PLOS ONE

RESEARCH ARTICLE

Machine learning prediction of dropping out of outpatients with alcohol use disorders

So Jin Park[1,2], Sun Jung Lee[1,2], HyungMin Kim[1,2], Jae Kwon Kim[1,2], Ji-Won Chun[1,2], Soo-Jung Lee[3], Hae Kook Lee[3], Dai Jin Kim[4]*, In Young Choi[1,2]*

1 Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul, South Korea, **2** Department of Biomedicine & Health Sciences, College of Medicine, College of Medicine, The Catholic University of Korea, Seoul, South Korea, **3** Department of Psychiatry, College of Medicine, The Catholic University of Korea, Seoul, Korea, **4** Department of Psychiatry, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, South Korea

* iychoi@catholic.ac.kr (IYC); kdj922@catholic.ac.kr (DJK)

Citation: Park SJ, Lee SJ, Kim H, Kim JK, Chun J-W, Lee S-J, et al. (2021) Machine learning prediction of dropping out of outpatients with alcohol use disorders. PLoS ONE 16(8): e0255626. https://doi.org/10.1371/journal.pone.0255626

Editor: Khanh N.Q. Le, Taipei Medical University, TAIWAN

Received: March 2, 2021

Accepted: July 19, 2021

Published: August 2, 2021

Data Availability Statement: Data cannot be shared publicly because permission to do so was not obtained from the institutional review board of the Catholic University of Korea and the participants. We also did not include the sharing of data in our informed consent procedures. For data access requests, interested researchers can contact the institutional review board of the Catholic University of Korea, for example, via https://eirb.cmcnu.or.kr/irb.do.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded

Abstract

Background

Alcohol use disorder (AUD) is a chronic disease with a higher recurrence rate than that of other mental illnesses. Moreover, it requires continuous outpatient treatment for the patient to maintain abstinence. However, with a low probability of these patients to continue outpatient treatment, predicting and managing patients who might discontinue treatment becomes necessary. Accordingly, we developed a machine learning (ML) algorithm to predict which the risk of patients dropping out of outpatient treatment schemes.

Methods

A total of 839 patients were selected out of 2,206 patients admitted for AUD in three hospitals under the Catholic Central Medical Center in Korea. We implemented six ML models—logistic regression, support vector machine, k-nearest neighbor, random forest, neural network, and AdaBoost—and compared the prediction performances thereof.

Results

Among the six models, AdaBoost was selected as the final model for recommended use owing to its area under the receiver operating characteristic curve (AUROC) of 0.72. The four variables affecting the prediction based on feature importance were the length of hospitalization, age, residential area, and diabetes.

Conclusion

An ML algorithm was developed herein to predict the risk of patients with AUD in Korea discontinuing outpatient treatment. By testing and validating various machine learning models, we determined the best performing model, AdaBoost, as the final model for recommended use. Using this model, clinicians can manage patients with high risks of discontinuing treatment and establish patient-specific treatment strategies. Therefore, our model can

by the Korea government (MSIT) (No. 2019R1A5A2027588). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

potentially enable patients with AUD to successfully complete their treatments by identifying them before they can drop out.

Introduction

According to a 2016 Korean epidemiological survey on mental illness, the lifetime prevalence of alcohol use disorders (AUDs), including alcohol dependence and abuse, was 12.2% (18.1% for men and 6.4% for women), which is the highest among mental disorders [1]. AUDs result in significant economic losses, various social problems such as alcohol-related crimes and accidents, and physical diseases such as alcohol-induced physical complications and alcohol-related dementia [2–4].

AUD is a disease with a higher recurrence rate than that of other mental illnesses [5–7]. To prevent recurrence, the disorder must be managed over a long time without stopping the treatment at all [8, 9]. Moreover, steady treatment can positively influence the treatment outcome, such as prevention of recurrence [10–13]. In other words, continuous follow-up from the patient is an important indicator of prognosis [14].

However, the rate of outpatient treatment duration in patients with AUD is significantly low. According to a domestic study, 91.7% of patients stopped follow-up within six months of discharge [14]. In other countries, 52–75% of patients receiving outpatient treatment for alcohol abuse and dependence discontinued the treatment upon the fourth installment [15–17]. Therefore, predicting and managing patients with AUDs likely to drop out of follow-up is of paramount importance.

With the aim of increasing the retention rate of treatments in patients with AUDs, factors affecting the continuous maintenance of outpatient care have been studied. It was found that age, sex, physical and mental comorbidities, hospitalization, family history, type of drugs, marital status, drinking volume, and drinking period were factors that affected continuous outpatient visits [18–20]. However, these studies were mostly prospective, and retrospective studies focused only on factor analysis, using traditional methodologies such as logistic regression [14].

In recent psychiatric research, machine learning (ML) models have been used to predict psychiatric disorders with high accuracy, which is useful for developing clinical decision support systems and identifying influential variables [21–24]. In the United States, a study predicted success in treatment of patients with substance use disorders. However, this study was not very meaningful as it compared the performance of various ML models rather than identifying factors [25]. In this regard, a study was conducted to predict the discontinuation of inpatient treatment for opioid abuse patients using the Treatment Episode Data Set—Discharges claim data from Substance Abuse and Mental Health Services Administration in the United States [26]. However, in psychiatry, success criteria for inpatient and outpatient treatment are defined differently, depending on the duration of treatment [25]. In other words, predicting treatment discontinuation in outpatients with AUD is challenging.

