Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Prediction of acute suicidal ideation in young adults using multi-dimensional scales: A graph neural network approach

Kyu Sung Choi
Graduate School of Medical Science and Engineering, Korea Advanced Institute for Science and Technology

Byung Hoon Kim
Department of Bio and Brain Engineering, Korea Advanced Institute for Science and Technology

Sung Hwan Kim
Graduate School of Medical Science and Engineering, Korea Advanced Institute for Science and Technology

Hong Jin Jeon
Samsung Medical Center, Sungkyunkwan University School of Medicine

Jong-Hoon Kim
Gachon University College of Medicine, Gil Medical Center

Joon Hwan Jang (jhjang602@snu.ac.kr)
Department of Human Systems Medicine, Seoul National University College of Medicine

Bumseok Jeong (bs.jeong@kaist.ac.kr)
Graduate School of Medical Science and Engineering, Korea Advanced Institute for Science and Technology

Research Article

Keywords: COVID-19, deep learning, depression, suicide, graph neural network

Posted Date: March 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-265513/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published at Scientific Reports on August 4th, 2021. See the published version at https://doi.org/10.1038/s41598-021-95102-7.

Loading [MathJax]/jax/output/CommonHTML/jax.js

Abstract

Precise remote evaluation of both suicide risk and psychiatric disorders is critical for suicide prevention as well as psychiatric well-being during COVID-19 crisis. Using questionnaires is an alternative to labour-intensive diagnostic interviews in a large general population, but previous models for predicting suicide attempts suffer from low sensitivity. We developed and validated a graph neural network model, *MindWatchNet*, which increased the prediction sensitivity of suicide risk in young adults (n = 17,482 for training; n = 14,238 for testing) using multi-dimensional questionnaires and suicidal ideation within 2 weeks as the prediction target. *MindWatchNet* achieved the highest sensitivity of 80.9% and an area under curve of 0.877 (95% confidence interval, 0.854–0.897). We demonstrated that multi-dimensional deep features covering depression, anxiety, resilience, self-esteem, and clinico-demographic information contribute to SI prediction. *MindWatchNet* might be useful in the remote evaluation of suicide risk in the general population of young adults for specific situations such as the COVID-19 pandemic.

Introduction

Suicide is the second leading cause of death in young adults (individuals 10–34 years old) in the US and is 2.5 times more frequent than homicides (48,344 vs. 18,830, respectively)[1]. In the last two decades, the total suicide rate increased to an all-time high of 35% in the US in 2018[1]. Suicidal ideation (SI) and suicide attempts (SAs), which are strong risk factors for completed suicide, are prevalent in the population (11–14% and 2.8–4.6%, respectively)[2]. Worldwide, the number of suicides is over 800,000 annually[3], and 60–70% of suicides die on the first or “index” attempt. Additionally, only approximately 30–40% of survivors received emergent hospital-level care[4,5]. Thus, accurate prediction of first SAs, or individuals with imminent suicide risk, followed by instantaneous intervention, would be effective in suicide prevention, leading to decreased mortality in young adults.

During pandemics such as the novel coronavirus disease 2019 (COVID-19) pandemic, remote mental health evaluation of self-isolating people to prevent viral spread is critical. To date, more than 43 million confirmed cases and one million deaths have been attributed to the COVID-19 pandemic, and global lockdowns are considered effective interventions to combat the virus[6]. However, these interventions, as well as the pandemic itself, can increase the potential for adverse outcomes on suicide risk[7]. As with the Spanish influenza pandemic in the US during 1918 – 19[8], the monthly suicidal rate increased by 16% in Japan during the second wave of the COVID-19 pandemic[9]. Symptoms of anxiety and depressive disorder markedly increased in the US during April–June 2020[10] compared with the same period in 2019[11]. Pre-existing psychiatric disorders are associated with increased SI as the psychological impact of the COVID-19 pandemic[12] and contribute to predicting future individuals with SI in young adult populations[13]. Moreover, younger adults reported having experienced disproportionately worse mental health outcomes and elevated SI than older adults[14]. Thus, the development of precise remote evaluation techniques of both suicide risk and psychiatric disorders is critical for suicidal prevention as well as psychiatric well-being when lockdowns are prolonged.

Loading [MathJax]/jax/output/CommonHTML/jax.js

However, there are many challenges for evaluating suicide risk in a large general population. In a pandemic situation, it is too labour intensive and clinician dependent to conduct structured interviews or scales for SI[15] to assess present and past mental health in an entire population. Moreover, there is a possibility of missing cases during screening with simple questionnaires in general population studies because most studies further evaluate cases only when they respond that they have SI, which could mask true patients at risk for suicide[16]. Existing prediction models[5,17–19] for suicidal behaviour achieved an accuracy over 80%, but, at the same time, these models had a very low sensitivity, which is due to low incidence of SAs in the general population. Specifically, the model sensitivity is more important than the specificity for monitoring suicide risk; individuals who attempt suicide are missed due to the low sensitivity of the monitoring system, which results in irreversible events, is more critical than missing individuals will not attempt suicide due a low specificity. Although knowing who is going to commit suicide is critical, it is difficult to predict who will attempt suicide with a prospective study. Active screening for COVID-19 risk[20], which extends the scope of the risk group, could reduce the likelihood of missing an infected person compared to traditional approaches that target people with obvious symptoms. Expanding the range of suicide risk to SI instead of SAs could increase the sensitivity of the prediction model and may contribute to reducing the number of individuals who may attempt suicide.

In many countries, including South Korea, a large population of young adults is obliged to have regular check-ups, including mental status examinations, for work or when entering a dormitory for college, which leads to the only portal to access individuals who may attempt suicide. We employed the multiple scales included in regular mental status examinations to predict imminent suicide risk. We used acute SI within 2 weeks as a surrogate marker for prospective imminent suicide risk[21]. To extract a good representation of acute SI from scales, multi-dimensional questionnaires evaluating depression, anxiety, resilience, and self-esteem levels were used as input features of the neural network. To overcome these challenges, we present a novel graph neural network (GNN)-based model that employs multi-dimensional scale-based prediction of depression and acute SI with a high sensitivity and specificity, which could promote the deployment of the model in the real world.

Results

Subject clinico-demographics

Across four centres, including university and secondary/tertiary hospitals, 31,720 (mean age, 23.64 ± 3.96 years old; 68.2% male) out of 32,250 participants responded to six self-report questionnaires: the Patient Health Questionnaire-9 (PHQ-9)[22], Generalized Anxiety Disorder-7 (GAD-7)[23], State-Trait Anxiety Inventory-State Anxiety (STAI-S, or STAI-X1)[24,25], the Resilience Appraisal Scale (RAS)[26], the Rosenberg Self-Esteem Scale (RSES)[27], and lifetime SA. The total number of positive acute SI cases was 306/31,720 cases (0.965%) across the 4 different institutions. The rate of acute SI differed between the training/validation and test sets (0.74% vs. 1.24%; $p<0.0001$). There was a difference in the ages of the individuals in the training/validation and test sets (mean age, 23.17 ± 4.17 vs. 24.23 ± 3.59 years old; $p < 0.0001$). The gender ratio was different between the training/validation and test sets (79.95% vs. 53.82% male; $p <$

Loading [MathJax]/jax/output/CommonHTML/jax.js

0.0001). For all five scales (i.e., the PHQ-9, GAD-7, STAI-S, RAS, and RSES), the distributions of the scores were different across centres (p<0.0001) as well as between the training/validation and test sets (STAI-S, p=0.01; all others, p < 0.0001). The clinico-demographic information is summarized in Supplementary Table 1. In brief (Supplementary Table 1), the incidence of lifetime SI was approximately 10 times higher than the incidence of acute SI. The incidence of lifetime SI, acute SI, and lifetime SAs in the total dataset were 2,641 (8.33%), 306 (0.97%), and 437 (1.38%) out of 31,720 participants, respectively. In total, 358 people received structured interviews using the Mini International Neuropsychiatric Interview (MINI)[28], of which 102 participants were diagnosed with major depressive episodes (MaDEs), accounting for 0.32% of all participants.

