

Prediction Models for Suicide Attempts among Adolescents Using Machine Learning Techniques

Jae Seok Lim^{1,*}, Chan-Mo Yang^{2,3,*}, Ju-Won Baek⁴, Sang-Yeol Lee², Bung-Nyun Kim³

¹Department of Oral and Maxillofacial Surgery, Chungbuk National University Hospital, Cheongju, ²Department of Psychiatry, School of Medicine, Wonkwang University, Iksan, ³Division of Child and Adolescent Psychiatry, Department of Psychiatry, Graduate School of Medicine, Seoul National University, Seoul, ⁴Dental Clinic Center, Chungbuk National University Hospital, Cheongju, Korea

Objective: Suicide attempts (SAs) in adolescents are difficult to predict although it is a leading cause of death among adolescents. This study aimed to develop and evaluate SA prediction models based on six different machine learning (ML) algorithms for Korean adolescents using data from online surveys.

Methods: Data were extracted from the 2011–2018 Korea Youth Risk Behavior Survey (KYRBS), an ongoing annual national survey. The participants comprised 468,482 nationally representative adolescents from 400 middle and 400 high schools, aged 12 to 18. The models were trained using several classic ML methods and then tested on internal and external independent datasets; performance metrics were calculated. Data analysis was performed from March 2020 to June 2020.

Results: Among the 468,482 adolescents included in the analysis, 15,012 cases (3.2%) were identified as having made an SA. Three features (suicidal ideation, suicide planning, and grade) were identified as the most important predictors. The performance of the six ML models on the internal testing dataset was good, with both the area under the receiver operating characteristic curve (AUROC) and area under the precision–recall curve (AUPRC) ranging from 0.92 to 0.94. Although the AUROC of all models on the external testing dataset (2018 KYRBS) ranged from 0.93 to 0.95, the AUPRC of the models was approximately 0.5.

Conclusion: The developed and validated SA prediction models can be applied to detect high risks of SA. This approach could facilitate early intervention in the suicide crisis and may ultimately contribute to suicide prevention for adolescents.

KEY WORDS: Adolescents; Suicide; Attempted suicide; Machine learning.

INTRODUCTION

Suicide is a major public healthcare issue, with an estimated 804,000 suicide deaths per year [1]. In particular, suicidal behavior in adolescents is a major concern

worldwide [2], with suicide being the second-leading cause of death among adolescents [1,3] and the leading cause of death among Korean adolescents specifically over the past 11 years [4]. Whereas 17.2% of US high school students seriously considered attempting suicide and 7.4% made a suicide attempt (SA) in 2017 [5], the proportion of adolescents who experienced suicidal ideation (SI) in Korea was 13.3%, and 3.1% made SAs [5,6]. Nevertheless, very few of the students at risk of suicide sought further help for their difficulties [7]. Considering the limited healthcare resources and low treatment rate [8,9], there is an urgent need for the early detection of a high risk for SA.

Although numerous studies have been conducted on identifying those at risk for SA, a previous meta-analysis revealed that their performance in predicting SA was not

Received: March 26, 2021 / **Revised:** May 18, 2021

Accepted: May 27, 2021

Address for correspondence: Sang-Yeol Lee
Department of Psychiatry, School of Medicine, Wonkwang University, 895 Muwang-ro, Iksan 54538, Korea
E-mail: psysangyeol@hanmail.net
ORCID: <https://orcid.org/0000-0003-1828-9992>

Bung-Nyun Kim
Division of Child and Adolescent Psychiatry, Department of Psychiatry, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea
E-mail: kbn1@snu.ac.kr
ORCID: <https://orcid.org/0000-0002-2403-3291>

*These authors contributed equally to this work.

© This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

far superior to predictions based on random chance [10]. This indicates the limitations of traditional models that predict SA through risk factors, which are reliant on researchers. Moreover, it suggests the need for a new predictive approach, such as machine learning (ML), which can reflect the simultaneous testing of various factors and the nonlinearity or innumerable complex interactions among predictors [11].

In recent years, an increasing amount of research applying ML techniques to suicide prediction has been conducted [12]. However, according to previous studies [12,13], several limitations remain, such as imbalanced datasets, ineffective sampling methods, and a low positive predictive value (PPV). The necessity of performing validation on external datasets and the application of various ML techniques, including deep neural networks (DNNs), have been demonstrated. Moreover, previous works on predicting adolescent SA using ML have focused largely on biased samples [3,14–17]. Therefore, the goal of this study was to develop six ML models to address the above issues in predicting SA among adolescents by using a nationally representative general population sample with close to half a million individuals.

METHODS

Study Population and Data Source

In this study, the analyzed data were obtained from the 2011–2018 Korea Youth Risk Behavior Survey (KYRBS), which was established in 2005 and is administered by the Ministry of Health and Welfare, the Korean Ministry of Education, and the Korea Centers for Disease Control and Prevention in South Korea [18]. The KYRBS was approved by the Institutional Review Board of the Korea Centers for Disease Control and Prevention (Certification Number: 11758). This government-approved statistical survey is an ongoing annual online national survey to assess the prevalence of health risk behaviors among Korean middle and high school students. The survey consists of 103 questions divided into 15 sections on socio-demographic characteristics, health-related behaviors, and mental and physical health [19]. All data used in this study were fully anonymized prior to accession. Participants with incomplete information were excluded. The samples in this study consisted of the following datasets: KYRBS 2011 ($n = 73,473$, response rate = 95.5%); KYRBS 2012 ($n = 72,228$, re-

sponse rate = 96.4%); KYRBS 2013 ($n = 70,354$, response rate = 96.4%); KYRBS 2014 ($n = 69,959$, response rate = 97.2%); KYRBS 2016 ($n = 63,741$, response rate = 96.4%); KYRBS 2017 ($n = 60,392$, response rate = 95.8%); and KYRBS 2018 ($n = 58,335$, response rate = 95.6%). KYRBS 2015 was excluded because drug experience data were missing. Thus, all of the descriptions regarding periods using a dash (e.g., 2011–2017 KYRBS) hereafter in the manuscript refer to datasets excluding KYRBS 2015. Ultimately, a total of 468,482 adolescents were included for the development of the ML models.

Features

The candidate features for the learning models were selected from a literature-based search following variables that were previously reported to be related to the risk of SA: socio-demographic variables (sex, grade, city type, academic achievement, family structure, family socioeconomic status, and education level of father and mother); health-related lifestyle factors (current smoking, current alcohol drinking, drug experience, physical activity, and body mass index [BMI]); psychological stress factors (sadness, stress, self-rated health, sleep satisfaction, and perceived body image), comorbidities (asthma, allergic rhinitis, and atopic dermatitis); and risk factors (suicide planning [SP] and SI) [10,20–22]. SA, as a dependent variable, was defined as a positive response to the question, “Have you attempted suicide in the past 12 months?” Subjects were asked to respond with either “no” or “yes.” Details about the survey questions are presented in the Supplementary methods (available online).

