unimodal feature–, bimodal feature–, and multimodal feature–based methods.

- Unimodal feature–based methods: Studies in this category [42–46] used features extracted from a single communication modality. For instance, Lopez-Otero et al. (2014) used vocal features only, such as Mel-frequency Cepstral Coefficients (MFCC) and short-time energy (Energy), in depression detection models built by the Gaussian Mixture Model (GMM) and Support Vector Regression (SVR).
- Bimodal feature–based methods. Studies in this category combined features of two modalities when building depression detection models (e.g., [47–50]). For example, Syed et al. [50] utilized vocal (e.g., spectral low-level features) and visual (e.g., head movement) features as the input of an SVR model for depression detection.
- Multimodal feature–based methods. This group of studies utilized features extracted from three or more communication modalities (e.g., [20,51,52]). For example, Yang et al. [52] considered vocal features (e.g., energy features), visual features (e.g., landmark features), and verbal features (e.g., word vector features) when training and testing depression detection models. However, they focused on individual-level features, not features at an interview question level.

Table 2 provides a summary and categorization of representative studies on automatic depression detection in multiple dimensions, including feature category, sample features, selected machine learning techniques, datasets, and reported performance. We make several observations from the table: 1) none of the previous studies has considered input features at a level of interview questions; 2) despite that some depression assessment questionnaires (e.g., PHQ-9) are designed to measure depression severity [5], the majority of those studies treated depression detection as a binary classification problem (i.e., depressed vs. non-depressed). It is widely recognized that people experience depression at different levels of severity [64,65], and thus the ability to identify depression severity can have significant implications for depression intervention; 3) despite a handful of studies using multimodal features for depression detection (e.g. [18,19,49,50,61],), none of them has provided theoretical explanations for the importance of different types of input features to depression detection; and 4) the previous studies have overlooked the semantic relatedness of different interview questions. As a result, there is a lack of understanding of the impact of question categorization on model performance.

3. The proposed method

This research aims to address the above-mentioned limitations by proposing a novel two-layered model for the detection of depression severity level that leverages multimodal features extracted from question-level responses to the questions in depression assessment interviews. The proposed model combines the strengths of semi-structured interviews and automated depression detection. In addition, it classifies interview questions into meaningful categories and further investigates the importance of different question categories for depression detection.

3.1. Dataset and interview questions

Our dataset was drawn from the DAIC-WOZ database [66], which consisted of audio and video data collected from semi-structured interviews for depression screening. Each interview consisted of three parts, which proceeded in the following order: interpersonal interviews, Wizard-of-Oz interviews, and automated interviews [66]. The interpersonal part aimed to build rapport with an interviewee and make him/her feel comfortable; the Wizard-of-Oz part focused on questions about symptoms and events related to depression and Post-Traumatic Stress Disorder (PTSD), if any; and the automated part, as a "cool-down" phase, intended to ensure that respondents would not leave the interview in a distressed state of mind [66].

There were a total of 97 interview questions in the database. Some questions such as those about contact information and profession were administered to all interviewees, whereas other questions, such as military services attended, depression experience, and intervention received previously, were personalized based on an interviewee's responses to earlier questions. As a result, each interviewee responded to about 55 questions on average. In addition, each interviewee was also asked to respond to the PHQ-8 questionnaire [67], and his/her PHQ-8 score served as the ground truth of his/her depression level. Our dataset included the responses to both interview questions and the PHQ-8 questionnaire from 142 interviewees. There were a total of 7866 responses to the interview questions. Nearly one-third of the selected interviewees had the PHQ-8 scores greater than 10, indicating a moderate or severe level of depression.

3.2. Multimodal features

The proposed method extracted multimodal features from the response to each interview question and used them as the input of depression detection models. We categorized the input features into vocal, visual, and verbal features, and further divided each category into several subcategories, as shown in Table 3. We selected the features by mainly drawing on related theories and previous empirical findings. Additionally, we incorporated several new features and introduce them later in this section.

We organized the vocal features used in previous depression detection literature (e.g., [35,36]) into three sub-categories: prosodic, voice quality, and spectral features. The vocal features were extracted by Collaborative Voice Analysis Repository (COVAREP) [68] with a 10 *ms* interval.

Visual features are important for physicians to gauge the depression level of an individual [69]. Based on the findings of previous studies [31,32,70,71], we selected facial action units (AUs), eye gaze, and head pose as three sub-categories of visual features, which were extracted by

Table 2
A Summary of Automatic Depression Detection Studies.

Categories			Features	Machine Learning Techniques	Dataset (size)	Performance[a]	Studies
A	V	T					
√			TEO-based features; cepstral, prosodic, spectral, and glottal features	GMM and SVM	Oregon Research Institute-collected (139)	Accuracy: 0.88 for males and 0.81 for females	[53]
√			MFCC; RASTA-PLP; SDC; energy and spectral features; prosodic features	GMM and SVR	AVEC 2013 (292)	RMSE: 10.06; MAE: 7.70	[44]
√			Mel-scale filter bank features	Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN)	AVEC 2016 (107)	F1 score: 0.52; precision: 0.35; recall: 1.00	[46]
√			VL-Formants; eGeMAPS; COVAREP feature set	SVM	DAIC-WOZ (142)	F1 score: 0.49 (male); 0.55 (female); 0.63 (overall)	[42]
	√		Body parts movement features	SVM	Clinic Interview (CI) (48)	Accuracy:0.97; F1-score: 0.89	[54]
	√		Eye movement features	SVM	Black Dog Institute (BDI)-collected (60)	Recall: 0.77; Accuracy: 0.75	[55]
	√		Head pose and movement features	SVM	BDI -collected (60)	Average recall: 0.73	[56]
	√		Fisher vector features	SVR	AVEC 2014 (83)	RMSE:8.91; MAE:7.08	[43]
		√	TF-IDF features	SVM	Bulletin Board System Collection (2071)	Accuracy: 0.8451; AUC: 0.9176	[57]
		√	GloVe representation features	SVM	DAIC-WOZ (142)	Accuracy: 0.857; F-score: 0.73	[45]
√	√		Facial actions and vocal prosody	SVM and Logistic regression	CI (57)	Accuracy: 0.88	[58]
√	√		MHH; EOH; LBP; LLDS; MFCCs	Partial Least Square regression (PLSR)	AVEC 2013 (292)	RMSE: 10.96; MAE: 8.72	[59]
√	√		F0-raw; intensity; loudness; MFCC; LBP-TOP features;	SVM	BDI -collected (60)	Accuracy : 0.917	[60]
√	√		Eye gaze; head pose; speaking behavior patterns	SVM	BDI -collected (60)	Accuracy: 0.88	[61]
√	√		TECC; MFCC; facial landmark features	SVM, G-PLDA, LR, and Random Forest (RF)	AVEC 2016 (107)	F1-score: 0.63; MAE: 5.3566; RMAE: 6.7334	[49]
√	√		MFCCs; LLDs; LDP; EOH; LPQ	PLSR and LR	AVEC 2014 (83)	RMSE: 7.43; MAE: 6.14	[47]
√	√		Prosodic features; COVAREP voice quality features; Spectral features; AUs; Eye gaze; Head pose	RF	AVEC 2017 (189)	RMSE: 6.97; MAE: 5.66	[62]
√	√		Turbulence features, LLDs; head movement, mouth openings, eyelid movements, fisher vector features	PLSR and SVR	AVEC 2017 (189)	RMSE: 6.34; MAE: 5.30	[50]
√		√	Prosody, speech-rate; syntactic, semantic features	SVR	AVEC 2014 (83)	RMSE: 9.21; MAE: 7.56	[48]
√	√	√	LDDs; facial expression; linguistic features, speech auxiliary behavior, and word affect features	GMM, Relevant Vector Machines, Gaussian Process Regression	AVEC 2017 (189)	RMSE: 6.02; MAE: 4.98	[63]
√	√	√	LLDs, spectral and energy features, voicing related features; 2D landmark features; word vector features	Deep Neural Network and Deep CNN	AVEC 2017 (189)	RMSE: 5.97; MAE: 5.16	[52]
√	√	√	COVAREP extracted and formant features; AUs; semantic features	RF, SGD and SVR	AVEC 2017 (189)	RMSE: 4.99; MAE: 3.96;	[51]
√	√	√	IS10 features, soundnet, interlocutor influence; facial expression features; word vector features	SVR and LSTM-RNN	AVEC 2017 (64)	Accuracy: 0.675 (arousal); 0.756 (valence); 0.509 (likability)	[20]
√	√	√	Prosodic, voice quality, and spectral features; syntax-POS tag; AUs	SVM	DAIC-WOZ (136)	Precision: 0.40; Recall: 0.96; F1-score: 0.49	[21]

Note: A, V, and T represent vocal, visual, and verbal features, respectively.
[a] This research presents the best reported performance of every previous study.

Table 3
The List of Selected Features.

