Social Media Images Can Predict Suicide Risk Using Interpretable Large Language-Vision Models

Yael Badian, MSc; Yaakov Ophir, PhD; Refael Tikochinski, MA; Nitay Calderon, MSc; Anat Brunstein Klomek, PhD; Eyal Fruchter, MD; and Roi Reichart, PhD

Abstract

Background: Suicide, a leading cause of death and a major public health concern, became an even more pressing matter since the emergence of social media two decades ago and, more recently, following the hardships that characterized the COVID-19 crisis. Contemporary studies therefore aim to predict signs of suicide risk from social media using highly advanced artificial intelligence (AI) methods. Indeed, these new AI-based studies managed to break a longstanding prediction ceiling in suicidology; however, they still have principal limitations that prevent their implementation in real-life settings. These include "black box" methodologies, inadequate outcome measures, and scarce research on non-verbal inputs, such as images (despite their popularity today).

Objective: This study aims to address these limitations and present an interpretable prediction model of clinically valid suicide risk from images.

Methods: The data were extracted from a larger dataset from May through June 2018 that was used to predict suicide risk from textual postings. Specifically, the extracted data included a total of 177,220 images that were uploaded by 841 Facebook users who completed a gold-standard suicide scale. The images were represented with CLIP (Contrastive Language-Image Pre-training), a state-of-the-art deep-learning algorithm, which was utilized, unconventionally, to extract predefined interpretable features (eg, "photo of sad people") that served as inputs to a simple logistic regression model.

Results: The results of this hybrid model that integrated theory-driven features with bottom-up methods indicated high prediction performance that surpassed common deep learning algorithms (area under the receiver operating characteristic curve [AUC] = 0.720, Cohen $d = 0.82$). Further analyses supported a theory-driven hypothesis that at-risk users would have images with increased negative emotions and decreased belongingness.

Conclusions: This study provides a first proof that publicly available images can be leveraged to predict validated suicide risk. It also provides simple and flexible strategies that could enhance the development of real-life monitoring tools for suicide.

J Clin Psychiatry 2024;85(1):23m14962

Author affiliations are listed at the end of this article.

Suicide, a leading cause of death,[1,2] has become an even more pressing public health concern following the crisis of the 2019 coronavirus disease (COVID-19).[3–6] The severe coping measures that were implemented during the crisis triggered a parallel pandemic of mental health hardships[7–13] and suicide behaviors,[14–18] and early detection of suicide risk has become an urgent task. Suicide detection, however, is not a trivial task.[19] In fact, 50 years of suicide research taught us that suicide prediction models typically produce prediction scores that are "only slightly better than chance" (area under the receiver operating characteristic [ROC] curve [AUC] range, 0.56–0.58).[20] This conclusion has been made for various types of suicide-related outcomes, including suicide ideation, behaviors, attempts, and deaths.

It has been only in the past decade when substantial improvements in suicide prediction started to emerge, following the "deep learning revolution"[21] in the field of artificial intelligence (AI).[22–26] A second historical change that contributed to these prediction improvements was the emergence of social media, which provided an unprecedented accessibility to valuable personal data.[27–32] Indeed, formal medical risk factors for suicide, such as prior suicide attempts and psychiatric diagnoses, may not appear explicitly in mundane postings online. However, social media behavior may well contain useful information about emotions and interpersonal relationships—

Editor's Note
We encourage authors to submit papers for consideration as a part of our Focus on Suicide section. Please contact Philippe Courtet, MD, PhD, at pcourtet@psychiatrist.com.

Clinical Points

- Artificial intelligence (AI) opens opportunities for suicide prediction from social media, but major gaps exist (eg, “black box” and scarce research on images).
- This study presents an interpretable, theory-inspired AI model that addresses these gaps and emphasizes the importance of emotions and relationships.
- The study provides proof that images can predict suicide risk alongside simple/flexible methods to conduct this prediction task.

two central psychosocial factors that received considerable research attention in suicidology.[1,33–35]

According to the interpersonal-psychological theory of suicidal behavior,[36,37] the emotional state of the person and their interpersonal and social relationships play a significant role in the formation and maintenance of suicide risk. These two pivotal factors are rooted in the seminal attachment theory by Bowlby and Ainsworth,[38] and they are typically interconnected, as relationships impact people’s emotions and vice versa. From a more practical, clinical perspective, the two evidence-based treatments that focus specifically on these two factors of emotions and relationships are interpersonal psychotherapy (IPT)[39] and attachment-based family therapy (ABFT).[40,41] Both of these therapeutic approaches have shown promising results in targeting and addressing emotional and relational aspects that contribute to suicidal behavior, thus emphasizing the potential of early identification of these factors for suicide prevention.

To date, however, the promising field of AI-based suicide prediction from social media suffers from key conceptual gaps that prevent its application to real-life settings. The first gap, which has also been the center of our previous research on this topic,[24] concerns the validity of the prediction models’ outcome measure. Typically, the outcome measures in these studies did not rely on reliable suicide assessments of research participants,[42] but on suicide labeling of social media texts (eg, “life sucks, I want to die”). Such outcome measures may be of value, but they are bound to generate multiple false predictions, mainly because most users refrain from disclosing their suicidal struggles online in an explicit manner.[43,44]

The second gap concerns the model’s inputs. Most of the research relied on computational linguistics, which now involves large language models, such as GPT-3,[45] to analyze social media texts.[27] Images, in contrast, were barely investigated in this context despite their popularity today in our daily communications online (eg, Instagram).[46–49] To our knowledge, none of the existing studies in the field managed to predict clinically validated suicide risk based solely on social media images.

The third gap concerns the difficulty to interpret the results of AI models. The complexity of deep learning classifiers constitutes a sort of “black box”[50] that leaves their operators with little understanding of how these models performed their predictions. In addition, image processing typically relies on purely bottom-up strategies whereby the features are extracted from the data in a fully automated way (ie, the researcher does not approach the visual task with a priori assumptions or hypotheses). These features usually reflect nuanced patterns in the input data and are therefore typically very hard to interpret. Thus, even if the final predictions of AI suicide models are relatively accurate, clinicians may refrain from utilizing them in real-life settings and researchers may not be able to derive new theoretical insights based on their performances.

Research Goal

The goal of this study is to address these conceptual gaps through the construction of an interpretable AI prediction model of clinically valid suicide risk based solely, and for the first time, on social media images. Our hypothesis was that social media images would contain information about the aforementioned risk factors of emotions and interpersonal relationships and therefore be of value to suicide prediction.