General Prediction Pipeline

Two general prediction pipelines were developed (Fig. 1). The first pipeline was solely generated from classic ML (CML) methods, namely logistic regression (LR), random forest (RF), artificial neural networks (ANN), support vector machines (SVM), and extreme gradient boosting (XGB) using the Caret package implemented in the R statistical software version 3.6.3 (RStudio). The second pipeline was generated from a DNN learning model using Keras, which is a deep learning library implemented in the R with TensorFlow as the backend. The developed pipeline randomly split the input dataset into training (70% of 2011–2017 KYRBS) and testing (30% of 2011–2017 KYRBS) datasets, while maintaining equal proportions of the class

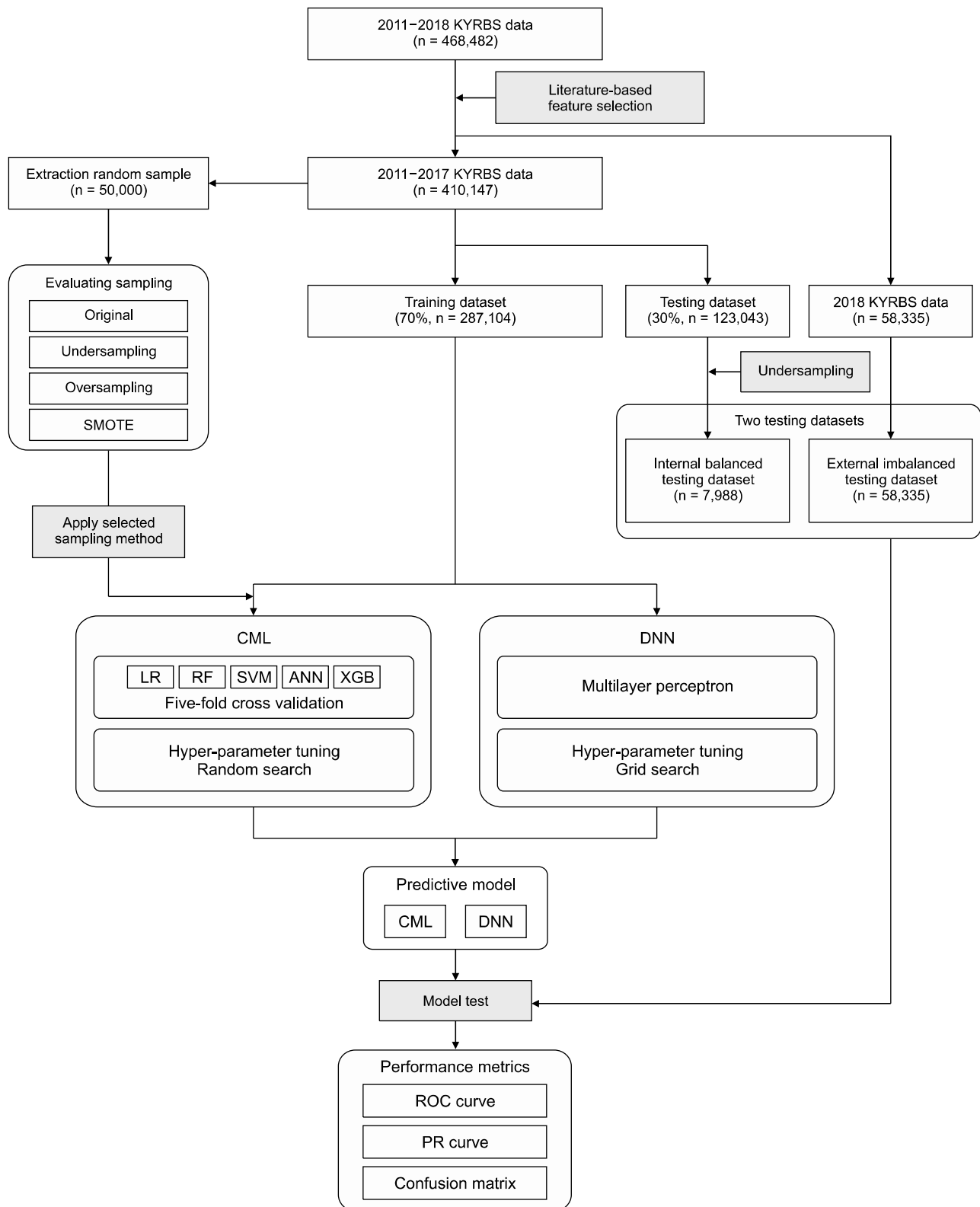


Fig. 1. Schematic of prediction model development.

KYRBS, Korea Youth Risk Behavior Survey; ROC, receiver operating characteristic; PR, precision–recall; LR, logistic regression; RF, random forest; ANN, artificial neural network; SVM, support vector machine; XGB, extreme gradient boosting; SMOTE, synthetic minority oversampling technique; CML, classic machine learning; DNN, deep neural network.