Category	Sub-category	Features	Description
Vocal Features	Prosodic features	F0 (a1)	Fundamental frequency of pitch
		VUV (a2)	Binary voicing decision
	COVAREP voice	NAQ (a3)	Normalized Amplitude Quotient
	quality features	QOQ (a4)	Quasi-Open Quotient
		H1H2 (a5)	Difference in amplitude of first two glottal harmonics
		PSP (a6)	Parabolic spectral parameter
		MDQ (a7)	Maxima dispersion quotient
		peakSlope (a8)	Maximum peaks at each scale for the middle part of the utterance
		Rd (a9)	Wavelet based features
		Rd_conf (a10)	The confidence of Rd
		Creak (a11)	Vocal fry or pulse phonation
	Spectral features	MCEP_0 – 24 (a12-a36)	Mel cepstral coefficient
		HMPDM_0 – 24 (a37-a61)	Harmonic model and phase distortion mean
		HMPDD_0 – 12 (a62-a74)	Harmonic model and phase distortion deviations
Visual Features	AUs	AU_confidence/AU_success (v1-v2)	Extraction confidence/success of AU features
		AU01 – 02/04 – 06/09 – 10/12/14 – 15/17/20/25 – 26_r (v3-v16)	Regression output of Inner brow raiser/Outer brow raiser/ Brow lower/Upper lid raiser/ Cheek raiser/Nose wrinkle/Upper lip raiser/Lip corner puller/ Dimpler/Lip corner depressor/Chin raiser/ Lip stretched/Lip part/Jaw drop
		AU04/12/15/23/28/45_c (v17-v22)	Binary outputs of AU4/AU12/AU15/Lip tightener/Lip suck/Blink
	Eye gaze	gaze_confidence/success (v23-v24)	Extraction confidence/success of gaze features
		x_0/ y_0/ z_0/ x_1/ y_1/ z_1 (v25-v30)	World coordinate space of both eyes
		x_h0 / y_h0/ z_h0/ x_h1/ y_h1 / z_h1 (v31-v36)	The gaze in head coordinate space
	Head pose	pose_confidence/success (v37-v38)	Extraction confidence/success of head pose features
		Tx/Ty/Tz (v39-v41)	The position coordinates
		Rx/Ry/Rz (v42-v44)	The head rotation coordinates
Verbal Features	Text	sen-num	The number of sentences
		avg-words	Average word count
		avg-adj	Average number of adjectives
		avg-adv	Average number of adverbs
		sentiment	Sentiment of responses
	Sentence	p-s-ratio	Positive sentence ratio
		n-s-ratio	Negative sentence ratio
	Word	prp-ratio	Reflexive pronoun ratio
		um-ratio	'um' word ratio
		sni-ratio	'sniffle' word ratio
		laugh-ratio	'laugh' word ratio
		sigh-ratio	'sigh' word ratio
		avg-ᴅ-ratio	Depression word ratio
		avg-p-ratio	Positive word ratio
		avg-n-ratio	Negative word ratio

OpenFace at a 66.67*ms* interval.

Previous depression detection studies (e.g., [13,72,73]) have suggested a number of verbal features, such as the number of sentences, average number of words per sentence, average word count, and sentiment. We categorized verbal features at word, sentence, and text levels. Importantly, based on recent depression studies and depression-related theories, we identified and incorporated several new verbal features, including depression word ratio, negative and positive sentence ratios, and reflective pronoun ratio, in this study. We extracted these features from the text transcripts of interview responses. The word level features were measured as a ratio of the number of corresponding type of words (e.g., depressive words) to the total word count in an interview response, and the sentence level features were normalized by the total number of sentences in an interview response. We briefly describe the rationales for including each of the new features as follows:

- There are cognitive symptoms of depression. For instance, absolutist thinking is regarded as cognitive distortion for depression by most cognitive therapies, and an elevated use of absolutist related words has been empirically found to be associated with depression [74]. There has been significant progress in developing depression lexicons using either the top-down or the bottom-up approach to support depression screening. One representative example of the top-down approach to depression lexicon development drew on "Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) and in the ICD-10 Classification of Mental and Behavioral Disorders: Clinical Descriptions and Diagnostic Guidelines classification systems" [76]. A typical example of the bottom-up approach to the lexicon development extracted depression phrases from the dependency relationships of metaphorical expressions (e.g., "depression is like…") in web search results, and combined those phrases with their synonyms, and phrases appearing in the definitions of the metaphors to create a depression lexicon [75]. The evaluation results with website and blog posts provided preliminary evidence for the positive effect of the lexicon on depression screening [75]. Nevertheless, depression lexicon is yet to be used to analyze interview responses for depression assessment. In this study, we introduced the depression word ratio as a verbal feature, defined as the number of depression words divided by the total word count in an interview response. Additionally, in view of the complementariness of different methods for lexicon development, we combined multiple depression lexicons in measuring this feature.
- Depression has a connotation of negative emotions. According to Cognitive Behavioral Theory of depression [77], the more negative that one feels, the more depressed he/she may become. Therefore, we introduced negative sentence ratio and positive sentence ratio, which were defined as the number of sentences expressing negative and positive sentiments divided by the total number of sentences in an interview response, respectively.
- According to self-awareness theory of reactive depression, self-focusing may be associated with depression [78]. Loss of self-worth

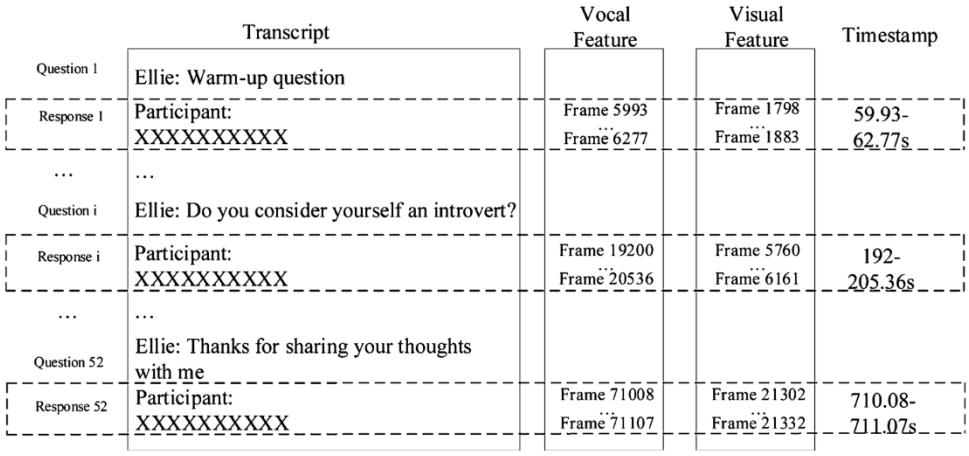


Fig. 1. Sample Interview Questions and Multimodal Data Alignment.

and self-esteem and getting stuck in self-derogation are attributable to depression. Thus, we introduce reflective pronoun ratio as a verbal feature, which is defined as the number of reflective pronouns normalized by the total word count in an interview response.

3.3. Question-level feature extraction

Previous studies on automatic depression detection employed input features at an individual level [21], which can be too coarse to reveal a person's true state of depression. In a clinical interview, an individual's responses to some questions can be more revealing of his/her depression state than other questions. Simply aggregating the input features extracted from all interview responses may fail to differentiate important indicators of depression. To address this critical limitation, we extracted input features from question-level responses in this study.

To support question-level analysis, we segmented the interview data by questions and aligned the multimodal data accordingly. Fig. 1 shows sample interview questions and alignment of multimodal data. Specifically, we segmented audio, video, and text data (transcript) by unit of responses to individual questions based on the start/end time of each question.

We extracted vocal, visual, and verbal features from the responses to individual questions separately. Because the original audio and video data were sampled by frame (i.e., 10 *ms* per frame for the vocal data and 66.67 *ms* for the visual data), the audio or video feature f_q of a response to question q was measured as the mean of all the frames falling between the start and end time of that response, as shown in Eq. (1):

$$f_q = \frac{\sum_{i=s_q}^{e_q} f_q^i}{e_q - s_q + 1} \quad (1)$$

where $i \in [s_q, e_q]$ denotes the index of a frame in the response to question q; and s_q and e_q denote the indexes of the starting and ending frames of the corresponding response, respectively.

The extraction of verbal features was relatively straightforward, which involved analyzing the text transcript of responses to individual interview questions.

3.4. Question categorization

One of the objectives of this research is to gain insights into the importance of individual questions to depression detection. Given the large number of interview questions, it makes both theoretical and practical senses to categorize questions.

We conducted thematic analysis of interview questions. Three researchers independently analyzed the themes of all interview questions first. Among them, one had background in psychology; another had been actively engaged in depression detection research for five years; and the third researcher was familiar with the DAIC-WOZ dataset. The researchers then discussed and resolved non-overlapping themes among the three sets of results via a face-to-face meeting. We finalized the themes of the questions based on group consensus and generated nine question categories, including depression, therapy, problem, social relationship, military, personality, sleep, emotion, and conversation. Table 4 lists the question categories, their descriptions, sample items, and the number of question items in each category.