To investigate this hypothesis, this study included a well-established measure of suicide risk. In addition, we implemented an interpretable and hybrid methodology that combined the common bottom-up, deep learning–based approach of the field with the more traditional, top-down, theory-driven approach. This integration of top-down strategies allowed us to develop a prediction model that relies on pre-defined features that can then be relatively easily interpreted and reconnected to suicide theories and therapies. In this way, the current study may enhance the discovery of new warning signs of suicide and the development of simple and relatively easy-to-use suicide monitoring tools.

METHODS

Data

The data of the current study were extracted from a larger dataset, which was collected by us during May through June 2018 for our previous research that focused on suicide prediction from textual postings.[24] Briefly, on completion of a consent form, participants recruited from Amazon MTurk crowdsourcing platform completed several, well-validated psychiatric and psychosocial assessment tools, including the gold-standard Columbia Suicide Severity Rating Scale (C-SSRS; see Supplementary Appendix 1).[51] In addition, the participants were asked to share their Facebook activity from the entire year that preceded the day they completed the questionnaires.

Specifically, for the purpose of this study, we gathered all the images that were uploaded by the participants to their Facebook accounts during this period and used them as inputs to an AI model that aims to predict the participant's scores on the C-SSRS. The chosen outcome measure from this scale (to be predicted by the images) was determined by the strict cutoff point for high suicide risk as suggested by the developers of the scale (Supplementary Appendix 1). This cutoff represents severe suicide ideation. After cleaning the data from bogus and poor-quality responses (Supplementary Appendix 1),[52] the final dataset included 841 high-quality respondents (83.4% female, mean age = 36.7 years) who uploaded together 177,220 images (mean [SD] = 124 [218.8]). Table 1 provides the descriptive statistics for the dataset.

Table 1.
Descriptive Statistics of the Cleansed Dataset (N = 841)

	High Risk for Suicide	Rest of the Sample
Participants, n	92	749
No. of images, mean (SD)	243.28 (225)	207.3 (218.3)
Age, mean (SD), y	32.71 (9.943)	37.16 (11.06)
Female %	79%	83.8%
Annual income, $	$44,347.25	$57,717.50

Extracting Interpretable Visual Features

The images were represented using the recent deep learning model of CLIP (Contrastive Language-Image Pre-training).[53] Briefly, CLIP is a multimodality deep neural network consisting of two components (encoders) that can represent images and texts as dense-numeric vectors. CLIP was trained in a bottom-up manner, to distinguish the right textual captions of the training images from tens of random captions sampled for each image (see Supplementary Appendix 1 for additional information about CLIP).

Importantly, in this study we used CLIP unconventionally, as a preliminary methodological step to extract visual features, which were predefined by us (see Supplementary Appendix 1 for the phrasing process of the requested features). Our purpose was to create a small set of basic visual features (eg, "a dark photo," "a photo of a person") by which we could represent the images. Notably, in contrast to other visual models, the features that were generated in the current study using CLIP are easily interpreted. They were formulated in advance by us in a simple language as straightforward sentences ("a photo of sad people"), and some of them were formulated specifically to reflect theory-driven components (ie, emotions and relationships) that can then be interpreted in the context of the available knowledge about suicide (see further explanation later in the Methods section).

To extract the features with CLIP, we defined 9 visual tasks (eg, determining the type of the relationships in the photo). For each visual task, we provided CLIP with a set of predefined complementing queries about visual properties that are related to the task. In return, CLIP assigned probability scores to these queries, whereby all the scores summed up to 1 (eg, "a photo of a family" = 0.1, "a photo of a couple" = 0.1, "a photo of friends" = 0.1, and "a photo of colleagues" = 0.7).

We clustered the 9 tasks into 3 clusters (Table 2). The first cluster consisted of general tasks that are relevant to all types of images and it targeted brightness, sentiment, and content. Brightness and sentiment consisted of two opposite queries (bright vs dark; positive vs negative), and content consisted of 5 queries measuring the presence of humans (person or people), animals, or other, non-living objects (images of text or inanimate objects).

Inspired by the attachment and interpersonal theories presented in the introduction, we conducted further analyses of images that were judged to contain human figures based on their CLIP probability scores. We hypothesized that such images could reveal valuable information about the emotions and relationships of the person who uploaded them. Since some images included only one human figure while others included more than one figure, we formulated two additional clusters of features (in addition to the previously described cluster that was relevant to all the images). The second cluster targeted images with one person and included the identity of the photographer (selfie/not selfie), the emotional state of the person (happy/sad), and his/her developmental stage (child, adult, elderly). The third cluster targeted images with people and included the identity of the photographer (selfie/not selfie), the emotional state of the people (happy/sad), and their relationships (romantic couples, families, friends, or work colleagues).

Altogether, the 3 clusters included 24 queries. For each image, we created a 24-dimensional feature-based vector representation according to the probability scores of the 24 queries (1A). To maintain a fixed number of features for all images, irrelevant queries received a probability score of zero. For example, queries about emotions and relationships received a probability score of zero when the images did not contain human figures. We then calculated an averaged 24-dimensional vector for each user based on her/his entire set of uploaded images. This user-level representation vector was subsequently served as an input to the suicide prediction model (logistic regression; Figure 1B)

Experimental Setup

The averaged representation vectors of the users described in the preceding text were entered as inputs (predictors) to a logistic regression machine learning model, which was trained to predict

Table 2.
CLIP Tasks and Queries[a]

Cluster	Task	Query	Probability Score		
			Image 1	Image 2	Image 3
Cluster 1: General visual features	Content	An image of one person	0.67	0.68	0.06
		An image of people	0.21	0.15	0.92
		An image of an animal	0.01	0.08	0.01
		An image of an object	0.01	0.04	0.01
		An image of text	0.10	0.05	0.01
	Brightness	A dark photo	0.79	0.10	0.03
		A bright photo	0.21	0.90	0.97
	Sentiment	An image of negative feeling	0.96	0.02	0.02
		An image of positive feeling	0.04	0.98	0.98
Cluster 2: Person characterization (relevant only for images of one person)	Photographer	The photo is a selfie	0.25	0.30	...
		The photo was taken by someone else	0.75	0.70	...
	Emotion	A photo of a sad person	0.99	0.05	...
		A photo of a happy person	0.01	0.95	...
	Development	A photo of a child	0.60	1.00	...
		A photo of an adult	0.11	0.00	...
		A photo of an old person	0.29	0.00	...
Cluster 3: People characterization (relevant only for images of people)	Photographer	The photo is a selfie	...	...	0.92
		The photo was taken by someone else	...	...	0.08
	Emotion	A photo of happy people	...	...	0.97
		A photo of sad people	...	...	0.03
	Relationship	A photo of a family	...	...	0.96
		A photo of friends	...	...	0.00
		A photo of colleagues	...	...	0.00
		A photo of a couple	...	...	0.04

Image 1

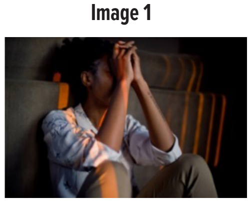


Image 2

Image 3

[a]The images were among those randomly selected from the Internet (from Pixabay.com) to test different phrasing alternatives for the queries until satisfactory results were obtained.
Abbreviation: CLIP = Contrastive Language-Image Pre-training.

cases of high suicide risk (Figure 1B). The training phase was conducted among a random subset of 70% of the data (n = 634), and the test phase was conducted among the rest of the sample (n = 207). To overcome potential biases of imbalanced datasets, we repeated this random split process 1,000 times.

