Identifying Key Predictive Features for Opioid Use Disorder Using Machine Learning

Suraiya Akhter[1*], John H. Miller[2]

[1] School of Business and Technology, Emporia State University, Emporia, Kansas, USA

[2] School of Engineering and Applied Sciences, Washington State University, Richland, Washington, USA

[*]Corresponding author

Email: sakhter1@emporia.edu

Abstract

Opioid Use Disorder (OUD) continues to pose a pressing public health challenge across the United States, highlighting the critical need for early and accurate risk assessment tools that facilitate prompt prevention and intervention efforts. Machine learning methods have emerged as valuable tools for parsing complex medical datasets and aiding in clinical decisions. However, their effectiveness and interpretability largely rely on the appropriateness and quality of selected input variables. In this work, we conducted a comprehensive comparison of three distinct feature selection strategies—Alternating Decision Tree (ADT)-based scoring, Cross-Validated Feature Evaluation (CVFE), and Hypergraph-Based Feature Evaluation (HFE)—to identify the most predictive indicators of OUD. The analysis was performed using data from the 2023 National Survey on Drug Use and Health (NSDUH), a dataset compiled by RTI International under the direction of the Substance Abuse and Mental Health Services Administration (SAMHSA). This dataset encompasses a broad spectrum of features related to demographics, behavior, mental health, and substance usage. Each feature selection method yielded a set of important predictors, which were subsequently used to train eXtreme Gradient Boosting (XGBoost) classification models. Their respective performances were evaluated and compared, with the model informed by CVFE-selected features achieving the best outcomes—demonstrating a predictive accuracy of 79.11% and an area under the receiver operating characteristic (ROC) curve of 0.8652. To enhance model transparency and interpretability, SHapley Additive exPlanations (SHAP) were employed to illustrate the influence of individual variables on model predictions. The findings highlight the crucial importance of effective feature selection in enhancing both model accuracy and interpretability, ultimately supporting the development of practical, data-driven approaches that may help healthcare providers assess OUD risk and tailor prevention strategies to individual needs.

Introduction

Since the early 2000s, the United States has experienced a dramatic escalation in both opioid use disorder (OUD) and overdose-related fatalities, culminating in the declaration of a nationwide public health emergency in 2017 [1, 2]. In that year alone, over 70,000 drug overdose deaths occurred, with a significant rise in cases involving fentanyl and its analogs [3]. Although opioid prescription rates have steadily declined since 2012, many individuals continue to receive long-term opioid treatments, which are strongly linked to heightened OUD risk [4-6]. In primary care, a substantial proportion of patients on chronic opioid therapy meet diagnostic criteria for OUD [5], and a nontrivial percentage of pain patients develop prescription-related OUD [6]. In the United States, millions are affected by OUD across both prescription and non-prescription opioid use [7, 8].

Traditional tools for identifying OUD risk—such as screening questionnaires and urine drug screens (UDS)—have important limitations. UDS can fail to detect drug use, especially when not directly observed, and self-reports are often inaccurate [9-11]. Consequently, many patients with OUD remain undiagnosed, leading to elevated risks for overdose and poor outcomes. Despite weak evidence supporting the long-term benefits of opioid therapy, healthcare providers frequently continue prescribing opioids for chronic pain conditions [12]. However, accurately identifying individuals vulnerable to OUD before starting opioid treatment remains a significant challenge. Established tools like the Opioid Risk Tool (ORT) and SOAPP-R have shown only limited utility [13]. Regression-based studies have attempted to identify predictive factors [14-19], but these approaches often struggle with issues such as class imbalance and poor generalizability [20, 21].

The application of machine learning techniques has gained considerable attention in the prediction of opioid use disorder (OUD), largely due to their ability to analyze and interpret high-dimensional datasets derived from electronic health records (EHRs). These algorithms have been increasingly utilized to identify individuals at elevated risk by drawing from diverse data environments, including clinical records and commercial data repositories, and by modeling patterns across medical history, healthcare utilization, and behavioral attributes [3, 22-27]. A broad array of factors—such as gender, racial or ethnic background, educational attainment, household income, marital and employment status, geographic context, history of substance use, psychiatric conditions, prior engagement with treatment services, chronic diseases, criminal justice interactions, and type of health insurance—can all influence patient outcomes. However, the intricate relationships among demographic, behavioral, and clinical variables continue to pose challenges for reliable risk differentiation. To date, limited research has systematically assessed the significance or comparative influence of these individual factors. There is a clear demand for advanced machine learning models capable of evaluating feature importance across heterogeneous population groups.

We hypothesize that a machine learning framework can be effectively used to autonomously discern and prioritize critical variables for accurate risk estimation, thereby enabling more customized and data-informed approaches to care. Such models could inform tailored treatment decisions and improve care for patients with—or at risk of—OUD and associated comorbidities. To address this, we used data from the 2023 National Survey on Drug Use and Health (NSDUH), conducted by RTI International for the Substance Abuse and Mental Health Services Administration (SAMHSA). We built a foundational machine learning pipeline aimed at forecasting OUD risk, with a strong emphasis on comprehensive feature selection methodologies. By systematically investigating both the individual impact of predictor variables, our goal is to optimize the relevance of clinical and behavioral information in guiding effective, person-centered care. This work aspires to contribute to the growing field of precision medicine and reinforce the role of machine learning in addressing public health challenges.

Materials and methods

An overview of the proposed methodology is depicted in Figure 1. The workflow begins with the collection of data for two groups—individuals diagnosed with opioid use disorder (positive cases) and those without the condition (negative cases). From this data, a wide range of potential features is generated. Feature evaluation techniques are then applied, including Alternating Decision Tree (ADT) [28, 29], Cross-Validated Feature Elimination (CVFE) [30], and hypergraph-based feature evaluation (HFE) [31], to systematically eliminate features that contribute minimal or no predictive value. The refined feature subsets are subsequently utilized to train a machine learning algorithm, enabling the evaluation of model performance in predicting OUD outcomes.