Prediction of MaDEs: external validation

For the test set (centre 4; Seoul National University (SNU)) with true MaDE labels (n = 64), the MaDE prediction model achieved a sensitivity, specificity, accuracy, and area under the receiver operating characteristic (ROC) curve (AUC) of 90.91%, 82.76%, 84.06%, and 0.934 (95% confidence interval (CI), 0.874–0.986), respectively, when using logistic regression with least absolute shrinkage and selection operator (LASSO) regularization and 90.90%, 67.24%, 71.01%, and 0.937 (95% CI, 0.881–0.987) when using a support vector machine (SVM), whereas the graph isomorphism network (GIN)-MaDE model achieved values of 96.55%, 95.00%, 95.65%, and 0.996 (95% CI, 0.988-1.000), respectively (Table 1).

Loading [MathJax]/jax/output/CommonHTML/jax.js

Table 1
Model performance of prediction of major depressive episodes (MaDE) and acute suicidal ideation: model comparison with external validation

Prediction	Author/ Data Source	No. of patients (positive/total, %)	Model	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
MaDE	Current study	102[a]/31,720 (0.32%)	Logistic regression with LASSO	*90.91*	82.76	84.06	0.934 (0.874–0.986)
			SVM	*90.90*	67.24	71.01	0.937 (0.881–0.987)
			GIN-MaDE	**96.55**	**95.00**	**95.65**	**0.996 (0.988–1.000)**
Acute SI	Jung et al[33]	7,443/59,984 (12.4%)	Logistic regression	*78.2*	77.6	77.9	0.851
			SVM	*78.4*	78.9	78.7	0.853
	Current study	306/31,720 (0.97%)	Logistic regression	*49.71*	98.41	97.78	0.740 (0.711–0.771)
			SVM	*15.03*	99.81	98.72	0.574 (0.552–0.597)
			GIN_u1	70.52	**85.44**	**85.25**	0.869 (0.842–0.891)
			GIN_u2	79.77	79.43	79.44	0.861 (0.835–0.886)
			GIN_ SMOTE	79.19	77.33	77.35	0.851 (0.826–0.874)
			Ensemble of GINs (*Mind WatchNet*)	**80.92**	80.57	80.57	**0.877 (0.854–0.897)**

Loading [MathJax]/jax/output/CommonHTML/jax.js

Note: Bold numbers indicate the best metrics among graph neural network models. Italic numbers indicate lower sensitivity of logistic regression, and SVM models. [a]A total of 358 people received structured interviews, of which 102 participants were diagnosed with MaDE, accounting for 0.32% of the total 31,720 participants. The results from a state-of-the-art study by Jung et al[33] were cited, where the confidence intervals of AUC were not reported.

Abbreviations: MaDE, major depressive episode; SI, suicidal ideation; AUC, area under the curve; LASSO, least absolute shrinkage and selection operator; SVM, support vector machine; GIN, graph isomorphism network.

Prediction of acute SI: external validation

Using conventional algorithms, the model achieved a sensitivity, specificity, accuracy, and AUC of 49.71%, 98.41%, 97.78%, and 0.740 (95% CI, 0.711–0.771), respectively, when using logistic regression with LASSO regularization and 15.03%, 99.81%, 98.72%, and 0.574 (95% CI, 0.552–0.597) when using an SVM. In contrast, the sensitivity, specificity, accuracy, and AUC of the ensemble of three GIN models, or *MindWatchNet*, were 80.9%, 80.6%, 80.6%, and 0.877 (95% CI, 0.854–0.897), respectively. For each model in the ensemble, the sensitivity, specificity, accuracy, and AUCs were as follows: 79.2%, 77.3%, 77.4% and 0.851 (95% CI, 0.826–0.874) for the GIN-synthetic minority over-sampling technique (SMOTE) model, respectively; 70.5%, 85.4%, 85.3% and 0.869 (95% CI, 0.842–0.891) for GIN-u1; and 79.8%, 79.4%, 79.4% and 0.851 (95% CI, 0.826–0.874) for GIN-u2 (Table 1).

Attention plots and interpretation

The raw averaged attention plots without normalization are given in Fig. 2 for the test set (Fig. 2a). In the attention plot comparing questionnaire items by using row-wise normalization (Fig. 2b), a high score (i.e., 4 points) for item 2 of the PHQ-9 (i.e., PHQ_2 – feeling down, depressed, or hopeless) was the most salient positive feature (i.e., a feature that increases the prediction score of acute SI) among 19 items of the questionnaires. The 2nd most salient positive feature was a high total score for the STAI-S (i.e., the 4th quartile group), which represents a high level of anxiety. For low scores (i.e., 1 point), the most salient negative feature (i.e., a feature that decreases the prediction score of acute SI) was a low PHQ_2 score, which means that a low PHQ_2 score is the most significant feature associated with reduced SI the most. For intermediate scores (i.e., 2–3 points), the most salient positive features were PHQ_8, PHQ_6, and PHQ_9 (i.e., psychomotor symptoms, low self-esteem, and suicidal thoughts, respectively) to a similar degree. Moreover, for point 1, the RAS total score was the most salient negative feature for SI, which means that the 3rd quartile group (because of reversed order) of RAS total points had decreased SI.

In the attention plot comparing binary items by using column-wise normalization (Fig. 2c), attention values were highest for lifetime SA (odds ratio (OR), 8.51), presence of MaDE (OR, 3.46), type of institution (OR, 2.55), and female sex showed slightly increased attention values (OR, 1.15).

Loading [MathJax]/jax/output/CommonHTML/jax.js

In the plot using the L1-norm of the attention vector, obtained for each column of the 19 questionnaire items (Fig. 2d), the attention values were the highest for PHQ_2 (1.0), and STAI_X1_total had the 2nd highest attention values (0.825); and PHQ_5 (poor appetite or overeating) and PHQ_6 had the 3rd and 4th highest attention values (0.797, and 0.777), respectively. Moreover, the attention plots for the training/validation set showed nearly identical results to those for the test set (Supplementary Figs. 3a-d).

Ablation study for PHQ-9 item 9

Because PHQ_9 is related to acute SI, the model performance with and without PHQ_9 were obtained for comparison: the sensitivity, specificity, accuracy, and AUC were 79.77%, 77.85%, 77.87%, and 0.869 (95% CI, 0.846–0.891), respectively, for the model without PHQ_9. There was no significant difference in the AUCs between models with and without PHQ_9 (AUC = 0.877 vs. 0.869; $p = 0.150$) (Supplementary Fig. 2).

