Check for updates

OPEN ACCESS

EDITED BY Areej Alhothali, King Abdulaziz University, Saudi Arabia

REVIEWED BY Tao Xu, Northwestern Polytechnical University, China Gengyu Lyu, Beijing University of Technology, China

*CORRESPONDENCE Xuetao Tian Xttian@bnu.edu.cn

RECEIVED 24 April 2025 ACCEPTED 20 June 2025 PUBLISHED 18 July 2025

CITATION

Liu Y, Zheng J, Zeng Y, Luo F and Tian X (2025) Graph learning based suicidal ideation detection via tree-drawing test. *Front. Psychiatry* 16:1617650. doi: 10.3389/fpsyt.2025.1617650

COPYRIGHT

© 2025 Liu, Zheng, Zeng, Luo and Tian. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Graph learning based suicidal ideation detection via tree-drawing test

Ye Liu¹, Jiashuo Zheng¹, Yang Zeng², Fang Luo² and Xuetao Tian^{2,3*}

¹School of Future Technology, South China University of Technology, Guangzhou, China, ²Faculty of Psychology, Beijing Normal University, Beijing, China, ³Engineering Research Center of Integration and Application of Digital Learning Technology, Ministry of Education, Beijing, China

Introduction: Adolescent suicide is a critical public health concern worldwide, necessitating effective methods for early detection of high suicidal ideation. Traditional detection methods, such as self-report scales, suffer from limited accuracy and are susceptible to personal concealment. Automatic methods based on artificial intelligence techniques are more accurate, while they often lack scalability due to strict data requirements. In order to achieve a balance between accuracy and scalability, this paper introduces the Tree-Drawing Test (TDT) as an effective tool for suicidal ideation detection, and proposes a novel graph learning approach to enable its automatic application.

Methods: The proposed method first constructs a semantic graph based on psychological features annotated automatically from tree-drawing images, and leverages a Graph Convolutional Network (GCN) model to realize individual suicidal ideation detection. To evaluate this method, a real dataset of 806 students from primary and secondary school in Shaanxi Province, China, is collected, and some metrics including macro-F1, G-mean, and false positive rate are used.

Results: The results demonstrate that the proposed method significantly outperforms traditional machine learning and convolution neural network approaches. The ablation study demonstrates the effectiveness of feature "leaves and fruits" in detecting suicidal ideation. Further experiments demonstrate that the proposed method remains stable even when the model is disturbed, such as when a tree-drawing image cannot be fully represented.

Discussion: The proposed method highlights its effectiveness in large-scale suicidal ideation screening, as it not only achieves high detection performance but also maintains model stability while remaining flexible and adaptable.

KEYWORDS

suicidal ideation detection, tree-drawing test, projective test, graph learning, graph convolutional network

1 Introduction

Adolescent suicide has been a public health concern. Data from the World Health Organization (WHO) shows that more than 1.5 million adolescents and young adults aged 10 to 24 years died in 2021 (1), and suicide has been the fourth leading cause of death among young people from the ages of 15 to 29 (2). Meanwhile, approximately one-third of adolescents experiencing suicidal ideation progress to developing suicide plans, and approximately 60% of those have attempted suicide (3). In this case, early detection and intervention of suicidal ideation through large-scale screening is crucial to promoting healthy adolescent development.

Current tools for detecting suicidal ideation mainly contain selfreport scales, artificial intelligence (AI)-driven detection methods and projective tests as shown in Figure 1. Although self-report scales [e.g., Beck Scale (4)] facilitate efficient large-scale screening for suicidal ideation, their results are susceptible to inaccuracies and self-



FIGURE 1

Flowchart of suicidal ideation detection. (a) Flowchart of suicidal ideation detection via self-report scales. (b) Flowchart of suicidal ideation detection via automatic AI model. (c) Flowchart of suicidal ideation detection via the drawing projective tests.

presentation biases, particularly those arising from intentional selfconcealment of sensitive information. AI-driven detection methods, which utilize multi-modal data such as text (5-7), audio (8-10), video (11, 12), electroencephalogram (EEG) (13, 14), etc., demonstrate superior accuracy but lower scalability, as they require specialized data collection process. For instance, researchers collected video and audio data from clinical interviews to analyze smile and gaze behavior, and used SVM method to classify suicidal patients, psychiatric patients and control groups (11). Another study collected restingstate fMRI data from depressed patients in clinical suicidal crisis, constructed a feature mask via two-sample t-test on regional connectivity, and designed a semisupervised clustering framework using the mask for suicidal ideation prediction (15). Projective tests like the Tree-Drawing Test (TDT) have been applied to identify individuals with psychological states like depressive disorders (16) and dissociative identity disorder (17). By leveraging ambiguous stimuli to uncover latent emotions, these methods mitigate subjective bias, enhance response authenticity, and facilitate the scalability of psychological evaluation. However, this approach heavily relies on expert interpretation and cannot achieve a timely diagnosis.

To address this challenge of supporting large-scale screening, computer vision methods, such as image processing and Convolutional Neural Network (CNN), have been applied to the TDT successfully (18–20). However, since they focus more on the texture features of tree drawing images and lack the prior knowledge of psychology, it is difficult for them to capture the highly relevant characteristics between TDT and psychological states. Against this background, this paper aims to leverage graph learning methods to employ explainable semantic features related to suicidal ideation and achieve powerful detection capabilities simultaneously, thereby promote TDT's application in the field of suicidal ideation detection.

2 Materials and methods

2.1 Dataset

To conduct this research, we established a dataset involving lots of students from primary and secondary schools in Shaanxi Province, China. There are totally 806 participants (392 males and 414 females). Each participant completed the Beck Scale for Suicide Ideation and drew a tree on A4-sized white paper as part of the Tree-Drawing Test. To ensure the effectiveness of scale, we evaluated itemtotal correlation and refined it to nine items, achieving a Cronbach's α of 0.89, indicating high internal consistency and reliability (21, 22). The scoring ranges from 0 to 10 points for each item, with a total score ranging from 0 to 90 for the nineitem scale. The score distribution is displayed in Figure 2. The average score for suicidal ideation across all subjects was 17.19 (SD = 7.01). To categorize the subjects into groups based on their level of suicidal ideation, we applied a threshold of mean plus standard deviation, a common practice in psychological research (23, 24). Consequently, 94 subjects were classified into the high suicidal ideation group, while the remaining 712 were categorized as low suicidal ideation.



For the Tree-Drawing Test, we selected 98 significant features that can be grouped into 12 classes according to the current psychological research (25–28). These features are shown in Table 1, they were manually labeled by three graduate students specializing in psychometrics. If a certain feature was present in an image, it was marked as "1"; if absent, it was marked as "0". The labeling process was conducted independently by each annotator,

resulting in an initial agreement rate of 94%. Any disagreements were resolved through discussion among annotators to reach consensus final labels. Notably, the data collection was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Beijing Normal University (protocol code: 202112300092, and date of approval: December 30, 2021).

2.2 Baselines

To identify the high suicidal ideation, any binary classifier of machine learning algorithms can be effectively applied to features, extracted by automatic tree features extraction model or manually labeled from images. In this study, we adopted some machine learning (ML) models, including Logistic Regression (LR) (29), Decision Tree (DT) (30), Support Vector Machine (SVM) (31) and Random Forest (RF) (32) to implement binary classification. Besides, the suicidal ideation detection on images can be realized by deep learning techniques directly. Convolutional Neural Networks (CNN) is a common framework in deep learning, and also widely used in image classification. We selected some classic CNN models, such as AlexNet (33), VGG16 (34), Inception (35) and ResNet (36) as our baselines. To make a comparison with state-of-the-art graph learning models, we also selected GAT (37), HAN (38) and Simple-HGN (S-HGN) (39) as baselines.