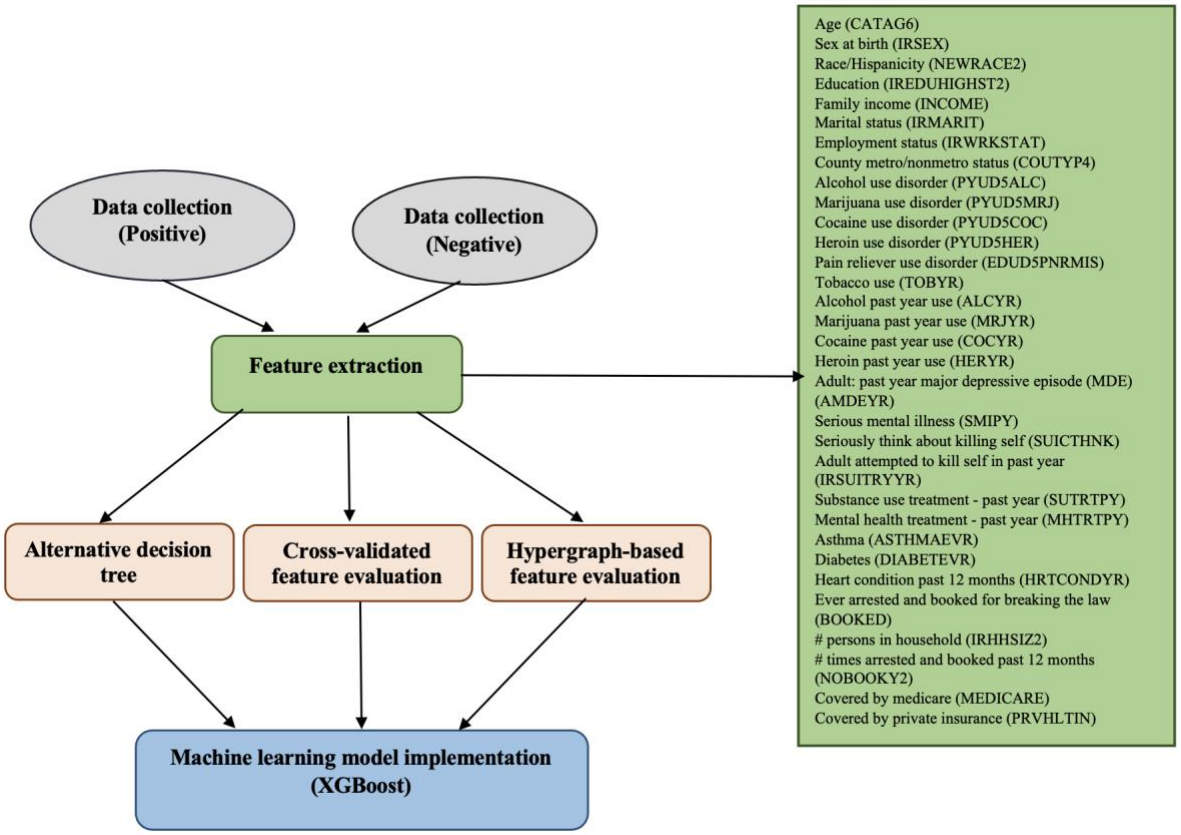


Figure 1 - Workflow illustrating the procedure for opioid use disorder (OUD) detection.

Study population and features

This study employed data from the 2023 edition of the National Survey on Drug Use and Health (NSDUH), administered by RTI International on behalf of the Substance Abuse and Mental Health Services Administration (SAMHSA) [32]. The dataset comprises an extensive set of variables covering demographic details, behavioral patterns, mental health indicators, and substance use information. The final dataset included 791 participants diagnosed with opioid use disorder and 42,607 participants without the condition. To mitigate the issue of class imbalance, random undersampling was applied to the non-OUD group, reducing it to 791 cases and achieving a balanced dataset across both classes. To build the predictive model, 80% of the dataset was allocated for training purposes, and the remaining 20% was set aside for performance evaluation. A detailed breakdown of the variables, their categorical groupings, associated demographic distributions, and statistical significance values is presented in Table 1. Features exhibiting a *p*-value below 0.05 were regarded as statistically significant indicators. The target outcome—opioid use disorder—was defined based on the variable UD5OPIANY (Opioid Use Disorder - Past Year Users), where a value of 1 (Yes) indicated classification as an OUD case. In total, 32 features were shortlisted for analysis, all of which are listed in Supplementary Table S1 **(Supplementary Material)**.

Table 1 – Overview of participants' demographic and clinical profiles.

Characteristics		All subjects (N = 1582)	CVD risk (N = 791)	No CVD risk (N = 791)	p-value [a]
CATAG6 – Age Category, N (%)	2 (18-25 Years Old)	426 (26.93%)	159 (20.1%)	267 (33.75%)	<0.0001
	3 (26-34 Years Old)	320 (20.23%)	162 (20.48%)	158 (19.97%)	
	4 (35-49 Years Old)	467 (29.52%)	263 (33.25%)	204 (25.79%)	
	5 (50-64 Years Old)	202 (12.77%)	115 (14.54%)	87 (11.0%)	
	6 (65 or Older)	167 (10.56%)	92 (11.63%)	75 (9.48%)	
IRSEX – Sex at Birth, N (%)	1 (Male)	703 (44.44%)	354 (44.75%)	349 (44.12%)	0.8396
	2 (Female)	879 (55.56%)	437 (55.25%)	442 (55.88%)	
NEWRACE2 -Race/Hispanicity, N (%)	1 (NonHisp White)	921 (58.22%)	471 (59.54%)	450 (56.89%)	0.0321

	2 (NonHisp Black/Afr Am)	195 (12.33%)	93 (11.76%)	102 (12.9%)	
	3 (NonHisp Native Am/AK Native)	41 (2.59%)	29 (3.67%)	12 (1.52%)	
	4 (NonHisp Native HI/Other Pac Isl)	8 (0.51%)	4 (0.51%)	4 (0.51%)	
	5 (NonHisp Asian)	48 (3.03%)	19 (2.4%)	29 (3.67%)	
	6 (NonHisp more than one race)	73 (4.61%)	41 (5.18%)	32 (4.05%)	
	7 (Hispanic)	296 (18.71%)	134 (16.94%)	162 (20.48%)	
IREDUHIGHST2 - Education, N (%)	1 (Fifth grade or less grade completed)	13 (0.82%)	9 (1.14%)	4 (0.51%)	<0.0001
	2 (Sixth grade completed)	4 (0.25%)	1 (0.13%)	3 (0.38%)	
	3 (Seventh grade completed)	7 (0.44%)	6 (0.76%)	1 (0.13%)	
	4 (Eighth grade completed)	15 (0.95%)	11 (1.39%)	4 (0.51%)	
	5 (Ninth grade completed)	33 (2.09%)	27 (3.41%)	6 (0.76%)	
	6 (Tenth grade completed)	39 (2.47%)	28 (3.54%)	11 (1.39%)	
	7 (Eleventh or Twelfth grade completed, no diploma)	127 (8.03%)	84 (10.62%)	43 (5.44%)	
	8 (High school diploma/GED)	487 (30.78%)	256 (32.36%)	231 (29.2%)	
	9 (Some college credit, but no degree)	340 (21.49%)	182 (23.01%)	158 (19.97%)	
	10 (Associate's degree (for example, AA, AS))	139 (8.79%)	76 (9.61%)	63 (7.96%)	
	11 (College graduate or higher)	378 (23.89%)	111 (14.03%)	267 (33.75%)	
INCOME – Total Family Income, N (%)	1 (Less than $20,000)	380 (24.02%)	251 (31.73%)	129 (16.31%)	<0.001
	2 ($20,000 - $49,999)	460 (29.08%)	245 (30.97%)	215 (27.18%)	
	3 ($50,000 - $74,999)	220 (13.91%)	107 (13.53%)	113 (14.29%)	

	4 ($75,000 or More)	522 (33.0%)	188 (23.77 %)	334 (42.23 %)	
IRMARIT – Marital Status, N (%)	1 (Married)	530 (33.5%)	218 (27.56 %)	312 (39.44 %)	<0.001
	2 (Windowed)	61 (3.86%)	47 (5.94%)	14 (1.77%)	
	3 (Divorced or Separated)	220 (13.91%)	144 (18.2%)	76 (9.61%)	
	4 (Never Been Married)	771 (48.74%)	382 (48.29 %)	389 (49.18 %)	
IRWRKSTAT – Employment Status, N (%)	1 (Employed full time)	639 (40.39%)	247 (31.23 %)	392 (49.56 %)	<0.001
	2 (Employed part time)	216 (13.65%)	92 (11.63%)	124 (15.68 %)	
	3 (Unemployed)	124 (7.84%)	73 (9.23%)	51 (6.45%)	
	4 (Other (incl. not in labor force)	603 (38.12%)	379 (47.91 %)	224 (28.32 %)	
COUTYP4 – County Metro/Nonmetro Status, N (%)	1 (Large Metro)	672 (42.48%)	314 (39.7%)	358 (45.26 %)	0.0806
	2 (Small Metro)	607 (38.37%)	317 (40.08 %)	290 (36.66 %)	
	3 (Nonmetro)	303 (19.15%)	160 (20.23 %)	143 (18.08 %)	
PYUD5ALC - Alcohol Use Disorder in the Past Year, N (%)	0 (No)	530 (33.5%)	165 (20.86 %)	365 (46.14 %)	<0.001
	1 (Yes)	316 (19.97%)	205 (25.92 %)	111 (14.03 %)	
	91 (Never Used Alcohol)	225 (14.22%)	95 (12.01%)	130 (16.43 %)	
	93 (Did Not Use Alcohol Past 12 Months or Used)	511 (32.3%)	326 (41.21 %)	185 (23.39 %)	
PYUD5MRJ - Marijuana Use Disorder in the Past Year, N (%)	0 (No)	224 (14.16%)	130 (16.43 %)	94 (11.88%)	<0.001
	1 (Yes)	291 (18.39%)	208 (26.3%)	83 (10.49%)	
	91 (Never Used Marijuana)	583 (36.85%)	208 (26.3%)	375 (47.41 %)	