Validity of the labels for acute SI: comparison study

Among the subjects in the test set, only $n = 792$ of 13,408 subjects completed the Korea Advanced Institute of Science and Technology (KAIST) Scale for Suicide Ideation (KSSI). Spearman's rank correlation coefficient 1) between the predicted scores by *MindWatchNet* and the total KSSI score was ρ_{pred}=0.719 ($p < 0.0001$) and 2) between PHQ_9 and the total KSSI score was ρ_{PHQ}=0.446 ($p < 0.0001$). In the comparison of the correlation coefficients,ρ_{pred} was larger than ρ_{PHQ} ($p < 0.0001$). A scatter plot between the raw predicted scores (i.e., the model output *before* applying the sigmoid function) by *MindWatchNet* and the total KSSI score is shown in Fig. 3.

Discussion

We developed a GNN model to predict acute SI within 2 weeks, which showed improved sensitivity compared to baseline models, and validated it in an external test set: the sensitivity, specificity, accuracy and AUC were 80.9%, 80.6%, 80.6%, and 0.877 (95% CI, 0.854–0.897), respectively, using an ensemble of GIN models with different sampling methods, or *MindWatchNet*; these values were 15.03%, 99.81%, 98.72%, and 0.574 (95% CI, 0.552–0.597), respectively, using an SVM. Specifically, *MindWatchNet*, based on a GIN to predict SI, improved the sensitivity significantly at the cost of slight reductions in the specificity and accuracy. The low sensitivities of the baseline models prevented the prediction of individuals who may attempt suicide before committing suicide, resulting in irreversible events[5]. In contrast, *MindWatchNet* achieves a significant increase in the sensitivity compared to previous baseline models, allowing more accurate prediction of individuals who may attempt suicide, suggesting that this model can potentially be of great help in the real world.

Our model achieved a good performance by incorporating the following three factors. 1) The GNN extracts a good graph embedding. The GIN[29], a variant of a spatial GNN specifically for graph classification, extracts an even better representation from the graph than other GNNs such as graph convolutional networks (GCNs) because GINs are equivalent to generalized convolutional neural networks (CNNs) for non-Euclidean data that can be represented as graph structures, such as brain connectivity[30,31]. 2) An ensemble method using under-sampling and over-sampling (i.e., SMOTE for nominal and continuous features (SMOTE-NC)[32]) was

Loading [MathJax]/jax/output/CommonHTML/jax.js

designed to handle class imbalance issues. 3) Rich information from multi-dimensional scales and subject clinico-demographic information for large multi-centre datasets were used; 7 questionnaires covering domains such as depression, anxiety, resilience, and self-esteem, which are obtained from $n = 31,720$ individuals across 4 centres including universities and hospitals. Jung et al[33] reported that the baseline models showed good performance in predicting SI in the past 12 months in a young population, with approximately 13 positive cases compared to the current data (12.4% vs. 0.97%, see Table 1). However, it is challenging to predict acute SI within 2 weeks. In the present study, having severe class imbalance, the SVM without the ensemble method, which is a baseline model, could not extract a good representation of the positive cases, resulting in a much lower sensitivity (~ 15%) than *MindWatchNet*, while a specificity and accuracy of nearly 99% were achieved. This finding suggests that dealing with class imbalance, such as with the ensemble method, should be considered to prevent prediction bias towards the majority class (i.e., the model always predicts SI-negative). It probably does not matter what kind of model is used, but this analysis is beyond the scope of the current study. Interestingly, our model can show not only feature importance but also the association among features. Although PHQ_2 and STAI-S are features having the highest saliency value, the former was associated with other items of the PHQ-9 and the latter was associated with resilience and self-esteem (Fig. 2d).

We predicted MaDEs as a pseudo-label before the prediction of acute SI because pre-existing psychiatric disorders such as major depressive disorder (MDD) have been known to increase suicide risk[34], which would be helpful in accurately predicting acute SI. In MaDE prediction, all the conventional and GIN models achieved AUCs and sensitivities over 90%. This finding suggests that both the PHQ-9 and other scales, including the GAD-7, contributed to predicting the MaDE labels. MaDE pseudo-labels were used as input to predict acute SI. Although the presence of a MaDE is 3.46 times more likely to indicate an individual with SI than its absence (Fig. 2c), its low saliency may be indirectly associated with SI via its association with various PHQ-9 items and GAD_7 (“Feeling afraid, as if something awful might happen”) (Figs. 2d and Supplementary Figs. 3d). Interestingly, lifetime SA achieve both the highest OR among the binary items (Fig. 2c) and a higher saliency score than MaDE. In addition, MaDE can be accurately predicted with conventional or GIN models. The results suggest that both gathering SA information and predicting MaDE with a model, instead of structural interviews for diagnosis, is an efficient approach for survey-based screening for suicide risk. Moreover, nearly identical attention plots for the training/validation set (Supplementary Fig. 3) and test set (Fig. 2) might suggest that the common “scale and clinico-demographic signature” of acute SI was extracted by using the GIN, which models the relationship between the scale items and clinico-demographic information in graph-structured data.

In the attention plots, the model recognized the salient items among the multi-dimensional questionnaires and other information (Fig. 2). Specifically, when comparing 19 questionnaire items, several PHQ-9 items (e.g., items 2, 4, 5, 6, and 8) and the total RAS and STAI-S scores showed high saliency values. Among these features with high saliency, depressed mood (PHQ_2) and high state and trait anxiety (STAI-S total score) were the two most salient features. The first 2 items of the PHQ-9 provide the two cardinal symptoms of depression, i.e., PHQ_1 (anhedonia) and PHQ_2 (depressed mood or hopelessness)[35]. Depressed mood (PHQ_2) mediates negative life events associated with SI[22], and its severity is also strongly associated with

Loading [MathJax]/jax/output/CommonHTML/jax.js

SI[36]. In the network analysis of depressive symptoms, hopelessness (PHQ_2) was the most central criterion or central node (specifically, the highest betweenness centrality) in the symptom network, showing a strong connection between PHQ_2 and PHQ_9 (suicide), as well as PHQ_2 and PHQ_6 (worthlessness), and a moderate connection between PHQ_2 and PHQ_1 (anhedonia)[37], which are all salient features for SI in the attention plot (Figs. 2b and Supplementary Figs. 3b). In another network analysis of anxiety and depressive symptoms, the same symptom network was revealed, which represents the connections between PHQ_2 and PHQ_1; PHQ_2 and PHQ_6; and PHQ_2 and PHQ_9, making PHQ_2 a central node[38]. In addition, psychomotor symptoms (PHQ_8, i.e., "moving or speaking so slowly that other people could have noticed or the opposite") was another salient feature for acute SI (Fig. 2b). In a large population-based longitudinal study, anxiety disorders were found to be independent risk factors for suicidal behaviours (i.e., SI and SA), and an increased risk of SA in combination with a mood disorder was found[2]. In our results, a high STAI-S total score was associated with increased acute SI, which is consistent with previous studies showing that both state and trait anxiety increase the risk of suicide risk[39,40]. It has been reported that resilience protects against symptoms of anxiety and depression and strongly influences the associations between symptoms and lifestyle factors[41], which is also consistent with the findings that low resilience is strongly associated with mild depression and psychological resilience is linked to social support[42], and might lead to an increased risk of SI compared to non-depressed subjects. Moreover, low resilience was a risk factor for suicidal behaviours[43]. In our study, a high RAS total score was associated with decreased SI, and vice versa, which is also consistent with a previous study showing that high resilience is one of the most protective features for SAs[26,44].

