Identifying Predictors for Substance Consumption Pattern Using Machine Learning Techniques

Bijoy Chhetri, Lalit Mohan Goyal, and Mamta Mittal

1 Introduction

A drug is prescribed by the medical practitioner for remedies for some illness. Among the various classes of pharmaceutical drugs, the psychoactive drug contains a substance that influences mental functionalities. When these types of drugs are consumed above the prescribed dose, it is an abuse of a drug. Further, such drugs are an abuse-able psychoactive drug whose effects on the mental state are considerably high and gives a pleasant or interesting experience that some people choose to take them for a reason other than healing ailments. Such drugs of abuse have a chemical substance that influences part of mental functioning which is recreational, and enjoyable to cause preference to take multiple times leading to addiction. But, apart from substance consumption is enjoyable, the addiction intimate various kinds of diseases. These chemically or naturally available drugs are categorically called substances and as per Diagnostic and Statistical Manual of Mental Disorders (DSM) [1], the disorder caused by it is classified as a substance use disorder. Out of the various classes of drugs, depressants are those substances that include alcohol, amyl nitrite, benzodiazepines, tranquilizers, and opiates such as heroin and methadone. These classes of substances are used to reduce arousal and stimulations. The stimulants on the other hand consist of amphetamines, nicotine, cocaine powder, crack cocaine, and caffeine. These substances when taken speed up the movement of chemicals from the body to the brain making a person more energetic, alert, and too confident. The third broad category of drug classification is hallucinogens. These substances cause the person to hallucinate on sensations and images that appear to be real although they

B. Chhetri (✉) · L. M. Goyal
J C Bose University of Science and Technology, YMCA, Faridabad, India

M. Mittal
Delhi Skill and Entrepreneurship University, Dwarka,
New Delhi, India

B. Agarwal et al. (eds.), *Proceedings of International Conference on Intelligent Cyber-Physical Systems*, Algorithms for Intelligent Systems,
https://doi.org/10.1007/978-981-16-7136-4_9

are not in reality. Cannabis, ecstasy, ketamine, Lysergic acid Diethylamide (LSD), and mushrooms are the substances that fall under this category. Some other classes like Volatile Substance Abuse (VSA) which is used through glue-sniffing, inhalant abuse, and solvent abuse for the deliberate inhalation of volatile substances to achieve intoxication also exist. There are shreds of evidence on the use of these substances in high quantity for social or recreational purposes, including fun, stress suppressor, or to feel different. People of all age groups have been taking this for various reasons irrespective of gender or ethnicity. Several reasons are associated with early drug use including psychological, social, individual, environmental, and economic factors [2]. These aspects are likewise linked with several behaviors [3, 4] which are social and personal, but the consequences are highly problematic.

When data records of various international and national agencies are looked at, consumption of both illicit and licit substances has been so often that World Health Organization reports more deaths globally due to tobacco and opioids than because of hepatitis or AIDS [5]. National Survey in India [6] reported that of 14.6% alcoholics, 2.8% use cannabis and 1.14% use opioids either prescribed or illicitly accessed. The US reports the same figure as 43.4 million (17.9%) and the young are more influenced. The fatality rate is also very high due to opioid overdose, especially heroin consumption [7]. On a similar note, the UK too has a large number of cases, and various strategies have been put in place to see increasing patterns of drug use and its threats [8]. World Drug Report [9] highlights record-high use of prescription drugs and opioids along with enormous users of common drugs like alcohol and tobacco globally. To address this rate of consumption, several case-based researches are done and drug-dependent disorderly behaviors are studied and tackled. However, in the heterogeneity of substance consumption patterns [10], opiates for stimulants have some differences in the behavior of consumers. The cause of neurobehavior and even their consumption patterns have been very distinct [11, 12] in personality involvement. With each class of substance, the pleasure circuit is caused to move very differently in different individuals [13]. The substance-specific uses are fundamentally agreed [14] on the choice, and preferences of individuals like alcohol seekers will have a different profile than that of LSD. This apart, the initial onset of one of them could lead to the addiction of another and the seeking tendencies cause one to ON and the other one to OFF [15]. For example, substances opiate and stimulant addictions have opposite effects such that the former activates inhibitory and sedative circuit of the brain whereas the latter arouses and creates lots of excitement, thus, personality trait plays a major role in the selection. The initial onset of marijuana at a young age could lead to another substance intake [16] which is also dependent on how a person perceives the substance. Thus, the identification of layers of influencing factors that could justify variable consumption patterns is required. This phenomenon of seeking one substance after another has prevailing psychological thoughts which have been somehow disseminated by the Machine Learning (ML) model applied on the various clinical or non-clinical datasets revealing some of the remarkable facts on substance consumption as presented in Table 1.