	93 (Did Not Use Marijuana Past 12 Months or Used Less Than 6 Days)	484 (30.59%)	245 (30.97%)	239 (30.21%)	
PYUD5COC - Cocaine Use Disorder in the Past Year, N (%)	0 (No)	85 (5.37%)	58 (7.33%)	27 (3.41%)	<0.001
	1 (Yes)	62 (3.92%)	59 (7.46%)	3 (0.38%)	
	91 (Never Used Cocaine)	1078 (68.14%)	401 (50.7%)	677 (85.59%)	
	93 (Did Not Use Cocaine in the Past 12 Months)	357 (22.57%)	273 (34.51%)	84 (10.62%)	
PYUD5HER - Heroin Use Disorder in the Past Year, N (%)	0 (No)	15 (0.95%)	14 (1.77%)	1 (0.13%)	<0.001
	1 (Yes)	105 (6.64%)	105 (13.27%)	0 (0.00%)	
	91 (Never Used Heroin)	1334 (84.32%)	555 (70.16%)	779 (98.48%)	
	93 (Did Not Use Heroin in the Past 12 Months)	128 (8.09%)	117 (14.79%)	11 (1.39%)	
EDUD5PNRMIS - Pain Reliever Use Disorder, Py Misusers, N (%)	0 (No)	47 (2.97%)	26 (3.29%)	21 (2.65%)	<0.001
	1 (Yes)	328 (20.73%)	328 (41.47%)	0 (0.00%)	
	91 (Never Misused Pain Relievers)	1057 (66.81%)	342 (43.24%)	715 (90.39%)	
	93 (Did Not Misuse Pain Reliever in the Past 12 Months)	150 (9.48%)	95 (12.01%)	55 (6.95%)	
TOBYR - Any Tobacco - Past Year Use, N (%)	0 (Did not use in the past year)	930 (58.79%)	351 (44.37%)	579 (73.2%)	<0.001
	1 (Used within the past year)	652 (41.21%)	440 (55.63%)	212 (26.8%)	
ALCYR - Alcohol - Past Year Use, N (%)	0 (Did not use in the past year)	556 (35.15%)	315 (39.82%)	241 (30.47%)	<0.001
	1 (Used within the past year)	1026 (64.85%)	476 (60.18%)	550 (69.53%)	
MRJYR - Marijuana - Past Year Use, N (%)	0 (Did not use in the past year)	977 (61.76%)	405 (51.2%)	572 (72.31%)	<0.001

	1 (Used within the past year)	605 (38.24%)	386 (48.8%)	219 (27.69%)	
COCYR - Cocaine - Past Year Use, N (%)	0 (Did not use in the past year)	1435 (90.71%)	674 (85.21%)	761 (96.21%)	<0.001
	1 (Used within the past year)	147 (9.29%)	117 (14.79%)	30 (3.79%)	
HERYR - Heroin - Past Year Use, N (%)	0 (Did not use in the past year)	1479 (93.49%)	689 (87.1%)	790 (99.87%)	<0.001
	1 (Used within the past year)	103 (6.51%)	102 (12.9%)	1 (0.13%)	
AMDEYR - Adult: Past Year Major Depressive Episode (MDE), N (%)	1 (Yes)	321 (20.29%)	242 (30.59%)	79 (9.99%)	<0.001
	2 (No)	1261 (79.71%)	549 (69.41%)	712 (90.01%)	
SMIPY – Serious Mental Illness (SMI), N (%)	0 (No Past Year SMI)	1321 (83.5%)	578 (73.07%)	743 (93.93%)	<0.001
	1 (Past Year SMI)	261 (16.5%)	213 (26.93%)	48 (6.07%)	
SUICTHNK - Seriously Think About Killing Self Past 12 Months, N (%)	1 (Yes)	200 (12.64%)	153 (19.34%)	47 (5.94%)	<0.001
	2 (No)	1377 (87.04%)	635 (80.28%)	742 (93.81%)	
	97 (Refused)	5 (0.32%)	3 (0.38%)	2 (0.25%)	
IRSUITRYYR - Adult Attempted to Kill Self in Past Year, N(%)	0 (No)	1531 (96.78%)	747 (94.44%)	784 (99.12%)	<0.001
	1 (Yes)	51 (3.22%)	44 (5.56%)	7 (0.88%)	
SUTRTPY - RC-RCVD Substance Use Treatment - Past Year, N (%)	0 (No)	1245 (78.7%)	490 (61.95%)	755 (95.45%)	<0.001
	1 (Yes)	337 (21.3%)	301 (38.05%)	36 (4.55%)	
MHTRTPY - RC-RCVD Mental Health Treatment - Past Year, N (%)	0 (No)	938 (59.29%)	370 (46.78%)	568 (71.81%)	<0.001
	1 (Yes)	644 (40.71%)	421 (53.22%)	223 (28.19%)	
ASTHMAEVR- Ever Told Had Asthma, N (%)	1 (Yes)	248 (15.68%)	143 (18.08%)	105 (13.27%)	<0.001

	2 (No)	471 (29.77%)	298 (37.67%)	173 (21.87%)	
	94 (Don't Know)	7 (0.44%)	4 (0.51%)	3 (0.38%)	
	97 (Refused)	3 (0.19%)	2 (0.25%)	1 (0.13%)	
	99 (Legitimate Skip)	853 (53.92%)	344 (43.49%)	509 (64.35%)	
DIABETEVR - Ever Told Had Diabetes/Sugar Diabetes, N (%)	1 (Yes)	173 (10.94%)	112 (14.16%)	61 (7.71%)	<0.001
	2 (No)	546 (34.51%)	329 (41.59%)	217 (27.43%)	
	94 (Don't Know)	7 (0.44%)	4 (0.51%)	3 (0.38%)	
	97 (Refused)	3 (0.19%)	2 (0.25%)	1 (0.13%)	
	99 (Legitimate Skip)	853 (53.92%)	344 (43.49%)	509 (64.35%)	
HRTCONDYR - Had Heart Condition Past 12 Months, N (%)	1 (Yes)	83 (5.25%)	60 (7.59%)	23 (2.91%)	<0.001
	2 (No)	67 (4.24%)	42 (5.31%)	25 (3.16%)	
	5 (Yes, logically assigned from skip pattern)	11 (0.7%)	9 (1.14%)	2 (0.25%)	
	98 (Blank (No Answer))	10 (0.63%)	6 (0.76%)	4 (0.51%)	
	99 (Legitimate Skip)	1411 (89.19%)	674 (85.21%)	737 (93.17%)	
BOOKED - Ever Arrested and Booked for Breaking the Law, N (%)	1 (Yes)	393 (24.84%)	293 (37.04%)	100 (12.64%)	<0.001
	2 (No)	1172 (74.08%)	486 (61.44%)	686 (86.73%)	
	3 (Yes, logically assigned)	14 (0.88%)	9 (1.14%)	5 (0.63%)	
	97 (Refused)	3 (0.19%)	3 (0.38%)	0 (0.00%)	
IRHHSIZ2 - # Persons in Household (HH), N (%)	1 (One person in household)	213 (13.46%)	111 (14.03%)	102 (12.9%)	0.0068
	2 (Two people in household)	490 (30.97%)	241 (30.47%)	249 (31.48%)	
	3 (Three people in household)	331 (20.92%)	186 (23.51%)	145 (18.33%)	