TABLE 1 Classes and details of tree-drawing features.

Class	Features
overall	huge tree; small tree; tree above; tree below; tree on the left; tree on the right; tilted trunk or crown; cut at the edge of the tree and paper; cut above; cut on the left; cut on the right; cut below
line	thick lines; light lines; obvious differences in strength; sloppy lines; mainly short lines; mainly long lines; mainly trembling wavy lines
special mark	snakelike; symmetrical
special tree type	dead tree; simplified tree; pine tree; theme loss; multiple trees
canopy	closed canopy; open canopy; zoned canopy; large canopy; small canopy; squashed canopy; full canopy; emphasized canopy lines; light canopy lines; multi-layered canopy lines; cloud-like canopy lines; ringed canopy line; quivering shaped canopy lines; circular canopy lines; black shadows in the canopy; chaos in the canopy; blank canopy; detailed depicted canopy; short wavy lines drawn in the canopy
branch	broken branches; crossing branches; vigorously growing branches; drooping branches; patchwork branches; single-line branches; parallel branches; open branch ends; sharp branch ends
leaf and flower	fruits or flowers in the canopy; leaves; fallen leaves
trunk	small trunk; long trunk; parallel trunk; wide at the top and narrow at the bottom; wide at the bottom of the trunk; scars on the trunk; black shadow on the trunk; depiction of bark; completely blank trunk; emphasis on trunk edge lines; light trunk edge lines; trunk with small twigs
junction	trunk sealed at the top; trunk open at the top; trunk forms a 'M' shape at the top; trunk sealed at the bottom; trunk transitionally sealed at the bottom (drawing whisker-like roots); trunk open at the bottom; trunk directly connects to the branches and the junction is hollow
root	drawn roots; fibrous roots; sharp roots; crossing roots; overly drawn roots; exposed roots
ground	drawn ground line; sloping ground line; wavy ground line; hilly ground line; emphasized ground; paper's base as ground line (only if depicted flowers and plants); ground line through the trunk
attachment	drawn attachments; sun; clouds; flowers under the tree; grass under the tree; birds or bird's nests; houses or people; words; wind and rain



FIGURE 3

Overall framework of the proposed model, which consists of three modules: (1) automatic feature extraction, (2) semantic graph construction, and (3) suicidal ideation detection. Such a method formalizes individual suicidal ideation detection as a node classification task in graph. Based on graph learning, the images from Tree-Drawing Test can be better represented via combining psychological tree features and deep learning techniques, leading to a better performance on suicidal ideation detection.

2.3 Proposed model

2.3.1 Overview

In this study, we proposed a new model to achieve automatic suicidal ideation detection based on TreeDrawing Test, as illustrated in Figure 3. The model consists of three modules. The first one is automatic feature extraction, determining which tree features are contained in Tree-Drawing Test images. The second one is semantic graph construction, it aims at establishing a graph, where the weight of edge represents the correlation between psychological tree feature and individual image. The third one is suicidal ideation detection, which realizes the detection of individual suicidal ideation based on the Tree-Drawing Test and graph node classification. Such a method transforms traditional image classification task to graph node classification task. Based on Graph Convolutional Network (GCN), the images from Tree-Drawing Test can be better represented, leading to a better performance on suicidal ideation detection. For clearer presentation, Table 2 displayed the main notations and descriptions.

2.3.2 Automatic tree features extraction

Based on the manually-labeled features, we fine-tuned a ResNet network to achieve automatic tree features extraction through multi-label classification of 98 image features. Specifically, we modified the output of the last fully-connected layer in ResNet34 (36) as 98 parallel linear layers. The model performed binary classification for each feature, and outputted a 98-dimensional

Symbols	Notations and descriptions
G	Image-feature semantic graph
ν	Vertices set
\mathcal{V}_F	Feature vertices set
\mathcal{V}_{I}	Image vertices set
ε	Edge set
Α	Adjacency matrix
х	Initial representation of all vertices
$\mathbf{H}^{(l)}$	Feature representation at the <i>l</i> -th layer
$\mathbf{W}^{(l)}$	Trainable weight matrix of the <i>l</i> -th layer
\mathbf{y}^{j}	The ground truth label for the <i>j</i> -th vertice
$\hat{\mathbf{y}}^{j}$	The prediction for the <i>j</i> -th vertice
W _c	The weight of class <i>c</i>
\mathbf{z}_i^j	The ground-truth of manual annotation whether the <i>j</i> -th image sample has the <i>i</i> -th feature
$\hat{z}^{j}_{i} w^{i}_{f}$	The probability that the model predicts the <i>j</i> -th image sample has the <i>i</i> -th feature The weight of minority class for the <i>i</i> -th feature

TABLE 2 Notations and descriptions of the proposed model.

vector that indicates whether the feature exists in the input image. It is trained by cross entropy loss, which is commonly used for classification tasks. For the *i*-th feature, the loss $\mathcal{L}_{feature}^{i}$ is shown in Equation 1. The total loss L^{1} is the mean of the losses of each feature, as shown in Equation 2.

$$\mathcal{L}_{feature}^{i} = -\frac{1}{M} \sum_{j=1}^{M} (\mathbf{z}_{i}^{j} \log{(\hat{\mathbf{z}}_{i}^{j})} + (1 - \mathbf{z}_{i}^{j}) \log{(1 - \hat{\mathbf{z}}_{i}^{j})}).$$
(1)

$$\mathcal{L}^{1} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_{feature}^{i}, i = 1, 2, ..., N.$$
(2)

Here, $\hat{\mathbf{z}}_i^j$ is the probability that the model predicts the *j*-th image sample has the *i*-th feature and \mathbf{z}_i^j represents the ground-truth of manual annotation whether the *j*-th image sample has the *i*-th feature. *M* is the number of training images, and *N* is 98 since there are 98 predefined tree features.

2.3.3 Semantic graph construction

To link the above psychological tree features and images, a semantic graph is constructed and further used to explore intrinsic semantic information contained in images. Generally, a graph is composed of a finite number of vertices and edges between them, usually represented as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{G} denotes the graph, \mathcal{V} is the set of vertices in graph \mathcal{G} , v_i represents the *i*-th node and \mathcal{E} is the set of edges in graph G. In this model, we take both features and images as vertices to build the image-feature semantic graph, that is, $\mathcal{V} = \mathcal{V}_F \cup \mathcal{V}_I$, where \mathcal{V}_F and \mathcal{V}_I respectively represent the set of feature vertices and the set of image vertices. If the number of images and features are $|\mathcal{V}_F|$ and $|\mathcal{V}_I|$ respectively, then the total number of vertices is $|\mathcal{V}| = |\mathcal{V}_F| + |\mathcal{V}_I|$. Moreover, the edge set \mathcal{E} of semantic graph is generated according to the results of automatic tree features extraction. When the *i*-th image ($v_i \in V_I$) has the *j*-th feature ($v_i \in V_F$), there is an edge $e_{i,i} \in \mathcal{E}$ between image vertex v_i and feature vertex v_i in the graph \mathcal{G} . In graph theory, the adjacency matrix is commonly used to represent the relationship between vertices. For semantic graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the corresponding adjacency matrix is an $|\mathcal{V}|$ by $|\mathcal{V}|$ square matrix $\mathbf{A} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ with each element generated as Equation 3.

$$\mathbf{A}(i,j) = \begin{cases} 1, & \text{if } e_{i,j} \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases}, & \text{where } i, j = 1, 2, \dots, |\mathcal{V}|. \end{cases}$$
(3)

If there is an edge $e_{i,j} \in \mathcal{E}$ between vertex v_i and vertex v_j in graph G, then the element in the *i*-th row and the *j*-th column of adjacency matrix **A** is 1, otherwise it is 0. In this way, the semantic knowledge of Tree-Drawing Test images can be fully preserved in the semantic graph. Consequently, the semantic graph is composed of 904 nodes and 15379 edges, and the edge density is approximately 0.04, indicating that the semantic graph is sparse.

