EUR REV MED PHARMACOL SCI
2025; 29: 528-538

PREDICTING SUICIDAL BEHAVIOR: A MACHINE LEARNING MODEL

A.-M. VEJNOVIĆ[1,2], V. TATALOVIĆ[3,4], M. VUJOVIĆ[5], N. JANKOVIĆ[5], A. DICKOV[1,2], V.M. JOVIN[6], V. KNEŽEVIĆ[1,2], I. KELEČEVIĆ[1,2]

• • •

[1]Department of Psychiatry and Psychological Medicine, Faculty of Medicine, University of Novi Sad, Novi Sad, Serbia
[2]Clinic for Psychiatry, University Clinical Center of Vojvodina, Novi Sad, Serbia
[3]Department of Plastic and Reconstructive Surgery, Faculty of Medicine, University of Novi Sad, Novi Sad, Serbia
[4]Clinic for Plastic and Reconstructive Surgery, University Clinical Center of Vojvodina, Novi Sad, Serbia
[5]Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
[6]Department of Pharmacology, Toxicology, and Clinical Pharmacology, Faculty of Medicine, University of Novi Sad, Novi Sad, Serbia

Corresponding Author
Ana-Marija Vejnović, MD; e-mail: ana-marija.vejnovic@mf.uns.ac.rs

ABSTRACT – ***Objective:*** Timely response and identification of risk factors are critical components of suicide prevention, enabling appropriate intervention. By applying machine learning to clinical data analysis, predictive models can be developed to classify patients into high- and low-risk groups, which can help prevent loss of life. This study aims to develop a machine learning model to categorize patients based on their risk of suicide attempts.

Materials and Methods: We analyzed data from 301 respondents hospitalized at the Clinic of Psychiatry for either suicidal behavior or non-suicidal reasons. They were divided into two groups according to the presence of suicidal behavior. By using machine learning methods, we analyzed the influence of 18 observed features on the development of suicidal behavior. A predictive model was developed to classify individuals into two risk categories – high and low – for the observed population. The k-Nearest Neighbors (kNN) algorithm was used to train the model.

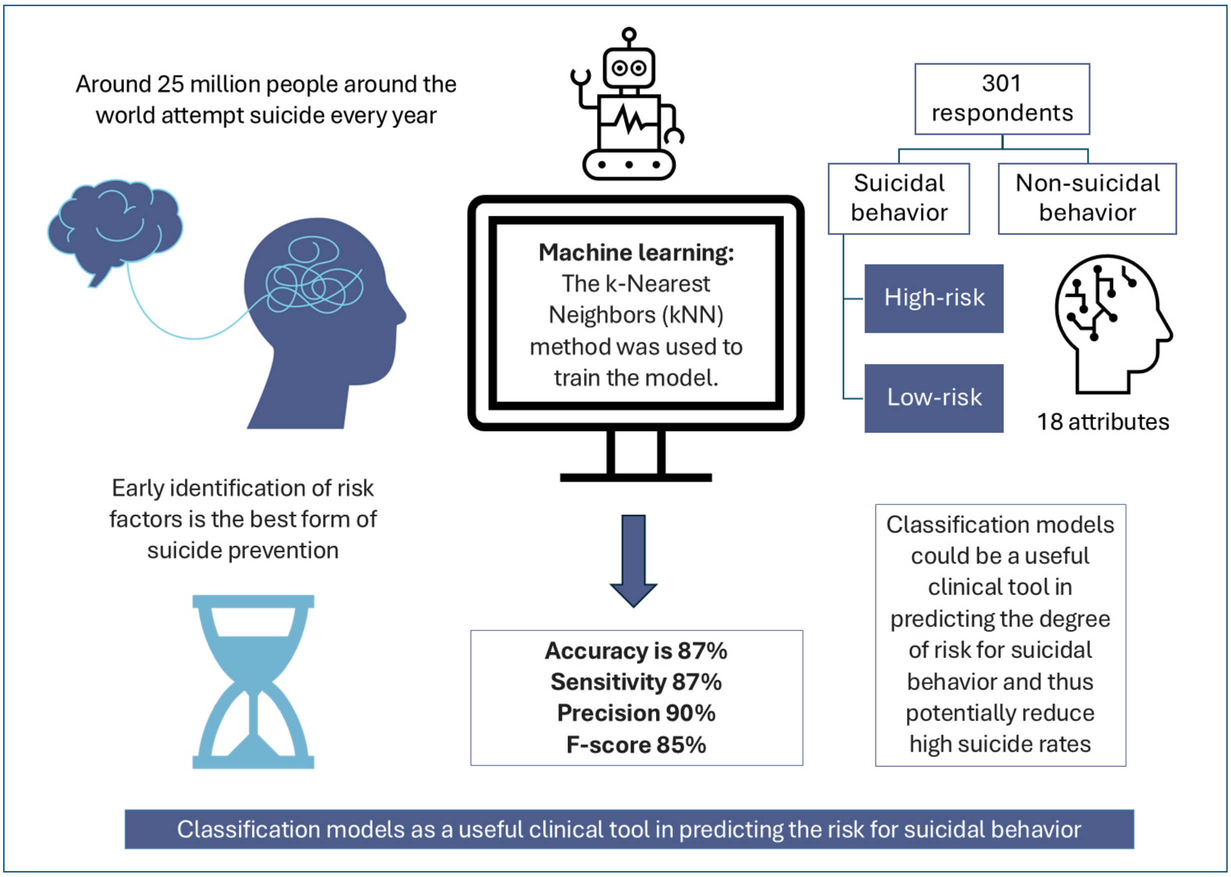
Results: The kNN-based model achieved a classification accuracy of 87%, with sensitivity at 87%, precision at 90% and an F-score of 85% for the tested sample.

Conclusions: Early identification of risk factors for suicidal behavior remains the most effective strategy for prevention. Classification models may serve as valuable clinical tools for assessing suicide risk, potentially contributing to a reduction in suicide rates in Vojvodina. Expanding the database could improve the functions of the obtained classifier and enable its introduction into medical practice.

KEYWORDS: Suicide attempt, Algorithm, Classifier, Risk.