Hence, knowing the fact that substance intake varies from person to person depending on their personality traits and consumption patterns, this article presents

Table 1 Various existing studies performed machine learning approach

Authors	Study and sample	Method used	Predictors	Outcome	Remark
[17]	Cohort/adolescents from Canada and Australia	Supervised learning	Personality, cognition, and alcohol attitude variables	Prevalence of alcohol drinking	AUC: 0.86
[18]	Cross-sectional/cocaine user and control	Supervised learning and regression	Impulsivity traits	Prevalence of cocaine and high impulsivity	AUC: 0.91
[13]	Cross-sectional/heroin user and Control	Supervised learning and regression with elastic net	Personality, impulsivity	Heroin (H) and Amphetamine (A) have distinct seeking and personality traits	AUC: 0.86(H) AUC: 0.71(A)
[19]	Cohort/smokers	CART supervised learning	Clinical, executive functions	Prevalence of cigarette and delay of execution	Accuracy: 0.80
[20]	Cohort/twitter users	Ensemble supervised	Tense level, arousal level, restlessness	Prevalence of alcohol more prominent	Accuracy: 0.89
[21]	Trail/cocaine dependent	Reinforcement learning	Deprivation	Sensitive to drug	R: 0.46
[16]	Cross-sectional/survey data	Supervised learning	Income, employment, early onset drug	Marijuana seeks to be driving the other substances	AUC: 0.89
[22]	Cross-sectional/drug user and non-user	Artificial neural network	Personality, demography	Alcohol user	Accuracy: 98.7

Label: AUC: Area under the Receivers Operating Characteristics

the use of the ML model to identify the baseline markers to understand the nature of substance intake and their behavioral association with an individual. Various reasons are likely to predict the initial onset of substance till it becomes an abuse. It is also the likelihood of one substance to induce another substance to cause poly-drug influence. The socio-economic parameters, demography, and location may also influence the cause. Although each of these aspects has a considerable amount of contribution to increase the threat of substance intake, neither one is the only origin of conclusion. Therefore, this research is aimed at analyzing the complex relationship among these dynamics and identifying patterns of consumption, their key predictor if any.

The relationships obtained are the measure of substance prevalence and consumption patterns which are the potential markers of risk assessment. Because of the fact that substance consumption may lead to psychological disorders in the later stage, therefore, these predictors along with potential clinical observation can help health workers and officials in planning out an appropriate intervention and precautionary approach.

The article is divided into various sections where Sect. 2 has a methodology explaining the approach used to carry out this research. Section 3 is a result section that highlights key findings obtained while carrying out this research. The primary findings are discussed in Sect. 4 which is followed by a conclusion and references.

2 Proposed Methodology

A systematic diagram that explains the methodology is presented in Fig. 1. The dataset acquired from the repository is subjected to rigorous pre-processing of data that are required. The first phase is data transformation where the categorical data is transformed into numerical values to facilitate the requirements of ML models. Data curation is also performed so as to obtain the required field for the classification problem. The subclass where similar substances are included is also prepared for association analysis. The insignificant columns are also removed.