2.3.4 Suicidal ideation detection based on treedrawing test and graph learning

Since each image vertex has its label, where images of the individuals with low suicidal ideation belong to class 0 and the

others belong to class 1, the suicidal ideation detection is transformed into a node classification task on the constructed semantic graph. The graph convolutional network (GCN) has a great expressive power to learn the node representations and has achieved a superior performance in node classification (40). Therefore, the graph convolutional network is adopted to achieve the final classification results by the following three steps:

2.3.4.1 Initial representation

Fundamentally, a GCN takes a graph together with a set of feature vectors as input, where each node is associated with its own feature vector. In this model, the initial feature vector is obtained by node2vec (41), which learns low-dimensional embedding for nodes in image-feature semantic graph via applying random walks on semantic graph starting at a target node. Specifically, it first calculates the transition probabilities between nodes, then generate the walk sequence by biased random walk, and finally obtains the representation through the Skip-gram method (42). It enables structurally-similar vertices to have similar representations. Assume that the feature of each vertex is represented as a *k*-dimensional vector, finally we obtained the set of all vertices representations $\mathbf{X} \in \mathbb{R}^{|V| \times k}$.

2.3.4.2 Graph convolutional network for node representation

To integrate the information of psychological tree features (i.e. the feature vertices) into the image representations (i.e. the embedding of image vertices), we leveraged Graph Convolutional Networks (GCN) (43), which can learn both topology structure of the graph and semantic features of vertices. GCN implements the node representation process as shown in Equation 4.

$$\mathbf{H} = f_{\theta}(\mathbf{X}, \mathbf{A}) \,. \tag{4}$$

Here, **X** is the initial vertex embeddings, θ is the learnable parameter in GCN model, **A** is the adjacency matrix of imagefeature semantic graph, and $\mathbf{H}^{(L)}$ is the representation of all unlabeled image vertices. The key of GCN model is to obtain a good node representation. To achieve this, GCN can update the embedding of vertex by aggregating the information from its corresponding neighbors via a multi-layer structure, as shown in Figure 4. If GCN model has *L* layers, every GCN layer updates the node features according to Equation 4.

$$\mathbf{H}^{(l+1)} = \sigma(\hat{\mathbf{A}} \mathbf{H}^{(l)} \mathbf{W}^{(l)}), \text{ where } l = 0, 1, ..., L - 1,$$
(5)

where $\hat{\mathbf{A}} \in \mathbb{R}^{|V| \times |V|}$ is the normalized adjacency matrix calculated by $\hat{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$, \mathbf{D} represent graph degree matrix, which is a diagonal matrix with each diagonal element as $\mathbf{D}(i, i) = \sum_{i=1}^{|V|} \mathbf{A}(i, j)$. And $\mathbf{W}^{(l)} \in \mathbb{R}^{d_l \times d_{l+1}}$ is a trainable weight matrix of the *l*th layer and σ represents an activation function. $\mathbf{H}^{(l)} \in \mathbb{R}^{|V| \times d_l}$ is the feature representation at the *l*-th layer. Notably, the initial state, $\mathbf{H}^{(0)}$, is set as \mathbf{X} , i.e., $\mathbf{H}^{(0)} = \mathbf{X}$.

2.3.4.3 Node classification

The output of the *L*-th GCN layer is $\mathbf{H}^{(L)}$, and node classification is realized by applying a fully-connected layer with softmax activation



function on it as formulated by Equation 6.

$$\mathbf{y} = softmax(\mathbf{W}\mathbf{H}^{(L)} + \mathbf{b}), \tag{6}$$

where $\mathbf{W} \in \mathbb{R}^{2 \times d_L}$, $\mathbf{b} \in \mathbb{R}^{2 \times 1}$, $\mathbf{y} \in \mathbb{R}^{|\mathcal{V}| \times 1}$. The parameters of the GCN are trained by the commonly-used cross entropy loss as shown in Equation 7.

$$\mathcal{L}^{2}(\hat{\mathbf{y}}, \mathbf{y}) = -\frac{1}{T} \sum_{j=1}^{T} (\mathbf{y}^{j} \log{(\hat{\mathbf{y}}^{j})} + (1 - \mathbf{y}^{j}) \log{(1 - \hat{\mathbf{y}}^{j})}), \quad (7)$$

where $\hat{\mathbf{y}}$ and \mathbf{y} are the predicted values and ground truth labels of image vertices, that is the prediction and real labels of Tree-Drawing Test image about suicidal ideation, respectively. *T* is the number of training image vertices with labels in the semantic graph. $\hat{\mathbf{y}}^{j}$ indicates the prediction for the *j*-th vertice. The whole GCN model is optimized through back propagation.

2.3.5 Cost-sensitive strategy for class-imbalanced issue

It is worth noting that the number of individuals with high suicidal ideation is usually much smaller than that of individuals with low suicidal ideation, so there is an class-imbalance issue for suicidal ideation detection. Similarly, automatic tree feature extraction also suffer from class-imbalance issue. To address issue of the class-imbalanced distribution during both automatic feature extraction and suicidal ideation detection, cost-sensitive strategy is employed by leveraging weighted cross entropy loss, which is shown in Equation 8. By giving minority class with a larger weight, costsensitive strategy would penalize more if incorrect prediction is achieved for these minority class. As a result, the prediction accuracy for these minority class is improved. For the binary classification process of suicidal ideation detection, by using costsensitive strategy, the loss function of GCN model in Equation 7 is revised as follows:

$$\mathcal{L}_{imbalance}^{2}(\hat{\mathbf{y}}, \mathbf{y}) = -\frac{1}{T} \sum_{j=1}^{T} (w_{c} \mathbf{y}^{j} \log{(\hat{\mathbf{y}}^{j})} + (1 - \mathbf{y}^{j}) \log{(1 - \hat{\mathbf{y}}^{j})}).$$
(8)

Here, w_c is the weight of class c, where the *j*-th sample belongs to, i.e., $c = \mathbf{y}^j$. When the *j*-th image vertice represents sample with high suicidal ideation (class 1 and $\mathbf{y}^j = 1$), a larger weight w_1 is assigned to the minority class sample. Through assigning a larger weight w_c for each sample in minority classes, the penalty on minority class in the loss function can be adjusted to alleviate the class-imbalanced issues and further detect more effective individuals with high suicidal ideation as well as tree features.

Similarly, in the multi-label classification process of automatic tree feature extraction, the cost-sensitive strategy is employed to handle the class-imbalanced tree features distribution, the loss function of automatic tree features extraction model in Equation 1 for the *i*-th feature is revised as follows:

$$\mathcal{L}_{feature,imbalance}^{i} = -\frac{1}{M} \sum_{j=1}^{M} M(w_{f}^{i} \mathbf{z}_{i}^{j} \log{(\hat{\mathbf{z}}_{i}^{j})} + (1 - \mathbf{z}_{i}^{j}) \log{(1 - \hat{\mathbf{z}}_{i}^{j})}).$$
(9)

Here, w_f^i is the weight of minority class for the *i*-th feature. For the *i*-th feature classification, if images with the *i*-th feature belongs to the minority class, a larger weight w_f^i is assigned to images with label $\mathbf{z}_i^i = 1$. If images without *i*-th feature belongs to the minority class, a larger weight weight w_f^i is assigned to images without *i*-th feature, i.e., images with label $\mathbf{z}_i^i = 0$.