INTRODUCTION

Suicide is a significant public health problem and one of the leading causes of premature death among young adults. Every year, around 25 million people around the world attempt suicide without fatal outcomes[1-3]. Recent research shows that there is a rising trend in both suicide rates and suicide attempts[1,4]. In 2018, the suicide rate in Serbia was 13.5 per 100,000 inhabitants, while the rate in Vojvodina was higher at 16.4 per 100,000 inhabitants[5]. Apart from representing a social and economic problem, suicide attempts are also one of the most significant risk factors for repeated suicide attempts and completed suicides[6-11]. It is estimated that 10-15% of patients who attempt suicide will eventually die by suicide[12].



Graphical Abstract. We analyzed data from 301 respondents hospitalized at the Clinic of psychiatry for either suicidal behavior or non-suicidal reasons. They were divided into two groups according to the presence of suicidal behavior. Using machine learning methods, we analyzed the influence of 18 observed features on the development of suicidal behavior. A predictive model was developed to classify individuals into two risk categories - high and low, for the observed population. The k-Nearest Neighbors (KNN) algorithm was used to train the model.

A suicide attempt is an act of intentional self-harm that does not result in death. Studies[3,5,13] have shown that individuals who attempt suicide are usually younger females and often suffer from personality disorders, depression, or alcoholism. Suicidal behavior may also be influenced by genetic, epidemiological, and personal factors, such as family history of suicide, previous suicide attempts, marital status, employment, level of education, financial status, physical comorbidities, prior psychiatric treatment, and psychological status[3,5,13].

The most common types of suicide attempts are low-lethal methods such as deliberate self-poisoning with medications, followed by self-harm with sharp objects, hanging, jumping from a height, and strangulation[14]. In Serbia, most acute poisonings among adults are caused by the intentional drug ingestion (primarily psychotropic and cardiotropic drugs) for suicidal purposes, although they can also occur accidentally[14].

The scope and severity of this problem have aroused the interest of many authors and indicate the importance of finding a screening model to prevent suicide attempts and fatalities[15]. In clinical practice, psychiatric interviews are supplemented by standardized clinical instruments to assess suicide risk more precisely. Numerous scales and questionnaires are used both in clinical settings and scientific research. One widely recognized tool is the Columbia Suicide Severity Rating Scale (C-SSRS), which was designated the gold standard for suicide risk assessment by the U.S. Food and Drug Administration in 2012[16]. In our region, Dr. Miklos Biro made a significant contribution to the understanding of the phenomenon of suicide, who developed a suicide risk assessment scale adapted to the local population[17]. The advantages of using scales are increasing the objectivity of assessment and systematized data collection; however, the disadvantages of their use are unreliability for long-term monitoring and dependence on the human factor in terms of expert training for adequate application[16].

Prediction of suicide attempts is still far from a reliable level[6,15,18,19]. Despite the great effort to improve prediction over the last 50 years, a meta-analysis of 365 studies showed that known risk factors only marginally improve predictive accuracy compared to chance[15]. The main reason for poor prediction is that most studies analyzed isolated factors as predictors[6,15]. More accurate prediction of suicide attempts may require complex combinations of numerous risk factors. Accurate prediction likely requires multifactorial models that traditional statistical methods struggle to accommodate. Although statistical significance was obtained using these simpler methods of data analysis, the data had limited clinical significance or were not adequately utilized. The emergence of artificial intelligence offers possible solutions to this problem.

Standard analytical methods can confirm that individual features used in the database affect the output size, but none of them can be sufficient to predict suicide risk. For this reason, more complex machine learning methods are employed to achieve higher precision through feature combinations – specifically to classify individuals into high- or low-risk groups for suicide attempts.

Machine learning enables the cross-analysis of numerous complex variables, creating algorithms that test a wide range of relationships between observed factors and suicide attempts. We assume that such algorithms can more reliably predict suicide risk and serve as practical tools in clinical practice. These models could alert healthcare professionals to elevated risk levels based on data that is entered into the electronic health record as part of the standard procedure of taking anamnesis and clinical examination. The use of machine learning is a useful tool in many areas of medicine, supported by research across various domains[20-22].

The aim of this study is to develop a predictive model for assessing suicide attempt risk among people hospitalized at the Clinic of psychiatry of the University Clinical Center of Vojvodina using the k-Nearest Neighbor (kNN) method.

MATERIALS AND METHODS

Training set and their features

A retrospective observational study was conducted using data extracted from the electronic database of the Clinical Information System of the University Clinical Center of Vojvodina (UCCV). The study period spanned from January 1, 2020, to December 31, 2021 and included a total of 305 patients, divided into two groups:

1) Group 1, high-risk group for suicide attempts, included 167 subjects with a history of suicide attempts. Inclusion criteria were hospitalization at the UCCV Clinic of Psychiatry following a non-fatal suicide attempt during the examined period. Exclusion criteria included: isolated intoxication with alcohol as the only agent, outpatient examination without indication for hospitalization, and more than 50% missing data across the observed parameters.

2) Group 2, low-risk group for suicide attempts, included 138 control patients without a history of suicidal behavior. Inclusion criteria were hospitalization at the UCCV Clinic of Psychiatry during the same examined period for non-suicidal reasons and no prior history of suicidal behavior. This group was selected as a randomized sample of subjects hospitalized during the corresponding period. The exclusion criterion was any information of a suicide attempt or suicidal thoughts.

To identify patients with a history of suicide attempts, diagnostic codes from the 10th revision of the International Classification of Diseases (ICD-10) were used. These included codes related to acute self-poisoning and intentional self-harm by various methods: F10-9.0, X60-X84.9. The following characteristics variables were analyzed: a) sociodemographic: gender, age, employment status, education level, marital status, number of household members, number of children, place of residence, financial status, and adverse childhood experience (e.g., abuse/neglect); b) biological: alcohol use, previous suicide attempts, prior psychiatric treatment, previous hospitalizations at the Clinic of Psychiatry, family history of psychiatric diseases, physical health conditions, and leading and accompanying diagnoses according to ICD-10; c) psychological: personality characteristics (aggression/impulsivity and emotional instability) and d) general features related to current hospitalization: day and month of admission, length of hospital stay, method of suicide attempt, and stated reason for suicide attempt.

Ethics statement

The research was approved by the University Clinical Center of Vojvodina Ethics Committee (approval number 00-230, dated December 29, 2021) and was conducted following the Helsinki Declaration as revised in 2013.