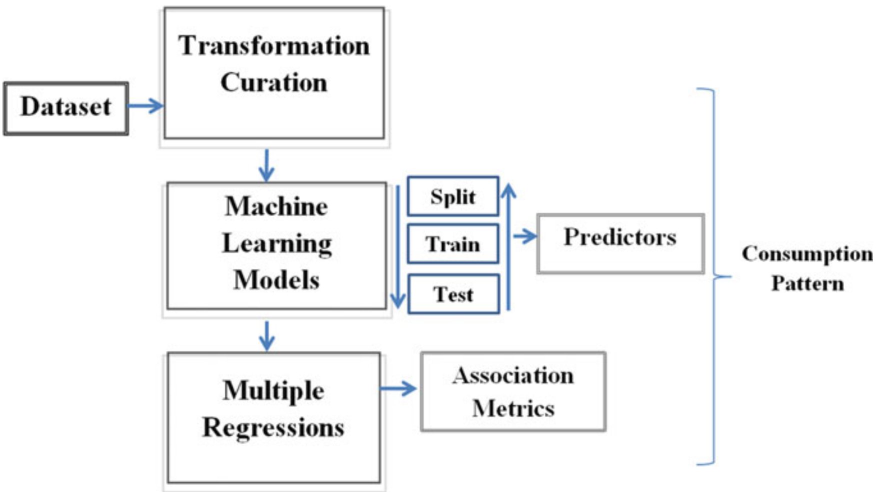


Fig. 1 Systematic diagram of proposed methodology

2.1 Dataset

The research is based on existing data [23] which is available in the UCI repository. The dataset contains 1185 $\times$ 31 values derived from the cross-sectional survey on personality trait parameters and substance consumption information of 18 licit and illicit substances. The categorical values are quantified. The personality traits are measured with 5 parameters to measure personality in terms of Neuroticism (N), Extraversion (E), Openness (O) to experience, Agreeableness (A), and Conscientiousness (C) which is popularly known as NEO-5 personality traits. It also contains a measure of Impulsivity (BIS-11) and Sensation Seeking measure (ImpSS) that is the standard battery of assessment used by psychologists to measure disinhibition, distraction, consciousness, and insufficiencies in decision-making. The dataset also contains little demographic information, i.e. level of education, age, gender, country of residence, and ethnicity. In addition, the dataset contains the response from the participants about their intake of 18 licit and illicit classes of substances. Alcohol, amphetamines, amyl nitrite, benzodiazepine, cannabis, chocolate, cocaine, caffeine, crack, ecstasy, heroin, ketamine, legal highs, LSD, methadone, mushrooms, nicotine, and VSA are included. A dummy drug (Semeron) is also used to rule out the over claim from the respondents. The information about consumption of each of these substances is collected through a questionnaire being asked to select one of the answers from the option of never used, used over a decade ago, or in the last decade, year, month, week, or a day suggesting the use patterns of the individual who has responded with the survey questionnaire. The dataset is curated and used in this study to suit the objective of the research. Out of 18 substances, the effect of VSA is not considered as it has insignificant statistics as compared to other substance use. The three classes of substances, namely depressants, stimulants, and hallucinogens, are made based on the standard composition of the substance. The outcome of multiple classifiers, multivariate analysis, and correlation analysis is presented with evaluation and comparisons with other literature. The paper takes a perspective to seek the likelihood of substance pattern due to one substance following the other in continuation. The outcome variables in terms of predictors signifying the likelihood are sought from the dataset.

2.2 Classifiers

The ML-based classifiers are employed to classify each of the respondents into substance user and non-user categories. Table 2 has a detailed list of the predictor variables used for classification on the ML models. The important features are extracted from the ML model to further understand their impact through multivariate analysis. The predictors' variable values are split into training and test cases. The training set is used for training the classifiers and test sets are kept for validation. The usual

Table 2 Summary of the predictor variables

Outcome: user/non-user			
Predictor variables	N	Mean	Levels
Gender Male Female	943 942	NA	2
Age 18–24 25–34 35–44 45–54 55–64 65+	643 481 356 294 93 18	42.44	6
Educational background Left school before 16 yrs Left school at 16 yrs Left school at 17 yrs Left school at 18 yrs Studied but No certificate Diploma Degree Masters and above	28 99 30 100 506 270 480 372	NA	8
Personality factors			
Neuroticism (N score)	1875	23.92	6
Extraversion (E score)		27.52	
Openness (O score)		33.64	
Agreeableness (A score)		30.87	
Conscientiousness		29.44	
Impulsivity and sensation			
BIS	1875		4
ImpSS	1875		
Potential drivers within the group of substances	NA		

percentage of training and test sets is 80–20 of all the responded values. The prediction showed low accuracy in the initial stage; it has a high number of non-users of any of the substances making it an imbalanced dataset. The authors, however, have used downsampling to balance the dataset, and classifications are performed. The datasets are used to train and test DT, RF, and K-NN classification algorithms. A Multivariate Regression (MR) analysis is also performed using Minitab to evaluate the association of various classes of substances. Python packages under the Scikit Learn library are used to build the classifiers model and perform classification. For all of the three classifiers, classification algorithms are run up to 100 times, and the average of the performance is measured using a 95% confidence level on the statistically significant parameters.

2.3 Evaluation of Performance

The results predicted by each of the classifiers are compared against the observed outcome to obtain the ground truth. The analysis is done through a confusion matrix which refers to relationships among the predicted and the observed true values. Here, true positive refers to the prediction of the substance user who is an actual user of the substance. False positive informs about substance abuse which is absent in the true case. Similarly, false negative is the predicted result which is incorrectly predicted as a non-user of substance which is a negative prediction as is false positive. However, certain cases of false predictions are considered and sensitivity percentage is computed. In addition, specificity which is a percentage of true negative cases is also taken as a measure to evaluate the classifiers. To visualize the performance, an Area under the Receivers Operating Characteristics curve is also assessed using various threshold values. In general, a range above 80% is considered as fair and good to predict the outcome, whereas 90% and above is targeted throughout this research. MR is evaluated using squared error measurement. The set of predictor variables, while they are identified, is evaluated statistically with a significance value of $p < 0.001$.

