

Received November 4, 2020, accepted November 20, 2020, date of publication December 2, 2020, date of current version December 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3041895

# Prediction of Cocaine Inpatient Treatment Success Using Machine Learning on High-Dimensional Heterogeneous Data

JOSE TAPIA-GALISTEO<sup>1,2</sup>, JOSE M. INIESTA<sup>1</sup>, CARMEN PÉREZ-GANDÍA<sup>1,2</sup>,  
GEMA GARCÍA-SÁEZ<sup>1,2</sup>, DIEGO URGELÉS PUÉRTOLAS<sup>3</sup>, FRANCISCO J. IZQUIERDO<sup>4</sup>,  
AND M. ELENA HERNANDO<sup>1,2</sup>, (Senior Member, IEEE)

<sup>1</sup>Centro de Tecnología Biomédica, Bioengineering and Telemedicine Group, ETSI de Telecomunicación, Universidad Politécnica de Madrid, 28040 Madrid, Spain

<sup>2</sup>Networking Research Center on Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), 28029 Madrid, Spain

<sup>3</sup>Research and Development, Clínica Nuestra Señora de la Paz, OH San Juan de Dios, 28043 Madrid, Spain

<sup>4</sup>IBM Analytics, 28002 Madrid, Spain

Corresponding author: Jose Tapia-Galisteo (jose.tapia.galisteo@upm.es)

This work was supported in part by the Autonomous Community of Madrid through the Pre-Doctoral Grant BOCM/286/30112018, and in part by the European Social Fund.

**ABSTRACT** The high prevalence of drug addiction is a major health challenge that pressures healthcare systems to respond with cost-effective treatments. To improve the treatment success of drug-dependent patients, it is necessary to identify the main associated risk factors for dropping out of treatment. Previous research shows disparate results due to the wide variety of approaches employed, the different and/or poorly defined metrics used, and the different target populations under study. This article presents the design and selection of a predictive model to estimate success of inpatient cocaine treatment based on a high-dimensional heterogeneous set of characteristics, with the aim of learning new associations between independent characteristics. We evaluated different feature selection techniques and machine learning algorithms to design the best predictive model in terms of accuracy, area under the receiver operating characteristic curve, recall, specificity, F1-score, and Matthews correlation coefficient. Random Forest was the top-performing model with a characteristic set consisting of 11 features selected with a wrapper evaluator and the Best First algorithm, achieving 82% accuracy, 0.81 of area under the receiver operating characteristic curve, 0.96 of recall, 0.47 of specificity, 0.89 of F1-measure and 0.53 of Matthews correlation coefficient. The predictive model's performance was enhanced by combining multiple dimensions with variables referring to previous treatments, mental exploration, cognitive functioning, personality, consumption habits, and pharmacological treatment. We have refined the use of machine learning techniques to predict drug addiction treatment success, which could represent a new step in treatment management especially when included in clinical decision support systems.

**INDEX TERMS** Cocaine addiction, feature selection, high-dimensional heterogeneous data, machine learning, predictive model, treatment success.

## NOMENCLATURE

ML	Machine learning.	ROC	Receiver operating characteristic curve.
SVM	Support Vector Machine.	AUC	Area under the receiver operating characteristic curve.
RF	Random Forest.	TP	True positives.
LR	Logistic regression.	TN	True negatives.
MLP	Multilayer perceptron.	FP	False positives.
CFS	Correlation-based Feature Selection.	FN	False negatives.
IWSS	Incremental Wrapper Subset Selection.	MCC	Matthews correlation coefficient.
		CS <sub>WBF</sub>	Characteristic set obtained by the wrapper evaluator with the Best First algorithm.

The associate editor coordinating the review of this manuscript and approving it for publication was F. K. Wang.

- CS<sub>WG</sub> Characteristic set obtained by the wrapper evaluator with the Genetic Search algorithm.
- CS<sub>IWSS</sub> Characteristic set obtained by the IWSS evaluator.
- CS<sub>CFBF</sub> Characteristic set obtained by the Correlation-based Feature Selection with the Best First algorithm.
- CS<sub>CFG</sub> Characteristic set obtained by the Correlation-based Feature Selection with the Genetic Search algorithm.

## I. INTRODUCTION

The high prevalence of cocaine addiction is a major health challenge that leads to pressure on healthcare systems to respond with cost-effective treatments. The latest European Drug Report by the European Monitoring Centre for Drugs and Drug Addiction [1] shows that, after cannabis, cocaine is the second most widely used illegal drug in Europe, being consumed by 3.9 million adults (15-64 years old) and 2.6 million young adults (15-34 years old) in 2018. Spain is the sixth largest cocaine consuming country in the European Union, with a prevalence of 2.8% among young adults [2]. It is estimated that 18 million Europeans have used cocaine in their lifetime. Cocaine was cited as the primary drug by about 73,000 patients entering specialized drug treatment in 2017 and by more than 33,000 patients in specialized drug treatment for the first time. Estimates of public spending on drugs in Europe range from 0.01-0.5% of gross domestic product.

US data from 2018 [3] shows an estimated 5.5 million people aged 12 or older have used cocaine in the past. Among them, about 977,000 had a cocaine use disorder (0.4% of the population) and 19% of those people received treatment for illicit drug use at a specialized facility.

International standards for the treatment of drug use disorders indicate that essential treatment services should be available within different healthcare systems [4] which include outreach services, brief psychosocial interventions, diagnostic assessment, outpatient psychosocial treatment, pharmacological treatment, services for the management of drug-induced acute clinical conditions, inpatient services for the management of severe withdrawal, and long-term residential services. In Europe, most treatments for drug addiction are provided on an outpatient basis whereas only a small portion are delivered as an inpatient service, mainly in hospital-based residential centers (e.g. psychiatric hospitals) [1]. Treatment pathways followed by patients are often characterized by the use of a variety of different services, by relapses that lead them to begin treatment again, and by different lengths of stay.

A high dropout rate is one of the main limitations of treatment success. Some studies report that among outpatient treatments the dropout rate ranges from 23-50% [5], [6], whereas in residential treatments the dropout rate ranges from 17-57% [7], [8]. Treatment completion is associated with higher levels of abstinence and fewer relapses [9],

as well as lower crime rates [10] and better employment levels [11]–[13]. On the other hand, leaving treatment is associated with greater legal and financial difficulties [14] and entails a high cost to society [15] as well as to loved ones.

According to a systematic review published in 2013, the most consistent dropout risk factors identified were cognitive deficits, low adherence to treatment, personality disorders, and younger age [9]. Cognitive deficits are extensive among patients in addiction treatment and, in addition to an increased risk of dropout [16]–[18], have been associated with some personality disorders [19].

The generation of successful predictive models would lead to a better understanding of the behavior of patients with drug addiction problems and could help to identify the most important risk factors that impede treatment. In medical and psychological research, it is generally found that one size does not fit all [20], and in many situations interactions between independent heterogeneous variables can offer potential outcomes that help to better understand the factors that explain treatment dropout.

Machine learning (ML) methods are useful to learn the relationships that exist amongst data points, and this can be applied to clinical datasets for the purpose of developing robust risk models [21]. The development of decision support systems based on predictive machine learning models has become a subject of enormous scientific interest, and this can be observed in various fields of medicine, finding systems that allow the identification of prescriptions with high risk of medication error [22], the prediction of metastasis in gastric cancer [23], the detection of pediatric autism [24], the prediction of coronary artery disease [25], the treatment of large kidney stones [26] or the prediction of heart failure [27] among others.

Recent work has shown the success of ML methods in clinical psychology and psychiatry that have explicitly focused on learning statistical functions from multidimensional datasets to make generalizable predictions about individuals [28]. At the same time, applications of ML in computational psychiatry [29] and computational neurosciences [30] are emerging.

ML methods, particularly supervised learning, are increasingly used in addiction psychiatry for informing medical decisions and further investigations of the potential applications of ML in precision psychiatry and neuroscience are warranted [31]. In fact, there is already evidence in some fields of addiction, such as the case of alcohol-dependent patients, in which machine learning models seem to be more accurate than psychologists in predicting treatment outcomes in abstinence programs [32].

The aim of this research was to develop a predictive model using ML with high-dimensional heterogeneous data to estimate success in inpatient treatment for those with cocaine addiction, as a first step towards developing a decision support system to assist clinicians in charge of referring and guiding patients in their treatment pathway. The main

contributions of the research work proposed in this article can be summarized as follows:

- To our knowledge, this is the first research study using ML techniques on a European population sample with cocaine use disorder.
- The patient sample size is the largest that has been used in this context focused on the specific case of patients with cocaine use disorder (253 subjects).
- To our knowledge, this is the first research study that has considered the characteristics of the previous treatments together with mental, cognitive, personality, and pharmacological variables in cocaine addiction. The large number of high-dimensional heterogeneous variables considered in this article allows for a more accurate representation of the context of the patients, as well as the possible synergies between these dimensions for prediction.
- In this work, in addition to evaluating the performance of ML algorithms, different feature selection techniques are also used and compared. The best results are contrasted with those shown in the literature.
- The predictive model developed in this research work is discussed in terms of possible benefits obtained from their use in clinical practice when included in a decision support system, improving the possibilities of reducing the treatment dropout rate, the number of future readmissions, and the consequent waiting lists. This would lead to a better use of health care resources and would have a positive impact in economic and social aspects.

The paper is structured as follows. Section II reviews the background and work related to the application of ML techniques in drug addictions. Section III describes the materials and methods used, begins with a description of the dataset and the characteristics, continues with the processing and filtering criteria applied to the data, the feature selection methods and the selected ML algorithms, and ends with the proposed validation methodology to evaluate the results. Section IV presents the results obtained with the ML algorithms and the different sets of characteristics. Section V includes the discussion based on the results obtained. Section VI presents the conclusions of the paper.

## II. BACKGROUND AND RELATED WORK

The research on ML techniques applied to the field of addictions is still quite limited. According to a recent systematic review about the applications of ML in addiction studies, previous research works could be divided according to the type of addiction, differentiating between smoking cigarettes, alcohol, cocaine, opioids, multiple substance use, internet addiction, and game addiction. Focusing on the papers on drug use, this review includes 14 papers published between 2012 and 2018, most of them developed in North America and to a lesser extent in Europe and Asia, and with sample sizes ranging from 22 to 228,405 subjects [31].

The subtypes of ML methods used in these research works are classification, regression, ensemble, multiple comparison of algorithms, clustering, and direct reinforcement learning, with most being supervised learning methods. Among the algorithms used are Support Vector Machine (SVM), decision trees, Random Forest (RF), artificial neural networks, classification and regression trees, Naïve Bayes, discriminant analysis, logistic regression (LR), penalized regression, nearest neighbors, elastic net, K-means clustering, K-medoids clustering and Q-learning. Among the algorithms and methods of feature selection used are correlations, regression coefficients, analysis of variance, chi-square test, grid search, filters, wrappers, information gain and Pearson's chi-squared test. Model evaluation methods included k-fold cross-validation, the receiver operating characteristic (ROC) curve, chi-squared test, leave-one-out cross-validation, variance analysis and multiple comparisons with Bonferroni correction.

A wide range of characteristics, combined in different ways, are used in these studies, including substance problems, route of administration, frequency, age at first use, demographics, psychopathology, personality, risk, cognitive, peer pressure, motives, attitudes, impulsivity, psychiatric problems, clinical, executive function, functional magnetic resonance imaging, arousal level, tense level, restlessness level, brain regions, family history, life events and genetic heritability.

In the field of cocaine addiction, Deane *et al.* [7] used a binary logistic regression for predicting dropouts in the first 3 months of residential drug and alcohol treatment in Australia. The final sample consisted of 618 participants with different primary drugs of abuse. Predictor variables used were age, gender, primary drug of concern, criminal involvement, psychological distress, drug cravings, self-efficacy to abstain, spirituality, forgiveness of self and others, and life purpose. The overall model accuracy was 61.6%, with 76.5% accuracy for being able to predict dropouts and 42.3% for nondropouts.

Anh *et al.* [33] used the least absolute shrinkage and selection operator regression to classify US individuals with cocaine dependence by impulsivity variables, both self-reported and from neurocognitive tasks. Despite the small sample size (31 cocaine-dependent and 23 healthy subjects), the area under the receiver operating characteristic curve (AUC) obtained was 0.912 in the test set, showing the subjects with cocaine dependence higher scores on motor and non-planning trait impulsivity, and poor response inhibition, discriminability and decision making.

Acion *et al.* [34] used a single dataset of 99,013 patients to compare different ML models for prediction of successful outpatient treatment focused on Hispanic Americans with different substance addictions. They considered 28 predictor variables (10 patient characteristics, 3 treatment characteristics, source of referral, summary of type of problematic substance, and mental health problems) and compared 5 ML algorithms (LR, penalized regression, RF, deep learning

neural networks, and Super Learning). All the algorithms were evaluated using the AUC. The best results were provided by the Super Learning and RF algorithms, obtaining AUCs of 0.820 and 0.816, respectively.