	4 (Four people in household)	295 (18.65%)	133 (16.81%)	162 (20.48%)	
	5 (Five people in household)	140 (8.85%)	56 (7.08%)	84 (10.62%)	
	6 (6 or more people in household)	113 (7.14%)	64 (8.09%)	49 (6.19%)	
NOBOOKY2 - # Times Arrested and Booked Past 12 Months, N (%)	0 (None)	302 (19.09%)	214 (27.05%)	88 (11.13%)	<0.001
	1 (One time)	55 (3.48%)	46 (5.82%)	9 (1.14%)	
	2 (Two times)	23 (1.45%)	20 (2.53%)	3 (0.38%)	
	3 (Three or more times)	12 (0.76%)	12 (1.52%)	0 (0.00%)	
	994 (Don't Know)	1 (0.06%)	1 (0.13%)	0 (0.00%)	
	997 (Refused)	3 (0.19%)	3 (0.38%)	0 (0.00%)	
	998 (Blank (No Answer))	14 (0.88%)	9 (1.14%)	5 (0.63%)	
	999 (Legitimate Skip)	1172 (74.08%)	486 (61.44%)	686 (86.73%)	
MEDICARE - Covered by Medicare, N (%)	1 (Yes)	277 (17.51%)	184 (23.26%)	93 (11.76%)	<0.001
	2 (No)	1276 (80.66%)	594 (75.09%)	682 (86.22%)	
	94 (Don't know)	6 (0.38%)	1 (0.13%)	5 (0.63%)	
	97 (Refused)	1 (0.06%)	1 (0.13%)	0 (0.00%)	
	98 (Blank (No Answer))	22 (1.39%)	11 (1.39%)	11 (1.39%)	
PRVHLTIN - Covered by Private Insurance, N (%)	1 (Yes)	731 (46.21%)	267 (33.75%)	464 (58.66%)	<0.001
	2 (No)	815 (51.52%)	508 (64.22%)	307 (38.81%)	
	94 (Don't Know)	12 (0.76%)	4 (0.51%)	8 (1.01%)	
	98 (Blank (No Answer))	24 (1.52%)	12 (1.52%)	12 (1.52%)	

[a] Chi-Squared test

N = Number of individuals

Evaluation of features

Prior to model development, removing irrelevant or low-importance features is critical to enhance predictive accuracy. In this study, we applied three distinct feature selection methods—Alternating Decision Tree (ADT) [28, 29], Cross-Validated Feature Elimination (CVFE) [30], and Hypergraph-based Feature Evaluation (HFE) [31]—to systematically discard less informative variables. The ADT algorithm combines the straightforward interpretability of a traditional decision tree with the superior prediction capability afforded by boosting methods. It encodes knowledge as a tree structure composed of interconnected decision stumps, a technique often employed in boosting frameworks. Unlike conventional trees where branches are mutually exclusive, ADTs allow overlapping branches, enabling more flexible decision paths. The model's root is a prediction node that assigns a numerical score, followed by layers of decision nodes representing predicate conditions or tests. These layers alternate between prediction and decision nodes throughout the tree. Importantly, prediction nodes appear both as the root and as terminal nodes (leaves), distinguishing ADTs from classic decision trees and highlighting their unique architecture and operational logic.

The ADT constructs a collection of rules, each comprising a prerequisite, a condition, and two corresponding scores. Conditions take the form of predicates structured as "attribute <comparison> value," while prerequisites are logical conjunctions of such conditions. Rule evaluation proceeds through nested conditional statements, using the associated scores to generate predictions for individual data points. The process begins with a root rule, whose prerequisite and condition are both set to "true," and whose scores are computed based on the weighted distribution of training samples. Initially, all training instances are assigned equal weights, calculated as the reciprocal of the total number of samples. The algorithm then proceeds iteratively, generating new rules by selecting the optimal pairing of prerequisites and conditions that minimize a specific objective function, denoted as z. This function quantifies how effectively a rule separates positive from negative instances, guiding the search for the best rule to add. At each step, the scores for the newly created rule are updated using a boosting strategy, and the weights of training samples are adjusted in accordance with the accuracy of the new rule's predictions. The training cycle continues until a stopping criterion is met, which may be a predetermined maximum number of iterations or a threshold for minimal accuracy improvement. The ensemble of rules produced by the ADT forms an alternating decision tree, characterized by prediction nodes containing numeric values, and structured according to the logical prerequisites defined by each successive rule. Notably, the ADT model employs only a subset of the total available features in its construction. In our

implementation, we evaluated 50 randomly chosen values for the boosting iteration parameter, *B*. The resulting ADT-derived decision tree is displayed in Supplementary Figure S1, highlighting the subset of 32 candidate features selected. This process yielded a final set of 10 features, detailed in Supplementary Table S2 (**Supplementary Material**).

In addition, we incorporated the cross-validated feature evaluation (CVFE) approach together with a hypergraph-based technique to refine the initial pool of features. The step-by-step process of the CVFE method is illustrated in Figure 2. Initially, the dataset was randomly divided into *c* distinct subsets. For each subset, the XGBoost algorithm was applied to determine the most influential features, with model hyperparameters optimized through a grid search strategy. Feature selection was conducted separately on every subset, after which an intersected feature set—comprising features common across all subsets—was identified. This entire procedure was repeated *e* times, generating *e* intersected feature sets in total. To finalize the feature selection, any feature appearing in at least ($p \times 100$)% of these intersected sets was included in the ultimate chosen feature set. Table 2 presents the count of features identified in these subsets across various configurations of the parameters *c*, *e*, and *p*. The detailed lists of these key features can be found in Supplementary Tables S3-S7, located in the **Supplementary Material**.

