Predicting involuntary hospitalisation in psychiatry: a machine learning investigation

Benedetta Silva\textsuperscript{ab}, Mehdi Gholam\textsuperscript{ce}, Philippe Golay\textsuperscript{ad}, Charles Bonsack\textsuperscript{a} & Stéphane Morandi\textsuperscript{ab}

a. Community Psychiatry Service, Department of Psychiatry, Lausanne University Hospital and University of Lausanne, Switzerland.

b. Cantonal Medical Office, General Directorate for Health of Canton of Vaud, Department of Health and Social Action (DSAS), Avenue des Casernes 2, 1014 Lausanne, Switzerland.

c. Epidemiology and Psychopathology Research Unit, Department of Psychiatry, Lausanne University Hospital and University of Lausanne, Switzerland.

d. General Psychiatry Service, Department of Psychiatry, Lausanne University Hospital and University of Lausanne, Switzerland.

e. Ecole polytechnique fédérale de Lausanne EPFL, School of Basic Sciences, Institute of Mathematics, Lausanne, Switzerland.

Corresponding Author: Benedetta Silva, Department of Psychiatry, Lausanne University Hospital, Consultations de Chauderon, Place Chauderon 18, 1003 Lausanne, Switzerland. ORCID iD: 0000-0002-8896-8467 Email: benedetta.silva@chuv.ch

Short Title: Involuntary admission and machine learning

This peer-reviewed article has been accepted for publication but not yet copyedited or typeset, and so may be subject to change during the production process. The article is considered published and may be cited using its DOI.

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.
Abstract

Background: Coercion in psychiatry is a controversial issue. Identifying its predictors and their interaction using traditional statistical methods is difficult, given the large number of variables involved. The purpose of this study was to use machine-learning models to identify socio-demographic, clinical and procedural characteristics that predict the use of compulsory admission on a large sample of psychiatric patients.

Methods: We retrospectively analysed the routinely collected data of all psychiatric admissions that occurred between 2013 and 2017 in the canton of Vaud, Switzerland (N=25584). The main predictors of involuntary hospitalisations were identified using two machine-learning algorithms: Classification and Regression Tree (CART) and Random Forests (RFs). Their predictive power was compared with that obtained through traditional logistic regression. Sensitivity analyses were also performed and missing data were imputed through Multiple Imputation using Chain Equations.

Results: The three models achieved similar predictive balanced accuracy, ranging between 68% and 72%. CART showed the lowest predictive power (68%) but the most parsimonious model, allowing to estimate the probability of being involuntarily admitted with only three checks: aggressive behaviours, who referred the patient to hospital and primary diagnosis. The results of CART and RFs on the imputed data were almost identical to those obtained on the original data, confirming the robustness of our models.

Conclusions: Identifying predictors of coercion is essential to efficiently target the development of professional training, preventive strategies and alternative interventions. Machine-learning methodologies could offer new effective tools to achieve this goal, providing accurate but simple models that could be used in clinical practice.

Keywords: Coercion; involuntary hospitalisation; predicting factor; machine learning.
1. INTRODUCTION

Coercion in psychiatry is a widely discussed and controversial issue. In fact, although coercive measures are justified by the need to preserve the patient’s health and safety and to protect others from their potentially harmful behaviours [1], their use is in strong opposition to the principle of patient’s autonomy and self-determination also defended by the UN Convention on the Rights of Persons with Disabilities [2].

Furthermore, despite the increasing political and ethical attention to patients’ human rights and the scarce evidence of coercion’s benefits [3], its use is increasing almost everywhere [4-6], with great variations among countries [7, 8]. A recent study compared the annual incidence of involuntary hospitalisation for 22 countries across Europe, Australia and New Zealand between 2008 and 2017 showing that time trends in annual rates raised in 11 out of 18 countries [6]. Moreover, Rains et al. (2019) found that incidences varied strikingly, ranging from 282 per 100’000 inhabitants in Austria to 14.5 per 100’000 inhabitants in Italy. Higher annual incidence rates were associated with lower absolute poverty, higher gross domestic product and per capita health care expenditure, larger proportion of foreign-born individuals and higher number of psychiatric beds [6].

In previous studies, other area- and system-related characteristics have been associated with an increased risk of involuntary hospitalisation. Rates of compulsory admission were higher in socio-economically deprived and densely populated urban areas, with higher proportions of young adults and ethnic minorities [9-12]. In addition, several aspects of the referral process, such as being referred outside regular service hours [13, 14], by a general hospital [13, 15], a general practitioner [15, 16], or someone who does not know the patient [17], and having a contact with the police at the time of admission [17], were identified as predictors of involuntary hospitalisation.

However, the most robust results point to the patient’s clinical profile as key determinant. People affected by schizophrenic, bipolar or organic mental disorders, with poor insight, low treatment adherence, high level of psychotic symptoms and aggressiveness, and a history of previous involuntary treatment were more likely to be detained [3, 11, 14-35]. On the contrary, affective disorders, anxiety disorders and substance-related
Accepted manuscript: Authors' Copy

disorders were associated with a lower risk of coercion [13, 15, 19, 27, 29, 30, 32]. Recently, a meta-analysis confirmed that previous involuntary hospitalisation and diagnosis of psychotic disorders were associated with the greatest risk of involuntary psychiatric admission [36].