The research on cocaine use by Mete *et al.* [35] is focused on medical imaging. It used a SVM algorithm to classify brain images of 93 cocaine-dependent participants and 69 healthy controls in the US population, obtaining F-measures of 0.88 and 0.89, sensitivities of 0.83 and 0.90, and specificities of 0.83 and 0.89 in 10-fold cross-validation and leave-one-out approaches, respectively. Sakoglu *et al.* [36] also used an SVM algorithm for both feature selection/reduction and classification, based on fMRI data obtained from 58 cocaine-dependent participants and 25 healthy subjects from the US population while performing a stop signal task. The aim was to determine whether dynamic functional connectivity features were more successful than static functional connectivity features in classification of cocaine-dependent patients and healthy controls. Based on dynamic functional connectivity, participants were successfully classified with 95% accuracy, whereas static functional connectivity yielded only 81%. Visual, sensorimotor, default mode, executive control networks, amygdala, and insula played the most significant role in classification.

Rish *et al.* [37] used fMRI data to investigate the effects of methylphenidate on the brain activity of US individuals with cocaine use disorders. The sample used included 18 subjects with cocaine use disorder and 16 control subjects. They used different classifiers, including naïve Bayes, nearest neighbour, linear discriminant analysis, LR, linear SVM, decision trees, and RF. They found that the classification error was 10-20% lower classifying subjects with cocaine use disorders under placebo than under methylphenidate, obtaining the best result with the linear discriminant analysis algorithm.

Yip *et al.* [38] used fMRI data to identify a brain-based predictor of cocaine abstinence by using connectome-based predictive modeling, a machine learning approach optimized for neuroimaging data, in order to identify networks that underlie specific behaviors. The sample used included 98 US subjects with cocaine use disorder. The algorithms were applied pre-treatment and post-treatment and leave-one-out cross validation was conducted. Abstinence was predicted during treatment with a significant correspondence between predicted and actual abstinence values ( $r=0.49$ ,  $df=52$ ). Connectivity strength did not change with treatment, and post-treatment assessment also significantly predicted abstinence during follow-up ( $r=0.34$ ,  $df=39$ ). Network strength in the independent sample predicted treatment response with 64% accuracy by itself and 71% accuracy when combined with baseline cocaine use.

Panlilio *et al.* [39] used unsupervised machine learning techniques on drug test results of patients with opioid and cocaine addiction problems, including hierarchical clustering of categorical results and K-means longitudinal clustering

of quantitative results. The sample used included 426 US subjects. They identified four clusters of use, categorized into opioid use, cocaine use, dual use (opioid and cocaine), and partial/complete abstinence. Contingency management increased membership in clusters with lower levels of drug use and fewer symptoms of substance use disorder.

SVM algorithms are widely used in psychiatry [21] due to their origin in early multivariate pattern recognition [43] that aimed to automatically discover regularities in multivariate data to fulfil a goal [44]. The main SVM applications in psychiatry focus on neuroimaging analysis, either for brain disorders [40], schizophrenia [41], or imaging biomarkers for neurological and psychiatric disease [42].

Literature comparisons are difficult due to the diverse nature of the studies undertaken, mainly caused by the inclusion of different combinations of drug abuse with mental problems and other diseases, the amount of substances consumed by the patients, the predictor variables used in the analysis, the heterogeneity of the samples and study designs, and the different metrics to show the performance of the predictive models. This heterogeneity may explain the contradictions and non-significant results that constitute the main important gaps in the research [9], and the need to conduct further research to represent the potential applications of machine learning methods in the field of addiction psychiatry [31].

### III. MATERIALS AND METHODS

#### A. DATA EXTRACTION

The usual therapeutic process for drug addiction patients in Spain begins with outpatient treatment. If abstinence from substance use becomes difficult, the patient is referred to inpatient treatment where he or she will remain in the hospital for a period of time [2]. It is necessary that referring physicians complete a patient referral report to recommend enrolment in the inpatient center, which includes information on health, psychological, pharmacological, family environment, social, and occupational aspects.

In the region of Madrid, the referral reports are formatted as tables and free text fields. For the extraction and structuring of the information, a collaborative effort with the psychiatrists was necessary in which the variables of interest and the metrics used for each of them were defined. A total of 120 characteristics were identified as potential predictors of treatment success. The initial dataset was created manually by a group of 3 researchers who were responsible for reading the referral reports and converting them into a structured database.

The target variable selected for prediction was success in inpatient treatment, which was defined as the completion of inpatient treatment. Cases in which the patient dropped out of treatment prematurely, including voluntary dropout against professional advice or expulsion from the treatment center for failure to comply with rules of conduct, were considered unsuccessful treatment.



## B. COHORT DESCRIPTION

The structured dataset included referral reports from 253 patients admitted to the cocaine addiction unit at the Clínica Nuestra Señora de la Paz (CNSP).

The distribution of men-women in the population sample was 80%-20%, respectively (see Table 1), which was lower for men than the average reported for European countries (85%). The proportion of patients using the inhaled administration route was similar to European figures (25%) while the nasal sniffed administration route (74.7%) was higher than the European figures (68%) [1]. Most patients had a previous history of polydrug use: 15.41%, 76.28% and 82.82% had consumed heroin, cannabis, or alcohol, respectively, at some point in their lives.

## C. DATA DESCRIPTION

Data extracted from the referral reports was grouped into 8 dimensions:

### Socio-demographic

1. **factors**(4 characteristics): sex; age at time of referral; nationality; and educational level.
2. **Mental exploration and cognitive functioning** (18 characteristics): appearance; awareness; orientation; attention; memory; psychomotricity; perception; thought; language; affectivity; appetite; sleep; sexual behavior; intelligence; numeracy; judgment capacity; abstract thinking; and volition.
3. **Personality assessment** (1 characteristic): followed DSM IV-TR criteria [45].
4. **Social, cognitive, and affective skills** (9 characteristics): communication; assertiveness; planning and organization; problem identification and analysis; decision making for troubleshooting; self-esteem; self-criticism; knowledge and control of emotions; and tolerance of frustration.
5. **Substance use habits and severity of drug usage history** (62 characteristics): substance used (cocaine, heroin, cannabis, alcohol, tobacco, other psychoactive substances); consumption habits of each substance (administration route at entry and current, frequency at entry and current, last month's frequency, last year's frequency, age of regular use, starting age, where, when and with whom drugs were used, and drug-related activities); and year when the patient first asked for help.
6. **Pharmacological treatment** (14 characteristics): number and types of prescribed psychiatric drugs (categorized into antidepressants, anxiolytics, antipsychotics, sedative antipsychotics, mood stabilizers, atomoxetine, buprenorphine or naloxone, bupropion, disulfiram, methadone, topiramate and trazodone or mirtazapine); and other non-psychopharmaceutical prescription drugs.
7. **Previous addiction treatments** (7 characteristics): number of previous outpatient and/or inpatient

treatments (previous hospitalization programs were differentiated by whether treatment was at the CNSP or in other inpatient facilities); number of previous treatments in day centers; maximum and total length of stay in inpatient treatments; and maximum abstinence time after drug addiction treatment.

8. **Other individual parameters related to personal health** (5 characteristics): body mass index; heart rate; HIV serology; Global Assessment of Functioning score (GAF, American Psychiatric Association [46]); and outpatient treatment center from which the patient was referred.

Data on treatment success was extracted from the electronic medical record.

## D. DATA PROCESSING AND FILTERING CRITERIA

A reduction in the dimensionality of the characteristics was performed, eliminating those highly affected by missing values or not providing information due to their low variability. Exclusion criteria were those variables with: >25% missing values; >90% of records in a single category; >95% of categories as a percentage of the number of records; and a minimum variation coefficient >0.1. An additional step was to eliminate highly correlated variables using the absolute value of Pearson's correlation coefficient and taking  $\geq 0.6$  as a cut-off point for a high correlation. As a result, the dimensionality of the set of characteristics was reduced from 120 to 60.

The selected characteristics and the treatment success rate are shown in Table 1. The characteristics marked with \*\* showed a statistically significant difference between both groups of patients (treatment success and dropouts) with  $p < 0.01$ , and those marked with \* showed a statistically significant difference with  $p < 0.05$ .

For subsequent steps, outlier values in continuous fields were replaced by a cut-off value at a distance of 3 standard deviations from the mean and continuous fields were then normalized to a common scale with a mean value of 0.0 and a standard deviation value of 1.0. Also, the nominal fields were transformed into numerical values.

## E. MACHINE LEARNING METHOD

The large number of available characteristics suggested applying feature selection techniques to obtain the optimal set. We evaluated 4 ML algorithms (RF, LR, multilayer perceptron (MLP) neuronal network, and SVM) that are widely used in predicting therapeutic outcomes in drug addiction and psychiatry, in combination with 3 different methods of feature selection (wrappers, filters, and hybrid methods). We used the Weka data mining tool [47], version 3.8.3, which provided all the algorithm implementations utilized here.

RF is a recursive partitioning method that consists of a large number of individual decision trees that operate as an ensemble, being able to evaluate a number of predictors even in the presence of complex interactions [48]. RF is more

TABLE 1. List of characteristics after data processing and filtering.

Variable	Values	Total N (%)
<b>Socio-demographic factors</b>		
Sex	Female	20.16%
	Male	79.84%
Age	Integer [20, 71]	-----
Education	Unspecified	1.97%
	Without studies	17%
	Elementary studies	44.27%
	Upper secondary studies	26.88%
	University degree	9.88%
<b>Mental exploration and cognitive functioning</b>		
Appearance	No apparent alterations	71.15%
	Congruent with state	26.48%
	Altered	2.37%
Awareness	No apparent alterations	76.28%
	Congruent with state	23.32%
	Altered	0.4%
Orientation	No apparent alterations	87.75%
	Congruent with state	11.07%
	Altered	1.18%
Attention	No apparent alterations	60.47%
	Congruent with state	29.25%
	Altered	10.28%
Memory	No apparent alterations	67.59%
	Congruent with state	24.11%
	Altered	8.3%
Psychomotricity	No apparent alterations	83.79%
	Congruent with state	10.28%
	Altered	5.93%
Perception	No apparent alterations	80.63%
	Congruent with state	15.81%
	Altered	3.56%
Thought	No apparent alterations	55.33%
	Congruent with state	33.6%
	Altered	11.07%
Language*	No apparent alterations	83.40%
	Congruent with state	12.25%
	Altered	4.35%
Affectivity	No apparent alterations	20.16%
	Congruent with state	49.41%
	Altered	30.43%
Appetite	No apparent alterations	51.78%
	Congruent with state	30.83%
	Altered	17.39%
Sleep	No apparent alterations	31.23%
	Congruent with state	42.29%
	Altered	26.48%
Sexual behavior	No apparent alterations	70.36%
	Congruent with state	20.55%
	Altered	9.09%
Numeracy	No apparent alterations	88.54%
	Congruent with state	11.07%
	Altered	0.39%
Judgment capacity	No apparent alterations	73.91%
	Congruent with state	23.32%
	Altered	2.77%
Abstract thinking	No apparent alterations	83.40%
	Congruent with state	15.81%
	Altered	0.79%
Volition	No apparent alterations	52.18%
	Congruent with state	33.2%
	Altered	14.62%
<b>Personality assessment</b>		
Personality assessment**	Normal	33.99%
	Mild dysfunction	45.06%
	Moderate dysfunction	15.02%
	Severe dysfunction	5.93%
<b>Social, cognitive, and affective skills</b>		
Communication	Very low	0.79%
	Low	18.58%

TABLE 1. (Continued.) List of characteristics after data processing and filtering.

