ELSEVIER

EClinicalMedicine
Published by THE LANCET

Research Paper

Using Machine Learning to Classify Individuals With Alcohol Use Disorder Based on Treatment Seeking Status

Mary R. Lee [a,*], Vignesh Sankar [a], Aaron Hammer [a], William G. Kennedy [b], Jennifer J. Barb [c], Philip G. McQueen [c], Lorenzo Leggio [a,d]

[a] Section on Clinical Psychoneuroendocrinology and Neuropsychopharmacology, National Institute on Alcohol Abuse and Alcoholism, National Institute on Drug Abuse, National Institutes of Health, Bethesda, MD, USA
[b] Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA, USA
[c] Center for Information Technology, National Institutes of Health, Bethesda, MD, USA
[d] Center for Alcohol and Addiction Studies, Department of Behavioral and Social Sciences, Brown University, Providence, RI, USA

ARTICLE INFO

ABSTRACT

Objective: The authors used a decision tree classifier to reduce neuropsychological, behavioral and laboratory measures to a subset of measures that best predicted whether an individual with alcohol use disorder (AUD) seeks treatment.

Method: Clinical measures (N = 178) from 778 individuals with AUD were used to construct an alternating decision tree (ADT) with 10 measures that best classified individuals as treatment or not treatment-seeking for AUD. ADT's were validated by two methods: using cross-validation and an independent dataset (N = 236). For comparison, two other machine learning techniques were used as well as two linear models.

Results: The 10 measures in the ADT classifier were drinking behavior, depression and drinking-related psychological problems, as well as substance dependence. With cross-validation, the ADT classified 86% of individuals correctly. The ADT classified 78% of the independent dataset correctly. Only the simple logistic model was similar in accuracy; however, this model needed more than twice as many measures as ADT to classify at comparable accuracy.

Interpretation: While there has been emphasis on understanding differences between those with AUD and controls, it is also important to understand, within those with AUD, the features associated with clinically important outcomes. Since the majority of individuals with AUD do not receive treatment, it is important to understand the clinical features associated with treatment utilization; the ADT reported here correctly classified the majority of individuals with AUD with 10 clinically relevant measures, misclassifying <7% of treatment seekers, while misclassifying 38% of non-treatment seekers. These individual clinically relevant measures can serve, potentially, as separate targets for treatment.

Funding: Funding for this work was provided by the Intramural Research Programs of the National Institute on Alcohol Abuse and Alcoholism (NIAAA), National Institute on Drug Abuse (NIDA) and the Center for Information Technology (CIT).

Research in Context: Evidence Before This Study: Less than 10% of persons who meet lifetime criteria for Alcohol Use Disorder (AUD) receive treatment. As the etiology of AUD represents a complex interaction between neurobiological, social, environmental and psychological factors, low treatment utilization likely stems from barriers on multiple levels. Given this issue, it is important from both a research and clinical standpoint to determine what characteristics are associated with treatment utilization in addition to merely asking individuals if they wish to enter treatment. At the level of clinical research, if there are phenotypic differences between treatment and nontreatment-seekers that directly influence outcomes of early-phase studies, these phenotypic differences are a potential confound in assessing the utility of an experimental treatment for AUD. At the level of clinical practice, distinguishing between treatment- and nontreatment-seekers may help facilitate a targeted treatment approach. Previous efforts to understand the differences between these populations of individuals with AUD leveraged the multidimensional data collected in clinical research settings for AUD that are not well suited to traditional regression methods.

Added Value of This Study: Alternating decision trees are well suited to deep-phenotyping data collected in clinical research settings as this approach handles nonparametric, skewed, and missing data whose relationships are nonlinear. This approach has proved to be superior in some cases to conventional clinical methods to solve

---

* Corresponding author.
  E-mail address: leemary@mail.nih.gov (M.R. Lee).

diagnostic problems in medicine. We used a decision tree classifier to understand treatment- and non-treatment seeking group differences. The decision tree classifier approach chose a subset of factors arranged in an alternating decision tree that best predicts a given outcome. Assuming that the input measures are clinically relevant, the alternating decision tree that is generated has clinical value. Unlike other machine learning approaches, in addition to its predictive value, the nodes in the tree and their arrangement in a hierarchy have clinical utility. With the "if-then" logic of the tree, the clinician can learn what features become important and which recede in importance as the logic of the tree is followed. The decision tree classifier approach reduced 178 characterization measures (both categorical and continuous) in multiple domains to a decision tree comprised of 10 measures that together best classified subjects by treatment seeking status (yes/no).

Implications After All the Available Evidence: We leveraged a large data set comprised of 178 clinical measures and using the decision tree approach, we have reduced these to a subset of 10 measures that accurately classified individuals with alcohol dependence by treatment utilization. From this analysis, drinking behavior variables and depression measures are strong treatment seeking predictors. Having identified a cluster of factors that predicts treatment seeking, we can assess the influence of these factors directly on the clinical study outcome measures themselves. In clinical practice these factors can be separate targets for treatment. In clinical research, the group differences my directly influence research outcomes for treatment of AUD.

1. Introduction

Alcohol Use Disorder (AUD), as defined in the Diagnostic and Statistical Manual of Mental disorders, fifth edition (DSM-5), is highly prevalent and is a leading cause of mortality and morbidity worldwide [1,2]. Less than 10% of persons who meet lifetime criteria for AUD receive any type of treatment [3]. As the etiology of AUD represents a complex interaction between neurobiological, social, environmental and psychological factors, low treatment utilization likely stems from barriers on multiple levels. Large-scale epidemiological surveys identified multiple societal, demographic and personal factors influencing treatment-seeking status [4,5]. However, the fact remains that the majority of people with AUD do not receive treatment [3]. Given this issue, it is important from both a research and clinical standpoint to determine what characteristics are associated with treatment utilization in addition to merely asking individuals if they wish to enter treatment.

