

RESEARCH

Open Access



BERT and BERTopic for screening clinical depression on open-ended text messages collected through a mobile application from older adults

Moo-Kwon Chung^{1†}, Sang Yup Lee^{2†}, Taeksoo Shin^{3†}, Ji Young Park⁴, Sangwon Hwang⁵, Min-Hyuk Kim⁶, Jinhee Lee⁶, Kyoung-Joung Lee⁷, Hyo-Sang Lim⁸, Erdenebayar Urtnasan⁹, YeonSu Jung¹⁰, Dan-Kyung Kim¹¹, Eunji Shin¹² and Jin-kyung Lee^{13*}

Abstract

Background Despite the high suicide rate in South Korea, older adults are reluctant to see a psychiatrist. Recently, text mining has gained popularity to detect depression in social media posts, but older adults rarely use social media. However, more than 90% of them use smartphones. South Korea has also made a public effort to utilize a mobile application to manage chronic health problems. In these situations, this study explores the possibility of screening the risk of depression through textual data reporting major stressors collected from older adults via a mobile application.

Methods We collected the data regarding stress and depressive symptoms through our mobile application. Pre-trained Bidirectional Encoder Representations from Transformers (BERT)-based Natural Language Processing (NLP) models were utilized, using Python and the Hugging Face Transformers. A total of 1,332 text messages collected from 230 participants were analyzed using BERT modeling to detect clinical depression, as screened by the PHQ-9. For Korean data, we used KcBERT and KLUE BERT. BERTopic and dynamic BERTopic were used to see what stress topics appeared among a high-risk group and how they changed.

Results The results demonstrate that KcBERT (precision = .89, recall = .86, F1 score = .87) was slightly better than KLUE BERT (precision = .81, recall = .78, F1 score = .79), although both performed well in identifying clinical depression. In BERTopic results, hierarchical clustering were re-grouped into four categories: financial problems, family-oriented stressful situations, physical and mental health problems, and work-related or acutely stressful situations. Dynamic BERTopic results show longitudinal changes. While event-related words such as family death or disease diagnosis were found more often for the cases when depression risk increased, words related to continued stressful situations appeared more often when the risk remained high.

[†]Moo-Kwon Chung, Sang Yup Lee and Taeksoo Shin contributed equally to this work.

*Correspondence:

Jin-kyung Lee

2jinkyung.lee@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Conclusion These results imply that collecting respondents' reports regarding stressful experiences can be useful to screen the risk of clinical depression. Including this function within a smartphone application publicly administered by community health care professionals can help monitor mental health in older adults. It can approach a hidden high-risk population suffering from depression in the community, providing enriched information about their risk factors.

Keywords BERT, Text Mining, Topic Modeling, Depression, Older Adults

Introduction

Depression is one of the most prevalent mental health problems in the world. For example, 3.8% of the population experience depressive disorders in the world. However, the proportion of those experiencing depressive symptoms is 5.7% among older adults over the age of 60 years [1]. Older adults with mild behavioral impairment are more vulnerable to the risk of depression [2]. These statistics might even be underestimated because many people experiencing depressive symptoms are reluctant to visit a doctor and remain undiagnosed. One previous research conducted in the United States reports that 31.45% of adult participants experiencing a level of clinical depression when utilizing the Patient Health Questionnaire-9 (PHQ-9) were never diagnosed with a depressive disorder at a clinic [3]. Untreated depression is likely to develop suicidal thoughts longitudinally. In South Korea, the suicide rate has been at the top in the OECD countries for a decade [4]. Although the national Korean statistics reveal that suicide rates by age peak in older adults compared to younger age groups [5], older Korean adults are much less likely to seek professional help for their mental health problems because of a strong stigma in their generation.

Traditional depression screening methods, which consist of self-report surveys and in-depth interviews with a mental health professional, have been a long-time standard in the field due to their high validity. However, these traditional methods are limited in the fact that they are only available when a person visits a clinic and is willing to be evaluated by a mental health professional. Although people may suffer from depressive symptoms, limitations such as these contribute to a high rate of undiagnosed depression. To detect a hidden high-risk aged population, we need to develop a more accessible way to monitor depressive symptoms. Also, the questions used in a self-report survey as a traditional screening tool for depression reveal obvious intentions to monitor one's depressive symptoms. It is possible that a responder intentionally hides or underreports their symptoms when answering the questions.

Recently, the growing interest in text mining from social media posts has emerged as an academic attempt to find an alternative way to identify a hidden population with

a high risk of clinical depression [6, 7]. Collecting texts describing what they thought, felt, or experienced has the benefit of revealing their psychological state without the pressure of taking a survey and/or being evaluated. This method might offer us more opportunities to identify a hidden high-risk population suffering from depression. Previous studies have consistently demonstrated high accuracy rates of sentiment analysis for detecting depression using supervised machine-learning models on social media posts [8–12]. For example, Aliman et al. [8] show 81% accuracy from the logistic regression model, even though their SVM model indicates only 69% accuracy. Jain et al. [9] show 77.12% accuracy from the SVM model and 79.00% from the logistic regression model. Ghosal and Jain [10] show 70% accuracy from logistic regression and 71.1% from the XGB model. Sujithra et al. [11] show 74.22% accuracy from the XGB classifier model, 74.78% from the logistic regression model, and 75.12% from the random forest model. Obagbuwa et al. [12] show 95.2% accuracy from the random forest model, 96.1% from the XGB model, 96.2% from the SVM model, and 96.3% from the logistic regression model. Although the best model varies among the studies, these results suggest that text mining has the potential to be an unintrusive alternative way to screen for undiagnosed depression without triggering significant defensive psychological reactions, especially from the high-risk population who were highly reluctant to visit psychiatric clinics due to the taboo surrounding depression.

The problem is that it is not practical to implement further medical diagnoses or interventions for the users who were identified as high-risk individuals based on their posts on a social network. This is mainly because their social media posts were uploaded to freely share about their daily thoughts and lives, without a specific aim to be monitored for their mental health. Furthermore, despite older adults' increased vulnerability to developing clinical depression, they do not actively upload on social media in most cases. Instead, a mobile application could be a more efficient way for older populations to approach the general population to monitor their health risks and recommend further diagnosis, counseling, or treatment [13, 14]. In South Korea, a nationwide public effort has steadily attempted to develop a mobile application for

monitoring and caring for chronic diseases and general health conditions. As a result, a mobile application is regarded as a helpful tool for older Korean adults to monitor their health conditions. We think a mobile application managed by community health professionals can be a suitable medium to extend the accessibility of medical services to the hidden high-risk population with clinical depression, in addition to monitoring the progress of severe chronic diseases for a high-risk older population, regardless of their residential area.

Older Korean people have fewer psychological barriers to being connected to and cared for by professionals in local public health centers through a mobile application. Older adults have even become skilled at using smartphones to communicate with other people during COVID-19. Stress is a significant cause of physical and mental health problems. Compared to a strong stigma among older adults to visit a psychiatrist, their psychological barrier is low to reporting stressful experiences in their daily lives [13]. Considering the significant association between stress and depressive symptoms, this study starts with the question of whether older adults' open-ended text responses to our weekly survey asking about stress can help to identify people who are at a high risk of suffering from clinical depression. If their open-ended text messages about weekly stress help to detect who is at high risk of suffering from depression, we will also explore what word topics are frequently found in older Korean adults. This method may shed light on a novel, noteworthy approach to identifying a hidden high-risk population in an aging society.

Methods

Recruitment and data collection

This research protocol adhered to the Declaration of Helsinki Statement. After we explained the information about our research project in our recruitment process, people who voluntarily agreed to participate individually provided us with their written consent before participating in any data collection process. All procedures involving human subjects in the research project were approved by Yonsei University Mirae Campus Institutional Review Board (1,041,849–202010-SB-151–02).

In this study, we recruited the participants from a local community, not limiting it to a clinical sample. It was mainly because we aim to expand our benefits towards the general population instead of focusing solely on specialized clinical care. Our inclusion criteria were individuals aged 55 or older, with no alcohol or substance use disorders, and with no cognitive impairments. With support from the collaboration with Samsung Health, we developed a mobile application that connects to an Android smartphone and a Galaxy smartwatch to

gather the data exclusively for research purposes. However, we did not receive support from Apple, resulting in our study utilizing the Android operating system as the main platform. Thankfully, according to Gallup Korea, unlike the younger generation, more than 90% of older Korean adults over the age of 55 years utilize smartphones, and 80% of those individuals choose Android phones [15]. Because of this, we thought that studying the trends and patterns among Android users may apply to the general population over the age of 55 years. Regarding survey measures, we designed face-to-face interviews and mobile app survey questionnaires to monitor weekly stress and monthly depressive symptoms in older adults [16]. Of the participants who completed the baseline face-to-face interview during their one-day visit from December 2020 to May 2021, 411 older adults agreed to participate in our longitudinal data collection via this mobile application installed on their Android smartphones. All participants provided written informed consent before participating in this longitudinal data collection. Mobile survey data were collected from March 2021 to March 2023 through our mobile application. Our research protocol, including the mobile app survey questions used in this study, was published in Lee et al. [16].