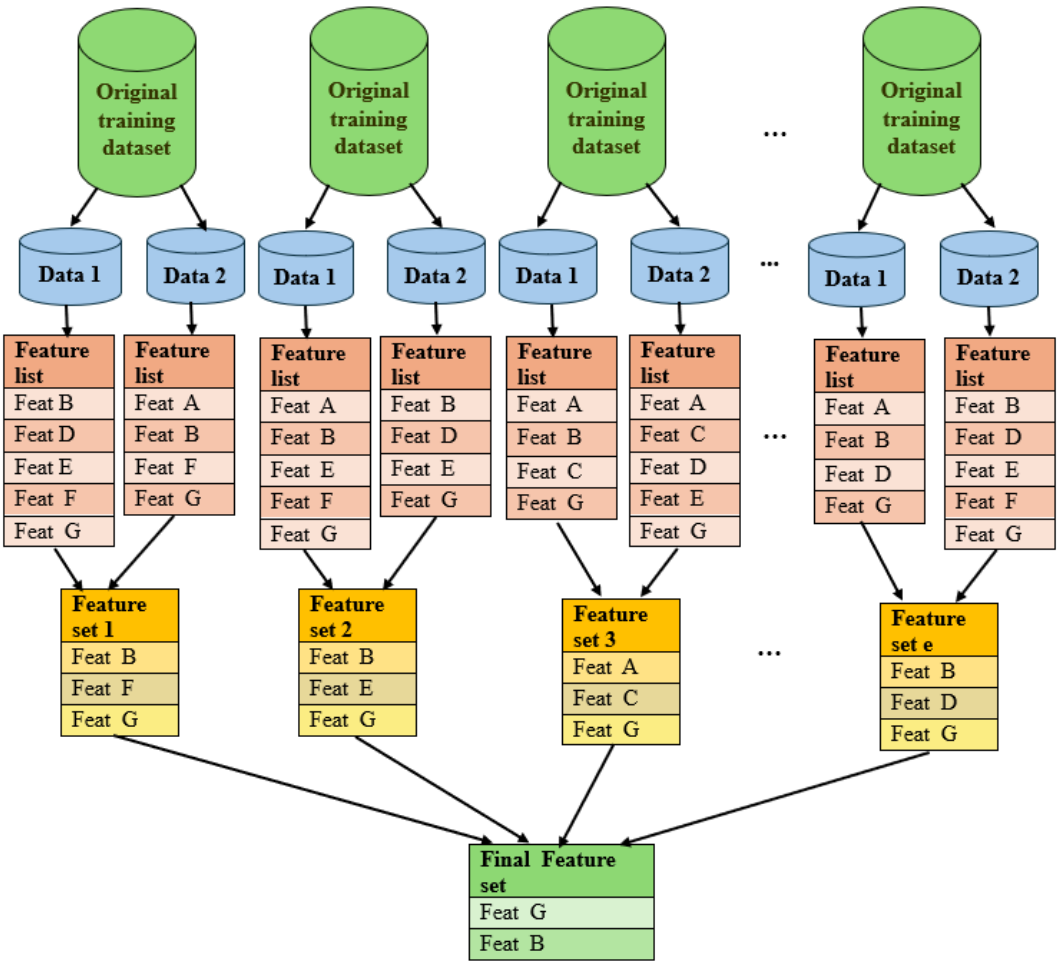


Figure 2 - Illustration of the feature selection process employing the cross-validated feature evaluation (CVFE) method.

The hypergraph-based feature evaluation (HFE) technique involves generating a hypergraph structure from the set of input features. Unlike standard graphs, which are denoted as G(V, E) with vertices (V) connected by binary edges (E), a hypergraph generalizes this concept by permitting each edge—referred to as a hyperedge—to link multiple vertices simultaneously. Formally, a hypergraph is represented as G(V, E), where V is the set of nodes and E comprises the hyperedges, each of which is a subset of V. Figure 3 provides a visual comparison between a traditional graph and a hypergraph, while Supplementary Figure S2 (**Supplementary Material**) presents the hypergraph structure derived from the full training dataset used in this study. In the HFE framework, feature relevance is determined by assigning weights to the hyperedges associated with specific feature values. These importance scores are computed using the principle of hypergraph cut conductance minimization [31], resulting in a feature rating vector that reflects the relative influence of each feature. The scores in this vector are ranked, and the top z features are selected. The value of z is defined as $\beta \times m$, where m is the total number of features and β represents the desired proportion of features to retain. In our analysis, the resulting number of selected features for

different values of β are listed in Table 2. Supplementary Tables S8-S10 (**Supplementary Material**) provide the specific feature subsets obtained for β values of 25, 50, and 75, respectively.

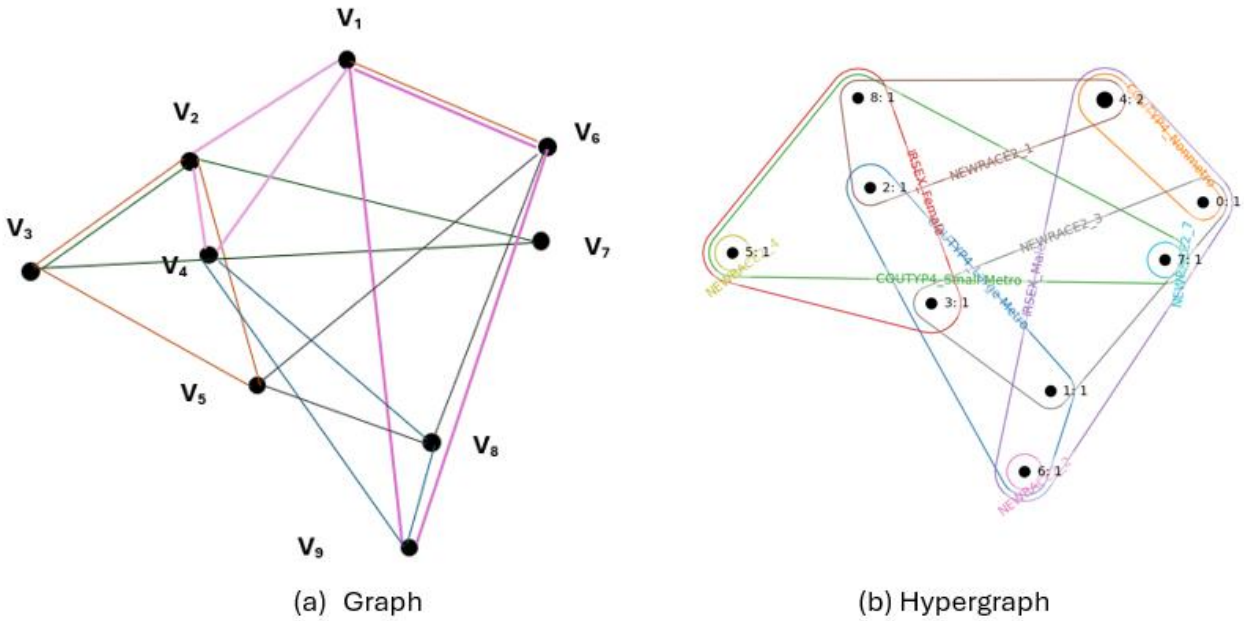


Figure 3 – Visualization comparing the structural differences between a standard graph and a hypergraph.

Code and data availability

All experimental datasets and corresponding code can be accessed through the following repository: https://github.com/suraiya14/OUDP, as also referenced in the **Supplementary Material**.

Results

Following the reduction of the original feature set using three independent feature selection strategies, we developed individual predictive models based on the refined subsets, utilizing the XGBoost algorithm [33]. To interpret and quantify the impact of each feature on model outcomes, we applied the Shapley Additive Explanations (SHAP) method [34]. SHAP assigns values that measure the marginal contribution of each variable to the prediction outputs produced by the machine learning model.

Evaluation of model performance

The XGBoost algorithm was trained using multiple feature subsets obtained through ADT, CVFE and HFE methodologies in conjunction with the training data. Model performance on the test dataset was measured using the metrics defined in Equations 1-4, where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. Among the evaluation criteria, accuracy reflects the ratio of correctly predicted cases to the overall number of observations in the dataset, serving as a key indicator of the classifier's predictive capability.

$$Test_{Acc} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Test_{recall} = \frac{TP}{TP+FN} \tag{2}$$

$$Test_{precision} = \frac{TP}{TP+FP} \tag{3}$$

$$Test_{F-1} = 2 \times \frac{(Test_{precision} \times Test_{recall})}{(Test_{precision} + Test_{recall})} \tag{4}$$

In addition, we evaluated the model using recall and precision metrics. Recall assesses how effectively the model identifies genuine positive instances, whereas precision represents the ratio of correctly predicted positives to the total number of positive predictions made by the model. To provide a balanced measure of both recall and precision, we computed the F1 score, which is the harmonic mean of the two. We also assessed the model's binary classification performance using the Area Under the Receiver Operating Characteristic Curve (AUC). Higher AUC values indicate stronger discriminative capability, with a score of 1 representing perfect classification and 0.5 corresponding to performance no better than random guessing.

Table 2 - Performance evaluation on the test set, detailing the count of selected features and associated metrics: accuracy, precision, recall, F1 score, and AUC for each feature selection method.