Currently, no studies that predict treatment dropout rates in outpatients with AUDs in Korea have been published. Therefore, this study aimed to predict the discontinuation of outpatient treatment in patients with AUD via ML. The obtained model can aid personalized patient management so that patients with AUDs can maintain treatment steadily. Ultimately, the model could prevent recurrence and increase the success rate of treatments for such patients.

Materials and methods

Experimental data

This was a multicenter retrospective study of patients visiting Seoul St. Mary's Hospital, Uijeongbu St. Mary's Hospital, and Bucheon St. Mary's Hospital in Korea. The study considered data from 2,206 patients, hospitalized between January 2006 and March 2020, for mental and behavioral disorders due to alcohol use (F10, ICD-10). The study protocol was approved by the institutional review board of the Catholic University of Korea (IRB No. XC20WIDI0079K). Data were collected through the Clinical Data Warehouse (CDW), which incorporates eight affiliated hospitals under the Catholic Medical Center (CMC) in Korea. The CDW is a database that has completely anonymized approximately 15 million electronic medical records, and can extract data based on research characteristics. Thus, this study proceeded with consent exemption. The selection of the participants follows the process shown in Fig 1.

For this study, patients with AUD were defined as those with a hospitalization period of more than two weeks. Thus, the study focused on patients requiring continuous outpatient treatment. If more than two hospitalizations had occurred for more than two weeks, the first hospitalization was defined as the date of hospitalization. Follow-up success in patients with AUD was defined as outpatient visits at least once a month, for six months after discharge

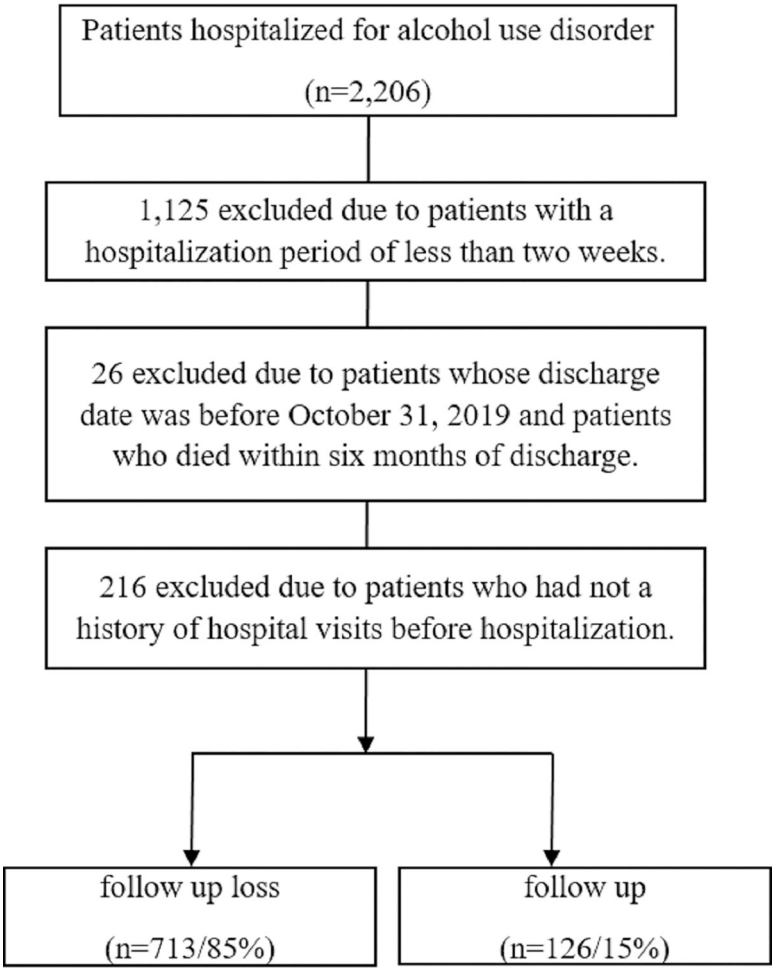

Fig 1. Flow chart of inclusion of subjects.

https://doi.org/10.1371/journal.pone.0255626.g001

[14]. Therefore, patients with discharge dates before October 31, 2019 were excluded. Patients who died within six months of discharge were excluded, and only those who had a history of hospital visits before hospitalization were included to determine whether they had comorbidities. Consequently, the final dataset comprised 839 patients.

Variable selection

In this study, we selected 11 variables based on prior research [16–18] and consultation with clinicians. Furthermore, we computed the variance inflation factor and constructed the variables without multicollinearity between them. The final variables determined were age, sex, length of hospitalization, address, medical department, comorbidities diagnosed within a year before hospitalization (diabetes, liver disease, depressive disorder, and anxiety disorder), outpatient treatment for AUD before hospitalization, and prescription of naltrexone.

Statistical analysis

The data set of 839 patients was divided into 'follow-up' and 'follow-up loss' groups, depending on the duration of outpatient treatment and the number of outpatient visits. The follow-up group consists of patients who visited outpatient care for more than six months and more than once a month. In all other cases, it was classified as a follow-up loss group. As a result, the follow-up group and follow-up loss group consisted of 126 (15%) and 713 (85%) patients respectively. We performed chi-square tests for 11 categorical variables to determine the differences between the groups(p-value < 0.05, chisq.test in R). To evaluate the model, we split the data into 2 datasets: 80% for training and 20% for testing [27]. The collected data had a class imbalance problem of 85:15. The data imbalance problem was solved before training the model. When the ML model was applied to a highly imbalanced dataset, most learners exhibited a bias towards the majority classes, while ignoring minority classes [28]. Therefore, we applied oversampling to the training set to balance the classes [29]. Subsequently, we applied the ML model using 11 predictors from the training set. The ML models used were logistic regression [30], support vector machine [31], k- nearest neighbor (KNN) [32], random forest [33], neural network [34], and AdaBoost decision tree [35]. The evaluation of the models considered accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUROC). The development of a machine learning model follows the process shown in Fig 2. The ML analysis was implemented in Python (version 3.8), and the analysis package Scikit-learn was used [36]. The ML algorithms used in the analysis are described below.