In research settings, proof-of-concept human laboratory studies that include alcohol administration and/or self-administration are a cost-effective approach to study alcohol-related pharmacological and behavioral effects of alcohol and to screen new medications that have shown promise in animal studies [6]. In these human laboratory studies, alcohol may be administered with the study medication to determine the safety of the combination; in addition, early efficacy of the putative treatment on drinking or craving measures may be assessed. Given the ethical concern of administering alcohol to treatment-seeking patients, only nontreatment-seeking subjects are typically studied [7]. If there are phenotypic differences between treatment and nontreatment-seekers that directly influence outcomes of early-phase studies, these phenotypic differences are a potential confound in assessing the utility of an experimental treatment for AUD. Therefore, understanding the phenotypic differences between these populations provides the opportunity to take into account potential confounding variables in these early phase clinical studies.

Further, distinguishing between treatment- and nontreatment-seekers may help facilitate a targeted treatment approach [8]. This is consistent with recent attempts to use deep phenotyping to characterize neurofunctional domains in AUD [9]. Although recovery from AUD is possible without treatment [10–13], the likelihood of this diminishes with age [10,14]. Therefore, identifying those individuals reluctant to seek treatment for AUD early in the course of their illness is important since educational, behavioral and/or pharmacological interventions can be used in primary care or specialized settings to help patients recover. We found previously that there are personal attributes that differ between treatment compared to nontreatment-seeking individuals who were diagnosed with DSM-IV alcohol dependence (AD) [8]. These differences persisted even after controlling for drinking behavior. Perhaps these characteristics, such as impulsivity and depression, offer

potential as independent targets for treatment. In this vein, the Project MATCH Research Group attempted to match 10 variables, (gender, AUD type, cognitive impairment, concept understanding, meaning seeking, motivation, psychiatric severity, sociopathy and support for drinking) defined a priori, to each of 3 treatment modalities for AUD. [15]. The results were disappointing as only one variable, psychiatric severity, significantly interacted with treatment assignment to reduce variance in drinking outcomes.

Previous efforts to understand the differences between these populations of individuals with AUD leveraged the multidimensional data collected in clinical research settings for AUD [8], [16]. However, these studies [8], [16] were descriptive and exploratory with no correction for multiple comparisons. This multidimensional data is comprised of measures that are relevant to the complex, heterogeneous, multifactorial attributes related to the etiology and clinical course of addiction [17]. In clinical research in the addiction field, there are a myriad of neuropsychosocial factors whose relationships are nonlinear. Traditional regression methods are not well suited to analyze such data given the high degree of collinearity between the variables in these datasets describing characteristics of individuals with a disorder that is heterogeneous in presentation. Decision trees, on the other hand, are well suited to this kind of clinical data because this approach handles nonparametric, skewed, missing data as well as dividing complex input data into subgroups that serve to simplify the relationship between input and outcome variables. This approach has been superior, in some cases, to conventional clinical methods to solve diagnostic problems in medicine [18].

Therefore, in the present study, we used a decision tree classifier to understand treatment- and non-treatment seeking group differences. The decision tree classifier approach chooses a subset of factors arranged in an alternating decision tree (ADT) that best predicts a given outcome. Assuming that the input measures are clinically relevant, the ADT that is generated also has clinical value. Unlike other machine learning approaches, in addition to its predictive value, the nodes in the tree and their arrangement in a hierarchy are understandable for the clinician. With the "if-then" logic of the tree, the clinician can learn what features become important and which recede in importance as the logic of the tree is followed. The decision tree classifier approach reduced 178 characterization measures (both categorical and continuous) in multiple domains (see Supplementary Material) to a decision tree comprised of 10 measures that together best classified subjects by treatment seeking status (yes/no). Given the gender imbalance in our sample, we also constructed separate decision trees for men and women to investigate possible biases associated with gender. Decision trees were validated by cross-validation on the training set of patients with AUD, and by validation on a second, independent dataset of individuals with AUD. For comparison to the ADT approach, two linear models

(logistics and simple logistics) and two machine learning techniques (random forest and random tree) were also employed.

2. Methods

2.1. Recruitment

Data were combined from three screening and evaluation protocols conducted at the National Institute on Alcoholism and Alcohol Abuse (NIAAA) Intramural Research Program at the National Institutes of Health (NIH) Clinical Center in Bethesda, MD. Data were collected from 2008 to 2018, under three protocols which were approved by the appropriate NIH Institutional Review Board. All participants provided written informed consent. Under these protocols, participants were recruited via advertisement on university campuses, newspapers, internet and word of mouth. First, participants were assessed with a phone interview. Preference (yes/no) for treatment was first evaluated by a trained screener over this phone screening interview. The phone screen was also used to exclude individuals with the likelihood of a history of severe mental illness or substance dependence, other than alcohol or nicotine.

2.2. In-person Screen

Those individuals meeting these preliminary criteria were assessed during an out- or in-patient screening visit at the NIH Clinical Center, depending on their clinical needs. In the main dataset, participants (N = 778, N = 554 males) met DSM-IV criteria for AD (this corresponds to a DSM-5 diagnosis of moderate or severe Alcohol Use Disorder (AUD)). Preference (yes/no) for treatment was confirmed and determined by in-person clinician interview. For the independent validation dataset, treatment and nontreatment-seeking subjects (N = 236, 162 males) were recruited subsequent to those in the main dataset. They were evaluated with the same measures except they were assessed with the DSM-5 SCID. Table 1 includes demographics for the main and independent validation datasets.

Recent and lifetime alcohol consumption were assessed using the timeline follow-back (TLFB) [19] and lifetime drinking history [20], respectively. The Structured Clinical Interview for DSM Disorders (SCID) [21] was used to assess for psychiatric comorbidities and to confirm a DSM-IV diagnosis of Alcohol Dependence (AD) for the main dataset (N = 778). For the validation dataset, (N = 236), the SCID for DSM-5 was used. Only those subjects who met criteria for DSM-5 AUD, moderate or severe were included in the validation dataset. All subjects

completed an assessment battery of measures in the following domains: cognitive (Wechsler Abbreviated Scale of Intelligence [22]), mood (Comprehensive Psychopathological Rating Scale (CPRS) [23]), impulsivity (Barratt Impulsivity Scale [24]/UPPS-P [25]), personality (NEO-5-PI-R [26]), aggression (Buss-Perry [27]), and early life stress and childhood trauma (ELSQ [28]/CTQ [29]).