In the ablation study of the PHQ_9, it was related to acute SI, the model performance without PHQ_9 showed no significant difference in term of the AUC compared to the model with this item (AUC = 0.869 vs. 0.877, respectively; p = 0.150), which guarantees that the model did not "cheat" to predict acute SI using only PHQ_9. In the validation study of the true labels for acute SI, the model prediction score showed a higher correlation with the KSSI score (i.e., it is a more accurate proxy for acute SI) than PHQ_9 (ρ=0.719 vs. 0.664, respectively, p = 0.005; see the Validity of the labels for acute SI section in the Results section). Originally, the PHQ-9 was designed for screening depression and to assess severity, not to assess suicide risk[22]. Interestingly, in a recent validation study, Na et al.[45] showed that PHQ_9 is an insufficient assessment tool for suicide risk and SI because of the limited utility in certain clinico-demographic and clinical subgroups, which is in line with our results. Our results indicate that our model-based predictions resulting from multi-dimensional information are more valid than those from only a single question (i.e., PHQ_9 and acute SI label) and is an alternative to a structured interview or a scale for suicide risk. While PHQ_9 itself may not be a valid measure for SI, our results (Fig. 2b) suggest that intermediate scores (i.e., 2–3 points) for this item should not be overlooked. This strategy should also apply to the PHQ_6 (worthlessness) and PHQ_8 (psychomotor symptoms) (Fig. 2b).

It is worth noting that this multi-dimensional scale dataset was collected before the outbreak of COVID-19, and the specific representation of mental illness, including depression and anxiety, evoked by consequences of the COVID-19 pandemic may not be reflected by the scales used in the present study. Further research is

Loading [MathJax]/jax/output/CommonHTML/jax.js

WatchNet during the COVID-19 pandemic. In addition, the true

labels for acute SI may be improved if we obtain the labels for suicidal behaviour from reference to standards, such as structured interviews by clinicians for all subjects; however, this process is time consuming, impractical, and requires large amounts of research funding.

This study has several limitations. As prediction of major depressive episode using small dataset can lead to overfitting, the benefit of the pseudo-label[46] of MaDE to predict SI should be confirmed in future studies. The type of institution cannot be generalized to other types of data obtained from workplaces. Although there was relatively low saliency of the type of institution (Fig. 2c) compared to lifetime SA or MaDE, its value for each individual might not be meaningful and must be interpreted carefully. Although beyond the scope of the current study, exploring the impact of edge and sparsity definitions on performance is necessary. To generalize the results of young adults to other populations, further studies of a wide range of ages are needed. Longitudinal cohort studies are needed to investigate factors that can predict future SAs or new SI cases. Verification studies are needed to determine whether predicting SI instead of SAs is effective in preventing SAs in the real world.

In conclusion, we developed and validated a deep-learning-based compensatory tool by using extracted deep features from multi-dimensional self-report questionnaires covering depression, anxiety, resilience, self-esteem, and clinico-demographic information of a large dataset to instantaneously predict suicide risk and monitor responses to suicide prevention strategies, which might be useful in remote clinical practice in the general population of young adults for specific situations such as the COVID-19 pandemic.

Methods

Dataset

Across the four centres in the Research Consortium for Young Adulthood Depression, young adults between the ages of 18 and 34 years old who underwent medical examinations, including mental health measurements and were enrolled in the study between Jan 1, 2018, and Dec 31, 2019 were included. The research protocol for the present study was approved by Korea Advanced Institute of Science and Technology (KAIST) Institutional Review Boards. The study protocol was performed in accordance with the relevant guidelines. Informed consent was obtained from all the participants. Anxiety disorders are known as independent risk factors for suicidal behaviour (i.e., SI and SAs) and increase the risk of SA when combined with mood disorders such as MDD[2]. A history of SAs is considered a crucial predictor of future suicidal behaviour[47,48]. Although knowing the presence of MDD should be very helpful[49] to improve the diagnostic performance of the prediction model for acute SI, structured interviews given by psychiatrists in large populations are less cost effective. Here, the MINI[28] was performed on a portion of the participants by four psychiatrists (S.H.K., S.H.Y., D.H.K. and M.S.K.) in centres 1–3 and via a web-based version in centre 4. Then, using data with labels for MaDE, a prediction model (GIN-MaDE) was trained. Finally, the MaDE pseudo-labels were predicted by the trained GIN-MaDE network for participants without MaDE labels (detailed in the *Semi-supervised learning-based input features: pseudo-labels for the MaDE* section) and were used to predict acute SI. The presence of acute SI was determined when the participant responded “yes” to the

Loading [MathJax]/jax/output/CommonHTML/jax.js

n the past 2 weeks?”. To develop the model, self-report

questionnaires and other clinical data from three independent institutions were obtained: centre 1 was KAIST (n = 17,322); centre 2 was Gachon University Hospital (Gachon) (n = 69); and centre 3 was Samsung Medical Center (SMC) in Seoul (n = 91). For external validation, questionnaires and other clinical data obtained from centre 4 (Seoul National University (SNU), n = 14,238) were used. All the data were anonymized prior to combining the data from the four institutions. All the descriptions of the self-report questionnaires are detailed in the Supplementary material. The overall workflow for constructing the graph-structured dataset is illustrated in Fig. 1a.

GIN as a graph neural network

A GIN, which is a variant of a GNN with equal representative/discriminative power for graph-structured data, such as the Weisfeiler-Lehman (WL) test – one of the most powerful existing tests for distinguishing a broad class of graphs[50], was developed for graph classification and has achieve state-of-the-art performance[29]. More specifically, for each node, *v*, graph convolution aggregates neighbouring – or nodes connected by weighted edges – node features, $\sum_{u\in \mathrm{N}(v)} p_u^{(k-1)}$ (see Eqs. 1 and 2 in Supplementary material) and combines the aggregation with the node feature of the previous hidden layer, $p_v^{(k-1)}$, to update the node feature at the current *k*th hidden layer, $p_v^{(k)}$. Next, for each node, multi-layer perceptron (MLP) layers elevate the node feature to a high-dimensional latent space (i.e., from the dimension of the node features of the hidden layer to the dimension of the MLP layers; $\mathrm{R}^{C^{(k-1)}} \rightarrow \mathrm{R}^{C^{(k)}}$, $\mathrm{where} C^{(k)}$ denotes the dimension of the node features of the *k*th hidden layer). For each hidden graph convolutional layer, all the updated node features were summed to make a graph feature of the *k*th hidden layer, $p_G^{(k)}$, which is known as sum-pooling. For the graph-level readout, all *K* graph features from the hidden layers were concatenated to make a final graph feature, p_G (Fig. 1), extracting an excellent graph representation[29] for positive and negative cases of acute SI. Finally, p_G is fed to the final classifier to calculate the sigmoid prediction score of acute SI. The overall model architecture is illustrated in Fig. 1, and the mathematical equations are described in the Supplementary material.