The results of this process (described in the Results section) were then compared with two control baseline models; that is, alternative, common deep learning methods that could have been used to predict suicide risk from images. The first baseline was the commonly used yet uninterpretable computer-vision model of ResNet,[54] which was implemented in related works.[55,56] The second baseline comprised an ablative model that utilized only the image encoder component of CLIP, thus producing uninterpretable numerical representations of the images.* These comparisons allowed us to assess the contribution of our interpretable and hybrid strategy beyond the strength of the bottom-up backbone of CLIP itself and the common model of ResNet.

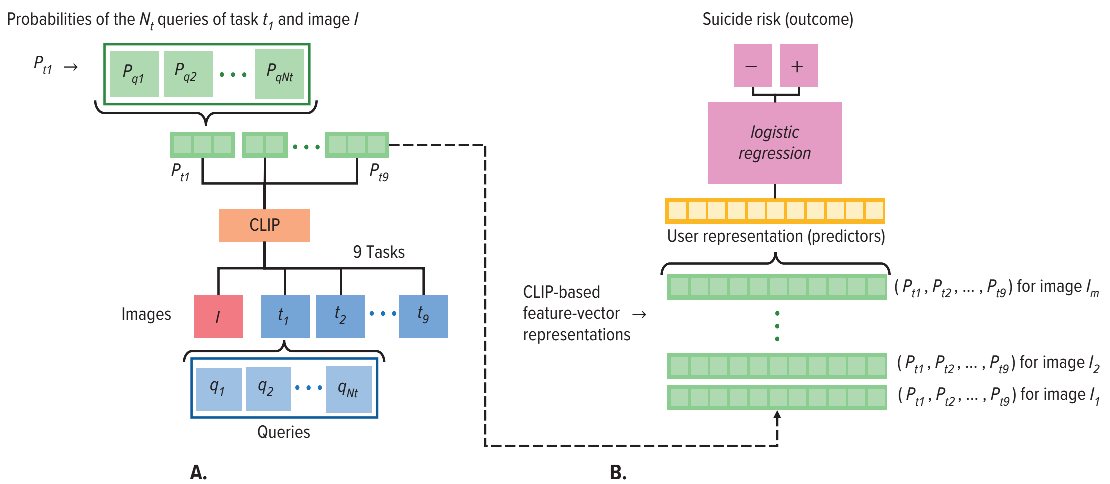
RESULTS

Table 3 presents the results of 4 models: the proposed hybrid model of the current study, the 2 deep learning baselines described in the previous section (ResNet and the ablative model), and a random baseline, which represents chance level results (ie, random scores of

*Note that in ablation analysis, the proposed model is compared to its variants in which most model components are kept fixed except for ≥1 component of interest.

Figure 1.

Illustration of the Extraction of the Features Using CLIP (Contrastive Language-Image Pre-training)[a]



[a]A task t was defined by its queries: $t=\{q_j\}_{j=1}^{N_t}$ for which N_t is the total number of queries in the task. (A) In each task, CLIP assigns a probability score vector to each query: $(p_1, \dots, p_{N_t}) = CLIP\ (I,t)$. The probabilities sum up to 1 and represent the degree of association between the query and the image. Overall, we defined 9 tasks containing together a total of 24 queries/features. (B) For each user, we averaged the 24-dimensional visual feature vectors of all her/his images and used the averaged vector as the input to the logistic regression model.

Table 3.

Main Results[a,b]

Model Type	AUC	95% CI	Cohen *d*	F1 Score	NPV	PPV (Precision)	Sensitivity (Recall)	Specificity	ECE
Random baseline	0.500	0.497–0.503		0.196	0.892	0.122	0.511	0.503	0.400
Bottom-up baseline with ResNet	0.623	0.621–0.625	0.44	0.275	0.923	0.203	0.508	0.711	0.045
Bottom-up baseline with CLIP	0.696	0.694–0.698	0.72	0.338	0.929	0.272	0.493	0.818	**0.038**
Hybrid model of this study	**0.720**	**0.716–0.724**	**0.82**	**0.363**	**0.934**	**0.295**	**0.515**	**0.829**	0.041

[a]All the scores represent the mean score obtained from 1,000 random splits of the data (see the Methods section).
[b]Boldface indicates the best-performing model according to the metric specified in the column.
Abbreviations: AUC = area under the ROC curve, CLIP = Contrastive Language-Image Pre-training, ECE = expected calibration error, F1 score = the harmonic mean of the precision and the recall scores of the positive class of high suicide risk, NPV = negative predictive value, PPV = positive predictive value, ROC = receiver operating characteristic.

suicide risk between 0 and 1). To evaluate the accuracy of the models, we used the standard measure of AUC.

In addition, we computed the following measures that require a specific decision threshold: negative predictive value (NPV), positive predictive value (PPV), precision, sensitivity (recall), specificity, and F1 scores (the harmonic mean of the precision and recall scores for the positive class of high suicide risk). The values of these measures reside in the [0, 1] interval, whereby high values indicate better prediction performances. The specific decision threshold for these measures was selected to maximize the F1 score when the sensitivity score ranges between 0.45 and 0.55. Expected calibration errors (ECEs) are presented as well to evaluate the extent to which the probability assigned by a model to the positive class (high suicide risk) indeed reflects the likelihood that a user reported such a high risk.[57] The values of the ECE reside in the [0, 1] interval as well, and high values indicate improved model's calibration.

