Check for updates

Predicting suicide attempts among Norwegian adolescents without using suicide-related items: a machine learning approach

E. F. Haghish[1]*, Nikolai O. Czajkowski[1,2] and Tilmann von Soest[1,3]

[1]Department of Psychology, Faculty of Social Sciences, University of Oslo, Oslo, Norway, [2]Department of Mental Disorders, Division of Mental and Physical Health, Norwegian Institute of Public Health (NIPH), Oslo, Norway, [3]Norwegian Social Research (NOVA), Oslo Metropolitan University, Oslo, Norway

Introduction: Research on the classification models of suicide attempts has predominantly depended on the collection of sensitive data related to suicide. Gathering this type of information at the population level can be challenging, especially when it pertains to adolescents. We addressed two main objectives: (1) the feasibility of classifying adolescents at high risk of attempting suicide without relying on specific suicide-related survey items such as history of suicide attempts, suicide plan, or suicide ideation, and (2) identifying the most important predictors of suicide attempts among adolescents.

Methods: Nationwide survey data from 173,664 Norwegian adolescents (ages 13–18) were utilized to train a binary classification model, using 169 questionnaire items. The Extreme Gradient Boosting (XGBoost) algorithm was fine-tuned to classify adolescent suicide attempts, and the most important predictors were identified.

Results: XGBoost achieved a sensitivity of 77% with a specificity of 90%, and an AUC of 92.1% and an AUPRC of 47.1%. A coherent set of predictors in the domains of internalizing problems, substance use, interpersonal relationships, and victimization were pinpointed as the most important items related to recent suicide attempts.

Conclusion: This study underscores the potential of machine learning for screening adolescent suicide attempts on a population scale without requiring sensitive suicide-related survey items. Future research investigating the etiology of suicidal behavior may direct particular attention to internalizing problems, interpersonal relationships, victimization, and substance use.

KEYWORDS

suicide attempt classification, risk factors, supervised machine learning, adolescents, survey

Introduction

Suicide attempt is defined as engaging in self-harm behavior with at least some intention to die (1) and its prevalence during adolescence is estimated to be up to 10% with a peak around the age of 16 (2). Cross-national studies also emphasize the high prevalence of suicide attempts during adolescence and early adulthood (1, 3). Thus, identifying adolescents at risk of attempting

suicide is a key step toward offering early interventions and mitigating its substantial public health impact. Yet, identifying at-risk adolescents is a challenging task, because suicide attempt risk is tied to a variety of individual, societal, cultural, and environmental factors (4) and despite decades of extensive research, the predictive power of identified risk factors remains weak and inconsistent (5). A systematic meta-analysis on clinical instruments revealed a low combined positive predictive value of 5.5% for suicidal behavior, concluding that the "high-risk" classification was not clinically useful (6). Adolescents' suicide attempts can also be unpremeditated or compulsive, making risk estimation even more difficult (7–9). As a result, recent reviews noted that prediction of suicidal behavior needs to account for the complex interplay of a multitude of risk factors, shifting the focus from risk factors to risk algorithms (5, 10).

Not only is the prediction of suicide attempt risk a challenging endeavor, but the sensitive subject matter also poses difficulties for assessment in the general population (11–13). Therefore, suicidal behavior is often not measured in population-based studies (12, 14, 15). Research has shown that prevalence of self-reported suicidal behavior in anonymous surveys is up to three times higher than non-anonymous surveys, further stressing sensitivity of suicide-related items (16, 17). Therefore, recent studies have stressed the importance of identifying individuals at risk for suicide attempts when suicide-related items, such as suicidal ideating and self-harm behavior are not assessed (18).

Recently, the potential of machine learning techniques to improve suicide attempt classification has been explored. Machine learning algorithms can make use of a multitude of variables and evaluate their complex interplay, resulting in promising improvements in the field of suicide research (19–21). However, these studies have focused on variables directly related to suicide, such as suicidal ideation and self-harm behaviors, and consistently identified them as the strongest predictors (19, 22–26). Moreover, there seems to be insufficient focus on the practicality of implementing machine learning algorithms at a population level, especially when suicide-related data are lacking. Although machine learning is regarded as a promising tool for predicting suicide attempts, a strong focus on suicide-related data may serve as a barrier to its implementation at scale. As a result, less is known about the performance of machine learning models in predicting suicide attempts when suicide-related items are not included (18).

The current study

In this study, we aim to answer two research questions. First, how accurately can machine learning models identify adolescents who self-report a recent suicide attempt without using sensitive suicide-related information such as self-harm, suicide plan, suicide ideation, or history of previous suicide attempts? Second, what are the most important personal, psychological, societal, and environmental variables associated with adolescents' recent suicide attempts? As noted in the literature, the risk factors of suicide attempts are usually studied in isolation and it is not clear what the most important predictors are, once hundreds of potential predictors are being taken into consideration (5, 10). Due to the low prevalence of suicidal behavior, richness and size of data is central to studying suicidal behavior with machine learning, both for classifying suicide attempt accurately as well as evaluating the predictors' importance; however, previous studies of machine learning frequently suffer from either

small sample sizes or using non-representative samples (23–29). We use nationally representative retrospective cross-sectional survey data from 173,664 Norwegian adolescents and analyze a multitude of items from a variety of psychological, sociological, and environmental domains to address our research questions.

Methods

Sample and procedure

This study utilizes the Ungdata surveys, a data collection scheme developed to administer youth surveys at municipal levels in Norway, assessing junior and senior high school students (grades 8–13, ages 13–18). The data were collected between 2014 and 2019 across most of municipalities in Norway, with students being invited to complete an electronic questionnaire in their classrooms. The current data analysis was performed on a subset of dataset, which included 169 items that were received by all participants ($n = 173,664$).

Measures

The analysis included 169 survey items. The full description and response options of these items can be found on the OSF repository of the study via https://osf.io/2qfnc/.

Recent suicide attempt

The outcome variable was recent suicide attempts, which was measured with a single self-report item, asking "*Have you tried to take your own life in the last 12 months?*," with the response options "Yes" and "No."

Socio-demographics

In addition to participants' gender and age, various socio-demographic factors were evaluated. These included socio-economic status, assessed using the second edition of the Family Affluence Scale (30, 31), parents' education level, adolescents' subjective assessment of their financial status in recent years, and the quality of their local environment. Participants also responded to items about future expectations such as higher education, career, and life quality, reflecting on their aspirations.