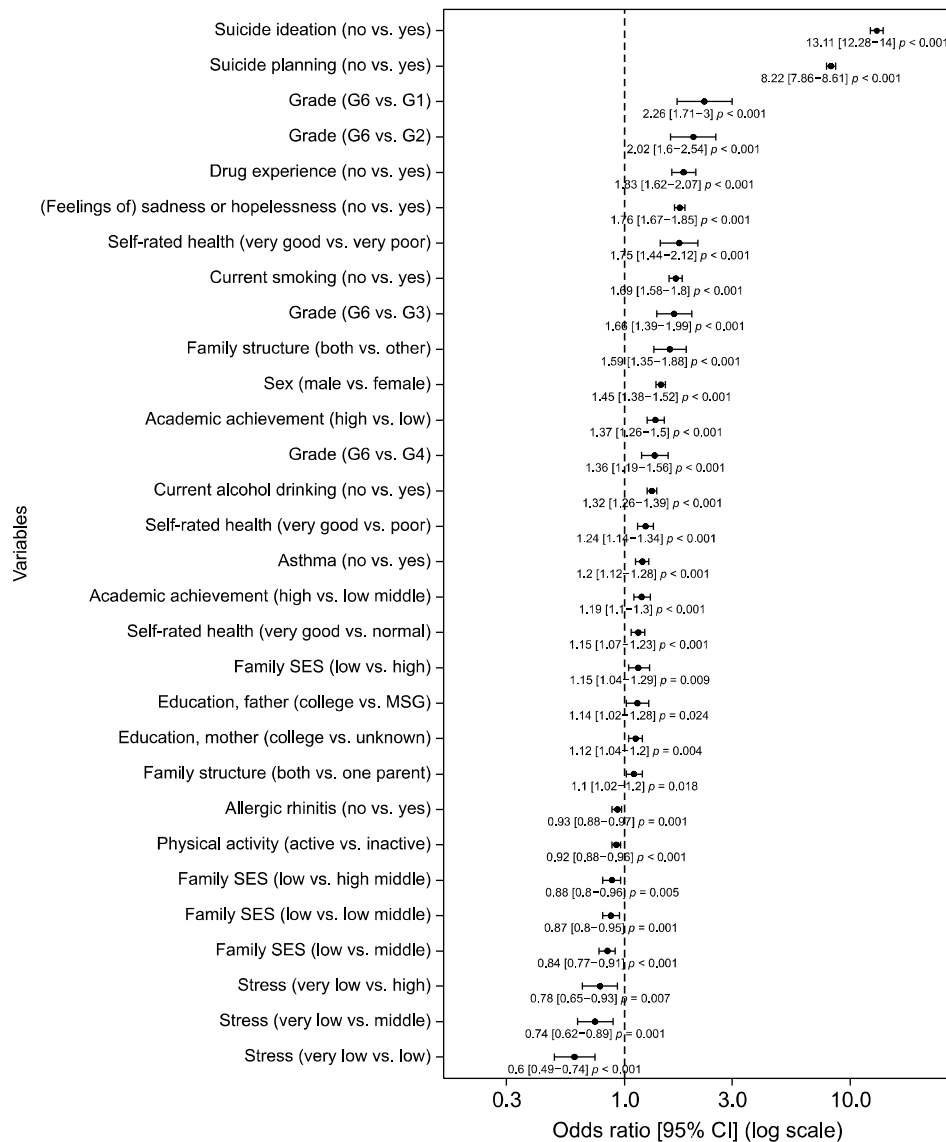


Fig. 2. Multivariate LR analysis to identify variables associated with SA. The forest plots indicate the odds ratios and confidence intervals of the variables associated with SA. The black dots indicate the adjusted odds ratios for the variables ($p < 0.05$) and the error bars indicate 95% confidence intervals. CI, confidence interval; MSG, middle school graduate or lower; LR, logistic regression; SA, suicide attempt; SES, socioeconomic status.

ratios in each split. The testing dataset was rebalanced using the undersampling method; the result was named the internal balanced testing dataset. The 2018 KYRBS dataset was used to validate the learning models externally, and was named the external imbalanced testing dataset. To assess the model's performance, the receiver operating characteristic (ROC) and precision–recall (PR) curves were plotted, and the respective area under the ROC curve (AUROC) and area under the PR curve (AUPRC) were obtained. A confusion matrix including the sensitivity, specificity, PPV and negative predictive value, as well as the F1 score, were obtained. Details about development of CML and DNN are presented in the Supplementary methods (available online).

RESULTS

We first summarized the baseline characteristics of the 2011–2017 and 2018 KYRBS datasets. Thereafter, we performed multivariate LR analysis to identify the variables associated with SA. We identified 13 features that were significantly associated with SA in both sets: sex, grade, academic achievement, family structure, education of father, education of mother, current smoking, current alcohol drinking, drug experience, sadness or hopelessness, self-rated health, SI, and SP (Fig. 2, Table 1, and Supplementary Table 1; available online). Moreover, the six features of age, city type, BMI, sleep, perceived body image, and atopic dermatitis were not significantly asso-

Table 1. Baseline characteristics of 2011–2017 KYRBS data

Features	Subgroup	Non-SA	SA	OR (univariate)	OR (multivariate)
Sex	Male	203,734 (97.7)	4,782 (2.3)	-	-
	Female	193,098 (95.8)	8,533 (4.2)	1.88 (1.82–1.95, $p < 0.001$)	1.45 (1.38–1.52, $p < 0.001$)
Age		15.4 ± 1.7	15.2 ± 1.7	0.91 (0.90–0.92, $p < 0.001$)	1.03 (0.98–1.09, $p = 0.242$)
Grade	G6	65,407 (97.6)	1,591 (2.4)	-	-
	G5	66,759 (97.3)	1,876 (2.7)	1.16 (1.08–1.24, $p < 0.001$)	1.10 (1.00–1.21, $p = 0.051$)
	G4	66,471 (97.1)	2,012 (2.9)	1.24 (1.16–1.33, $p < 0.001$)	1.36 (1.19–1.56, $p < 0.001$)
	G3	66,821 (96.4)	2,501 (3.6)	1.54 (1.44–1.64, $p < 0.001$)	1.66 (1.39–1.99, $p < 0.001$)
	G2	65,870 (96.0)	2,768 (4.0)	1.73 (1.62–1.84, $p < 0.001$)	2.02 (1.60–2.54, $p < 0.001$)
	G1	65,504 (96.2)	2,567 (3.8)	1.61 (1.51–1.72, $p < 0.001$)	2.26 (1.71–3.00, $p < 0.001$)
City type	Big city	178,508 (96.8)	5,934 (3.2)	-	-
	Medium small city	178,611 (96.8)	5,915 (3.2)	1.00 (0.96–1.03, $p = 0.839$)	1.02 (0.98–1.07, $p = 0.395$)
	Countryside	39,713 (96.4)	1,466 (3.6)	1.11 (1.05–1.18, $p < 0.001$)	1.04 (0.97–1.12, $p = 0.273$)
Academic achievement	High	47,198 (97.5)	1,216 (2.5)	-	-
	High middle	98,821 (97.5)	2,508 (2.5)	0.99 (0.92–1.06, $p = 0.671$)	1.00 (0.92–1.08, $p = 0.956$)
	Middle	111,509 (97.4)	2,966 (2.6)	1.03 (0.97–1.10, $p = 0.355$)	1.08 (0.99–1.17, $p = 0.075$)
	Low middle	95,787 (96.2)	3,779 (3.8)	1.53 (1.43–1.64, $p < 0.001$)	1.19 (1.10–1.30, $p < 0.001$)
	Low	43,517 (93.9)	2,846 (6.1)	2.54 (2.37–2.72, $p < 0.001$)	1.37 (1.26–1.50, $p < 0.001$)
Family structure	Living with both parents	370,490 (96.9)	11,807 (3.1)	-	-
	Living with one parent	22,978 (95.1)	1,182 (4.9)	1.61 (1.52–1.72, $p < 0.001$)	1.10 (1.02–1.20, $p = 0.018$)
	Other	3,364 (91.2)	326 (8.8)	3.04 (2.71–3.41, $p < 0.001$)	1.59 (1.35–1.88, $p < 0.001$)
Family SES	Low	14,577 (91.7)	1,324 (8.3)	-	-
	Low middle	59,663 (95.7)	2,652 (4.3)	0.49 (0.46–0.52, $p < 0.001$)	0.87 (0.80–0.95, $p = 0.001$)
	Middle	190,078 (97.3)	5,343 (2.7)	0.31 (0.29–0.33, $p < 0.001$)	0.84 (0.77–0.91, $p < 0.001$)
	High middle	102,080 (97.3)	2,804 (2.7)	0.30 (0.28–0.32, $p < 0.001$)	0.88 (0.80–0.96, $p = 0.005$)
	High	30,434 (96.2)	1,192 (3.8)	0.43 (0.40–0.47, $p < 0.001$)	1.15 (1.04–1.29, $p = 0.009$)
Education, father	College	183,870 (97.0)	5,679 (3.0)	-	-
	High school graduate	123,924 (96.8)	4,122 (3.2)	1.08 (1.03–1.12, $p < 0.001$)	1.04 (0.98–1.10, $p = 0.229$)
	Middle school graduate or less	13,123 (95.6)	604 (4.4)	1.49 (1.37–1.62, $p < 0.001$)	1.14 (1.02–1.28, $p = 0.024$)
	Unknown	75,915 (96.3)	2,910 (3.7)	1.24 (1.19–1.30, $p < 0.001$)	0.97 (0.90–1.04, $p = 0.351$)
Education, mother	College	154,410 (96.9)	4,867 (3.1)	-	-
	High school graduate	157,338 (96.9)	5,032 (3.1)	1.01 (0.97–1.06, $p = 0.476$)	0.97 (0.92–1.03, $p = 0.347$)
	Middle school graduate or less	12,406 (96.0)	511 (4.0)	1.31 (1.19–1.43, $p < 0.001$)	0.93 (0.82–1.05, $p = 0.248$)
	Unknown	72,678 (96.2)	2,905 (3.8)	1.27 (1.21–1.33, $p < 0.001$)	1.12 (1.04–1.20, $p = 0.004$)
Current smoking	No	369,320 (97.2)	10,793 (2.8)	-	-
	Yes	27,512 (91.6)	2,522 (8.4)	3.14 (3.00–3.28, $p < 0.001$)	1.69 (1.58–1.80, $p < 0.001$)
Current alcohol drinking	No	331,764 (97.4)	9,008 (2.6)	-	-
	Yes	65,068 (93.8)	4,307 (6.2)	2.44 (2.35–2.53, $p < 0.001$)	1.32 (1.26–1.39, $p < 0.001$)
Drug experience	No	393,830 (96.9)	12,672 (3.1)	-	-
	Yes	3,002 (82.4)	643 (17.6)	6.66 (6.10–7.26, $p < 0.001$)	1.83 (1.62–2.07, $p < 0.001$)
Body mass index	Optimal	213,093 (96.8)	7,101 (3.2)	-	-
	Underweight	97,022 (96.5)	3,479 (3.5)	1.08 (1.03–1.12, $p < 0.001$)	1.04 (0.97–1.10, $p = 0.247$)
	Overweight	45,102 (96.8)	1,478 (3.2)	0.98 (0.93–1.04, $p = 0.564$)	0.96 (0.89–1.04, $p = 0.304$)
	Obese	41,615 (97.1)	1,257 (2.9)	0.91 (0.85–0.96, $p = 0.002$)	0.95 (0.87–1.03, $p = 0.215$)
(Feelings of) sadness or hopelessness	No	289,978 (99.0)	2,831 (1.0)	-	-
	Yes	106,854 (91.1)	10,484 (8.9)	10.05 (9.64–10.48, $p < 0.001$)	1.76 (1.67–1.85, $p < 0.001$)
Stress	Very low	12,229 (98.4)	195 (1.6)	-	-
	Low	61,955 (99.3)	418 (0.7)	0.42 (0.36–0.50, $p < 0.001$)	0.60 (0.49–0.74, $p < 0.001$)
	Middle	170,494 (98.5)	2,513 (1.5)	0.92 (0.80–1.07, $p = 0.294$)	0.74 (0.62–0.89, $p = 0.001$)
	High	113,280 (95.8)	4,977 (4.2)	2.76 (2.39–3.19, $p < 0.001$)	0.78 (0.65–0.93, $p = 0.007$)
	Very high	38,874 (88.2)	5,212 (11.8)	8.41 (7.30–9.75, $p < 0.001$)	0.88 (0.74–1.06, $p = 0.176$)