Assertiveness	Media	41.50%
	High	33.99%
	Very high	5.14%
	Very low	6.72%
	Low	50.60%
Planning and organization	Media	29.64%
	High	11.46%
	Very high	1.58%
	Very low	13.04%
	Low	42.69%
Problem identification and analysis	Media	31.22%
	High	12.65%
	Very high	0.4%
	Very low	8.69%
	Low	43.87%
Decision making for troubleshooting	Media	35.97%
	High	11.07%
	Very high	0.4%
	Very low	11.46
	Low	61.27%
Self-esteem	Media	21.74%
	High	5.53%
	Very high	0%
	Very low	13.44%
	Low	56.92%
Self-criticism	Media	22.13%
	High	7.51%
	Very high	0%
	Very low	10.28%
	Low	35.97%
Knowledge and control of emotions	Media	34.38%
	High	16.21%
	Very high	3.16%
	Very low	16.21%
	Low	59.68%
Tolerance of frustration*	Media	20.55%
	High	3.56%
	Very high	0%
	Very low	18.97%
	Low	58.50%
<b>Substance use habits and severity of drug usage history</b>		
Cocaine administration route at entry	Oral	0%
	Nasal sniffing	74.70%
	Inhaling	24.51%
Current cocaine administration route*	Injecting	0.79%
	Oral	0.39%
	Nasal sniffing	79.05%
Current cocaine frequency	Inhaling	19.37%
	Injecting	1.19%
	Every day	16.6%
	4-6 days/week	11.85%
	2-3 days/week	28.06%
Cocaine frequency at entry	1 day/week	13.04%
	< 1 day/week	13.83%
	Did not use	16.6%
	Every day	32.02%
	4-6 days/week	22.52%
Last month's cocaine frequency	2-3 days/week	25.3%
	1 day/week	7.51%
	< 1 day/week	8.7%
	Did not use	3.95%
	Every day	19.76%
Last year's cocaine frequency	4-6 days/week	15.42%
	2-3 days/week	28.06%
	1 day/week	11.86%
	< 1 day/week	13.04%
	Did not use	11.86%
Last year's cocaine frequency	Every day	22.53%
	4-6 days/week	20.16%

**TABLE 1. (Continued.) List of characteristics after data processing and filtering.**

	2-3 days/week	31.22%
	1 day/week	7.51%
	< 1 day/week	16.6%
	Did not use	1.98%
Where-cocaine	Unspecified	16.6%
	Anywhere	17.39%
	On the street	22.13%
	In a pub and/or at work and/or in the car	25.3%
	At home	18.58%
When-cocaine	Unspecified	17.39%
	Anytime	37.55%
	In the morning and in the evening and at night	1.98%
	Weekends and holidays	3.16%
	Other combinations	39.92%
With whom-cocaine	Unspecified	14.22%
	Anyone	1.98%
	With family and/or couple	7.91%
	With friends and/or acquaintances	36.36%
	Alone	39.53%
Cocaine-related activities	Unspecified	18.97%
	Any	54.55%
	In a pub	9.09%
	At work or leisure time	17.39%
Starting age-cocaine	Integer [10,52]	-----
Cannabis use	Used	23.72%
	Not used	76.28%
<b>Pharmacological treatment</b>		
Psychopharmaceuticals	Not used	15.01%
	Consume 1	14.23%
	Consume 2	17%
	Consume 3	22.53%
	Consume 4	18.58%
	Consume 5	9.09%
	Consume 6 to 8	3.56%
	Not used	60.47%
Antidepressants	Consume 1	38.74%
	Consume 2	0.79%
	Not used	41.90%
Anxiolytics	Consume 1	42.29%
	Consume 2	14.23%
	Consume 3	1.58%
Antipsychotics	Not used	67.59%
	Consume 1	32.41%
Disulfiram	Not used	87.75%
	Consume 1	12.25%
Mood stabilizers	Not used	62.85%
	Consume 1	34.78%
	Consume 2	2.37%
Topiramate	Not used	86.56%
	Consume 1	11.86%
	Consume 2	1.58%
Trazodone or mirtazapine	Not used	80.63%
	Consume 1	19.37%
	Not used	67.60%
Non-psychopharmaceutical	Consume 1	15.41%
	Consume 2	7.11%
	Consume 3	5.53%
	Consume 4 to 10	4.35%
<b>Previous addiction treatments</b>		
Previous outpatient treatments	Integer [1,9]	-----
Previous treatments in day centers	Integer [0,3]	-----
Previous inpatient treatments (NOT in CNSP)	Integer [0,11]	-----

**TABLE 1. (Continued.) List of characteristics after data processing and filtering.**

Previous inpatient treatments (in CNSP)	Integer [0,4]	-----
Maximum length of stay in inpatient treatments*	Integer [0,35] (months)	-----
Maximum abstinence time after drug addiction treatment	Integer [0,252] (months)	-----
<b>Other individual parameters related to personal health</b>		
Body mass index**	Integer [15.4,42.7]	-----
Heart rate	Integer [50,110]	-----
Outpatient treatment center from which the patient was referred*	Integer [1,71]	-----
<b>Treatment Success</b>		
Treatment Success	Yes	71.54%
	No	28.46%

protective against overfitting in comparison with other tree algorithms since the low correlation between the trees of the forest protect each other from their individual errors.

LR is a regression method similar to the linear regression model but it is suited to models where the dependent variable is dichotomous. This model uses the maximum-likelihood ratio to determine the statistical significance of the variables [49]. LR imposes less stringent requirements than linear regression, in that it does not assume linearity of the relationship between the explanatory variables and the response variable and does not require Gaussian distributed independent variables [50].

MPL neural network is a combination of perceptrons stacked in several layers, to solve complex problems generating non-linearity classification rules [51]. The relationships between the perceptrons are defined by weights calculated using a given rule. Each layer can have a large number of perceptrons, and there can be multiple layers, so the MLP can quickly become a very complex system, which is the case of the Shallow Neural Network or Deep Neural Network.

SVM is a supervised ML algorithm which can be used for classification or regression problems [52]. It basically finds the hyper-plane that best differentiates the classes within the variable dimensional space. It is effective in high-dimensional spaces, even if the number of dimensions is greater than the number of samples, and it is also memory efficient.

For RF, LR, and MLP we used the basic implementations available in Weka and in the case of SVM we used the sequential minimal optimization algorithm for training a support vector classifier [53].

Feature selection processes combined a searching algorithm and an evaluator that scored each characteristic or set of characteristics. We tested the 3 main approaches of evaluators: filters, wrappers, and hybrid methods.

1. Filter methods score each subset of features using heuristics based on general characteristics of the data

rather than using a learning algorithm, as shown in Fig. 1. They are independent from algorithms and are computationally very fast, but only consider individual characteristics of features to identify their relative importance. Among the filter models we chose Correlation-based Feature Selection (CFS) which evaluates the worth of a feature subset according to a correlation based heuristic evaluation function. It works on the assumption that a characteristic is useful if it is class correlated or class predictive. The bias of the evaluation function is toward subsets that contain features that are highly correlated with the class and uncorrelated with each other. Redundant features should be screened out as they will be highly correlated with one or more of the remaining features. The calculation of the merit of a feature subset is based on a ratio, in which the numerator represents how predictive of the class a set of features is, and the denominator represents how much redundancy there is between the features. CFS’s feature subset evaluation function is as follows:

$$M_S = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k - 1)\bar{r}_{ff}}} \quad (1)$$

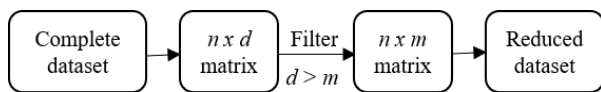


FIGURE 1. Filter feature selection method.

where  $M_S$  is the heuristic “merit” of a feature subset  $S$  containing  $k$  features,  $\bar{r}_{cf}$  is the mean feature-class correlation ( $f \in S$ ), and  $\bar{r}_{ff}$  is the average feature-feature intercorrelation [54].

- In the wrapper approach, the feature subset selection algorithm is integrated as a wrapper around the learning algorithm, as shown in Fig. 2. The feature subset selection algorithm performs a search for the optimal subset using the learning algorithm itself as part of the feature subset evaluation. In this way the learning algorithm is considered a black box. The learning algorithm is run on different sets of features from the original dataset, usually partitioned into internal training and holdout sets. The feature subset with the highest evaluation is chosen as the final set on which to run the learning algorithm. The wrapper method uses a 5-fold cross-validation as an evaluation function, so that the

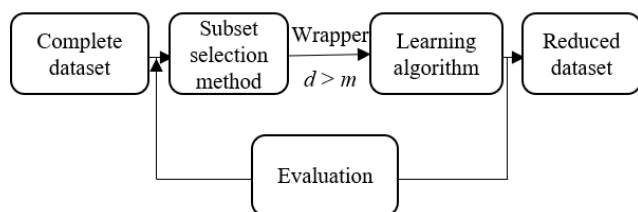


FIGURE 2. Wrapper feature selection method.

score assigned to a subset of features is the accuracy obtained in the cross-validation test set [55]. Due to the large number of iterations necessary to evaluate the performance of each possible set of characteristics, the wrapper method is computationally intensive but at the same time provides the best possible set of characteristics.

- Hybrid methods try to combine the best properties of filters and wrappers, where a filter method reduces the dimensionality of feature space and a wrapper method finds the optimal subset afterwards. We chose the Incremental Wrapper Subset Selection (IWSS) hybrid method, by first creating a ranking of features based on the correlation-based metric, and later running IWSS over the whole ranking, giving each attribute in the data set the chance of being selected, and selecting those attributes that improve performance for a given minimum number of folds out of the wrapper cross-validation folds. The main advantage of this approach is that it retains a great part of wrapper advantages, while reducing the computational cost of pure wrapper approaches. The main disadvantage of the IWSS algorithm is its greedy behavior, since the algorithm always tries the best ranked features first and once a feature is included in the selected set, it is maintained therein until the end of the search [56]. Table 2 shows the pseudo-code of IWSS method.

TABLE 2. Incremental wrapper based subset selection algorithm.

In	T (training set), FM (filter measure), C (classifier)
Out	S // The selected subset
1	list R = {} // The ranking
2	for each attribute $A_i \in T$
3	Score = FM( $A_i$ , class) // e.g. Correlation-based( $A_i$ , class)
4	insert $A_i$ in R according to Score
5	S = {R[1]}
6	BestData = evaluate(C,S,T)
7	for i = 2 to n // n = R.size()
8	$S_{aux} = S \cup \{R[i]\}$
9	AuxData = evaluate(C, $S_{aux}$ , T)
10	if (AuxData > BestData)
11	S = $S_{aux}$
12	BestData = AuxData

Weka default parameters were maintained in all the evaluators.

Since not all searching algorithms perform equally with different datasets, we decided to explore algorithms belonging to different categories. There are three main categories available for searching algorithms: “Exponential search”, “Sequential search” and “Random search” [57]. The main drawback of “Exponential search” is that it requires  $2^N$  combinations for  $N$  variables, which is too computationally intensive and time consuming, and for this reason we decided to focus on the other two categories.



In “Sequential Search”, features are sequentially added to an initially empty set (forward strategy) or features are sequentially removed from an initially complete set (backward strategy) until the addition or removal of a characteristic does not result in a higher evaluation. Its main problem is the nesting effect because the deleted features cannot be considered in later iterations. From this category we tested the Best First (BF) algorithm with a forward searching strategy [47], which searches the space of features subsets by greedy hill climbing augmented with a backtracking facility, being simple and fast. To prevent the BF search from exploring the entire feature subset search space, a stopping criterion is imposed, so that the search will end if a certain consecutive number of subsets show no improvement over the current best subset. Table 3 shows the pseudo-code of BF algorithm.

**TABLE 3. Best first search algorithm.**

```

closed list = [ ]
open list = [start node]
do{
  if open list is empty then{
    return no solution;
  }
  n = heuristic best node;
  if n == final node then{
    return path from start to goal node;
  }
  for each direct available node do{
    if node not in open and not in closed list do{
      add node to open list;
      set n as his parent node
    }
  }
  delete n from open list;
  add n to closed list;
} while (open list is not empty)
}

```

“Random search” methods could be categorized as global search because they try to generate an approximate and efficient solution instead of a more precise solution obtained with sequential search, but less costly in time and computational resources. From this category we tested Genetic Search (GS). These algorithms are based on the biological process of evolution through natural selection. Each individual in the sample is characterized by a set of binary variables called genes, and these are linked together in strings to form chromosomes, which could be understood as solutions. It uses a fitness function through which a fitness score is assigned to each individual, and this represents the probability that an individual is selected for reproduction. Based on the fitness score, two pairs of individuals are selected for reproduction, which would represent the role of parents. From the parents, offspring are generated by crossover and mutation, creating new individuals with combinations of the parents’ genes. The algorithm ends when the generated population converges, that is, when no significantly different offspring are produced from the previous generation. At this point is when the genetic

algorithm has provided a set of solutions to the initial problem [58]. Table 4 shows the pseudo-code of GS algorithm.

**TABLE 4. Genetic search algorithm.**

```

n: number of individuals in the population
X: fraction of population to be replaced by
crossover in each iteration
μ: mutation rate

// Initialize generation 0:
k = 0;
Pk = a population of n randomly-generated individuals;
// Evaluate Pk:
Compute fitness(i) for each i ∈ Pk;
do{
  // Create generation k+1:
  // 1. Copy:
  Select (1 - X) × n members of Pk and insert into Pk+1;
  // 2. Crossover:
  Select X × n members of Pk; pair them up; produce
  offspring; insert the offspring into Pk+1;
  // 3. Mutate:
  Select μ × n members of Pk+1; invert a randomly-selected
  bit in each;
  // Evaluate Pk+1:
  Compute fitness(i) for each i ∈ Pk+1;
  // Increment:
  k = k + 1;
}
while fitness of fittest individual in Pk is not high enough;
return the fittest individual from Pk;

```

In the case of the BF searching algorithm, we increased the allowed number of consecutive non-improving nodes before terminating the search from 5 to 15. For the GS algorithm we increased the number of generations to evaluate from 20 to 40.