Interpersonal relationships

Interpersonal relationships were assessed through a comprehensive set of items inquiring about sexual and romantic relationships, relationships with parents, relationships with teachers and peers, and activities on social media. In addition, adolescents were asked to report online friendships and their parents' supervision of their social media activities online.

Somatic health, mental health, and victimization

Several items measured somatic health, physical pain, and hospitalizations, along with other health aspects such as exercising habits and dietary practices. For example, participants reflected on their consumption of vegetables, fruits, junk food, and energy drinks, adherence to strict diets, and concerns about body weight. Several mental health instruments were also administered, including the measurement of depressive and anxiety symptoms using a short

version of the Hopkins Symptom Checklist (32). Self-esteem was assessed with items from the Self-Perception Profile for Adolescents (33), and loneliness was assessed using the short form of the revised UCLA Loneliness Scale (34). In addition to somatic and mental health evaluations, participants' exposure to victimization, including bullying and physical, verbal, or sexual victimization in settings such as family, school, or social media, was assessed (35).

Conduct problems and substance use

Conduct problems were assessed through a series of questions drawn from Olweus' scale of antisocial behavior (36) and the National Longitudinal Youth Survey (37). Participants were also queried about their use of substances such as alcohol, tobacco, snus (an oral tobacco product), cannabis, and other illegal substances. Additionally, the alcohol consumption of parents and friends was assessed, along with the respondents' perception of alcohol consumption norms within their family.

Other measures

Various other aspects of adolescents' lives were also assessed, including leisure activities, lifestyle, online and offline activities, participation in different youth clubs and organizations, individual values, and political attitudes.

Statistical analysis

The data were prepared and analyzed using R statistical software version 4.1 (38) and all the machine learning models were built using the h2o R package version 3.30.0.6 (39).

Procedure

To predict recent suicide attempts among adolescents, the Extreme Gradient Boosting (XGBoost; (40)) algorithm was examined, which is based on boosting decision-tree techniques. The algorithm was trained using 80% of total data, reserving 20% of the sample for testing. Instead of a random split, stratified random splitting was employed from the splitTools R package (41) to ensure the prevalence of suicide attempts remained consistent in both the training and testing datasets.

Class imbalance

The prevalence of suicide in the population is low, creating a severe imbalance between the classes (suicidal vs. non-suicidal), which can bias both the model and the measures used to evaluate its performance in favor of the majority class (42–44). A standard solution is to artificially balance the training dataset by either up-sampling the minority class or down-sampling the majority class. However, modifying the underlying distribution of the outcome breaches the principal machine learning assumption that the training and testing datasets are sampled from the same population (45). To resolve this issue, previous machine learning studies on suicide attempt prediction have often balanced the testing dataset as well [see for example, (26, 29)], rendering the model inapplicable to the real-world problem. Simply put, when the distributions of both the training and testing datasets differ from the underlying population distribution, the performance of the model on a real-world dataset with severe class imbalance would be unknown (45). Another approach suggests preserving the testing dataset to represent the real-world distribution (46) and calibrating the trained

model's probabilities before evaluating its performance on the testing dataset (45, 47, 48). This method addresses the imbalance problem while keeping the model pertinent to the actual outcome distribution. In accordance with this strategy, we balanced the training dataset by up-sampling the minority class through bootstrapping, calibrated the models' probabilities using monotonic transformation, provided by the h2o R package, and then assessed the performance of the selected models on the testing dataset.

Fine-tuning and model evaluation

To fine-tune the models, random search was employed (49) and early stopping strategies were implemented during the search to mitigate overfitting (50). Each model was assessed through 10-fold cross-validation and optimized and ranked according to the Area Under the Precision-Recall Curve (AUPRC). Traditionally, the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve is used to optimize and evaluate binary classifiers (51). Under severe class imbalance, however, AUC can be overly optimistic, making AUPRC the preferred measure for assessing the model's performance (52). Yet, to make our results comparable to other suicide prediction studies, we continue to report AUC alongside AUPRC. When the random search process stopped, the model with the highest AUPRC (on the training dataset) was selected and further examined with the unseen testing dataset, and metrics such as AUC, AUPRC, sensitivity, and specificity were computed and reported.

Predictor importance

To evaluate the predictors' importance, we employed SHapley Additive exPlanations [SHAP; (53)] method, which is inspired by Shapley values in cooperative game theory (54, 55). Initially, we will present a SHAP summary plot to depict the contribution of each variable across all subjects (rows in the dataset) within the test dataset. This procedure differs from determining predictor importance based on the gain of the loss function during the construction of decision trees. The advantage of this method is that it offers detailed insights into the importance of each variable throughout its entire range of values. Hence, SHAP also enhances the transparency of the model by illustrating how different values of each item influence the model's decision-making. We identify the top 15 predictors from a total of 168 items included in the model, ranking them according to the normalized mean absolute SHAP value in descending order [see (56, 57)]. By normalizing the mean absolute SHAP value, ranging from 0 to 1, we can compare the relative importance of the top predictors. It should be noted that the term "predictor" may be contentious, as we are utilizing retrospective cross-sectional data. In this context, we use "predictor" solely to refer to a survey item that provides the model with unique information, facilitating the evaluation of suicide attempt risk (i.e., the likelihood that an adolescent has attempted suicide within the past 12 months) and classification, without implying any causal relationship between the item and the outcome.

Missing data imputation

The average missing rates of the data was 2.64%. The regularized iterative Factorial Analysis for Mixed Data (FAMD) algorithm from the missMDA R package (58) was utilized to conduct a single imputation. This algorithm flexibly imputes numerical and factor variables (59, 60) and, importantly, offers a fast and scalable