Table 1. Continued

Features	Subgroup	Non-SA	SA	OR (univariate)	OR (multivariate)
Sleep	Very high	31,141 (97.9)	661 (2.1)	-	-
	High	78,404 (98.1)	1,547 (1.9)	0.93 (0.85–1.02, $p = 0.120$)	0.95 (0.85–1.06, $p = 0.364$)
	Middle	130,076 (97.4)	3,520 (2.6)	1.27 (1.17–1.39, $p < 0.001$)	0.97 (0.87–1.07, $p = 0.501$)
	Low	108,702 (96.4)	4,111 (3.6)	1.78 (1.64–1.94, $p < 0.001$)	0.97 (0.87–1.07, $p = 0.500$)
	Very low	48,509 (93.3)	3,476 (6.7)	3.38 (3.11–3.68, $p < 0.001$)	1.00 (0.90–1.11, $p = 0.928$)
Self-rated health	Very good	91,902 (97.9)	1,962 (2.1)	-	-
	Good	188,080 (97.6)	4,589 (2.4)	1.14 (1.08–1.21, $p < 0.001$)	0.96 (0.90–1.02, $p = 0.188$)
	Normal	92,739 (95.5)	4,405 (4.5)	2.22 (2.11–2.35, $p < 0.001$)	1.15 (1.07–1.23, $p < 0.001$)
	Poor	23,020 (91.6)	2,120 (8.4)	4.31 (4.05–4.59, $p < 0.001$)	1.24 (1.14–1.34, $p < 0.001$)
	Very poor	1,091 (82.0)	239 (18.0)	10.26 (8.84–11.86, $p < 0.001$)	1.75 (1.44–2.12, $p < 0.001$)
Perceived body image	Normal	139,224 (97.1)	4,129 (2.9)	-	-
	Very thin	18,751 (96.3)	722 (3.7)	1.30 (1.20–1.41, $p < 0.001$)	1.00 (0.90–1.12, $p = 0.950$)
	Thin	89,787 (97.1)	2,720 (2.9)	1.02 (0.97–1.07, $p = 0.397$)	0.95 (0.89–1.01, $p = 0.096$)
	Fat	130,864 (96.5)	4,802 (3.5)	1.24 (1.19–1.29, $p < 0.001$)	0.99 (0.93–1.04, $p = 0.657$)
	Very fat	18,206 (95.1)	942 (4.9)	1.74 (1.62–1.87, $p < 0.001$)	0.99 (0.89–1.10, $p = 0.893$)
Physical activity	Active	181,878 (96.6)	6,369 (3.4)	-	-
	Inactive	214,954 (96.9)	6,946 (3.1)	0.92 (0.89–0.96, $p < 0.001$)	0.92 (0.88–0.96, $p < 0.001$)
Asthma	No	361,579 (96.9)	11,584 (3.1)	-	-
	Yes	35,253 (95.3)	1,731 (4.7)	1.53 (1.46–1.61, $p < 0.001$)	1.20 (1.12–1.28, $p < 0.001$)
Allergic rhinitis	No	265,812 (96.9)	8,613 (3.1)	-	-
	Yes	131,020 (96.5)	4,702 (3.5)	1.11 (1.07–1.15, $p < 0.001$)	0.93 (0.88–0.97, $p = 0.001$)
Atopic dermatitis	No	301,656 (96.9)	9,643 (3.1)	-	-
	Yes	95,176 (96.3)	3,672 (3.7)	1.21 (1.16–1.25, $p < 0.001$)	1.01 (0.97–1.06, $p = 0.579$)
Suicide ideation	No	346,331 (99.6)	1,543 (0.4)	-	-
	Yes	50,501 (81.1)	11,772 (18.9)	52.32 (49.59–55.24, $p < 0.001$)	13.11 (12.28–14.00, $p < 0.001$)
Suicide planning	No	384,510 (98.7)	5,014 (1.3)	-	-
	Yes	12,322 (59.7)	8,301 (40.3)	51.66 (49.67–53.74, $p < 0.001$)	8.22 (7.86–8.61, $p < 0.001$)

Values are presented as number (%) or mean \pm standard deviation.

KYRBS, Korea Youth Risk Behavior Survey; SES, socioeconomic status; SA, suicide attempt; OR, odds ratio.

ciated with SA in both datasets (Supplementary Fig. 1; available online). There were no strong correlations among all features (Supplementary Fig. 2; available online).

The observed SA ratio was 3.2% (15,012/468,482), which was consistent with the imbalanced data. To determine the most efficient sampling method in the dataset with an imbalanced class distribution, we compared the performance of four sampling methods, namely the original, undersampling, oversampling, and SMOTE methods using the LR, RF, ANN, SVM, and XGB models. No clear differences in the performance metrics, including the ROC and PR curves, were observed among the sampling methods (Supplementary Fig. 3; available online). Therefore, we applied the undersampling method to rebalance the imbalanced dataset in the subsequent CML analysis.