The feature selection was made by means of a 10-fold cross-validation in all the techniques used. In the case of the wrapper method, a 5-fold cross-validation for feature selection was performed within each of the 10 folds [55]. For each feature selection method, we included in the final characteristic set those variables that were selected in at least 8 of the folds and we discarded all those not selected in any of the folds. Then we followed a sequential forward strategy adding the rest of the features one by one, selecting the most popular each time and stopping when accuracy decreased.

By combining the ML algorithms and the different feature selection techniques (evaluator and searching algorithm) we obtained 14 different sets of variables.

## F. EVALUATION

A 10 times repeated 10-fold stratified cross-validation was used to evaluate the performance of each ML algorithm with the 14 final characteristic sets obtained during the feature selection process. These results were also compared with a Zero Rule classifier, which is the simplest classifier since it predicts the majority class in the existing data.

For evaluation measures we used accuracy (2), AUC, recall (3), specificity (5), F1-score (6), and Matthews

correlation coefficient (MCC) (7), calculating for each the 95% confidence interval for the mean. In these mathematical equations we will refer to the number of true positives as TP, true negatives as TN, false positives as FP and false negatives as FN.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (7)$$

The AUC provides a better measure than accuracy, especially when measuring and comparing classification systems, as it compares the classifiers' performance across the entire range of class distributions and error costs [59].

F1-score is the harmonic mean of precision and recall, where precision measures the ability of a classifier to predict positive samples as positive and recall measures how many actual positive observations are predicted correctly. This parameter tries to overcome the disadvantage of accuracy against unbalanced sets, where a large majority class can introduce bias. F1-score ranges in [0,1], where the minimum is reached when all the positive samples are misclassified, that is, TP = 0, and the maximum for FN = FP = 0, that is for perfect classification. MCC is an alternative measure unaffected by the unbalanced datasets issue, being a contingency matrix method of calculating the Pearson product-moment correlation coefficient between actual and predicted values [60]. It takes values in the range [-1,+1], where -1 and +1 are the cases of perfect misclassification and perfect classification respectively, and 0 when it has no classification capacity.

An essential point in the management of drug rehabilitation treatments is the need to improve their performance, which means an increase in the number of patients who successfully complete the treatment and achieve rehabilitation. It is essential to accurately identify those patients for whom the treatment would work better. For this reason, among the requirements was to achieve the best possible accuracy while reducing the rate of false negatives so as not to classify a patient as a dropout when he or she could benefit from treatment. The reduction of false negatives leads to an increase in the rate of true positives or recall. Therefore, where applicable and where it would be beneficial, we manipulated the cost matrix in order to reduce the rate of false negatives at the cost of also reducing accuracy.

## IV. RESULTS

### A. FEATURE SELECTION

Table 5 shows to which extent a feature was selected for any combination of ML algorithm and feature selection strategy. Each cell represents the number of folds (from 0 to 10) in which the feature was selected during the 10-fold cross-validation. The first column is the feature name and the additional ones show the possible combinations of the type of evaluator (wrapper, IWSS, and CFS), the searching algorithm (BF or GS), and the learning algorithm (RF, LR, MLP, or SVM). Selected features in each combination are shown in bold and underlined.

The number of characteristics selected by each method ranges from 3 to 23. There are 10 characteristics selected in at least half of the methods, which are: judgment capacity, personality assessment, tolerance of frustration, cocaine administration route at entry, current cocaine administration route, antidepressants, antipsychotics, topiramate, previous treatments in day centers, and maximum length of stay in inpatient treatments. Of these characteristics, 'Personality assessment' was selected in all the final characteristic sets, and 'Current cocaine administration route' was selected in 13 out of the 14 final characteristic sets.

### B. LEARNING ALGORITHMS PERFORMANCE

Table 6 shows the results achieved by each algorithm tested with 10 times repeated stratified 10-fold cross-validation. Results are compared with the Zero Rule classifier, which obtains the same values for all the data sets by considering only the target variable as a classification criterion. CS<sub>WBF</sub>, CS<sub>WG</sub>, and CS<sub>IWSS</sub> correspond to characteristic sets obtained by the wrapper evaluator with the BF (WBF), the GS (WG) search algorithms and the IWSS evaluator, respectively. CS<sub>CFBF</sub> and CS<sub>CFG</sub> correspond to characteristic sets obtained by the CFS evaluator with the BF (CFBF) and the GS (CFG) search algorithms, respectively.

The feature selection process and the consequent elimination of redundant features improved the results obtained with the ML algorithms: practically all of them achieved a statistically significant improvement of the AUC and specificity with any of the selected characteristic sets compared to the Zero Rule classifier, except for the AUC of the LR with the CS<sub>CFBF</sub> and CS<sub>CFG</sub>. In the case of accuracy, only the characteristic sets obtained with the wrapper-BF combination achieved a statistically significant improvement in all the methods (RF, LR, MLP, and SVM) together with the IWSS and wrapper-G combination when used with SVM.

In terms of accuracy, AUC, F1-score and MCC the best result was achieved by the RF algorithm with the CS<sub>WBF</sub>. In terms of recall the best combination was achieved by the SVM algorithm equally with the CS<sub>WG</sub> and CS<sub>IWSS</sub>. In terms of specificity the best result was achieved by the MLP algorithm with the CS<sub>WG</sub>.

Our top-performing model was the RF algorithm with the CS<sub>WBF</sub>. It was only 2% lower than the combination with

TABLE 5. Features selected with each combination.

Variable	Wrapper				IWSS				CFS					
	BF		GS		RF		LR		BF	GS				
	RF	LR	MLP	SVM	RF	LR	MLP	SVM	RF	LR				
<b>Socio-demographic factors</b>														
Sex	0	1	1	0	2	<b>4</b>	4	2	0	0	0	0	0	5
Age	0	1	0	1	1	1	<b>10</b>	0	<b>3</b>	<b>4</b>	<b>1</b>	1	0	0
Education	0	0	0	0	3	0	0	3	0	0	1	0	0	1
<b>Mental exploration and cognitive functioning</b>														
Appear.	3	0	0	0	3	0	4	2	<b>2</b>	1	<b>1</b>	0	0	4
Awareness	3	1	1	1	2	1	<b>7</b>	2	<b>2</b>	0	<b>1</b>	0	0	1
Orientation	<b>7</b>	0	3	2	2	3	4	1	0	0	<b>1</b>	0	0	1
Attention	0	1	0	2	1	<b>4</b>	<b>7</b>	2	1	0	0	0	0	4
Memory	0	0	0	0	3	1	4	2	0	1	0	0	0	2
Psycho.	4	0	1	0	<b>5</b>	1	<b>3</b>	2	0	1	0	0	<b>1</b>	1
Perception	1	1	0	0	2	0	1	1	0	0	0	0	0	2
Thought	<b>9</b>	<b>2</b>	<b>6</b>	0	<b>7</b>	1	5	3	1	0	1	0	0	0
Language	0	1	3	0	<b>7</b>	1	6	1	1	0	0	0	<b>6</b>	<b>7</b>
Affection	0	0	0	0	2	2	2	1	1	0	1	0	0	3
Appetite	0	0	0	1	<b>7</b>	2	1	4	0	1	0	0	0	1
Sleep	0	<b>3</b>	0	0	5	1	<b>8</b>	1	0	1	0	0	0	1
Sex. Behav.	0	2	0	0	2	2	4	4	0	0	0	0	0	2
Numeracy	3	<b>3</b>	0	0	<b>8</b>	<b>5</b>	<b>3</b>	2	0	1	<b>2</b>	0	0	2
Judgment capacity	<b>9</b>	<b>2</b>	<b>7</b>	<b>4</b>	4	<b>3</b>	<b>6</b>	2	1	1	<b>2</b>	1	0	2
Abstract thinking	2	0	1	0	4	0	2	0	0	0	0	0	0	1
Volition	2	0	0	0	<b>10</b>	0	2	1	1	0	0	0	0	1
<b>Personality assessment</b>														
Personality assessment	<b>10</b>	<b>10</b>	<b>10</b>	<b>10</b>	<b>9</b>	<b>10</b>	<b>10</b>	<b>10</b>	<b>1</b>	<b>9</b>	<b>4</b>	<b>10</b>	<b>10</b>	<b>10</b>
<b>Social, cognitive, and affective skills</b>														
Communic.	0	0	0	0	1	1	1	0	1	0	0	0	<b>1</b>	2
Assertiven.	0	0	0	1	<b>8</b>	1	<b>8</b>	1	<b>2</b>	0	0	0	<b>2</b>	4
Planning and organi.	0	0	0	0	2	<b>4</b>	<b>8</b>	6	<b>4</b>	<b>1</b>	0	0	<b>3</b>	<b>8</b>
Problem identif. and analysis	0	0	0	0	1	<b>4</b>	2	0	<b>2</b>	0	<b>1</b>	0	<b>3</b>	4
Decision making	1	0	0	1	4	<b>4</b>	6	7	1	1	0	0	<b>1</b>	<b>8</b>
Self-esteem	0	0	0	1	0	3	0	7	0	0	0	0	0	4
Self-critici.	0	0	0	0	2	2	1	1	0	0	0	0	<b>1</b>	1
Control of emotions	1	1	0	0	1	0	5	1	1	<b>1</b>	0	0	0	0
Tolerance to frustrati.	0	0	0	0	<b>6</b>	2	<b>7</b>	0	<b>4</b>	<b>1</b>	<b>2</b>	0	<b>5</b>	<b>8</b>
<b>Substance use habits and severity of drug usage history</b>														
Cocaine admi. route at entry	<b>4</b>	<b>4</b>	<b>1</b>	2	4	<b>3</b>	<b>7</b>	1	<b>1</b>	<b>1</b>	<b>1</b>	0	0	2
Current cocaine admi. route	<b>4</b>	<b>9</b>	<b>8</b>	<b>9</b>	3	<b>10</b>	<b>10</b>	<b>10</b>	<b>2</b>	<b>8</b>	<b>2</b>	<b>10</b>	<b>10</b>	<b>10</b>
Current cocaine frequency	0	0	0	0	3	0	5	0	1	1	0	0	0	0
Cocaine frequency at entry	0	1	0	0	4	0	2	0	1	0	0	0	0	2
Last month cocaine frequency	0	0	0	1	2	0	3	1	0	0	0	0	0	1
Last year cocaine frequency	0	0	0	0	4	0	6	1	0	1	0	0	0	2
Where-cocaine	0	0	0	1	<b>5</b>	1	<b>6</b>	2	1	0	0	0	0	2

TABLE 5. (Continued.) Features selected with each combination.