FIGURE 1
The process of model development and evaluation.
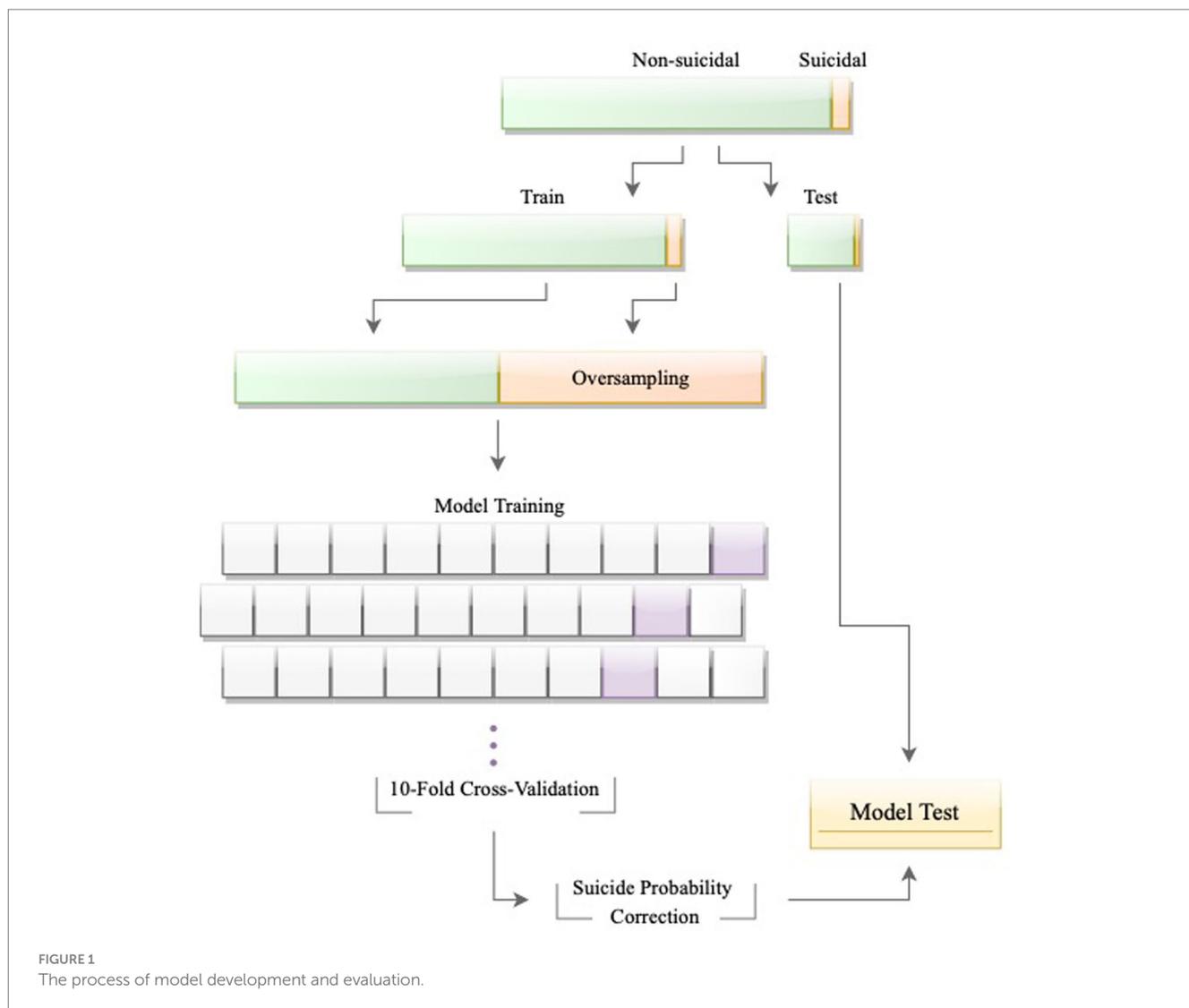
imputation solution for large datasets. Except for the descriptive statistics, all other analyses were carried out on the imputed dataset.

Results

The prevalence of self-reported recent suicide attempt in the sample was 4.66% ($n = 8,090$), with a male prevalence of 3.37% ($n = 2,824$) and a female prevalence of 5.87 ($n = 5,040$). There were 2.28% missing data on gender, and a Chi-square test revealed a significant difference between genders [χ^2 (1) $= 579.52$, $p < 0.001$]. According to Table 1, students in their third year of junior high school (age 15) and first year of senior high school (age 16) reported the highest prevalence of self-reported recent suicide attempts, both at a rate of 5.21%. Conversely, the lowest rate was observed among third-year senior high school students, at a rate of 3.20%. Students who reported no parental higher education had higher suicide attempt rates compared to others, as detailed in Table 1. Additionally, the majority of adolescents who reported a recent suicide attempt had not met with a psychologist (56.22%) or visited a health clinic for adolescents (69.67%), a school nurse (42.89%), or an accident and emergency ward (51.42%) in the past 12 months.

Model performance

Upon examination with the testing dataset, the best model achieved an AUPRC of 0.471 and an AUC of 0.921. The sensitivity and specificity of the model on the testing dataset were 0.77 and 0.90, respectively. This indicates that 77.0% of the adolescents who reported recent suicide attempt could be correctly classified, while simultaneously correctly classifying 90.0% of adolescents who did not report a recent suicide attempt. Despite the low prevalence of recent suicide attempts at the population level, this result indicates that the best-performing model could identify the majority of adolescents who reported a suicide attempt with high specificity.

Predictor importance

To address our second research question, we investigated the importance of the 168 items that were entered in the model as potential predictors. Table 2 shows the scaled relative importance of the top 15 predictors of suicide attempts, which were computed by normalizing mean absolute SHAP values. Of these, 7 items were

TABLE 1 Characteristics of the high-risk and no-risk suicide groups in the raw dataset.

Items (missing rate)	Reported recent suicide attempt (n = 8,090)	No reported recent suicide attempt (n = 165,574)
Gender (2.28%) *		
Boy	2,824 (3.37%)	80,983 (96.63%)
Girl	5,040 (5.87%)	80,863 (94.13%)
School year (1.58%) *		
Junior (age 13)	1,391 (4.12%)	32,395 (95.88%)
Junior (age 14)	1,673 (5.06%)	31,389 (94.94%)
Junior (age 15)	1,735 (5.21%)	31,571 (94.79%)
Senior (age 16)	1,760 (5.21%)	32,004 (94.79%)
Senior (age 17)	1,013 (3.93%)	24,792 (96.07%)
Senior (age 18)	359 (3.20%)	10,845 (96.80%)
Father education (13.59%) *		
Higher education	3,967 (3.93%)	96,891 (96.07%)
No higher education	2,757 (5.60%)	46,450 (94.40%)
Mother education (11.79%) *		
Higher education	4,518 (3.96%)	109,450 (96.04%)
No Higher education	2,466 (6.29%)	36,758 (93.71%)
Subjective SES (2.14%) *		
Well-off	2,164 (2.90%)	72,415 (97.10%)
Average	5,211 (5.58%)	88,124 (94.42%)
Poor	474 (23.37%)	1,554 (76.63%)