Logistic regression. Logistic regression is a model of association between a dependent variable and independent variables when the dependent variable is binary [37]. Logistic regression coefficients can be easily interpreted as indicators of variable importance [38]. Therefore, it is a widely used model in healthcare. Furthermore, logistic regression can improve performance by considering regularization (L1, L2), to prevent overfitting. However, logistic regression has limitations in terms of solving nonlinear problems because linearity between dependent and independent variables must be assumed. For this study, the LogisticRegression algorithm in Scikit-Learn is used to model the dataset and l2 regularization specified.

Support vector machine. SVM is available for both classification and regression analysis, and is a supervised learning model for pattern recognition and data analysis in ML. SVM is a method of computing hyperplanes that optimally separate data belonging to two classes [39]. In addition to linear classification, SVM also enables nonlinear classification using kernel tricks. However, for large-scale data, the training time is long and it is difficult to understand the individual effects in the final model. This study used the linearSVC in Scikit-Learn.

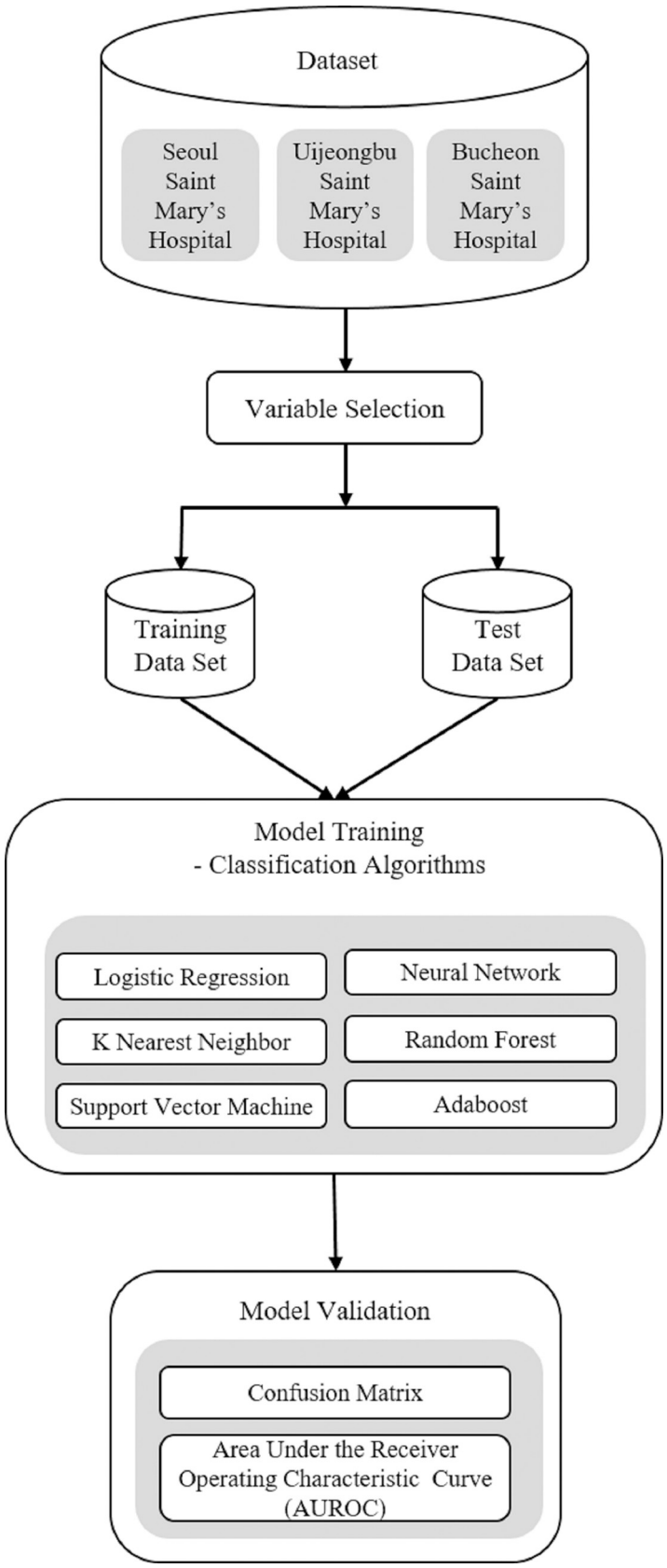

Fig 2. Block diagram of the process of the research analysis.

https://doi.org/10.1371/journal.pone.0255626.g002

K Nearest Neighbor. KNN is an algorithm that predicts output variables by referring to the nearest k neighboring points to a particular point. Both classification and regression models are possible. KNN can easily reflect newly accumulated data because it goes through a reasoning process without a learning process. However, the more the data and the larger the dimension, the more the increase in time and cost increase. In this work, we used the KNeighborsClassifier in Scikit-Learn with five neighbors using Euclidian distances. The Euclidean distance formula is shown below.

$$\mathrm{D}(\mathrm{X},\mathrm{Y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \text{ where } X = (x_1, x_2, \cdots, x_n) \text{ and } Y = (y_1, y_2, \cdots, y_n).$$

Random forest. Random forest is an ensemble method for learning multiple decision trees. It uses randomization techniques such as bagging to reduce the variance of performance. In addition, it solves the problem of overfitting, generating good prediction results. However, a large number of trees are required for accurate prediction, which can make the algorithm very slow and inefficient. Also, it is impossible to explain the associations between variables [40]. For this study, RandomForestClassifier in scikit-learn is used and gini impurity to measure the quality of a split.