2.4 Experimental settings

Lots of experiments are conducted to validate the effectiveness of the proposed method. These experiments are designed to address the following research questions:

• *RQ-1*: Does the proposed graph learning method outperform other baselines?

- *RQ-2*: Can automatic tree features extraction replace manual annotation for the suicidal ideation detection task?
- *RQ-3*: Does each class of tree features contribute to the individual suicidal ideation detection?
- *RQ-4*: How do different hyperparameters (i.e. cost-sensitive weight for suicidal ideation detection, cost-sensitive weight for feature extraction, and the number of layers in GCN) affect the performance of suicidal ideation detection?
- *RQ-5*: Does the graph learning method perform stable on different sizes of training set?

When comparing the proposed method with the baselines, features are firstly extracted by automatic feature extraction module for each image. Then, an image can be represented as a 98dimensional vector, where each dimension represents the presence or absence of the corresponding feature. For ML models, classification results are obtained by applying ML models on the extracted 98-dimensional vector. For the graph learning model represented by GCN, the extracted tree features are used to build the image-feature semantic graph, and then the model is applied for suicidal ideation detection task. In graph learning models, such as HAN and Simple-HGN, the image-feature semantic graph is represented as a heterogeneous graph, where image nodes and feature nodes are treated as two distinct types of nodes. The two types of edges are "image-has-feature" and "feature-exists in-image." For CNN models, suicidal ideation detection can be seen as a simple image classification task with original image as input.

Additionally, Table 3 shows the hyperparameters in the proposed method. Specifically, we used a multi-label learning model to implement automatic tree-drawing features extraction. By using the manuallyannotated features as labels, we trained the automatic tree features extraction model to obtain feature predictions for each image. During the training, the learning rate is set as 0.01, and the weight decay is set as 0.000001. For the class-imbalance issue in automatic tree feature extraction, the cost-sensitive weighted hyperparameter (w_f^i in Equation 9) is set to 4.0. Also, a two-layer GCN (L = 2) is employed as

Hyperparameter	Description	Value
L	The number of GCN layers.	2
activation function	A mathematical function used in neural networks to introduce non-linearity, allowing the network to learn complex patterns.	ReLU
learning rate	A hyperparameter that controls the step size of parameter updates during model training.	0.01
weight decay	A technique used to prevent overfitting by adding a penalty term to the loss function.	0.001
W _c	The weight of class c.	7.0
Wj	The weight of minority class for the i- th feature.	4.0

graph learning module, and the ReLU function is used as the activation function in Equation 5. To train GCN model, the learning rate is set as 0.01 and the weight decay is 0.001. For the class-imbalance issue in suicidal ideation detection, the cost-sensitive weighted hyperparameter (w_c in Equation 8) is set to 7.0, These hyperparameters are determined by sensitivity analysis as shown in section 3.4. Besides, after the semantic graph construction, the graph contains 904 vertices(|V| = 904), comprising 806 image vertices and 98 feature vertices. For performance evaluation, 80% of image vertices are used as the training set and the rest 20% of image vertices are utilized as the testing set. The performance on the testing set is reported.

2.5 Metrics

Regarding the task of suicidal ideation detection, it is more important to minimize missed cases with high suicidal ideation. Since suicidal ideation detection suffers from serious classimbalanced issue, the individuals who actually have high suicidal ideation (class 1) but are predicted as low suicidal ideation (class 0) are our main concern. Therefore, the following metrics, widely used in imbalanced classification, are employed to evaluate performance: precision of class 0 (precision₀), recall of class 1 (recall₁), macro average of F1 score (macro-F1), G-mean and false positive rate (FPR). They are defined in Equations 10–14 as follows (44):

precision₀ =
$$\frac{\text{TN}}{\text{TN} + \text{FN}}$$
, recall₀ = $\frac{\text{TN}}{\text{TN} + \text{FP}}$, (10)

precision₁ =
$$\frac{\text{TP}}{\text{TP} + \text{FP}}$$
, recall₁ = $\frac{\text{TP}}{\text{TP} + \text{FN}}$, (11)

$$F1_0 = 2 \times \frac{\text{precision}_0 \times \text{recall}_0}{\text{precision}_0 + \text{recall}_0},$$
 (12)

$$F1_{1} = 2 \times \frac{\text{precision}_{1} \times \text{recall}_{1}}{\text{precision}_{1} + \text{recall}_{1}},$$
$$\text{macro} - F1 = \frac{1}{2} \times (F1_{0} + F1_{1}),$$

$$G - mean = \sqrt{\frac{recall_1 \times TN}{TN + FP}},$$
(13)

$$FPR = \frac{FP}{FP + TN},$$
 (14)

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative respectively. $F1_0$ and $F1_1$ represent F1 scores of class 0 and class 1 respectively. Macro-F1 can evaluate the model's performance by treat all classes equally. G-mean is a comprehensive indicator of the recall of class 0 and class 1. For the above metrics, a higher value indicates better performance. The False Positive Rate (FPR) measures the model's tendency to incorrectly predict negative samples as positive. A lower FPR indicates better performance in identifying negative samples.

model	precision ₀	$recall_1$	macro- F1	G-mean	FPR	model*	precision ₀	$recall_1$	macro- F1	G-mean	FPR
LR	89.36%	21.05%	54.78%	43.22%	11.27%	LR*	89.93%	26.32%	56.68%	48.13%	11.97%
DT	89.58%	21.05%	56.22%	43.73%	9.15%	DT*	89.15%	26.32%	52.24%	46.17%	19.01%
SVM	88.96%	10.53%	53.98%	31.87%	3.52%	SVM*	90.13%	21.05%	60.88%	45.07%	3.52%
RF	88.96%	10.53%	53.98%	31.87%	3.52%	RF*	89.03%	10.53%	54.46%	31.98%	2.82%
AlexNet	88.17%	42.11%	44.09%	49.31%	42.25%	-	-	-	_	-	-
VGG16	86.89%	57.89%	35.35%	46.49%	62.68%	-	-	-	-	-	-
Inception	86.44%	57.89%	34.46%	45.60%	64.08%	-	-	-	-	-	-
ResNet	85.29%	47.37%	35.65%	43.99%	59.15%	-	-	-	-	-	-
GAT	89.74%	57.89%	42.60%	53.42%	50.70%	GAT*	91.09%	52.63%	50.52%	58.39%	35.21%
HAN	87.95%	47.37%	41.72%	49.35%	48.59%	HAN*	90.72%	52.63%	48.87%	57.11%	38.03%
S-HGN	88.76%	47.37%	44.09%	51.34%	44.37%	S-HGN*	90.00%	42.11%	50.71%	54.18%	30.28%
GCN	90.59%	57.89%	45.50%	56.03%	45.77%	GCN*	92.68%	52.63%	60.56%	65.00%	19.72%

TABLE 4 Performance comparison results.

Bold indicates the best result among all methods.

*represents the situations where the tree features are manually-annotated.

-represents the situations where the model implementation does not rely on manually-annotated tree features.