Altogether, the results indicated that the proposed hybrid model of this study produced good prediction performances (AUC = 0.720; 95% CI, 0.716–0.724; Cohen $d = 0.82$) surpassing what would be expected by mere chance (random baseline). The hybrid model significantly outperformed the commonly used model of ResNet and the ablative CLIP-based baseline ($t = 44.3$, $P < .0001$ and $t = 11.4$, $P < .0001$, respectively). This pattern was evident also across the remaining evaluation measures that required a decision threshold. The ECE scores indicted that all the examined models were near perfectly calibrated.

The obtained AUC score of our hybrid model, which is comparable to those of previous successful language-based

Table 4.
Associations Between CLIP Features and High Suicide Risk

Visual Features (Queries)	High Suicide Risk, Mean (SD)	Rest of the Sample, Mean (SD)	t-Score	*P* Value	Logistic Regression Model β	SE	Wald χ2	*P* Value
Sentiment (negative)	0.42 (0.09)	0.34 (0.10)	6.99	.000	0.517	0.072	51.509	.000
Brightness (dark)	0.50 (0.15)	0.41 (0.18)	5.27	.000	0.353	0.1258	7.873	.005
Photographer—people (selfie)	0.33 (0.07)	0.29 (0.08)	3.97	.000	0.301	0.0435	47.775	.000
Emotional state—person (sad)	0.47 (0.10)	0.41 (0.11)	5.00	.000	0.230	0.0644	12.748	.000
Developmental stage—person (child)	0.56 (0.16)	0.49 (0.16)	3.99	.000	−0.161	0.0751	1.667	.197
Photographer—person (selfie)	0.66 (0.16)	0.58 (0.17)	4.45	.000	−0.157	0.0839	3.502	.061
Relationships (friends)	0.27 (0.09)	0.23 (0.08)	4.12	.000	0.140	0.0450	9.705	.000
Developmental stage—person (elderly)	0.40 (0.12)	0.34 (0.11)	4.58	.000	0.130	0.0451	8.238	.000
Emotional state—people (sad)	0.30 (0.18)	0.41 (0.24)	−5.10	.000	0.120	0.1280	0.884	.347
Relationships (family)	0.25 (0.09)	0.29 (0.10)	−3.55	.001	−0.036	0.0587	23.115	.000
Content (people)	0.25 (0.07)	0.27 (0.09)	−2.95	.004	−0.031	0.0498	48.815	.000

Abbreviations: CLIP = Contrastive Language-Image Pre-training, SE = standard error.

models,[31] provides a first proof that AI models can predict validated suicide risk purely from social media images.[20] The comparison with the bottom-up baselines suggests that the integrative strategy of this study is highly beneficial in such a complicated task of suicide prediction from images. Indeed, the ablative baseline produced relatively good predictions as well, but its results cannot be interpreted. In the same vein, future studies may achieve even better predictions using stronger deep learning classifiers; however, this might come at the cost of interpretability. The benefits of our choice to prioritize a logistic regression classifier over other, more sophisticated deep learning models, for simplicity and interpretability, are presented in the next paragraph (see also the Discussion).

Relationships Between CLIP Features and Suicide Risk

A further analysis was conducted to explore the associations between the 24 CLIP-based features and the risk of suicide in the entire sample (Supplementary Appendix 1). A t-score comparison indicated that 11 features were significantly different between the high suicide risk group (n = 92) and the rest of the sample (n = 749). A standard multiple logistic regression analysis in which these 11 features were entered as simultaneous predictors of high suicide risk pointed to 8 significant features (Table 4). These 8 features addressed the two hypothesized factors of this study: emotions and interpersonal relationships.

The linearity assumption of the regression alongside the straightforward nature of the features that were inserted into the model facilitate the interpretability of its results. High risk participants had higher scores than the rest of the participants in features indicative of negative emotions (eg, dark or negative images, sad person) and lower scores in features indicative of relationships and belongingness (eg, images of people or family). They also had increased scores in features that might be indicative of loneliness (selfie images and images with an elderly person). One feature appeared in the opposite direction from our hypothesis (“a photo of friends”), but overall, the theory-inspired features seemed to have contributed significantly to the successful prediction of suicide risk (Table 4).

DISCUSSION

Despite substantial advancements in suicide predictions following the social media and deep learning revolutions,[27–30] current efforts to develop useful prediction models suffer from major conceptual gaps[31] (see the Introduction). This study combined top-down and bottom-up strategies to address these gaps and construct an interpretable prediction model of valid suicide risk from Facebook images.

Image-Based Prediction of Validated Suicide Risk

The first contribution of the study concerns the very input and outcome measure of our prediction model. To our knowledge, this study is the first to demonstrate high-quality predictions of clinically validated suicide risk based solely on images. Most of the studies on this topic implemented inadequate outcome measures of suicide risk (eg, postings with suicide-related content),[42] which are bound to produce high rates of false predictions (see the Introduction). In addition, most of the studies relied heavily on lingual inputs (eg, Tweets), while visual inputs were rarely investigated.[27] Not only does the current focus on images fill a major gap in the literature, but it is also most timely as Internet communication becomes more and more visual-based and as image-based social networks, such as Instagram, become highly popular.[46–49]

Integrative Methodological Approach

The second contribution concerns the integrative approach of the study. Suicide prediction from images is a complex task. The classification task involves a relatively abstract and elusive outcome, which is not obtained from the visual content (ie, the suicide label is established via an external assessment tool), and standard, bottom-up AI tools may not suffice, as shown in the comparisons with the control baselines (Table 3). Not only does the proposed method of this study produce improved predictions, but future studies may also leverage its flexible nature to explore a variety of top-down hypotheses regarding potential warning signs of suicide.

Interpretability, Simplicity, and Theoretical Context

The third contribution concerns the interpretability and simplicity of the proposed model. In this study, we used CLIP to formulate a parsimonious number of features, which were predefined by us, using self-explained queries about basic visual characteristics as well as theory-driven elements (eg, "bright photo," "photo of sad people"). This is noteworthy considering that typical visual classification tasks are usually addressed by "black box" deep neural network models,[50] which are difficult to interpret.[31] In addition, our non-conventional usage of CLIP provides researchers and clinicians, who do not always have extensive computational backgrounds, with a practical and easy-to-implement prediction method, using everyday language instructions.[53] Moreover, our choice to insert these plain CLIP-generated features into a relatively simple logistic regression model also contributed to the interpretability of the model. The linearity assumption of the model produces easy-to-understand relationships (eg, the more images with sad people a participant uploads, the greater the suicide risk is) and allows the usage of standard feature-importance tools, which could not be applied to more sophisticated, non-linear deep learning models.

Specifically for this study, the further regression analysis that was implemented on the entire sample lent support to our hypothesis that social media images would contain valuable information about emotions and interpersonal relationships. Utilizing the linear nature of our analysis, we could conclude that participants with high suicide risk had increased levels of negative emotions and loneliness in their images and decreased levels of interpersonal relationships and belonginess.