3 Result

This section presents the research findings in three modes; first, the classification algorithm is run to find the predictor variable among all the variables along with their importance. The ML classifiers are not very accurate in classifying the individual into their respective class. Therefore, an additional method of classifying the individual into each substance class using the ensemble method is also carried out and presented in a separate section. The third section has unique findings as to what extent the substances themselves drive the other substance consumption. To have a deeper understanding of every outcome, MR is performed with a statistically significant value of $p < 0.001$. The outcomes of classifiers as key predictor variables along with the importance of each variable are also presented along with the performance of ML models.

3.1 Classification Based on Three Classes of Substances

Based on the classification of substances [24], a total of 17 licit and illicit substances are categorized into three groups, namely Depressants, Stimulants, and Hallucinogens. The response value recorded under each category are taken all together to find correlation among the individual substance within the group like amphetamines, nicotine, cocaine powder, crack cocaine, caffeine, and chocolate falls under the same class of stimulants. The coefficient enabled the research to deepen down to find the most correlated preferences among the users. Further, a new target class is computed based on the preference of either HIGH or LOW. HIGH here signifies the respondents who have been taking substances for a month, but LOW is to signify that person is in abstinence for at least a month. For example, in a class of depressants, the three substances, namely benzodiazepine, alcohol, and meth are positively correlated with significance value ($p < 0.001$); thus a class of response variable (Depressant) is determined with the binary classification of HIGH or LOW as their indicators. The classifiers are trained and tested with the predictor variables considered, and the outcome evaluation metrics are presented in Table 3. Due to an imbalanced set of data in the stimulant class, where there is very minimal user classified as HIGH, the classifier was overfitted to predict with 99% of accuracy, which is not achieved in the other cases. Finally, the imbalanced set is managed using downsampling of the dataset to accord a result as shown.

Out of all the classifiers, RF seems to have scored high accuracy to predict the class of substances. The predictor variables as rated by RF have impulsivity factor on the top and relevance of education being very insignificant as far as substance intake likelihood is a concern. The predictor variables identified by the classifier(s) along with their percentage of importance are also presented in Table 4.

Table 3 Result of classifiers

Class of substance	Classifier	Precision	Recall	F1 score
Hallucinogens	K-NN	0.85	0.73	0.79
	RF	0.88	0.78	**0.83**
	DT	0.96	0.65	0.78
Stimulants	K-NN	0.65	0.81	0.72
	RF	0.74	0.81	**0.77**
	DT	0.85	0.65	0.74
Depressant	K-NN	0.87	0.67	0.76
	RF	0.75	0.85	**0.80**
	DT	0.89	0.65	0.75

Table 4 Relative importance of variables contributing to substance consumption

Predictor variables	Relative importance (%)
Impulsivity (ImpSS)	17
Impulsivity (BIS)	17
Openness	15
Extroversion	15
Neuroticism	12
Agreeable	5
Conscientiousness	4

MR analysis is performed on the target variables after removing less important predictor variables and considering only personality trait predictors (NEO) and impulsivity predictors (BIS and ImpSS) to test the regressed variables. The outcome tests the differences and associations to conclude an effect to the response variable or not and also validates the class of substances with a statistically significant p-value < 0.001 to support the assumption made at the beginning of this study.

3.2 *Classification Based on Individual Substance*

While the classification is done according to the class of substances, the viability of testing one's use of drugs cannot be mapped to the individual substances apart from the classification being made by the classifiers that has moderate accuracy; thus, it is difficult to find a final model of prediction. Therefore, to perform the classification synergistically, an ensemble method is adopted to combine the prediction of three ML models and lessen inconsistency in prediction as well as generality errors.

The training data is varied in every model using a random split of training and test data as well. The models are validated with a tenfold cross-validation technique to produce the performance as presented in Fig. 2 with performance metrics in Table 5.