Standard tests of clinical laboratory biomarkers related to AD were a hepatic panel: alanine aminotransferase (ALT); aspartate aminotransferase (AST); gamma-glutamyl transferase (GGT), alkaline phosphatase (Alk Phos), a mineral panel (calcium, magnesium, phosphorus, potassium and sodium), albumin, hemoglobin, mean corpuscular volume (MCV), hepatitis B antigen and hepatitis C antibody. Together, the in-person interview, the assessment battery and biomarkers provided a total of 178 measures in addition to their treatment seeking status.

2.3. Analysis

The alternating decision tree (ADT) method [30,31], specifically, the Waikato Environment for Knowledge Analysis (WEKA) was used [32] to classify persons with respect to their treatment-seeking status for AUD. An ADT consists of splitter nodes, each of which is an assessment measure (either dichotomous or continuous); the splitter node gives rise to two predictor nodes which assign a numerical score for each branch from the splitter node (Figs. 1.a, 2.a, 3.a). An ADT is constructed using a *boosting* algorithm that starts with the predictive node of a weak hypothesis and adds more and more nodes to strengthen the classification [30]. Measurement of treatment seeking status is determined from a linear combination of all predictor nodes in the tree added to the root score. The result is a numerical *score* for a given subject, referred to here as the ADT score. The value of the numerical score indicates classification into treatment (<0) or nontreatment (>0) –seeking groups. The size of the absolute magnitude of the score is a measure of the confidence in the classification into one of the two groups. The further the score from zero, the more likely that the score correctly classifies an individual subject by treatment-seeking status. Because the sample was predominantly male, separate ADTs were constructed and assessed for men and women.

Two different approaches were used to assess the accuracy of the ADT to classify subjects. In the first approach, the main data set described above (N = 778) was used in a cross-validation procedure. (In this report, the main data set is the training set for all the comparison methods discussEd.) The 179 attributes were entered into the WEKA tool: one attribute was treatment seeking status and the other 178 attributes were used to train the ADT. The accuracy was found by

Table 1
Group differences in demographic characteristics between nontreatment-seeking and treatment-seeking research participants with alcohol dependence.

	Primary data set			Validation data set		
	Nontreatment-seeking	Treatment-seeking	*p*-Value	Nontreatment-seeking	Treatment seeking	*p*-Value
N	169	609		90	146	
Age, years						
Mean (*SD*)	42.5 (12.2)	43.1 (10.2)		38.2 (12.8)	47.8 (11.1)	
Range	21.3–65.9	19.0–64.4	0.570[a]	21.1–61.0	23.1–68.9	<0.001[a]
Education, years						
Mean (*SD*)	13.5 (3.0)	13.8 (2.7)		14.3 (3.3)	13.7 (2.4)	
Range	2–24	1–24	0.197	2–21	7–20	0.736[a]
Intelligence quotient (IQ)						
Mean (*SD*)	91.6 (25.8)	97.4 (14.8)		102.1 (19.3)	97.8 (18.8)	
Range	58–141	57–110	0.001	56–148	55–144	0.121[a]
Gender						
Female: *n* (%)	38 (23%)	186 (30%)		32 (36%)	42 (29%)	
Male: *n* (%)	131 (77%)	423 (70%)	0.04[b]	58 (64%)	104 (69%)	0.275[b]
Race						
African American: *n* (%)	134 (73%)	237 (39%)		58 (64%)	58 (40%)	
Caucasian: *n* (%)	34 (20%)	318 (52%)		23 (26%)	69 (47%)	
Other: *n* (%)	11 (7%)	54 (9%)	<0.001[b]	9 (10%)	19 (13%)	0.001[b]

[a] Independent samples *t*-test.
[b] Chi-square.

a
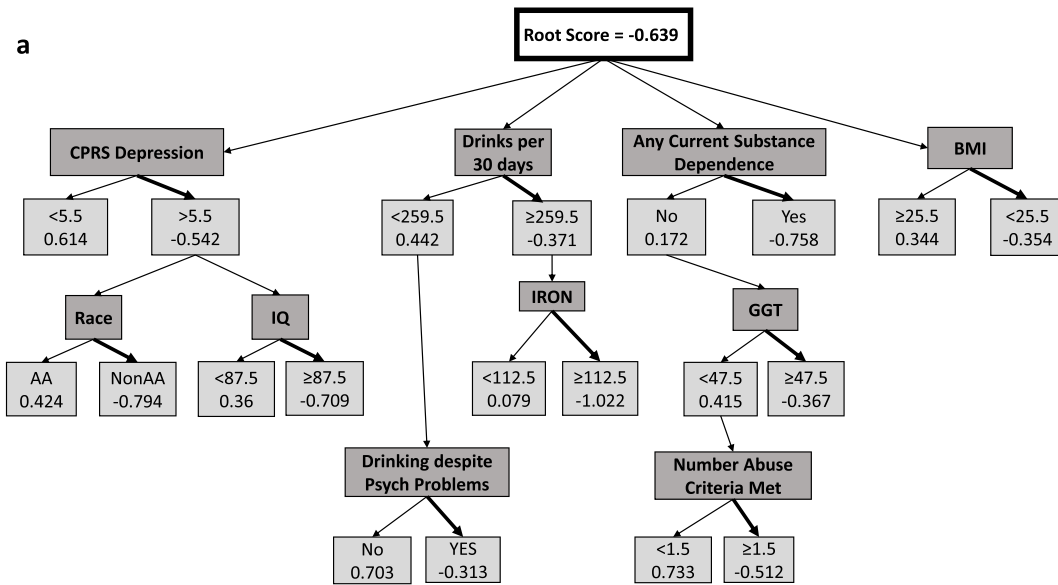
b: Confusion Matrix from Cross-validation with Training Subjects (N=778)

	Classified as TX	Classified as Non-TX
Real TX	566	43
Real Non-TX	65	104
Correctly classified	670 (86.1 %)	
Incorrectly classified	108 (13.8 %)	
Kappa=0.57; Balanced Accuracy=78%		

c: Confusion Matrix from Test with Independent Data (N=236)

	Classified as TX	Classified as Non-TX
Real TX	135	11
Real Non-TX	41	49
Correctly classified	184 (78.0%)	
Incorrectly classified	27 (22.0%)	
Kappa=0.51; Balanced Accuracy =73%		