First, we tested all models on the internal balanced testing dataset. The AUROCs and AUPRCs of all models were above 0.9, indicating that all models performed effec-

tively on the balanced dataset (Fig. 3). The relative importance of all features was calculated by the Boruta algorithm [23]. Three features, namely city type, asthma, and atopic dermatitis, were determined as irrelevant features for predicting SA (Fig. 4). In comparison to the multivariate regression analysis, the Boruta algorithm identified different feature sets as relevant predictors. However, the top three features, namely SP, SI, and grade, were identified as the most important predictors in both analyses (Figs. 2, 4).

Finally, we tested all models on the external imbalanced testing dataset. Although the AUROCs of all models were above 0.9, the AUPRCs of all models were below 0.5 (Fig. 5). The LR, ANN, and XGB models exhibited the best performance on the balanced and imbalanced testing datasets (Figs. 3, 5 and Table 2).

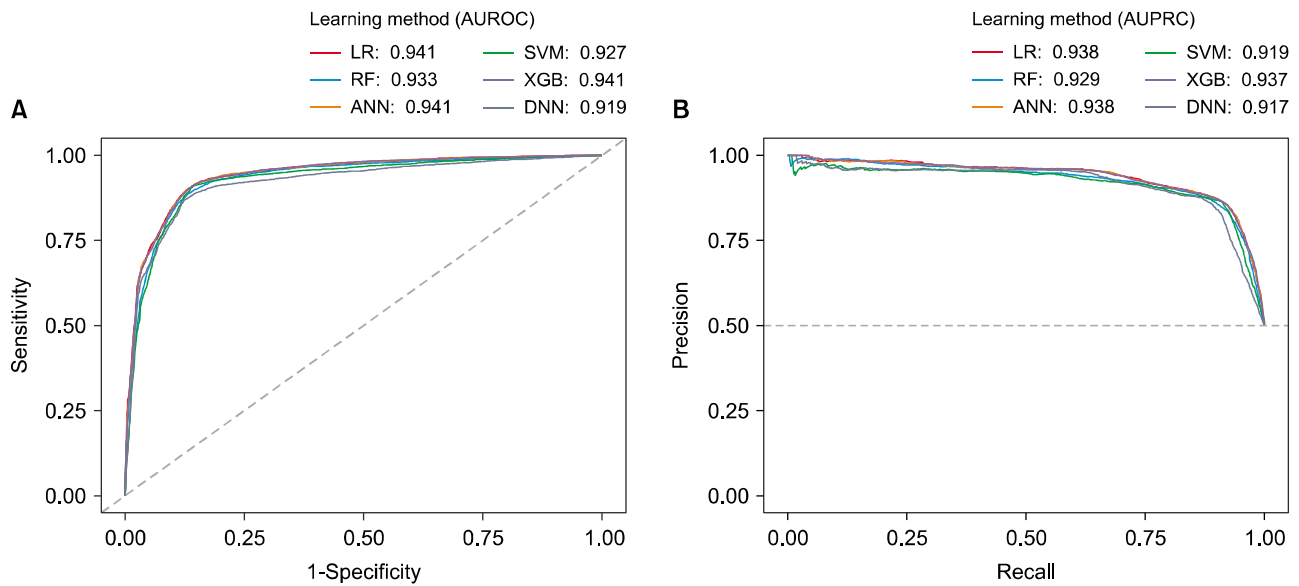


Fig. 3. ROC and PR curves plotted from internal balanced testing dataset. (A) ROC curves (B) PR curves.

ROC, receiver operating characteristic; PR, precision–recall; AUROC, area under the ROC curve; AUPRC, area under the PR curve; LR, logistic regression; RF, random forest; ANN, artificial neural networks; SVM, support vector machines; XGB, extreme gradient boosting; DNN: deep neural network.

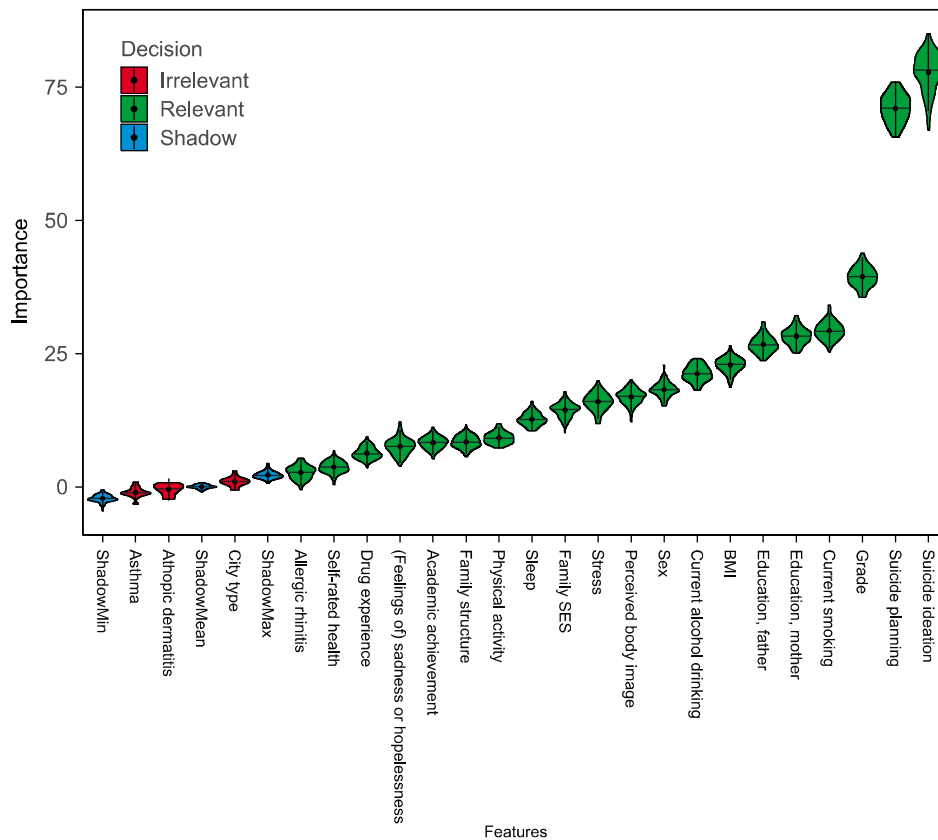


Fig. 4. Relative feature importance computed by Boruta algorithm. The blue violin plots correspond to the minimal, average, and maximum Z scores of a shadow attribute. The red and green violin plots represent the Z scores of rejected and confirmed attributes, respectively. The black dots and horizontal lines inside each violin plot represent the mean and median values, respectively. All features that received a lower relative feature importance than the shadow feature were defined as irrelevant for the prediction. For the training dataset, the irrelevant features (marked in red) were city type, asthma, and atopic dermatitis. BMI, body mass index; SES, socio-economic status.