*Chi-square test of independence is significant with $p < 0.001$.
Reported values are from the raw dataset, prior to the imputation.

related to the domain of mental health, assessing symptoms of depression, anxiety, self-esteem, life satisfaction, or use of mental health service (e.g., "felt unhappy, sad, or depressed," "felt worthless," "satisfied with my appearance," "felt scared for no reason," and "contact with a psychologist/psychiatrist"). The item "felt unhappy, sad, or depressed" was the most important predictor of suicide attempt. In addition to mental health, items in the domain of victimization (e.g., "being teased or threatened"), substance use ("smoking" and "having been offered cannabis"), interpersonal problems (e.g., "satisfied with my parents," "having nobody to share problems with"), and lastly, participants' age were also among the most important predictors of suicide attempts. These prominent predictors, thus, were reflecting on four main domains of mental health, especially internalizing problems and substance use, victimization, and interpersonal relationships.

We visualized the SHAP values to elucidate how the significant predictors are related to the model's output for all participants in the testing dataset. In Figure 2, each of the top 15 predictors is presented in a separate row, with color coding representing the normalized data point, ranging from blue (0) to red (1). Take, for example, the blue section of the "Smoking" item corresponds to low values such as "I've never smoked," while red dots symbolize high values such as "I smoke every day." In addition to the color range, Figure 2 also illustrates the SHAP contribution of each variable, which can be negative, positive, or zero. Positive SHAP values

signify importance for positive classification and vice versa, while SHAP values close to zero denote unimportance. For example, adolescents who have had frequent contact with a psychologist or psychiatrist (thus a red color) have a high and positive SHAP values, indicating that this information supported the model to distinguish adolescents who had reported suicide attempts from other adolescents. Regardless of the direction of the SHAP values, the higher the absolute value, the more important the predictor. In Figure 2, for example, the item "felt unhappy, sad or depressed" has a high overall SHAP value at both negative and positive sides, indicating that this item strongly supported the model's classification. Figure 2 also shows that while taking hundreds of items into account, having lower age (shown in blue color) is indicator of a higher suicide attempt risk compared to higher age (red color).

Discussion

The aim of this study was to identify adolescents who had attempted suicide during the past 12 months without utilizing other sensitive suicide-related survey items as predictors. In doing so, we employed population-based national data to fine-tune the XGBoost machine learning algorithm based on a multitude of psychological, social, and environmental survey items. The fine-tuned model reached an excellent AUC of 92.1% (61). This figure is near the highest AUC reported in existing literature, ranging from 0.716 to 0.925 (23, 25, 26, 28). With a high specificity of 0.90 and a high sensitivity of 0.77, the model could correctly classify 77% of adolescents reporting a recent suicide attempt as well as 90% of adolescents not reporting a recent suicide attempt. Such performance is promising and indicates the feasibility of classifying adolescents' suicide attempts at scale with substantial accuracy, yet without utilizing any suicide-related items. Consequently, our findings answer our first research question, providing substantial evidence that high classification accuracy for adolescents' recent suicide attempts at the population scale can indeed be achieved without the need to rely on suicide-related items. This result has important implications for the broader field and could guide future efforts in developing scalable, sensitive, and effective tools for identifying and addressing suicide risk in young populations.

To address our second research question, we pinpointed the most important predictors and provisionally grouped them within four main domains of internalizing problems, interpersonal relationships, victimization, and substance use. Our findings are in line with existing literature showing that internalizing problems, particularly depressive symptoms, are key predictors of suicidal behavior among adolescents (7, 8, 62, 63). Similar to our findings, Carballo et al., (62) also identified conflicts in interpersonal relationships to be among the major risk factors of suicidality among adolescents [see also (64)]. Moreover, a review by Calati et al. (65) underscores the importance of interpersonal factors for suicidal outcomes. Reviews have also repeatedly emphasized the importance of childhood maltreatment and bullying experiences for understanding suicidal behavior (7, 9, 66, 67), thereby supporting the role of victimization experiences. Finally, substance use and more specifically, smoking, was also among the most important predictors of adolescents' suicide attempt. This aligns with numerous studies that recognize smoking as a general risk factor, independent of other socio-demographic variables (68–71), although

the association between smoking and suicide has been debated (72). Our analysis accentuates that frequent smoking is indeed related to adolescent suicide attempts, even when accounting for hundreds of other items.

In our analysis, while focusing on the important predictors of suicide attempts, we also observed that certain mental health-related factors were not as influential in the model as might have been expected based on prior literature. This warrants a closer examination of these seemingly inconsistent findings. For example, looking at

TABLE 2 Top 15 predictors of adolescents' suicide attempt and their relative importance.

Items	Relative Importance
Felt unhappy, sad, or depressed	1.000
Contact with a psychologist/psychiatrist	0.621
Felt worthless	0.477
Disappointed with myself	0.393
Smoking	0.360
Satisfied with my appearance	0.338
Satisfied with my parents	0.319
Felt scared for no reason	0.309
Being teased or threatened	0.296
Talking to parents in case of a problem	0.239
Having nobody to share problems with	0.227
Felt hopelessness about the future	0.213
Having been offered cannabis	0.204
Age	0.203
Having a partner or reporting a breakup	0.203

Relative importance represents normalized means absolute SHAP value.

Tables 1 and 2, despite significant variations in the suicide attempt rate across socio-demographic variables such as age, gender, and socio-economic status, only age emerged as one of the important predictors. Moreover, among the top 15 predictors, age contributed among the least in terms of SHAP values. The fact that gender was not identified as an important predictor is particularly noteworthy. This contrasts with existing literature, including findings by Esang and Ahmed (69), where gender has been recognized as a significant factor. Given the marked prevalence differences between boys and girls in our dataset, the lack of gender's role in the model requires careful consideration. A plausible explanation for this inconsistency may lie in the complex interplay of internalizing problems, interpersonal problems, and victimization experiences. These domains may subsume much of the demographic differences related to suicide attempts, rendering gender a comparatively less important predictor. Finally, in line with the existing literature, externalizing problems and delinquency did not emerge among the most important predictors of adolescent suicide attempts (69).