Feature evaluation algorithm	**Configuration**	**Number of features**	$Test_{Acc}$	$Test_{Precision}$	$Test_{Recall}$	$Test_{F1}$	$Test_{AUC}$
ADT	$B = 50$	10	0.769	0.7931	0.7278	0.7591	0.8245

CVFE	(c = 2, e = 5, p = 0.2)	30	0.7911	0.7949	0.7848	0.7898	0.8652
	(c = 3, e = 5, p = 0.6)	29	0.788	0.7898	0.7848	0.7873	0.8644
	(c = 4, e = 10, p = 0.6)	29	0.7911	0.7949	0.7848	0.7898	0.8652
	(c = 8, e = 5, p = 0.6)	21	0.788	0.8013	0.7658	0.7832	0.8511
	(c = 10, e = 10, p = 0.2)	21	0.7722	0.7829	0.7532	0.7677	0.8449
HFE	β = 25	8	0.75	0.8110	0.6519	0.7228	0.8266
	β = 50	16	0.7816	0.7908	0.7658	0.7781	0.8382
	β = 75	24	0.7468	0.7532	0.7342	0.7436	0.8404

Legend:

$Test_{acc}$: Accuracy on the testing dataset

$Test_{precision}$: Precision on the testing dataset

$Test_{recall}$: Recall on the testing dataset

$Test_{F1}$: F1 score on the testing dataset

$Test_{AUC}$: AUC on the testing dataset

B: Number of boosting cycles

c: Number of disjoint sub-parts

e: Number of repeated runs

p: Proportions of repeated runs for extracting common features

β: percentage of feature selected

Table 2 presents a comparative assessment of XGBoost model performance using feature subsets derived from ADT, CVFE, and HFE selection methods. Correspondingly, Supplementary Figure S3 (**Supplementary Material**) displays the confusion matrices associated with each reduced feature configuration. Among the evaluated models, the subset generated through CVFE with parameters $c = 4$, $e = 10$, and $p = 0.6$ yielded superior predictive outcomes relative to the ADT- and HFE-based models. Notably, the highest-performing model correctly identified 124 out of 156 individuals diagnosed with opioid use disorder.

Features selected via the CVFE method

The highest-performing model in our analysis was the XGBoost classifier trained using the feature subset identified by the CVFE method, configured with parameters $c = 4$, $e = 10$, and $p = 0.6$. Figure 4 presents a SHAP summary bar chart, which ranks the importance of the selected features based on SHAP value analysis applied to this specific model. From an initial set of 32 candidate variables, the CVFE approach retained 29 features. Among them, the top 10 most influential features—ranked by their SHAP values—were: EDUD5PNRMIS (past-year misuse of pain relievers), PYUD5ALC (alcohol use disorder within the last year), CATAG6 (age group), ASTHMAEVR (history of asthma), SUTRTPY (receipt of substance use treatment in the past year), IREDUHIGHST2 (educational attainment), IRHHSIZ2 (number of household members), INCOME (total household income), IRMARIT (marital status), and NEWRACE2 (race/ethnicity).

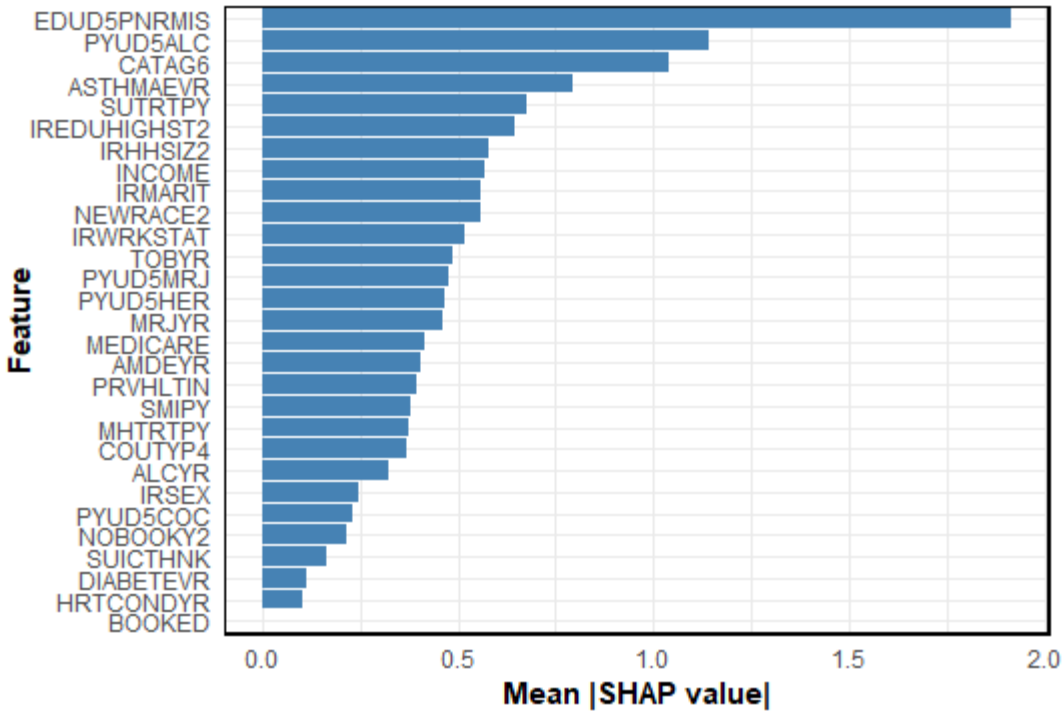
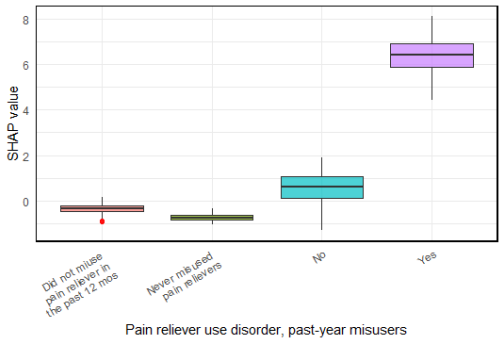


Figure 4 - The ranking of feature relevance as identified by the XGBoost framework. The horizontal axis represents the average magnitude of SHAP values for each variable, reflecting their typical influence on prediction outcomes. Features are arranged in decreasing order of predictive contribution.

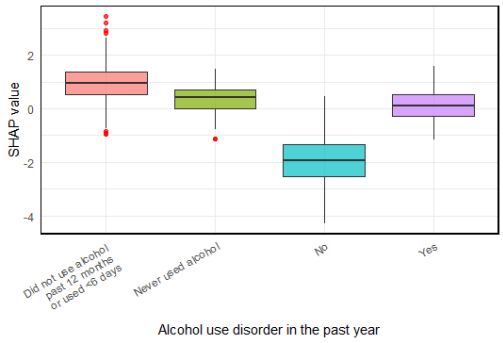
Analysis of feature contributions

Figure 5 presents boxplots illustrating the distribution of SHAP values for the categories within the top 10 categorical features identified in Figure 4. Since elevated SHAP values correspond to an increased opioid risk for patients, these plots allow direct interpretation of risk levels across feature categories. For instance, as shown in Figure 5(a), individuals who have never misused pain relievers exhibit a low opioid risk, whereas those classified as “Yes (misused)” show a substantially higher risk. Similarly, Figure 5(b) reveals that patients without an alcohol use disorder have a reduced opioid risk. The SHAP values for age categories, displayed in Figure 5(c), indicate that older individuals carry a higher opioid risk. Surprisingly, Figure 5(d) shows that a history of asthma is linked to a lower opioid risk. Recent substance use treatment correlates with an elevated opioid risk as demonstrated in Figure 5(e). Educational attainment at the college graduate level or above is associated with decreased risk (Figure 5(f)). Finally, factors such as larger household size, lower family income, widowed marital status, and being Non-Hispanic Native American or Alaska Native correspond to increased opioid risk, as shown in Figures 5(g) - 5(j).