3 Results

3.1 The proposed GCN method achieves better performance than baselines (RQ-1)

The performance of all methods are reported in Table 4 (left), with the cost-sensitive strategy consistently applied to both CNN models and graph learning models. From the results, both ML models and graph learning models have better result on macro-F1, demonstrating that these models perform well in predicting both high suicidal ideation class and low suicidal ideation class. Different from CNN models, which use convolution operations to extract the features of images, ML models and graph learning models were conducted based on the 98 tree-drawing features extracted from the images (see Section 2.3.2). Since these tree-drawing features are meaningful psychologically, they are relevant to individual suicidal ideation. They are well-suited for the task of suicidal ideation

detection, thereby improving the overall performance. In order to explain the results more intuitively, Figure 5 shows the confusion matrices of decision tree (DT), cost-sensitive DT (45) and GCN. It can be noticed that the TP of GCN is higher than that of DT, thereby the recall₁ is higher. Since recall₁ is a very important metric in our task, which measures the model's ability to recognize the individuals with high suicidal ideation, the ML methods are not competitive for this task. In this case, lower FPR is due to a small number of high suicidal ideation being identified. The highest Gmean also demonstrates this point. Conceivably, compared with CNN and ML models, different tree-drawing image nodes are connected through tree-drawing feature nodes in semantic graph, similarities and differences among different samples during training can be better captured by graph.

As for the comparison with other graph learning models, GCN outperforms in terms of all metrics. We hypothesize that, due to the limited number of node types and edge types in the image-feature



semantic graph, heterogeneous graph learning models are unable to fully demonstrate their advantages. Compared to GAT, which assigns adaptive edge weights through attention mechanisms, GCN still demonstrates superior performance. This discrepancy may suggest that maintaining equal edge weights is more suitable for the current task. Additionally, this could indicate that the current dataset size might not be sufficient for graph learning to learn the feature importance. Further exploration is warranted with a larger dataset in the future research.

3.2 Automatic extraction of tree-drawing features has room for improvement (RQ-2)

As mentioned in Section 2.3, we used a multi-label learning model to implement automatic tree-drawing features extraction. On the basis, we compared the models using automatically-extracted tree features with that using manually-annotated tree features for individual suicidal ideation detection. The results are reported in Table 4 (left vs. right). For a specific model, the suicidal ideation detection performance with manually-annotated tree features can be considered as idealized result. If the performance obtained using automatically-extracted tree features is closer to the idealized result, it means automatic feature extraction is more effective. From the results, it is apparent that individual suicidal ideation detection depends on accurate tree features for both machine learning models and graph learning models. In our method, the treedrawing features are represented as vertices, which only have connections with image vertices in the image-feature semantic graph. The representation of a tree-drawing image would heavily be affected by the quality of features based on the information propagation mechanism of GCN, leading to significant improvement in terms of macro-F1 and G-mean metrics with manually-annotated feature for GCN. Thus, there is still significant

ΤA	BLE	5	Feature	ab	lation	exper	iment
----	-----	---	---------	----	--------	-------	-------

room for improvement in automatic feature extraction in the future. Besides, it can be seen that GCN model outperforms ML models in suicidal ideation detection even under the setting of using manuallyannotated tree features, demonstrating the effectiveness of the proposed method again. In an ideal scenario, the performance of GCN still surpasses that of the ML models, CNN models, and other graph learning models. And, the ideal GCN model performs well in both macro-F1 (60.56%) and G-mean (65.00%), and also demonstrates good performance in recall₁ (52.63%) and FPR (19.72%). This indicates that the GCN model has the highest potential for suicidal ideation detection based on Tree-Drawing Test.

3.3 All classes of tree-drawing test features can contribute to suicidal ideation detection verified by ablation study (RQ-3)

As above-mentioned in Table 1, there is a total of 12 features classes. In order to prove the effectiveness of the features we selected, we conducted an ablation experiment. Based on the manually-annotated tree features, we removed each class of features respectively, and further conducted a GCN-based suicidal ideation detection experiment with the remaining features. The results are shown in Table 5. The "removed feature" column represents the removed feature class, and the first "none" row means that all features are retained. The results indicate that the performance of the model after removing a certain class of features is slightly inferior to that without removing any class of features, which can be explained to a certain extent that all 12 classes of features are effective for individual suicidal ideation detection. Moreover, it is worth noting that when we removed the feature class "leaf and flower", the performance had the most significant decline. In contrast, the performance reduction caused by removing feature class "root" or "attachment" is not obvious. We can initially

removed feature	precision ₀	recall ₁	macro-F1	G-mean	FPR
none	92.68%	52.63%	60.56%	65.00%	19.72%
overall	92.04%	52.63%	55.71%	62.09%	26.76%
line	91.89%	52.63%	54.81%	61.49%	28.17%
special mark	92.11%	52.63%	56.17%	62.38%	26.06%
special tree type	92.79%	57.89%	56.65%	64.80%	27.46%
canopy	92.45%	57.89%	54.38%	63.21%	30.99%
branch	91.67%	52.63%	53.49%	60.58%	30.28%
leaf and flower	90.91%	57.89%	46.74%	57.11%	43.66%
trunk	91.82%	52.63%	54.37%	61.18%	28.87%
junction	91.89%	52.63%	54.81%	61.49%	28.17%
root	92.62%	52.63%	60.04%	64.72%	20.42%
ground	92.73%	57.89%	56.19%	64.49%	28.17%
attachment	92.62%	52.63%	60.04%	64.72%	20.42%

Bold indicates the best result among all methods.

conclude that for the task of suicidal ideation detection, features in the class "leaves and flowers", such as "leaves" and "fruits", play a more important role in detecting suicidal ideation than other features. In existing tree-drawing test studies, researchers have identified numerous characteristics related to psychological states, such as the shape of the tree crown, the inclination of the trunk, and the overall size of the tree (16, 46). The importance of "leaves and fruits" validated by our study provides a new perspective to this field. In terms of metric recall₁, the value increased when the feature classes "special tree type", "canopy", "leaf and flower" or "ground" were removed, which may suggest that these features usually appear in the images of low suicidal ideation. In total, all the 12 classes of features have a certain effect on detecting individual suicidal ideation.

3.4 Sensitivity analysis helps determining the best hyperparameters (RQ-4)

3.4.1 Cost-sensitive weight for suicidal ideation detection

In reality, the number of individuals with low suicidal ideation is much larger than that with high suicidal ideation, resulting in a serious class-imbalance issue between class 0 (low suicidal ideation) and class 1 (high suicidal ideation) in the dataset. A cost-sensitive strategy with weighting factor w_c is introduced in Equation 8 to solve this problem. We tested the weight values in the range between 1 and 16 in steps of 1, since the ratio of samples with low suicidal ideation to those with high suicidal ideation is approximately 7.57:1. The performance with the change of weighting factor w_c is shown in Figure 6. When the w_c is 1 or 2, the values of G-mean and recall₁ are zero, which means that the model has difficulties in detecting the samples with high suicidal ideation, and the model gives wrong predictions to all samples with high suicidal ideation. In general, the overall performance increases firstly and then decreases. The model has a more stable performance when the w_c is between 6 and 9. Besides, it seems that the four metrics achieve the best results when w_c is 7. It suggests that the proposed method requires selecting w_c based on data distribution in practical applications.