Fig. 1. Entire Dataset: a) Alternate Decision Tree (ADT) classifying N = 778 individuals with alcohol dependence (AD) as treatment or nontreatment-seekers. 10 characterization measures are noted in dark gray boxes. Classification: Treatment-seeking: total score <0; Nontreatment-seeking: total score >0. CPRS: Comprehensive Psychopathological Rating Scale; BMI: body mass index; IQ: Wechsler Abbreviated Scale of Intelligence. b) Confusion Matrix from Cross-validation with Training Subjects. c) Confusion Matrix from Test with Independent Data (N = 115). Tx = Treatment-seeking (bold arrow); Non-Tx = Not treatment-seeking.

constructing ADT's with data from 10 randomly-chosen datasets. Each dataset consisted of 90% of the subjects. The accuracy of the ADT in each of the 10 datasets was tested on the remaining 10% of subjects. The reported accuracy was the average over the ten cycles. For the second approach, data from an independent group (the "validation group") of subjects (N = 236) who were recruited after those in the main group were recruited. This independent group had moderate or severe AUD (DSM-5) and this group was used for testing the classification accuracy of the ADT that was constructed from the main dataset. (In this report, this second independent dataset is the validation set for all the comparison methods discussEd.)

In order to assess how the ADT scores related to prediction status, a script was written in the JMP Statistical Discovery software version 11.1 (SAS headquarters, Cary, NC) using the ADT node values to calculate the ADT score for each individual patient in the independent validation dataset. Thus, this script can be used to predict whether an individual patient is likely to seek treatment or not. The script prompts the user to enter the 10 clinical variable parameters in the ADT and calculates an ADT score (according to the ADT used to generate the script, see Figs. 1.a, 2.a and 3.a). If one of the 10 clinical variable measures was missing for a given subject, then the weight for that variable was ignored in the final ADT score calculation.
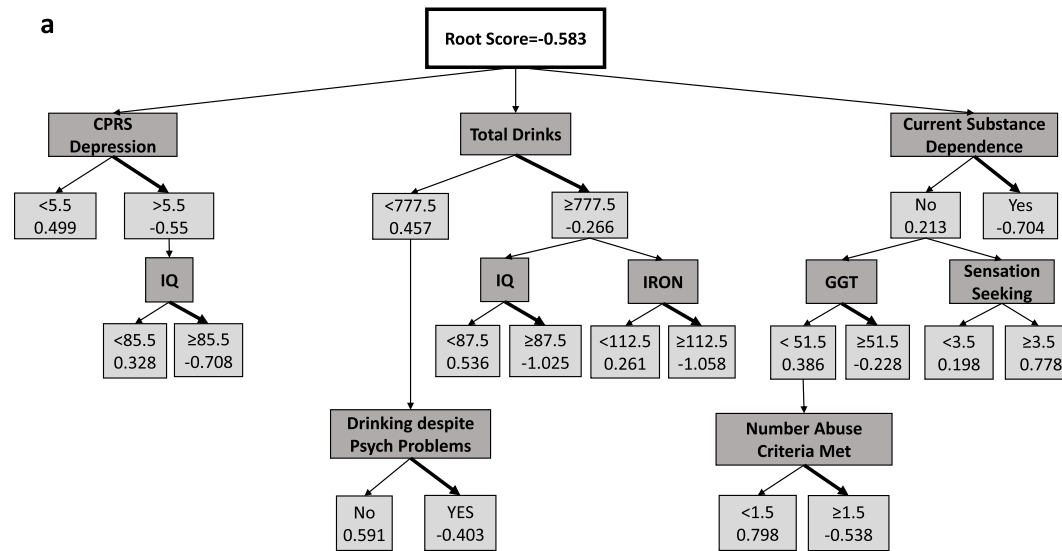
b: Confusion Matrix from Cross-validation with Training Subjects, Male Only (N=554)

	Classified as TX	Classified as Non-TX
Real TX	393	30
Real Non-TX	63	68
Correctly classified	461 (83.2%)	
Incorrectly classified	93 (16.7%)	
Kappa=0.49; Balanced Accuracy=73%		

c: Confusion Matrix from Test with Independent Dataset, Male Only (N=162)

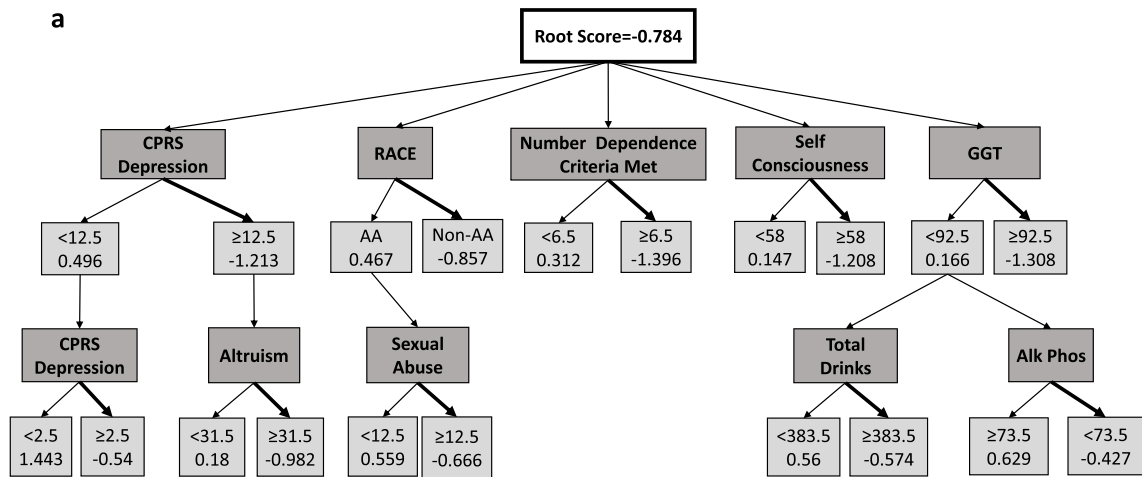|	Classified as TX	Classified as Non-TX
Real TX	87	17
Real Non-TX	20	38
Correctly classified	125 (77.2%)	
Incorrectly classified	22 (27.8%)	
Kappa=0.50; Balanced Accuracy=75%		

Fig. 2. Males only: a) Alternate Decision Tree (ADT) classifying N = 554 males with alcohol dependence (AD) as treatment or nontreatment-seekers. 10 characterization measures are noted in dark gray boxes. Classification: Treatment-seeking: total score <0; Nontreatment-seeking: total score >0. CPRS: Comprehensive Psychopathological Rating Scale; IQ: Wechsler Abbreviated Scale of Intelligence. b) Confusion Matrix from Cross-validation with Training Subjects. c) Confusion Matrix from Test with Independent Data (N = 79, males only). Tx = Treatment-seeking (bold arrow); Non-Tx = Not treatment-seeking.