In our mobile application, we asked participants every Sunday about stress in their lives over the past 7 days. Our weekly survey consisted of three questions. The first question asked each participant whether there was a life stressor that hindered their everyday lives. The response to the first question was coded as a binary variable. If a participant said yes to the first question, the next question was open to ask the participant to choose one category for the most severe stressor during the past week. Seven categories were given as categories of the most severe stressors: work, relationships, major life events, health problems, financial problems, extraordinary accidents, and uncategorized. Finally, an open-ended question followed, enabling the participant to freely type in the text as much as they wanted to describe their life stressors experienced during the past week. During the last week of every month, participants were invited to complete the monthly survey of PHQ-9 [17] through our mobile application.

Data analysis

To investigate whether the open-ended text responses about stressors can help to identify clinical depression, this research uses the Bidirectional Encoder Representations from Transformers (BERT) for data analysis. In 2017, the collaboration among Google Brain, Google Research, and the University of Toronto presented an innovative approach called "Transformer" in the fields of large language models [18]. One year later, Google AI

Language developed BERT [19]. While the original transformer model has both encoders and decoders, BERT uses only encoders, reducing the time and costs of handling the text data. By using only encoder parts, BERT focuses on improving the prediction of the words within the input sentence. Reading the input sentence bidirectionally (i.e., reading words left to right and vice versa), BERT gathers rich information about numeric vector representations for each word. This enhances its performance in predicting a masked word within the input sentence. BERT is a pre-trained model involving two tasks, namely, (1) masked language modeling and (2) next sentence prediction, and it is possible to perform fine-tuning for classification tasks [20]. Because of these characteristics, BERT is often used for sentiment analysis. Since our data has the issue of class imbalance, over-sampling and under-sampling methods under our supervision were used for the data preprocessing.

In the text preprocessing stage, all the collected open-ended textual responses underwent human review first. Three trained researchers cross-reviewed the data multiple times to ensure each participant's responses met the minimum quality prior to text mining analysis. The collected text data should contain relevant information regarding stressors. In addition, we checked that all essential parts were included consistently in the responses. If something delivering a necessary part for analysis (e.g., to whom, when, where, how, why) was answered in the previous response but was skipped in the following responses in a participant's description of the same stressor while their responding to our stress surveys longitudinally, we added it so that each participant's responses include required information at minimum level, following a consistent format. After that, the Kiwi morphological analyzer [21] was used to segment the text data into morphemes and to identify the base form of each morpheme. This process includes a text cleaning step to remove unnecessary symbols. To predict the risk of depression using the text data, we employed BERT-based text classification models that have been pre-trained using the Hugging Face Transformers library in Python, namely, KcBERT and KLUE BERT [22]. During model training, the document length was set to 20 tokens, as this configuration yielded superior classification performance on the test data relative to alternative lengths, such as 10 and 30 tokens. Approximately 95% of the documents were shorter than or equal to 20 tokens. For model training, 80% of the annotated text data was used. And the remaining data was used for model testing. In training data, the model's performance would be poor when the data imbalance is severe. To adjust this, we used random undersampling to reduce the size of the majority class and random oversampling to increase the

size of the minority class. We also fine-tuned KcBERT and KLUE BERT by using TFBertForSequenceClassification and the Adam optimizer with a learning rate of 0.00002. To mitigate the possible overfitting problem, we employed the early stopping technique with a patience of 3. This process utilized 10% of the training data as a validation dataset. We set the epoch to 10 and batch size to 32. For BERT-based classification models, we used the codes presented in Lee [23].

In addition to BERT-based classification models, we used BERTopic to visualize what word topics appear in the text data identified as high risk. BERTopic utilizes BERT embeddings for clustering topics, thereby producing interpretable information from text analysis. BERT is suitable for model evaluation, but BERT itself does not produce interpretable information from text analysis. However, using BERTopic enables it to visualize which words belong to the same topic, the number of topics that appear, and which topics can be merged. For BERTopic analysis, we use the KcBERT-based Beomi package from Hugging Face to handle Korean text data. For BERTopic, we used the codes presented in Lee [23]. Overall, Fig. 1 illustrates the entire process from data collection to data analyses of the present study.

Lastly, dynamic topic modeling with BERTopic was implemented for the entire text data used in this study ($N = 1,332$) to explore the longitudinal changes in topics found in textual responses. For Dynamic BERTopic modeling, we used the codes presented in Grootendorst [24]. Figure 4 illustrates the changes in topics over time. Furthermore, by matching depression scores reported in that corresponding month and comparing it to the previous depression score, we classified each case into five types: (1) constantly no risks (PHQ-9 total score: 0–4), (2) constantly mild risks (PHQ-9 total score: 5–9), (3) decreased risk (previous PHQ-9 level > current PHQ-9 level), (4) increased risk (previous PHQ-9 level < current PHQ-9 level), and (5) constantly high risks (PHQ-9 total score: 10+). After that, we explored topics by type. The results of frequently observed topics, especially for (4) the textual responses with an increased risk of depression and (5) the textual responses with a constantly high risk of depression, are presented in Fig. 5.

Results

BERT model results for detecting depression risks by textual responses

Of 411 total participants who agreed to install our mobile application to collect longitudinal data, 388 participants (94.40%) answered weekly stress surveys 20,820 times on our mobile application during the overall data collection period. Out of 20,820 responses from 388 participants, 19,117 responses (91.82%) from 385 participants

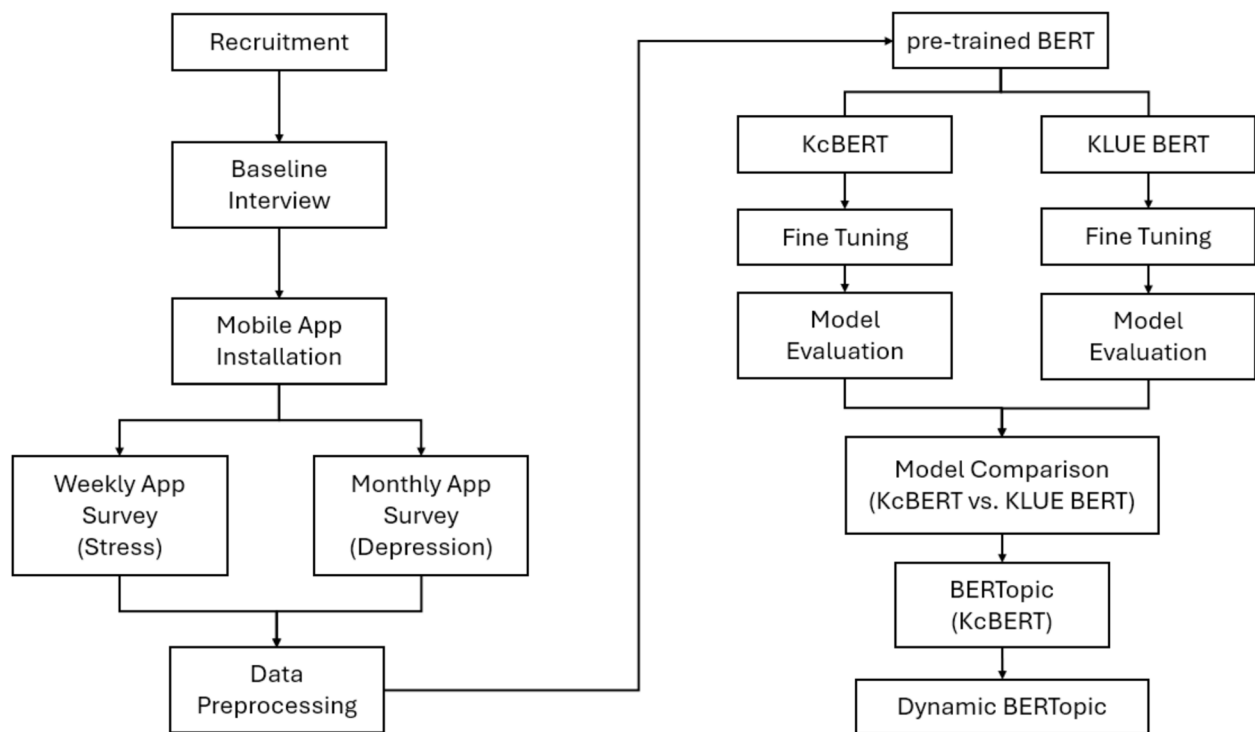


Fig. 1 Data collection and analysis process

reported that they did not experience any severe stress that hindered their daily lives. However, 1,703 responses (8.18%) from 230 participants indicated that they experienced at least one major life stressor. These 1,703 responses also had text responses for describing their stressors. Among 1,703 text responses, it was possible to match 1,332 responses (78.21%) with monthly PHQ-9 scores in the corresponding month, but 371 (21.78%) failed to complete a monthly PHQ-9 survey.