Limitations and strengths

Several limitations of the present study must be mentioned. Most notably, the assessment of suicide attempts through a single self-report item lacks detailed information about the nature, seriousness, and context of the attempt (73). However, as noted at the outset of the paper, collecting such sensitive items in nation-wide surveys from hundreds of thousands of adolescents can be difficult, in contrast to clinical samples or smaller population-based samples. Yet, despite the absence of detailed data to distinguish between severe and less severe suicide attempts, the high performance and generalizability of our model provide assurance that the accurate identification of adolescents who have self-reported a recent suicide attempt is feasible, a finding of

FIGURE 2
SHAP contribution of the top 15 predictors.
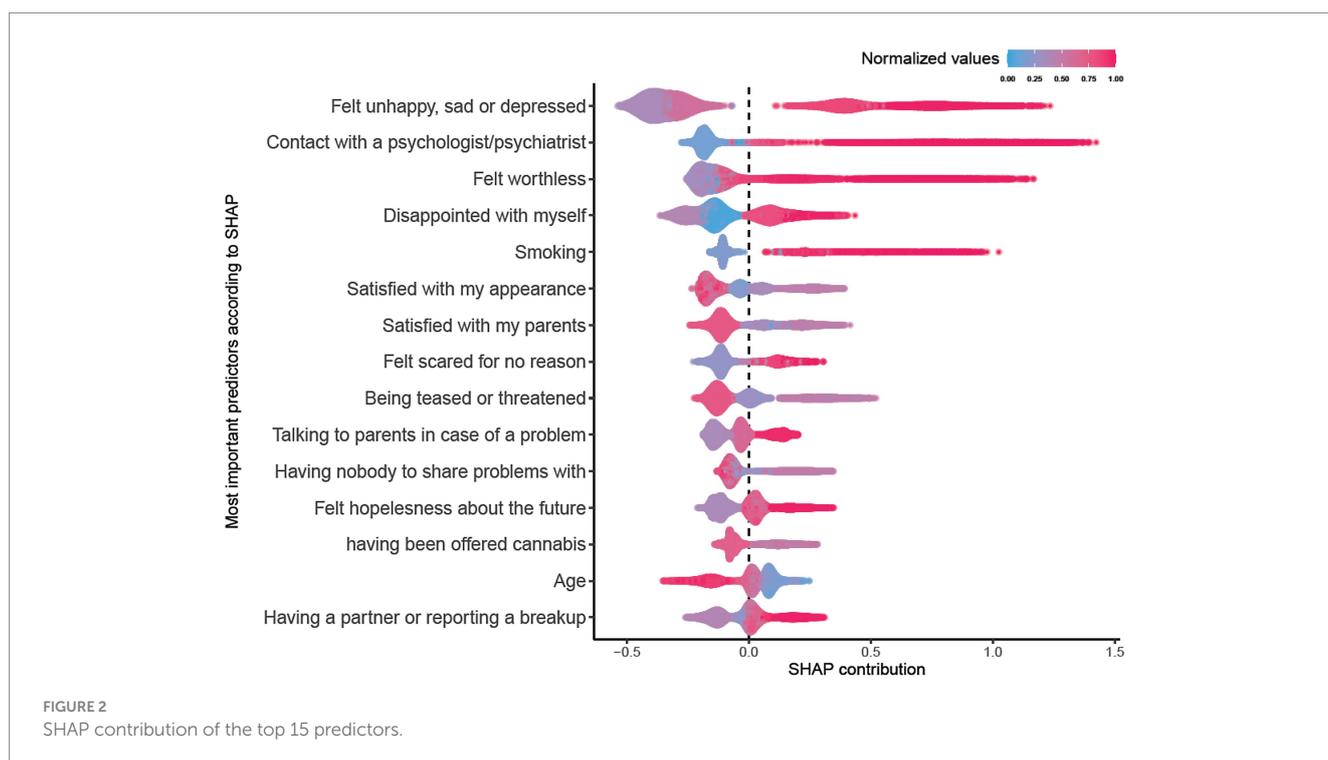
considerable significance. Additionally, our study employs retrospective cross-sectional data, limiting our understanding of the temporal development or severity of the predictors. While we have identified key psychological, social, and environmental predictors through comprehensive item analysis, the cross-sectional nature of our data precludes any inference about causality between these predictors and suicide attempts. However, as mentioned above, it is worth noting that our findings align well with existing literature that has established similar results in longitudinal studies. Emerging evidence suggests that risk scores derived from machine learning binary classification models using retrospective data can accurately predict future suicide attempts, including for those attempting for the first time [see (74, 75)]. Interestingly, these estimated risk scores also appear to be indicative of the severity of the suicide attempt. Recent longitudinal work demonstrates that higher estimated risk scores are inversely associated with the likelihood that adolescents will inconsistently report a prior lifetime suicide attempt in a 2-year follow-up assessment (76). These findings indicate that although our model is trained with retrospective cross-sectional data and cannot clarify causal relationships, yet, it provides real-world applications for identifying adolescents at risk of suicide attempts. Finally, we did not evaluate the model fairness, an issue increasingly emphasized in contemporary computer science literature. Warnings have been issued that even highly accurate models may underperform for minority groups, thereby perpetuating systematic inequality in fields such as social and health sciences (77). Future studies should aim to ascertain that models do not discriminate against specific socio-demographic groups, particularly if the model is intended for practical application (78).

Our paper also has a few significant strengths. The large sample size is particularly advantageous, as the low prevalence of suicide attempts necessitates a substantial number of participants to train robust classification models effectively. Additionally, the extensive array of survey items enables a comprehensive analysis, encompassing a myriad of potential risk and protective factors. This is particularly crucial given the complex nature of suicidal behavior (79), allowing the machine learning algorithms to account for intricate interactions between variables to refine predictive accuracy. Methodologically, we employed rigorous techniques to manage class imbalance and to evaluate and rank key predictors using a model-agnostic approach.

Conclusion

The utility of models requiring suicide-specific information for risk estimation of suicide attempts is questionable because such data are challenging to collect on a population scale, especially from adolescents. Our study demonstrates that it is possible to classify adolescents reporting recent suicide attempts with high accuracy without depending on such sensitive items. A novel contribution of our study is the identification of key factors—namely, internalizing problems, interpersonal relationships, victimization, and substance use—as important predictors of suicide attempts in this population, while accounting for a myriad of other items across different domains. A focus on these aspects is warranted when identifying at-risk adolescents, particularly from longitudinal research, which can offer invaluable insights into the causal dynamics between identified factors and adolescence suicide attempts.