Table 4
Categorization and Description of Interview Questions.

Category	Description	Sample Items	Question Count
Depression	Previous depression experience	Depression diagnosis, symptoms, time of diagnosis	5
Therapy	Previous experience with depression therapy	Effects, motivation, usefulness	6
Problem	Problems encountered in one's life that cannot be solved effectively	Regrets, disturbing thoughts, argument with someone, difficulty in finding a job	31
Social relationship	A strong, deep, or close association or acquaintance with other individuals.	Family, kids, hometown, roommates, parent	14
Military	Military and post-military experience	Motivation, time of serving, transition to civilian life	6
Personality	Individual differences in characteristic patterns of thinking, feeling, and behaving.	best qualities, temper control, introvert	10
Sleep	Sleep quality	Time to fall asleep, poor sleep effect	3
Emotion	An affective state of consciousness in which joy, sorrow, fear, hate, or the like, is experienced.	Feeling, happy, mad, sad, fun, experience	20
Conversation	The opening and closing routines in a conversation	Greeting, wrap-up remark	2

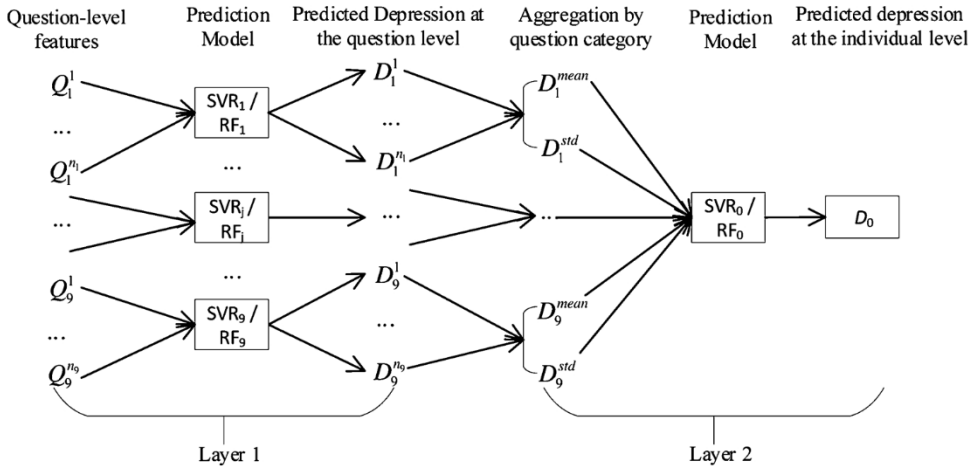


Fig. 2. A Two-Layered Depression Detection Model.

3.5. Building depression detection models

We have designed a novel two-layered model for depression severity detection (see Fig. 2). The first layer consists of nine sub-models, with each model corresponding to one category of interview questions (see Table 4). Each sub-model takes question-level features Q_j^i (i.e., 133 vocal, visual, and verbal features) as the input and generates the predicted depression level D_j^i as the output, where Q_j^idenotes vocal, visual, and verbal features extracted from the response to the question i in the question category j ($j=1\ldots9$); n_j represents the total number of questions in category j; and SVRj and RFj ($j=1\ldots9$) represent Support Vector Regressor and Random Forest models for the prediction of depression, respectively.

We chose SVR and RF by following the previous studies that shared a similar research goal [20,43,49]. The total sample size for training and testing the nine sub-models was 7,866. The sub-models were built for individual categories of interview questions. The sample size for each sub-model varied for two main reasons. First, the number of questions differed among different question categories. Second, the numbers of interviewees who answered questions in different categories also varied, depending on whether the questions were applicable to individual interviewees during an interview. We trained and tested each sub-model using five-fold cross-validation by randomly dividing the data into five sections with equal sample size, with one section retained as the validation data for testing a sub-model, and the remaining four sections used as the training data. This cross-validation process was repeated five times (i.e., folds), with each of the five data sections used exactly once as the validation data. The results of the five folds were then averaged to produce a single model prediction.

In the second layer, the input features were the aggregated outputs of the first layer. Specifically, the input features consisted of nine pairs of means (D_j^{mean}) and standard deviations (D_j^{std}) of the predicted depression levels at the first layer, one pair from each sub-model (i.e., question category). We chose mean and standard deviation instead of other values (e.g., median, mode, range) for two reasons. First, the former have been more frequently used in data analytics and modeling than the latter [79–81]. Second, we empirically experimented with mode, median, and range in the second-layer models, which resulted in inferior model performance. SVR_0 and RF_0 were Support Vector Regressor and Random Forest models for detecting depression at an individual level in the second layer. The data size for this layer was 142. Similarly, we applied five-fold cross-validation to train and test the models in this layer.

4. Evaluation and results

4.1. Experimental procedure

We extracted question-level multimodal features of every interviewee using several python packages such as NLTK and pandas based on the question and response alignment method, as illustrated in Fig. 1. Then, we trained and tested the proposed two-layered model in two steps. In the first step, we created nine sub-models, one for each of the nine pre-defined question categories. In every sub-model, we trained SVR and RF models for the prediction of the question-level depression scores. In the second step, we trained and tested both SVR and RF models at an individual level, with the inputs being the means and standard deviations of the predicted depression scores in step 1 for all the individual's responses to questions in each category. In addition, we also trained and tested baseline models for depression detection. Finally, we performed sensitivity analyses on the best-performed models to examine the importance of input features and question categories to model performance. The detailed experimental procedure is shown in Fig. 3.

4.2. Evaluation setting and metrics

We chose RF and SVR as the machine learning techniques for building depression detection models because they had been commonly

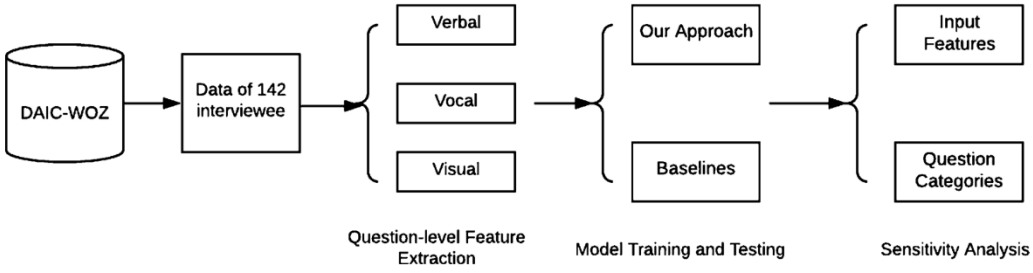


Fig. 3. Experimental Procedure.

used in depression detection research [20]. To assess the impacts of multimodal features on depression detection, we compared the proposed depression detection models (O_RF and O_SVR) against models using unimodal or bimodal features. In addition, we chose two types of baseline methods. The first type of baseline (B1) models was Random Forest models that employed vocal features only, or visual features only, or both vocal and visual (i.e., bimodal) features extracted from the same data source. The second type of baseline (B2_SVR and B2_RF) models was developed with individual-level data, which aims to evaluate the effect of the proposed two-layered, question-level model. Moreover, B2 incorporated the additional verbal features (see Section 3.2).

By following previous studies [82], we used Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) as the metrics for assessing model performance in detecting depression levels, which are defined in Eqs. (2) and (3):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{3}$$

where P_i and O_i denote the depression level predicted by a model and the actual depression level of the i_{th} interviewee, respectively, and n denotes the total number of interviewees. MAE measures the average difference between the predicted and actual depression levels, and RMSE is the standard deviation of residuals (i.e., prediction errors) that indicates how spread out those residuals are. The reported performance of each model was the averaged results of 5-fold cross-validations.

To gain insights into their importance to depression detection, we performed sensitivity analyses on both input features and individual question categories. To determine the importance of each question category to depression detection performance, we focused on the sensitivity of RMSE, which has shown advantages over MAE in revealing model performance differences [82].

4.3. Depression detection results

Table 5 reports MAE and RMSE of the constructed depression detection models. Because B1 did not consider verbal features, there was no performance reported for the corresponding settings, as denoted by '-'in the table. The best performances across the different models are highlighted for each feature setting in Table 5. Table 6 presents the results of a paired-sample *t*-test between the performances of our proposed model and the baseline models. Fig. 4 plots the best results of B1, B2, and our proposed model under different feature settings.

Table 5
The Performance of Depression Detection Models.

(a) MAE

Input Features	B1	B2_SVR	B2_RF	O_SVR	O_RF
A	5.72	4.94	4.81	4.1	4.23
V	6.12	5.4	5.32	4.43	4.09
T	–	4.72	4.75	4.41	4.35
A + V	5.66	5.12	4.79	3.95	4.12
A + T	–	4.71	4.72	3.94	4.2
V + T	–	5.11	5.05	4.24	4.2
A + V + T	–	4.65	4.71	3.84	4.1

(b) RMSE

Input Features	B1	B2_SVR	B2_RF	O_SVR	O_RF
A	7.78	6.38	5.86	5.49	5.39
V	6.97	6.8	6.35	5.62	5.11
T	–	5.88	5.84	5.58	5.42
A + V	7.05	6.76	5.82	5.32	5.28
A + T	–	5.80	5.58	5.3	5.37
V + T	–	6.5	6.05	5.44	5.24
A + V + T	–	5.79	5.75	4.96	5.22

Table 6
T-test Results of the Performances of the Proposed Model and Baseline Models.