Semi-supervised learning-based input features: pseudo-labels for MaDE

MaDE labels are important information to predict acute SI; however, only a fraction of MaDE labels were available because only a fraction of subjects, 294 individuals in the training/validation set and 64 individuals in the test set, completed the MINI. Following the pseudo-labelling strategy frequently used in semi-supervised learning[46], we generated pseudo-labels for MaDE via other questionnaires and clinico-demographic information, such as gender and type of institution, using the GIN-MaDE network prior to training the GINs for predicting acute SI. Details are described in the Supplementary method section.

Prediction of acute SI: subsampling strategy

To overcome the intrinsic challenge of SI prediction or the sparsity of positive cases of acute SI (i.e., the class imbalance problem), we utilized not only data augmentation for balancing the data but also ensembles of models with different subsamplings. First, a GIN model was developed to predict acute SI

Loading [MathJax]/jax/output/CommonHTML/jax.js

using the MaDE pseudo-labels as an additional input feature. Most machine learning models built on imbalanced datasets give predictions that are biased towards the majority class (i.e., negative cases); hence, the model will always predict a case as a negative case even if it is a positive case. Specifically, to obtain different decision boundaries to be ensembled, which may largely depend on the subsampled data distribution, we built three different GIN models with different subsampling strategies: 1) GIN-u1 (under-sampling of the majority class with a balance ratio of 10); 2) GIN-u2 (under-sampling of the majority class with a balance ratio of 5); and 3) GIN-SMOTE[32] (over-sampling of the minority class with a balance ratio of 1), where the majority and minority classes has negative and positive SI labels, respectively, and the balance ratio is defined as the ratio of negative to positive cases in the subsampled data from the training set.

For the training and validation sets, datasets from centres 1–3 (SMC, Gachon, and KAIST) were used, and a dataset from centre 4 (SNU) was used for the test set. Note that the test set was never augmented.

Ensemble model

After training each of the three GIN models defined above, the best model for each subsampling strategy was saved at the epoch when the validation loss was minimized: GIN-u1-best, GIN-u2-best, and GIN-SMOTE-best. Next, the final ensemble GIN model was obtained using the three best models. Specifically, the sigmoid prediction scores from the best models were averaged to obtain the final prediction score of the ensemble model, which is a process known as “soft voting”[51,52].

Evaluation

For the prediction of MaDE and acute SI, the sensitivity, specificity, and accuracy were calculated for all the models. To evaluate the diagnostic performance of the models, a ROC analysis was performed to obtain the AUC, and DeLong’s method was used to compare the AUCs. For the comparison with conventional algorithms, logistic regression with LASSO and an SVM (detailed in the Supplementary material) were used for the prediction of MaDEs and acute SI. All statistical analyses were performed using R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria).

Ablation study for PHQ_9

Because PHQ_9 (“Thoughts that you would be better off dead or of hurting yourself in some way”) is related to acute SI, including PHQ_9 as a predictor could be redundant. Moreover, response to PHQ_9 has been reported to be a moderate predictor of a subsequent SA or death[53]. However, in studies for the validation of PHQ_9 using the Structured Clinical Interview for DSM Disorders (SCID) assessment as the reference standard, it had a good sensitivity, specificity, and negative predictive value but a low positive predictive value (PPV) in irritable bowel disease (20.8%)[54] and neurological disorders such as epilepsy (39.1%), migraine (54.5%), multiple sclerosis (41.7%), and stroke (57.1%)[55]. Here, to test the benefit of the inclusion of PHQ_9, the performance of the model without PHQ_9 was also assessed and compared with that of the model including PHQ_9. Specifically, the ROC comparison of the best GIN-based model with and without PHQ_9 was performed using DeLong’s method. The saliency plots were also compared with and without PHQ_9 using the best GIN-based model.

Validity of the labels for acute SI: comparison study

Loading [MathJax]/jax/output/CommonHTML/jax.js

Self-report instruments for the assessment of suicidal thinking, such as the Beck Scale for Suicidal Ideation, could be a reliable quantitative reference for acute SI[56–58]. The KSSI[13] is a comprehensive scale to evaluate suicide risk. The KSSI score for the previous 2 weeks was significantly correlated with the Beck Scale for Suicidal Ideation score (Kendall's $\tau = 0.35$, $p < 0.001$) in our previous study[13]. To investigate the reliability of the model prediction score, we also compared Spearman's correlation coefficients between the KSSI total score and the prediction score or PHQ_9.

Attention plots and interpretation

To interpret what the ensemble model "thinks" is important for the prediction of acute SI, we calculated the saliency/attention values, which are defined as the gradient of the input with respect to the model output, $\frac{\partial y}{\partial x_i}$, where *y* is the linear output of the prediction model and x_i is the *i*th input node feature for $i = 0, 1, \cdots, N$ (*N*=the number of nodes in the graph), i.e., how much the output changes when we change the input values. The set of attention plots was obtained for both the test set (Figs. 2a-d) and the training/validation set (Supplementary Figs. 3a-d).

Declarations

Author Contribution

KSC, SHK, BHK, and BJ designed the study Design. SHK, H-JJ, JHK, and JHJ collected data. KSC, SHK, BHK, H-JJ, JHK, JHJ, and BJ analyzed data. KSC, SHK, JHJ, and BJ wrote the manuscript. All authors contributed to data interpretation and critically reviewed the final manuscript.

Data availability

Due to potentially identifying information, the data that support the findings of this study are not publicly available, but can be obtained under the condition both on reasonable request to corresponding authors and the permission of Institutional Review Board.

Code availability

Codes will be uploaded soon on https://github.com/kyuchoi/YAD_survey.

Acknowledgements

We thank and express our gratitude to members of Research Consortium for Young Adulthood Depression (https://www.kaistbrain.org/) for collecting data and specially thank to Haeorm Park, Ph.D. candidate, Minchul Kim, M.D., Seokho Yoon, M.D., Minseob Kim, M.D., and Geumsook Shim, M.D., Ph.D. for providing structured interviews for MINI and managing data.

Ethics declarations

The authors disclose no conflicts of interest related to this work.