Neural network. Neural networks are methods for predicting the values of target variables after learning, using numerous interconnected nodes within each layer consisting of input layers, hidden layers, and output layers. Recently, it has been a popular model in various fields due to its good predictive performance. Neural networks are prone to problems such as overfitting and underfitting due to their limited hidden layers and the complexity of learning. Furthermore, the results derived are difficult to explain.

AdaBoost. This is the first practical boosting algorithm studied by Freund and Schapire in 1997. This method leads to several weak learners and shows high performance because it creates a new predictor by relatively increasing the weight of poorly classified training samples based on errors from previous learners. Furthermore, the method has the advantage of being a tree-based algorithm, from which we can obtain the importance of variables that affect prediction, making it interpretable. This study used AdaBoostClassifier in scikit-learn.

Results

We divided patients with AUDs as those who dropped out of outpatient treatment after discharge and those who received continuous treatment. A basic statistical analysis was conducted to determine which patient characteristics differed between the groups. Sex, address, medical department, depressive disorder, outpatient treatment for AUD before hospitalization, and prescription of naltrexone showed significant differences between the two groups (Table 1).

According to the results of the follow-up loss group, 82.7% of men and 74.2% of patients were found to not have depression. In addition, 87.8% of patients were not prescribed naltrexone and 63.3% of them were in the Gyeonggi Province. Lastly, 54.4% of patients opted for outpatient treatment for alcohol use disorder before hospitalization.

Table 2 shows the performance results of the machine-learning models analyzed using the 11 predictors. To compare the performance of the six models, we obtained the accuracy, specificity, sensitivity, and AUROC values. In this study, we selected the final model based on the AUROC value, considering both sensitivity and specificity, owing to the high class imbalance present in the data. Moreover, a grid search was performed with five-fold cross-validation to find the optimal hyperparameters of the ML models.

Table 1. Patient characteristics.

	Follow-up (n = 126)	Follow-up loss (n = 713)	P-value
Length of hospitalization			0.406
Under 28d	437 (61.3%)	77 (61.1%)	
29-56d	136 (19.1%)	25 (19.8%)	
57-70d	110 (15.4%)	15 (11.9%)	
Over 70d	30 (4.2%)	9 (7.1%)	
Sex			0.008*
Male	91 (72.2%)	590 (82.7%)	
Female	35 (27.8%)	123 (17.3%)	
Age			0.058
Under 29	9 (7.1%)	22 (3.1%)	
30–39	22 (17.5%)	96 (13.5%)	
40–49	29 (23.0%)	201 (28.2%)	
50–59	30 (23.8%)	216 (30.3%)	
60+	36 (28.6%)	178 (25.0%)	
Address			0.04*
Seoul	37 (29.4%)	144 (20.2%)	
Gyeonggi	75 (59.5%)	451 (63.3%)	
Other	14 (11.1%)	118 (16.5%)	
Medical department			0.01*
Psychiatry	111 (88.1%)	546 (76.6%)	
Gastroenterology	9 (7.1%)	104 (14.6%)	
Other	6 (4.8%)	63 (8.8%)	
Outpatient treatment for alcohol use disorder before hospitalization			0.000*
No	35 (27.8%)	325 (45.6%)	
Yes	91 (72.2%)	388 (54.4%)	
Diabetes			0.087
No	109 (86.5%)	654 (91.7%)	
Yes	17 (13.5%)	59 (8.3%)	
Liver disease			0.224
No	107 (84.9%)	569 (79.8%)	
Yes	19 (15.1%)	144 (20.2%)	
Depressive disorder			0.006*
No	78 (61.9%)	529 (74.2%)	
Yes	48 (38.1%)	184 (25.8%)	
Anxiety disorder			0.053
No	104 (82.5%)	635 (89.1%)	
Yes	22 (17.5%)	78 (10.9%)	
Naltrexone			0.000*
No	93 (73.8%)	626 (87.8%)	
Yes	33 (26.2%)	87 (12.2%)	

* $p<0.05$

https://doi.org/10.1371/journal.pone.0255626.t001

Fig 3 shows the receiver operating characteristic (ROC) curves for the six models. The lower area of the curve drawn with each point is the AUROC value. The results show that the larger the area, greater is the AUROC value, and the higher is the performance. The results of the ROC curves show that AdaBoost has the highest AUROC of 0.72.

Table 2. The performance of machine learning algorithms.

Model	AUROC	Accuracy	Sensitivity	Specificity
Logistic Regression	0.6914	0.6130	0.7058	0.6026
SVM	0.6797	0.7023	0.6470	0.7086
KNN	0.6166	0.6726	0.5294	0.6887
Random Forest	0.6365	0.7380	0.4705	0.7682
Neural Network	0.6891	0.7440	0.5294	0.7682
AdaBoost	0.7241	0.6428	0.7647	0.6291

https://doi.org/10.1371/journal.pone.0255626.t002

Table 3 shows the comparison of sampling methods to address data imbalance. First, we compared oversampling and undersampling. Upon comparing AdaBoost performance with random oversampling and random undersampling, the accuracy with of oversampling was higher by 0.065. Therefore, oversampling was considered as the preferred method. However, there is a possibility of overfitting because random oversampling is a method of randomly replicating data from the minority class. Next, we performed a comparison with other oversampling methods. The Synthetic Minority Oversampling Technique (SMOTE) method does not simply replicates minority data but uses the KNN algorithm to generate synthetic data. However, SMOTE does not take into consideration neighboring examples can be from other classes. This can increase the overlapping of classes and can introduce additional noise. After application of the SMOTE, performance was low. In other words, random oversampling showed the best results for our data.