To assess the power of the ADT method, four other classifier algorithms available in the WEKA tool were considered for comparison: [1] random forest, [2], random tree, [3] logistic regression, and [4] simple logistics model. The random forest method [33] started by constructing 100 decision trees using a subset of attributes chosen at random. The number of attributes in each tree was determined by: $\log_2 (178) + 1$, or 8 attributes. Classification of an individual subject was determined by the majority declaration of the trees. The random tree approach used a random search technique to construct a single large decision tree, in this case, with 576 nodes. The WEKA implementation of the logistic regression method included a "bridge estimator" to control convergence of the logistic probabilities [34]. The simple

logistics model was similar to logistic regression, except that only those variables that have the most effect on the least square error were kept [35].

3. Results

For the main dataset of 778 subjects, ADT found 10 measures that classified persons by treatment seeking status. These measures reflected quantity and quality of drinking, depression and drinking-related psychological problems, IQ, race, BMI as well as substance dependence (Fig. 1.a). Specifically, current alcohol consumption, measured in number of drinks in past 30 days (corresponding to ~>8 drinks per day),

a

b: Confusion Matrix from Cross-validation with Training Subjects, Female Only (N=224)

	Classified as TX	Classified as Non-TX
Real TX	175	11
Real Non-TX	10	28
Correctly classified	202 (90.2%)	
Incorrectly classified	22 (9.8%)	
Kappa=0.67; Balanced Accuracy=84%		

c: Confusion Matrix from Test with Independent Female Data (N=74)

	TX	Non-TX
Real TX	39	3
Real Non-TX	18	14
Correctly classified	53 (71.6%)	
Incorrectly classified	21 (21.1%)	
Kappa=0.40; Balanced Accuracy=69%		

Fig. 3. Females only: a) Alternate Decision Tree (ADT) classifying N = 224 females with alcohol dependence (AD) as treatment or nontreatment-seekers. 10 characterization measures are noted in dark gray boxes. Classification: Treatment-seeking: total score <0; Nontreatment-seeking: total score >0. CPRS: Comprehensive Psychopathological Rating Scale; GGT: gamma-glutamyl transferase; Alk Phos: alkaline phosphatase. b) Confusion Matrix from Cross-validation with Training Subjects. c) Confusion Matrix from Test with Independent Data (N = 36, females only). Tx = Treatment-seeking (bold arrow); Non-Tx = Not treatment-seeking.

any current substance dependence as assessed by DSM-IV, drinking despite psychological problems and alcohol abuse all classified individuals toward treatment seeking. Similarly, BMI <25.5, an IQ >87.5, depression score on the CPRS depression subscale, non-African American race classified toward treatment seeking. Elevated serum iron, GGT which are generally associated with excessive alcohol consumption were also associated with treatment seeking. Using the cross-validation method, 674/778 (86%) were classified correctly (Fig. 1.b). With the independent dataset (N = 236), 88/236 (78%) were correctly classified (Fig. 1.c). ADT scores generated for individual subjects in the independent validation dataset, (tabulated using a R script), ranged from −5.2 to 2.8.

The ADT for the males of the main sample (N = 554) (Fig. 2.a), contained nodes that were almost identical to that the non-gender-specific tree (Fig. 1.a), with the addition of *sensation seeking* from the

NEO-5-PI-R as a predictor node (low sensation seeking associated with treatment seeking) and absence of race and BMI as nodes. The correct classification rate using cross-validation for males was 83% (Fig. 2. b). With the independent validation dataset (N = 162 males), the correct classification rate was 77% (Fig. 2.c). For this ADT, the scores for the males in the independent validation dataset ranged from −4.9 to 2.6.

The ADT for females of the main sample (N = 224) (Fig. 3.a) included nodes for personality measures of self-consciousness and altruism, both of which were positively associated with treatment seeking. A history of sexual abuse was also associated with treatment seeking, as was the number of dependence criteria. The female only ADT classified 90% of females correctly using cross-validation (Fig. 3.b). With the independent dataset, (N = 74 females), it classified almost 72%

correctly (Fig. 3.c). For this ADT, the scores for the female patients in the independent dataset ranged from − 7.3 to 2.7

To test whether large ADT scores were associated with a higher confidence in the classification, the absolute values of the non-gender-specific ADT scores from the independent validation dataset (ADT_absscore) were grouped into 4 categories, as shown in Fig. 4. The distribution of ADT scores for the independent validation dataset is shown in Fig. 4.a. The proportion of subjects, in the independent validation data set, whose treatment seeking status was correctly assigned in each ADT score category for all subjects is shown in Fig. 4.b. Nearly 40% of patients with ADT_absscore <1 were mis-classified. We did not find a clear pattern of amount of missing data vs mis-classification. Fig. S1 shows the number of missing clinical variable entries per subject for the independent validation data set for each category of ADT score, in both correctly and incorrectly assigned groups.

Results of comparison analyses are summarized in Table 2. Each alternative method was inferior to the ADT method with respect to accuracy and kappa statistic as a classifier, except for the simple logistic model. In addition, the comparison decision tree approaches produced either a single large, complex decision tree with 576 nodes or a large number of small decision trees. Accordingly, neither random tree, random forest nor logistic regression was clinically useful. In contrast, the ADT approach did aim to reduce the large number of measures from the dataset into a streamlined battery composed of a subset of measures. The simple logistic model approached the ADT in accuracy and kappa statistic, but it required 24 variables (derived from 8 instruments requiring approximately 4–5 h testing time) to make a prediction. This is in contrast to 4 instruments in the ADT that can be completed in about 1½ hours.