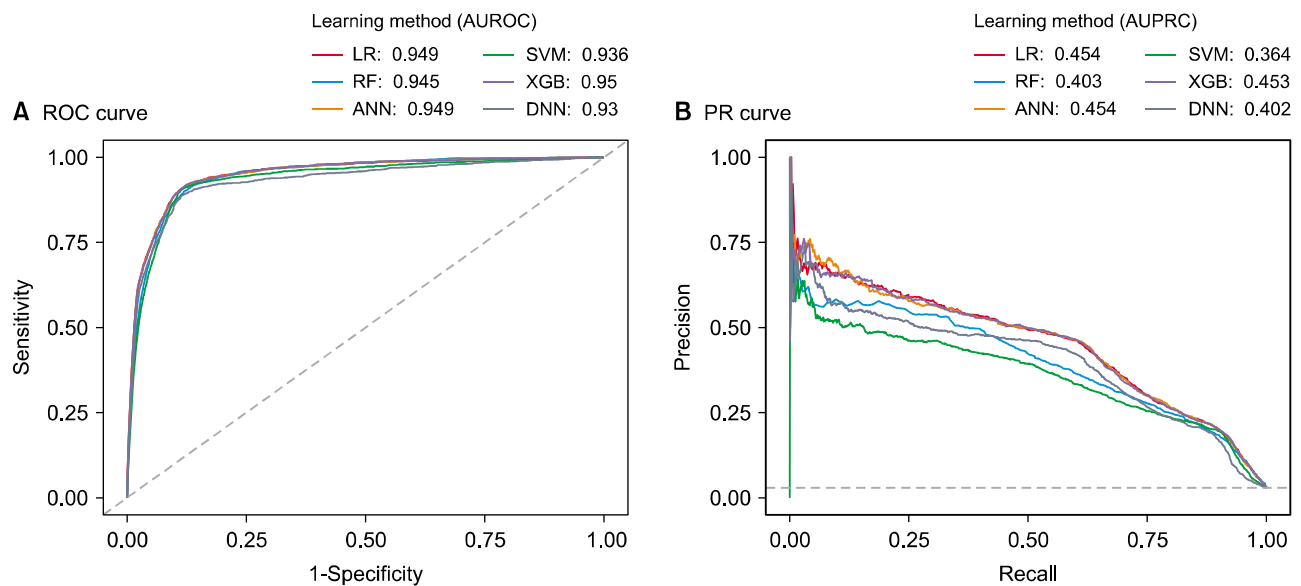


Fig. 5. Performance of learning models evaluated with external imbalanced testing dataset. (A) ROC curves (B) PR curves.

ROC, receiver operating characteristic; PR, precision–recall; AUROC, area under the ROC curve; AUPRC, area under the PR curve; LR, logistic regression; RF, random forest; ANN, artificial neural networks; SVM, support vector machines; XGB, extreme gradient boosting; DNN: deep neural network.

Table 2. Confusion matrix of prediction models in internal balanced and external imbalanced testing datasets

Dataset	Model	Accuracy	Sensitivity	Specificity	PPV	NPV	F1
Internal balanced	LR	0.88	0.91	0.85	0.86	0.91	0.89
	RF	0.88	0.90	0.85	0.86	0.89	0.88
	ANN	0.88	0.92	0.85	0.86	0.91	0.89
	SVM	0.88	0.91	0.86	0.86	0.91	0.89
	XGB	0.88	0.91	0.85	0.86	0.91	0.89
	DNN	0.87	0.89	0.86	0.86	0.88	0.87
External imbalanced	LR	0.97	0.60	0.98	0.46	0.99	0.52
	RF	0.97	0.51	0.98	0.42	0.99	0.46
	ANN	0.97	0.61	0.98	0.46	0.99	0.52
	SVM	0.95	0.64	0.96	0.31	0.99	0.42
	XGB	0.97	0.61	0.98	0.46	0.99	0.52
	DNN	0.97	0.58	0.98	0.43	0.99	0.50

PPV, positive predictive value; NPV, negative predictive value; LR, logistic regression; RF, random forest; ANN, artificial neural networks; SVM, support vector machines; XGB, extreme gradient boosting; DNN: deep neural network.

DISCUSSION

All six prediction models exhibited comparable accuracy, and the range of the AUROC results of 0.93 to 0.95 indicate excellent categorization in terms of predictive performance [24]. Existing studies on SA prediction using ML only in adolescent samples are extremely rare and the classification performance metrics have not been effectively described [14,15]. Some inevitable limitations exist in this study when compared to other methods in predict-

ing SA owing to several fundamental differences, such as the sample type (mostly patients), sample size, study design, and age group (mostly adults) [12,17,20,22,25–28]. Despite these limitations, our prediction models exhibited superior performance; several possible reasons are provided for this. First, to the best of our knowledge, no previous studies involved adolescent SA groups of over 15,000 individuals among a total sample size of almost half a million. Moreover, considering that suicidal behavior was more common among non-responders in pre-

vious studies [29,30], our web-based data collection method, resulting in a higher response rate of over 95% compared to the previous research range of 71% to 88%, may have affected SA prevalence and model performance. This methodological advantage may have contributed to reducing the probability of the underreporting of suicidal behavior that may occur during face-to-face interviews [31–34]. In terms of the research design, several approaches demonstrated high performance even in a longitudinal study [3]; however, Walsh *et al.* reported that prospective SA prediction accuracy improved going from 2 years to 1 week and that the predictor importance shifted over time [28]. Moreover, it was suggested that distal risk factors may exhibit limitations in reflecting the recent risk of suicide in a 10-year longitudinal study [3]. In line with this, the possibility of the limited test–retest reliability of SA in the prospective design and inconsistent reporting over time could have led to a performance difference with our results [30,35].

In Korea, despite suicide being the leading cause of adolescent death, access to treatment is very limited. According to a survey of mental illness in child and adolescent in 2019 [9], despite the fact that 17.6% of respondents reported SI, only 17% of that subset met with specialists, and Yang and Lee [36] reported that 86.7% of those who experienced adolescent depression did not receive treatment at that time. The treatment rate is very low compared to that in the United States, where most adolescents with SI (80.2%), SP (87.5%), and SA (94.2%) are treated in some way [21]. This difference may partially originate from the lack of the necessary budget for child and adolescent mental health policies or problems with accessibility of mental health services [8,9]. Obviously, adolescent SAs are not easy to predict, with 66% of SI groups not attempting suicide and even 20% of no SP groups attempting suicide [21]. Considering this context, our SA prediction model, which attempts to include all subjects regardless of SI or SP, can be applied in a valuable manner to prioritize high-risk groups and the efficient use of limited resources.

The relative feature importance calculated by the Boruta algorithm revealed grade as the third risk factor following SI and SP. In the multivariate LR analysis, compared to G6, G1, G2, and G3 sequentially exhibited a high association with an increased risk of SA. This can be understood in a similar context to the study of Nock *et al.*

[21], whereby the prevalence of SA remained very low ($< 1\%$) until 12 years of age and increased sharply through 15 years of age, as well as the work of Voss *et al.* [37], which demonstrated that suicidal behavior began to increase after the age of 10 years and increased most rapidly at the ages of 13 to 14 years. This indicates that the transition from late childhood to early adolescence is a critical period for a high risk of suicide, suggesting the need for sensitive screening.

