RAC-001 https://doi.org/10.5080/61upk.ozt412

AN ALTERNATIVE APPROACH FOR DETECTING PROBLEMATIC ALCOHOL USE: DEVELOPING A NEW RISK SCORE MODEL FOR UNIVERSITY STUDENTS USING AI-ASSISTED MACHINE LEARNING

Şükrü Alperen Korkmaz[1], **Pınar Mutlu**[2], **Sibel Oymak**[3], **Gamze Çan**[3]

[1]*Çanakkale Onsekiz Mart University, Faculty of Medicine, Department of Psychiatry, Çanakkale, Türkiye*
[2]*Çanakkale Onsekiz Mart University, Center for Combating Addiction Application and Research, Çanakkale, Türkiye*
[3]*Çanakkale Onsekiz Mart University, Faculty of Medicine, Department of Public Health, Çanakkale, Türkiye*

BACKGROUND AND AIM: Problematic alcohol use in university students is a critical public-health issue because it is associated with academic disruption, high-risk behaviors, and psychological burden. While established tools (e.g., AUDIT, CAGE, RAPS4-QF) are widely used, student populations may exhibit early functional harms (academic decline, dangerous situations, financial strain, mood and sleep problems) before classical dependence patterns become apparent. Therefore, a brief, multidimensional, and practically implementable risk score that reflects both consumption intensity and real-world impairment may improve early detection and targeted intervention.

We aimed to develop and validate an AI-assisted Alcohol Risk Score for university students, and to examine its reliability, convergent validity with established instruments, and machine-learning (ML) classification performance for actionable risk stratification.

METHODS (Ethics Committee Approval must be obtained and the number should be specified.): A cross-sectional survey was conducted in collaboration with a university Center for Combating Addiction. The questionnaire captured a wide range of alcohol-related domains, including demographics and contextual variables, alcohol use patterns (lifetime, past year/month/week; weekly frequency; standard drink amount; binge drinking), and alcohol-related psychosocial/health factors (e.g., academic impact, dangerous situations, help-seeking, mood changes, perceived harms, and intention to quit). The overall conceptual framework deliberately extended beyond consumption to include functional impairment, psychological correlates, and readiness-to-change indicators—dimensions particularly salient in student populations. Ethics approval was obtained from the institutional ethics committee (05.06.2024, 2024/05-08).

AI-assisted item selection and model construction:

The initial candidate pool included 72 raw items, consolidated and refined into 59 unique items after removing conceptually duplicate or non-informative candidates. These items were evaluated using an AI-assisted scoring approach (OpenAI ChatGPT 4.5 API) based on predefined criteria reflecting screening suitability, clinical relevance, clarity, and discriminatory potential. From this AI-ranked pool, a preliminary set of 20 items was created, followed by expert review by four authors; items with strong consensus and relevance were retained, resulting in a final 15-item risk score model.

Scoring and risk stratification:

The model includes binary (0/1) items and two frequency/intensity items scored from 0–4, generating a total score range of 0–27. Risk categories were defined as 0–5 (low risk), 6–12 (moderate risk), and 13–27 (high risk), designed to create actionable strata for screening and triage.

Validation and machine learning evaluation:

Convergent validity was assessed via associations with established screening tools (AUDIT, CAGE, and RAPS4-QF). Internal consistency reliability was tested using Cronbach's alpha and McDonald's omega. For classification performance, multiple ML algorithms were evaluated (logistic regression, support vector machines, random forest, and k-nearest neighbors). Models were assessed using accuracy and class-wise F1 scores; additional evaluation included cross-validation and ROC-AUC estimates via a one-vs-rest strategy. Feature contributions were examined using permutation importance to identify the most informative items in final risk score.

RESULTS:A total of 599 university students were included (mean age 20.84 ± 3.27 years), with 66.8% women; 37.2% were first-year students and 42.6% were enrolled in medical school. Past-month alcohol consumption was reported by 45% of participants; 9.8% consumed alcohol three or more days per week during same period. Binge drinking (≥5 standard drinks at a time; ≥4 for women) was observed in 6.7% overall, with a significantly higher rate in men than women.

Risk score distribution and clinical gradients:

The new 15-item risk score produced a mean of 3.65 ± 3.64, with men scoring higher than women (4.37 ± 3.75 vs 3.27 ± 3.54, p<0.001). Based on predefined thresholds, 74.6%

(n=447) were classified as low risk, 22.9% (n=137) as moderate risk, and 2.5% (n=15) as high risk. Proportion of women was higher in the low-risk category, while men were relatively more represented in the moderate-risk group; the high-risk category proportion was similar across genders.