We ran the sensitivity analysis to see if participants who did not complete the PHQ-9 survey systematically (for $n = 371$, 21.78%) differ from those who did (for $n = 1,332$, 78.21%). The sensitivity analysis results demonstrated that missingness in the monthly PHQ-9 surveys is more likely to happen when the participant was a male ($OR = 0.53$, $p < 0.001$), at an older age ($OR = 1.07$, $p < 0.001$), less educated ($OR = 0.83$, $p < 0.001$), and having less low mood each participant daily reported during that week ($OR = 0.71$, $p < 0.01$). Considering that female, older, less educated older adults, and experiencing more depressed mood were more likely to have higher scores in PHQ-9 total scores, we were not able to see consistent directions in the association between our sample characteristics related to missingness in PHQ-9 surveys and their assuming PHQ-9 scores. In addition, there was no significant difference in the monthly PHQ-9 response missingness by stress

level ($OR = 1.00$, $p = 0.804$), the frequencies of drinking alcohol ($OR = 0.97$, $p = 0.694$), the status of living with a partner ($OR = 1.10$, $p = 0.677$), and whether the participant had ever experienced a depression episode(s) based on the MINI interview in our annual face-to-face interviews ($OR = 0.88$, $p = 0.411$). These results imply that the missingness in monthly PHQ-9 surveys caused some loss of text data, but using complete cases did not create critical issues in analyzing the association between our text data and PHQ-9 scores.

Of 1,332 text responses with PHQ-9 scores, we assigned 80% of the data as train data and 20% as test data. If a PHQ-9 score in a monthly survey was equal to or greater than 10, it was coded as 1 (a high-risk group in clinical depression). Otherwise, if a PHQ-9 score in a monthly survey was smaller than 10, it was coded as 0 (no risk group). As a preliminary analysis, we ran the logistic regression. We wanted to see whether the source of the most severe stressor is associated with the prevalence of depression. The preliminary results demonstrated that the cases when the participants experienced a severe level of stress due to a financial problem ($OR = 3.66$, $p < 0.001$), an unexpected accident ($OR = 3.05$, $p < 0.01$), a major life event ($OR = 2.80$, $p < 0.01$), or a relationship problem ($OR = 2.13$, $p < 0.01$) had higher risks of depression compared to the cases when the most severe stressor being uncategorized.

Table 1 Results of predicting clinical depression using BERT modeling

	KcBERT	KLUE BERT
Precision		
Normal	0.89 (95% CI: 0.84, 0.94)	0.84 (95% CI: 0.78, 0.90)
Depressed	0.90 (95% CI: 0.83, 0.97)	0.78 (95% CI: 0.68, 0.89)
Overall	0.89 (95% CI: 0.85, 0.93)	0.81 (95% CI: 0.75, 0.86)
Recall		
Normal	0.95 (95% CI: 0.92, 0.98)	0.91 (95% CI: 0.87, 0.96)
Depressed	0.77 (95% CI: 0.69, 0.86)	0.64 (95% CI: 0.53, 0.75)
Overall	0.86 (95% CI: 0.82, 0.90)	0.78 (95% CI: 0.73, 0.83)
F1 score		
Normal	0.92 (95% CI: 0.92, 0.98)	0.87 (95% CI: 0.83, 0.91)
Depressed	0.83 (95% CI: 0.77, 0.89)	0.70 (95% CI: 0.63, 0.77)
Overall	0.87 (95% CI: 0.84, 0.90)	0.79 (95% CI: 0.75, 0.83)

By utilizing BERT models for Korean text data, we have investigated how much text responses regarding weekly stress predict the risk of depression. Precision, recall, and F1-score were used as the model evaluation metrics. Table 1 presents the results of depression detection from KcBERT and KLUE BERT models. As shown in

Table 1, the KcBERT model performs slightly better than the KLUE BERT model. In the KcBERT model, the precision for the non-depressed is 0.89 (95% CI: 0.84, 0.94), while the precision for the depressed is 0.90 (95% CI: 0.83, 0.97). However, the recall (0.95; 95% CI: 0.92, 0.98) and the f1-score (0.92; 95% CI: 0.92, 0.98) for the non-depressed were higher than the recall (0.77; 95% CI: 0.69, 0.86) and the f1-score (0.83; 95% CI: 0.77, 0.89) for the depressed. In the KLUE BERT model, the precision (0.84; 95% CI: 0.78, 0.90), recall (0.91; 95% CI: 0.87, 0.96), and f1-score (0.87; 95% CI: 0.83, 0.91) for the non-depressed were higher than the precision (0.78; 95% CI: 0.68, 0.89), recall (0.64; 95% CI: 0.53, 0.75), and f1-score (0.70; 95% CI: 0.63, 0.77) for the depressed (Table 1).

BERTopic results for high-risk groups

Next, we ran the BERTopic to create word clusters when clinical depression was detected. Figure 2 presents the bar chart of the 16 overall topic clusters from the BERTopic. It shows the list of words for each topic, a good tool for visualizing the topics interpretably. The number of word topics is automatically generated, but some highly similar topics can be merged based on shared characteristics. Figure 3 illustrates the hierarchical