4. Discussion

We used a WEKA data mining tool to reduce 178 clinical measures obtained in a clinical research setting to a subset of 10 that correctly classified over 85% of individuals with alcohol dependence (DSM-IV) by treatment seeking status (Fig. 1.b). These 10 measures became splitter nodes in an ADT that was used to predict treatment-seeking status in an independent dataset of subjects with an equivalent DSM-5 diagnosis: moderate or severe AUD (N = 236). Validation using an independent dataset is considered the gold standard for machine learning-generated classification systems [36]. With this independent dataset, subjects were classified with high accuracy, namely correct classification of 78% of individuals (Fig. 1.c). The four other comparison classification methods, two machine-learning and two logistic-type methods

Table 2
Accuracy and kappa statistic for each Random Forest, Random Tree and a logistic model compared to the ADT approach (bold).

	Cross validation		Independent validation	
	Accuracy	Kappa	Accuracy	Kappa
Random forest	84.8%	0.43	73.3%	0.36
Random tree	77.9%	0.37	70.0%	0.33
Logistic model	78.0%	0.40	72.5%	0.42
Simple logistic model	85.9%	0.57	77.5%	0.50
ADT	**86.1**%	**0.57**	**78.0**%	**0.51**

yielded either inferior accuracy and kappa statistic or required many more measures in order to classify.

The 10 nodes used by the ADT were not defined a priori, but were found by the hypothesis-free, data-driven ADT algorithm to best classify the group. The 10 splitter nodes in the ADT for the 778 individuals (Fig. 1.a), focused largely on drinking behavior, psychological symptoms and substance addiction: drinks per 30 days, drinking-related psychological problems, number of DSM-IV alcohol abuse criteria met, current substance dependence, as well as laboratory values of GGT and serum iron whose elevations are associated with excessive drinking. Of note, the 10 splitter nodes in the ADT that best classified treatment-seekers overlapped with measures that were significantly different between treatment- and nontreatment-seeking groups in the previous exploratory study [8], and the direction of the effect remained the same. Previous studies using data from community samples also found that treatment utilization was associated with comorbid mood, personality and drug use disorders [10] [37]. Notably, the factors from these community studies such as mood and substance use were also nodes in the ADT generated from the current study's research population despite the fact that in the clinical research setting, subjects with comorbid mental illness and/or substance use disorders are less prevalent.

Studies on racial disparities in treatment utilization for AUD are mixed with some indicating racial disparities and others indicating no racial differences in treatment utilization [38]. In the present study, race was one of the 10 nodes in the ADT and African American race was associated with not seeking treatment. However, following the logic of the ADT, the latter was true only in the context of a higher level of depression symptoms. It did not play a classifying role in the context of lower depression scores. Further, when race was omitted as one of the 178 measures, the classification accuracy, kappa value and nodes in the ADT were essentially unchanged indicating that race was not a strong classifier of treatment utilization, in contrast to depression or drinking behavior.

a

b
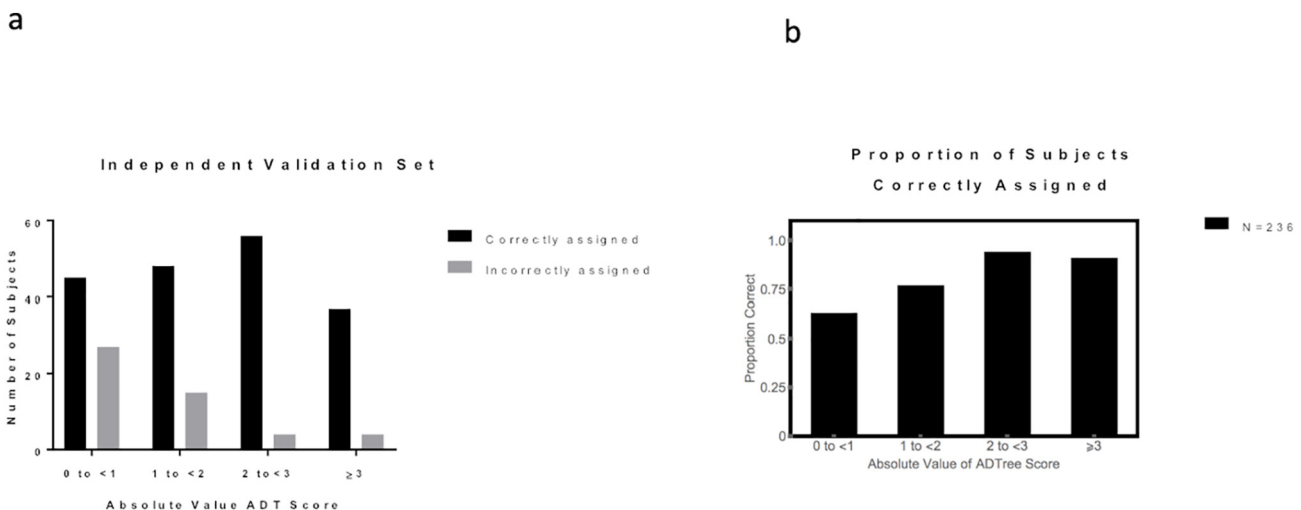
Fig. 4. Magnitude of ADT score for the independent validation dataset (N = 236) vs a) number of subjects correctly and incorrectly assigned to treatment seeking status in the independent validation dataset for all subjects; b) proportion of subjects correctly classified in each ADT score group.

For datasets where the group distribution is skewed, as is the case in the present study, (treatment- vs nontreatment seekers in the main dataset was 78 vs 22%; in the validation dataset, 62 vs 38%, respectively) the kappa statistic is considered a better evaluation of the accuracy of the classifier compared to the percentage of cases classified accurately. The kappa statistic compares the observed accuracy (proportion of cases classified accurately) to random chance (expected accuracy). The kappa statistics for the main dataset confusion matrix (Fig. 1.b) and for the validation dataset matrix (Fig. 1.c) were in the 'moderate' agreement range [39] compared to the random forest and random tree methods which were in 'fair' agreement. Kappa statistic for the two logistic methods was in the moderate ranges. The ADT generated using the WEKA program tended to overclassify nontreatment seekers as treatment seekers. Importantly, it misclassified very few treatment seekers (7%). Further, the balanced accuracy, of the main and validation datasets was 78 and 73%, respectively which denotes a classification accuracy of the majority-approximately 3 out of 4 subjects-overall.