Missing data

The original database consisted of 305 patients, including 167 classified as high-risk and 138 as low risk for suicide attempts. Each patient was initially described by 25 features/attributes. Furthermore, the database was modified in order to approach the creation of a mathematical model. Due to a high proportion of missing data (over 40%) for several variables (educational status, financial status, history of abuse/neglect, aggressiveness/impulsivity, emotional instability, and positive family history), these features were excluded in model development. Although the method of suicide attempt was considered during data analysis, it was not used in model training. For features with a lower percentage of missing values, data imputation strategies were applied: introducing a new category (for variables: reason for suicide attempt, accompanying diagnosis, marital status, employment status, and previous treatment) or replacing the missing data with a mean value (for variables: number of children, number of household members, number of hospitalizations, and number of suicide attempts). Additionally, four respondents were excluded from the database due to missing values in key variables: gender, place of residence, and number of hospitalization days. These cases, which represented less than 1% of the total sample, were excluded to avoid introducing potential bias or irregularities, as their inclusion would not significantly enhance model performance.

After preprocessing, the final database consisted of 301 samples described by 18 attributes, namely: gender, age, employment status, marital status, number of hospitalization days, number of suicide attempts, number of hospitalizations, previous psychiatric treatment, physical comorbidities, number of household members, number of children, month of suicide attempt, day of the week, place of residence, alcohol use, leading diagnosis, accompanying diagnosis and reason for suicide attempt.

Preparation for modeling

Python programming language was used to implement methods of automatic classification for categorizing individuals into high- or low-risk groups for suicide attempts. The SciPy, NumPy, and Pandas libraries were used for data analysis, matplotlib and seaborn for visual representation. The k-Nearest Neighbors (kNN) algorithm was implemented using scikit-learn, a widely used machine learning library.

Boxplot is one of the statistical tools used to graphically represent the distribution of data. By analyzing the boxplot, the dynamic and interquartile range, median, and outliers can be observed. The dynamic range represents the difference between the maximum and minimum values of the observed attribute, while the interquartile range represents the range in which the middle 50% of the values are found[23]. Outliers represent those values that deviate significantly from other values within the observed attribute.

Statistical analysis

Descriptive statistics were used to summarize the dataset: categorical variables were presented as frequencies and percentages, while numerical variables were described using mean values and standard deviations. Comparison of the values of numerical features between the two groups was performed using the Student's *t*-test, while testing the difference in frequencies of attributive features was performed using the chi-square (c^2) test. A *p*-value of less than 0.05 was considered statistically significant. Statistical data analyses were performed using the computer program JASP (Jeffrey's Amazing Statistics Program), version 0.16.1.

Creating a prediction model

The method used to solve the classification problem is the k-Nearest Neighbors (kNN) algorithm. The obtained results are presented and analyzed using the confusion matrix, which provides information on the accuracy, precision, sensitivity, and F-measure of the trained model. This matrix serves as a tool for assessing the quality of the model by illustrating how reliably it can predict the class of an unknown sample[23]. It shows the relationship between predicted and actual sample classes and is presented in Table I.

True Positive (TP) refers to samples where both actual and predicted labels are 1 (in this research, patients classified as high-risk for suicide attempts). True Negative (TN) denotes samples where both the actual and predicted labels are 0 (indicating low-risk patients). False Positive (FP) represents samples with an actual label of 0 but a predicted label of 1, meaning low-risk patients are incorrectly classified as high-risk. Conversely, False Negative (FN) refers to samples with an actual label of 1 but a predicted label of 0, indicating high-risk patients incorrectly classified as low risk. Based on these values, basic measures of classifier evaluation such as accuracy, precision, sensitivity, and specificity can be calculated. Sensitivity (TP Rate/Recall) shows the proportion of correctly

Table I. Confusion matrix.

		True values	
	The entire population	**A real positive**	**A real negative**
Predicted values	Predicted value of positive	TP	FP
	Predicted value of negative	FN	TN

TP: true positive, FP: false positive, FN: false negative, TN: true negative.

predicted high-risk samples out of all actual high-risk cases. Precision shows the proportion of correctly predicted high-risk samples out of all samples classified by the model as high-risk. Accuracy shows the percentage of successfully classified samples in relation to the total number of samples. F-score is the harmonic mean of precision and sensitivity, providing a balanced measure of model performance[23].

The relationship between sensitivity and specificity is graphically represented using the Receiver Operating Characteristic (ROC) curve. More precisely, the dependence of the share of correctly classified positives in all positives (True Positive Rate - TPR) on the share of false positives in all negatives (False Positive Rate - FPR) is graphically represented. It is the same as the value of specificity subtracted from 1, for different decision thresholds.

To quantify the effectiveness of the classifier based on the ROC curve, we calculated the Area Under the Curve (AUC)[23].

For sample classification, the kNN algorithm was used. This method assigns a class label to an unlabeled sample based on the class membership of its nearest neighbors – those with the highest degree of similarity with the observed sample[23]. When applied in binary classification, the value of k is typically chosen as an odd number to ensure a clear majority vote for class assignment. The kNN algorithm belongs to the group of lazy learning methods, meaning it defers computation until classification is required[23,24].

An important part of creating machine learning algorithms is the choice of hyperparameters. The two most important hyperparameters for the kNN classifier are k, the observed number of neighbors, and the distance metric, the method used to measure similarity between samples. The Hamming distance was used for binary vectors. Two binary vectors of equal length are compared, where the Hamming distance is the number of different binary pairs. It is expected that the hamming metric will prove to be the best when optimizing the parameters because the categories have been transformed into dummy variables due to the needs of the algorithm. For a feature with k categories, k − 1 binary variables were created, with 1 indicating the presence of a category and 0 its absence.

To create the classification model, the database was split into a training set and an independent test set, with the latter comprising 20% of the total sample. The test was reserved for final evaluation to ensure a more objective assessment of the model's generalization capability. Additionally, we extracted a validation set in order to determine the optimal hyperparameters. The training set is used to train the classifier for different combinations of hyperparameters, which are then evaluated against the validation set. After determining the optimal hyperparameters, we performed the final evaluation of the classifier on the test set, which was not used either in the training or for determining the hyperparameters, ensuring an objective assessment of the classifier that gives significantly more realistic results. To divide the training set into training and validation subsets, we used the cross-validation method with N folds. In each iteration, N – 1 folds were used for training, and one fold for validation. In this way, the problem of using some samples only for validation purposes is overcome, as well as the problem of fitting one of the sample subsets. In other words, each sample will at some point be part of both the training and the validation set[25]. The final confusion matrix for each hyperparameter combination was obtained by summing the confusion matrices from all cross-validation iterations.