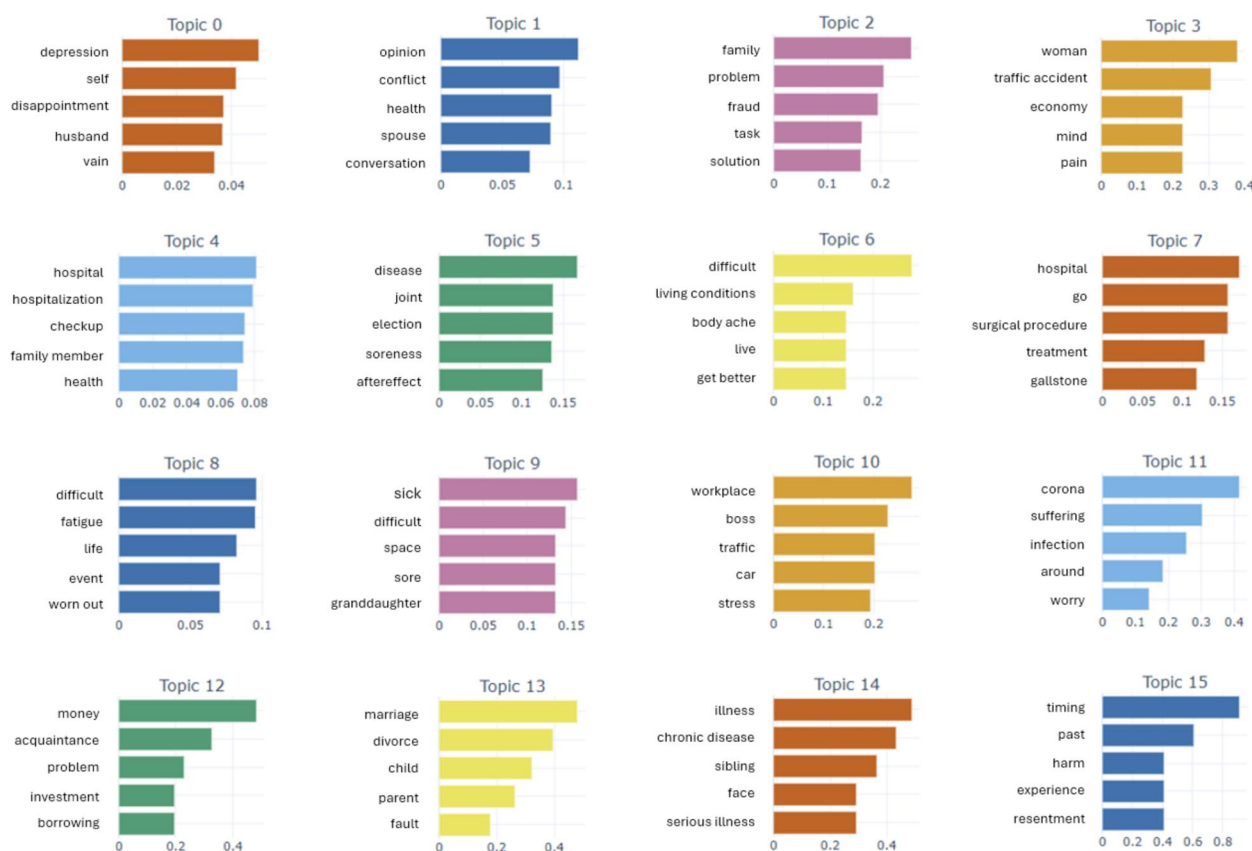


Fig. 2 Bar chart from the BERTopic

clustering results from BERTopic, indicating which topics can be incorporated. This graph displays the topics with the same colors based on semantic similarity. Semantic similarity means how close the two topics are in meaning. In addition to the results, we also conducted an external validation process with expert reviews. In the human validation process, we wanted to ensure whether the identified topics are meaningfully incorporated to capture distinctions in developing depressive symptoms. As a result, we applied a similarity threshold of 0.9 based on the colored graph and human validation. As shown in Fig. 3, we re-grouped 16 topics into four large categories. Firstly, we called the group of topics 15, 2, and 12 ‘financial problems.’ Secondly, topics 6, 9, 13, and 14 were grouped as ‘family-related stressful situations with physical symptoms.’ Thirdly, topics 5, 7, 4, 1, 0, 8, and 11 were included as the group of ‘physical and mental health problems.’ Lastly, topics 10 and 3 belonged to the group called ‘work-related or acutely stressful situations.’

Dynamic BERTopic to explore longitudinal changes

Furthermore, we illustrated longitudinal changes of BERT topics over time using dynamic BERTopic modeling with all textual responses ($N=1,332$). The results presented 23 topics in total, and the most frequently appeared 12 topics among them are health problems (T0), stressful incidents (T1), conflicts with a spouse (T2), problems faced by children (T3), family’s death or fight with serious illness (T4), concerns related to a spouse (T5), COVID-19 or traumatic events (T6), parent’s disease (T7), one’s own sickness (T8), job stress (T9), issues with family in law (T10), family hospitalization (T11), and arguments with family (T12). As anticipated, their frequencies changed over time. Figure 4 illustrates the longitudinal changes in

PHQ-9 total scores and BERT topics. As shown in Fig. 4, no single topic was directly matched with the fluctuation in the PHQ-9 total scores. However, we observed that depression levels reported by participants fluctuated over time, and frequently observed topics were also changed longitudinally. Among frequently observed topics, those related to health problems and stressful incidents (e.g., car accidents or the stock market) were more commonly reported by older adults.

Lastly, we examined the topics when the risk of depression increased and the topics when the risk of depression was constantly high (Fig. 5). When increasing depression risks, the topic of health problems (T0) was the most frequently found. Then, the topics related to stressful incidents (T1), family’s death or fight with illness (T4), parent’s disease (T7), problems faced by children (T3), COVID-19 or traumatic events (T6), one’s own sickness (T8), conflict with a spouse (T2), concerns with a spouse (T5), and job stress (T9) were followed in turn. When high depression risks constantly remained, the topic related to health problems (T0) was the most frequently reported, and the topic related to stressful incidents (T1) and the topic related to problems faced by children (T3) were followed.

Discussion

This study investigated whether open-ended responses describing life stress can help to detect the risk of clinical depression in older Korean adults. Although the risk of clinical depression increases with age, it is difficult to approach a potential high-risk group for timely treatment of depressive symptoms unless they are willing to seek professional help for mental health services. So far, social media posts have been investigated, but we think