3.4.2 Cost-sensitive weight for feature extraction

For each feature, there is also a class-imbalance issue. For example, closed tree canopy is more common in tree images, so there are more images with the "closed canopy" feature than that without it in our dataset. And only a very small number of images contain houses or people, so the number of images without "houses or people" feature is much larger than that with it. Therefore, we introduced the weighting factor (named w_f^i) in Equation 9 during the automatic features extraction. For each feature, we gave a larger w_f^i to the feature class with fewer samples when performing the multi-classification task. Following the same range of w_c in Figure 6, we tried 5 different w_f^i values for each feature class and conducted the automatic tree features extraction experiment, and the results are shown in Figure 7. The specific approach is using the manuallyannotated features as labels and applying automatic tree features extraction model. During the training, w_f^i is applied to the class with fewer samples for each feature. Overall, the feature extraction model performs the best when w_f^i is 4.0.

After we obtained automatic feature extraction results based on the models with five different w_f^i , the image-feature semantic graph is constructed and the suicidal ideation task is conducted respectively. The results are shown in Table 6. It can be noticed





that GCN and DT perform better when w_f^i is 4, and the best w_f^i for LR, SVM and RF is 2. From the results, we can find that the performance with w_f^i as 1 is worse than that of 2, 4 and 8, such result demonstrates that the cost-sensitive strategy does improve the performance of individual suicidal ideation detection. Meanwhile, there are opposite effects when the w_f^i is too large.

3.4.3 The number of GCN layers

Since the number of layers in GCN also has great influence in the performance of our suicidal ideation detection model, we experimented on GCN with 1, 2, 3, and 4 layers. The results are shown in Table 7. It shows that the best performance is achieved when the number of layers is 2. Such a result is also consistent with

TABLE 6	Suicidal	ideation	detection	based	on	automatic	feature	coding	with	different w	f.
---------	----------	----------	-----------	-------	----	-----------	---------	--------	------	-------------	----

Model	W _f	Precision ₀	$Recall_1$	Macro-F1	G-mean	FPR
GCN	1	89.36%	47.37%	46.06%	52.93%	40.85%
	2	90.32%	52.63%	47.24%	55.80%	40.85%
	4	90.59%	57.89%	45.50%	56.03%	45.77%
	8	90.00%	52.63%	46.02%	54.79%	42.96%
	16	89.58%	47.37%	46.85%	53.56%	39.44%
LR	1	89.29%	21.05%	54.33%	43.05%	11.97%
	2	88.71%	26.32%	50.28%	45.15%	22.54%
	4	89.36%	21.05%	54.78%	43.22%	11.27%
	8	88.54%	5.26%	50.84%	22.70%	2.11%
	16	89.24%	10.53%	56.09%	32.33%	0.70%
DT	1	88.74%	10.53%	52.63%	31.52%	5.63%
	2	88.81%	15.79%	52.67%	37.58%	10.56%
	4	89.58%	21.05%	56.22%	43.73%	9.15%
	8	88.98%	26.32%	51.44%	45.76%	20.42%
	16	84.47%	15.79%	39.41%	31.10%	38.73%
SVM	1	88.61%	5.26%	51.21%	22.78%	1.41%
	2	88.96%	10.53%	53.98%	31.87%	3.52%

(Continued)

TABLE 6 Continued

Model	W _f	Precision ₀	$Recall_1$	Macro-F1	G-mean	FPR
	4	88.67%	5.26%	51.61%	22.86%	0.70%
	8	88.08%	5.26%	48.84%	22.20%	6.34%
	16	88.75%	5.26%	52.02%	22.94%	0%
RF	1	88.96%	10.53%	53.98%	31.87%	3.52%
	2	88.96%	10.53%	53.98%	31.87%	3.52%
	4	88.54%	5.26%	50.84%	22.70%	2.11%
	8	88.82%	10.53%	53.06%	31.63%	4.93%
	16	87.86%	10.53%	48.62%	30.20%	13.38%

previous researches that stacking more layers in GCN will lead to worse performance, due to vanishing gradients and oversmoothing (47).

3.5 The Proposed GCN model exhibits stability as training data changes (RQ-5)

3.5.1 Stability of the model with different training sizes

In the suicidal ideation task, the image vertices on the graph are divided into 645 training vertices and 161 testing vertices, and GCN in graph learning module is trained with training data via back-propagation. Therefore, in order to show the stability of our proposed method with different training sizes, we conducted the experiments with different size of training data. Keeping the 161 testing vertices unchanged, 580 (90%), 516 (80%), 451 (70%), 387 (60%), 322 (50%) training vertices are randomly selected from 645 training vertices and used to construct the image-feature semantic graph respectively. Then the GCN model is trained on these graphs separately. The results are shown in Figure 8. It is easy to find that the experiment results do not fluctuate as the number of training vertices changes, which effectively proves the stability of the model.

3.5.2 Stability of the model on different edge sizes

Considering the situation where insufficient individual expression or inaccurate automatic feature extraction leads to missing features, the semantic graph may miss some edges, i.e., the connections between tree-drawing image and its features. Thus, it is also necessary to explore the model stability on different numbers of missing edges. To achieve it, we randomly kept edges to simulate the situation where tree features cannot be extracted. We retained 90%, 80%, 70%, 60% and 50% of the edges in the image-feature semantic graph and then performed suicide ideation detection on the newly generated graphs, the results are shown in Figure 9.

From the results, we can notice that when some edges are removed, the overall performance is relatively stable as the edge size decreases. Besides, macro-F1 and G-mean show a downward trend, and this trend become apparent when the edge size is less than 80%. It can be concluded that the absence of edges does have a slight impact on the performance of suicidal ideation detection, and this impact becomes significant when degree of absence increases to a certain value. Such results indicate that our proposed method is still capable of detecting suicidal ideation when less than 20% characteristics of the tree-drawing image cannot be fully represented. Moreover, such a result once again demonstrates that automatic features extraction should be further studied to improve the performance of the proposed method in the future.

4 Discussion

This study employs projective test methodologies, specifically the Tree-Drawing Test (TDT), which effectively uncovers individuals' subconscious psychological states while minimizing susceptibility to social desirability and subjective control biases. Compared to techniques requiring high-precision equipment and technical expertise, such as electroencephalography, the TDT necessitates only paper and pencil,

TABLE 7	The	performance	of the	e GCN	with	different	number	of	layers.
---------	-----	-------------	--------	-------	------	-----------	--------	----	---------

Layer number	Precision ₀	Recall ₁	Macro-F1	G-mean	FPR
1	92.66%	57.89%	55.73%	64.17%	28.87%
2	92.68%	52.63%	60.56%	65.00%	19.72%
3	92.11%	52.63%	56.17%	62.38%	26.06%
4	91.67%	52.63%	53.49%	60.58%	30.28%



enabling broader implementation and lower operational costs. This study focuses on how to automatically detect individual suicidal ideation based on the TDT. Beyond machine learning and image processing techniques, we further leverage graph learning to implement automated detection from projective responses. This approach integrates psychological tree features with graph learning techniques, improving the performance of suicidal ideation detection and thereby advancing the application of the TDT in the field of mental health.

Next, it is worth noting that the theoretical basis of TDT is the psychoanalytic genre (48), which heavily relies on experts' interpretations. To address this issue, existing researches have explored various coding systems to link drawing features with specific mental states (16, 49–52). For instance, trunk width, trunk base opening, and branch ends size are significantly associated with schizophrenia (49). Canopy area, canopy height, canopy width, trunk width and total tree area are related to depressive symptoms (16). Roots, truncated tree, flattened crown, and bizarre tree are considered as the important predictors for mental disorders (50). In our ablation study, "leaves and fruits" demonstrates its effectiveness in detecting suicidal ideation. Generally speaking, in projective tree-drawing test, "leaves and fruits" typically correspond to an individual's connection with their environment and personal growth and aspirations (53), and thus some



researches have adopted "leaves" and "fruit" to predict individual depression (54). This is inherently consistent with our findings. This series of research suggests that projection tests contain rich individual differences, and the relevant features are worth continuously exploring.