Loading [MathJax]/jax/output/CommonHTML/jax.js

References

1. Xu, J., Murphy, S., Kochanek, K. & Arias, E. Mortality in the United States, 2018. NCHS Data Brief, no 355. National Center for Health Statistics, Hyattsville, MD(2020).
2. Sareen, J. *et al.* Anxiety disorders and risk for suicidal ideation and suicide attempts: a population-based longitudinal study of adults. *Arch Gen Psychiatry.* **62**, 1249–1257 (2005).
3. Organization, W. H. Preventing suicide: A global imperative(World Health Organization, 2014).
4. Bostwick, J. M., Pabbati, C., Geske, J. R. & McKean, A. J. Suicide Attempt as a Risk Factor for Completed Suicide: Even More Lethal Than We Knew. *American Journal of Psychiatry.* **173**, 1094–1100 (2016).
5. Zheng, L. *et al.* Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records. *Translational Psychiatry.* **10**, 1–10 (2020).
6. Flaxman, S. *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature.* **584**, 257–261 (2020).
7. Reger, M. A., Stanley, I. H. & Joiner, T. E. Suicide Mortality and Coronavirus Disease 2019-A Perfect Storm? *JAMA Psychiatry.* **77**, 1093–1094 (2020).
8. Wasserman, I. M. The impact of epidemic, war, prohibition and media on suicide: United States, 1910–1920. *Suicide Life Threat Behav.* **22**, 240–254 (1992).
9. Tanaka, T. & Okamoto, S. Increase in suicide following an initial decline during the COVID-19 pandemic in Japan.*Nat Hum Behav*,1–10(2021).
10. National Center for Health Statistics, C.f.D.C.a.P. Anxiety and Depression. Household Pulse Survey. Vol. 2021 (2020).
11. National Center for Health Statistics, C.f.D.C.a.P. Early release of selected mental health estimates based on data from the January–June 2019 national health interview survey. in *National Center for Health Statistics*, Vol. 2021 (2020).
12. Hao, F. *et al.* Do psychiatric patients experience more psychiatric symptoms during COVID-19 pandemic and lockdown? A case-control study with service and research implications for immunopsychiatry. *Brain, behavior, and immunity.* **87**, 100–106 (2020).
13. Shim, G. & Jeong, B. Predicting Suicidal Ideation in College Students with Mental Health Screening Questionnaires. *Psychiatry Investig.* **15**, 1037–1045 (2018).
14. Czeisler, M. E. *et al.* Mental Health, Substance Use, and Suicidal Ideation During the COVID-19 Pandemic - United States, June 24–30, 2020. *Mmwr-Morbidity and Mortality Weekly Report.* **69**, 1049–1057 (2020).
15. Beck, A. T., Steer, R. A. & Ranieri, W. F. Scale for Suicide Ideation: psychometric properties of a self-report version. *J Clin Psychol.* **44**, 499–505 (1988).
16. Kliem, S., Lohmann, A., Mossle, T. & Brahler, E. German Beck Scale for Suicide Ideation (BSS): psychometric properties from a representative population survey. *BMC Psychiatry.* **17**, 389 (2017).
17. Barak-Corren, Y. *et al.* Predicting Suicidal Behavior From Longitudinal Electronic Health Records. *Am J Psychiatry.* **174**, 154–162 (2017).

Loading [MathJax]/jax/output/CommonHTML/jax.js

18. Belsher, B. E. *et al.* Prediction Models for Suicide Attempts and Deaths: A Systematic Review and Simulation. *JAMA Psychiatry.* **76**, 642–651 (2019).
19. Oh, J., Yun, K., Hwang, J. H. & Chae, J. H. Classification of Suicide Attempts through a Machine Learning Algorithm Based on Multiple Systemic Psychiatric Scales. *Frontiers in Psychiatry.* **8**, 192 (2017).
20. Lee, D. & Lee, J. Testing on the move: South Korea's rapid response to the COVID-19 pandemic. *Transportation Research Interdisciplinary Perspectives.* **5**, 100111 (2020).
21. Rogers, M. L., Ringer, F. B. & Joiner, T. E. The association between suicidal ideation and lifetime suicide attempts is strongest at low levels of depression. *Psychiatry Res.* **270**, 324–328 (2018).
22. Kroenke, K., Spitzer, R. L. & Williams, J. B. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* **16**, 606–613 (2001).
23. Spitzer, R. L., Kroenke, K., Williams, J. B. & Lowe, B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med.* **166**, 1092–1097 (2006).
24. Hahn, D. W. Korean adaptation of Spielberger's STAI (K-STAI). *Kor J Health Psychol.* **1**, 1–14 (1996).
25. Spielberger, C. State-trait anxiety inventory. The Corsini encyclopedia of psychology.*Hoboken: Wiley***1**(2010).
26. Johnson, J., Gooding, P. A., Wood, A. M. & Tarrier, N. Resilience as positive coping appraisals: Testing the schematic appraisals model of suicide (SAMS). *Behaviour Research and Therapy.* **48**, 179–186 (2010).
27. Rosenberg, M. Rosenberg self-esteem scale (RSE). *Acceptance and commitment therapy. Measures package.* **61**, 18 (1965).
28. Yoo, S. W. *et al.* Validity of Korean version of the mini-international neuropsychiatric interview. *Anxiety and Mood.* **2**, 50–55 (2006).
29. Xu, K., Hu, W., Leskovec, J. & Jegelka, S. How powerful are graph neural networks? arXiv preprint arXiv:1810.00826(2018).
30. Kim, B. H. & Ye, J. C. Understanding Graph Isomorphism Network for rs-fMRI Functional Connectivity Analysis. *Front Neurosci.* **14**, 630 (2020).
31. Yang, Z. K., Chen, C. S., Li, H. W., Yao, L. & Zhao, X. J. Unsupervised Classifications of Depression Levels Based on Machine Learning Algorithms Perform Well as Compared to Traditional Norm-Based Classifications.*Frontiers in Psychiatry***11**(2020).
32. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research.* **16**, 321–357 (2002).
33. Jung, J. S. *et al.* Prediction models for high risk of suicide in Korean adolescents using machine learning techniques. *PLoS one.* **14**, e0217639 (2019).
34. Isometsä, E. Suicidal behaviour in mood disorders—who, when, and why? *The Canadian Journal of Psychiatry.* **59**, 120–130 (2014).
35. Lowe, B., Kroenke, K. & Grafe, K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *J Psychosom Res.* **58**, 163–171 (2005).
36. Garlow, S. J. *et al.* Depression, desperation, and suicidal ideation in college students: Results from the American Foundation for Suicide Prevention College Screening Project at Emory University. *Depression*

Loading [MathJax]/jax/output/CommonHTML/jax.js

and Anxiety. **25**, 482–488 (2008).

37. Kendler, K. S., Aggen, S. H., Flint, J., Borsboom, D. & Fried, E. I. The centrality of DSM and non-DSM depressive symptoms in Han Chinese women with major depression. *J Affect Disord.* **227**, 739–744 (2018).
38. Beard, C. *et al.* Network analysis of depression and anxiety symptom relationships in a psychiatric sample. *Psychol Med.* **46**, 3359–3369 (2016).
39. Choi, H. Y. *et al.* A Study on Correlation between Anxiety Symptoms and Suicidal Ideation. *Psychiatry Investigation.* **8**, 320–326 (2011).
40. Ohring, R. *et al.* State and trait anxiety in adolescent suicide attempters. *J Am Acad Child Adolesc Psychiatry.* **35**, 154–157 (1996).
41. Skrove, M., Romundstad, P. & Indredavik, M. S. Resilience, lifestyle and symptoms of anxiety and depression in adolescence: the Young-HUNT study. *Soc Psychiatry Psychiatr Epidemiol.* **48**, 407–416 (2013).
42. Choi, Y. *et al.* The relationship between levels of self-esteem and the development of depression in young adults with mild depressive symptoms. *Medicine.* **98**, e17518 (2019).
43. Roy, A., Sarchiapone, M. & Carli, V. Low resilience in suicide attempters: Relationship to depressive symptoms. *Depression and Anxiety.* **24**, 273–274 (2007).
44. Kim, S. M. *et al.* Resilience as a Protective Factor for Suicidal Ideation among Korean Workers. *Psychiatry Investigation.* **17**, 147–156 (2020).
45. Na, P. J. *et al.* The PHQ-9 Item 9 based screening for suicide risk: a validation study of the Patient Health Questionnaire (PHQ)-9 Item 9 with the Columbia Suicide Severity Rating Scale (C-SSRS). *J Affect Disord.* **232**, 34–40 (2018).
46. Lee, D. H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. in Workshop on challenges in representation learning, Vol. 3(ICML, 2013).
47. Mann, J. J. A current perspective of suicide and attempted suicide. *Ann Intern Med.* **136**, 302–311 (2002).
48. Park, E. H., Hong, N., Jon, D. I., Hong, H. J. & Jung, M. H. Past suicidal ideation as an independent risk factor for suicide behaviours in patients with depression. *International Journal of Psychiatry in Clinical Practice.* **21**, 24–28 (2017).
49. Gili, M. *et al.* Mental disorders as risk factors for suicidal behavior in young people: A meta-analysis and systematic review of longitudinal studies. *Journal of Affective Disorders.* **245**, 152–162 (2019).
50. Babai, L. Graph isomorphism in quasipolynomial time. in Proceedings of the forty-eighth annual ACM symposium on Theory of Computing 684–697(2016).
51. Sharma, A. & Verbeke, W. J. Improving Diagnosis of Depression with XGBOOST Machine Learning Model and a Large Biomarkers Dutch Dataset (n = 11,081). *Frontiers in Big Data.* **3**, 15 (2020).
52. Wolpert, D. H. Stacked Generalization. *Neural Netw.* **5**, 241–259 (1992).
53. Simon, G. E. *et al.* Does Response on the PHQ-9 Depression Questionnaire Predict Subsequent Suicide Attempt or Suicide Death? *Psychiatric Services.* **64**, 1195–1202 (2013).