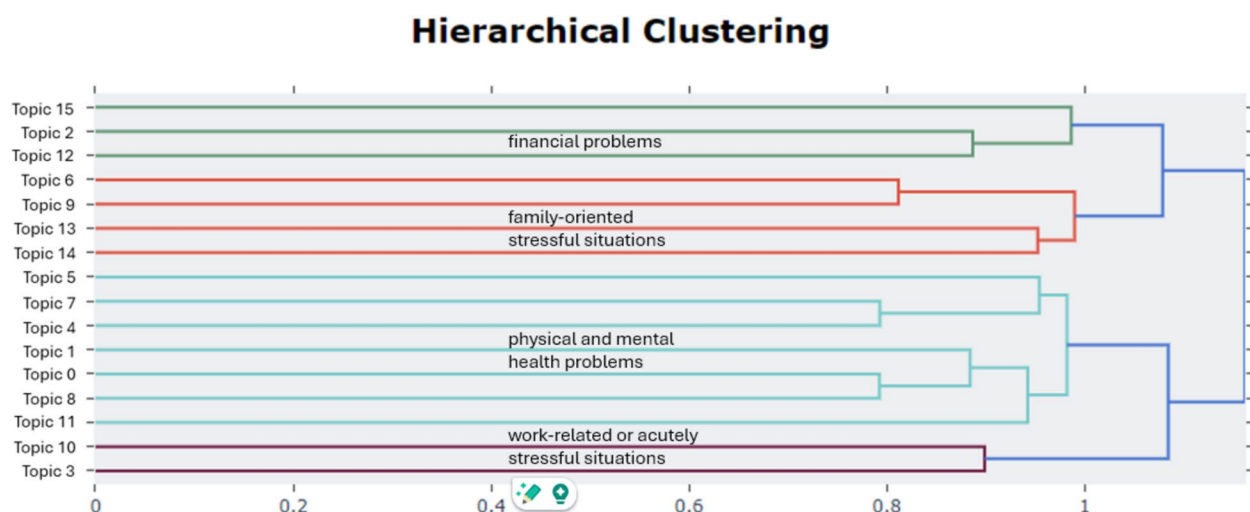


Fig. 3 Hierarchical clustering from the BERTopic

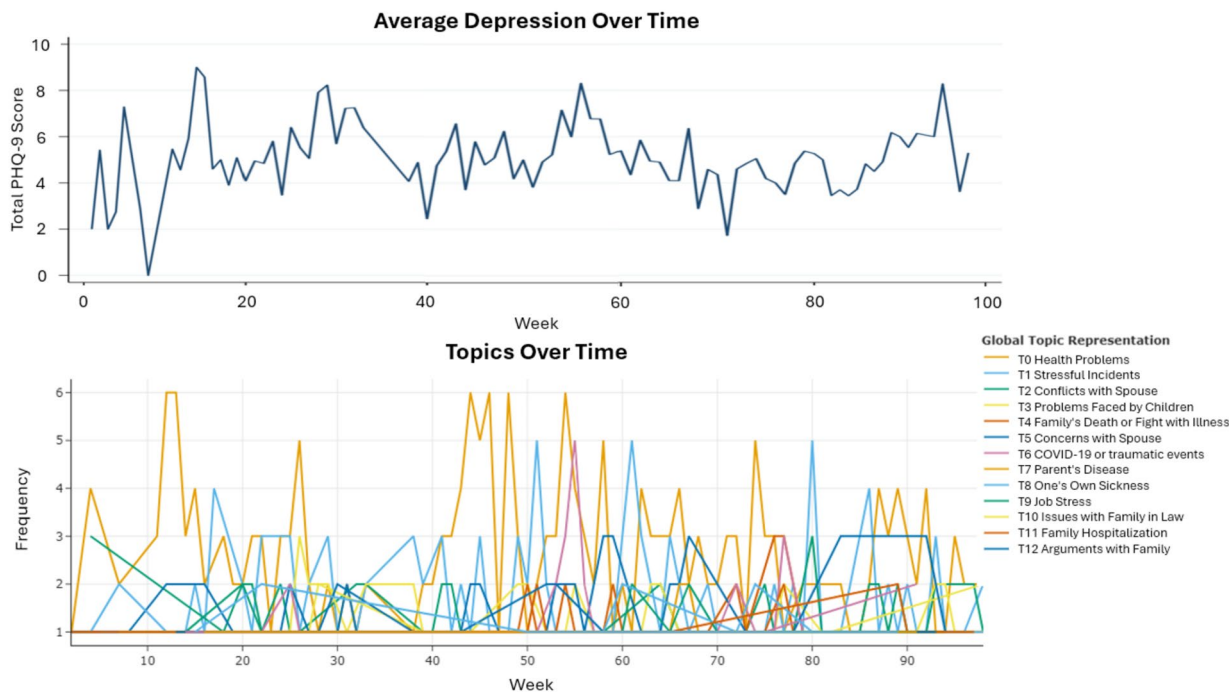


Fig. 4 Dynamic BERT topics over time

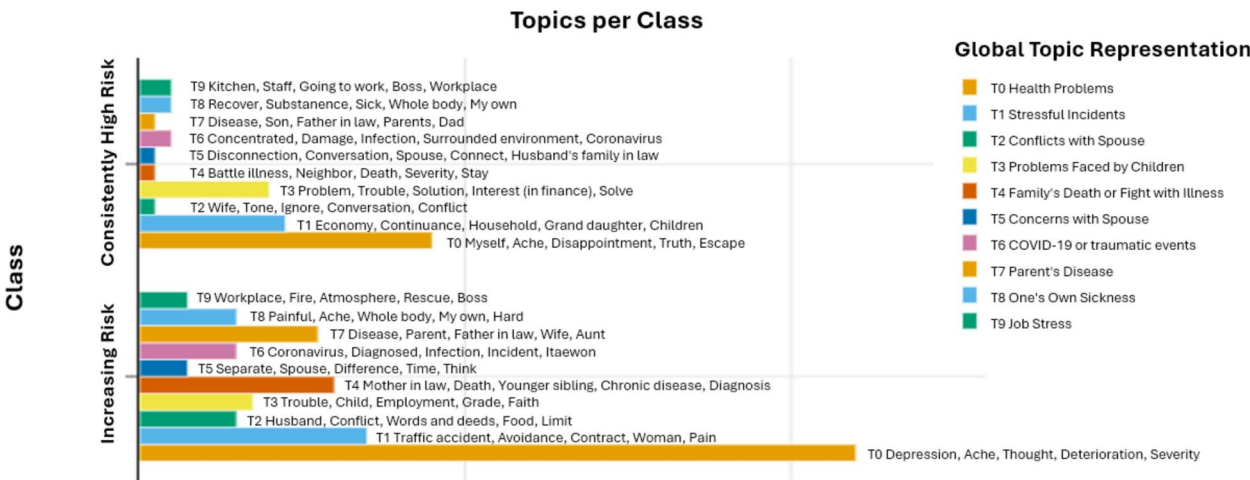


Fig. 5 Dynamic BERT topics during periods of constantly high or increasing depression risk

a mobile application managed by community health professionals would be a more efficient and accessible way to approach a hidden risk group to provide professional help and follow-up treatments. In this study, we investigated whether the risk of clinical depression is predicted by textual responses in which older adults express their stress as if they were confiding in friends via smartphones. Our significant results and derived implications using BERT modeling are as follows.

First, BERT modeling with text data collected through a mobile application in a natural setting has the potential to predict clinical depression among older adults. Generally, older adults tend to pay more attention to the things that stir positive emotions but have blunted responses to negative stimuli [25]. Also, older adults are likely to think of their depressive symptoms as a natural process of aging and rarely open up about their mental health problems to professionals. In contrast, they are likelier to open up

their non-medical problems to others [26]. In this study, we saw that only 1,703 out of 20,820 responses reported severe stress. This shows the tendency of older adults to present themselves as psychologically okay to others. However, once they reported that they experienced a severe level of stress, there was lower reluctance to type in plain text to describe what stress they faced. Compared to weekly open-ended surveys regarding stress, monthly PHQ-9 surveys had more missing data. For example, in our data analysis, only 78.16% ($n = 1,331$) out of 1,703 responses were available for PHQ-9 scores. This reveals that older adults had a lower burden of describing their life stress than completing a PHQ-9 questionnaire. In this context, text mining has noticeable potential to screen the risk of clinical depression in older adults. A friendly mobile application allowing respondents to report their stressors would help monitor their mental health risks and promptly introduce them to available mental health services in the community.

Second, when using text mining analysis for early detection to identify the risk group of depression in older adults, appropriately handling imbalanced data is the key to enhancing the quality of developing a prediction algorithm. According to the National Mental Health Survey 2021 in South Korea, the 12-month prevalence of clinical depression is 1.8% to 3.1% among the Korean population aged 50 to 70 [27]. Generally, the expected prevalence of clinical depression in older adults is 31.74% worldwide, and it decreases to 17.05% in developed countries [28]. Based on this, text mining for monitoring the risk of clinical depression for the general population inherently has the issue of appropriately handling imbalanced data. In our data, text mining using the raw imbalanced data did not display a good performance in identifying the high-risk group of depression. Previous studies demonstrate that Synthetic Minority Oversampling Technique (SMOTE) or Easy Data Augmentation (EDA) and undersampling strategies can improve the performance of classification models with imbalanced data [29, 30]. However, applying data preprocessing such as automatic over-sampling and under-sampling methods to the raw text data did not significantly improve our model performance due to the data characteristics. Instead, the human review process before data analysis, which ensures that the text data includes relevant information and fulfills the minimum quality level, significantly enhances the classification model's performance. This might be partly because some older male adults tended to report only a few words in their answers, and this short text message would not have been enough to figure out the context in which the responding words were indicated. This may hinder the classification of at-risk and non-risk groups for depression when determining

it based on the reported words. However, when sorting all the text messages by each participant and monitoring the patterns in the text responses longitudinally, human professionals could catch what words were abbreviated and intuitively distinguish whether the remaining word (even if this is just one word) was used positively or negatively. In addition, since the population of older Korean adults is more diverse in their education levels and physical functions (e.g., eyesight weakening in older adults), it would be more efficient if we could gather the data not only by their typing text but also by recording their voice to report their life stressors. Thankfully, current cutting-edge technology enables translating voice data into text data. Linking this function to text-mining would lessen the barriers to approaching participants' stress experiences. By gathering richer information about the context from which their life stress originated, we expect to further improve the performance of the text mining model. Additionally, human review can also be helpful after data analysis. In the present study, for instance, we proceeded with a human review process after data analysis to think about the implications the produced results brought to us in more depth. In the BERTopic modeling, human professionals monitored the results automatically created by algorithms to ensure whether the results make sense and the implications are reasonably meaningful.

Third, the significant findings of the present study show that KcBERT performs better than KLUE BERT in analyzing the Korean text data collected in our research project. KLUE BERT is a pre-trained model for the Korean language based on the BERT model [22]. KLUE BERT was trained by five publicly available Korean datasets: Modu Corpus including formal articles and dialogues, Korean portion from CC-100 using web crawling, Namu-wiki encyclopedia written in Korean languages, Newscrawl including 12,800,000 news articles, and Petition including public petitions posted to the Blue House [22]. While formal languages primarily trained KLUE BERT, informal languages mainly trained KcBERT. The full name of KcBERT is "Korean comments BERT." Based on the need to analyze spoken Korean languages, KcBERT was developed as a pre-trained Korean BERT model using the threads of replies to online news articles [31]. Considering the significant characteristics of the two models, the reason why the KcBERT showed better performance than KLUE BERT would come from the fact that our data collected in this research project are close to text messages between people, including informal spoken Korean languages, rather than well-refined formal Korean languages found in official articles. More research is recommended to see the replicability of this finding with other data.

Fourth, for older Korean adults with a high risk of clinical depression, 16 topics appeared regarding significant

life stressors in total. The results of BERTopic can be used to prepare the contents regarding what additional information would be provided and what community organizations would be connected as a follow-up in the mobile application. When a high risk of developing or suffering from clinical depression is detected in the mobile application, these contents can be displayed to a high-risk individual so that this hidden high-risk individual can navigate possible supports in the community proactively. In our results, 16 topics were clustered into four groups. Four categories, namely, financial problems, family problems with physical symptoms, health problems, and work-related or acutely stressful situations. These results align with previous research investigating the associations between stress and depression using traditional research methods. For example, a recent systematic review [32] elaborates that financial stress is positively associated with depression. The review [32] also reports that a subjective psychological burden from financial issues is riskier to depression than an objective financial burden (e.g., the amount of debt). In our study, financial stressors are often related to unexpected traumatic events with monetary issues (e.g., experiences of monetary fraud from others or financial problems with acquaintances borrowing money). Stress from these situations would have brought more detrimental effects on older adults' mental health because these cases involved breaking trust in close relationships, in addition to the financial problem itself.