Fig 4 shows the feature importance results obtained using the Gini index [41] in the AdaBoost decision tree. The top four variables that affect patients with AUDs dropping out of treatment are: length of hospitalization, age, region and diabetes.

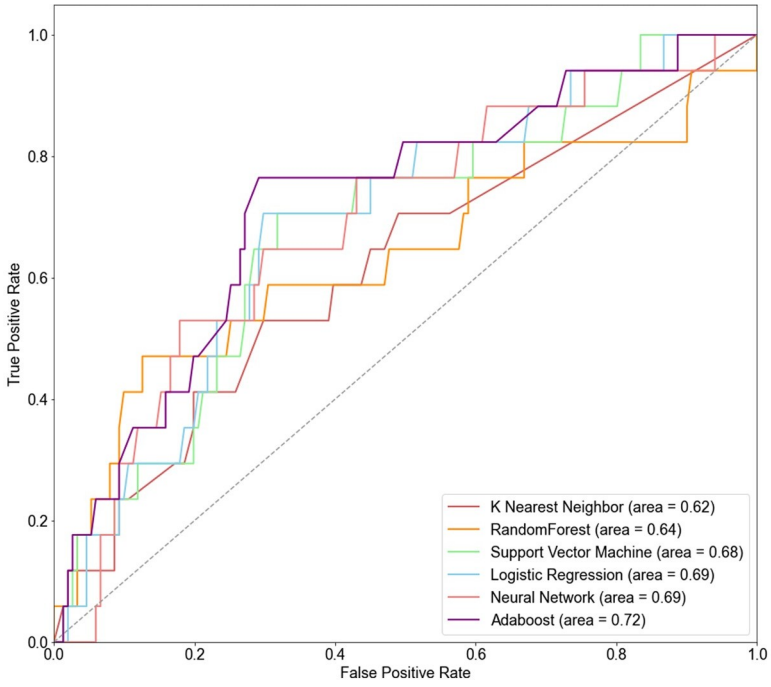

Fig 3. ROC curves of six different machine learning models.

https://doi.org/10.1371/journal.pone.0255626.g003

Table 3. Comparison of imbalanced data set sampling methods.

Method	AUROC	Accuracy	Sensitivity	Specificity
Random Undersampling	0.6505	0.5773	0.7058	0.5629
Random Oversampling	0.7241	0.6428	0.7647	0.6291
SMOTE	0.6427	0.5952	0.5294	0.6026

https://doi.org/10.1371/journal.pone.0255626.t003

According to Fig 5, 61% of all patients were hospitalized within a month. In particular, the proportion of patients who stopped outpatient treatment early was relatively higher than those who did not when the length of hospitalization was 14–25 days.

Discussion

We developed an algorithm to predict the patients who dropped out of outpatient treatment after discharge from AUD in Korea. This study is relevant as it is the first to use ML. Among the six models, the AdaBoost decision tree had the highest AUROC.

The following are the top four variables affecting discontinuation of outpatient treatment, derived from AdaBoost.

First, the length of hospitalization for AUDs can have an impact on predicting whether outpatient treatment is discontinued. Therefore, special care is needed that patients receive continuous treatment in consideration of the their length of hospitalization. According to the analysis, most of the patients who stopped treatment were hospitalized within 25 days. Based on this information, special care is needed to patients within 25 days of hospitalization. Second, the age at the time of hospitalization for AUD was related to whether treatment was discontinued after discharge. This result was consistent with existing studies that showed increased motivation for continuous treatment with age [42]. Management measures will be needed to increase the willingness of patients in their 40-50s to continue treatment.

Third, the residential area acted as a variable affecting the discontinuation of outpatient treatment. Prior studies also indicate that the higher the accessibility to residential areas and hospitals, the higher the possibility of continuous treatment [43]. As residents outside Seoul

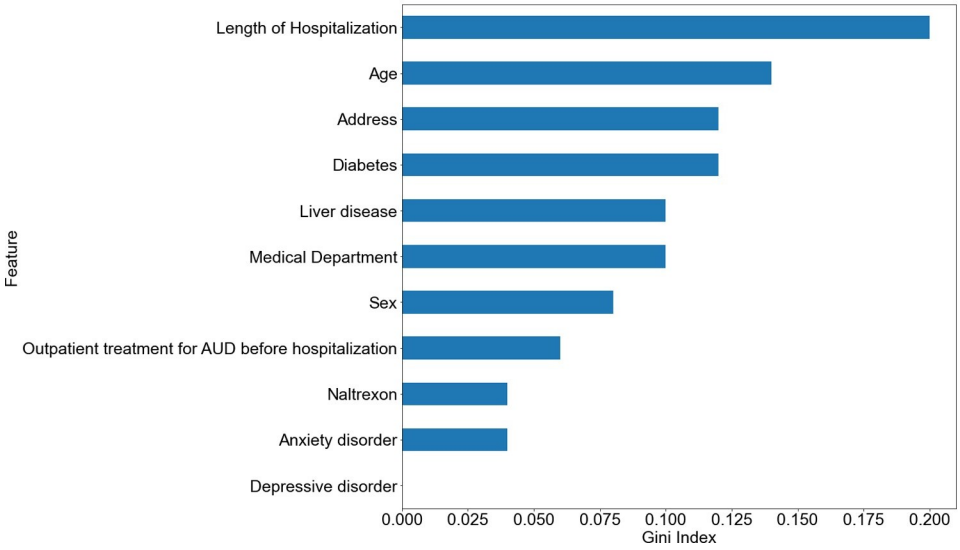

Fig 4. Feature importance of AdaBoost decision tree.