Due to the gender imbalance in the sample, we constructed separate ADT's for males and females. The ADT for women consisted of distinctly different measures (nodes) that predicted treatment seeking for females. Different personality factors from the NEO-5-PI-R appeared in each gender-specific ADT. For men, sensation seeking (subscale of the extraversion factor of the NEO-5-PI-R) was one of 10 nodes and was associated with treatment seeking. Otherwise the male ADT was minimally changed from the ADT for the entire sample. In contrast, the female-only ADT included NEO-5-PI-R subfactors, altruism and self-consciousness, both of which were positively associated with treatment seeking. In the female ADT, a history of sexual abuse was a node and associated with treatment seeking. IQ was not represented in the ADT for women whereas depression as measured by CPRS appeared in 2 successive nodes. While the ADT for women alone was comprised of different nodes, as described, the classification accuracy of the ADT for women patient in the independent dataset (N = 74) was inferior (Fig. 3.c) to that of the non-gender-specific ADT constructed from the main dataset.

In the previous descriptive paper [8] many differences between treatment and nontreatment seekers with AUD persisted even when controlling for drinking patterns and quantity of alcohol consumption. When we removed these drinking measures from the analysis, the tree had many of the same nodes, but the kappa statistic when evaluating the independent data set fell to chance. There was the similar result when CPRS depression was removed from the analysis. This indicates that, although some of the 10 variables in the ADT are moderately correlated with each other (Supplementary Material), for the ADT analysis, (and possibly in clinical practice applications), drinking behavior variables and depression measures persist as strong treatment seeking predictors.

Machine learning approaches have been applied to guide clinical care in the treatment of alcohol use disorder [40]and substance use disorder [41]. Specifically, Connor and colleagues, compared a decision tree approach to conventional linear techniques to predict *relapse* after treatment for AUD using demographic, addiction severity, drinking behavior, craving, health and psychological data. Their classification accuracy (77%)[41]with a decision tree approach and validation with an independent dataset was similar to that reported here (78%,Fig. 1.c) With conventional techniques, the prediction accuracy was less than chance which is also similar to what we found when applying logistic regression to the main dataset.

In our clinical research setting for AUD, beyond simply asking patients whether they wish treatment for AUD, we examined with a broad array of clinical measures in multiple domains. Then we used these measures to construct an ADT that identified a cluster of factors that predicted treatment seeking. We can then assess the influence of these factors directly on the clinical study outcome measures of interest in the AUD clinical research setting. For example, we have identified that a higher level of drinking, other substance dependence, depression

and IQ is associated with seeking treatment. Therefore, nontreatment-seeking individuals who participate in human lab studies are, by inference, less depressed, drink less, have less comorbid substance dependence and lower IQ. These characteristics may affect the outcome of the experimental intervention such as alcohol consumption or cue-induced craving; therefore, the conclusions regarding clinical effectiveness drawn from these studies may not be generalizable to other populations with AUD. The extent to which these characteristics can interact with the mechanisms underlying a given experimental intervention should be considered as a source of variance in this domain of clinical research.

Despite the machine learning approach, this analysis has several limitations. The subjects were predominantly male, therefore, conclusions regarding females should be validated with a larger dataset of females. The dataset was imbalanced with respect to treatment- and nontreatment-seekers; this undoubtedly contributed to the poor accuracy classifying the nontreatment compared to treatment seekers. It is important to follow up on this finding with a more balanced dataset to determine whether this approach performs well to classify each group. With a more numerically balanced dataset, data from the clinical research setting can be leveraged to inform treatment in the primary care setting as the clinical research setting is composed of individuals with AUD who seek or do not seek treatment. This is in contrast to community AUD *treatment programs* where there are minimal numbers of nontreatment seekers. In effect, the AUD patients in a clinical research setting may be more representative of AUD patients presenting to the primary care setting, notwithstanding the different motivations underlying attendance at a primary care vs clinical research clinic.

Although the nodes closest to the root might be construed as being more important for classification, one should really make inferences about the likelihood of treatment seeking status by considering the magnitude of the absolute values of the score, which is a linear combination of the value of many nodes. This is a novel approach to identify characteristics associated with treatment utilization for AUD. The ADT machine learning technique misclassified very few treatment seekers; therefore, the clinical characteristics identified herein are important features of this group that can be used perhaps to guide patient management and treatment. In addition, the structure of the ADT can indicate how groupings of measures could indicate sub-phenotypes of AUD along the treatment-seeking spectrum.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eclinm.2019.05.008.

Author Contributions

MRL, VS, AH collected the data, researched literature and prepared figures. MRL, WGK, JJB, PGM interpreted and analyzed the data. MRL, JJB, PGM, LL prepared the manuscript. All authors read and approved the final version of the manuscript.

Declaration of Competing Interest

Dr. Leggio reports other from MCA, UK, other from Routledge, outside the submitted work. The other authors declare that they have nothing to disclose. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Acknowledgments

Funding

References

[1] Grant BF, Goldstein RB, Saha TD, Chou SP, Jung J, Zhang H, et al. Epidemiology of DSM-5 alcohol use disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions III. JAMA Psychiat 2015;72(8):757–66.

[2] Laramee P, Leonard S, Buchanan-Hughes A, Warnakula S, Daeppen JB, Rehm J. Risk of all-cause mortality in alcohol-dependent individuals: a systematic literature review and meta-analysis. EBioMedicine 2015;2(10):1394–404.

[3] Abuse S. Mental Health Services Administration (SAMHSA). Key substance use and mental health indicators in the United States: results from the 2015 national survey on drug use and health. Technical report SMA 17-5044. US Department of Health and Human Services, SAMHSA; 2017.

[4] Grant BF. Barriers to alcoholism treatment: reasons for not seeking treatment in a general population sample. J Stud Alcohol 1997;58(4):365–71.

[5] Zemore SE, Mulia N, Ye Y, Borges G, Greenfield TK. Gender, acculturation, and other barriers to alcohol treatment utilization among Latinos in three National Alcohol Surveys. J Subst Abuse Treat 2009;36(4):446–56.

[6] Litten RZ, Falk DE, Ryan ML, Fertig J, Leggio L. Advances in pharmacotherapy development: human clinical studies. Handb Exp Pharmacol 2018:579–613.