Furthermore, considering the associations among negative emotions, abnormal mental state and suicidal ideation, previous studies provide a certain basis for automatic suicidal ideation detection. However, in traditional machine learning methods that initially depend on feature recognition, the accuracy of this feature recognition process significantly impacts model performance. The current study demonstrates that employing graph learning approaches can, to a certain extent, address the issue of performance stability. This may be because graph learning allows for the modeling of complex relationships and interactions between tree-drawing images and tree features (55). This characteristic ensures robust stability of the model under high performance, even when the training set changes. At the same time, the unique "imagefeature" semantic graph structure can better explain the inference of the proposed method. Moreover, graph learning models are inherently flexible and adaptable, capable of incorporating new information dynamically. When new image features are added, the graph model can naturally expand by adding new nodes, but traditional machine learning models can only be redesigned and trained (56, 57). Additionally, more and more graph modeling techniques are beginning to focus on the interpretability of model's decisions (38, 39, 58). Although these complex models did not yield good results on our dataset, they may become future solutions for automated analysis of projection tests as data accumulates.

Finally, there are also some limitations to be explored in the future. First, regarding sample selection, this study only included 806 primary and middle school students from Shaanxi Province, resulting in a small sample size with limited geographic and age diversity. The symbolic meanings of tree drawings may vary across cultural and age groups, so the model's applicability in different cultural contexts requires further consideration. Second, this study adopted a cross-sectional design, capturing data at a single time point. However, suicidal ideation often exhibits dynamic development, and data from a single time point may be significantly influenced by the testing context and participants' current psychological states. Future research should consider integrating projective tests into longitudinal designs, capturing the temporal dynamics of suicidal ideation and allowing for a more comprehensive understanding of its progression over time. Next, due to the small data size and the nature of GCN model, this study did not conduct an in-depth analysis of the importance of each tree-drawing image feature. In future research, with the application of advanced graph models supported by a larger dataset, it is expected that we can explain the effectiveness of these features more thoroughly. Finally, the performance of automated recognition of image features in the treedrawing test still needs improvement, to support more accurate individual suicidal ideation detection.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Institutional Review Board of Beijing Normal University. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because the study was conducted as part of a national education monitoring task, and the relevant educational systems and schools were fully informed about the research objectives and procedures.

Author contributions

YL: Writing – original draft, Funding acquisition, Formal analysis, Methodology. JZ: Methodology, Writing – original draft. YZ: Writing – original draft. FL: Writing – review & editing, Conceptualization, Resources. XT: Funding acquisition, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work is supported by the National Natural Science Foundation of China (No. 62306118, 62207002), the Guangdong Provincial Key Laboratory of Human Digital Twin (2022B1212010004), the Fundamental Research Funds for the Central Universities (2023ZYGXZR105), and Engineering Research Center of Integration and Application of Digital Learning Technology, Ministry of Education (1411003).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

10.3389/fpsyt.2025.1617650

References

1. World Health Organization. Adolescent and young adult health. (2023). Available at: https://www.who.int/news-room/fact-sheets/detail/adolescents-health-risks-and-solutions

2. World Health Organization. *Suicide*. (2023). Available at: https://www.who.int/ news-room/fact-sheets/detail/suicide

3. Wasserman D, Carli V, Iosue M, Javed A, Herrman H. Suicide prevention in childhood and adolescence: a narrative review of current knowledge on risk and protective factors and effectiveness of interventions. *Asia-Pac Psychiatry*. (2021) 13: e12452. doi: 10.1111/appy.12452

4. Beck AT, Steer RA, Ranieri WF. Scale for suicide ideation: Psychometric properties of a self-report version. J Clin Psychol. (1988) 44:499–505. doi: 10.1002/1097-4679(198807)44:4<499::AID-JCLP2270440404>3.0.CO;2-6

5. Bello HJ, Palomar-Ciria N, Baca-García E, Lozano C. Suicide classification for news media using convolutional neural networks. *Health Commun.* (2023) 38:2178–87. doi: 10.1080/10410236.2022.2058686

6. McManus KF, Stringer JM, Corson N, Fodeh S, Steinhardt S, Levin FL, et al. Deploying a national clinical text processing infrastructure. *J Am Med Inf Assoc.* (2024) 31:727–31. doi: 10.1093/jamia/ocad249

7. Heckler WF, Feijó LP, de Carvalho JV, Barbosa JLV. Thoth: An intelligent model for assisting individuals with suicidal ideation. *Expert Syst Appl.* (2023) 233:120918. doi: 10.1016/j.eswa.2023.120918

8. Gideon J, Schatten HT, McInnis MG, Provost EM. Emotion recognition from natural phone conversations in individuals with and without recent suicidal ideation. In: *Interspeech* Graz, Austria: ISCA (2019).

9. Belouali A, Gupta S, Sourirajan V, Yu J, Allen N, Alaoui A, et al. Acoustic and language analysis of speech for suicidal ideation among us veterans. *BioData Min.* (2021) 14:1-17. doi: 10.1186/s13040-021-00245-y

10. Iyer R, Nedeljkovic M, Meyer D. Using voice biomarkers to classify suicide risk in adult telehealth callers: retrospective observational study. *JMIR Ment Health.* (2022) 9:e39807. doi: 10.2196/39807

11. Eigbe N, Baltrusaitis T, Morency L-P, Pestian J. (2018). Toward visual behavior markers of suicidal ideation, in: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Piscataway, NJ, USA: IEEE. pp. 530–4. IEEE.

12. Onie S, Li X, Glastonbury K, Hardy RC, Rakusin D, Wong I, et al. Understanding and detecting behaviours prior to a suicide attempt: A mixed-methods study. *Aust New Z J Psychiatry*. (2023) 57:1016–22. doi: 10.1177/00048674231152159

13. Prasad DK, Liu S, Chen S-HA, Quek C. Sentiment analysis using eeg activities for suicidology. *Expert Syst Appl.* (2018) 103:206–17. doi: 10.1016/j.eswa.2018.03.011

14. Hasey G, Colic S, Reilly J, MacCrimmon D, Khodayari A, DeBruin H, et al. Detection of suicidal ideation in depressed subjects using resting electroencephalography features identified by machine learning algorithms. *Biol Psychiatry.* (2020) 87:S380–1. doi: 10.1016/j.biopsych.2020.02.974

15. Dai Z, Shen X, Tian S, Yan R, Wang H, Wang X, et al. Gradually evaluating of suicidal risk in depression by semi-supervised cluster analysis on resting-state fmri. *Brain Imaging Behav.* (2021) 15:2149–58. doi: 10.1007/s11682-020-00410-7

16. Gu S, Liu Y, Liang F, Feng R, Li Y, Liu G, et al. Screening depressive disorders with tree-drawing test. *Front Psychol.* (2020) 11:1446. doi: 10.3389/fpsyg.2020.01446

17. Morris MB. The diagnostic drawing series and the tree rating scale: An isomorphic representation of multiple personality disorder, major depression, and schizophrenia populations. *Art Ther.* (1995) 12:118–28. doi: 10.1080/07421656.1995.10759142

18. Anica PF, Lucian CV. Image processing techniques, interpretation of parameters and psychological significance by the tree-drawing test. In: *International Conference "SUPERVISION IN PSYCHOTHERAPY" 2nd editionAt.* Timisoara, Romania: IPCS (2019).