Metrics	Model 1	Model 2	t-statistics (Model 1 – Model 2)
MAE	O_SVR	B1	−10.038***
	O_SVR	B2_SVR	−5.973***
	O_RF	B1	−11.108**
	O_RF	B2_RF	−7.530***
RMSE	O_SVR	B1	−9.222***
	O_SVR	B2_SVR	−4.705**
	O_RF	B1	−7.572**
	O_RF	B2_RF	−5.928***

Notes: ***: p < .001, **: p < .01; *: p < .05.

Tables 5 and 6 show that both O_SVR and O_RF outperform the baseline models B1 and B2 across all the combinations of input features. Therefore, the results provide strong evidence for the positive effects of question-level features on depression level detection. When combining all the three types of input features, O_SVR outperformed O_RF in terms of both MAE and RMSE. Fig. 4 also shows the best performance of the two baseline and the proposed models under different feature settings. A comparison of these models shows that the O_SVR model with multimodal features outperformed the baseline model B1 in terms of MAE by 32.2 % and in terms of RSME by 28.8 %; and outperformed baseline B2 in terms of MAE by 17.4 % and in terms of RMSE by 13.7 %.

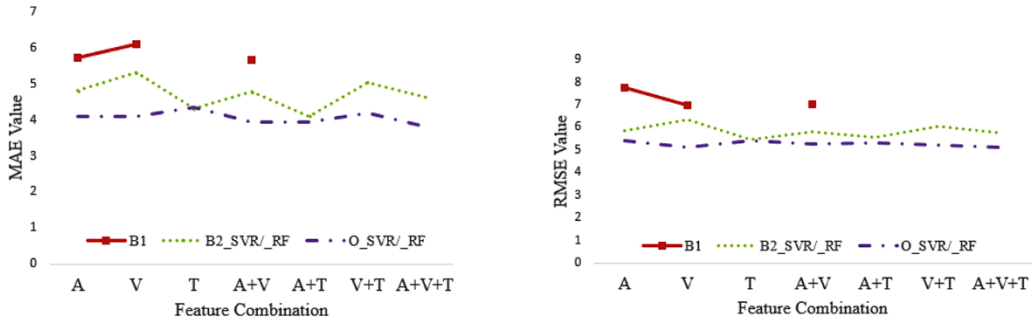
To gain insights into the effects of multimodal features, we compared the performance of our proposed model across different feature settings based on 30 runs by varying the model parameters. The *t*-test results, as reported in Table 7, show that the models built with multimodal features outperform with bimodal ($p < .001$) and unimodal counterparts ($p < .01$), demonstrating the benefits of incorporating multimodal features in depression detection.

To test the effects of new verbal features (depression word ratio, negative and positive sentence ratios, and reflective pronoun ratio), as introduced in this study, we compared the performances of the models with and without those new verbal features (Table 8). The results of paired-sample *t*-test reveal that incorporating the new verbal features consistently improved depression detection performance ($p < .05$) when using verbal features only or combination of verbal and visual features.

4.4. Feature importance and ranking

To examine the relative importance of individual features to depression detection, we performed a post-modeling sensitivity analysis of the input features. Based on the results of the RF model (O_RF), we ranked the features in a decreasing order of their importance to depression detection. Table 9 lists the top-30 most important features along with their importance scores, with the top 5 being depression word ratio, Mel cepstral coefficients 11, 6, and 8, and Action Unit 20. It is worth noting that the top-5, top-10, and top-30 features all include the features of three modalities.

To validate the effects of the top features on depression detection, we re-trained O_RF with the top n features only, with n ranging from 10 to 60 with an interval of 10. The results, as shown in Fig. 5, revealed that using the top n features only (_Selected) led to improved depression detection performance in terms of MAE and RMSE compared with using the entire feature set (_Whole). In addition, the MAE initially declined as the number of input features increased, and then leveled off after the number of features reached 40. Specifically, the best model with the top 40 features outperformed the best model with the entire set of features by 42.4 % in terms of MAE and by 37.5 % in terms of RSME,



(a) MAE **(b) RMSE**

Fig. 4. Performance Comparison of the Best-performed Models.

Table 7
T-test Results of Performance Comparison between Different Modalities.

Models	Performance	Feature setting 1	Feature setting 2	t-statistic (Feature settings 1–2)
O_SVR	MAE	Multimodal	Unimodal	−29.559***
		Multimodal	Bimodal	−13.788***
		Bimodal	Unimodal	−12.403***
	RMSE	Multimodal	Unimodal	−104.697***
		Multimodal	Bimodal	−60.020***
		Bimodal	Unimodal	−24.064***
O_RF	MAE	Multimodal	Unimodal	−10.951***
		Multimodal	Bimodal	−18.345***
		Bimodal	Unimodal	−4.184***
	RMSE	Multimodal	Unimodal	−5.857**
		Multimodal	Bimodal	−13.304***
		Bimodal	Unimodal	−1.016

Notes: ***: $p < .001$, **: $p < .01$.

respectively. These results provide strong evidence for the importance of feature selection to the performance of automated depression detection.

4.5. Importance of interview question categories

To examine the importance of individual question categories to depression detection, we conducted a sensitivity analysis by building multimodal depression detection models that excluded the data of each question category alternately. Fig. 6 presents the results in a bar chart. The dotted line in the chart denotes the RMSE of the model when the interview questions of all categories were included, which serves as a reference point. The height of a bar above the dotted line is proportional to the degree of performance reduction resulting from removing data in the corresponding question category. Based on the results of side-by-side comparisons, the problem-related questions led to the largest performance reduction, implying that the problem category was most influential on model performance. The next four most important question categories included depression, military, personality, and emotion. In contrast, interview questions of the therapy and social relationship categories had the least impact on the depression detection performance.

Table 9
Top-30 Features and their Importance Scores.

Feature	Score	Feature	Score	Feature	Score	Feature	Score
avg-ᴅ-ratio	43.42	AU05	24.34	AU15_r	22.50	AU45_c	21.06
MCEP_11	26.88	laugh-ratio	24.24	AU25_r	22.46	AU04_c	20.71
MCEP_6	25.89	Gaze confidence	24.23	HMPDM_9	22.41	AU06_r	19.61
MCEP_8	25.88	MCEP_10	24.19	MCEP_13	22.33	AU09_r	19.60
AU20_r	25.49	AU confidence	23.65	AU23_c	22.13	MCEP_3	19.54
MCEP_5	24.93	AU01_r	23.34	MCEP_22	21.56	AU28_c	19.37
MCEP_9	24.58	AU10_r	23.26	MCEP_14	21.54		
avg-p-ratio	24.43	AU12_r	23.12	AU14_r	21.45		

5. Discussion

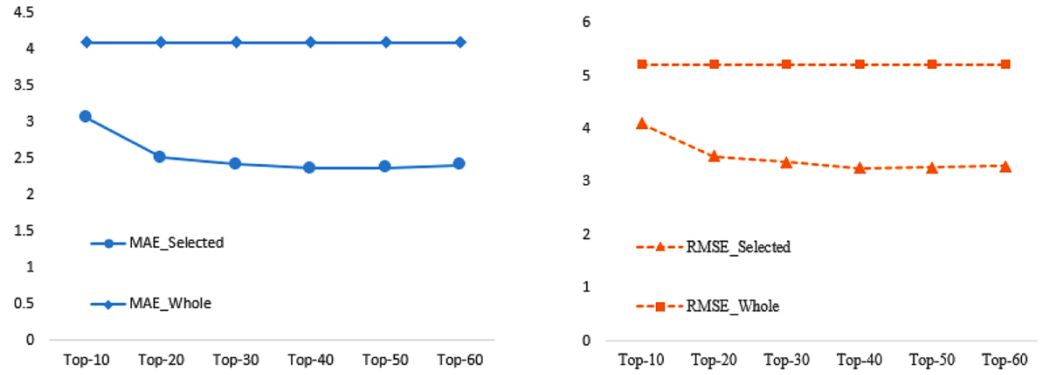
Depression is a major mental disorder that can cause serious harms to individuals and huge economic loss to themselves, their families, and the society at large [83]. Automatic and effective depression detection can enable early intervention and reduce negative consequences. In this study, we developed depression detection models with multimodal features by exploiting rich information embedded in interview responses. More importantly, we introduced a two-layered model for automatic depression detection by identifying categories of interview questions and detecting depression at a question level first prior to making a final detection decision at an individual level.

5.1. Major findings

This study produces several major findings. First, incorporating more multimodal features into depression detection models as predictive variables led to significant performance improvement. Second, input features extracted from interview responses at a question level were more effective for depression detection than their individual-level

Table 8
Comparisons of Model Performances with vs. without New Verbal Features.