This conclusion is noteworthy considering the centrality of these two psychosocial factors in the formation and maintenance of suicide risk.[1,33–35] As mentioned in the Introduction, the emotional state of the person and their interpersonal and social relationships are crucial components of mental health, as they relate to the elementary psychological concept of attachment formulated by John Bowlby and further characterized by Mary Ainsworth.[38] Unstable and insecure human relationships are intertwined with negative and painful emotions, and, together, they have a significant impact on the person's overall well-being. As such, they stand at the core of the interpersonal-psychological theory of suicidal behavior[36,37] as well as the evidence-based psychological interventions that are typically applied in the field—interpersonal psychotherapy (IPT)[39] and attachment-based family therapy (ABFT).[40,41]

Moreover, beyond the direct connection to principal ideas in suicidology, the specific themes that emerged in the regression analysis of this study may also shed light on specific suicide-related behaviors on social media. We already know from our previous studies that users at risk rarely post explicit suicidal manifestations online.[24,43] This study suggests that these users may find other, more subtle ways to express their emotional struggles using less troubling (but nevertheless valuable) signs, such as images indicative of negative emotions or loneliness. In this way, this study illustrates how deep learning–based research can also be used to explore top-down hypotheses about non-verbal warning signs that are crucial to suicide prevention. Unfortunately, some people, especially children and adolescents, die by suicide without obvious warning signs, such as explicit manifestations of suicide ideation or communications with health services.[58] We therefore hope that our findings would encourage further AI research that will aim to uncover additional subtle, non-verbal clues, both in real-life settings offline and on social media, thus improving our ability to understand and prevent suicide behaviors in real-life settings.[31]

Limitations

The current study has several limitations. First, despite the well-established measure of suicide risk and the rigid data-quality protocol that were implemented, the chosen ground truth criterion may still be subjected to biases and inaccuracies, as it relies solely on self-report responses of crowdsourcing participants from afar (ie, with no personal communication). Second, the strength of the current study, which focused on images only, may also serve as one of its weaknesses as suicide predictions may benefit from joint processing of both visual and lingual contents (eg, Facebook images and texts)[55,59] as well as other social media features (eg, likes and comments) and publicly available data (eg, available sociodemographic characteristics of users). Third, although we considered two key theory-driven risks from an influential theory of suicide and evidence-based treatments, our set of visual features was not exhaustive, as these sources include additional risk factors and as other suicide theories may also be of relevance to suicide prediction from images (see Supplementary Appendix 1 for more information). Finally, the ecological validity of the study is somewhat limited, as its MTurk-based sample may not fully represent the general population (eg, the final sample consisted of significantly more female than male participants).[52] Further research that will examine other social media (eg, Instagram) as well as diverse populations, including clinical populations (eg, hospitalized patients), is therefore crucially needed.

CONCLUSION

Without underestimating these limitations, we believe that future studies may build on the encouraging results of this study and its relatively simple methodologies (that overcome the typical "black box" problem of AI research) to keep developing the promising field of AI-suicide research. Further studies are encouraged to keep harnessing the abundant (non-verbal) information people upload to their social media accounts for suicide prevention. In this study, we presented an interpretable and flexible prediction model of validated suicide risk from Facebook images that was inspired by key interventions and theory in suicidology. It is therefore our hope that these characteristics will encourage researchers to utilize our hybrid approach to uncover new warning signs of suicide and, perhaps, develop effective, real-life monitoring tools that will eventually contribute to the global efforts to reduce suicide behaviors around the world.

Article Information

Published Online: November 29, 2023. https://doi.org/10.4088/JCP.23m14962

Submitted: June 7, 2023; accepted September 21, 2023.

To Cite: Badian Y, Ophir Y, Tikochinski R, et al. Social media images can predict suicide risk using interpretable large language-vision models. *J Clin Psychiatry.* 2024;85(1):23m14962

Author Affiliations: Technion—Israel Institute of Technology, Faculty of Data and Decision Sciences, Haifa, Israel (Badian, Ophir, Tikochinski, Calderon, Reichart); University of Cambridge, Centre for Human Inspired Artificial Intelligence, Cambridge, UK (Ophir); Reichman University, Baruch Ivcher School of Psychology, Herzliya, Israel (Klomek); Technion—Israel Institute of Technology, Rappaport Faculty of Medicine, Haifa, Israel (Fruchter).

Corresponding Author: Yaakov Ophir, PhD, Technion—Israel Institute of Technology, Technion City, Haifa 3200003 (yaakovophir@gmail.com).

Mrs Badian and Dr Ophir contributed equally to the article.

Author Contributions: Conceptualization: Badian, Calderson, Ophir, Reichart. Consultation: Klomek, Fruchter. Methodology: Badian, Calderon, Tikochinski. Investigation: Badian, Calderon, Tikochinski. Visualization: Tikochinski and Badian. Funding acquisition: Reichart and Ophir. Supervision: Reichart. Writing: Ophir. Review: Reichart, Klomek, Fruchter. Editing: Ophir, Reichart, Klomek.

Relevant Financial Relationships: The authors declare no conflict of interests.

Funding/Support: A grant for data science by VATAT, the Israeli higher education council was used to partially fund the scholarships of Yael Badian and Nitay Calderon.

Role of the Funders/Sponsors: The funding agent (VATAT) had no role in the conduct or publication of the study.

Ethical Approval Statement: The procedures of the study comply with the ethical standards of the Helsinki Declaration of 1975, as revised in 2008. All procedures were approved by the Ethics for Research on Human Subjects Committees of the Technion—Israel Institute of Technology.

Data and Materials Availability: The social media materials of the study are identifiable and therefore cannot be shared.

Supplementary Material: Available at Psychiatrist.com.