RESULTS

Between January 1, 2020, and December 31, 2021, a total of 167 individuals were hospitalized at the UCCV Clinic of Psychiatry due to suicide attempts. During the same period, we collected data for 138 control patients, with no history of suicidal behavior. In Group 1, 46.1% were male (77/167), 53.9% (90/167) were female, with a mean age of 39.44 ± 16.23 years. In Group 2, 43.5% (60/138) were male, 56.5% (78/138) were female, with a mean age of 40.56 ± 15.98 years. There is no significant difference in these characteristics between the groups ($p>0.05$).

Table II shows data that most often appear within the high-risk and low-risk categories for suicide attempt.

The largest number of respondents tried to commit suicide by less lethal methods, with self-poisoning being the most prevalent (77.2%, 129/167), followed by self-harm with a sharp object (14.4%, 24/167), hanging (4.8%, 8/167), and jumping from a height (3.6%, 6/167).

Since self-poisoning proved to be the most common method of suicide attempts, we analyzed the values of several variables in this group of respondents in more detail. Respondents in this group were most often hospitalized at least once in the earlier period, except for 11/129 patients (the self-poisoning subgroup) for whom it was their first hospitalization at the time of the examination. Notably, many respondents in this group had a history of repeated self-poisoning. Only 15/129 patients had no history of suicide attempts in the past.

Table II. Presentation of features that occurred most frequently within the risk factors of high and low risk groups (modal value).

	High risk (Group 1)	Low risk (Group 2)
Gender	Female	Female
Work status	Unemployed	Unemployed
Marital status	Single	Single
Earlier treatment	Yes	Yes
A month of hospital admission	September	October
Day of the week	Monday	Thursday
Place	City	City
History of alcohol abuse	Yes	No
Leading diagnosis	F1	F2

People who attempted suicide were most often treated for the leading diagnosis of mental and behavioral disorders due to alcohol use (F1) (68/138, 49.3%), while neurotic, stress-related and somatoform disorders (F4) (26/138, 18.8%) and mood (affective) disorders (F3) (15/138, 10.9%) occurred somewhat less frequently.

The highest frequency of suicide attempts was recorded in September, January, and November, respectively, while the most common days are Monday, Friday, and Wednesday, respectively.

Given the high proportion of patients with a history of alcohol abuse and dependence, the high-risk group was further subdivided into individuals with and without such a history. Comparative analysis was then conducted between these subgroups.

Among high-risk individuals without a history of harmful alcohol abuse/dependence, the most common leading diagnosis is the F4 category. Additionally, this subgroup most lacked physical comorbidities. Regarding previous suicide attempts, it can be noted that the most common value for the number of previous suicide attempts is zero, indicating no prior attempts, while a single previous suicide attempt was observed somewhat less frequently.

Analyzing respondents who belong to the high-risk patient category and who have previously attempted suicide, the rate is twice as high as that of people without a history of alcohol abuse or dependence. It can be noted that more than 10% of respondents from the high-risk category, who have a history of alcohol abuse/dependence, have attempted suicide 3 times in the past. The most common leading diagnosis in high-risk patients who consume alcohol is diagnosis F1 (occurs in 63 patients), while, for example, diagnoses F4 and F6 (disorders of adult personality and behavior) occur much less frequently.

Due to the large number of missing data, several characteristics were not included in the construction of the suicidal behavior prediction model, but we used standard statistical methods to represent them. Within the group of high-risk patients, 57 respondents have the characteristics of an emotionally unstable personality based on psychological testing, which is statistically significantly higher than the control group (c^2=8.469, p<0.01, Kramer's V coefficient=0.33), while this information is missing for 110 respondents. Out of 167 subjects, 29 show aggressive traits, which is not statistically significantly different from the control subjects (c^2=1.078, p>0.05, Kramer's V coefficient=0.09). For only one respondent, we have recorded information about an adverse childhood experience (experience of abuse/neglect), while for all other respondents, this information is unknown or not recorded in medical history.

Suicidal behavior prediction model

By observing the accumulated confusion matrices for different combinations of hyperparameters, accuracy and F-measure were calculated to guide the selection of optimal parameters for the final model. Hamming distance metric and k=5 yielded the best performance, achieving 94% accuracy and a 94% F-measure. These results indicate a high level of classification reliability. The final confusion matrix is presented in Table III.

After determining the optimal hyperparameters, the model was trained using the entire training set and tested on a previously separat-

Table III. Final confusion matrix.

	True positive	True negative
Predicted value of positive	104	0
Predicted value of negative	15	121