In terms of family-related stressors and depression, our findings consistently support the findings of previous research. For example, a recent systematic review and meta-analysis [33] analyzing the association between family functionality and depression demonstrates that family dysfunction is positively associated with depression ($OR = 3.72$). Those who experienced family dysfunction are more likely to have greater social vulnerability (i.e., one's susceptibility to physical, relational, and environmental factors), and this increases the risk of depression compared to those who did not experience family dysfunction [33]. Furthermore, those with family dysfunction are likely to develop chronic diseases, which are linked to the heightened risk of depression [33]. According to our data, the word lists for this category (Topics 6, 9, 13, and 14) reveal stressful situations that occurred in family relationships as a chronic stressor. At the same time, our findings also show a high level of stress when caring for ill family members.

Concerning physical and mental health problems, previous research demonstrates a bidirectional relationship between health problems and depression in older adults. On the one hand, the likelihood of experiencing visual or audio impairments increases as people age [34]. These

physical impairments worsen communication with others and heighten the risk of social isolation or social withdrawal, which increases the risk of depression in older adults [35]. On the other hand, older adults with depression have less motivation to manage their health conditions [36]. The feeling of helplessness and hopelessness that older adults with depression often experience can worsen their physical health conditions as they exhibit poor compliance with medical advice [34]. In particular, depression negatively influences inflammatory functions in the body and increases the risk of cardiovascular disease and diabetes mellitus [37]. More severe white matter hyperintensities were found in older adults with depression compared to those without depression [38, 39]. White matter hyperintensities indicate the degree of damage to the nerve cells. Thus, white matter hyperintensities are related to cerebrovascular diseases and executive functional deficits [34]. Interestingly, the BERTopic results of the present study also reveal that chronic illnesses (e.g., joint problems) as well as severe health problems (e.g., emergency followed by surgical procedures or coronavirus) make older adults psychologically vulnerable. Words listed in Topics 0 and 8 include depression, disappointment, fatigue, or worn-out, which implies the bidirectional relationship between depression and health problems. These findings reflect that depression with feelings of helplessness or hopelessness significantly worsens their quality of life.

In the matter of work-related or acute stress and depression in older adults, a classical review of work-related stress and depression reports that acute stressful situations at work contribute to depression [40]. Another empirical study investigating individuals never diagnosed with clinical depression at clinics [41] demonstrates that the participants experiencing depressive symptoms showed rapid reactivity to stress as well as exaggerated recovery from stress in cortisol responses when they were exposed to acute stress. These patterns are related to heightened risks of HPA axis dysregulation and allostatic load. Those findings from previous research support the BERTopic results of the present study that words regarding work-related stress and words regarding acute stress belong to the same category, and this is one of the prevalent categories found in older Korean adults suffering from depressive symptoms.

When examining longitudinal changes in topic frequencies over time, we observed fluctuations in both the depression scores and the frequencies of topics. In our data, the average PHQ-9 scores usually fluctuated within the level of mild depression ($PHQ-9 < 10$). Yet, we think it is worth monitoring. According to previous research that conducted a systematic review and meta-analysis [42], individuals with subthreshold depression have a high risk

of developing major depression. This result consistently appeared over different age groups and sample types (in both a community sample and a clinical sample) [42]. In the present study, the longitudinal changes in depression scores did not appear following a change in a single topic. Instead, there was a notable trend: average depression scores tended to increase around periods when topics regarding health problems, stressful incidents, and arguments with family were reported more frequently. This aligns with previous research supporting the dynamic longitudinal association between the occurrence of stressful life events and the increased risk of depressive symptoms [43, 44]. Our results suggest that text mining might have the potential to monitor depression risks longitudinally.

We also explored the specific words associated with periods when depression risks either increased or remained consistently high. In both cases, the topic related to health problems (T0) was the most commonly reported, followed by the topic regarding stressful incidents (T1). However, distinct patterns emerged beyond these. In particular, when the severity of depression increased, the topics related to family deaths or fighting illness, parents' disease, or specific events causing acute stress were frequently found. In contrast, when depression risk remained consistently high, the topic words were more associated with ongoing and chronic stressors, such as "economy," "continuance," and "household," which appeared in this case. These results align with previous review studies suggesting that chronic stressors tend to have more detrimental effects on mental and physical health than acute stressors. However, events causing acute stress are also risky in health conditions because of worsening pre-existing diseases [45, 46]. Of course, more research is recommended, considering that the relationship between stressful life events and depression is highly complex. Some stressful events can be more potent dependently (i.e., affected by their behaviors or thoughts) than other events, and the effect of a stressful life event on the development of depression can vary by individual and/or environmental risk factors [45, 47]. To explore more detailed background information, text mining can also be applied to a digital version of a traditional structured interview (e.g., the Life Events and Difficulties Schedule [48]) by using a mobile as a digital platform to identify key threatening life events that contribute to worsening mental health.

This paper focused on the sample that responded to our weekly stress and monthly depression surveys. However, the umbrella research project of our data used here primarily aims to collect longitudinal multimodal data to develop an algorithm predicting the risk of geriatric depression. The recruitment of the research project

occurred within the cohort of the Korean Genome and Epidemiology Study (KoGES)[16]. In South Korea, the prevalence of smartphones is high (> 90%) even among older adults. Thus, recruiting smartphone users for our research did not restrict the representativeness of our sample too much. Our sample includes various education levels and socioeconomic statuses (i.e., older adults who did not finish elementary school to older adults who graduated from graduate school; and older adults who worked in agriculture for their whole life to older adults who lived in an apartment and enjoyed urban life). Health-seeking behaviors also varied in this sample. Of course, older adults living in a different culture where smartphone use is not as prevalent might have different characteristics. Still, we believe that using a mobile application to collect older adults' open-ended reports regarding their major stressors would be practical for monitoring their mental health worldwide. This is mainly because smartphones can be intuitively designed so that people from diverse educational, socioeconomic, and digital literacy backgrounds easily learn and use a mobile application to monitor their health.

Limitations

The present study has several limitations. First, due to technical issues, our mobile application was initially only available for installation on Android phones. Since most older Korean adults use Android phones, we could still cover a representative portion of the older adult population. However, it would be more beneficial if future research could cover both Android and iPhone users. Second, our mobile application received text messages directly typed by the participants. This might have been particularly challenging for older adults with impaired vision or literacy issues, making it difficult to respond to our mobile surveys. Further research is recommended to develop a mobile application that enables participants to report their stressors in either text or voice and then converts all their responses into textual data. Third, since our sample was a community sample with the primary research purpose of helping a hidden high-risk population, we had small numbers of cases of clinical depression. Going forward, a more nuanced classification system would help to analyze longitudinal changes in finer detail. The current level of longitudinal analysis was our best. However, we believe that future research, collecting more clinical cases and developing a more nuanced classification system, could produce more useful information to analyze changes in topics and depression severity in greater depth. Fourth, the BERT model is an embedding technique based on contextual language encoder representations. Compared to other text mining methods, it is more potent at capturing context and

nuances in language. Despite its strength over other text mining methods, there may be potential limitations in providing sufficient contextual information, especially for short text responses. To keep meaningful details omitted in the repeated text data collection, we recommend a voice-to-text transformation technology in the data collection and a human review process in the data preprocessing stage. In addition, machine learning models may have possible issues, such as false positives or negatives. Also, for older adults in different cultures, the generalizability of the current study's findings may be limited by environmental factors, such as the extent to which society supports digital literacy and/or smartphone access among older adults.

Implications

Despite several limitations, the present study contributes to providing empirical evidence that collecting open-ended text messages describing major stressors could be valuable for the detection of incidences of clinical depression among the general community-dwelling older adult population. Although the population of older adults is known to be more likely to experience clinical depression, they remain a highly reluctant group to seek professional help for managing their mental health issues. The present study's findings have discovered that describing their stressors is far less burdensome on this population. In this situation, a mobile application that utilizes digital technology to collate real-time data regarding one's stressors can function as a platform for venting negative emotions and managing stress. At the same time, this technology enables the remote monitoring of depressive symptoms beyond their residential areas. Thus, patients with mobility issues would have lower burdens to access care. In South Korea, a high proportion of the elderly population lives in rural areas where medical services are more limited or insufficient. However, the availability of professional healthcare resources rapidly decreases in rural areas, making this population more vulnerable to accessing appropriate care in times of need. In this situation, text mining would be an invaluable tool to expand opportunities for early detection and timely intervention, namely, two keys to successfully treating geriatric depression.