When-cocaine	0	0	0	0	5	0	5	2	<b>4</b>	0	<b>1</b>	0	0	4
With whom-cocaine	0	0	0	1	0	0	2	0	0	0	0	0	0	0
Cocaine-related activities	0	0	0	0	3	1	6	1	<b>3</b>	1	1	0	0	1
Starting age-cocaine	0	1	0	1	4	1	3	0	0	0	0	0	0	0
Cannabis use	1	<b>2</b>	1	0	5	1	6	2	0	0	0	0	0	3
<b>Pharmacological treatment</b>														
Psychopa.	0	0	0	2	4	0	2	1	0	0	0	0	0	4
Antidepress.	0	<b>5</b>	0	<b>4</b>	<b>6</b>	<b>4</b>	<b>6</b>	4	<b>2</b>	0	<b>1</b>	0	0	2
Anxiolytics	0	1	0	1	1	0	1	2	0	0	0	0	0	7
Antipsych.	1	<b>2</b>	0	0	<b>5</b>	<b>3</b>	3	3	<b>9</b>	<b>3</b>	<b>4</b>	0	0	<b>6</b>
Disulfiram	<b>10</b>	<b>3</b>	<b>6</b>	0	<b>9</b>	<b>5</b>	<b>7</b>	3	0	0	1	0	0	4
Mood stabilizers	0	0	0	0	4	0	6	0	<b>3</b>	0	0	0	0	4
Topiramate	2	<b>8</b>	4	<b>7</b>	<b>5</b>	<b>7</b>	4	<b>7</b>	0	<b>3</b>	0	<b>4</b>	<b>5</b>	<b>7</b>
Trazodone or mirtaza.	1	<b>3</b>	0	0	<b>6</b>	3	4	5	<b>1</b>	0	0	0	0	1
Non-psych.	0	0	0	1	2	0	<b>8</b>	0	<b>6</b>	0	0	0	<b>2</b>	<b>6</b>
<b>Previous addiction treatments</b>														
Previous outpatient treatments	<b>1</b>	0	0	1	1	<b>3</b>	<b>7</b>	2	2	<b>1</b>	<b>1</b>	0	0	0
Previous treatments day centers	<b>9</b>	<b>9</b>	<b>10</b>	2	<b>8</b>	<b>6</b>	<b>8</b>	3	<b>3</b>	0	0	0	<b>1</b>	1
Previous inpatient treatments (NOT in CNSP)	1	1	0	0	<b>6</b>	1	<b>9</b>	3	0	<b>1</b>	0	0	0	0
Previous inpatient treatments (in CNSP)	<b>2</b>	<b>3</b>	1	0	<b>5</b>	2	3	4	0	0	<b>1</b>	0	0	0
Max. stay in inpatient treatments	<b>1</b>	<b>2</b>	1	0	<b>5</b>	1	<b>7</b>	0	<b>9</b>	<b>3</b>	<b>6</b>	3	<b>1</b>	1
Max. abs. time after addiction treatment	0	<b>3</b>	0	0	3	2	<b>8</b>	3	<b>1</b>	0	0	0	0	0
<b>Other individual parameters related to personal health</b>														
BMI	0	2	1	0	3	0	<b>7</b>	0	<b>7</b>	<b>6</b>	<b>6</b>	6	<b>5</b>	5
Heart rate	0	1	0	1	3	0	5	0	1	1	0	0	0	0
Outpatient treatment center from which the patient was referred	0	0	0	0	2	0	<b>9</b>	0	0	0	0	0	<b>10</b>	<b>9</b>
<b>Selected features in each technique</b>														
Total	11	16	7	5	19	16	23	3	21	13	18	3	17	11

the best recall measure but higher in AUC (20%), specificity (23%), F1-score (3%), MCC (17%), and accuracy (5.3%). Similarly, it was 5% lower than the combination with the best specificity measure but higher in AUC (3%), recall (10%), F1-score (5%), MCC (12%), and accuracy (5.6%).

TABLE 6. Learning algorithm results with each characteristic set.

Set	CS <sub>WBF</sub>	CS <sub>WG</sub>	CS <sub>IWSS</sub>	CS <sub>CFBF</sub>	CS <sub>CFG</sub>
<b>Zero Rule</b>					
Accur.	71.6 [71.3,71.80]				
AUC	0.50				
Recall	1.00				
Spec.	0.00				
F1	0.83				
<b>Random Forest</b>					
Accur.	<b>82.1 *</b> [81.0,83.2]	75.7 [74.5,77.0]	75.0 [73.8,76.2]	70.6 [69.5,71.7]	71.8 [70.6,73.0]
AUC	<b>0.81 *</b> [0.79,0.83]	0.72 * [0.70,0.74]	0.71 * [0.69,0.73]	0.67 * [0.65,0.69]	0.65 * [0.62,0.68]
Recall	0.96 [0.95,0.97]	0.94 [0.93,0.95]	0.95 [0.94,0.96]	0.93 [0.92,0.94]	0.91 [0.90,0.92]
Spec.	0.47 * [0.44,0.50]	0.30 * [0.27,0.33]	0.25 * [0.22,0.28]	0.14 * [0.12,0.16]	0.23 * [0.20,0.26]
F1	<b>0.89 *</b> [0.88,0.90]	0.85 [0.84,0.86]	0.84 [0.83,0.85]	0.82 [0.81,0.83]	0.82 [0.81,0.83]
MCC	<b>0.53</b> [0.50,0.56]	0.33 [0.29,0.37]	0.30 [0.26,0.34]	0.13 [0.09,0.17]	0.20 [0.16,0.24]
<b>Multilayer perceptron</b>					
Accur.	78.2 * [77.0,79.4]	76.5 [75.0,78.0]	73.5 [72.0,75.0]	71.4 [69.9,72.9]	67.8 [66.2,69.5]
AUC	0.72 * [0.70,0.74]	0.78 * [0.76,0.80]	0.72 * [0.70,0.74]	0.68 * [0.66,0.70]	0.63 * [0.61,0.65]
Recall	0.90 [0.89,0.91]	0.86 [0.84,0.88]	0.84 [0.82,0.86]	0.82 [0.80,0.84]	0.80 [0.78,0.82]
Spec.	0.48 * [0.45,0.51]	<b>0.52 *</b> [0.49,0.55]	0.48 * [0.45,0.51]	0.46 * [0.42,0.50]	0.38 * [0.35,0.41]
F1	0.85 [0.84,0.86]	0.84 [0.83,0.85]	0.82 [0.81,0.83]	0.80 [0.79,0.81]	0.78 [0.77,0.79]
MCC	0.43 [0.40,0.46]	0.41 [0.37,0.45]	0.33 [0.29,0.37]	0.28 [0.24,0.32]	0.19 [0.15,0.23]
<b>Logistic regression</b>					
Accur.	77.7 * [76.5,78.9]	75.0 [73.6,76.5]	74.2 [72.9,75.4]	67.3 [65.7,68.9]	70.9 [69.6,72.2]
AUC	0.68 * [0.66,0.70]	0.69 * [0.67,0.71]	0.73 * [0.71,0.75]	0.60 [0.57,0.63]	0.62 [0.59,0.65]
Recall	0.96 [0.95,0.97]	0.90 [0.88,0.92]	0.91 [0.90,0.92]	0.78 [0.76,0.80]	0.83 [0.81,0.85]
Spec.	0.33 * [0.30,0.36]	0.38 * [0.35,0.41]	0.32 * [0.29,0.35]	0.39 * [0.35,0.43]	0.40 * [0.36,0.44]
F1	0.86 [0.85,0.87]	0.84 [0.83,0.85]	0.83 [0.82,0.84]	0.77 [0.76,0.78]	0.80 [0.79,0.81]
MCC	0.39 [0.35,0.43]	0.33 [0.29,0.37]	0.29 [0.25,0.33]	0.19 [0.15,0.23]	0.25 [0.21,0.29]
<b>Support Vector Machine</b>					
Accur.	77.5 * [76.4,78.7]	76.8 * [75.8,77.9]	76.8 * [75.8,77.9]	71.8 [70.4,73.1]	73.4 [72.0,74.8]
AUC	0.63 * [0.61,0.65]	0.61 * [0.59,0.62]	0.61 * [0.59,0.62]	0.62 * [0.60,0.63]	0.63 * [0.61,0.65]
Recall	0.97 [0.96,0.98]	<b>0.98</b> [0.97,0.98]	<b>0.98</b> [0.97,0.98]	0.85 [0.83,0.87]	0.88 [0.86,0.90]
Spec.	0.29 * [0.26,0.32]	0.24 * [0.21,0.27]	0.24 * [0.21,0.27]	0.39 * [0.35,0.43]	0.37 * [0.34,0.40]
F1	0.86 [0.85,0.87]	0.86 [0.85,0.87]	0.86 [0.85,0.87]	0.81 [0.80,0.82]	0.83 [0.82,0.84]
MCC	0.38 [0.34,0.42]	0.36 [0.32,0.40]	0.36 [0.32,0.40]	0.27 [0.23,0.31]	0.30 [0.26,0.34]

\* Statistically significant improvement with p<0.01 over the results obtained with a Zero Rule classifier. Accur.: accuracy; AUC: area under the receiver operating characteristic curve; Spec.: Specificity; F1: F1-score; MCC: Matthews correlation coefficient.

The F1-score achieved by RF-CS<sub>WBF</sub> combination obtained a statistically significant improvement over the Zero Rule. Similarly, in the case of the MCC, this combination is in the

range of a moderate and higher-level correlation between the predicted and real values of the target variable.

Fig. 3 shows the best AUC graphs for the two classes of the target variable (treatment success or dropout).

For the best performing algorithm (RF-CS<sub>WBF</sub>), we manipulated the cost matrix in order to reduce the rate of false negatives at the cost of also decreasing the accuracy and the rest of the parameters. Table 7 shows the results obtained for different cost balances between false positives and false negatives. The cost balance 1.9/1.0 obtained a recall of 0.98, which would allow the RF algorithm to reduce the number of false negatives by 2% with an overall accuracy loss of 5.43%. But on the other hand, the specificity, the MCC and the F1-score were also reduced, losing this last parameter its statistical significance.

TABLE 7. Results of rf algorithm manipulating cost matrix.

FN/FP	Accuracy	Recall	Specificity	F1-score	MCC
<b>Cost-sensitive Random Forest</b>					
1.0/1.0	82.1 * [81.0,83.2]	0.96 [0.95,0.97]	0.47 * [0.44,0.50]	0.89 * [0.88,0.90]	0.53 [0.50,0.56]
1.3/1.0	80.2 * [79.0,81.4]	0.97 [0.96,0.98]	0.38 * [0.34,0.42]	0.88 * [0.87,0.89]	0.47 [0.43,0.51]
1.9/1.0	76.7 * [75.7,77.7]	0.98 [0.97,0.99]	0.24 * [0.21,0.27]	0.86 [0.85,0.87]	0.36 [0.32,0.40]

\* Statistically significant improvement with p<0.01 over the results obtained with a Zero Rule classifier.

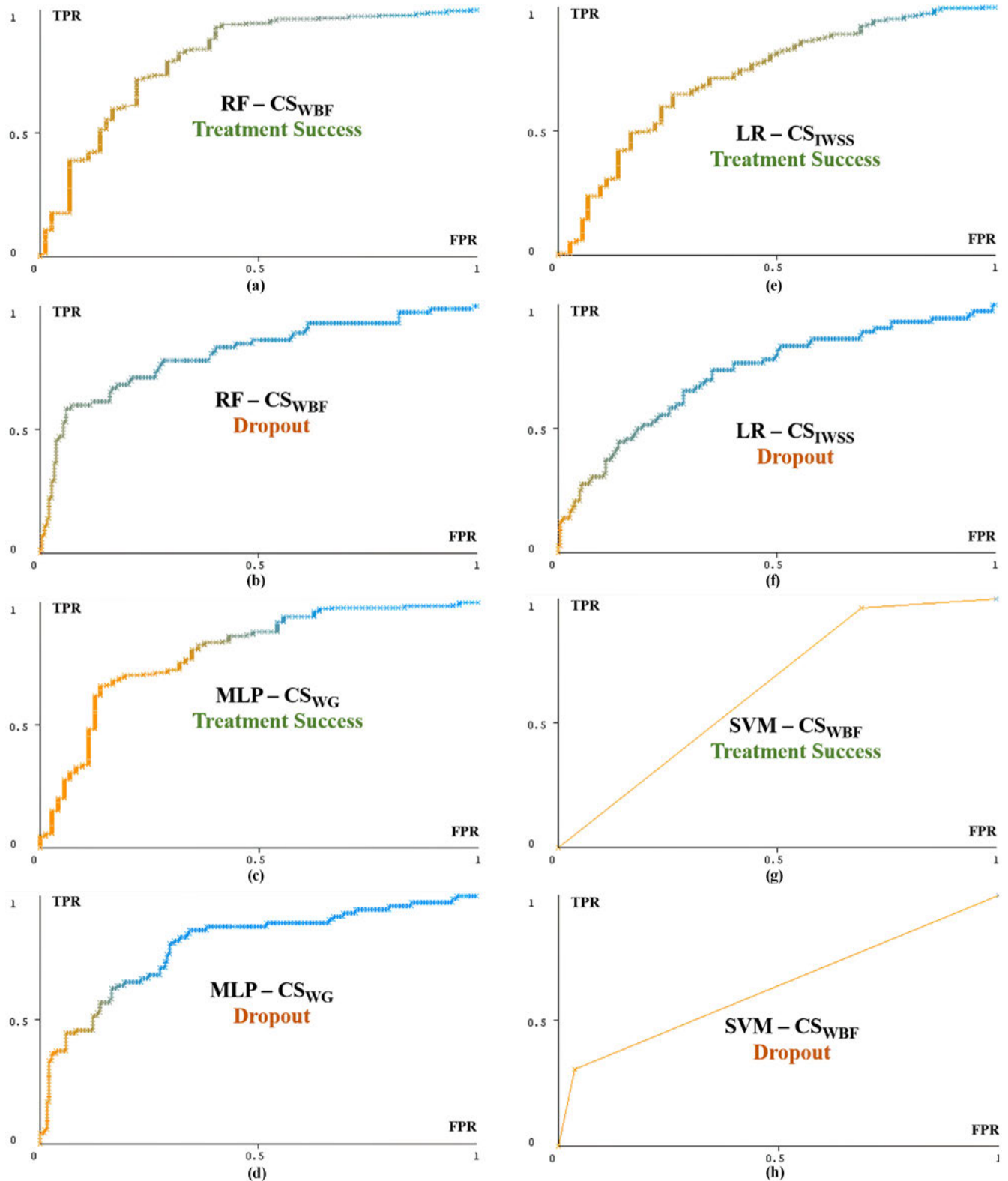
Fig. 4 shows the relative importance of the eleven CS<sub>WBF</sub> characteristics in the RF predictive model. This importance is calculated based on average impurity decrease and number of nodes using the feature, and ranges from 0.25 to 0.37.