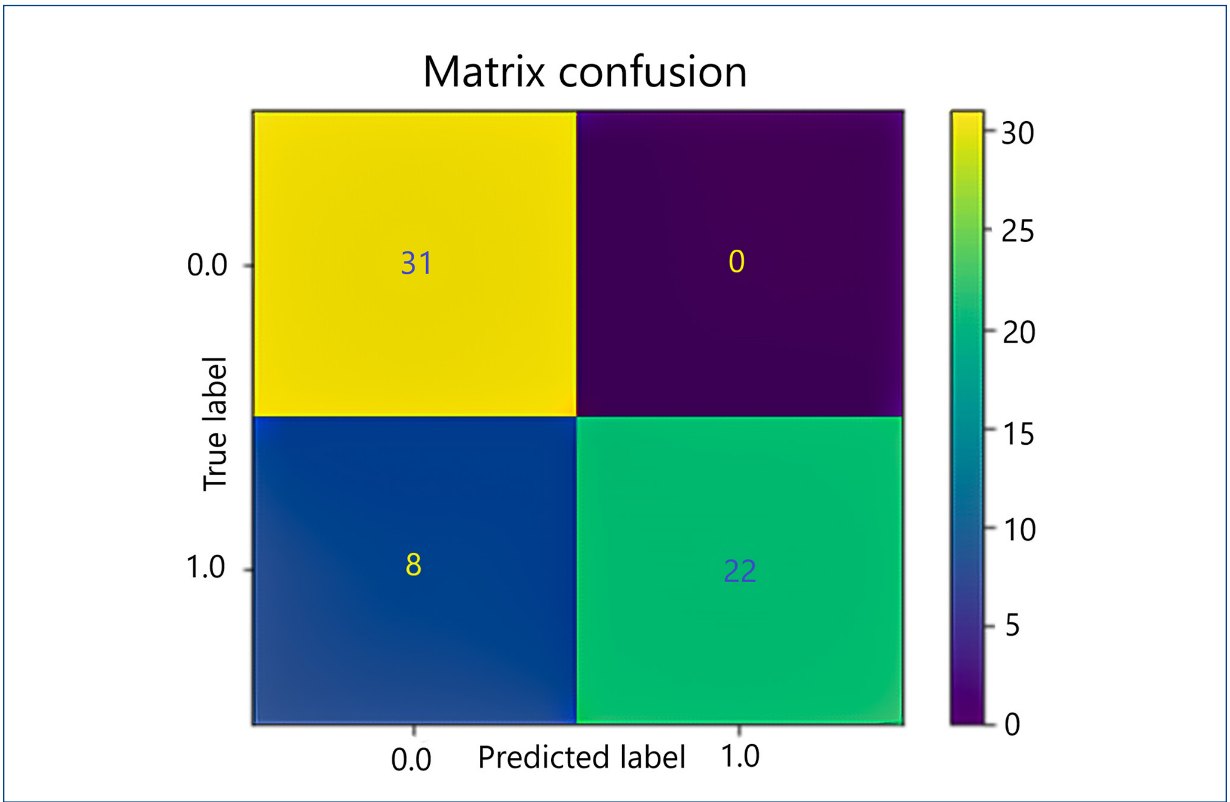


Figure 1. Final confusion matrix obtained on an independent test set.

ed independent test set for objective evaluation. The confusion matrix and the results obtained using the kNN classifier are presented above (Figure 1).

Using the kNN algorithm, the classification accuracy is 87%, sensitivity 87%, precision 90% and an F measure 85% for the examined sample (Table IV).

The quantification of the classifier's success is presented in Figure 2. It can be seen that the AUC value for our classifier is 0.91.

DISCUSSION

Research on the prediction of suicide attempts remains limited in our region, according to the assessment of several local authors[3,26,27]. The significance of this study lies in our effort to apply the predictive capabilities of machine learning algorithms to assess the risk of suicide attempts. Although earlier research by Franklin et al[15] showed that model predictions up to 2017 were poor, with AUC values=0.56-0.58, and that no single risk factor was superior in risk estimation, our model achieved an AUC of 0.91 – a similar value obtained in later studies conducted across

Table IV. Values obtained for the accuracy, precision, sensitivity and F-measure of the trained model, obtained based on the confusion matrix.

	Percentage
Accuracy	87%
Precision	90%
Sensitivity	87%
F-Measure	85%

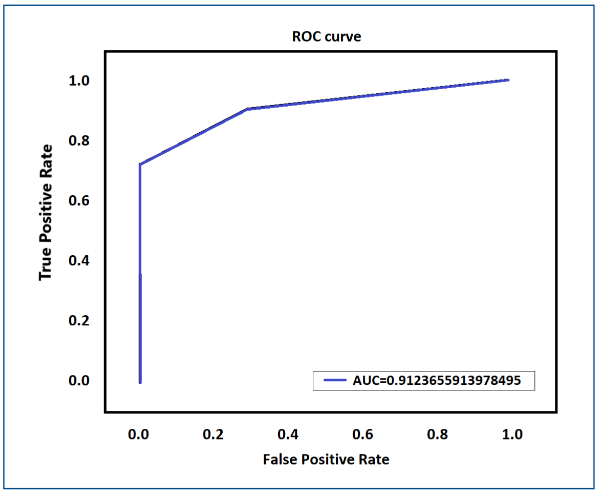


Figure 2. Display of ROC curve and AUC values; ROC: Receiver-operating characteristic curve; AUC: area under the curve.

the US[28], as well as in Walsh et al[29]. Authors who used the kNN classifier generally obtained worse classifier performance than us[30]. However, they also used other classifiers such as Random Forest (RF)[28] and Support Vector Machine (SVM)[30]. A study conducted in Sweden used neural networks as a classifier that achieved a slightly lower result than ours, with an AUC value of 0.87[31]. Compared to a previous study[32] conducted with the kNN classifier, we obtained a higher AUC value (0.91 *vs.* 0.72), but a slightly lower accuracy (87% *vs.* 94.7%).

A more accurate assessment could help reduce the suicide rate, which is the highest rate in Vojvodina, and at one point, it was even twice as high as in other regions of Serbia[33].

Research has shown that the implementation of the suicide prediction models has significantly improved the identification of individuals who may be at risk of attempting suicide. However, a critical area for future research remains in determining the time when a person might commit suicide[34]. On the other hand, the contribution of the model to predict who may be at risk is valuable, but it is insufficient to assess the immediate risk[29].

Although the field of artificial intelligence is an increasingly popular and promising area, its shortcomings are noted by other authors[35]. Analyzing suicide attempts is especially important, since prior attempts are repeatedly cited in the literature as a major predictor of future suicidal behavior[6-11]. Research conducted in Spain showed that patients who committed suicide in almost 32% of cases have a history of at least one previous suicide attempt[36], which supports the fact that patients after a failed suicide attempt can conclude how they can increase the chance of a fatal outcome and apply it.

The difference in the distribution between the genders indicates that women attempt suicide more frequently than men, which is in agreement with data from the literature[3,5,27,37,38]. The claim that women try to commit suicide more often while men succeed in it more often, which is confirmed in the literature and marked as "gender paradox"[39], is explained by the fact that women more often use low-lethal methods of attempts such as self-poisoning or cutting veins, while men more often use high-lethal methods such as hanging, using firearms or jumping from a height[30]. Although our results show that city dwellers more often carry out suicide attempts, a large study conducted in the territory of England, Canada, America, and Australia shows that suicides are more characteristic of rural areas[40]. Psychiatric services are less available in rural areas, and we assume that people from rural areas would turn to a psychiatrist less frequently in case of suicidal thoughts due to more pronounced stigmatization in smaller areas. This may be the reason for the higher rate of suicide committed in rural areas, according to the results of other studies, in contrast to suicide attempts, which were the subject of our research[41]. The most frequently cited reasons for suicide attempts were partner and family problems, which were the most frequent motives in studies conducted in America[42], Iran[43], and Serbia[3].