References

1. Turecki G, Brent DA, Gunnell D, et al. Suicide and suicide risk. *Nat Rev Dis Primers*. 2019;5(1):74.
2. Hedegaard H, Curtin SC, Warner M. *Suicide Rates in the United States Continue to Increase*. Hyattsville, MD: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics; 2018.
3. Moutier C. Suicide prevention in the COVID-19 era: transforming threat into opportunity. *JAMA Psychiatry*. Published online October 16, 2020.
4. Klomek AB. Suicide prevention during the COVID-19 outbreak. *Lancet Psychiatry*. 2020;7(5):390.
5. Banerjee D, Kosagisharaf JR, Sathyanarayana Rao TS. 'The dual pandemic' of suicide and COVID-19: a biopsychosocial narrative of risks and prevention. *Psychiatry Res*. 2021;295:113577.
6. Gunnell D, Appleby L, Arensman E, et al; COVID-19 Suicide Prevention Research Collaboration. Suicide risk and prevention during the COVID-19 pandemic. *Lancet Psychiatry*. 2020;7(6):468–471.
7. Samji H, Wu J, Ladak A, et al. Review: mental health impacts of the COVID-19 pandemic on children and youth–a systematic review. *Child Adolesc Ment Health*. 2022;27(2):173–189.
8. Zaninotto P, Iob E, Demakakos P, et al. Immediate and longer-term changes in the mental health and well-being of older adults in england during the COVID-19 pandemic. *JAMA Psychiatry*. 2022;79(2):151–159.
9. Vizheh M, Qorbani M, Arzaghi SM, et al. The mental health of healthcare workers in the COVID-19 pandemic: a systematic review. *J Diabetes Metab Disord*. 2020;19(2):1967–1978.
10. Yao H, Chen J-H, Xu Y-F. Patients with mental health disorders in the COVID-19 epidemic. *Lancet Psychiatry*. 2020;7(4):e21.
11. Li LZ, Wang S. Prevalence and predictors of general psychiatric disorders and loneliness during COVID-19 in the United Kingdom. *Psychiatry Res*. 2020;291:113267.
12. Ren X, Huang W, Pan H, et al. Mental health during the Covid-19 outbreak in China: a meta-analysis. *Psychiatr Q*. 2020;91(4):1033–1045.
13. Taquet M, Geddes JR, Husain M, et al. 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. *Lancet Psychiatry*. 2021;8(5):416–427.
14. Ellis WE, Dumas TM, Forbes LM. Physically isolated but socially connected: psychological adjustment and stress among adolescents during the initial COVID-19 crisis. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement.* 2020;52(3):177–187.
15. Murata S, Rezeppa T, Thoma B, et al. The psychiatric sequelae of the COVID-19 pandemic in adolescents, adults, and health care workers. *Depress Anxiety*. 2021;38(2):233–246.
16. Zhang L, Zhang D, Fang J, et al. Assessment of mental health of Chinese primary school students before and after school closing and opening during the COVID-19 pandemic. *JAMA Netw Open*. 2020;3(9):e2021482.
17. Sakamoto H, Ishikane M, Ghaznavi C, et al. Assessment of suicide in Japan during the COVID-19 pandemic vs previous years. *JAMA Netw Open*. 2021;4(2):e2037378.
18. Dubé JP, Smith MM, Sherry SB, et al. Suicide behaviors during the COVID-19 pandemic: a meta-analysis of 54 studies. *Psychiatry Res*. 2021;301:113998.
19. Bruffaerts R, Demyttenaere K, Hwang I, et al. Treatment of suicidal people around the world. *Br J Psychiatry*. 2011;199(1):64–70.
20. Franklin JC, Ribeiro JD, Fox KR, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull*. 2017;143(2):187–232.
21. Sejnowski TJ. *The deep learning revolution*. Mit Press; 2018.
22. Roy A, Nikolitch K, McGinn R, et al. A machine learning approach predicts future risk to suicidal ideation from social media data. *NPJ Digit Med*. 2020;3(1):78.
23. Coppersmith G, Leary R, Crutchley P, et al. Natural language processing of social media as screening for suicide risk. *Biomed Inform Insights*. 2018;10:1178222618792860.
24. Ophir Y, Tikochinski R, Asterhan CSC, et al. Deep neural networks detect suicide risk from textual Facebook posts. *Sci Rep*. 2020;10(1):16685.
25. Zirikly A, Resnik P, Uzuner O, et al. CLPsych 2019 shared task: predicting the degree of suicide risk in Reddit posts. *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*. 2019 2019:24-33.
26. Zheng L, Wang O, Hao S, et al. Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records. *Transl Psychiatry*. 2020;10(1):72.
27. Bernert RA, Hilberg AM, Melia R, et al. Artificial intelligence and suicide prevention: a systematic review of machine learning investigations. *Int J Environ Res Public Health*. 2020;17(16):5929.
28. Burke TA, Ammerman BA, Jacobucci R. The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: a systematic review. *J Affect Disord*. 2019;245:869–884.
29. Resnik P, Foreman A, Kuchuk M, et al. Naturally occurring language as a source of evidence in suicide prevention. *Suicide Life Threat Behav*. 2021;51(1):88–96.
30. Ji S, Pan S, Li X, et al. Suicidal ideation detection: a review of machine learning