Future research should address the limitations of our current study and align its findings with existing theoretical frameworks on suicidal behavior. Subsequent investigations should also examine the latent factorial structure of these predictors and consider whether different machine learning algorithms would identify similar important items. There is also an opportunity to explore the use of stacked ensemble meta-learners, constructed by combining multiple machine learning models, for achieving superior classification performance.

Data availability statement

The data used in the current research is not open-access. However, researchers can apply to access the dataset via Ungdata.no website. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Ethical committee at the Department of Psychology, University of Oslo. The studies were conducted in accordance with the local legislation and institutional requirements. Informed consent for participation in this study was provided by the participants.

Author contributions

EFH conceptualized the study and the research methods, carried out the data cleaning, missing data imputation, data analysis and visualization, as well as literature review and writing of the draft. NC and TVS were engaged in all steps of the study and provided input regarding multiple theoretical, technical, practical challenges, provided feedback on several drafts, and helped with revising the manuscript. All authors approved the final version of the manuscript for submission.

Funding

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.

Publisher's note

References

1. Nock MK, Borges G, Bromet EJ, Alonso J, Angermeyer M, Beautrais A, et al. Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *Br J Psychiatry*. (2008) 192:98–105. doi: 10.1192/bjp.bp.107.040113

2. Evans E, Hawton K, Rodham K, Deeks J. The prevalence of suicidal phenomena in adolescents: a systematic review of population-based studies. *Suicide Life Threat Behav*. (2005) 35:239–50. doi: 10.1521/suli.2005.35.3.239

3. World Health Organization. (2019). Suicide data. Available at: https://www.who.int/teams/mental-health-and-substance-use/data-research/suicide-data

4. Bae SM, Lee SA, Lee S-H. Prediction by data mining, of suicide attempts in Korean adolescents: a national study. *Neuropsychiatr Dis Treat*. (2015) 11:2367–75. doi: 10.2147/NDT.S91111

5. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull*. (2017) 143:187. doi: 10.1037/bul0000084

6. Carter G, Milner A, McGill K, Pirkis J, Kapur N, Spittal MJ. Predicting suicidal behaviours using clinical instruments: systematic review and meta-analysis of positive predictive values for risk scales. *Br J Psychiatry*. (2017) 210:387–95. doi: 10.1192/bjp.bp.116.182717

7. Greening L, Stoppelbein L, Fite P, Dhossche D, Erath S, Brown J, et al. Pathways to suicidal behaviors in childhood. *Suicide Life Threat Behav*. (2008) 38:35–45. doi: 10.1521/suli.2008.38.1.35

8. Lewis AJ, Bertino MD, Bailey CM, Skewes J, Lubman DI, Toumbourou JW. Depression and suicidal behavior in adolescents: a multi-informant and multi-methods approach to diagnostic classification. *Front Psychol*. (2014) 5:766. doi: 10.3389/fpsyg.2014.00766

9. Reed KP, Nugent W, Cooper RL. Testing a path model of relationships between gender, age, and bullying victimization and violent behavior, substance abuse, depression, suicidal ideation, and suicide attempts in adolescents. *Child Youth Serv Rev*. (2015) 55:128–37. doi: 10.1016/j.childyouth.2015.05.016

10. Walsh CG, Ribeiro JD, Franklin JC. Predicting risk of suicide attempts over time through machine learning. *Clin Psychol Sci*. (2017) 5:457–69. doi: 10.1177/2167702617691560

11. Bailey E, Mühlmann C, Rice S, Nedeljkovic M, Alvarez-Jimenez M, Sander L, et al. Ethical issues and practical barriers in internet-based suicide prevention research: a review and investigator survey. *BMC Med Ethics*. (2020) 21:1–16. doi: 10.1186/s12910-020-00479-1

12. Mckenzie SK, Li C, Jenkin G, Collings S. Ethical considerations in sensitive suicide research reliant on non-clinical researchers. *Res Ethics*. (2017) 13:173–83. doi: 10.1177/1747016116649996

13. Mozaffor M, Nurunnabi A, Shahriah S. Ethical issues in suicide research. *J Psychiatr Assoc Nepal*. (2020) 9:5–9. doi: 10.3126/jpan.v9i1.31307

14. Diamond GS, O'Malley A, Wintersteen MB, Peters S, Yunghans S, Biddle V, et al. Attitudes, practices, and barriers to adolescent suicide and mental health screening: a survey of Pennsylvania primary care providers. *J Prim Care Community Health*. (2012) 3:29–35. doi: 10.1177/2150131911417878

15. Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Iii HD, et al. Datasheets for datasets. *Commun ACM*. (2021) 64:86–92. doi: 10.1145/3458723

16. Safer DJ. Self-reported suicide attempts by adolescents. *Ann Clin Psychiatry*. (1997) 9:263–9. doi: 10.3109/10401239709147808

17. Turner CF, Ku L, Rogers SM, Lindberg LD, Pleck JH, Sonenstein FL. Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. *Science*. (1998) 280:867–73. doi: 10.1126/science.280.5365.867

18. Burke TA, Jacobucci R, Ammerman BA, Alloy LB, Diamond G. Using machine learning to classify suicide attempt history among youth in medical care settings. *J Affect Disord*. (2020) 268:206–14. doi: 10.1016/j.jad.2020.02.048

19. 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) 245:869–84. doi: 10.1016/j.jad.2018.11.073

20. Healy BC. Machine and deep learning in MS research are just powerful statistics–no. *Mult Scler J*. (2021) 27:663–4. doi: 10.1177/1352458520978648

21. Ley C, Martin RK, Pareek A, Groll A, Seil R, Tischer T. Machine learning and conventional statistics: making sense of the differences. *Knee Surg Sports Traumatol Arthrosc*. (2022) 30:753–7. doi: 10.1007/s00167-022-06896-6

22. Grendas LN, Chiapella L, Rodante DE, Daray FM. Comparison of traditional model-based statistical methods with machine learning for the prediction of suicide behaviour. *J Psychiatr Res*. (2022) 145:85–91. doi: 10.1016/j.jpsychires.2021.11.029

23. Mann JJ, Ellis SP, Waternaux CM, Liu X, Oquendo MA, Malone KM, et al. Classification trees distinguish suicide attempters in major psychiatric disorders: a model of clinical decision making. *J Clin Psychiatry*. (2008) 69:2693. doi: 10.4088/jcp.v69n0104