The eleven characteristics from CS<sub>WBF</sub> are from 5 different dimensions: 3 from mental exploration and cognitive functioning (thought, judgment capacity, and orientation); 1 from personality assessment; 2 from substances of consumption (cocaine administration route at entry and current cocaine administration route); 1 from pharmacological treatment (disulfiram); and 4 from previous treatment pathway (previous outpatient treatments, previous treatments in day centers, previous inpatient treatments in CNSP, and maximum length of stay in inpatient treatments).

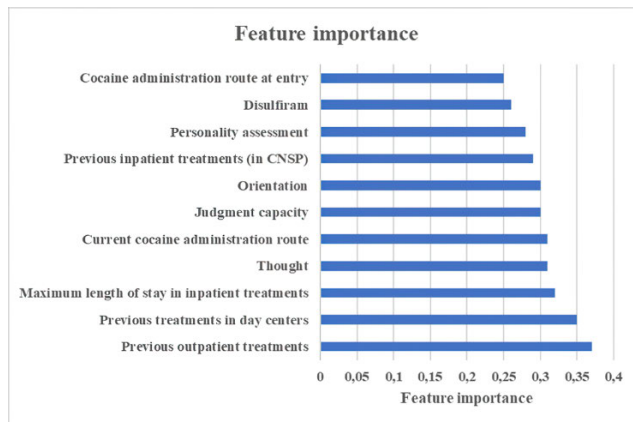
## V. DISCUSSION

In our testing, the best predictive model combined RF and the CS<sub>WBF</sub>, which presented the highest accuracy (82.12%), F1-score (0.89), MCC (0.53), and AUC (0.81), as well as a high recall measure (0.96). In general, an AUC from 0.7 to 0.8 is considered acceptable, from 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding [61]. From this point of view, the RF algorithm with the CS<sub>WBF</sub> can be considered as excellent. The F1-score parameter presents a more optimistic view of the results than the MCC parameter, and this is because MCC gives equal importance to all the parameters of the confusion matrix, or in other words, it gives the same importance to both classes of the target variable. In the case of our top-performing model, probably due to the





**FIGURE 3.** (a),(b) AUC obtained with RF- $CS_{WBF}$  for Treatment Success and Dropout target, respectively; (c),(d) AUC obtained with MLP- $CS_{WG}$  for Treatment Success and Dropout target, respectively; (e),(f) AUC obtained with LR- $CS_{IWSS}$  for Treatment Success and Dropout target, respectively; (g),(h) AUC obtained with SVM- $CS_{WBF}$  for Treatment Success and Dropout target, respectively; TPR: true positive rate; FPR: false positive rate; RF: Random Forest; MLP: multilayer perceptron neural network; LR: logistic regression; SVM: Support Vector Machine;  $CS_{WBF}$ : characteristic set obtained by the wrapper evaluator with the Best First algorithm;  $CS_{WG}$ : characteristic set obtained by the wrapper evaluator with the Genetic Search algorithm;  $CS_{IWSS}$ : characteristic set obtained by the IWSS evaluator.



**FIGURE 4.** Attribute importance for the Random Forest algorithm with the characteristic set obtained by the wrapper evaluator with the Best First algorithm ( $CS_{WBF}$ ).

unbalanced classes, it has better performance in recognizing positive cases (patients who successfully complete treatment) than negative cases (patients who abandon treatment), just as it has a much lower rate of false negatives than false positives. This fits with the initial criterion, which was to give priority to reducing false negatives so that patients who might benefit from treatment would not be classified as dropouts.

The strategy of sacrificing accuracy in favor of recall allowed us to achieve the design criteria of minimizing false negatives, in other words, minimizing the number of patients erroneously classified as dropouts. Depending on the selected point of the RF ROC curve, the algorithm achieved an accuracy in the range of 76.69-82.12%, a recall measure in the range of 0.96-0.98, a specificity in the range of 0.24-0.47, an F1-score in the range of 0.86-0.89 and an MCC in the range of 0.36-0.53.

The treatment of patients addicted to drugs, including cocaine addicts, presents an opportunity for improvement as can be seen in the high level of dropout. These statistics result from the inherent complexity of this type of patient, most of whom suffer from associated mental problems and unfavorable environments. In a similar population, our model suggests that using the developed predictor to decide the referral of patients to inpatient treatment would result in a significant reduction of the dropout rate by 36.82% (from 28.46% to 17.98% of the total admitted patients), with a false negative rate of 4%. However, if the false negative rate is reduced by 2% instead, a dropout rate reduction of 17.11% (from 28.46% to 23.59%) is still obtained. Dropout reduction has a direct impact on the waiting time for treatment access and also on the number of future treatment re-admissions, as treatment completion is associated with higher levels of abstinence and fewer relapses [9]. From a clinical or management point of view, it would mean a clear improvement in the management of resources, with a greater proportion of patients making profitable use of the treatment and all the subsequent synergies in their reintegration into society.

The fact that patients who complete treatment are less likely to relapse, or that relapses are less severe, has

important and enormous economic and social implications that are difficult to quantify. On the one hand, there are the costs generated by drug users who impact on non-drug users, both the pain and emotional distress generated in loved ones, and that generated in potential crime victims. On the other hand, there are the economic costs associated with medical complications generated by drug use, such as the spread of HIV, sexually transmitted diseases, and hepatitis B and C. The medical consequences are broad and pervasive: infectious diseases, cardiovascular effects, respiratory effects, gastrointestinal effects, musculoskeletal effects, kidney damage, liver damage, neurological effects, mental health effects, hormonal effects, cancer, prenatal effects, other health effects, and mortality. The impact of lost productivity of drug users should also be considered, as the nation loses the productive capacity of that person over what could have been a long career. In the United States, drug-related health care spending was estimated at \$12.862 billion in 1998, 69% of which was caused by this loss of productivity. Other related impacts would be the costs of criminal justice, infrastructure, private security, and the social welfare administration [62]. As can be seen, the implications and total costs of drug use are immense and complex to measure, especially those related to emotional well-being and mental health, but an improvement in the opportunities for social reintegration of patients with drug addiction problems would bring relief in all the aspects mentioned above.

The feature selection methodology made it possible to reduce the number of variables from 60 to only 11 in the best characteristic set. This set includes variables from 5 of the 8 possible dimensions. The relevance of high heterogeneous dimensionality can be seen not only in this set but also in the others that have been generated, as all of them include variables related to multiple dimensions and most include variables from all the categories, which confirms the importance of evaluating them together rather than separately. This fact highlights the interactions generated between the heterogeneous variables considered, showing the influence of mental health, cognitive functioning, and the patient's personality on possible treatment success, as well as the importance of previous behaviors in similar treatments, beyond parameters related to substance use. The possibility of employing other methodologies for feature selection or reduction of dimensionality, such as those already mentioned in section II, remains open as possible future work.

The heterogeneous composition of all characteristic sets supports our initial hypothesis about the need to include information related to different areas of the patient in order to optimize the treatment pathway of drug-addicted patients. The 'Personality assessment' characteristic was found to be the most relevant because it was included in all the created sets and, additionally, it was the most important variable in 8 of them. Personality disorder has been one of the main dropout risk factors identified in previous studies [9].

In this work, no relationship was found between treatment success and younger age, as defined in previous research

about dropout risk factors [9]. On the other hand, it does confirm the relationship between treatment success and certain disorders of cognitive function [9], in this case orientation, thought, and judgment capacity, and the effect of disulfiram in the treatment of cocaine-addicted patients [63].

Compared to the work of Deane *et al.* [7] our results are superior in accuracy (82.12% vs 61.6%), both with the logistic regression and with the other models. The most significant predictor variables in that study were the primary drug of concern and forgiveness of self. The first variable cannot be compared with our study since our subjects are all cocaine users, and the second was not among the data available in the referral reports. Even considering the difference in population size employed, which are populations from different regions, and Deane's multi-substance approach, the performance achieved by our machine learning techniques and the set of variables employed is an improvement in the predictive capacity of treatment success.

In the case of Anh *et al.* [33], they attempted to classify US subjects with cocaine dependence and healthy subjects using data on self-reported impulsivity and from neurocognitive tasks. On the one hand, our work tries to classify only individuals with cocaine dependence, so the comparison between the AUC of both works would not be logical or fair. On the other hand, our study does not include data on direct impulsivity, but related variables such as 'Judgment capacity' or 'Thought' are found in the  $CS_{WBF}$  used with the RF algorithm.

The work of Acion *et al.* [34] is the most similar to the one presented in this article since it is also based on the comparison of different ML algorithms to predict success in outpatient treatment. Acion *et al.* focused on Hispanic Americans with addictions to different substances, used a larger database than ours (99,013), and a smaller number of variables in the initial dataset (28). Despite this, the RF algorithm also showed a great performance, positioning itself as the second option in terms of AUC (0.816), and presenting practically identical values to the best model obtained with Super Learning (0.82). This could imply that the RF algorithm fits better than other algorithms to heterogeneous data with complex interactions such as those that characterize drug dependence within addiction psychiatry. In the study by Acion *et al.*, the length of stay in treatment was also established as one of the most important predictors.

Few studies have considered the variables of the treatments themselves [9] and to our knowledge none have considered the characteristics of the previous treatments together with mental, cognitive, personality, and pharmacological variables in cocaine addiction. We found that among the most important characteristics within the RF predictive model were those related to the patient's previous treatment pathway. This indicates that their previous behavior in similar treatments related to cocaine use can be of great help when selecting the next step in their treatment. Some variables of mental exploration, cognitive functioning, and personality have contributed in a similar magnitude to the predictive result achieved, as well as

consumption habits and drugs to a lesser extent. Individually these variables do not have the capacity to provide the same results as when considered together.

The power of the tested classifier is highly dependent upon the population characteristic set used for its creation. To our knowledge, the size of our subject sample is the largest that has been used in this context in the specific case of European subjects with cocaine use disorders. Although it would be very beneficial for the generalization of the data obtained to be able to use larger sample sizes than those presented in this article, due to the inherent high complexity of addiction psychiatry, data of these dimensions are not easily accessible. In fact, this can also be seen in the related works mentioned above, where in many cases they use samples much smaller than ours. This is also related to the growing need for development in the field of psychiatry and specifically in addiction psychiatry, as these are fields of medicine that are not yet as developed as other specialties and still have a great deal of room for improvement. In any case, whenever possible it would be very useful to conduct further research with a different and larger population than that used in this article to determine whether the current results can be generalized.

Longer length of stay in residential drug treatment has been previously associated with more favorable outcomes, but optimal durations are likely to be a function of patient and problem characteristics [64]. Some patients drop out in advanced stages of treatment and represent a great loss due to the health resources consumed. According to the evidence, after 37.37 days of treatment there is a reliable change in psychological recovery and well-being [64] and, in the case of late dropouts, shorter residential treatment with longer aftercare may be beneficial [65], both in terms of cost and clinical outcomes [66]. On the other hand, late dropouts could also be subjected to a pre-treatment phase to improve the stage of change, which has been shown to be a good predictor of treatment completion [67]. For these reasons it would also be of great interest for future research to explore whether predictive models would be able to differentiate between early and late dropouts.

A further step in our research will be to develop a decision support system to assist clinical professionals in charge of referring patients to cocaine addiction inpatient treatment in the next phase of their treatment pathway. The integration of the implemented predictor with a technological platform would facilitate its use by clinical professionals and would allow the final performance of the classifier to be more easily assessed.

## VI. CONCLUSION

The use of ML algorithms together with feature selection techniques on a set of heterogeneous variables of different dimensions has shown promising results for the prediction of therapeutic success in cocaine addicted patients. The best performance was achieved by the Random Forest algorithm with an accuracy of 82.12%, an excellent AUC of 0.81, a recall measure of 0.96, an F1-measure of 0.89 and an MCC

of 0.53 on average. As seen in previous research works, this algorithm has again demonstrated a great capacity for adaptation to datasets with complex relationships, an inherent condition in the field of psychiatry and drug addiction.

Interactions between variables of different dimensions have proved to enhance the predictor's performance, establishing synergies between characteristics of the previous treatments together with mental, cognitive, personality, and pharmacological variables. This article has confirmed the relationship between treatment success and certain variables also established in the literature, such as cognitive function disorders, personality disorders, length of stay in previous treatments and the effect of disulfiram in the treatment of patients addicted to cocaine.