https://doi.org/10.1371/journal.pone.0255626.g004

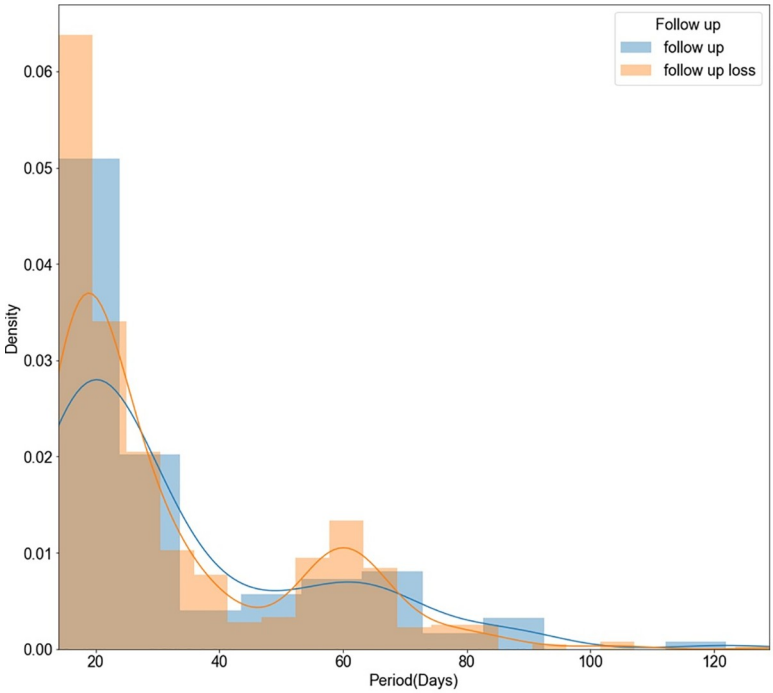

Fig 5. Density plot of length of hospitalization.

https://doi.org/10.1371/journal.pone.0255626.g005

and Gyeonggi Province have limited access to hospitals, a system to monitor them in connection with primary institutions will be needed.

Fourth, the presence of diabetes affects the compliance of outpatient treatment. According to the data, the proportion of patients with diabetes in the follow-up group was higher than that in the follow-up loss group. Excessive alcohol consumption by diabetic patients can worsen blood sugar control, which can be fatal to diabetes treatment and, if severe, can lead to death [44]. Diabetes is a chronic disease; therefore, continuous management is needed. However, drinking and diabetes self-management performance are negatively correlated [45]. Patients with diabetes feel the need for alcohol treatment in terms of diabetes management. Moreover, it is estimated that they are more willing and demanding of outpatient treatment. Therefore, continuous care should be maintained in patients without diabetes than in patients with diabetes.

This study however has several limitations. First, there is a problem with inaccuracy regarding the information on patients' comorbidities. Due to the nature of the retrospective study, the patient's disease can be identified by the diagnostic code patients received at the hospital where they received treatment. Therefore, it may be challenging to identify all undiagnosed comorbidities in hospitals. However, this study addressed the limitations by using medical records in the hospital from one year before hospitalization.

Second, the data for the study did not include socioeconomic factors and variables related to addiction treatment. Socioeconomic factors such as the patient's religion, marital status, and occupation, as well as the degree of alcoholism and treatment process, are known to affect the continuous outpatient treatment of patients with AUD [46–48]. However, the CDW that collected the data was then not collecting this information or was under construction. Future studies on predictive models that consider various factors will further reflect clinical reality.

This is the first study in Korea to develop an algorithm that predicts whether patients with AUD will drop out of treatment using ML methods. The final selected AdaBoost algorithm showed higher accuracy than that of traditional models, such as regression analysis. The algorithm could identify key variables affecting treatment discontinuation. This model can be used to develop a clinical decision support system. In other words, this study allows clinicians to assist patients with AUDs in receiving continuous treatment. Moreover, this ML model predicts discontinuation of outpatient treatment in patients with AUD and identifies its factors, but also has the potential to be applicable to other substances.

Finally, our data for analysis included only structured data, and did not include unstructured data elements such as clinical notes. In future studies, including both unstructured and structured data may further improve prediction accuracy.

Acknowledgments

We thank the Information Service Team of Catholic Medical Center for helping us analyze the data.

Author Contributions

Conceptualization: In Young Choi.

Data curation: So Jin Park.

Formal analysis: So Jin Park.

Funding acquisition: In Young Choi.

Investigation: So Jin Park, Ji-Won Chun, In Young Choi.

Methodology: So Jin Park, Jae Kwon Kim.

Project administration: In Young Choi.

Resources: Ji-Won Chun, Soo-Jung Lee, Hae Kook Lee, Dai Jin Kim, In Young Choi.

Software: So Jin Park, HyungMin Kim.

Supervision: In Young Choi.