Input Features	Without new verbal features				With new verbal features			
	MAE		RMSE		MAE		RMSE	
	O_SVR	O_RF	O_SVR	O_RF	O_SVR	O_RF	O_SVR	O_RF
T	4.5	4.39	5.6	5.48	4.41	4.35	5.58	5.42
A + T	4.26	4.29	5.41	5.44	3.94	4.2	5.3	5.37
V + T	4.27	4.28	5.49	5.32	4.24	4.2	5.44	5.24
A + V + T	4.02	4.19	5.09	5.23	3.84	4.1	4.96	5.22



(a) MAE **(b) RMSE**

Fig. 5. Performance Comparison between the Selected Features and the Whole Feature Set.

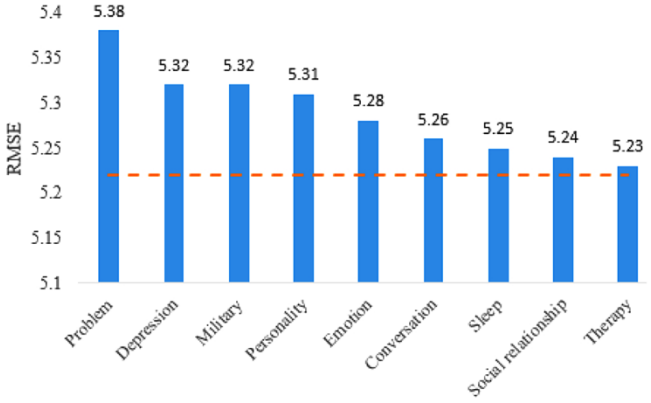


Fig. 6. Performance of Individual Question Categories.

counterparts. Third, depression word ratio, Mel cepstral coefficients, and Action Unit 20 were among the most important features for depression detection, spanning across all three types of modalities. Fourth, among different question categories, the *problem* category was the most effective in facilitating depression detection, followed by the categories of depression, military, and personality. Research suggests that problems encountered in one's life that cannot be solved effectively, such as regrets, disturbing thoughts, argument with someone, and difficulty in finding a job, may lead to depression [84–86]. Therefore, responses to the problem-related questions such as "*is there anything you regret?*" and "*do you have disturbing thoughts?*" become important to detection of depression levels.

5.2. Additional tests

Given the variation in the number of interview questions administered to individual interviewees (mean = 55, std. = 9.02), we further examined how sensitive depression detection performance was to the number of interview questions. To this end, we identified two subgroups of interviewees: one consisting of the top 30 interviewees who were asked to respond to the most questions (an average of 67 questions), while the other consisting of the bottom 30 interviewees who were asked the least number of questions (an average of 43 questions) during an interview. Then, we compared the absolute error of the prediction results of the best-performed model (i.e., SVR in the A + V + T setting) between the two subgroups. The results of a paired-sample *t*-test show that the mean difference between the two subgroups is not statistically significant ($p > .05$). Therefore, the number of interview questions did not appear to affect the performance of depression level prediction.

We conducted an additional test to examine whether the importance of question categories is contingent upon their question counts. We produced two rankings of the question categories: one was based on question counts and the other was based on question importance scores, both in a descending order. A comparison between the two rankings shows that the rank positions of some questions but not others were consistent between the two rankings. For instance, the 'problem' category contained the most questions and was the most important question category. However, 'depression' and 'military' were ranked as the second and third categories in terms of importance, yet were only the fifth and seventh in terms of question count, respectively. We further performed Spearman's correlation analysis between the two rankings, which did not yield any significant correlation ($\rho = 0.2678, p > .05$). Therefore, the importance of question categories did not appear to correlate with their question counts.

5.3. Research contributions

This study makes multi-fold research contributions. First, it combines the strengths of data analytics and interview methods in detection of depression levels. Second, it integrates a rich set of multimodal features, including verbal, vocal, and visual features, in developing depression detection models, which improves detection performance. The sensitivity analysis not only reveals the importance of individual input features to the detection of depression level for the first time, but also highlights the significance of considering multimodal features in depression detection. Further, this study introduces several novel and effective verbal features such as depression word ratio, negative and positive sentence ratios, and reflective pronoun ratio. Third, the proposed method extracts input features from interview responses at a question level. Those fine-grained features have shown to contribute more to performance improvement compared with individual-level features. Last but not least, the categorization of interview questions based on thematic analysis and the related sensitivity analysis shed light on the importance of different question categories to automatic depression detection.

5.4. Practical implications

The findings of this study have important practical implications. The improved performance of depression detection can enable timely depression intervention and increase user acceptance of depression detection models. The development of different categories of interview questions and the analyses of their effects on depression detection provide valuable suggestions on how to improve the design of interview questions in support of automatic depression detection. Additionally, the examination of a rich set of multimodal features and their importance to depression detection provides practical guidance on feature selection for building automatic depression detection models, which

can both empower novice users and assist mental health professionals in depression assessment.

This study has a few limitations that offer opportunities for future research. First, the research findings are generated based on the analysis of a single dataset. It would be helpful to validate the generality of the research findings with a different dataset. Second, limited by the data accessibility, the size of the dataset is relatively small, which does not warrant employing advanced deep learning techniques. The model performance is expected to benefit from a larger dataset. Third, in addition to SVR and RF, other machine learning techniques may be explored to improve the performance of depression detection models. Furthermore, although the present study utilized linguistic features extracted from interview responses, similar to existing studies [87,88], it did not analyze and incorporate deep-level semantics of interview responses, which is potentially helpful to depression detection and is worth exploring in future.

CRediT author statement

All the authors have contributed to the conceptualization, methodology, validation, investigation, writing – original draft, and writing – review & editing of the reported study. In addition, **Guohou Shan** has also contributed to software and formal analysis, and **Lina Zhou** and **Dongsong Zhang** have contributed to funding acquisition.

Acknowledgement

This research was partially supported by the National Science Foundation under Grants [grant numbers SES-1527684, CNS-1704800]. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the above funding agency.

Appendix A

Vocal features

The vocal features listed (a1-a74) were extracted using COVERAP.[2] Their detailed descriptions and range of values are provided below.

- **F0 (a1)**: Fundamental frequency of pitch. F0 represents the periodicity of the speech signal, which is primary acoustic correlated of pitch, affected by the frequency of vocal fold vibration at the glottis. For a typical adult woman, the range of F0 is between 165HZ and 255HZ, while the range is from 85HZ to 180HZ for a typical adult man.
- **VUV (a2)**: Binary voicing decision (0 or 1). VUV indicates whether the current segment is voiced or unvoiced.
- **NAQ (a3)**: Normalized Amplitude Quotient. NAQ parameterizes the glottal closing phase, which is defined as the ratio between the maximum of the glottal flow and the minimum of its derivative, then normalized with respect to the fundamental frequency.[3] The values of NAQ fell into the range between 0 and 0.35 in this study.
- **QOQ (a4)**: Quasi-Open Quotient. QOQ describes the relative open time of the glottis. It is defined as the duration during which the glottal flow is 50 % above the minimum flow, normalized to the pitch period. [3] For a normal speaker, the values of QOQ varied from 0.75 to 0.40.
- **H1H2 (a5)**: Difference in amplitude of first two glottal harmonics. H1H2 is computed as the difference between the log-magnitude of the first and second harmonics. [3] Depending the pitch of speakers, the values of H1H2 varied in the range between 50.08 and 30.01 in this study.
- **PSP (a6)**: Parabolic spectral parameter. PSP is a frequency domain parameter for the quantification of the glottal volume velocity waveform. It is based on fitting a parabolic function to the low-frequency part of a pitch-synchronously computed spectrum of the estimated glottal flow.[4] Based on speakers' fundamental pitch, the PSP value can vary from 0 to 3.
- **MDQ (a7)**: Maxima dispersion quotient. MDQ is used to discriminate breathy to tense voice. It is the ratio of mean distances of maxima locations over the local glottal period duration.[5] Its values fall into the range between 0 and 1.
- **peakSlope (a8)**: Maximum peaks at each scale for the middle part of the utterance. peakSlope is a parameter for identifying breathy to tense voice qualities in a given speech segment using measurements from the wavelet transform. It is calculated using the slope coefficient of the regression line on peak amplitudes.[6] Its value ranges from -1 to 0.
- **Rd (a9)**: Wavelet-based features. Rd is a shape parameter, closely related to the effective pulse declination time of the falling branch. It is calculated by Rd = Td(F0/110), where Td is the pulse decline time and F0 is the frequency pitch.[7] Its values varied in the range of 0–2.50.
- **Rd_conf (a10):** The confidence of Rd, which varies between 0 and 1.
- **Creak (a11):** Vocal fry or pulse phonation. Creak voice is a special kind of phonation. It is calculated based on the process of: (1) extracting the excitation characteristics of creak, highlighting the main acoustic features; (2) filtering the characteristic inputs to a binary classifier to classify whether it is creaky voice or not (0 or 1).[8]
- **MCEP_0 − 24 (a12-a36):** Mel cepstral coefficient. MCEP_0 − 24 are used to measure speech quality. They are calculated based on the Mel frequency scale and critical-band filtering.[9] Their values varied in the range of -10 to 10.
- **HMPDM_0 − 24 (a37-a61):** Harmonic model and phase distortion mean. HMPDM_0 − 24 are a group of features calculated using harmonic model on voices combined with phase distortion mean.[10] The values of HMPDM variables varied in the range of -4 and 4 in this study.
- **HMPDD_0 − 12 (a62-a74):** Harmonic model and phase distortion deviations. HMPDD_0 − 12 are a group of features calculated using harmonic model on voices combined with distortion deviations. [10] The values of the HMPDD variables varied in the range of -3 to 3 in this study.