The medical diagnosis of a disease is a well-known class-imbalanced ML task, in which the majority of the patients are healthy [38]. When applying ML to class-imbalanced datasets, learners will typically overestimate the majority group (for example, the non-SA group), leading to more frequent misclassification of the minority group (for example, the SA group). Although numerous solutions have been proposed for classification of imbalanced data, they are mostly related to data-level techniques [39]. Data-level techniques attempt to reduce the level of imbalance through various data sampling methods, such as undersampling the majority class, oversampling the minority class, or combining the over- and undersampling techniques (for example, SMOTE) [40]. Each method has its own set of advantages and disadvantages (such as a loss of valuable information in undersampling or increased training time in oversampling), suggesting that an evaluation of the sampling methods is required to determine an efficient method for subsequent analysis. As undersampling exhibited the shortest execution time without loss of performance in our randomly extracted dataset (Supplementary Fig. 2; available online), we used the undersampling method to rebalance the 2011–2017 KYRBS dataset.

Although the ROC curve is the most popular evaluation metric for binary classification, the interpretation of ROC curves requires special caution when using an imbalanced dataset. Because the ROC curve does not measure the effects of the negative class (for example, the non-SA group), it does not reflect the minority class effectively. Compared to the ROC curve, the PR curve provides a PPV, which is directly influenced by the class imbalance, thereby enabling a more reliable measurement of classification performance [41]. Using the AUPRC, we determined that the LR, ANN, and XGB models exhibited superior performance over the RF, SVM, and DNN models on the imbalanced testing dataset, implying that the PR

curve can provide a more realistic representation of the model performance in datasets with an imbalanced class distribution.

In recent years, deep learning methods have become increasingly popular as a tool for all healthcare analytics applications, particularly in medical image classification [42]. Their recent extensive application can be attributed to the increased availability of electronic health records as well as improvements in hardware and software [43-46]. Despite these advances and the usefulness of these methods for certain tasks such as classifying medical images, deep learning is not suitable for all clinical data problems. There are several reasons for this: 1) although the interpretability of deep learning has been improved, current deep learning models still behave as black boxes and fail to provide explanations for their predictions [47]; 2) the subject of deep learning with class-imbalanced data is understudied [48,49]; and 3) the majority of state-of-the-art deep learning models generally require millions of parameters and operations to achieve acceptable accuracy [50]. In our study, the DNN model provided a lower AUPRC than CML approaches such as the LR, RF, ANN, and XGB methods, suggesting that CML models can be trained faster and exhibit better overall performance compared to the DNN. Moreover, CML provides interpretable models, a factor which is important for clinical decision making, allowing clinicians to understand the relative importance of variables (for example, the *p* values and odds ratios in LR and feature importance in the Boruta algorithm) in the overall prediction. The appropriate application of CML may provide simpler and faster methods with interpretability for clinical data modeling.

Several limitations of the current study should be mentioned. First, owing to the cross-sectional nature of this study, the capability for causal inference was limited. Of course, even a longitudinal study design in SA prediction must address the ethical problem of making the decision whether or not to intervene when a subject is identified as being at high risk for SA. Nevertheless, a further prospective study should investigate the applicability of ML models to future SA prediction by transforming these web-based data into a cohort or longitudinal research design. Second, considering that the data used in this study were based on retrospective self-reports, as opposed to face-to-face interviews, the data may have been affected by recall bias and therefore could be vulnerable

to underreporting [51]. Third, although this study targeted representative adolescents on a large scale, because a school-based approach was used, adolescents outside of school, which constitutes approximately 1% to 1.7% of adolescents each year [52], or who did not meet the middle to high school student age requirement were not considered.

Adolescence is a critical period for the onset and intervention of suicidal behavior. Despite the fact that many high-risk adolescents remain untreated, many difficulties exist in predicting SA at this time, and few studies have been conducted on this topic. This study offers clinical significance by expanding on the applicability of existing SA prediction studies. The developed and validated SA prediction model using ML in this study could be applied to screening high-risk groups for SA in the entire adolescent population. In this manner, it is expected that the high-risk SA group will be prioritized so that limited resources can be introduced and treatment access can be increased.

■ Funding

This paper was supported by Wonkwang University in 2022.

■ Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

■ Author Contributions

Organized the project: Jae Seok Lim, Chan-Mo Yang, and Bung-Nyun Kim. Contributed to collecting the web-survey data and literature review: Ju-Won Baek. Contributed to data management, data analysis, and visualization: Jae Seok Lim. Contributed to the study conceptualization and analysis plan: Chan-Mo Yang. Contributed useful comments and to writing the manuscript: Sang-Yeol Lee. Contributed to drafting of the manuscript: Jae Seok Lim, Chan-Mo Yang, and Ju-Won Baek. All authors critiqued the work for intellectual content and approved it for submission.

■ ORCID

Jae Seok Lim	https://orcid.org/0000-0001-5755-7529
Chan-Mo Yang	https://orcid.org/0000-0002-4959-7595
Ju-Won Baek	https://orcid.org/0000-0002-4459-6534
Sang-Yeol Lee	https://orcid.org/0000-0003-1828-9992