Validation: So Jin Park, Sun Jung Lee, HyungMin Kim, Jae Kwon Kim, Ji-Won Chun, Soo-Jung Lee, Hae Kook Lee, Dai Jin Kim, In Young Choi.

Visualization: So Jin Park, Sun Jung Lee.

Writing – original draft: So Jin Park.

Writing – review & editing: So Jin Park.

References

1. Shin JW, Lee SY, Shin HW, Yoon KJ, Lee SH, Son CK. patient survey [Internet]. Sejong: Ministry of Health and Welfare; 2017 Sep [updated 2017 Sep 25; cited 2020 Apr 14]. 2016.
2. Giesbrecht N, Cukier S, Steeves DAN. Collateral damage from alcohol: implications of'second-hand effects of drinking'for populations and health priorities. 2010;
3. Collins SS. Associations Between Socioeconomic Factors and Alcohol Outcomes. Alcohol Research: Current Reviews, 38 (1), 83–94. 2016. PMID: 27159815
4. Witkiewitz K, Kranzler HR, Hallgren KA, O'Malley SS, Falk DE, Litten RZ, et al. Drinking risk level reductions associated with improvements in physical health and quality of life among individuals with alcohol use disorder. Alcohol Clin Exp Res. 2018; 42(12):2453–65. https://doi.org/10.1111/acer.13897 PMID: 30395350

5. Brandon TH, Vidrine JI, Litvin EB. Relapse and relapse prevention. Annu Rev Clin Psychol. 2007; 3:257–84. https://doi.org/10.1146/annurev.clinpsy.3.022806.091455 PMID: 17716056

6. Goldstein RZ, Bechara A, Garavan H, Childress AR, Paulus MP, Volkow ND. The neurocircuitry of impaired insight in drug addiction. Trends Cogn Sci. 2009; 13(9):372–80. https://doi.org/10.1016/j.tics.2009.06.004 PMID: 19716751

7. Dandaba M, Serra W, Harika-Germaneau G, Silvain C, Langbour N, Solinas M, et al. Predicting relapse in patients with severe alcohol use disorder: The role of alcohol insight and implicit alcohol associations. Addict Behav. 2020; 107:106433. https://doi.org/10.1016/j.addbeh.2020.106433 PMID: 32289744

8. Lenaerts E, Matheï C, Matthys F, Zeeuws D, Pas L, Anderson P, et al. Continuing care for patients with alcohol use disorders: a systematic review. Drug Alcohol Depend. 2014; 135:9–21. https://doi.org/10.1016/j.drugalcdep.2013.10.030 PMID: 24314854

9. Knox J, Hasin DS, Larson FRR, Kranzler HR. Prevention, screening, and treatment for heavy drinking and alcohol use disorder. The Lancet Psychiatry. 2019; 6(12):1054–67. https://doi.org/10.1016/S2215-0366(19)30213-5 PMID: 31630982

10. McKay JR. Continuing care research: What we have learned and where we are going. J Subst Abuse Treat. 2009; 36(2):131–45. https://doi.org/10.1016/j.jsat.2008.10.004 PMID: 19161894

11. Johannessen DA, Nordfjærn T, Geirdal AØ. Substance use disorder patients' expectations on transition from treatment to post-discharge period. Nord Stud Alcohol Drugs. 2020; 37(3):208–26.

12. Murthy P, Chand P, Harish MG, Thennarasu K, Prathima S. Outcome of alcohol dependence: The role of continued care. Indian J community Med Off Publ Indian Assoc Prev Soc Med. 2009; 34(2):148. https://doi.org/10.4103/0970-0218.51226 PMID: 19966963

13. Malick R. Prevention of substance use disorders in the community and workplace. Indian J Psychiatry. 2018; 60(Suppl 4):S559. https://doi.org/10.4103/psychiatry.IndianJPsychiatry_24_18 PMID: 29540931

14. Kim KH, An IS. Affecting Factors for Continuing Outpatient Care After Inpatient Care for Patients with Alcohol Dependence in Korea: A Population-based. Korea Inst Heal Soc Aff. 2015; 35(4):116–30.

15. Baekeland F, Lundwall L. Dropping out of treatment: a critical review. Psychol Bull. 1975; 82(5):738. https://doi.org/10.1037/h0077132 PMID: 1103201

16. Elbreder F, Pillon C, Laranjeira R. Alcohol dependence: analysis of factors associated with retention of patients in outpatient treatment. Alcohol Alcohol. 2011; 46(1):74–6. https://doi.org/10.1093/alcalc/agq078 PMID: 21118901

17. Simpson DD, Joe GW, Rowan-Szal GA. Drug abuse treatment retention and process effects on follow-up outcomes. Drug Alcohol Depend. 1997; 47(3):227–35. https://doi.org/10.1016/s0376-8716(97)00099-9 PMID: 9306048

18. Seong SK, Bang YW, Haham W. A follow-up study of inpatients by the telephone interview. J Korean Neuropsychiatr Assoc. 1993; 32(5):698–706.

19. Kim JS, Han SI, Kim KS. Clinical variables affecting relapse of alcoholism. J Korean Neuropsychiatr Assoc. 1994; 33(4):817–24.

20. Saarnio P. Factors associated with dropping out from outpatient treatment of alcohol-other drug abuse. Alcohol Treat Q. 2002; 20(2):17–33.