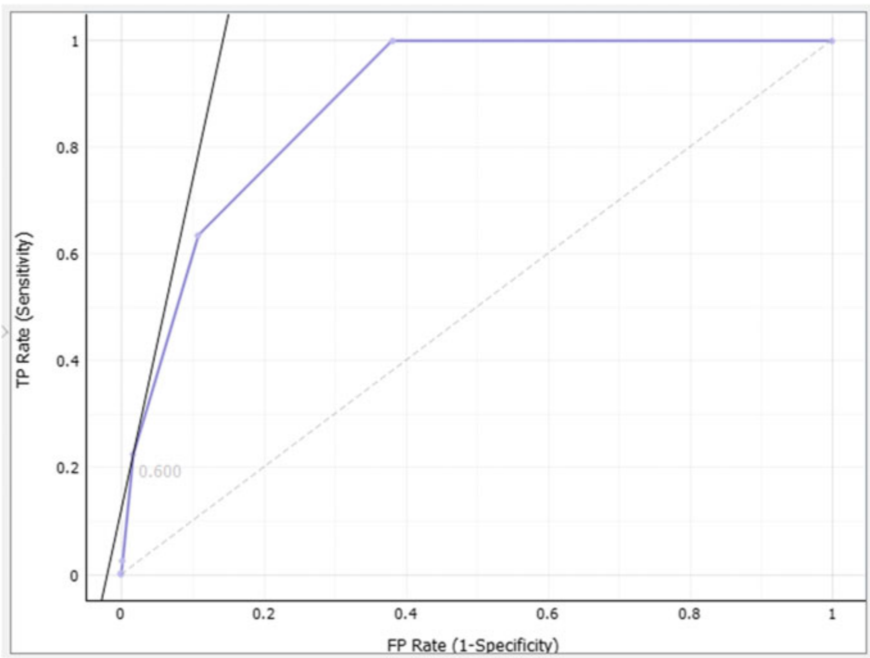


Fig. 2 Ensemble model AUC curve

Table 5 Performance of ensemble model

AUC	CA	F1	Precision	Recall
0.973	0.936	0.932	0.939	0.936

3.3 Potential Drivers for Substance Consumption

Various researches have concluded the fact that one's initial onset of driver substance could lead to their future increase in consumption thereby causing more harm. Therefore, to understand the consumption pattern and such driver substances, an analysis has been performed based on the objective to find a substance that could be a potential marker inherent within the group of substances driving an urge. It is found that both the statement of early onset and carrying further are significantly true with $p < 0.001$, and it was also surprisingly found that the drugs that belong to the same class have minimal correlation and association ($r << -0.01$) when they are analyzed within the group. However, the substance that belongs to a different class tends to have a higher association ($r > 0.21$). A MANOVA is performed with three classes of substances and the substance which is present in another class of substances. It is found that heroin is the leading driver in the group of depressants. It would lead to the consumption of other substances belonging to another group of substances like stimulants. Table 6 has a significant list of substances that falls under the potential driver to another class of substances with their significance value.

There are quite a few substances that are closely related to the hallucinogen class like cocaine, and amphetamines belonging to stimulants. Stimulants though have indicated a very limited association with depressants except with heroin. But, when analyzed along with hallucinogen class, cannabis and LSD has a high chance of leading to consumption of stimulants. Heroin also has been an active substance to have a strong association with hallucinogens. Methadone, an opiate medicine, is also of depressant class, which has a high influence on hallucinogens. Among the stimulants, crack cocaine has a high likelihood of driving it toward hallucinogens. Thus, there are some driver substances that are causing consumption patterns to

Table 6 Association of various driver substances in the pattern of consumption

Class	Name of substance	T statistics	F	p
Hallucinogens	Crack	0.97096	28.11	0.001
	Amphetamines	0.99295	6.64	0.001
	Heroin	0.96.73	38.54	0.001
	Methadone Opiate	0.9967	8.87	0.001
Stimulant	Cannabis	0.98818	11.23	0.001
	LSD	0.98303	16.97	0.001
	Heroin	0.9785	20.88	0.001
Depressant	Crack	0.98564	13.68	0.001

follow with the personality traits, as visually presented in Fig. 3a–c, and along with the key features of traits being the key predictors.

The regression model for various classes of substances highlights that the consumption pattern suggests sufficient association with the personality trait variables. The R2 measure of the regression model shows quite a high value with a significance measure of $p < 0.001$. As shown in Fig. 3a, it is a model building plot of the hallucinogen class and the predictor variables showing the impact of adding a new variable. The variation in the form of regression coefficient function is obtained, which is fitted on the curve to plot residual versus the fitted values. The final regression report suggests impulsivity being highly associated with the class of substances. Similarly, in Fig. 3b, stimulants are plotted against the predictors. There are some statistically insignificant values encountered in the analysis ($p > 0.001$), aside from the variable impulsivity which is evident in the variation bar plots, where R2 values are significant to fit the information to a model with low variation.

When the regression is plotted against the depressant class of the substance, each of the variables has a significant value ($p < 0.001$); thus the variation is significant as shown in Fig. 3c. The value of impulsivity is high while it is regressed with all the predictor variables, whereas whenever new information is to be predicted, the score of E contributes to the maximum in order to fit the model.