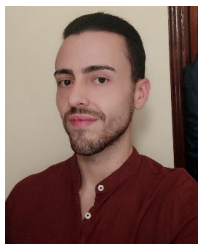
Reducing treatment drop-out rates for people with drug addiction problems would mean better management and use of medical and care resources, improving opportunities for social reintegration and involving major economic and social benefits. With the increasing use of ML methods in addiction psychiatry for informing medical decisions, the use of these techniques to predict treatment success could represent a new step in treatment management especially when included in clinical decision support systems.

## REFERENCES

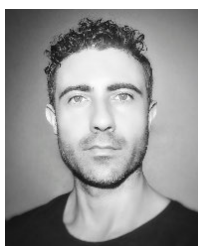
- [1] European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). (Jun. 2019) *European Drug Report 2019: Trends and Developments*. Publications Office of the European Union, Luxembourg. Accessed: Jul. 23, 2020. [Online]. Available: [https://www.emcdda.europa.eu/publications/edr/trends-developments/2019\\_en](https://www.emcdda.europa.eu/publications/edr/trends-developments/2019_en)
- [2] European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). (2019). *Spain Country Drug Report 2019*. Government Delegation for the National Plan on Drugs, Ministry of Health, Social Policy and Equality, Spain. Accessed: Jul. 23, 2020. [Online]. Available: <https://www.emcdda.europa.eu/system/files/publications/11353/spain-cdr-2019.pdf>
- [3] Substance Abuse and Mental Health Services Administration. (2019). *Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health*. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration, Rockville, MD, USA, HHS Publication No. PEP19-5068, NSDUH Series H-54. Accessed: Jul. 23, 2020. [Online]. Available: <https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.pdf>
- [4] World Health Organization and United Nations Office on Drugs and Crime. (2020). *International Standards for the Treatment of Drug Use Disorders*. Revised Edition Incorporating Results of Field-Testing. License: CC BY-NC-SA 3.0 IGO. Accessed: Jul. 23, 2020. [Online]. Available: <https://www.who.int/publications/i/item/international-standards-for-the-treatment-of-drug-use-disorders>
- [5] G. A. Dakof, M. Tejada, and H. A. Liddle, "Predictors of engagement in adolescent drug abuse treatment," *J. Amer. Acad. Child Adolescent Psychiatry*, vol. 40, no. 3, pp. 274–281, Mar. 2001, doi: [10.1097/00004583-200103000-00006](https://doi.org/10.1097/00004583-200103000-00006).
- [6] R. K. McHugh, H. W. Murray, B. A. Hearon, E. M. Pratt, M. H. Pollack, S. A. Safren, and M. W. Otto, "Predictors of dropout from psychosocial treatment in opioid-dependent outpatients," *Amer. J. Addictions*, vol. 22, no. 1, pp. 18–22, Jan. 2013, doi: [10.1111/j.1521-0391.2013.00317.x](https://doi.org/10.1111/j.1521-0391.2013.00317.x).
- [7] F. P. Deane, D. J. Wootton, C.-I. Hsu, and P. J. Kelly, "Predicting dropout in the first 3 months of 12-step residential drug and alcohol treatment in an Australian sample," *J. Stud. Alcohol Drugs*, vol. 73, no. 2, pp. 216–225, Mar. 2012, doi: [10.15288/jsad.2012.73.216](https://doi.org/10.15288/jsad.2012.73.216).
- [8] D. B. Samuel, D. M. LaPaglia, L. M. Maccarelli, B. A. Moore, and S. A. Ball, "Personality disorders and retention in a therapeutic community for substance dependence," *Amer. J. Addictions*, vol. 20, no. 6, pp. 555–562, Nov. 2011, doi: [10.1111/j.1521-0391.2011.00174.x](https://doi.org/10.1111/j.1521-0391.2011.00174.x).
- [9] H. H. Brorson, E. Ajo Arnevik, K. Rand-Hendriksen, and F. Duckert, "Drop-out from addiction treatment: A systematic review of risk factors," *Clin. Psychol. Rev.*, vol. 33, no. 8, pp. 1010–1024, Dec. 2013, doi: [10.1016/j.cpr.2013.07.007](https://doi.org/10.1016/j.cpr.2013.07.007).
- [10] H. D. Kleber et al., "Treatment of patients with substance use disorders, American psychiatric association," *Amer. J. Psychiatry*, vol. 164, no. 4, pp. 123–125, Apr. 2007.
- [11] M. J. Stark, "Dropping out of substance abuse treatment: A clinically oriented review," *Clin. Psychol. Rev.*, vol. 12, no. 1, pp. 93–116, Jan. 1992, doi: [10.1016/0272-7358\(92\)90092-M](https://doi.org/10.1016/0272-7358(92)90092-M).
- [12] G. A. Zarkin, L. J. Dunlap, J. W. Bray, and W. M. Wechsberg, "The effect of treatment completion and length of stay on employment and crime in outpatient drug-free treatment," *J. Subst. Abuse Treat.*, vol. 23, no. 4, pp. 261–271, Dec. 2002, doi: [10.1016/S0740-5472\(02\)00273-8](https://doi.org/10.1016/S0740-5472(02)00273-8).
- [13] A. M. Arria, "Drug treatment completion and post-discharge employment in the TOPPS-II interstate cooperative study," *J. Subst. Abuse Treat.*, vol. 25, no. 1, pp. 9–18, Jul. 2003, doi: [10.1016/S0740-5472\(03\)00050-3](https://doi.org/10.1016/S0740-5472(03)00050-3).
- [14] A. I. Alterman, J. R. McKay, F. D. Mulvaney, and A. T. McLellan, "Prediction of attrition from day hospital treatment in lower socioeconomic cocaine-dependent men," *Drug Alcohol Depend.*, vol. 40, no. 3, pp. 227–233, Mar. 1996, doi: [10.1016/0376-8716\(95\)01212-5](https://doi.org/10.1016/0376-8716(95)01212-5).
- [15] D. D. Simpson, G. W. Joe, G. A. Rowan-Szal, and J. M. Greener, "Drug abuse treatment process components that improve retention," *J. Subst. Abuse Treat.*, vol. 14, no. 6, pp. 565–572, Nov. 1997, doi: [10.1016/S0740-5472\(97\)00181-5](https://doi.org/10.1016/S0740-5472(97)00181-5).
- [16] E. Aharonovich, D. S. Hasin, A. C. Brooks, X. Liu, A. Bisaga, and E. V. Nunes, "Cognitive deficits predict low treatment retention in cocaine dependent patients," *Drug Alcohol Dependence*, vol. 81, no. 3, pp. 313–322, Feb. 2006, doi: [10.1016/j.drugaldep.2005.08.003](https://doi.org/10.1016/j.drugaldep.2005.08.003).
- [17] E. Aharonovich, A. C. Brooks, E. V. Nunes, and D. S. Hasin, "Cognitive deficits in marijuana users: Effects on motivational enhancement therapy plus cognitive Behavioral therapy treatment outcome," *Drug Alcohol Dependence*, vol. 95, no. 3, pp. 279–283, Jun. 2008, doi: [10.1016/j.drugaldep.2008.01.009](https://doi.org/10.1016/j.drugaldep.2008.01.009).
- [18] G. Teichner, M. D. Horner, J. C. Roitzsch, J. Herron, and A. Thevos, "Substance abuse treatment outcomes for cognitively impaired and intact outpatients," *Addict. Behav.*, vol. 27, no. 5, pp. 751–763, Sep. 2002, doi: [10.1016/S0306-4603\(01\)00207-6](https://doi.org/10.1016/S0306-4603(01)00207-6).
- [19] A. B. Morgan and S. O. Lilienfeld, "A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function," *Clin. Psychol. Rev.*, vol. 20, no. 1, pp. 113–136, Jan. 2000, doi: [10.1016/S0272-7358\(98\)00096-8](https://doi.org/10.1016/S0272-7358(98)00096-8).
- [20] D. P. MacKinnon and L. J. Luecken, "How and for whom? Mediation and moderation in health psychology," *Health Psychol.*, vol. 27, no. 2, pp. S99–S100, Mar. 2008, doi: [10.1037/0278-6133.27.2\(Suppl.\).S99](https://doi.org/10.1037/0278-6133.27.2(Suppl.).S99).
- [21] R. C. Deo, "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, Nov. 2015, doi: [10.1161/CIRCULATIONAHA.115.001593](https://doi.org/10.1161/CIRCULATIONAHA.115.001593).
- [22] J. Corny, A. Rajkumar, O. Martin, X. Dode, J.-P. Lajonchère, O. Billuart, Y. Bézie, and A. Buronfosse, "A machine learning-based clinical decision support system to identify prescriptions with a high risk of medication error," *J. Amer. Med. Inform. Assoc.*, vol. 27, no. 11, pp. 1688–1694, Nov. 2020, doi: [10.1093/jamia/ocaa154](https://doi.org/10.1093/jamia/ocaa154).
- [23] Q.-X. Feng, C. Liu, L. Qi, S.-W. Sun, Y. Song, G. Yang, Y.-D. Zhang, and X.-S. Liu, "An intelligent clinical decision support system for preoperative prediction of lymph node metastasis in gastric cancer," *J. Amer. College Radiol.*, vol. 16, no. 7, pp. 952–960, Jul. 2019, doi: [10.1016/j.jacr.2018.12.017](https://doi.org/10.1016/j.jacr.2018.12.017).
- [24] B. Wingfield, S. Miller, P. Yogarajah, D. Kerr, B. Gardiner, S. Seneviratne, P. Samarasinghe, and S. Coleman, "A predictive model for paediatric autism screening," *Health Informat. J.*, Mar. 2020, Art. no. 146045821988782, doi: [10.1177/1460458219887823](https://doi.org/10.1177/1460458219887823).
- [25] D. Velusamy and K. Ramasamy, "Ensemble of heterogeneous classifiers for diagnosis and prediction of coronary artery disease with reduced feature subset," *Comput. Methods Programs Biomed.*, vol. 198, Jan. 2021, Art. no. 105770, doi: [10.1016/j.cmpb.2020.105770](https://doi.org/10.1016/j.cmpb.2020.105770).
- [26] T. Shabaniyan, H. Parsaei, A. Aminsharifi, M. M. Movahedi, A. T. Jahromi, S. Pouyesh, and H. Parvin, "An artificial intelligence-based clinical decision support system for large kidney stone treatment," *Australas. Phys. Eng. Sci. Med.*, vol. 42, no. 3, pp. 771–779, Sep. 2019, doi: [10.1007/s13246-019-00780-3](https://doi.org/10.1007/s13246-019-00780-3).