Loading [MathJax]/jax/output/CommonHTML/jax.js

54. Litster, B. *et al.* Validation of the PHQ-9 for Suicidal Ideation in Persons with Inflammatory Bowel Disease. *Inflamm Bowel Dis.* **24**, 1641–1648 (2018).
55. Altura, K. C. *et al.* Suicidal ideation in persons with neurological conditions: prevalence, associations and validation of the PHQ-9 for suicidal ideation. *Gen Hosp Psychiatry.* **42**, 22–26 (2016).
56. Barnhofer, T. *et al.* Mindfulness-based cognitive therapy as a treatment for chronic depression: A preliminary study. *Behaviour Research and Therapy.* **47**, 366–373 (2009).
57. Crane, C. *et al.* Comfort from suicidal cognition in recurrently depressed patients. *J Affect Disord.* **155**, 241–246 (2014).
58. Hirsch, J. K. & Conner, K. R. Dispositional and explanatory style optimism as potential moderators of the relationship between hopelessness and suicidal ideation. *Suicide and Life-Threatening Behavior.* **36**, 661–669 (2006).

Figures

Loading [MathJax]/jax/output/CommonHTML/jax.js

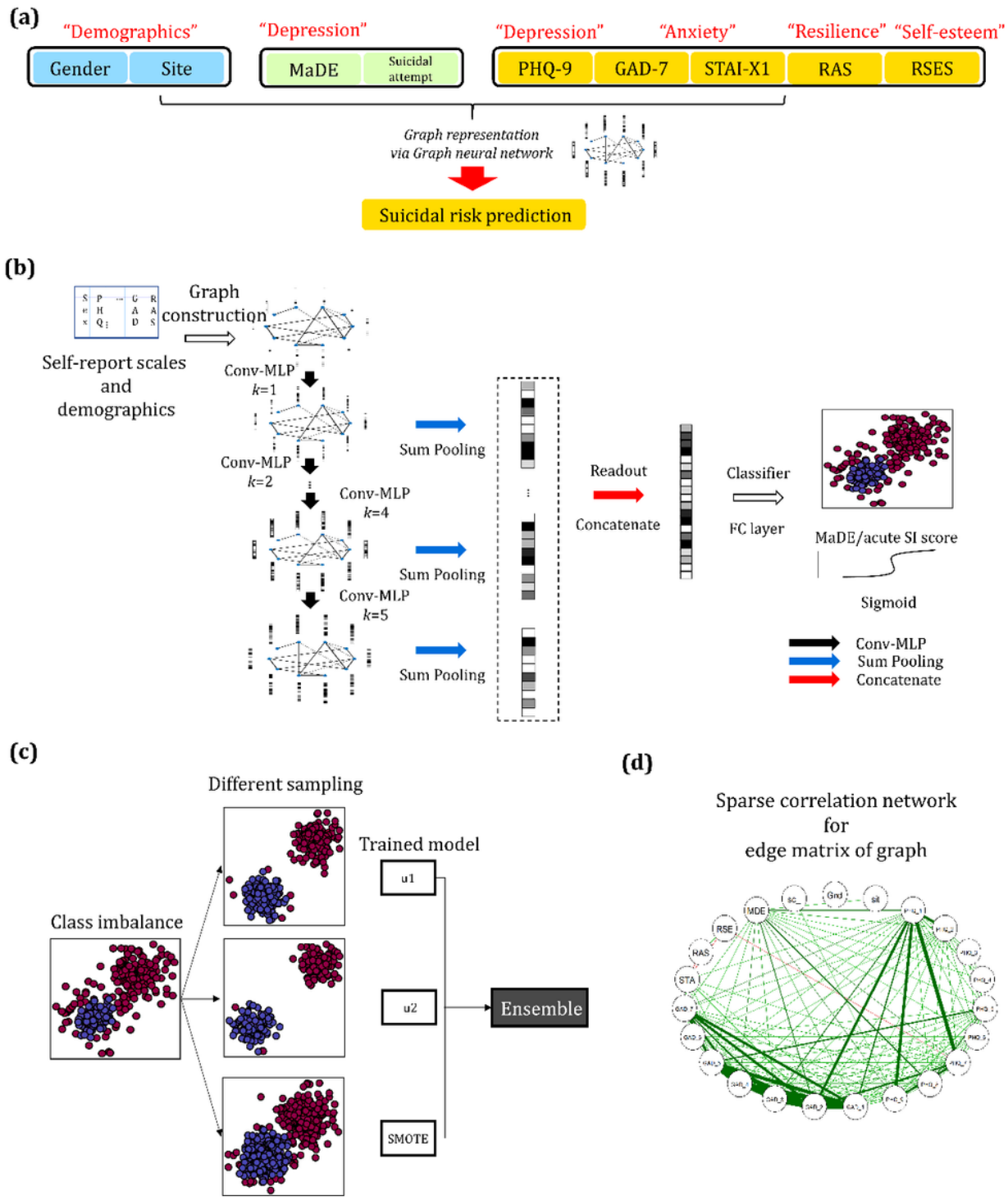

Figure 1

Overall architecture of MindWatchNet based on graph isomorphism network (GIN): (a) Overall workflow for constructing the graph-structured dataset. A single subject corresponds to a single graph. (b) Graph constructed on multi-dimensional self-report questionnaires and clinico-demographic information, which were used as input node features and fed into 5 graph convolutional layers combined with MLP layers. In each layer, the graph representation is obtained after graph pooling, concatenated into a latent feature connected layer to output a sigmoid prediction of MaDEs or