[2] Degottex, Gilles, John Kane, Thomas Drugman, Tuomo Raitio, and Stefan Scherer. "COVAREP—A collaborative voice analysis repository for speech technologies." In 2014 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp. 960 − 964. IEEE, 2014.

[3] Drugman, T. and Dutoit, T., 2010. Glottal-based analysis of the Lombard effect. In Eleventh Annual Conference of the International Speech Communication Association.

[4] Alku, P., Strik, H. and Vilkman, E., 1997. Parabolic spectral parameter—a new method for quantification of the glottal flow. Speech Communication, 22(1), pp.67 − 79.

[5] Kane, J. and Gobl, C., 2013. Wavelet maxima dispersion for breathy to tense voice discrimination. IEEE Transactions on Audio, Speech, and Language Processing, 21(6), pp.1170 − 1179.

[6] Kane, J. and Gobl, C., 2011. Identifying regions of non-modal phonation using features of the wavelet transform. In Twelfth Annual Conference of the International Speech Communication Association.

[7] Fant, G., 1995. The LF-model revisited. Transformations and frequency domain analysis. Speech Trans. Lab. Q. Rep., Royal Inst. of Tech. Stockholm, 2(3), p.40.

[8] Kane, J., Drugman, T. and Gobl, C., 2013. Improved automatic detection of creak. Computer Speech & Language, 27(4), pp.1028 − 1047.

[9] Kubichek, R., 1993, May. Mel-cepstral distance measure for objective speech quality assessment. In *Proceedings of IEEE Pacific Rim Conference on Communications Computers and Signal Processing* (Vol. 1, pp. 125 − 128). IEEE.

[10] Pantazis, Y., Rosec, O. and Stylianou, Y., 2008. On the properties of a time-varying quasi-harmonic model of speech. In Ninth Annual Conference of the International Speech Communication Association.

Visual features

The visual features listed were extracted using OPENFACE.[11] A total of 22 features were extracted from facial action unit (AU). All the AU features except v1 and v2 were measured as intensity (regression) and/or presence (binary classification). The values of regression output features ranged from -1 to 1 and those of binary classification features ranged from 0 to 1. For eye gaze features (v23-v36), their outputs were represented as four three-dimension vectors. The first two vectors denote the world coordinate space describing the gaze direction of both eyes, and the last two vectors describe the gaze in head coordinate space.[12] For the representation of head pose, there were six features (v37-v44) depicting the head position coordinates and head rotation coordinates. The position was measured in millimeters and rotation in radians and in Euler angle convention. The values of all ratio variables ranged from 0 to 1. All features were represented as continuous variables unless otherwise noted. Their detailed descriptions and range of values are provided below.

- **AU_confidence (v1)**: Extraction confidence of AU features. Its values ranged from 0 to 1.
- **AU_success (v2)**: Extraction success of AU features. Its values are binary (0 or 1).
- **AU01 (v3):** Regression output of inner brow raiser.
- **AU02 (v4):** Regression output of outer brow raiser.
- **AU04 (v5)/(v17):** Regression/binary classification output of brow lower.
- **AU05 (v6):** Regression output of upper lid raiser.
- **AU06 (v7):** Regression output of cheek raiser.
- **AU09 (v8):** Regression output of nose wrinkle.
- **AU10 (v9):** Regression output of upper lip raiser.
- **AU12 (v10)/(v18):** Regression/binary classification output of lip corner puller.
- **AU14 (v11):** Regression output of dimpler.
- **AU15 (v12)/(v19):** Regression/binary classification output of lip corner depressor.
- **AU17 (v13):** Regression output chin raiser.
- **AU20 (v14):** Regression output of lip stretched.
- **AU25 (v15):** Regression output of lip part.
- **AU26 (v16):** Regression output of jaw drop.
- **AU23 (v20):** Binary classification output of lip tightener.
- **AU28 (v21):** Binary classification output of lip suck.
- **AU45 (v22):** Binary classification output of blink.
- **gaze_confidence (v23):** Eye gaze feature extraction confidence. Its values ranged from 0 to 1.
- **gaze_success (v24):** Eye gaze feature extraction success. It took binary values (0 or 1).
- **x_0/ y_0/ z_0/ x_1/ y_1/ z_1 (v25-v30):** World coordinate space of both eyes. Their values range from -1 to 1.
- **x_h0 / y_h0/ z_h0/ x_h1/ y_h1 / z_h1 (v31-v36):**The gaze in head coordinate space. Their values range from -1 to 1.
- **pose_confidence (v37):** The extraction confidence of head pose features. Its values ranged from 0 to 1.
- **pose_success (v38):** The extraction success of head pose features. It took binary values (0 or 1).
- **Tx/Ty/Tz (v39-v41):** Position coordinates. Its values ranged from -1 to 1.
- **Rx/Ry/Rz (v42-v44):** Head rotation coordinates. Its values ranged from -1 to 1.

Verbal features

We extracted 15 verbal features from depression interview transcripts. Their detailed descriptions and range of values are provided below.

- **Sen-num:** The number of sentences. It was computed as the number of sentences contained in a response to an interview question. Its values ranged from 2 to 37 in this study.
- **Avg-words:** Average word count. It was computed as the average word count of the sentences in a response. Its values fell into the range between 0.5 and 49.3 in this study.
- **Avg-adj:** Average number of adjectives. It was computed as the average number of adjectives in sentences of a response. Its values ranged from 0 to 1.5 in this study.
- **Avg-adv:** Average number of adverbs. It was computed as the average number of adverbs in sentences of a response. Its values varied in the range of 0–5.5 in this study.
- **Sentiment:** Expression of positive and negative sentiments. It was measured as the aggregated count of sentiment words. Its values ranged from 0 to 7.5 in this study.
- **P-s/N-s -ratio:** Positive/negative sentence ratio. It is defined as the ratio of sentences expressing positive/negative sentiment.
- **Prp-ratio:** Reflexive pronoun ratio. It is defined as the ratio of reflexive pronoun count to the total number of words in a response.
- **Um-, sni-, laugh-, and sigh- ratio:** The ratio of 'um', 'sniffle', 'laugh', and 'sigh' words, respectively. Each of them was measured as the count of the corresponding type of word divided by the total number of words in a response.
- **Avg-D-, avg-p-, and avg-n- ratio:** The ratio of depression related, positive, and negative words, respectively. Each of them was measured as the count of the corresponding type of word divided by the total number of words in a response.

[11] Baltrušaitis, T., Robinson, P. and Morency, L.P., 2016, March. OPENFACE: an open source facial behavior analysis toolkit. In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 1−10). IEEE.