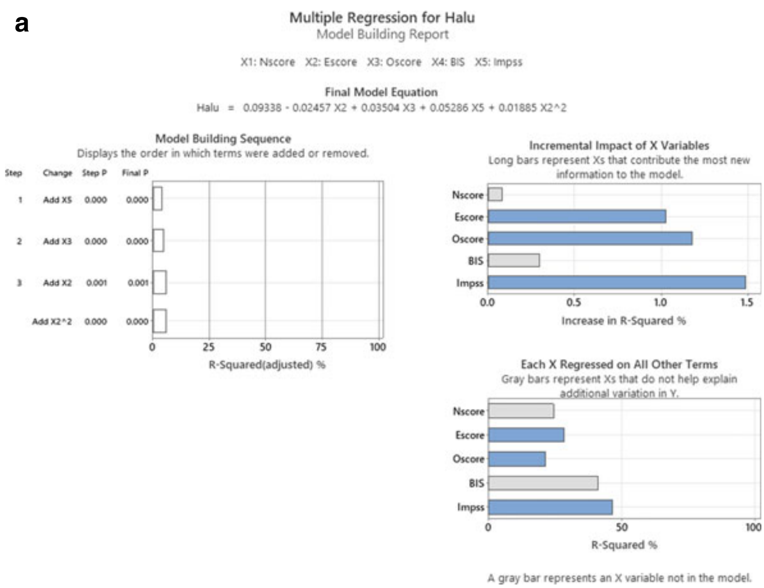
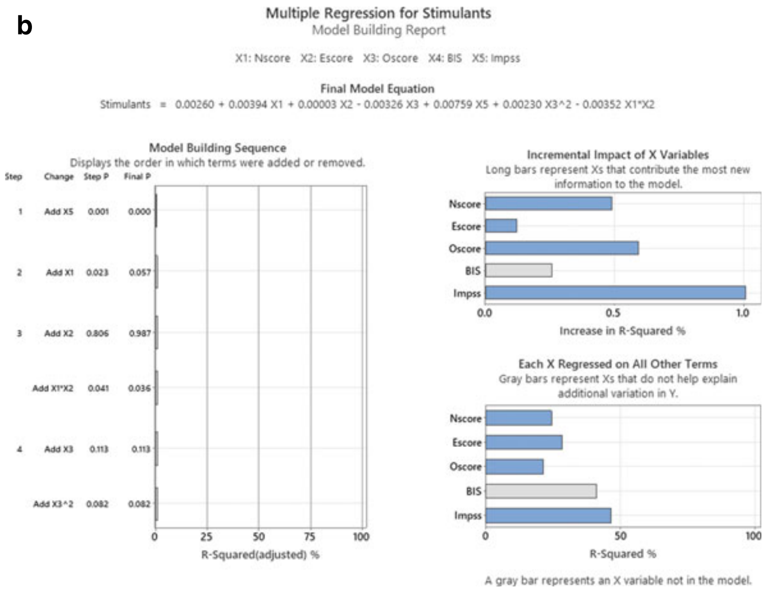


Fig. 3 **a** Multiple regression of halucinagens class. **b** Multiple regression of stimulant class. **c** Multiple regression of depressant class



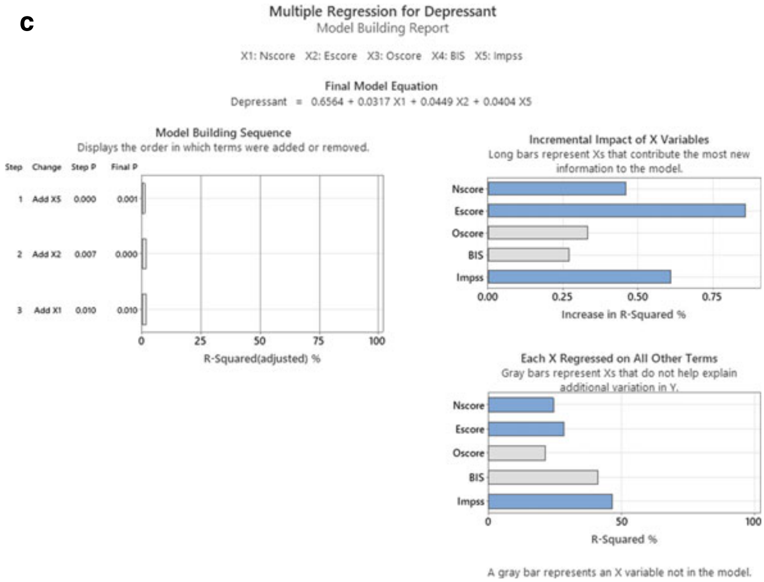


Fig. 3 (continued)

4 Discussion

This study presents a case on analysis of an existing dataset of substance use and its association with personality traits along with demographic variables. Findings show a positive association between consumption patterns and their predictor variables that influence an individual to consume substances. Initially, classification of substance usage as User/Non-User is performed using ML methods. Further, the substances are studied in association with personality traits and impulsivity along with some covariates as predictors to analyze multivariate kinds of patterns in their consumption, inter- and intra-class relationship as well as within individual substances. The existing literature [25, 26] has concluded the profile of personality traits to each Pleiades of substances. Similarly, a neural network applied to the same dataset has concluded results with the accurate classification of alcohol [22] consumption (AUC: 98%). This work has been able address major gaps that remained unaddressed, specifically (a) latent predictor variables which are addressed as drivers throughout the research are analyzed concurrently with multiple dimensions of personality traits within the substance-dependent individuals to gain insight into association among the various class of substance; (b) it identified a common predictor variable from multivariate patterns of consumption that classified them into user and non-user in new samples; and (c) it provided common patterns of evidence that an individual with certain endophenotypes correlates to substance intake.