methods and applications. *IEEE Trans Comput Soc Syst*. 2021;8(1)214–226.
31. Ophir Y, Tikochinski R, Brunstein Klomek A, et al. The Hitchhiker's Guide to computational linguistics in suicide prevention. *Clin Psychol Sci*. 2022;10(2):212–235.
32. Ophir Y, Rosenberg H, Lipshits-Braziler Y, et al. *"Digital Adolescence": The Effects of Smartphones and Social Networking Technologies on Adolescents' Well-Being. Online Peer Engagement in Adolescence*. Routledge; 2020:122–139.
33. Stravynski A, Boyer R. Loneliness in relation to suicide ideation and parasuicide: a population-wide study. *Suicide Life Threat Behav*. 2001;31(1):32–40.
34. Bennardi M, Caballero FF, Miret M, et al. Longitudinal relationships between positive affect, loneliness, and suicide ideation: age-specific factors in a general population. *Suicide Life Threat Behav*. 2019;49(1):90–103.
35. Rogers ML, Joiner TE. Rumination, suicidal ideation, and suicide attempts: a meta-analytic review. *Rev Gen Psychol*. 2017;21(2):132–142.
36. Ribeiro JD, Joiner TE. The interpersonal-psychological theory of suicidal behavior: current status and future directions. *J Clin Psychol*. 2009;65(12):1291–1299.
37. Joiner T. *Why People Die by Suicide*. Harvard University Press; 2009.
38. Bretherton I. *The Origins of Attachment Theory: John Bowlby and Mary Ainsworth. Attachment Theory*. Routledge; 2013:45–84.
39. Markowitz JC, Weissman MM. Interpersonal psychotherapy: principles and applications. *World Psychiatry*. 2004;3(3):136–139.
40. Ewing ESK, Diamond G, Levy S. Attachment-based family therapy for depressed and suicidal adolescents: theory, clinical model and empirical support. *Attach Hum Dev*. 2015;17(2):136–156.
41. Diamond G, Diamond GM, Levy S. Attachment-based family therapy: theory, clinical model, outcomes, and process research. *J Affect Disord*. 2021;294:286–295.
42. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ Digit Med*. 2020;3(1):43.
43. Ophir Y, Asterhan CSC, Schwarz BB. The digital footprints of adolescent depression, social rejection and victimization of bullying on Facebook. *Comput Human Behav*. 2019;91:62–71.
44. Ophir Y. SOS on SNS: adolescent distress on social network sites. *Comput Human Behav*. 2017;68:51–55.
45. Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*. 2020.
46. Lee E, Lee J-A, Moon JH, et al. Pictures speak louder than words: motivations for using instagram. *Cyberpsychol Behav Soc Netw*. 2015;18(9):552–556.
47. Herring S, Dainas A. *"Nice Picture Comment!" Graphicons in Facebook Comment Threads*. 2017.
48. Tang Y, Hew KF. Emoticon, emoji, and sticker use in computer-mediated communication: a review of theories and research findings. *Int J Commun*. 2019;13:27.
49. Bai Q, Dan Q, Mu Z, et al. A systematic review of emoji: current research and future perspectives. systematic review. *Front Psychol*. 2019;10:2221.
50. Alishahi A, Chrupała G, Linzen T. Analyzing and interpreting neural networks for NLP: a report on the first BlackboxNLP workshop. *Nat Lang Eng*. 2019;25(4):543–557.
51. Posner K, Brown GK, Stanley B, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *Am J Psychiatry*. 2011;168(12):1266–1277.
52. Ophir Y, Sisso I, Asterhan CSC, et al. The turker blues: hidden factors behind increased depression rates among Amazon's Mechanical Turkers. *Clin Psychol Sci*. 2020;8(1):65–83.
53. Radford A, Kim JW, Hallacy C, et al. *Learning Transferable Visual Models From Natural Language Supervision*. PMLR; 2021:8748–8763.
54. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016:770–778.
55. Ramírez-Cifuentes D, Freire A, Baeza-Yates R, et al. Detection of suicidal ideation on social media: multimodal, relational, and behavioral analysis. *J Med Internet Res*. 2020;22(7):e17758.
56. Ma Y, Cao Y. *Dual Attention based Suicide Risk Detection on Social Media*. IEEE; 2020:637–640.
57. Huang Y, Li W, Macheret F, et al. A tutorial on calibration measurements and calibration models for clinical prediction models. *J Am Med Inform Assoc*. 2020;27(4):621–633.
58. Rodway C, Tham S-G, Turnbull P, et al. Suicide in children and young people: can it happen without warning? *J Affect Disord*. 2020;275:307–310.
59. Ribeiro JD, Huang X, Fox KR, et al. Predicting imminent suicidal thoughts and nonfatal attempts: the role of complexity. *Clin Psychol Sci*. 2019;7(5):941–957.

The Journal of Clinical Psychiatry

Supplementary Material

Article Title: Social Media Images Can Predict Suicide Risk Using Interpretable Large Language-Vision Models

Author(s): Yael Badian, MSc; Yaakov Ophir, PhD; Refael Tikochinski, MA; Nitay Calderon, MSc; Anat Brunstein Klomek, PhD; Eyal Fruchter, MD; and Roi Reichart, PhD

DOI Number: https://doi.org/10.4088/JCP.23m14962

LIST OF SUPPLEMENTARY MATERIAL FOR THE ARTICLE

DISCLAIMER

This Supplementary Material has been provided by the author(s) as an enhancement to the published article. It has been approved by peer review; however, it has undergone neither editing nor formatting by in-house editorial staff. The material is presented in the manner supplied by the author.

Social Media Images Can Predict Suicide Risk Using Interpretable Large Language-Vision Models

Supplementary Appendix 1

The suicide risk outcome measure of the study

The participants' risk of suicide was assessed with the well-established and well-researched CSSRS – the Columbia Suicide Severity Rating Scale.[1] The CSSRS has high predictive validity of suicide risk[2,3] and it consists of 6 categorical (yes/no) items. The first two items measure the very existence of a suicide risk, that is the risk that the person is experiencing any level of suicidal thoughts, whether these thoughts are concrete and highly dangerous, or 'just' passive and abstract death wishes. The remaining four items measure the severity of this general risk, and they are shown to the respondents only if the first two items indicated that they are at a (general) suicide risk. These items address concrete ideation to engage in active suicide behaviors, such as when the person reports of having a specific method or a plan to act on their suicidal thoughts. Notably, a positive answer to one or more of these four items indicates that the person is at a relatively high risk of suicide. In the current study, we therefore used this stricter cut-off point for a *high suicide risk* as our primary outcome (to be predicted by Facebook images).

The sample of the study

Of the initial sample of 2,685 MTurk users, 462 participants did not provide a working Facebook ID, 102 participants did not upload images to their timeline, and 341 participants failed implanted quality checks we developed to detect inattentive and bogus crowdsourcing respondents.[4] We also removed users who uploaded a relatively small number of images to their Facebook account (i.e., users who had less than 39 images – the median number of images in the sample) to ensure that our further computational analyses will be based on a substantial amount of visual data for each participant. The final and cleansed dataset included 841 high-quality respondents (83.4% female, average age = 36.7) who uploaded together 177,220 images (*M* =

124, *SD* = 218.8). Corresponding with previous studies that documented increased levels of mental health issues on MTurk (e.g., [4-6]), relatively high proportions of the current sample were classified as 'participants at high suicide risk' (10.93%).

Table 1 (in the body of the article) provides the descriptive statistics of the dataset. Complimentary description and statistical analyses of the entire sample of high-quality respondents who uploaded at least one accessible image (*N* = 1697) are available by the authors upon request. For further, detailed information about the complete dataset, see in Ophir, Tikochinski, et al., 2020.[7]

The vison-language model of CLIP

The Facebook images were represented using the recently developed deep learning model of CLIP (Contrastive Language-Image Pre-training; see the Supplementary Material).[8] CLIP is a multi-modality deep neural network consisting of two components (encoders) that can represent images and texts as dense-numeric vectors. CLIP was trained in a bottom-up manner, to match the right textual captions with their corresponding images using tens of randomly sampled options. The developers of CLIP collected 400 million pairs of images and texts from various sources on the internet. To ensure a large variety of visual concepts, each text sample had to include one word from a set of 500K queries. This set of queries consisted of all the words that occurred at least 100 times in the English version of Wikipedia.