In our study, the most common day of suicide attempts was Monday, which can be seen in most other studies as well[44,45]. Given that the majority of patients who attempted suicide were unemployed, this is a possible consequence of feelings of loneliness at the beginning of the week and the working day, which the authors from Glasgow cite as one of the risk factors for suicide attempts[46]. September was found to be the most common month of attempts, which is consistent with studies, where autumn and spring are cited as the seasons with the most recorded suicide attempts[45]. What could influence this distribution are bioclimatic factors, such as temperature and amount of sunlight. Serotonin is one of the neurotransmitters responsible for the hypothesis of mood disorders, and its level in the body depends on climate changes and exposure to light. On the other hand, the peak of suicidality can also be attributed to a social factor, specifically an increase in workload and stress before the summer months and vacations, as well as a sudden return to high demands and jobs that were less during the summer[47,48].

In 2010, Molnar et al[49] found that disorders of adult personality and behavior (F6) were the most common leading diagnosis in patients with suicide attempts. In our study this was not the case, but the most common diagnosis was mental and behavioral disorders due to alcohol use (F10). It can be explained by the fact that we also took into account self-poisoning with alcohol in combination with other substances for suicide attempts, which are also the most frequently recorded types of suicide attempts in our study[49]. According to some authors, overdoses of alcohol and "recreational drugs" are also included in intentional self-poisoning and suicide attempts[50]. However, in a certain number of patients, there was still a characteristic of emotional instability and aggressiveness, which supports the results of Molnar et al[49] regarding the occurrence of F6 diagnosis in such patients.

Patients who attempted suicide were most often hospitalized for an average of two days, while control group patients were hospitalized for an average of seven days. In studies conducted in other countries, the average length of hospitalization after a suicide attempt is significantly longer and lasts 36 days[51]. Through a more detailed analysis of the data, it was determined that discharges after a short hospitalization were made at the patient's personal request, which the legal regulations within the Act on the Protection of Persons with Mental Disorders allow in Serbia. In this way, patients can stop treatment against the doctor's advice if it is determined that they do not currently threaten their health, safety, or life.

Patients who attempted suicide in a higher percentage report that they consume alcohol compared to the control group of patients, which was also shown in studies conducted in Brazil in 2020[52], Iran in 2020[43], but also in those studies conducted much earlier in Norway[53]. Under the influence of alcohol, disinhibition occurs, and the ability to reason is reduced during drunkenness, which is one of the possible causes of suicide attempts in substance abuse[54]. Another explanation may be the association of alcoholism with depression. Studies confirm the frequent occurrence of dual diagnosis, and the etiology of their occurrence goes in both directions[55].

Limitations

The main limitation of this study is that, in addition to the small sample, we considered only non-fatal outcomes of suicidal behavior to create the model. In this way, our analysis did not include objectively positive respondents who committed suicide and for whom we can say with greater certainty that they had suicidal intent. Another limitation of our study is that only one classifier was used. The kNN algorithm is simple to implement, and since it uses local information for classification, the procedure is highly adaptable, and the complexity of the actual class distributions is not important. The disadvantage of the algorithm is that it is computationally demanding, so increasing the dimensionality would increase the computational complexity. If the database were to increase, the kNN classifier would not be the right choice, because the whole process would be slower. Other classifiers could improve the prediction performance. Algorithms that could be used when working with larger amounts of data include neural networks and machines based on support vectors. Support vector machines show good performance on larger data sets. Also, one of the advantages is that the results are stable and repeatable. On the other hand, neural networks are a powerful tool for solving classification problems, but for neural networks to show their quality, a large amount of data is needed. The time required to optimize a neural network model can be long because there are no clear guidelines for the selection of parameters when forming a neural network, but this problem is solved by a combination of different parameters until satisfactory results are achieved. One more limitation of this study could be the division of subjects into groups based on X diagnosis. Namely, the ICD-10 classification within this code does not differentiate between self-harm with suicidal intent and without such intent. In the research, we classified each respondent with a code of self-harm into a group with a suicide attempt, although this information was not clearly indicated in the medical history for everyone.

CONCLUSIONS

Based on the results obtained, it can be concluded that the developed model could be capable of predicting the level of suicide attempt risk among individuals with mental disorders requiring hospitalization, achieving an accuracy of 87%. This finding highlights the potential of machine learning-based classification models as valuable tools in clinical practice, enabling early identification of high-risk individuals. By integrating predictive models into routine psychiatric assessment, healthcare professionals may be better equipped to intervene proactively, thereby contributing to a reduction in suicide rates. Expanding the database could enhance the classifier's functions and enable the application of other algorithms that demonstrated advantages over the one used in this study, thereby introducing the model into medical practice.

In addition to a larger number of analyzed patients that would serve for better training of the model, it is necessary to point out the need for a better interview structure that would include more characteristics, which have already been shown to be significant risk factors through literature, but also for the purpose of finding new risk factors.

Conflict of Interest

The authors declare that they have no conflict of interest to disclose.

Funding

The study received no funding.

Authors' Contributions

AMV, VT, MV, NJ: Conceptualization, study design, article writing (definitive version). AMV, VT: Definition of conclusions. AMV, VT, VMJ, AD: Criticism and review. All authors: Literature review, supervision, article writing (first draft).

Informed Consent
The requirement for informed consent was waived by the Ethics Committee, as the study involved retrospective analysis of anonymized data obtained from the institutional database under the authority of the University Clinical Center of Vojvodina.

Ethics Approval
The research was approved by the University Clinical Center of Vojvodina Ethics Committee (approval number 00-230, dated December 29, 2021) and was conducted following the Helsinki Declaration as revised in 2013.

ORCID ID
A.-M. Vejnovic: 0000-0002-2467-542X
V. Tatalovic: 0009-0008-1942-706X
A. Dickov: 0000-0002-1954-6746
V. Mijatovic Jovin: 0000-0002-3200-0676
V. Knezevic: 0000-0002-3934-0710

Availability of Data and Materials
De-identified aggregate data are available upon reasonable request; individual-level data are not publicly shared due to patient privacy.

AI Disclosure
During the final editing process, this manuscript underwent proofreading and grammar checking using Microsoft Copilot, an AI-powered writing assistant. The authors reviewed and approved all AI-suggested edits to ensure accuracy and integrity of the scientific content.