The dataset contained a poly-drug assessment with personality traits; the only common predictor variable in all the categories of substance is impulsivity as in similar research [27]. It is strongly related to one psychological desire to perform any task without thinking about its result. When impulsivity goes too far, it strongly relates to aggressive behavior, restlessness, and most importantly gets easily distracted by influences, thus opening a wide scope of indulging in substance consumption [18]. Apart from this result, it also suggests that lifestyle personality factors like Neuroticism, Extraversion, and Openness has an equal stake in the development of substance consumption habits as well as their impact may lead to other comorbidities like Hypersensitivity or Conduct disorder, etc. [13]. The study also produces far-ranging evidence of one drug that has the chance of leading to another. ML models like K-NN, RF, and DT can understand substance consumption patterns, and relatively important variables are obtained. The consumption patterns show some of the inherent driving factors that are driving users toward the other class of substance. For example, someone is keen on prescribed opiates for some time; the desire of getting into illicit substances like LSD and Cannabis cannot be overruled and various profiles have already been identified in case of substance use disorder [28]. There is a prevalence of polysubstance consumption and such patterns even lead to mental illnesses like depression and anxiety [21, 29]. In contrast with other previous findings, the education and gender variables are not significant to drive people into substance consumption.

5 Conclusion

The ML approach has classified the substance user with acceptable accuracy and their relative predictors identified from the personality traits and demographical details to discover a consumption pattern. The study also produces some evidence on the inherent property of one or more substances to drive into the use of other substances. The dominant predictor variable is impulsivity along with the significant importance of personality trait variables. Depressant substances such as alcohol and its seeking sensation may not lead to heroin consumption but can have a strong association with hallucinogens such as cannabis. It can be established that variables like NEO that define personality trait evidence are indispensable for the estimated definition of substance consumption. The results are quite eye-opening. The truth of consumption patterns as revealed through data of substance intake seeks utmost attention. The multi way of influences in consumption pattern can be reduced while tapping the key predictors at the early stage using ML models and appropriate intervention is used to reduce the risks. Though the highest care is taken to carry out this research study, a few limitations exist: facilitation of individual consumption pattern with the development of the singular prevention method can be devised. The imbalance dataset can be corrected and can enrich the precision in the likelihood of time of consumption, which can also be taken up in future directions. In summary, the multiple variables are responsible for predicting markers that classify them to some substance consumption pattern or dependence and also that there are inherent capabilities of one substance to switch ON other substance intakes.

References

1. Pomeroy, E.C., Anderson, K.: The DSM-5 has arrived (2013)
2. Finn, K.: Why marijuana will not fix the opioid epidemic. Mo. Med. **115**, 191–193 (2018)
3. Zou, Z., Wang, H., Uquillas, F.D.O., Wang, X., Ding, J., Chen, H.: Definition of substance and non-substance addiction. Substance and Non-substance Addiction 21–41
4. Poria S., Gelbukh A., Agarwal B., Cambria E., Howard N.: Common sense knowledge based personality recognition from text. In: Castro F., Gelbukh A., González M. (eds) Advances in Soft Computing and Its Applications, MICAI 2013. Lecture Notes in Computer Science, vol. 8266. Springer, Berlin (2013). https://doi.org/10.1007/978-3-642-45111-9_42
5. Peacock, A., Leung, J., Larney, S., Colledge, S., Hickman, M., Rehm, J., Ali, R.: Global statistics on alcohol, tobacco and illicit drug use: 2017 status report. Addiction **113**(10), 1905–1926 (2017)
6. Ambekar, A., Agrawal, A., Rao, R., Mishra, A.K., Khandelwal, S.K., Chadda, R.K.: Magnitude of substance use in India. New Delhi: Ministry of Social Justice and Empowerment, Government of India (2019)
7. Abuse, S.: Mental health services administration. In: Key Substance Use and Mental Health Indicators in the United States: Results From the 2016 National Survey on Drug Use and Health (HHS Publication No. SMA 17–5044, NSDUH Series H-52). Rockville, MD: Center for Behavioral Health Statistics and Quality. Substance Abuse and Mental Health Services Administration (2017)
8. Drug misuse and dependence. www.assets.publishing.service.gov.uk (2017)