In practice, CLIP uses representation vectors to evaluate the similarity between images and texts and assigns probabilities to each candidate caption based on its similarity to the image. It then selects the caption that achieved the highest probability score as the correct caption of the given image. This training allows CLIP to be used for various sub-tasks, such as extracting visual features from an image. For example, to detect whether an image is bright or dark, researchers can provide CLIP with the image and a set of captions (queries) – "a bright image" and "a dark image". CLIP then assigns probabilities to each one of the queries (e.g., "a bright image" = 0.7 and "a dark image" = 0.3). Based on these probabilities, which sum up to 1, the researchers can determine which one of the queries is most likely to be correct for this image ("a bright image").

Supplemented information about the extraction of the interpretable visual features

As explained in the main article, this study used CLIP in an unconventional way as a preliminary methodological step to extract visual features, which were predefined by us, in advance, in a top-down manner (for details about the extraction process, see the Method section). This is in contrast to common uses of CLIP, as CLIP is typically utilized for solving end-to-end tasks, such as object detection or segmentation.[9]

It should be noted here that the exact verbal phrasing of the queries affects the probability scores generated by CLIP. For example, the score of the query "a bright image" can differ from the score of "the image is bright". To ensure that the chosen queries of the current study were well phrased, we randomly selected 10 images from the Internet (i.e., not from our dataset, see examples in Table 2), and applied CLIP to test different phrasing alternatives until we received satisfactory results. We conducted this fine-tuning phrasing of the queries on external images from the Internet to overcome the potential problem of overfitting, which might have occurred if we were to conduct it on original images from our dataset.

Supplemented information about the extraction of the theory-driven features

Aside from the key theory-driven features discussed in the main article, we reviewed published lists of risk factors of suicide by leading health establishments, such as the World Health Organization (WHO), the Centers for Disease Control and Prevention (CDC), and the National Institute of Mental Health (NIMH),[10] and searched for additional risks that might be evident in social media images. For each risk factor (e.g., prior suicide attempts, drug abuse, and psychiatric diagnoses), we phrased matching visual queries (e.g., the person in the image abuses drugs), but CLIP could not perform well with these queries (nor could we, as human experts), probably because the images did not contain such blatant risk factors. The only theory-driven features we could extract from these social media images targeted emotions and relationships, as hypothesized by the interpersonal-psychological theory of suicidal behavior,[11,12] as well as the evidence-based treatments – the interpersonal psychotherapy (IPT)[13] and the attachment-based family therapy (ABFT).[14,15] Further studies that will find ways to consider stronger theory-driven risks as potential predictors are therefore encouraged, as they might achieve even better results than the obtained prediction scores of this study.

The prediction performance measure (AUC scores)

To evaluate the prediction performance of the various models of the study, we used the standard measure of AUC – the Area Under the Receiver Operating Characteristic (ROC) curve. The AUC measure is most appropriate for such a class-imbalanced dataset,[16] since it provides a single holistic value that reflects the relations between correct predictions of suicide (true-positive) and incorrect predictions of suicide (false-positive) at all potential classification thresholds.

Supplemented information about the t-test comparisons of the visual features

As mentioned in the Results section, a further analysis was conducted to explore the associations between the 24 CLIP-based features and the risk of suicide in the entire sample. The first step of this analysis included a t-test comparison of the mean probability scores of the visual features between the high suicide risk group (N = 92) and the rest of the sample (N = 749), using an FDR correction for multiple tests.[17] This procedure yielded 17 significant features. However, 6 of these 17 features had to be removed from the final table of differences (Table 4 of the main article) because 6 CLIP tasks involved only two (opposite) queries that sum up to the probability of one, thus creating redundant duplicates (e.g., the t scores of ‘happy people’ were the same as of ‘sad people’).

References

1. Posner K, Brown GK, Stanley B, et al. The Columbia–Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *American Journal of Psychiatry*. 2011;168(12):1266-1277.

2. Drapeau CW, Nadorff MR, McCall WV, Titus CE, Barclay N, Payne A. Screening for suicide risk in adult sleep patients. *Sleep Medicine Reviews*. 2019/08/01/ 2019;46:17-26. doi:https://doi.org/10.1016/j.smrv.2019.03.009

3. Weber AN, Michail M, Thompson A, Fiedorowicz JG. Psychiatric emergencies: assessing and managing suicidal ideation. *Medical Clinics*. 2017;101(3):553-571.

4. Ophir Y, Sisso I, Asterhan CSC, Tikochinski R, Reichart R. The turker blues: Hidden factors behind increased depression rates among Amazon's Mechanical Turkers. *Clinical Psychological Science*. 2020;8(1):65-83.

5. Arditte KA, Çek D, Shaw AM, Timpano KR. The importance of assessing clinical phenomena in Mechanical Turk research. *Psychological Assessment*. 2016;28(6):684.

6. McCredie MN, Morey LC. Who Are the Turkers? A Characterization of MTurk Workers Using the Personality Assessment Inventory. *Assessment*. 2018:1073191118760709.

7. Ophir Y, Tikochinski R, Asterhan CSC, Sisso I, Reichart R. Deep neural networks detect suicide risk from textual facebook posts. *Scientific Reports*. 2020/10/07 2020;10(1):16685. doi:10.1038/s41598-020-73917-0

8. Radford A, Kim JW, Hallacy C, et al. Learning transferable visual models from natural language supervision. PMLR; 2021:8748-8763.

9. Khan S, Naseer M, Hayat M, Zamir SW, Khan FS, Shah M. Transformers in vision: A survey. *ACM computing surveys (CSUR)*. 2022;54(10s):1-41.

10. Franklin JC, Ribeiro JD, Fox KR, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychological bulletin*. 2017;143(2):187.

11. Ribeiro JD, Joiner TE. The interpersonal-psychological theory of suicidal behavior: Current status and future directions. *Journal of clinical psychology*. 2009;65(12):1291-1299.

12. Joiner T. *Why people die by suicide*. Harvard University Press; 2009.

13. Markowitz JC, Weissman MM. Interpersonal psychotherapy: principles and applications. *World Psychiatry*. 2004;3(3):136.

14. Ewing ESK, Diamond G, Levy S. Attachment-based family therapy for depressed and suicidal adolescents: theory, clinical model and empirical support. *Attachment & human development*. 2015;17(2):136-156.

15. Diamond G, Diamond GM, Levy S. Attachment-based family therapy: Theory, clinical model, outcomes, and process research. *Journal of Affective Disorders*. 2021/11/01/ 2021;294:286-295. doi:https://doi.org/10.1016/j.jad.2021.07.005

16. Jeni LA, Cohn JF, De La Torre F. Facing imbalanced data--recommendations for the use of performance metrics. IEEE; 2013:245-251.

17. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*. 1995;57(1):289-300.