Crucially, risk categories demonstrated meaningful clinical differentiation: moderate/high-risk students reported more psychological complaints, higher smoking rates, and greater family alcohol consumption (all p<0.001). The high-risk category showed striking elevation in harm-related outcomes, including academic impairment, alcohol-related medical needs, dangerous situations or legal issues, inability to attend classes, missing exams/assignments, neglecting professional responsibilities, financial difficulties, sleep disturbances, alcohol-related mood changes, and perceiving alcohol as a serious problem (all p<0.05). Notably, binge drinking in the last month was 80% in the high-risk group vs 4.9% in the low-risk group (p<0.001), supporting the model's ability to detect a clinically severe subgroup despite its small prevalence.

Reliability and convergent validity:

The risk score demonstrated strong internal consistency (Cronbach's alpha = 0.811; McDonald's omega = 0.831). Convergent validity was robust, with high correlation to AUDIT (r = 0.861, p<0.001) and substantial correlations to RAPS4-QF (r = 0.793, p<0.001) and CAGE (r = 0.631, p<0.001), indicating alignment with established screening constructs while preserving broader functional coverage.

Machine learning performance:

Among tested algorithms, logistic regression achieved the best overall performance, with 93.5% accuracy and strong class discrimination (F1 low = 0.97; F1 moderate = 0.94; F1 high = 1.00). Other algorithms also performed well but were consistently lower (SVM accuracy 90.9%; random forest 88.3%; KNN 80.5%). ROC-AUC analyses demonstrated high separability, particularly for low risk (AUC=0.96), while maintaining strong discrimination for moderate (AUC=0.88) and high risk (AUC=0.93). Permutation importance highlighted the model's most influential indicators: feelings of guilt/regret, average drinking amount, financial difficulties, and physical health impact, supporting both interpretability and face validity.

CONCLUSIONS:This study developed and validated a brief, AI-assisted 15-item alcohol risk score that captures multidimensional nature of problematic alcohol use among university students by combining consumption intensity with functional, psychosocial, and harm-related indicators. The model demonstrated high reliability, strong convergent validity with established tools, and excellent ML classification performance, particularly using logistic regression. Importantly, risk categories were not merely statistical strata; they mapped onto clinically meaningful gradients of academic impairment, dangerous behaviors, psychological complaints, and other adverse outcomes—features that are central to early intervention in student populations.

Because the tool is concise, interpretable, and grounded in real-world consequences, it offers a pragmatic pathway for scalable screening, early detection, and triage in university health services. Beyond identifying severe dependence signals, this model is positioned to detect earlier-stage risk profiles where prevention and brief interventions may yield the highest impact. Future research should test external validity across diverse universities and longitudinally examine whether baseline risk scores predict incident harms and service utilization.

Keywords: risky alcohol use, alcohol use disorder, artificial intelligence, machine learning, logistic regression, screening

Table 1. The 15-item new risk model

Items	Score Range
1. Have you used alcohol in the last year?	No = 0, Yes = 1
2. How many days a week did you use alcohol in the last month?	None = 0, 1–2 days = 1, 3–4 days = 2, 5+ days= 3
3. How much alcohol do you consume on average at a time (standard drink)?	None = 0, 1–2 standard = 1, 3–4 standard = 2, 5–9 standard = 3, 10 + standard = 4
4. Have you ever felt that you needed medical (psychological or physical) help with alcohol use?	No=0, Yes=1
5. Do you experience physical health problems due to alcohol use?	No=0, Yes=1
6. Have you ever been in dangerous situations after drinking alcohol (e.g., accidents, injuries, offences, unprotected sex)?	No=0, Yes=1
7. Does your alcohol use affect your attendance at classes/work?	No=0, Yes=1
8. Do you experience financial challenges due to alcohol use?	No=0, Yes=1
9. Do you think there is a relationship between alcohol use and your depression or anxiety symptoms?	No=0, Yes=1

Table 1. Continued

Items	Score Range
10. Do you think you should give up alcohol in the future?	No=0, Yes=1
11. How many times this year have you been unable to stop drinking alcohol after starting to drink? (binge drinking)	None = 0, 1–2 hands = 1, 3–5 hands = 2, 6–10 hands = 3, 10 + hands = 4
12. How many times this year have you felt guilty or regretful after drinking alcohol?	None = 0, 1–2 hands = 1, 3–5 hands = 2, 6–10 hands = 3, 10 + hands = 4
13. Do you feel like you are doing something bad or wrong because you drink?	No=0, Yes=1
14. In the last year, has a friend or family member told you about things you have said or done while drinking that you cannot remember (memory loss, loss of consciousness, etc.)?	No=0, Yes=1
15. Have you ever drunk alcohol first thing in the morning to wake up and feel refreshed or to relieve hangover symptoms?	No=0, Yes=1
Total score	*0- 27*

Table 2. The new risk model's performance

Algorithm	Accuracy	F1 (low)	F1 (medium)	F1 (high)
Logistic Regression	93.5%	0.97	0.94	1.00
SVM	90.9%	0.93	0.85	1.00
Random Forest	88.3%	0.91	0.78	1.00
KKN	80.5%	0.86	0.59	1.00

SVM: Support Vector Machines; KNN: K-Nearest Neighbors.