Bung-Nyun Kim <https://orcid.org/0000-0002-2403-3291>

REFERENCES

- World Health Organization. *Preventing suicide: a global imperative* [Internet]. Geneva: World Health Organization; 2014 Aug 17 [cited at 2020 Jul 15]. Available from: <https://www.who.int/publications/i/item/9789241564779>.
- Centers for Disease Control and Prevention. *Web-based injury statistics query and reporting system* [Internet]. Atlanta: Centers for Disease Control and Prevention; 2017 [cited at 2020 Jul 15]. Available from: <https://www.cdc.gov/injury/wisqars/index.html>.
- Miché M, Studerus E, Meyer AH, Gloster AT, Beesdo-Baum K, Wittchen HU, et al. *Prospective prediction of suicide attempts in community adolescents and young adults, using regression methods and machine learning*. *J Affect Disord* 2020;265:570-578.
- Statistics Korea. *2019 Statistics on the youth* [Internet]. Daejeon: Statistics Korea; 2019 May 1 [cited at 2020 Jul 15]. Available from: <http://kostat.go.kr/portal/eng/pressReleases/13/3/index-board?bmode=read&bSeq=&aSeq=377344&pageNo=1&rowNum=10&navCount=10&currPg=&searchInfo=&sTarget=title&sTxt=>
- Centers for Disease Control and Prevention. *Youth risk behavior survey data summary & trends report 2007-2017* [Internet]. Atlanta: Centers for Disease Control and Prevention; 2018 [cited at 2020 Jul 15]. Available from: <https://www.cdc.gov/healthyyouth/data/yrbs/pdf/trendsreport.pdf>.
- Korea Centers for Disease Control and Prevention. *The 15th Korea youth risk behavior survey* [Internet]. Cheongju: Korea Centers for Disease Control and Prevention; 2019 [cited at 2020 Jul 15]. Available from: <http://www.kdca.go.kr/yhs/>.
- Carlton PA, Deane FP. *Impact of attitudes and suicidal ideation on adolescents' intentions to seek professional psychological help*. *J Adolesc* 2000;23:35-45.
- Roh S, Lee SU, Soh M, Ryu V, Kim H, Jang JW, et al. *Mental health services and R&D in South Korea*. *Int J Ment Health Syst* 2016;10:45.
- Korea Mental Health R&D Project. *Prevalence and risk factors of psychiatric disorders in child and adolescent population-school based research* [Internet]. Seoul: Korea Mental Health R&D Project; [cited at 2020 Aug 6]. Available from: http://www.mhrrd.re.kr/xe/?module=file&act=procFileDownload&file_srl=2764&sid=587cf4c29ddbde42a97a3c21803c016c&module_srl=137.
- 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;143:187-232.
- McArdle JJ, Ritschard G. *Contemporary issues in exploratory data mining in the behavioral sciences*. Taylor & Francis; 2013.
- 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.
- Kessler RC, Bossarte RM, Luedtke A, Zaslavsky AM, Zubizarreta JR. *Suicide prediction models: a critical review of recent research with recommendations for the way forward*. *Mol Psychiatry* 2020;25:168-179.
- Hill RM, Oosterhoff B, Do C. *Using machine learning to identify suicide risk: a classification tree approach to prospectively identify adolescent suicide attempters*. *Arch Suicide Res* 2020;24:218-235.
- Bae SM, Lee SA, Lee SH. *Prediction by data mining, of suicide attempts in Korean adolescents: a national study*. *Neuropsychiatr Dis Treat* 2015;11:2367-2375.
- Carson NJ, Mullin B, Sanchez MJ, Lu F, Yang K, Menezes M, et al. *Identification of suicidal behavior among psychiatrically hospitalized adolescents using natural language processing and machine learning of electronic health records*. *PLoS One* 2019;14:e0211116.
- Walsh CG, Ribeiro JD, Franklin JC. *Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning*. *J Child Psychol Psychiatry* 2018;59:1261-1270.
- Kim Y, Choi S, Chun C, Park S, Khang YH, Oh K. *Data resource profile: the Korea youth risk behavior web-based survey (KYRBS)*. *Int J Epidemiol* 2016;45:1076-1076e.
- Korea Centers for Disease Control and Prevention. *Korea youth risk behavior web-based survey* [Internet]. Cheongju: Korea Centers for Disease Control and Prevention; [cited at 2020 Aug 6]. Available from: <https://www.kdca.go.kr/yhs/home.jsp>.
- Husky MM, Olsson M, He JP, Nock MK, Swanson SA, Merikangas KR. *Twelve-month suicidal symptoms and use of services among adolescents: results from the National Comorbidity Survey*. *Psychiatr Serv* 2012;63:989-996.
- Nock MK, Green JG, Hwang I, McLaughlin KA, Sampson NA, Zaslavsky AM, et al. *Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: results from the National Comorbidity Survey Replication Adolescent Supplement*. *JAMA Psychiatry* 2013;70:300-310.
- Passos IC, Mwangi B, Cao B, Hamilton JE, Wu MJ, Zhang XY, et al. *Identifying a clinical signature of suicidality among patients with mood disorders: a pilot study using a machine learning approach*. *J Affect Disord* 2016;193:109-116.
- Kursa MB, Rudnicki WR. *Feature selection with the Boruta package*. *J Stat Softw* 2010;36:1-13.
- Šimundić AM. *Measures of diagnostic accuracy: basic definitions*. *EJIFCC* 2009;19:203-211.
- Hettige NC, Nguyen TB, Yuan C, Rajakulendran T, Baddour J, Bhagwat N, et al. *Classification of suicide attempters in schizophrenia using sociocultural and clinical features: a machine learning approach*. *Gen Hosp Psychiatry* 2017;47:20-28.
- 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;69:23-31.
27. Simon GE, Johnson E, Lawrence JM, Rossom RC, Ahmedani B, Lynch FL, et al. Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. *Am J Psychiatry* 2018;175:951-960.
 28. Walsh CG, Ribeiro JD, Franklin JC. Predicting risk of suicide attempts over time through machine learning. *Clin Psychol Sci* 2017;5:457-469.
 29. Svensson T, Inoue M, Sawada N, Iwasaki M, Sasazuki S, Shimazu T, et al. The association between complete and partial non-response to psychosocial questions and suicide: the JPHC Study. *Eur J Public Health* 2015;25:424-430.
 30. Mars B, Cornish R, Heron J, Boyd A, Crane C, Hawton K, et al. Using data linkage to investigate inconsistent reporting of self-harm and questionnaire non-response. *Arch Suicide Res* 2016;20:113-141.
 31. 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;35:239-250.
 32. 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;280:867-873.
 33. Greist JH, Gustafson DH, Stauss FF, Rowse GL, Laughren TP, Chiles JA. A computer interview for suicide-risk prediction. *Am J Psychiatry* 1973;130:1327-1332.
 34. Levine S, Ancill RJ, Roberts AP. Assessment of suicide risk by computer-delivered self-rating questionnaire: preliminary findings. *Acta Psychiatr Scand* 1989;80:216-220.
 35. Christl B, Wittchen HU, 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;10:253-263.
 36. Yang CM, Lee SY. Effect of untreated depression in adolescence on the suicide risk and attempt in male young adults. *Korean J Psychosom Med* 2020;28:29-35.
 37. Voss C, Ollmann TM, Miché M, Venz J, Hoyer J, Pieper L, et al. Prevalence, onset, and course of suicidal behavior among adolescents and young adults in Germany. *JAMA Netw Open* 2019;2:e1914386.
 38. Mostafizur Rahman M, Davis DN. Addressing the class imbalance problem in medical datasets. *Int J Mach Learn Comput* 2013;3:224-228.
 39. Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. *J Big Data* 2019;6:27.
 40. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res* 2002;16:321-357.
 41. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015;10:e0118432.
 42. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med* 2019;25:24-29.
 43. Kruse CS, Stein A, Thomas H, Kaur H. The use of electronic health records to support population health: a systematic review of the literature. *J Med Syst* 2018;42:214.
 44. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: a system for large-scale machine learning. In: *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation*; Nov 2-4, 2016; Savannah, USA. p. 265-283.
 45. Collet F. Keras [Internet]. 2015 [cited at 2020 Jul 16]. Available from: <https://keras.io>.
 46. Chetlur S, Woolley C, Vandermersch P, Cohen J, Tran J, Catanzaro B, et al. cuDNN: efficient primitives for deep learning [Internet]. *arXiv*; 2014 Oct 3 [updated 2014 Dec 18; cited at 2020 Jul 16]. Available from: <https://arxiv.org/abs/1410.0759v3>.
 47. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 2019;1:206-215.
 48. Wang S, Liu W, Wu J, Cao L, Meng Q, Kennedy PJ. Training deep neural networks on imbalanced data sets. In: *2016 International Joint Conference on Neural Networks (IJCNN)*; Jul 24-29, 2016; Vancouver, Canada.
 49. Khan SH, Hayat M, Bennamoun M, Sohel FA, Togneri R. Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE Trans Neural Netw Learn Syst* 2018;29:3573-3587.
 50. Sejnowski TJ. The unreasonable effectiveness of deep learning in artificial intelligence. *Proc Natl Acad Sci U S A* 2020;117:30033-30038.
 51. Ritter PL, Stewart AL, Kaymaz H, Sobel DS, Block DA, Lorig KR. Self-reports of health care utilization compared to provider records. *J Clin Epidemiol* 2001;54:136-141.
 52. Korean Educational Development Institute. *Statistics on adolescents outside school in Korea* [Internet]. Jincheon: Korean Educational Development Institute; 2019 Mar 6 [cited at 2020 Aug 6]. Available from: <https://kess.kedi.re.kr/post/6678743>.