19. Anica F, Lucian CV. Automatic image processing of tree drawings for psychological tests. Int Conf Legal Med Cluj. (2020) 2:166-71.

20. Maliki I, Firmansyah AR. (2023). Personality detection based on tree drawing using convolutional neural network, in: 2023 International Conference on Informatics Engineering, Science & Technology (INCITEST), Piscataway, NJ, USA: IEEE. pp. 1–6.

21. Nunnally J. Psychometric Theory 3E. McGraw-Hill series in psychology. New York, USA: Tata McGraw-Hill Education (1994).

22. Cronbach LJ. Coefficient alpha and the internal structure of tests. *psychometrika*. (1951) 16:297–334. doi: 10.1007/BF02310555

23. Reynolds CR. Behavior assessment system for children. Corsini Encyclopedia Psychol. (2010), 1–2.

24. Tian X, Jing L, Luo F, Zhang S. Shyness trait recognition for schoolchildren via multi-view features of online writing. *IEEE Trans Affect Comput.* (2023) 14:509–22. doi: 10.1109/TAFFC.2021.3077410

25. Zhang T, Zhang H. Uncover the secrets of your personality: House-Tree-Person Drawing Test (*China federation of literary and art circles publishing corporation*). (2007).

26. Sorge A, Saita E. Assessment of suicide and self-harm risk in foreign offenders. evaluating the use of tree-drawing test. *Mediterr J Clin Psychol.* (2021) 9:1–18. doi: 10.13129/2282-1619/mjcp-3024

27. Xiong D, Lin R, Ge H, Jiang R, Li L, Yang L, et al. (2023). Development and efficiency research of distressed mood and suicide risk test system with house-treeperson (htp) drawing, in: *International Conference on Man-Machine-Environment System Engineering*, Berlin, German: Springer. pp. 259–64. Springer.

28. Yang G, Zhao L, Sheng L. Association of synthetic house-tree-person drawing test and depression in cancer patients. *BioMed Res Int.* (2019) 2019:1478634. doi: 10.1155/2019/1478634

29. Hosmer DW Jr., Lemeshow S, Sturdivant RX. Appl Logistic Regression. (2013). Hoboken, NJ, USA: John Wiley & Sons. doi: 10.1002/9781118548387

30. Quinlan JR. Induction of decision trees. Mach Learn. (1986) 1:81-106. doi: 10.1007/BF00116251

31. Schölkopf B, Smola AJ. Learning with kernels: support vector machines, regularization, optimization, and beyond. Cambridge, Massachusetts, USA: MIT press (2002).

32. Breiman L. Random forests. Mach Learn. (2001) 45:5–32. doi: 10.1023/A:1010933404324

33. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst.* (2012) 25:1097–105. doi: 10.1145/3065386

34. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. (2014). doi: 10.48550/arXiv.1409.1556

35. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. (2015). Going deeper with convolutions, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Piscataway, NJ, USA: IEEE. pp. 1–9.

36. He K, Zhang X, Ren S, Sun J. (2016). Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Piscataway, NJ, USA: IEEE. pp. 770–8.

37. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. *arXiv preprint arXiv:1710.10903.* (2017). doi: 10.48550/arXiv.1710.10903

38. Wang X, Ji H, Shi C, Wang B, Ye Y, Cui P, et al. (2019). Heterogeneous graph attention network, in: *The world wide web conference*, New York, NY, USA: ACM. pp. 2022–32.

39. Lv Q, Ding M, Liu Q, Chen Y, Feng W, He S, et al. (2021). Are we really making much progress? revisiting, benchmarking and refining heterogeneous graph neural networks, in: *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery* & data mining, New York, NY, USA: ACM. pp. 1150–60.

40. Bhatti UA, Tang H, Wu G, Marjan S, Hussain A. Deep learning with graph convolutional networks: An overview and latest applications in computational intelligence. *Int J Intelligent Syst.* (2023) 2023:8342104. doi: 10.1155/2023/8342104

41. Grover A, Leskovec J. Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (2016). p. 855–64.

42. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. (2013). doi: 10.48550/arXiv.1301.3781

43. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907. (2016). doi: 10.48550/arXiv.1609.02907

44. Sokolova M, Japkowicz N, Szpakowicz S. (2006). Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation, in: *Australasian joint conference on artificial intelligence*, Berlin, German: Springer. pp. 1015–21. Springer.

45. Ting KM. An instance-weighting method to induce cost-sensitive trees. *IEEE Trans Knowledge Data Eng.* (2002) 14:659–65. doi: 10.1109/TKDE.2002.1000348

46. Guo H, Feng B, Liu T, Zhao R, Fan H, Dong Z, et al. Tree imagery in drawing tests for screening mental disorders: A systematic review and meta-analysis. In: *Research Squere* (2024). preprint.

47. Li Q, Han Z, Wu X-M. (2018). Deeper insights into graph convolutional networks for semisupervised learning, in: Proceedings of the AAAI conference on artificial intelligence.

48. Pan T, Zhao X, Liu B, Liu W. (2022). Automated drawing psychoanalysis via house-tree-person test, in: 2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI), Piscataway, NJ, USA. pp. 1120–5. IEEE.

49. Kaneda A, Yasui-Furukori N, Saito M, Sugawara N, Nakagami T, Furukori H, et al. Characteristics of the tree-drawing test in chronic schizophrenia. *Psychiatry Clin Neurosci.* (2010) 64:141–8. doi: 10.1111/j.1440-1819.2010.02071.x

50. Guo H, Feng B, Ma Y, Zhang X, Fan H, Dong Z, et al. Analysis of the screening and predicting characteristics of the house-tree-person drawing test for mental disorders: A systematic review and meta-analysis. *Front Psychiatry*. (2023) 13:1041770. doi: 10.3389/fpsyt.2022.1041770

51. Stanzani Maserati M, Matacena C, Sambati L, Oppi F, Poda R, De Matteis M, et al. The tree-drawing test (koch's baum test): a useful aid to diagnose cognitive impairment. *Behav Neurol.* (2015) 2015:534681. doi: 10.1155/2015/534681

52. Guo Q, Yu G, Wang J, Qin Y, Zhang L. Characteristics of house-tree-person drawing test in junior high school students with depressive symptoms. *Clin Child Psychol Psychiatry*. (2023) 28:1623–34. doi: 10.1177/13591045221129706

53. Koch C. The Tree Test; the tree-drawing test as an aid in psychodiagnosis. New York, USA: Grune & Stratton (1952).

54. Hu Y, Chen J-d. Application of projective tree drawing test in adolescents with depression. *Chin J Clin Psychol.* (2012) 20:185–7. doi: 10.16128/j.cnki.1005-3611.2012.02.041

55. Xia F, Sun K, Yu S, Aziz A, Wan L, Pan S, et al. Graph learning: A survey. *IEEE Trans Artif Intell*. (2021) 2:109–27. doi: 10.1109/TAI.2021.3076021

56. Zhang S, Tong H, Xu J, Maciejewski R. Graph convolutional networks: a comprehensive review. *Comput Soc Networks*. (2019) 6:1–23. doi: 10.1186/s40649-019-0069-y

57. Gao H, Wang Z, Ji S. (2018). Large-scale learnable graph convolutional networks, in: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, New York, NY, USA: ACM. pp. 1416–24.

58. Wang X, He X, Cao Y, Liu M, Chua T-S. (2019). Kgat: Knowledge graph attention network for recommendation, in: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, New York, NY, USA: ACM. pp. 950–8.