Using digital technology, text mining enables real-time screening of mental health in older adults by vividly capturing their living experiences. This has critical implications for coping with depression among the geriatric community. First, it helps to identify specific targets within the community regarding geriatric depression. Secondly, it provides empirical evidence to develop strategic plans to protect high-risk groups. Thirdly, it enables an interdisciplinary and integrative approach

to support people with depression in managing various socio-demographic and psychological factors that influence their depressive symptoms. In particular, highlighting these hidden at-risk groups and understanding their specific needs through digital technology would benefit the community-based case management system. In South Korea, the community-based case management system is the main integrated geriatric support model. To provide effective case management for older adults with a risk of depression, it is crucial to find out when and what they need care. If our approach is integrated into a mobile application that older adults can easily access, it can provide more integrative care and support from the beginning.

Considering the ongoing demographic shift toward a more ageing population and the consequent surges in health care expenditures, the need for more preventive and community-based health care models has gained significant attention from scholars and policymakers worldwide [48, 49]. These models, which are built on a multidisciplinary network including public health agencies, nonprofit social service providers, and residents' associations, are proven effective in dealing with health-care disparities and strengthening early screening, particularly through the social prescribing model [48–51]. In several European countries, multidisciplinary community care models have been adopted not only to address physical health needs but also to respond to the growing burden of mental health problems in vulnerable populations [49, 52, 53]. This method is particularly promising for improving the effectiveness of mental health interventions for community-dwelling older adults with geriatric depression—a prevalent yet often underdiagnosed condition among older adults with limited access to mental health services [49, 54]. In this context, digital health technologies, primarily through smartphone applications integrated with natural language processing and text mining, would offer a powerful mechanism for advancing mental health care in community settings.

To execute this approach, we recommend that policymakers prioritize developing and deploying smartphone-based mental health screening tools at the community level. These tools should be carefully designed with several features in consideration of the challenges older adults may face in using digital interfaces. Key features such as voice-to-text input, large and intuitive interfaces, and culturally responsive design elements would substantially improve usability and user friendliness. Public health care providers can also play a central role in implementing these tools by collaborating in the development of service maps, facilitating connections with nearby mental health professionals, and effectively coordinating referrals. In turn,

public health institutions can serve as administrators of digital platforms, integrating these applications into existing health surveillance systems. By enabling the collection of real-time and open-ended data on current stressors and mood changes in users' natural settings, these digital tools support community healthcare workers in identifying at-risk individuals early and providing timely interventions. Furthermore, these models can significantly transform public health interventions for mental health issues with appropriate policy support and infrastructure. Developing a smartphone-based screening tool for mental health issues could be applied in much wider contexts, incorporating a broader range of age groups, SES, and different sociocultural backgrounds in the digital era.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12889-025-23337-4>.

Supplementary Material 1: Pseudocode and Python Implementation. To facilitate replication, codes for the BERT models used in this study are included in the Supplementary Material.

Acknowledgements

We appreciate all the participants for the precious time and openness they willingly shared with us for research purposes. We also appreciate administrative help from the Institute for Poverty Alleviation and International Development at Yonsei University and many research assistant students who worked hard with the hope of making a better society by supporting our research regarding mental health for older adults.

Authors' contributions

M-KC is the principal investigator for the research project. M-KC supervised how the present study can contribute to enhancing community-level mental health with relevant professionals. SYL analyzed the text data and supervised the methodological parts of this study. TS initiated the idea for the present study. JYP provided additional advice on the implications of this study. M-KC, TS, JYP, SH, M-HK, JL, K-JL, HSL, EU, and Jk-L contributed to the design and the implementation of the longitudinal research project regarding geriatric depression. YSJ, DKK, and ES contributed to data collection from the participants and data preprocessing used in this manuscript. J-kL proceeded with this research project longitudinally from data collection to data analysis and interpretation, and mainly wrote this manuscript.

Funding

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2020S1 A5 A2 A03045088).

Data availability

The datasets analyzed in the present study are not available in the public repository to protect sensitive information from the participants. Requests regarding the datasets should be directly addressed to the corresponding author.

Declarations

Ethics approval and consent to participate

All the participants voluntarily provided their written informed consent before their participation in this study. All procedures involving human subjects in the research project were approved by Yonsei University Mirae Campus Institutional Review Board (1041849–202010-SB-151–02).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Global Public Administration, Yonsei University Mirae Campus, Wonju, Republic of Korea. ²Department of Communication, Yonsei University Sinchon Campus, Seoul, Republic of Korea. ³Department of Business Administration, Yonsei University Mirae Campus, Wonju, Republic of Korea. ⁴Department of Social Welfare, Sangji University, Wonju, Republic of Korea. ⁵Department of Precision Medicine, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea. ⁶Department of Psychiatry, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea. ⁷President, Songho University, Hoengseong, Republic of Korea. ⁸Division of Software, Yonsei University Mirae Campus, Wonju, Republic of Korea. ⁹Institute of Yonsei AI Data Convergence Science, Yonsei University Mirae Campus, Wonju, Republic of Korea. ¹⁰Interdisciplinary Program in Comparative Literature, Yonsei University Sinchon Campus, Seoul, Republic of Korea. ¹¹Department of Public Administration, Korea University Seoul Campus, Seoul, Republic of Korea. ¹²Division of Health Administration, Yonsei University Mirae Campus, Wonju, Republic of Korea. ¹³Mo-Im Kim Nursing Research Institute, Yonsei University Sinchon Campus, Seoul, Republic of Korea.

Received: 21 January 2025 Accepted: 27 May 2025

Published online: 10 June 2025

References

- World Health Organization. Depressive disorder (depression). 2023. <https://www.who.int/news-room/fact-sheets/detail/depression>. Accessed 11 May 2025.
- Elefante C, Brancati GE, Pistolesi G, Amadori S, Torrigiani S, Baldacci F, Ceravolo R, Ismail Z, Lattanzi L, Perugi G. The impact of mild behavioral impairment on the prognosis of geriatric depression: preliminary results. *Int Clin Psychopharmacol*. 2024;39(5):305–12.
- Handy A, Mangal R, Stead TS, Coffee RL Jr, Ganti L. Prevalence and impact of diagnosed and undiagnosed depression in the United States. *Cureus*. 2022;14(8):e28011.
- Organisation for Economic Co-operation and Development. Suicide rates. 2022. <https://www.oecd.org/en/data/indicators/suicide-rates.html>. Accessed 11 May 2025.
- Statistics Korea. Statistics of causes of death by age groups. 2024. https://kosis.kr/statHtml/statHtml.do?orgId=101&dtbld=DT_1B34E01&conn_path=l2. Accessed 11 May 2025.
- Babu NV, Kanaga EGM. Sentiment analysis in social media data for depression detection using artificial intelligence: a review. *SN Comput Sci*. 2022;3:74.
- William D, Suhartono D. Text-based depression detection on social media posts: a systematic literature review. *Procedia Comput Sci*. 2021;179:582–9.
- Aliman G, Nivera TFS, Olazo JCA, Ramos DJP, Sanchez CDB, Amado TM, Arago NM, Jorda RL Jr, Virrey GC, Valenzuela IC. Sentiment analysis using logistic regression. *J Comput Innov Eng Appl*. 2022;7(1):35–40.
- Jain P, Srinivas KR, Vichare A. Depression and suicide analysis using machine learning and NLP. *J Phys Conf Ser*. 2022;2161: 012034.
- Ghosal S, Jain A. Depression and suicide risk detection on social media using fasttext embedding and xgboost classifier. *Procedia Comput Sci*. 2023;218:1631–9.
- Sujithra M, Rathika J, Velvadivu P, Marimuthu M. An intellectual decision system for classification of mental health illness on social media using computational intelligence approach. *J Ubiquitous Comput Commun Technol*. 2023;5(1):23–35.
- Obagbuwa IC, Danster S, Chibaya OC. Supervised machine learning models for depression sentiment analysis. *Front Artif Intell*. 2023;6: 1230649.
- Dietz M, Aslan I, Schiller D, Flutura S, Steinert A, Klebbe R, André E. Stress annotations from older adults-exploring the foundations for mobile ML-based health assistance. In: *Proceedings of the 13th EAI International*