References

[1] A.T. Beck, B.A. Alford, Depression: Causes and Treatment, University of Pennsylvania Press, 2009.
[2] R.C. O'Connor, M.K. Nock, The psychology of suicidal behaviour, Lancet Psychiatry 1 (1) (2014) 73–85.
[3] P.E. Greenberg, A.A. Fournier, T. Sisitsky, C.T. Pike, R.C. Kessler, The economic burden of adults with major depressive disorder in the United States (2005 and 2010), J. Clin. Psychiatry 76 (2) (2015) 155–162 2015.
[4] S. Khan, J. Peña, Playing to beat the blues: linguistic agency and message causality effects on use of mental health games application, Comput. Human Behav. 71 (2017) 436–443.
[5] K. Kroenke, R.L. Spitzer, J.B. Williams, The PHQ-9: validity of a brief depression severity measure, J. Gen. Intern. Med. 16 (9) (2001) 606–613.
[6] J. Nordgaard, L.A. Sass, J. Parnas, The psychiatric interview: validity, structure, and subjectivity, Eur. Arch. Psychiatry Clin. Neurosci. 263 (4) (2013) 353–364.
[7] A. Bowling, Mode of questionnaire administration can have serious effects on data quality, J. Public Health 27 (3) (2005) 281–291.
[8] R.S. McGinnis, E.W. McGinnis, J. Hruschak, N.L. Lopez-Duran, K. Fitzgerald, K.L. Rosenblum, M. Muzik, Wearable sensors and machine learning diagnose anxiety and depression in young children, IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) (2018) 410–413.
[9] M. De Choudhury, S. Counts, E. Horvitz, Social media as a measurement tool of depression in populations, Proceedings of the 5th Annual ACM Web Science Conference (2013) 47–56.
[10] Rui Wang, Weichen Wang, Alex daSilva, Jeremy F. Huckins, William M. Kelley, Todd F. Heatherton, Andrew T. Campbell, Tracking depression dynamics in college students using mobile phone and wearable sensing Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2 (1) (2018) 43.
[11] Z. Xu, K. Bai, S. Zhu, Taplogger: inferring user inputs on smartphone touchscreens using on-board motion sensors, Proceedings of the Fifth ACM Conference on Security and Privacy in Wireless and Mobile Networks (2012) 113–124.
[12] M. De Choudhury, M. Gamon, S. Counts, E. Horvitz, Predicting depression via social media, ICWSM 13 (2013) 1–10.
[13] Jesper Pedersen, J.T.M. Schelde, E. Hannibal, K. Behnke, B.M. Nielsen, M. Hertz, An ethological description of depression, Acta Psychiatr. Scand. 78 (3) (1988) 320–330.
[14] Y. Yang, C. Fairbairn, J.F. Cohn, Detecting depression severity from vocal prosody, IEEE Trans. Affect. Comput. 4 (2) (2012) 142–150.
[15] Zhihong Zeng, Maja Pantic, Glenn I. Roisman, Thomas S. Huang, A survey of affect recognition methods: audio, visual, and spontaneous expressions, IEEE Trans. Pattern Anal. Mach. Intell. 31 (1) (2008) 39–58.
[16] Yu Zhu, Yuanyuan Shang, Zhuhong Shao, Guodong Guo, Automated depression diagnosis based on deep networks to encode facial appearance and dynamics, IEEE

Trans. Affect. Comput. 9 (4) (2017) 578–584.
[17] Namunu C. Maddage, Rajinda Senaratne, Lu-Shih Alex Low, Margaret Lech, Nicholas Allen, Video-based detection of the clinical depression in adolescents, in 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2009) 3723–3726.
[18] S. Alghowinem, R. Goecke, M. Wagner, G. Parker, M. Breakspear, Eye movement analysis for depression detection, in 2013 IEEE International Conference on Image Processing (2013) 4220–4224.
[19] L. Wen, X. Li, G. Guo, Y. Zhu, Automated depression diagnosis based on facial dynamic analysis and sparse coding, IEEE Trans. Inf. Forensics Secur. 10 (7) (2015) 1432–1441.
[20] Shizhe Chen, Qin Jin, Jinming Zhao, Shuai Wang, Multimodal multi-task learning for dimensional and continuous emotion recognition, Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge (2017) 19–26.
[21] M. Morales, S. Scherer, R. Levitan, A linguistically-informed fusion approach for multimodal depression detection, Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic (2018).
[22] D.E. Ness, S.F. Kiesling, Language and connectedness in the medical and psychiatric interview, Patient Educ. Couns. 68 (2) (2007) 139–144.
[23] J.E. Wiersma, P. van Oppen, D.J. Van Schaik, A.J. Van der Does, A.T. Beekman, B.W. Penninx, Psychological characteristics of chronic depression: a longitudinal cohort study, J. Clin. Psychiatry 72 (3) (2011) 288–294.
[24] T. Matsudaira, T. Kitamura, Personality traits as risk factors of depression and anxiety among Japanese students, J. Clin. Psychol. 62 (1) (2006) 97–109.
[25] Elise Sloan, Kate Hall, Richard Moulding, Shayden Bryce, Helen Mildred, Petra K. Staiger, Emotion regulation as a transdiagnostic treatment construct across anxiety, depression, substance, eating and borderline personality disorders: a systematic review, Clin. Psychol. Rev. 57 (2017) 141–163.
[26] J.M. Romano, J.A. Turner, Chronic pain and depression: does the evidence support a relationship? Psychol. Bull. 97 (1) (1985) 18.
[27] M.H.M. De Moor, A.L. Beem, J.H. Stubbe, D.I. Boomsma, E.J.C. De Geus, Regular exercise, anxiety, depression and personality: a population-based study, Prev. Med. 42 (4) (2006) 273–279.
[28] L. Zhou, An empirical investigation of deception behavior in instant messaging, IEEE Trans. Prof. Commun. 48 (2) (2005) 147–160.
[29] J.B. Persons, The Case Formulation Approach to Cognitive-behavior Therapy, Guilford Press, 2012.
[30] Kathryn L. Schaefer, Jacqueline Baumann, Brendan A. Rich, David A. Luckenbaugh, Carlos A. Zarate Jr, Perception of facial emotion in adults with bipolar or unipolar depression and controls, J. Psychiatr. Res. 44 (16) (2010) 1229–1235.
[31] C.J. Ranelli, R.E. Miller, Behavioral predictors of amitriptyline response in depression, Am. J. Psychiatry (1981).
[32] T.E. Joiner, K.A. Timmons, Depression in its interpersonal context, Handbook of Depression 2 (2002), pp. 322–339.
[33] V.N. Giri, Nonverbal communication theories, Encyclopedia of Communication Theory, (2009), pp. 690–694.
[34] H. Stassen, S. Kuny, D. Hell, The speech analysis approach to determining onset of improvement under antidepressants, Eur. Neuropsychopharmacol. 8 (4) (1998) 303–310.
[35] A. Nilsonne, Speech characteristics as indicators of depressive illness, Acta Psychiatr. Scand. 77 (3) (1988) 253–263.
[36] J.K. Darby, N. Simmons, P.A. Berger, Speech and voice parameters of depression: a pilot study, J. Commun. Disord. 17 (2) (1984) 75–85.
[37] M. Hamilton, A rating scale for depression, J. Neurol. Neurosurg. Psychiatr. 23 (1) (1960) 56.
[38] Albert F.G. Leentjens, Frans R.J. Verhey, Richel Lousberg, Harro Spitsbergen, Frederik W. Wilmink, The validity of the Hamilton and Montgomery-Åsberg depression rating scales as screening and diagnostic tools for depression in Parkinson's disease, Int. J. Geriatr. Psychiatry 15 (7) (2000) 644–649.
[39] A.T. Beck, R.A. Steer, G.K. Brown, Beck Depression Inventory-II 78 (1996), pp. 490–498 San Antonio (2).
[40] A.J. Rush, M.H. Trivedi, H.M. Ibrahim, T.J. Carmody, B. Arnow, D.N. Klein, J.C. Markowitz, P.T. Ninan, S. Kornstein, R. Manber, M.E. Thase, The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression, Biol. Psychiatry 54 (5) (2003) 573–583.
[41] K. Kroenke, R.L. Spitzer, J.B. Williams, B. Löwe, The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review, Gen. Hosp. Psychiatry 32 (4) (2010) 345–359.
[42] Nicholas Cummins, Bogdan Vlasenko, Hesam Sagha, Björn Schuller, Enhancing speech-based depression detection through gender dependent vowel-level formant features, In Conference on Artificial Intelligence in Medicine in Europe, Cham, Springer, 2017, pp. 209–214.
[43] A. Dhall, R. Goecke, A temporally piece-wise fisher vector approach for depression analysis, International Conference on Affective Computing and Intelligent Interaction (ACII), September, 2015, pp. 255–259.
[44] P. Lopez-Otero, L. Dacia-Fernandez, C. Garcia-Mateo, A study of acoustic features for depression detection, In 2nd International Workshop on Biometrics and Forensics, March, 2014, pp. 1–6.
[45] P. Lopez-Otero, L.D. Fernández, A. Abad, C. Garcia-Mateo, Depression detection using automatic transcriptions of de-identified speech, Proc. Interspeech 2017 (2017) 3157–3161.
[46] X. Ma, H. Yang, Q. Chen, D. Huang, Y. Wang, Depaudionet: an efficient deep model for audio based depression classification, In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, October, 2016, pp. 35–42.
[47] A. Jan, H. Meng, Y.F.B.A. Gaus, F. Zhang, Artificial intelligent system for automatic depression level analysis through visual and vocal expressions, IEEE Trans. Cogn. Dev. Syst. 10 (3) (2017) 668–680.
[48] M.R. Morales, R. Levitan, Speech vs. Text: a comparative analysis of features for depression detection systems, In 2016 IEEE Spoken Language Technology Workshop (SLT), December, 2016, pp. 136–143.
[49] M. Nasir, A. Jati, P.G. Shivakumar, S. Nallan Chakravarthula, P. Georgiou, Multimodal and multiresolution depression detection from speech and facial landmark features, In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge (2016) 43–50.
[50] Z.S. Syed, K. Sidorov, D. Marshall, Depression severity prediction based on biomarkers of psychomotor retardation, In Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge, October, 2017, pp. 37–43.
[51] Y. Gong, C. Poellabauer, Topic modeling based multi-modal depression detection, In Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge, October, 2017, pp. 69–76.
[52] L. Yang, D. Jiang, X. Xia, E. Pei, M.C. Oveneke, H. Sahli, Multimodal measurement of depression using deep learning models, In ProceEdings of the 7th Annual Workshop on Audio/Visual Emotion Challenge, October, 2017, pp. 53–59.
[53] L. Low, Detection of clinical depression in adolescents' speech during family interactions, IEEE Trans. Biomed. Eng. 58 (3) (2011) 574–586.
[54] J. Joshi, A. Dhall, R. Goecke, J.F. Cohn, Relative body parts movement for automatic depression analysis, In 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, September, 2013, pp. 492–497.
[55] S. Alghowinem, R. Goecke, M. Wagner, G. Parker, M. Breakspear, Eye movement analysis for depression detection, In 2013 IEEE International Conference on Image Processing, September, 2013, pp. 4220–4224.
[56] S. Alghowinem, R. Goecke, M. Wagner, G. Parkerx, M. Breakspear, Head pose and movement analysis as an indicator of depression, In 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, September, 2013, pp. 283–288.
[57] Y.C. Shen, T.T. Kuo, I.N. Yeh, T.T. Chen, S.D. Lin, Exploiting temporal information in a two-stage classification framework for content-based depression detection, Pacific-Asia Conference on Knowledge Discovery and Data Mining, April, Springer, Berlin, Heidelberg, 2013, pp. 276–288.
[58] J.F. Cohn, T.S. Kruez, I. Matthews, Y. Yang, M.H. Nguyen, M.T. Padilla, F. Zhou, F. De la Torre, Detecting depression from facial actions and vocal prosody, In 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, September, 2009, pp. 1–7.
[59] H. Meng, D. Huang, H. Wang, H. Yang, M. Ai-Shuraifi, Y. Wang, Depression recognition based on dynamic facial and vocal expression features using partial least square regression, October, In Proceedings of the 3rd ACM International Workshop on Audio/visual Emotion Challenge (2013) 21–30.
[60] J. Joshi, R. Goecke, S. Alghowinem, A. Dhall, M. Wagner, J. Epps, G. Parker, M. Breakspear, Multimodal assistive technologies for depression diagnosis and monitoring, J. Multimodal User Interfaces 7 (3) (2013) 217–228.
[61] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Hyett, G. Parker, M. Breakspear, Multimodal depression detection: fusion analysis of paralinguistic, head pose and eye gaze behaviors, IEEE Trans. Affect. Comput. 9 (4) (2016) 478–490.
[62] F. Ringeval, B. Schuller, M. Valstar, J. Gratch, R. Cowie, S. Scherer, S. Mozgai, N. Cummins, M. Schmitt, M. Pantic, Avec 2017: Real-life depression, and affect recognition workshop and challenge, In Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge (2017) 3–9.
[63] T. Dang, B. Stasak, Z. Huang, S. Jayawardena, M. Atcheson, M. Hayat, P. Le, V. Sethu, R. Goecke, J. Epps, Investigating word affect features and fusion of probabilistic predictions incorporating uncertainty in AVEC 2017, In Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge (2017) 27–35.
[64] L.M. McDermott, K.P. Ebmeier, A meta-analysis of depression severity and cognitive function, J. Affect. Disord. 119 (1-3) (2009) 1–8.
[65] N.L. Tefera, L. Zhou, A Scorecard Method for Detecting Depression in Social Media Users, (2018).
[66] J. Gratch, R. Artstein, G.M. Lucas, G. Stratou, S. Scherer, A. Nazarian, R. Wood, J. Boberg, D. DeVault, S. Marsella, D.R. Traum, The Distress Analysis Interview Corpus of Human and Computer Interviews, In LREC, 2014, pp. 3123–3128.
[67] K. Kroenke, T.W. Strine, R.L. Spitzer, J.B. Williams, J.T. Berry, A.H. Mokdad, The PHQ-8 as a measure of current depression in the general population, J. Affect. Disord. 114 (1-3) (2009) 163–173.
[68] G. Degottex, J. Kane, T. Drugman, T. Raitio, S. Scherer, COVAREP—A collaborative voice analysis repository for speech technologies, In 2014 Ieee International Conference on Acoustics, Speech and Signal Processing (Icassp) (2014) 960–964.
[69] V. Jain, J.L. Crowley, A.K. Dey, A. Lux, Depression estimation using audiovisual features and fisher vector encoding, In Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge (2014) 87–91.
[70] G.E. Schwartz, P.L. Fair, P. Salt, M.R. Mandel, G.L. Klerman, Facial muscle patterning to affective imagery in depressed and nondepressed subjects, Science 192 (4238) (1976) 489–491.
[71] T. Baltrušaitis, P. Robinson, L.P. Morency, Openface: an open source facial behavior analysis toolkit, In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV) (2016) 1–10.
[72] H. Ellgring, Non-verbal Communication in Depression, Cambridge University Press, 2007.
[73] S. Dham, A. Sharma, A. Dhall, Depression Scale Recognition From Audio, Visual and Text Analysis, arXiv preprint arXiv:1709.05865 (2017).
[74] M. Al-Mosaiwi, T. Johnstone, In an absolute state: elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation, Clin. Psychol. Sci. 6 (4) (2018) 529–542.