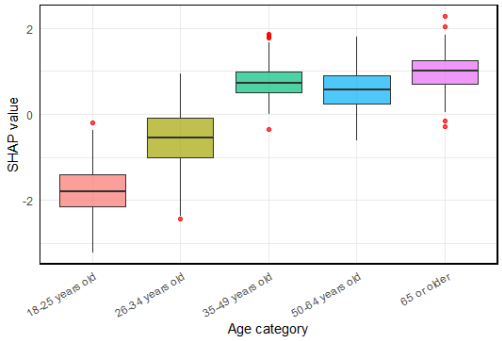
230



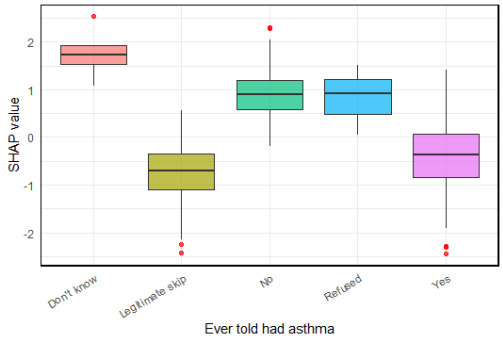
(a) EDUD5PNRMIS



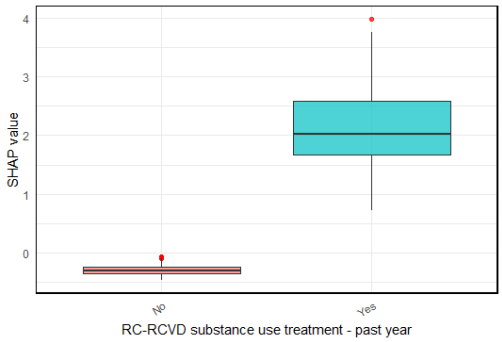
(b) PYUD5ALC



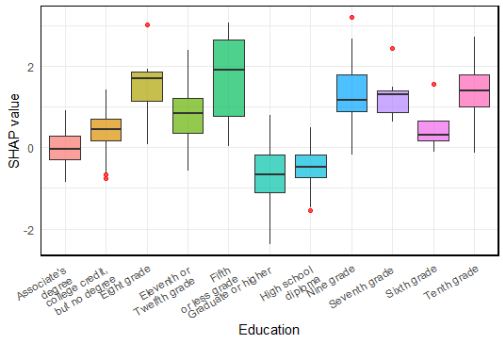
(c) CATAG6



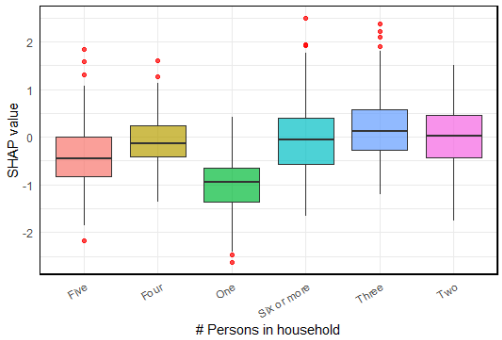
(d) ASTHMAEVR



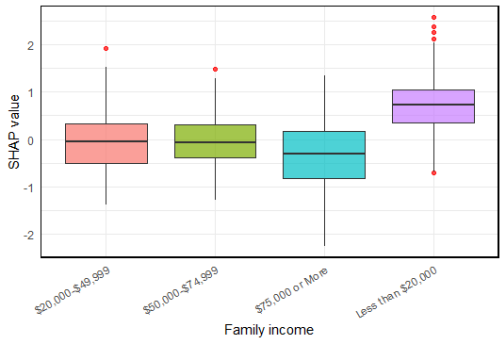
(e) SUTRTPY



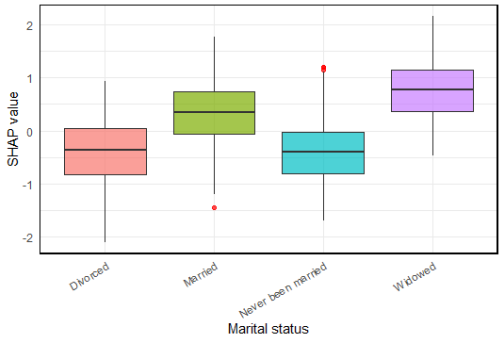
(f) IREDUHIGHST2



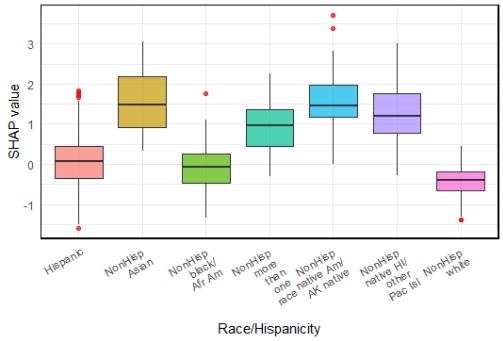
(g) IRHHSIZ2



(h) INCOME



(i) IRMERIT



(j) NEWRACE2

Figure 5 - SHAP value distribution for the top 10 categorical features selected by the model.

Discussion

Tailoring predictions to individual patients is critical for enhancing care in those vulnerable to opioid use disorder (OUD). To meet this goal, we constructed a machine learning prediction framework built on a selectively refined set of features. By removing unnecessary and redundant variables, we improved the model's clarity and interpretability. Utilizing this streamlined feature set, we assessed the performance of the XGBoost algorithm. Our approach effectively pinpointed important predictors linked to OUD, offering valuable insights for personalized risk evaluation, clinical decision-making, and the development of focused therapeutic strategies. Aligning with prior research, our model recognized age and ethnicity as among the most significant predictors. For example, a study utilizing a hierarchical attention transformer model also emphasized these features within the top ten predictors [3]. However, that analysis lacked a detailed hierarchy of feature importance, which limited its explanatory power. Similarly, Hasan et al. [22] employed several machine learning approaches—including logistic regression,

random forest, decision trees, and gradient boosting—on extensive healthcare claims data. Their highest-performing random forest model identified age, heroin poisoning, and drug abuse as key predictors. Our results are consistent with these observations; notably, younger age was linked to a decreased estimated risk for opioid use disorder, indicating a potential protective effect in specific scenarios. Additionally, Banks et al. [23] constructed a personalized prediction model for opioid-related outcomes, where age, substance use, and alcohol consumption ranked as the most influential features. The convergence of these findings supports the robustness and clinical relevance of our predictive model.

Importantly, our model highlighted pain reliever misuse as the foremost predictor for opioid use disorder. This finding is consistent with prior research underscoring the pivotal influence of prescription opioid misuse in the onset of OUD [35-37]. Recognizing such critical predictors is of great clinical importance, as it can guide early detection efforts, improve risk assessment, and shape targeted intervention plans. Additionally, the other key features identified by our model carry substantial potential to enhance individualized patient management and improve the efficient distribution of healthcare resources. While our results are encouraging, there are important limitations to consider. The models were developed using historical datasets, some of which contained missing data that were addressed through imputation methods [32]. This preprocessing step may have introduced bias, potentially affecting the calculated feature importance values. Furthermore, features with limited representation due to small sample sizes might have disproportionately influenced the findings. Additionally, the current model does not support time-to-event or longitudinal prediction tasks, limiting its application for monitoring the progression of opioid use disorder over time. Future work should focus on enhancing the model to accommodate these types of analyses.

In conclusion, we established and validated a machine learning framework to pinpoint critical determinants of OUD using data from the NSDUH survey. While the model shows encouraging predictive capability, there remains room for refinement, especially regarding improvements in AUC and other performance indicators. Future efforts will aim to integrate more detailed, high-dimensional datasets—such as multi-omics profiles and behavioral health metrics—to boost the model's accuracy. Additionally, we intend to create a practical, intuitive application to promote clinical implementation and maximize translational benefits. This research advances the development of personalized strategies for opioid misuse prevention and treatment, ultimately aiming to enhance patient care outcomes.

Supporting information

The supplementary material for this article is provided herewith the manuscript.

(DOCX)

Authors' contributions

SA: Conceptualization, data collection, formal analysis, software implementation, validation, visualization, and writing manuscript. JHM: Conceptualization, supervision, reviewing analyses, and editing the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Ethical approval statement

The study was conducted in accordance with the Declaration of Helsinki.