24. Miché M, Studerus E, Meyer AH, Gloster AT, Beesdo-Baum K, Wittchen H-U, et al. Prospective prediction of suicide attempts in community adolescents and young adults, using regression methods and machine learning. *J Affect Disord*. (2020) 265:570–8. doi: 10.1016/j.jad.2019.11.093

25. Shen Y, Zhang W, Chan BSM, Zhang Y, Meng F, Kennon EA, et al. Detecting risk of suicide attempts among Chinese medical college students using a machine learning algorithm. *J Affect Disord*. (2020) 273:18–23. doi: 10.1016/j.jad.2020.04.057

26. van Vuuren C, van Mens K, de Beurs D, Lokkerbol J, van der Wal M, Cuijpers P, et al. Comparing machine learning to a rule-based approach for predicting suicidal behavior among adolescents: results from a longitudinal population-based survey. *J Affect Disord*. (2021) 295:1415–20. doi: 10.1016/j.jad.2021.09.018

27. Delgado-Gomez D, Baca-Garcia E, Aguado D, Courtet P, Lopez-Castroman J. Computerized adaptive test vs. decision trees: development of a support decision system to identify suicidal behavior. *J Affect Disord*. (2016) 206:204–9. doi: 10.1016/j.jad.2016.07.032

28. Lopez-Castroman J, Perez-Rodriguez Mde L, Jaussent I, Alegria AA, Artes-Rodriguez A, Freed P, et al. Distinguishing the relevant features of frequent suicide attempters. *J Psychiatr Res*. (2011) 45:619–25. doi: 10.1016/j.jpsychires.2010.09.017

29. Metzger M, Tvardik N, Gicquel Q, Bouvry C, Poulet E, Potinet-Pagliaroli V. Use of emergency department electronic medical records for automated epidemiological surveillance of suicide attempts: a French pilot study. *Int J Methods Psychiatr Res*. (2017) 26:e1522. doi: 10.1002/mpr.1522

30. Currie CE, Elton RA, Todd J, Platt S. Indicators of socioeconomic status for adolescents: the WHO health behaviour in school-aged children survey. *Health Educ Res*. (1997) 12:385–97. doi: 10.1093/her/12.3.385

31. Torsheim T, Cavallo F, Levin KA, Schnohr C, Mazur J, Niclasen B, et al. Psychometric validation of the revised family affluence scale: a latent variable approach. *Child Indic Res*. (2016) 9:771–84. doi: 10.1007/s12187-015-9339-x

32. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins symptom checklist (HSCL): a self-report symptom inventory. *Behav Sci*. (1974) 19:1–15. doi: 10.1002/bs.3830190102

33. Wichstrøm L. Harter's self-perception profile for adolescents: reliability, validity, and evaluation of the question format. *J Pers Assess*. (1995) 65:100–16. doi: 10.1207/s15327752jpa6501_8

34. Russell D, Peplau LA, Cutrona CE. The revised UCLA loneliness scale: concurrent and discriminant validity evidence. *J Pers Soc Psychol*. (1980) 39:472. doi: 10.1037/0022-3514.39.3.472

35. Stefansen K, Hegna K, Valset K, von Soest T, Mossige S. Vold mot "homofil" ungdom. Forekomst og fortolkninger. *Sosiologi i Dag*. (2009) 39:43–71.

36. Olweus D. Prevalence and incidence in the study of antisocial behavior: definitions and measurements In: MW Klein, editor. *Cross-national research in self-reported crime and delinquency*. Netherlands: Springer (1989). 187–201.

37. Windle M. A longitudinal study of antisocial behaviors in early adolescence as predictors of late adolescent substance use: gender and ethnic group differences. *J Abnorm Psychol*. (1990) 99:86–91. doi: 10.1037/0021-843X.99.1.86

38. R Core Team. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing (2013).

39. H2O.ai. (2021). *h2o: R Interface for H2O*. R package (3.30.0.6). Available at: https://CRAN.R-project.org/package=h2o

40. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. Xgboost: Extreme gradient boosting. *R Package Version 0.4–2*. (2015) 1:1–4.

41. Mayer M. (2021). splitTools: Tools for data splitting (0.3.1). Available at: https://CRAN.R-project.org/package=splitTools

42. Carrington AM, Fieguth PW, Qazi H, Holzinger A, Chen HH, Mayr F, et al. A new concordant partial AUC and partial c statistic for imbalanced data in the evaluation of machine learning algorithms. *BMC Med Inform Decis Mak*. (2020) 20:1–12. doi: 10.1186/s12911-019-1014-6

43. Galar M, Fernandez A, Barrenechea E, Bustince H, Herrera F. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Trans Syst Man Cybern Part C Appl Rev*. (2011) 42:463–84. doi: 10.1109/TSMCC.2011.2161285

44. Jeni L. A., Cohn J. F., De La Torre F. (2013). Facing imbalanced data—recommendations for the use of performance metrics. 2013 Humaine association conference on affective computing and intelligent interaction Washington, DC IEEE Computer Society, 245–251

45. Dal Pozzolo A, Caelen O, Bontempi G. When is Undersampling effective in unbalanced classification tasks? In: A Appice, PP Rodrigues, V Santos Costa, C Soares, J Gama and A Jorge, editors. *Machine learning and knowledge discovery in databases*. Heidelberg: Springer International Publishing (2015). 200–15.

46. Provost F. (2000). Machine learning from imbalanced data sets 101. Proceedings of the AAAI'2000 workshop on imbalanced data sets, 68, 1–3

47. Dal Pozzolo A., Caelen O., Johnson R. A., Bontempi G. (2015). Calibrating probability with undersampling for unbalanced classification. 2015 IEEE Symposium Series on Computational Intelligence Cape Town IEEE, 159–166

48. Wallace B. C., Dahabreh I. J. (2012). Class probability estimates are unreliable for imbalanced data (and how to fix them) in *2012 IEEE 12th international conference on data mining*. (pp. 695–704). doi: 10.1109/ICDM.2012.115

49. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *J Mach Learn Res*. (2012) 13:281–305.