Loading [MathJax]/jax/output/CommonHTML/jax.js

acute SI score (0-1). (c) Three different GIN models with different subsampling strategies (i.e., GIN-SMOTE, GIN-u1, and GIN-u2) were ensembled to obtain the final MindWatchNet to overcome class imbalance. (d) Sparse correlation network for the edge matrix of the graph: pairwise correlation coefficients were obtained between the categorical variables of each node, representing each questionnaire item and subject information feature and were used as edge matrix to construct a graph. All subjects share the same edge matrix to construct a graph. Note that the PHQ-9 and GAD-7 show positive intra- and inter-correlations, whereas the RAS and RSES total scores are negatively correlated with the STAI-S total score. The sparsity of the graph edges was controlled by setting the threshold to 0.6. Note: Thickness of lines indicates degree of correlation coefficients in (d): i.e. a strong correlation between nodes is indicated as a thick line connecting the nodes. Abbreviations: MaDE, major depressive episode; SI, suicidal ideation; SMOTE, synthetic minority oversampling technique; u1, and u2, undersampling strategy 1, and 2, respectively

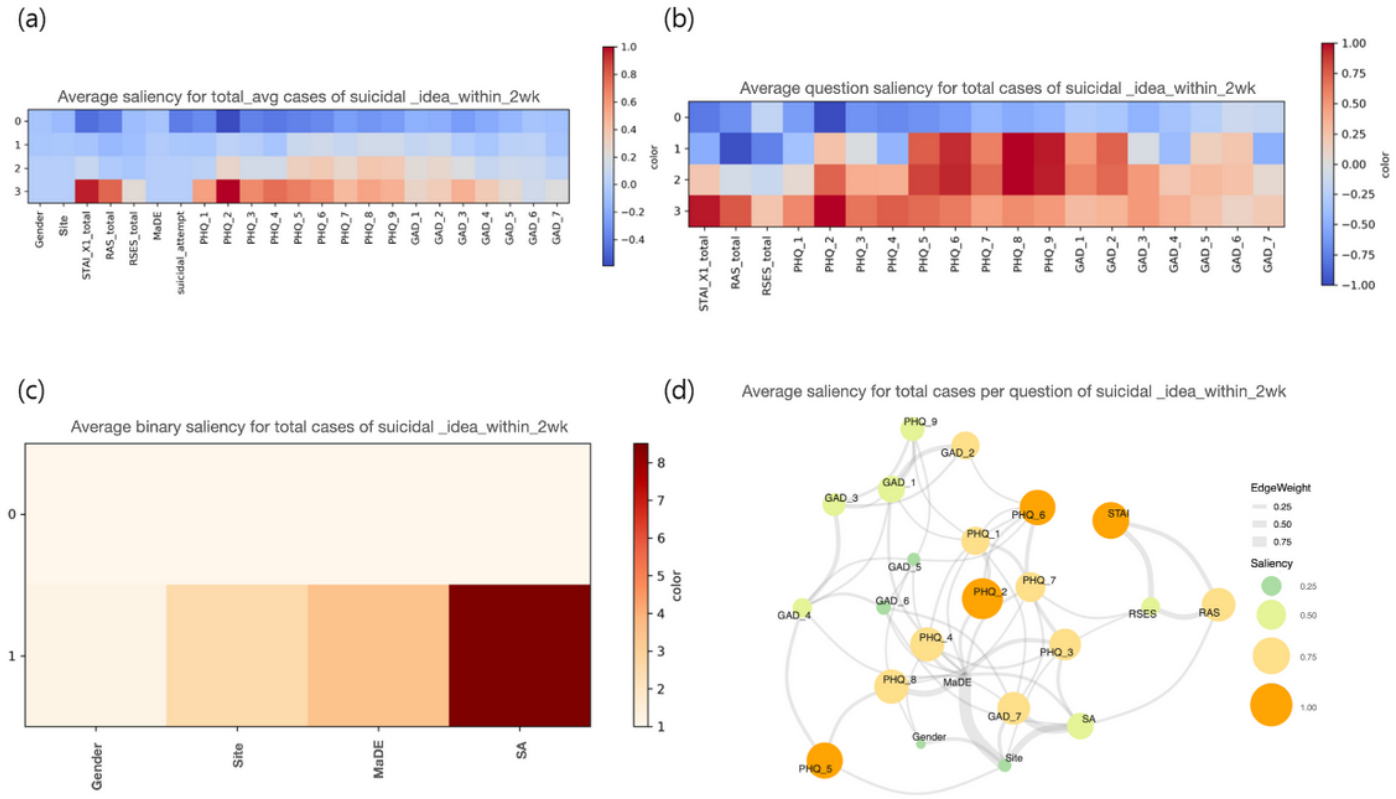

Figure 2

Attention plots. (a) The raw averaged attention plots, (b) the attention plot comparing the questionnaire items, (c) the attention plot comparing the binary items, (d) the attention plot for the mixed Gaussian model-based graphical network (generated with the “mgm” R package for visualization) for the questionnaire items on the test set (n=14,238).

Loading [MathJax]/jax/output/CommonHTML/jax.js

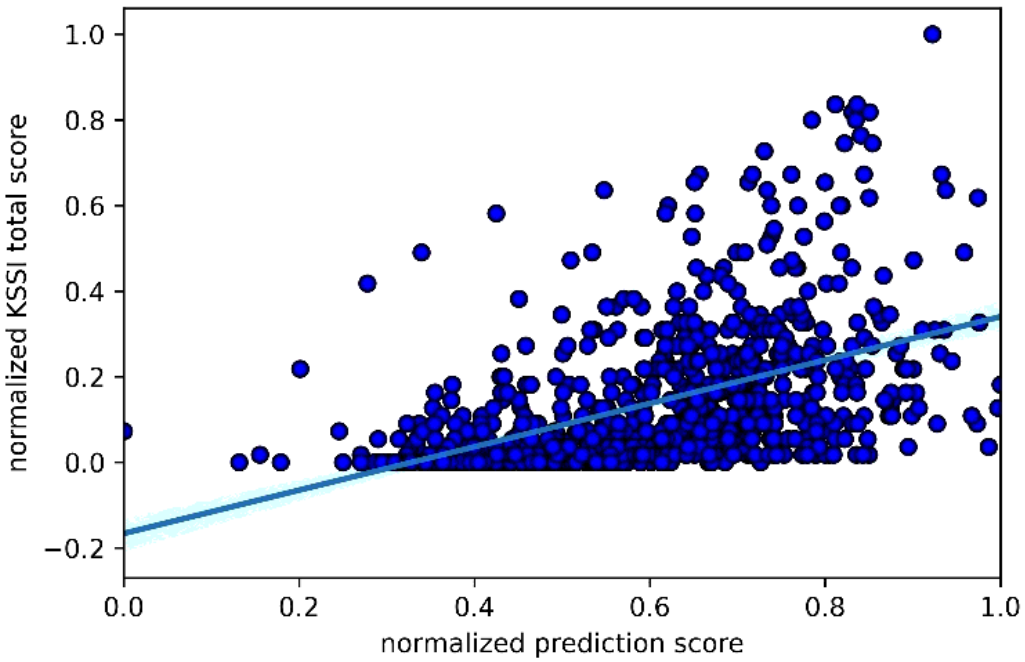

Figure 3

Scatter plot showing the normalized raw prediction score of the model and KSSI total score (n=792 of 13,408; the part of the test set with KSSI scores). Spearman's correlation coefficient between the prediction score of the model and the KSSI total score was ρ_pred=0.719 (p<0.0001).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplfinalFeb22.docx

Loading [MathJax]/jax/output/CommonHTML/jax.js