References

1. Garbin C, Marques N, Marques O. Machine learning for predicting opioid use disorder from healthcare data: a systematic review. Computer Methods and Programs in Biomedicine. 2023;236:107573.

2. Services USDoHaH. What is the U.S. Opioid Epidemic? 2017 [June 27, 2025]. Available from: https://www.hhs.gov/opioids/index.html.

3. Kashyap A, Callison-Burch C, Boland MR. A deep learning method to detect opioid prescription and opioid use disorder from electronic health records. International journal of medical informatics. 2023;171:104979.

4. Bohnert AS, Guy Jr GP, Losby JL. Opioid prescribing in the United States before and after the Centers for Disease Control and Prevention's 2016 opioid guideline. Annals of internal medicine. 2018;169(6):367-75.

5. Guy GP, Zhang K, Schieber LZ, Young R, Dowell D. County-level opioid prescribing in the United States, 2015 and 2017. JAMA internal medicine. 2019;179(4):574-6.

6. Higgins C, Smith BH, Matthews K. Incidence of iatrogenic opioid dependence or abuse in patients with pain who were exposed to opioid analgesic therapy: a systematic review and meta-analysis. British journal of anaesthesia. 2018;120(6):1335-44.

7. Keyes KM, Rutherford C, Hamilton A, Barocas JA, Gelberg KH, Mueller PP, et al. What is the prevalence of and trend in opioid use disorder in the United States from 2010 to 2019? Using multiplier approaches to estimate prevalence for an unknown population size. Drug and alcohol dependence reports. 2022;3:100052.

8. Kirson NY, Shei A, Rice JB, Enloe CJ, Bodnar K, Birnbaum HG, et al. The burden of undiagnosed opioid abuse among commercially insured individuals. Pain medicine. 2015;16(7):1325-32.

9. Katz NP, Sherburne S, Beach M, Rose RJ, Vielguth J, Bradley J, et al. Behavioral monitoring and urine toxicology testing in patients receiving long-term opioid therapy. Anesthesia & Analgesia. 2003;97(4):1097-102.

10. Heit HA, Gourlay DL. Urine drug testing in pain medicine. Journal of pain and symptom management. 2004;27(3):260-7.

11. Mallya A, Purnell AL, Svrakic DM, Lovell AM, Freedland KE, Gott BM, et al. Witnessed versus unwitnessed random urine tests in the treatment of opioid dependence. The American Journal on Addictions. 2013;22(2):175-7.

12. Chapman CR, Lipschitz DL, Angst MS, Chou R, Denisco RC, Donaldson GW, et al. Opioid pharmacotherapy for chronic non-cancer pain in the United States: a research guideline for developing an evidence-base. The Journal of Pain. 2010;11(9):807-29.

13. Jones T, Moore T. Preliminary data on a new opioid risk assessment measure: the Brief Risk Interview. Journal of Opioid Management. 2013;9(1):19-27.

14. Ciesielski T, Iyengar R, Bothra A, Tomala D, Cislo G, Gage BF. A tool to assess risk of de novo opioid abuse or dependence. The American journal of medicine. 2016;129(7):699-705. e4.

15. Dufour R, Mardekian J, Pasquale MK, Schaaf D, Andrews GA, Patel NC. Understanding predictors of opioid abuse: predictive model development and validation. Am J Pharm Benefits. 2014;6(5):208-16.

16. Edlund MJ, Martin BC, Fan M-Y, Devries A, Braden JB, Sullivan MD. Risks for opioid abuse and dependence among recipients of chronic opioid therapy: results from the TROUP study. Drug and alcohol dependence. 2010;112(1-2):90-8.

17. Ives TJ, Chelminski PR, Hammett-Stabler CA, Malone RM, Perhac JS, Potisek NM, et al. Predictors of opioid misuse in patients with chronic pain: a prospective cohort study. BMC health services research. 2006;6:1-10.

18. Rice JB, White AG, Birnbaum HG, Schiller M, Brown DA, Roland CL. A model to identify patients at risk for prescription opioid abuse, dependence, and misuse. Pain Medicine. 2012;13(9):1162-73.

19. Turk DC, Swanson KS, Gatchel RJ. Predicting opioid misuse by chronic pain patients: a systematic review and literature synthesis. The Clinical journal of pain. 2008;24(6):497-508.

20. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002;16:321-57.

21. Japkowicz N, editor The class imbalance problem: Significance and strategies. Proc of the Int'l Conf on artificial intelligence; 2000.

22. Hasan MM, Young GJ, Patel MR, Modestino AS, Sanchez LD, Noor-E-Alam M. A machine learning framework to predict the risk of opioid use disorder. Machine Learning with Applications. 2021;6:100144.

23. Banks TJ, Nguyen TD, Uhlmann JK, Nair SS, Scherrer JF. Predicting opioid use disorder before and after the opioid prescribing peak in the United States: A machine learning tool using electronic healthcare records. Health informatics journal. 2023;29(2):14604582231168826.

24. Warren D, Marashi A, Siddiqui A, Eijaz AA, Pradhan P, Lim D, et al. Using machine learning to study the effect of medication adherence in Opioid Use Disorder. PLoS One. 2022;17(12):e0278988.

25. Dong X, Deng J, Rashidian S, Abell-Hart K, Hou W, Rosenthal RN, et al. Identifying risk of opioid use disorder for patients taking opioid medications with deep learning. Journal of the American Medical Informatics Association. 2021;28(8):1683-93.

26. Ellis RJ, Wang Z, Genes N, Ma'ayan A. Predicting opioid dependence from electronic health records with machine learning. BioData mining. 2019;12:1-19.

27. Burgess-Hull AJ, Brooks C, Epstein DH, Gandhi D, Oviedo E. Using machine learning to predict treatment adherence in patients on medication for opioid use disorder. Journal of Addiction Medicine. 2023;17(1):28-34.

28. Akhter S, Miller JH. BPAGS: a web application for bacteriocin prediction via feature evaluation using alternating decision tree, genetic algorithm, and linear support vector classifier. Frontiers in Bioinformatics. 2024;3:1284705.

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

30. Yang M-R, Wu Y-W. A Cross-Validated Feature Selection (CVFS) approach for extracting the most parsimonious feature sets and discovering potential antimicrobial resistance (AMR) biomarkers. Computational and Structural Biotechnology Journal. 2023;21:769-79.

31. Misiorek P, Janowski S. Hypergraph-based importance assessment for binary classification data. Knowledge and Information Systems. 2023;65(4):1657-83.

346 32. Quality CfBHSa. 2023 National Survey on Drug Use and Health Public Use File Codebook, Substance Abuse and
347 Mental Health Services Administration 2024 [updated June 27, 2025]. Available from:
348 https://www.samhsa.gov/data/system/files/media-puf-file/NSDUH-2023-DS0001-info-codebook_v1.pdf.
349 33. Chen T, Guestrin C, editors. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd
350 international conference on knowledge discovery and data mining; 2016.
351 34. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global
352 understanding with explainable AI for trees. Nature machine intelligence. 2020;2(1):56-67.
353 35. Ferguson E, Lewis B, Teitelbaum S, Reisfield G, Robinson M, Boissoneault J. Longitudinal associations between pain
354 and substance use disorder treatment outcomes. Journal of substance abuse treatment. 2022;143:108892.
355 36. Wei Y-JJ, Schmidt S, Fillingim RB, Brock G, Schmidt S, Winterstein AG. Unrelieved pain and risk of opioid use
356 disorder or overdose in older adults prescribed opioids. Pain. 2022:10.1097.
357 37. Rudolph KE, Williams NT, Diaz I, Forrest S, Hoffman KL, Samples H, et al. Pain management treatments and opioid
358 use disorder risk in medicaid patients. American Journal of Preventive Medicine. 2024;67(6):878-86.

359