50. Ying X. An overview of overfitting and its solutions. *J. Phys. Conf. Ser.* (2019) 1168:22022. doi: 10.1088/1742-6596/1168/2/022022

51. Provost F, Fawcett T, Kohavi R. The case against accuracy estimation for comparing induction algorithms. *Proc 15th Internat Conf Mach Learn.* (1998) 1998:445–53. doi: 10.5555/645527.657469

52. Davis J, Goadrich M. (2006). The relationship between precision-recall and ROC curves. Proceedings of the 23rd International Conference on Machine Learning (ICML '06), 233–240

53. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions in *31st conference on neural information processing systems (NIPS 2017)*. (2017). Available at: https://proceedings.neurips.cc/paper_files/paper/2017

54. Shapley LS. A value for n-person games In: *Classics in game theory*: Princeton, Princeton University Press (1997).

55. Winter E. The shapley value. *Handb Game Theory Econ. Applicat.* (2002) 3:2025–54. doi: 10.1016/S1574-0005(02)03016-3

56. Lundberg S. M. (2022). SHAP (SHapley additive exPlanations) [computer software]. Available at: https://github.com/slundberg/shap

57. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intellig.* (2020) 2:56–67. doi: 10.1038/s42256-019-0138-9

58. Josse J, Husson F. MissMDA: a package for handling missing values in multivariate data analysis. *J Stat Softw.* (2016) 70:1–31. doi: 10.18637/jss.v070.i01

59. Audigier V, Husson F, Josse J. A principal component method to impute missing values for mixed data. *Adv Data Anal Classificat.* (2016) 10:5–26. doi: 10.1007/s11634-014-0195-1

60. Hunt LA. Missing data imputation and its effect on the accuracy of classification In: F Palumbo, A Montanari and M Vichi, editors. *Data science.* New York: Springer (2017). 3–14.

61. Šimundić AM. Measures of Diagnostic Accuracy: Basic Definitions. *EJIFCC.* (2009) 19:203–11.

62. Carballo JJ, Llorente C, Kehrmann L, Flamarique I, Zuddas A, Purper-Ouakil D, et al. Psychosocial risk factors for suicidality in children and adolescents. *Eur Child Adolesc Psychiatry.* (2020) 29:759–76. doi: 10.1007/s00787-018-01270-9

63. Lohner J, Konrad N. Deliberate self-harm and suicide attempt in custody: distinguishing features in male inmates' self-injurious behavior. *Int J Law Psychiatry.* (2006) 29:370–85. doi: 10.1016/j.ijlp.2006.03.004

64. Toprak S, Cetin I, Guven T, Can G, Demircan C. Self-harm, suicidal ideation and suicide attempts among college students. *Psychiatry Res.* (2011) 187:140–4. doi: 10.1016/j.psychres.2010.09.009

65. Calati R, Ferrari C, Brittner M, Oasi O, Olié E, Carvalho AF, et al. Suicidal thoughts and behaviors and social isolation: a narrative review of the literature. *J Affect Disord.* (2019) 245:653–67. doi: 10.1016/j.jad.2018.11.022

66. Cha CB, Franz PJ, Guzmán EM, Glenn CR, Kleiman EM, Nock MK. Annual research review: suicide among youth – epidemiology, (potential) etiology, and treatment. *J Child Psychol Psychiatry.* (2018) 59:460–82. doi: 10.1111/jcpp.12831

67. Darke S, Torok M, Kaye S, Ross J. Attempted suicide, self-harm, and violent victimization among regular illicit drug users. *Suicide Life Threat Behav.* (2010) 40:587–96. doi: 10.1521/suli.2010.40.6.587

68. Clarke DE, Eaton WW, Petronis KR, Ko JY, Chatterjee A, Anthony JC. Increased risk of suicidal ideation in smokers and former smokers compared to never smokers: evidence from the Baltimore ECA follow-up study. *Suicide Life Threat Behav.* (2010) 40:307–18. doi: 10.1521/suli.2010.40.4.307

69. Esang M, Ahmed S. A closer look at substance use and suicide. *Am J Psychiat Residents' J.* (2018) 13:6–8. doi: 10.1176/appi.ajp-rj.2018.130603

70. Rowan AB. Adolescent substance abuse and suicide. *Depress Anxiety.* (2001) 14:186–91. doi: 10.1002/da.1065

71. Vijayakumar L, Kumar MS, Vijayakumar V. Substance use and suicide. *Curr Opin Psychiatry.* (2011) 24:197–202. doi: 10.1097/YCO.0b013e3283459242

72. Kessler RC, Borges G, Sampson N, Miller M, Nock MK. The association between smoking and subsequent suicide-related outcomes in the National Comorbidity Survey panel sample. *Mol Psychiatry.* (2009) 14:1132–42. doi: 10.1038/mp.2008.78

73. Christl B, Wittchen H-U, Pfister H, Lieb R, Bronisch T. The accuracy of prevalence estimations for suicide attempts. How reliably do adolescents and young adults report their suicide attempts? *Arch Suicide Res.* (2006) 10:253–63. doi: 10.1080/13811110600582539

74. Haghish EF, Czajkowski NO. Reconsidering false positives in machine learning binary classification models of suicidal behavior. *Curr Psychol.* (2023) 2023. doi: 10.1007/s12144-023-05174-z

75. Haghish EF, Laeng B, Czajkowski NO. Are false positives in suicide classification models a risk group? Evidence for "true alarms" in a population-representative longitudinal study of Norwegian adolescents. *Front Psychol.* (2023a) 14:1216483. doi: 10.3389/fpsyg.2023.1216483

76. Haghish E. F., Czajkowski N. O., Walby F. A., Qin P., Laeng B. (2023b). Suicide attempt risk predicts adolescents' inconsistent self-reported suicide attempts: A machine learning approach using longitudinal data. Manuscript submitted for peer-review.

77. Abebe R., Barocas S., Kleinberg J., Levy K., Raghavan M., Robinson D. G. (2020). Roles for computing in social change. Proceedings of the 2020 conference on fairness, accountability, and transparency New York, NY, 252–260

78. Yuan M, Kumar V, Ahmad MA, Teredesai A. Assessing fairness in classification parity of machine learning models in healthcare. *ArXiv Preprint ArXiv:210203717.* (2021). doi: 10.48550/arXiv.2102.03717

79. de Beurs D, Bockting C, Kerkhof A, Scheepers F, O'Connor R, Penninx B, et al. A network perspective on suicidal behavior: understanding suicidality as a complex system. *Suicide Life Threat Behav.* (2021) 51:115–26. doi: 10.1111/sltb.12676