[7] Acion L, Zwick J, Rojas Saunero LP, Arndt S. Distinctions between seeking-and non-seeking-treatment research participants: implications for clinical trials effectiveness. Am J Drug Alcohol Abuse 2017;43(6):628–30.

[8] Rohn MC, Lee MR, Kleuter SB, Schwandt ML, Falk DE, Leggio L. Differences between treatment-seeking and nontreatment-seeking alcohol-dependent research participants: an exploratory analysis. Alcohol Clin Exp Res 2017;41(2):414–20.

[9] Kwako LE, Schwandt ML, Ramchandani VA, Diazgranados N, Koob GF, Volkow ND, et al. Neurofunctional domains derived from deep behavioral phenotyping in alcohol use disorder. Am J Psychiatry 2019:appiajp201818030357.

[10] Cohen E, Feinn R, Arias A, Kranzler HR. Alcohol treatment utilization: findings from the National Epidemiologic Survey on Alcohol and Related Conditions. Drug Alcohol Depend 2007;86(2):214–21.

[11] Dawson DA, Grant BF, Stinson FS, Chou PS, Huang B, Ruan W. Recovery from DSM-IV alcohol dependence: United States, 2001–2002. Addiction 2005;100 (3):281–92.

[12] Dawson DA. Correlates of past-year status among treated and untreated persons with former alcohol dependence: United States, 1992. Alcohol Clin Exp Res 1996; 20(4):771–9.

[13] Bischof G, Rumpf HJ, Meyer C, Hapke U, John U. Influence of psychiatric comorbidity in alcohol-dependent subjects in a representative population survey on treatment utilization and natural recovery. Addiction 2005;100(3):405–13.

[14] Hermos JA, Locastro JS, Glynn RJ, Bouchard GR, De Labry LO. Predictors of reduction and cessation of drinking in community-dwelling men: results from the normative aging study. J Stud Alcohol 1988;49(4):363–8.

[15] Allen JP, Mattson M, Miller W, Tonigan J, Connors G, Rychtarik R, et al. Matching alcoholism treatments to client heterogeneity. J Stud Alcohol 1997;58(1):7–29.

[16] Ray LA, Bujarski S, Yardley MM, Roche DJ, Hartwell EE. Differences between treatment-seeking and non-treatment-seeking participants in medication studies for alcoholism: do they matter? Am J Drug Alcohol Abuse 2017:1–8.

[17] Kwako LE, Momenan R, Litten RZ, Koob GF, Goldman D. Addictions neuroclinical assessment: a neuroscience-based framework for addictive disorders. Biol Psychiatry 2016;80(3):179–89.

[18] Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med 2001;23(1):89–109.

[19] Sobell LC, Sobell MB, Leo GI, Cancilla A. Reliability of a timeline method: assessing normal drinkers' reports of recent drinking and a comparative evaluation across several populations. Br J Addict 1988;83(4):393–402.

[20] Skinner HA, Sheu W-J. Reliability of alcohol use indices. The lifetime drinking history and the MAST. J Stud Alcohol 1982;43(11):1157–70.

[21] First M, Spitzer R, Gibbon M, Williams JB. Structured clinical interview for DSM-IV-TR axis I disorders, research version, non-patient edition. (SCID-I/NP). New York: Biometrics Research, New York State Psychiatric Institute; 2002 [November].

[22] Wechsler D. Wechsler abbreviated scale of intelligence. Psychological Corporation; 1999.

[23] Mattila-Evenden M, Svanborg P, Gustavsson P, Asberg M. Determinants of self-rating and expert rating concordance in psychiatric out-patients, using the affective subscales of the CPRS. Acta Psychiatr Scand 1996;94(6):386–96.

[24] Patton JH, Stanford MS, Barratt ES. Factor structure of the Barratt impulsiveness scale. J Clin Psychol 1995;51(6):768–74.

[25] Lynam DR, Smith GT, Whiteside SP, Cyders MA. The UPPS-P: assessing five personality pathways to impulsive behavior. West Lafayette, IN: Purdue University; 2006.

[26] Costa PT, McCrae RR. Normal personality assessment in clinical practice: the NEO personality inventory. Psychol Assess 1992;4(1):5.

[27] Buss AH, Perry M. The aggression questionnaire. J Pers Soc Psychol 1992;63(3): 452–9.

[28] McFarlane A, Clark CR, Bryant RA, Williams LM, Niaura R, Paul RH, et al. The impact of early life stress on psychophysiological, personality and behavioral measures in 740 non-clinical subjects. J Integr Neurosci 2005;4(1):27–40.

[29] Bernstein DP, Stein JA, Newcomb MD, Walker E, Pogge D, Ahluvalia T, et al. Development and validation of a brief screening version of the Childhood Trauma Questionnaire. Child Abuse Negl 2003;27(2):169–90.

[30] Freund Y, Mason L, editors. The alternating decision tree learning algorithm. icml; 1999.

[31] Optimizing the induction of alternating decision trees. In: Pfahringer B, Holmes G, Kirkby R, editors. Pacific-Asia conference on knowledge discovery and data mining. Springer; 2001.

[32] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. SIGKDD Explor Newslett 2009;11(1):10–8.

[33] Breiman L. Random forests. Mach Learn 2001;45(1):5–32.

[34] Le Cessie S, Van Houwelingen JC. Ridge estimators in logistic regression. Appl Stat 1992:191–201.

[35] Landwehr N, Hall M, Frank E. Logistic model trees. Mach Learn 2005;59(1–2): 161–205.

[36] Kohavi R, editor. A study of cross-validation and bootstrap for accuracy estimation and model selection, Ijcai 1995 [Montreal, Canada].

[37] Ilgen MA, Price AM, Burnett-Zeigler I, Perron B, Islam K, Bohnert AS, et al. Longitudinal predictors of addictions treatment utilization in treatment-naïve adults with alcohol use disorders. Drug Alcohol Depend 2011;113(2):215–21.

[38] Weisner C, Schmidt LA. Rethinking access to alcohol treatment. Recent Dev Alcohol 2001;15:107–36.

[39] Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 1977:363–74.

[40] Connor J, Symons M, Feeney G, Young RM, Wiles J. The application of machine learning techniques as an adjunct to clinical decision making in alcohol dependence treatment. Subst Use Misuse 2007;42(14):2193–206.

[41] 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.