9. United Nations Office on Drugs and Crime (UNODC). World drug report 2017. Vienna: UNODC (2017)
10. Badiani, A., Belin, D., Epstein, D., Calu, D., Shaham, Y.: Opiate versus psychostimulant addiction: the differences do matter. Nat. Rev. Neurosci. **12**(11), 685–700 (2011)
11. George, O., Koob, G.F.: Individual differences in prefrontal cortex function and the transition from drug use to drug dependence. Neurosci. Biobehav. Rev. **35** (2010)
12. Vassileva, J., Paxton, J., Moeller, F.G., Wilson, M.J., Bozgunov, K., Martin, E.M., Gonzalez, R., Vasilev, G.: Heroin and amphetamine users display opposite relationships between trait and neurobehavioral dimensions of impulsivity. Addict. Behav. **39**, 652–659 (2014)
13. Ahn, W.Y., Vassileva, J.: Machine-learning identifies substance-specific behavioral markers for opiate and stimulant dependence. Drug Alcohol Depend. **161**, 247–257 (2016)
14. Clark, S.L., Gillespie, N.A., Adkins, D.E., Kendler, K.S., Neale, M.C.: Psychometric modeling of abuse and dependence symptoms across six illicit substances indicates novel dimensions of misuse. Addict. Behav. **53**, 132–140 (2016)
15. Peters, J., Pattij, T., De Vries, T.J.: Targeting cocaine versus heroin memories: divergent roles within ventromedial prefrontal cortex. Trends Pharmacol. Sci. **34**(12), 689–695 (2013)
16. Wadekar, A.S.: Understanding opioid use disorder (OUD) using tree-based classifiers. Drug Alcohol Depend. **208**, 107839 (2020)
17. Afzali, M.H., Sunderland, M., Stewart, S., Masse, B., Seguin, J., Newton, N., Conrod, P.: Machine-learning prediction of adolescent alcohol use: a cross-study, cross-cultural validation. Addiction **114**(4), 662–671 (2019)
18. Ahn, W.Y., Ramesh, D., Moeller, F.G., Vassileva, J.: Utility of machine-learning approaches to identify behavioral markers for substance use disorders: impulsivity dimensions as predictors of current cocaine dependence. Front. Psych. **7**, 34 (2016)
19. Coughlin, L.N., Tegge, A.N., Sheffer, C.E., Bickel, W.K.: A machine-learning approach to predicting smoking cessation treatment outcomes. Nicotine Tob. Res. **22**(3), 415–422 (2020)
20. Liu, J., Weitzman, E.R., Chunara, R.: Assessing behavior stage progression from social media data. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. IEEE, pp. 1320–1333 (2017)
21. Wang, J.M., Zhu, L., Brown, V.M., De La Garza II, R., Newton, T., King-Casas, B., Chiu, P.H.: In cocaine dependence, neural prediction errors during loss avoidance are increased with cocaine deprivation and predict drug use. Biol. Psychiatry Cogn. Neurosci. Neuroimaging **4**(3), 291–299 (2019).
22. Kumari, D., Kilam, S., Nath, P., Swetapadma, A.: Prediction of alcohol abused individuals using artificial neural network. Int. J. Inf. Technol. **10**(2), 233–237 (2018)
23. Machine learning repository. https://archive.ics.uci.edu/ml/datasets/Drug+consumption
24. Drug wheel: https://adf.org.au/
25. Fehrman, E., Egan, V., Gorban, A. N., Levesley, J., Mirkes, E.M., Muhammad, A.K.: Personality traits and drug consumption. Springer International Publishing (2019)
26. Fehrman, E., Muhammad, A.K., Mirkes, E.M., Egan, V., Gorban, A.N.: The five factor model of personality and evaluation of drug consumption risk. In: Data Science. Springer, Cham, pp. 231–242 (2017)
27. Argyriou, E., Um, M., Carron, C., Cyders, M.A.: Age and impulsive behavior in drug addiction: a review of past research and future directions. Pharmacol. Biochem. Behav. **164**, 106–117 (2018)
28. Zhang, X., Huang, R., Li, P., Ren, Y., Gao, J., Mueller, J.F., Thai, P.K.: Temporal profile of illicit drug consumption in Guangzhou, China monitored by wastewater-based epidemiology. Environ. Sci. Pollut. Res. **26**(23), 23593–23602 (2019)
29. Chhetri, B., Goyal, L.M., Mittal, M., Gurung, S.: Consumption of licit and illicit substances leading to mental illness: a prevalence study. EAI Endorsed Trans. Pervasive Health Technol. **6**, 21 (2020)