- Conference on Pervasive Computing Technologies for Healthcare. 2019. <https://doi.org/10.1145/3329189.3329197>.
14. Laird B, Puzia M, Larkey L, Ehlers D, Huberty J. A mobile app for stress management in middle-aged men and women (calm): feasibility randomized controlled trial. *JMIR Form Res*. 2022;6(5): e30294.
15. Gallup Korea: 2012-2022 smartphone, smartwatch, and wireless ear-phone usage & brand survey. 2022. <https://www.gallup.co.kr/gallupdb/reportContent.asp?seqNo=1309>. Accessed 11 May 2025.
16. Lee JK, Kim MH, Hwang S, Lee KJ, Park JY, Shin T, Lim HS, Urtnasan E, Chung MK, Lee J. Developing prediction algorithms for late-life depression using wearable devices: a cohort study protocol. *BMJ Open*. 2024;14(6):e073290.
17. Löwe B, Kroenke K, Herzog W, Gräfe K. Measuring depression outcome with a brief self-report instrument: Sensitivity to change of the Patient Health Questionnaire (PHQ-9). *J Affect Disord*. 2004;81(1):61–6.
18. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. *Adv Neural Inf Process Syst*. 2017. <https://doi.org/10.48550/arXiv.1706.03762>.
19. Devlin J, Chang MW. Open sourcing BERT: state-of-the-art pre-training for natural language processing. 2018. <https://research.google/blog/open-sourcing-bert-state-of-the-art-pre-training-for-natural-language-processing/>. Accessed 11 May 2025.
20. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. *arXiv*. 2019. <https://doi.org/10.48550/arXiv.1810.04805>.
21. Lee MC. Python API for Kiwi. 2024. <https://github.com/bab2min/kiwipy>. Accessed 11 May 2025.
22. Park S, Moon J, Kim S, Cho WI, Han J, Park J, Song C, Kim J, Song Y, Oh T, et al. KLUKE: Korean language understanding evaluation. *arXiv*. 2021. <https://doi.org/10.48550/arXiv.2105.09680>.
23. Lee SY. Python text mining bible 2. Paju: Wikibook; 2023.
24. Grootendorst M. Dynamic topic modeling. 2024. https://maartengr.github.io/BERTopic/getting_started/topicsovertime/topicsovertime.html. Accessed 11 May 2025.
25. Charles ST, Mather M, Carstensen LL. Aging and emotional memory: the forgettable nature of negative images for older adults. *J Exp Psychol Gen*. 2003;132(2):310–24.
26. Murray J, Banerjee S, Byng R, Tylee A, Bhugra D, Macdonald A. Primary care professionals' perceptions of depression in older people: a qualitative study. *Soc Sci Med*. 2006;63(5):1363–73.
27. National Center for Mental Health. National Mental Health Survey 2021 (Government publications registration number 11–1352629-000065-01). Seoul: National Center for Mental Health; 2022.
28. Zenebe Y, Akele B, W/Selassie M, Necho M. Prevalence and determinants of depression among old age: a systematic review and meta-analysis. *Ann Gen Psychiatry*. 2021;20:55.
29. Danuri MSNM, Abd Rahman R, Mohamed I, Amin A. The improvement of stress level detection in twitter: imbalance classification using SMOTE. In: 2022 IEEE International Conference on Computing (ICOCO). 2022. <https://doi.org/10.1109/ICOCO56118.2022.10031684>.
30. Choi JW, Lee YJ, Lim CG, Choi HJ. A study on improving performance of software requirements classification models by handling imbalanced data. *Korea Inf Process Soc (KIPS) Trans: Softw Data Eng*. 2023;12(7):295–302.
31. Lee J. KcBERT: Korean comments BERT. In: Proceedings of the 32nd annual conference on human and cognitive language technology. 2020;10a:437–440.
32. Guan N, Guariglia A, Moore P, Xu F, Al-Janabi H. Financial stress and depression in adults: a systematic review. *PLoS One*. 2022;17(2): e0264041.
33. Guerrero-Muñoz D, Salazar D, Constaín V, Perez A, Pineda-Cañar CA, García-Perdomo HA. Association between family functionality and depression: a systematic review and meta-analysis. *Korean J Fam Med*. 2021;42(2):172–80.
34. Pocklington C. Depression in older adults. *Br J Med Pract*. 2017;10(1):a1007.
35. Bernabei V, Morini V, Moretti F, Marchiori A, Ferrari B, Dalmonte E, Ronchi DD, Rita Atti A. Vision and hearing impairments are associated with depressive-anxiety syndrome in Italian elderly. *Aging Ment Health*. 2011;15(4):467–74.
36. Evans M, Hammond M, Wilson K, Lye M, Copeland J. Treatment of depression in the elderly: effect of physical illness on response. *Int J Geriatr Psychiatry*. 1997;12(12):1189–94.
37. Win S, Parakh K, Eze-Nliam CM, Gottdiener JS, Kop WJ, Ziegelstein RC. Depressive symptoms, physical inactivity and risk of cardiovascular mortality in older adults: the cardiovascular health study. *Heart*. 2011;97(6):500–5.
38. Hickie I, Hickie I, Simons L, Naismith S, Simons J, Mccallum J, Pearson K. Vascular risk to late-life depression: evidence from a longitudinal community study. *Aust N Z J Psychiatry*. 2003;37(1):62–5.
39. Hickie I, Scott E, Wilhelm K, Brodaty H. Subcortical hyperintensities on magnetic resonance imaging in patients with severe depression—a longitudinal evaluation. *Biol Psychiatry*. 1997;42(5):367–74.
40. Tennant C. Work-related stress and depressive disorders. *J Psychosom Res*. 2001;51(5):697–704.
41. Fiksdal A, Hanlin L, Kuras Y, Gianferante D, Chen X, Thoma MV, Rohleder N. Associations between symptoms of depression and anxiety and cortisol responses to and recovery from acute stress. *Psychoneuroendocrinology*. 2019;102:44–52.
42. Lee YY, Stockings E, Harris M, Doi S, Page I, Davidson S, Barendregt J. The risk of developing major depression among individuals with subthreshold depression: a systematic review and meta-analysis of longitudinal cohort studies. *Psychol Med*. 2019;49(1):92–102.
43. Shriya A. The effect of lifetime cumulative adversity on change and chronicity in depressive symptoms and quality of life in older adults. *Int Psychogeriatr*. 2012;24(12):1988–97.
44. Kendler KS, Gardner CO. Depressive vulnerability, stressful life events and episode onset of major depression: a longitudinal model. *Psychol Med*. 2016;46(9):1865–74.
45. Cohen S, Murphy ML, Prather AA. Ten surprising facts about stressful life events and disease risk. *Annu Rev Psychol*. 2019;70(1):577–97.
46. Maier A, Riedel-Heller SG, Pabst A, Lupp A. Risk factors and protective factors of depression in older people 65+: a systematic review. *PLoS One*. 2021;16(5):e0251326.
47. Liu RT, Hamilton JL, Boyd SI, Dreier MJ, Walsh RF, Sheehan AE, Turnamian MR, Workman AR, Jorgensen SL. A systematic review and Bayesian meta-analysis of 30 years of stress generation research: clinical, psychological, and sociodemographic risk and protective factors for prospective negative life events. *Psychol Bull*. 2024;150(9):1021–69.
48. World Health Organization. Continuity and coordination of care: a practice brief to support implementation of the WHO framework on integrated people-centered health services. Geneva: World Health Organization; 2018.
49. Duncan F, Baskin C, McGrath M, Coker JF, Lee C, Dykxhoorn J, Adams EA, Gnani S, Lafortune L, Kirkbride JB, et al. Community interventions for improving adult mental health: mapping local policy and practice in England. *BMC Public Health*. 2021;21:1691.
50. Aughterson H, Baxter L, Fancourt D. Social scribing for individuals with mental health problems: a qualitative study of barriers and enablers experienced by general practitioners. *BMC Fam Pract*. 2020;21:194.
51. Bickerdike L, Booth A, Wilson PM, Farley K, Wright K. Social prescribing: less rhetoric and more reality. A systematic review of the evidence. *BMJ Open*. 2017;7:e013384.
52. Tops L, Beerten SG, Vandenbulcke M, Vermandere M, Deschodt M. Integrated care models for older adults with depression and physical comorbidity: a scoping review. *Int J Integr Care*. 2024;24(1):1–17.
53. Katschnig H, Hinkov H, Sooniste I, Simon L, Floritti A, McDonnell A, Madden T, O'Donnell E, de Carvalho A, Xavier M, et al. Joint action on mental health and well-being. Brussels: European Union; 2017.
54. Edwards-Smith A, Ajiboye A, Pywell S, Kenyon A, Routh F, Williams J. Adult mental health, major conditions and social prescribing: a rapid review. *Health Soc Care Community*. 2025. <https://doi.org/10.1155/hsc/2917260>.

Publisher's Note

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