- [27] L. Ali and S. A. C. Bukhari, "An approach based on mutually informed neural networks to optimize the generalization capabilities of decision support systems developed for heart failure prediction," *IRBM*, May 2020, doi: [10.1016/j.irbm.2020.04.003](https://doi.org/10.1016/j.irbm.2020.04.003).
- [28] D. B. Dwyer, P. Falkai, and N. Koutsouleris, "Machine learning approaches for clinical psychology and psychiatry," *Annu. Rev. Clin. Psychol.*, vol. 14, no. 1, pp. 91–118, May 2018, doi: [10.1146/annurev-clinpsy-032816-045037](https://doi.org/10.1146/annurev-clinpsy-032816-045037).
- [29] D. Bzdok and A. Meyer-Lindenberg, "Machine learning for precision psychiatry: Opportunities and challenges," *Biol. Psychiatry, Cognit. Neurosci. Neuroimag.*, vol. 3, no. 3, pp. 223–230, Mar. 2018, doi: [10.1016/j.bpsc.2017.11.007](https://doi.org/10.1016/j.bpsc.2017.11.007).
- [30] N. Yahata, K. Kasai, and M. Kawato, "Computational neuroscience approach to biomarkers and treatments for mental disorders," *Psychiatry Clin. Neurosci.*, vol. 71, no. 4, pp. 215–237, Apr. 2017, doi: [10.1111/pcn.12502](https://doi.org/10.1111/pcn.12502).
- [31] K. K. Mak, K. Lee, and C. Park, "Applications of machine learning in addiction studies: A systematic review," *Psychiatry Res.*, vol. 275, pp. 53–60, May 2019, doi: [10.1016/j.psychres.2019.03.001](https://doi.org/10.1016/j.psychres.2019.03.001).
- [32] M. Symons, G. F. X. Feeney, M. R. Gallagher, R. M. Young, and J. P. Connor, "Predicting alcohol dependence treatment outcomes: A prospective comparative study of clinical psychologists versus 'trained' machine learning models," *Addiction*, vol. 115, no. 11, pp. 2164–2175, Mar. 2020, doi: [10.1111/add.15038](https://doi.org/10.1111/add.15038).
- [33] W.-Y. Ahn, D. Ramesh, F. G. Moeller, and J. Vassileva, "Utility of machine-learning approaches to identify behavioral markers for substance use disorders: Impulsivity dimensions as predictors of current cocaine dependence," *Frontiers Psychiatry*, vol. 7, p. 34, Mar. 2016, doi: [10.3389/fpsy.2016.00034](https://doi.org/10.3389/fpsy.2016.00034).
- [34] L. Acion, D. Kelmansky, M. van der Laan, E. Sahker, D. Jones, and S. Arndt, "Use of a machine learning framework to predict substance use disorder treatment success," *PLoS ONE*, vol. 12, no. 4, Apr. 2017, Art. no. e0175383, doi: [10.1371/journal.pone.0175383](https://doi.org/10.1371/journal.pone.0175383).
- [35] M. Mete, U. Sakoglu, J. S. Spence, M. D. Devous, T. S. Harris, and B. Adinoff, "Successful classification of cocaine dependence using brain imaging: A generalizable machine learning approach," *BMC Bioinf.*, vol. 17, no. S13, p. 357, Oct. 2016, doi: [10.1186/s12859-016-1218-z](https://doi.org/10.1186/s12859-016-1218-z).
- [36] U. Sakoglu, M. Mete, J. Esquível, K. Rubia, R. Briggs, and B. Adinoff, "Classification of cocaine-dependent participants with dynamic functional connectivity from functional magnetic resonance imaging data," *J. Neurosci. Res.*, vol. 97, no. 7, pp. 790–803, Jul. 2019, doi: [10.1002/jnr.24421](https://doi.org/10.1002/jnr.24421).
- [37] I. Rish, P. Bashivan, G. A. Cecchi, and R. Z. Goldstein, "Evaluating effects of methylphenidate on brain activity in cocaine addiction: A machine-learning approach," *Proc. SPIE*, vol. 9788, Mar. 2016, Art. no. 978800, doi: [10.1117/12.2218212](https://doi.org/10.1117/12.2218212).
- [38] S. W. Yip, D. Scheinost, M. N. Potenza, and K. M. Carroll, "Connectome-based prediction of cocaine abstinence," *Amer. J. Psychiatry*, vol. 176, no. 2, pp. 156–164, Feb. 2019, doi: [10.1176/appi.ajp.2018.17101147](https://doi.org/10.1176/appi.ajp.2018.17101147).
- [39] L. V. Panlilio, S. W. Stull, J. W. Bertz, A. J. Burgess-Hull, W. J. Kowalczyk, K. A. Phillips, D. H. Epstein, and K. L. Preston, "Beyond abstinence and relapse: Cluster analysis of drug-use patterns during treatment as an outcome measure for clinical trials," *Psychopharmacology*, vol. 237, no. 11, pp. 3369–3381, Nov. 2020, doi: [10.1007/s00213-020-05618-5](https://doi.org/10.1007/s00213-020-05618-5).
- [40] M. R. Arbabshirani, S. Plis, J. Sui, and V. D. Calhoun, "Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls," *NeuroImage*, vol. 145, pp. 137–165, Jan. 2017, doi: [10.1016/j.neuroimage.2016.02.079](https://doi.org/10.1016/j.neuroimage.2016.02.079).
- [41] J. Kambeitz, L. Kambeitz-Ilanovic, S. Leucht, S. Wood, C. Davatzikos, B. Malchow, P. Falkai, and N. Koutsouleris, "Detecting neuroimaging biomarkers for schizophrenia: A meta-analysis of multivariate pattern recognition studies," *Neuropsychopharmacology*, vol. 40, no. 7, pp. 1742–1751, Jun. 2015, doi: [10.1038/npp.2015.22](https://doi.org/10.1038/npp.2015.22).
- [42] G. Orrù, U. Pettersson-Yeo, A. F. Marquand, G. Sartori, and A. Mechelli, "Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review," *Neurosci. Biobehav. Rev.*, vol. 36, no. 4, pp. 1140–1152, Apr. 2012, doi: [10.1016/j.neubiorev.2012.01.004](https://doi.org/10.1016/j.neubiorev.2012.01.004).
- [43] R. A. Fisher, "The statistical utilization of multiple measurements," *Ann. Eugenics*, vol. 8, no. 4, pp. 376–386, Aug. 1938, doi: [10.1111/j.1469-1809.1938.tb02189.x](https://doi.org/10.1111/j.1469-1809.1938.tb02189.x).
- [44] C. M. Bishop, "Pattern recognition and machine learning," in *Information Science and Statistics*. New York, NY, USA: Springer, 2006, p. 738.
- [45] A. P. Association, *Diagnostic and Statistical Manual of Medical Disorders*, 5th ed. Washington, DC, USA: A. P. Association, 2013.
- [46] M. Startup, M. C. Jackson, and S. Bendix, "The concurrent validity of the global assessment of functioning (GAF)," *Brit. J. Clin. Psychol.*, vol. 41, no. 4, pp. 417–422, Nov. 2002, doi: [10.1348/014466502760387533](https://doi.org/10.1348/014466502760387533).
- [47] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Amsterdam, The Netherlands: Elsevier, 2017.
- [48] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- [49] I. Kurt, M. Ture, and A. T. Kurum, "Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease," *Expert Syst. Appl.*, vol. 34, no. 1, pp. 366–374, Jan. 2008, doi: [10.1016/j.eswa.2006.09.004](https://doi.org/10.1016/j.eswa.2006.09.004).
- [50] A. Subasi and E. Erçelebi, "Classification of EEG signals using neural network and logistic regression," *Comput. Methods Programs Biomed.*, vol. 78, no. 2, pp. 87–99, May 2005, doi: [10.1016/j.cmpb.2004.10.009](https://doi.org/10.1016/j.cmpb.2004.10.009).
- [51] S. Haykin and S. S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York, NY, USA: Macmillan, 1994.
- [52] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: [10.1007/bf00994018](https://doi.org/10.1007/bf00994018).
- [53] J. C. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," Microsoft Res., Redmond, WA, USA, Tech. Rep. MSR-TR-98-14, Apr. 1998. Accessed: Jul. 23, 2020. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/sequential-minimal-optimization-a-fast-algorithm-for-training-support-vector-machines/>
- [54] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, Dept. Comput. Sci., Waikato Univ., Hamilton, New Zealand, 1999.
- [55] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, nos. 1–2, pp. 273–324, 1997, doi: [10.1016/S0004-3702\(97\)00043-X](https://doi.org/10.1016/S0004-3702(97)00043-X).
- [56] P. Bermejo, J. A. Gamez, and J. M. Puerta, "Improving incremental wrapper-based subset selection via replacement and early stopping," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 25, no. 05, pp. 605–625, Aug. 2011, doi: [10.1142/S0218001411008804](https://doi.org/10.1142/S0218001411008804).
- [57] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, Apr. 2005, doi: [10.1109/TKDE.2005.66](https://doi.org/10.1109/TKDE.2005.66).
- [58] D. Goldberg, K. Deb, and B. Korb, "Messy genetic algorithms: Motivation, analysis, and first results," *Complex Syst.*, vol. 3, no. 5, pp. 493–530, 1989.
- [59] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 3, pp. 299–310, Mar. 2005, doi: [10.1109/TKDE.2005.50](https://doi.org/10.1109/TKDE.2005.50).
- [60] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.
- [61] J. N. Mandrekar, "Receiver operating characteristic curve in diagnostic test assessment," *J. Thoracic Oncol.*, vol. 5, no. 9, pp. 1315–1316, Sep. 2010, doi: [10.1097/JTO.0b013e3181ec173d](https://doi.org/10.1097/JTO.0b013e3181ec173d).
- [62] W. S. Cartwright, "Economic costs of drug abuse: Financial, cost of illness, and services," *J. Substance Abuse Treatment*, vol. 34, no. 2, pp. 224–233, Mar. 2008, doi: [10.1016/j.jsat.2007.04.003](https://doi.org/10.1016/j.jsat.2007.04.003).
- [63] A. De Sousa and A. De Sousa, "Disulfiram in the management of cocaine dependence and other psychiatric disorders," in *Disulfiram*, 1st ed. Singapore: Springer, 2019, pp. 31–42, doi: [10.1007/978-981-32-9876-7](https://doi.org/10.1007/978-981-32-9876-7).
- [64] B. Turner and F. P. Deane, "Length of stay as a predictor of reliable change in psychological recovery and well being following residential substance abuse treatment," *Therapeutic Communities, Int. J. Therapeutic Communities*, vol. 37, no. 3, pp. 112–120, Sep. 2016, doi: [10.1108/TC-09-2015-0022](https://doi.org/10.1108/TC-09-2015-0022).
- [65] D. E. Baker, K. A. Edmonds, M. L. Calvert, S. M. Sanders, A. J. Bridges, M. A. Rhea, and S. Kosloff, "Predicting attrition in long-term residential substance use disorder treatment: A modifiable risk factors perspective," *Psychol. Services*, vol. 17, no. 4, pp. 472–482, Nov. 2020, doi: [10.1037/ser0000333](https://doi.org/10.1037/ser0000333).
- [66] K. A. Peterson, R. W. Swindle, C. S. Phibbs, B. Recine, and R. H. Moos, "Determinants of readmission following inpatient substance abuse treatment: A national study of VA programs," *Med. Care*, vol. 32, no. 6, pp. 535–550, Jun. 1994.
- [67] B. Pellissier, "Gender differences in substance use treatment entry and retention among prisoners with substance use histories," *Amer. J. Public Health*, vol. 94, no. 8, pp. 1418–1424, Aug. 2004, doi: [10.2105/AJPH.94.8.1418](https://doi.org/10.2105/AJPH.94.8.1418).



**JOSE TAPIA-GALISTEO** received the M.Sc. degree in telecommunication engineering and the M.Sc. degree in biomedical engineering from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2015 and 2016, respectively, where he is currently pursuing the Ph.D. degree in biomedical engineering. His research interests include machine learning and development of predictive models applied to health, intelligent data analysis, decision support systems in telemedicine services, and development of e-health and m-health tools for monitoring, prevention, and treatment of diseases.



**JOSE M. INIESTA** received the M.Sc. degree in bioengineering and telemedicine from the Universidad Politécnica de Madrid (UPM), Spain, and the degree in electronics engineering from the Universidad Tecnológica Nacional (UTN), Argentina. He has been a Researcher with the Biomedical Engineering and Telemedicine Centre, GBT, UPM, since 2008. His research interests include the design and development of e-health systems and the application of m-health technologies for

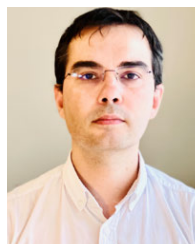
the support and promotion of healthy lifestyle habits and treatment of chronic diseases. He is a member of the Spanish Society of Biomedical Engineering (SEIB).



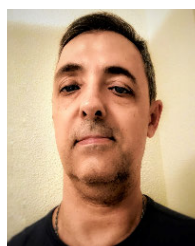
**CARMEN PÉREZ-GANDÍA** received the Ph.D. degree. She is currently a Telecommunication Engineer with the Universidad Politécnica de Madrid (UPM), Spain. She is also a Senior Researcher in intelligent data analysis, machine learning, and prediction model design, applied to cardiovascular diseases and mental health.



**GEMA GARCÍA-SÁEZ** received the Ph.D. degree in biomedical engineering. She is currently a Telecommunication Engineer and a Senior Researcher in biomedical engineering. Her research interests include design and evaluation of decision support systems for clinical care, e-health and m-health solutions for chronic patients' care, and development of monitoring scenarios.



**DIEGO URGELÉS PUÉRTOLAS** received the M.D. degree from the University of Zaragoza. He completed an Internship and a Residency in psychiatry at the Hospital de La Princesa, Madrid, Spain. He has several years of experience in treating addictions and affective disorders. He is currently responsible for the Bipolar Disorder Unit and the Head of the research at the Clinic Ntra. Sra. de La Paz, Madrid. His current research interests include bipolar disorders, addictions, and predictive medicine.



**FRANCISCO J. IZQUIERDO** received the B.Sc. degree in civil engineering from the Universidad Politécnica de Madrid. Since 2007, he has been working as a Technical Professional for IBM, involved in projects in the area of natural language processing. In 2010, he has been teaching computational linguistics in post-graduate courses in different Spanish Universities.



**M. ELENA HERNANDO** (Senior Member, IEEE) received the Ph.D. degree in Spain. She is currently a Telecommunication Engineer in Spain. She is also a Professor of biomedical engineering topics. Her main research interests include telemedicine services and the definition of decision support tools for patients and healthcare professionals.

...