[75] Y. Neuman, Y. Cohen, D. Assaf, G. Kedma, Proactive screening for depression through metaphorical and automatic text analysis, Artif. Intell. Med. 56 (1) (2012) 19–25.
[76] P.G.F. Cheng, et al., Psychologist in a pocket: lexicon development and content validation of a mobile-based app for depression screening, JMIR mHealth uHealth 4 (3) (2016) e88.
[77] A.T. Beck, Cognitive Therapy of Depression, Guilford press, 1979.
[78] T. Pyszczynski, J. Greenberg, Self-regulatory perseveration and the depressive self-focusing style: a self-awareness theory of reactive depression, Psychol. Bull. 102 (1) (1987) 122.
[79] U. Okkan, G. Inan, Statistical downscaling of monthly reservoir inflows for Kemer watershed in Turkey: use of machine learning methods, multiple GCMs and emission scenarios, Int. J. Climatol. 35 (11) (2015) 3274–3295.
[80] A. Colubri, T. Silver, T. Fradet, K. Retzepi, B. Fry, P. Sabeti, Transforming clinical data into actionable prognosis models: machine-learning framework and field-deployable app to predict outcome of Ebola patients, PLoS Negl. Trop. Dis. 10 (3) (2016) e0004549.
[81] M.J. Fard, S. Ameri, R. Darin Ellis, R.B. Chinnam, A.K. Pandya, M.D. Klein, Automated robot-assisted surgical skill evaluation: predictive analytics approach, Int. J. Med. Robot. Comput. Assist. Surg. 14 (1) (2018) e1850.
[82] T. Chai, R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature, Geosci. Model. Dev. 7 (3) (2014) 1247–1250.
[83] J. Bowlby, Loss: Sadness and Depression, Random House, 1998.
[84] L. Lecci, M.A. Okun, P. Karoly, Life regrets and current goals as predictors of psychological adjustment, J. Pers. Soc. Psychol. 66 (4) (1994) 731.
[85] D. Dooley, J. Prause, K.A. Ham-Rowbottom, Underemployment and depression: longitudinal relationships, J. Health Soc. Behav. (2000) 421–436.
[86] A.M. Hayes, C.G. Beevers, G.C. Feldman, J.P. Laurenceau, C. Perlman, Avoidance and processing as predictors of symptom change and positive growth in an integrative therapy for depression, Int. J. Behav. Med. 12 (2) (2005) 111.
[87] D. Zhang, L. Zhou, J.L. Kehoe, I.Y. Kilic, What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews, J. Manag. Inf. Syst. 33 (2) (2016) 456–481.
[88] N. Majumder, S. Poria, A. Gelbukh, E. Cambria, Deep learning-based document modeling for personality detection from text, IEEE Intell. Syst. 32 (2) (2017) 74–79.

Guohou Shan is a first year PhD student in the Department of Management Information System of Temple University. His research interests include Healthcare IT, blockchain, and online community. He has published and presented papers in Americas Conference in Information System, Workshop on e-Business, INFORMS, Journal of Computer Applications, etc.

Lina Zhou is a Professor in the Department of Business Information Systems and Operations Management at UNC Charlotte. Her research interests span the areas of social media analytics, deception detection, biomedical informatics, and intelligent mobile interface. She has (co-)authored articles published in journals such as *MIS Quarterly, Journal of Management Information Systems, Communications of the ACM, ACM Transaction, IEEE Transactions, Information and Management*, and *Decision Support Systems*.

Dongsong Zhang is a Belk Distinguished Professor in Business Analytics in the Department of Business Information Systems and Operations Management, at UNC Charlotte. He received his Ph.D. in Management Information Systems from the University of Arizona. His research interests include business intelligence, social media analytics, mobile HCI, and health IT. He has published approximately 150 research articles and received a dozen research grants and awards from National Science Foundation, National Institute of Health, and U.S. Department of Education, among other funding agencies.