REFERENCES

1) Centers for Disease Control and Prevention. WISQARS (web-based injury statistics query and reporting system) Injury Center CDC [Internet]. Centers for Disease Control and Prevention. Available from: https://www.cdc.gov/injury/wisqars/. [Accessed on February 12, 2022].
2) World Health Organization. Suicide data [Internet]. Available from: https://www.who.int/teams/mental-health-and-substance-use/suicide-data. [Accessed on February 12, 2022].
3) Tatalovic V, Vejnovic AM, Bjelan S, Bjelan S, Djozic I. Clinical profile of a person with experience of deliberate self-poisoning. OA Lib J 2024; 11: e11476.
4) Duffy ME, Twenge JM, Joiner TE. Trends in mood and anxiety symptoms and suicide-related outcomes among U.S. undergraduates, 2007-2018: evidence from two national surveys. J Adolesc Health 2019; 65: 590-598.
5) Institute for Public Health of Vojvodina. World Suicide Prevention Day [Internet]. Novi Sad: IZJZV; 2020 Sep 10. Available from: http://www.izjzv.org.rs/?lng=&cir=&link=3-15-2059. Serbian. [Accessed on February 12, 2022].
6) Ribeiro JD, Franklin JC, Fox KR, Bentley KH, Kleiman EM, Chang BP, Nock MK. Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. Psychol Med 2016; 46: 225-236.
7) Suominen K, Isometsä E, Suokas J, Haukka J, Achte K, Lönnqvist J. Completed suicide after a suicide attempt: a 37-year follow-up study. Am J Psychiatry 2004; 161: 562-563.
8) Suokas J, Suominen K, Isometsä E, Ostamo A, Lönnqvist J. Long-term risk factors for suicide mortality after attempted suicide - Findings of a 14-year follow-up study. Acta Psychiatr. Scand 2001; 104: 117-121.
9) Hayashi N, Igarashi M, Imai A, Osawa Y, Utsumi K, Ishikawa Y, Tokunaga T, Ishimoto K, Harima H, Tatebayashi Y, Kumagai N, Nozu M, Ishii H, Okazaki Y. Psychiatric disorders and clinical correlates of suicidal patients admitted to a psychiatric hospital in Tokyo. BMC Psychiatry 2010; 10: 109.
10) Runeson B, Haglund A, Lichtenstein P, Tidemalm D. Suicide risk after nonfatal self-harm. J Clin Psychiatry 2015; 77: 240-246.
11) Shah MMA, Sajib MWH, Arafat SMY. Demography and risk factor of suicidal behavior in Bangladesh: A cross-sectional observation from patients attending a suicide prevention clinic of Bangladesh. Asian J Psychiatr 2018; 35: 4-5.
12) Bjornaas MA, Hovda KE, Heyerdahl F, Skog K, Drottning P, Opdahl A, Jacobsen D, Ekeberg O. Suicidal intention, psychosocial factors and referral to further treatment: A one-year cross-sectional study of self-poisoning. BMC Psychiatry 2010; 10: 58.
13) Deveci A, Taskin EO, Erbay Dundar PE, Demet M, Kaya E, Ozmen E, Dinc G. Prevalence of suicide ideation and attempt in Manisa city centre. Tur J of Psy 2005; 16: 170-178.
14) Rezic T. [Poison Control Center Yearbook]. National Poison Control Center. Belgrade: Military Medical Academy; 2018. Serbian.
15) Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, Musacchio KM, Jaroszewski AC, Chang BP, Nock MK. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol Bull 2017; 143: 187-232.
16) Giddens JM, Sheehan KH, Sheehan DV. The Columbia-Suicide Severity Rating Scale (C-SSRS): has the "gold standard" become a liability? Innov Clin Neurosci 2014; 11: 66-80.
17) Biro M. [Suicide - psychology and psychopathology]. Belgrade: Nolit; 1982. Serbian.
18) Bentley KH, Franklin JC, Ribeiro JD, Kleiman EM, Fox KR, Nock MK. Anxiety and its disorders as risk factors for suicidal thoughts and behaviors: A meta-analytic review. Clin Psychol Rev 2016; 43: 30–46.
19) Chang BP, Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Nock MK. Biological risk factors for suicidal behaviors: a meta-analysis. Transl Psychiatry 2016; 6: e887.
20) Wang CH, Lv CY, Lin YF, Zhang WH, Tang XL, Zhao LX. Effect of Esketamine on perioperative anxiety and depression in women with systemic tumors based on big data medical background. Eur Rev Med Pharmacol Sci 2024; 28: 1797-1811.
21) Kim J, Kim SC, Kang D, Yon DK, Kim JG. Classification of Alzheimer's disease stage using machine learning for left and right oxygenation difference signals in the prefrontal cortex: a patient-level, single-group, diagnostic interventional trial. Eur Rev Med Pharmacol Sci 2022; 26: 7734-7741.
22) Al-Remawi M, Ali Agha ASA, Al-Akayleh F, Aburub F, Abdel-Rahem RA. Artificial intelligence and machine learning techniques for suicide prediction: Integrating dietary patterns and environmental contaminants. Heliyon 2024; 10: e40925.