21. Tate AE, McCabe RC, Larsson H, Lundström S, Lichtenstein P, Kuja-Halkola R. Predicting mental health problems in adolescence using machine learning techniques. PLoS One. 2020; 15(4):e0230389. https://doi.org/10.1371/journal.pone.0230389 PMID: 32251439

22. Iniesta R, Stahl D, McGuffin P. Machine learning, statistical learning and the future of biological research in psychiatry. Psychol Med. 2016; 46(12):2455–65. https://doi.org/10.1017/S0033291716001367 PMID: 27406289

23. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim H-C, et al. Artificial intelligence for mental health and mental illnesses: an overview. Curr Psychiatry Rep. 2019; 21(11):1–18. https://doi.org/10.1007/s11920-019-1094-0 PMID: 31701320

24. Cohen ZD, DeRubeis RJ. Treatment selection in depression. Annu Rev Clin Psychol. 2018; 14. https://doi.org/10.1146/annurev-clinpsy-050817-084746 PMID: 29494258

25. Acion L, Kelmansky D, van der Laan M, Sahker E, Jones D, Arndt S. Use of a machine learning framework to predict substance use disorder treatment success. PLoS One. 2017; 12(4):e0175383. https://doi.org/10.1371/journal.pone.0175383 PMID: 28394905

26. Gautam P, Singh P. A Machine Learning Approach to Identify Socio-Economic Factors Responsible for Patients Dropping out of Substance Abuse Treatment. Am J Public Health. 2020; 8(5):140–6.

27. Géron A. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media; 2019.

28. Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. J Big Data. 2019; 6 (1):1–54.

29. Sim J, Kim YA, Kim JH, Lee JM, Kim MS, Shim YM, et al. The major effects of health-related quality of life on 5-year survival prediction among lung cancer survivors: applications of machine learning. Sci Rep. 2020; 10(1):1–12. https://doi.org/10.1038/s41598-019-56847-4 PMID: 31913322

30. Thabtah F, Abdelhamid N, Peebles D. A machine learning autism classification based on logistic regression analysis. Heal Inf Sci Syst. 2019; 7(1):1–11. https://doi.org/10.1007/s13755-019-0073-5 PMID: 31168365

31. Pisner DA, Schnyer DM. Support vector machine. In: Machine Learning. Elsevier; 2020. p. 101–21.

32. Yao M, Vocational B. Research on learning evidence improvement for KNN based classification algorithm. Int J Database Theory Appl. 2014; 7(1):103–10.

33. Probst P, Wright MN, Boulesteix A. Hyperparameters and tuning strategies for random forest. Wiley Interdiscip Rev Data Min Knowl Discov. 2019; 9(3):e1301.

34. Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: A survey. Heliyon. 2018; 4(11):e00938. https://doi.org/10.1016/j.heliyon.2018.e00938 PMID: 30519653

35. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci. 1997; 55(1):119–39.

36. Hackeling G. Mastering Machine Learning with scikit-learn. Packt Publishing Ltd; 2017.

37. Vetter TR, Schober P. Regression: the apple does not fall far from the tree. Anesth Analg. 2018; 127 (1):277–83. https://doi.org/10.1213/ANE.0000000000003424 PMID: 29771712

38. Schober P, Vetter TR. Logistic regression in medical research. Anesth Analg. 2021; 132(2):365. https://doi.org/10.1213/ANE.0000000000005247 PMID: 33449558

39. Surówka G, Ogorzalek M. Resolution invariant wavelet features of melanoma studied by SVM classifiers. PLoS One. 2019; 14(2):e0211318. https://doi.org/10.1371/journal.pone.0211318 PMID: 30726260

40. Couronné R, Probst P, Boulesteix A-L. Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics. 2018; 19(1):1–14. https://doi.org/10.1186/s12859-017-2006-0 PMID: 29291722

41. Daniya T, Geetha M, Kumar KS. Classification And Regression Trees with Gini Index. Adv Math Sci J. 2020; 9(10):8237–47.

42. McKay JR, Foltz C, Leahy P, Stephens R, Orwin RG, Crowley EM. Step down continuing care in the treatment of substance abuse: Correlates of participation and outcome effects. Eval Program Plann. 2004; 27(3):321–31.

43. Schmitt SK, Phibbs CS, Piette JD. The influence of distance on utilization of outpatient mental health aftercare following inpatient substance abuse treatment. Addict Behav. 2003; 28(6):1183–92. https://doi.org/10.1016/s0306-4603(02)00218-6 PMID: 12834661

44. Engler PA, Ramsey SE, Smith RJ. Alcohol use of diabetes patients: the need for assessment and intervention. Acta Diabetol. 2013; 50(2):93–9. https://doi.org/10.1007/s00592-010-0200-x PMID: 20532803

45. Balhara YPS. Diabetes and psychiatric disorders. Indian J Endocrinol Metab. 2011; 15(4):274. https://doi.org/10.4103/2230-8210.85579 PMID: 22028998

46. Kweon Y, Lee H, Lee J, Lee C. A follow up study of alcoholic inpatients after alcoholism treatment program. J Korean Acad Addict Psychiatry. 2002; 6:114–9.

47. Schaefer JA, Ingudomnukul E, Harris AHS, Cronkite RC. Continuity of care practices and substance use disorder patients' engagement in continuing care. Med Care. 2005;1234–41. https://doi.org/10.1097/01.mlr.0000185736.45129.95 PMID: 16299435

48. Harris AHS, McKellar JD, Moos RH, Schaefer JA, Cronkite RC. Predictors of engagement in continuing care following residential substance use disorder treatment. Drug Alcohol Depend. 2006; 84(1):93–101. https://doi.org/10.1016/j.drugalcdep.2005.12.010 PMID: 16417977