23) Nosek T, Brkljac B, Despotovic D, Secujski M, Loncar-Turukalo T. [Practicum in machine learning]. Novi Sad; 2020. p. 1-3, 19-22, 31-35, 79-82. Serbian.
24) Rajaguru H, Prabhakar SK. KNN classifier and K-means clustering for robust classification of epilepsy from EEG signals A detailed analysis. Hamburg: Diplomica Verlag; 2017: 21.
25) Crnojevic V. [Shape recognition for engineers]. Novi Sad: Faculty of technical sciences Publishing; 2014. 266-268. Serbian.
26) Kelecevic I, Vejnovic AM, Mijatovic Jovin V. Deliberate self-poisoning and suicidality: a qualitative analysis of the factors that influence repetitive suicidal behavior. Hosp Pharm 2024; 11: 1546-1554.
27) Mijatović Jovin V, Prodanović D, Samojlik I, Petković S, Drvendžija Z, Vejnović AM, Dickov A, Vuković Ercegović G, Perković Vukćević N, Gvozdenović N. Repetitive suicide attempts by poisoning in Vojvodina, Serbia - Psychotropic drugs as main causes. Eur Rev Med Pharmacol Sci 2022; 26: 6090-6097.
28) García de la Garza Á, Blanco C, Olfson M, Wall MM. Identification of suicide attempt risk factors in a national US survey using machine learning. JAMA Psychiatry 2021; 78: 398-406.
29) Walsh CG, Ribeiro JD, Franklin JC. Predicting risk of suicide attempts over time through machine learning. Clin Psychol Sci 2017; 5: 457-469.
30) Wang N, Fan L, Shivtare Y, Badal V, Subbalakshmi K, Chandramouli R, Lee E. Learning models for suicide prediction from social media posts. In: Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access; Association for Computational Linguistics; 2021. p.87-92.
31) Fernandes AC, Dutta R, Velupillai S, Sanyal J, Stewart R, Chandran D. Identifying Suicide Ideation and Suicidal Attempts in a Psychiatric Clinical Research Database using Natural Language Processing. Sci Rep 2018; 8: 7426.
32) Parreco J, Rattan R. Machine learning models for prediction of reinjury after penetrating trauma. JAMA Surg 2018; 153: 184-186.
33) Dedić G. Gender differences in suicide in Serbia within the period 2006-2010. Vojnosanit Pregl 2014; 71: 265-270.
34) 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.
35) Pigoni A, Delvecchio G, Turtulici N, Madonna D, Pietrini P, Cecchetti L, Brambilla P. Machine learning and the prediction of suicide in psychiatric populations: a systematic review. Transl Psychiatry 2024; 14: 140.
36) Goñi-Sarriés A, Blanco M, Azcárate L, Peinado R, López-Goñi JJ. Are previous suicide attempts a risk factor for completed suicide? Psicothema 2018; 30: 33-38.
37) Ghoreishi SA, Mousavinasab N. Systematic review of researches on suicide and suicide attempt in Iran. Iran J Psychiatry 2008; 14: 115-121.
38) O'Connor RC, Wetherall K, Cleare S, Eschle S, Drummond J, Ferguson E, O'Connor DB, O'Carroll RE. Suicide attempts and non-suicidal self-harm: national prevalence study of young adults. BJPsych Open 2018; 4: 142-148.
39) Aaltonen KI, Isometsä E, Sund R, Pirkola S. Risk factors for suicide in depression in Finland: first-hospitalized patients followed up to 24 years. Acta Psychiatr Scand 2019; 139: 154-163.
40) Barry R, Rehm J, de Oliveira C, Gozdyra P, Kurdyak P. Rurality and risk of suicide attempts and death by suicide among people living in four English-speaking high-income countries: a systematic review and meta-analysis. Can J Psychiatry 2020; 65: 441-447.
41) Reccord C, Power N, Hatfield K, Karaivanov Y, Mulay S, Wilson M, Pollock N. Rural-urban differences in suicide mortality: an observational study in Newfoundland and Labrador, Canada. Can J Psychiatry 2021; 66: 918-928.
42) LaCroix JM, Colborn VA, Hassen HO, Perera KU, Weaver J, Soumoff A, Novak LA, Ghahramanlou-Holloway M. Intimate partner relationship stress and suicidality in a psychiatrically hospitalized military sample. Compr Psychiatry 2018; 84: 106-111.
43) Farzaneh E, Amani F, Sadeghi-Movahhed F, Mataei K. Factors affecting unsuccessful suicide among women referred to emergency room of Ardabil City Hospital, Ardabil, Iran. Int J Med Toxicol Forensic Med 2020; 10: 27412.
44) Kposowa AJ, D'Auria S. Association of temporal factors and suicides in the United States, 2000–2004. Soc Psychiatry Psychiatr Epidemiol 2010; 45: 433-445.
45) Beauchamp GA, Ho ML, Shan Y. Variation in suicide occurrence by day and during major American holidays. J Emerg Med 2014; 46: 776-781.
46) McClelland H, Evans JJ, Nowland R, Ferguson E, O'Connor RC. Loneliness as a predictor of suicidal ideation and behaviour: a systematic review and meta-analysis of prospective studies. J Affect Disord 2020; 274: 880-896.
47) Woo JM, Okusaga O, Postolache TT. Seasonality of suicidal behavior. Int J Environ Res Public Health 2012; 9: 531-547.
48) Yu J, Yang D, Kim Y, Hashizume M, Gasparrini A, Armstrong B, Honda Y, Tobias A, Sera F, Vicedo-Cabrera AM, Kim H, Íñiguez C, Lavigne E, Ragettli MS, Scovronick N, Acquaotta F, Chen B, Guo YL, de Sousa Zanotti Stagliori Coelho M, Saldiva P, Zanobetti A, Schwartz J, Bell ML, Diaz M, De la Cruz Valencia C, Holobâcă I, Fratianni S, Chung Y. Seasonality of suicide: a multi-country multi-community observational study. Epidemiology and Psychiatric Sciences. Cambridge University Press 2020; 29: e163.
49) Molnar S, Mihanovic M, Grah M, Kezic S, Filakovic P, Degmecic D. Comparative study on gene tags of the neurotransmission system in schizophrenic and suicidal subjects. Coll Antropol 2010; 34: 1427-1432.
50) Joksovic D, Segrt Z, Vucinic S. [Acute poisoning with chemical substances - basic principles of diagnosis, treatment and registration of acute poisoning]. Belgrade: Military Medical Academy, National Poison Control Center; 2006. Serbian.
51) Todorov L, Vulser H, Pirracchio R, Thauvin I, Radtchenko A, Vidal J, Guigui P, Limosin F, Lemogne C. Suicide attempts by jumping and length of stay in general hospital: a retrospective study of 225 patients. J Psychosom Res 2019; 119: 34-41.
52) Marcon G, Massaro Carneiro Monteiro G, Ballester P, Cassidy RM, Zimerman A, Brunoni AR, von Diemen L, Hauck S, Passos IC. Who attempts suicide among medical students? Acta Psychiatr Scand 2020; 141: 254-264.
53) Rossow I, Romelsjo A, Leifman H. Alcohol abuse and suicidal behaviour in young and middle aged men: differentiating between attempted and completed suicide. Addiction 1999; 94: 1199-1207.
54) Jadhav KS, Boutrel B. Prefrontal cortex development and emergence of self-regulatory competence: the two cardinal features of adolescence disrupted in context of alcohol abuse. Eur J Neurosci 2019; 50: 2274-2281.
55) Kuria MW, Ndetei DM, Obot IS, Khasakhala LI, Bagaka BM, Mbugua MN, Kamau J. The association between alcohol dependence and depression before and after treatment for alcohol dependence. ISRN Psychiatry 2012: 482802.