Among socio-demographic characteristics, being a young man, homeless, unemployed, with a migration background and a low educational level were identified as being associated with involuntary admission [3, 9, 11, 13, 14, 18, 19, 22, 26, 28, 30, 31, 33]. However, these findings are less consistent, with several studies showing the opposite [16, 25, 37] or finding no association [15, 20, 27, 32, 38].

Most of the above mentioned studies aimed at identifying predictors of involuntary hospitalisation using traditional statistical methods, mainly regression models [11, 15-18, 20, 21, 24-27, 29-31, 38], which made the statistical interaction between risk factors difficult to model and interpret [13, 39, 40]. Machine learning (ML) may overcome this limitation. ML algorithms are “trained” to find patterns in large amounts of data in order to make predictions based on new data and reveal previously “unseen” non-linear interactions between variables [41, 42]. A recent German study used a decision-tree-generating algorithm (CHAID Chi-Square Automatic Interaction Detector), to detect risk factors for involuntary admission on a weighted sample of 10'171 inpatients [13]. Beside main diagnosis, this study identified several modifiable predictors of involuntary hospitalisation, such as the absence of outpatient treatment prior to admission or admission outside of regular service hours. Consequently, several preventive measures were suggested to address these factors and reduce the rate of involuntary inpatient treatment [13]. In 2020, Karash et al. improved this first predictive decision tree model by optimizing machine learning techniques and broadening the data set to include environmental socioeconomic data [43]. Three main subgroups of patients were identified as at highest risk of involuntary admission: 1. patients with an organic mental disorder, retired, admitted outside of regular service hours and living in assisted housing; 2. patients with suicidal tendencies but no affective disorder, with unclear previous suicide attempts, or living in areas with high unemployment rates; 3. Psychotic patients, living in densely built areas with a large proportion of small or one-person households [43]. ML algorithms were also employed by Günther et al. (2020) to predict the use of coercive measures (seclusion, restraint, involuntary medication) in a sample of 370 forensic offender patients with schizophrenia [41]. Overall, ten factors out of a set of 569 potential variables were identified as best
predictors of coercion with a balanced accuracy of 73.3%. Previously, another Swiss study on 393 hospitalised patients had also shown that machine learning was useful in predicting coercion, reaching results comparable to binary logistic regression but with better generalizability [44]. Further studies were recommended by the authors to evaluate the potential of these methods on larger samples.

The purpose of this study was to use machine-learning models to identify which characteristics predicted the use of compulsory admission on a large sample of psychiatric patients (N=25584).

2. METHODS

2.1 Study setting

This study was set in the canton of Vaud (794’384 inhabitants in 2017), a mixed urban-rural region of Western Switzerland. The canton is divided into four psychiatric district, each one served by a psychiatric hospital. In 2015, the four psychiatric hospitals provided a total of 0.6 psychiatric beds per 1′000 inhabitants [15], below the total number for Switzerland (0.9/1’000 inhabitants) and for Europe (0.7/1’000 inhabitants) [45, 46]. On the contrary, according to data from the Swiss Health Observatory, in 2016 the canton of Vaud registered the highest rate of involuntary admissions in Switzerland (2.3 per 1’000 inhabitants), well above the national average (1.6 per 1’000 inhabitants) [47, 48] and among the highest in Europe [6].

Involuntary admissions in Switzerland are regulated at federal level by the Article 426 of the Swiss Civil Code (CC) which states that “a person suffering from a mental disorder or mental disability or serious neglect (the patient) may be committed to an appropriate institution if the required treatment or care cannot be provided otherwise.”.

Therefore, need of treatment is the legal requirement for involuntary admission. Lack of capacity to consent to treatment or the fact of posing an imminent danger to oneself or others are not required to decide on an involuntary hospitalisation. However, danger to oneself or others is the prerequisite to detain a person suffering from a mental disorder who has voluntarily entered an institution and wishes to leave (Art. 427 CC). In this particular case, detention may last a maximum of three days, at the end of which the patient may leave the institution, unless a compulsory admission is ordered.
The federal law is then executed at cantonal level, with important regional differences. In the canton of Vaud, only the Adult Protection Authority (APA) and medical doctors designated by the Department of Health and Social Action (DSAS), namely general practitioners, psychiatrists, paediatricians and on-call doctors, are allowed to order an involuntary admission [49]. APA orders have no time limit, but a mandatory revaluation after 6 months, 12 months and then every year. Instead, involuntary admissions ordered by medical doctors cannot exceed 6 weeks, unless the APA issues an extension order.

2.2 Study design

This study retrospectively analysed the routinely collected data of all psychiatric admissions that occurred between the 1st January 2013 and the 31st December 2017 in the four psychiatric hospitals of the canton of Vaud, Switzerland. Available data included socio-demographic characteristics, like gender, age, marital status and nationality. Several clinical characteristics were also available, as primary diagnosis, aggregated into ten main categories based on the ICD-10 system, a secondary diagnosis of addiction and/or personality disorders, and the 12 item-level scores of the Health of the Nations Outcome Scale (HoNOS) [50, 51], which measure the patients’ mental and social functioning. Within the four hospital, an additional observer-rated item on psychotropic medication compliance is also routinely assessed. This was also included in the analyses. Finally, we took into account the available information on the referral process, such as who referred the patient, to which hospital, the legal status of the hospitalisation (voluntary or involuntary) and whether they were admitted during regular service hours (Monday/Friday from 8:00 a.m. to 18:00 p.m.) or outside regular service hours (Monday/Friday from 18:00 p.m. to 8:00 a.m., weekends and public holidays).

All data were anonymized and the Human Research Ethics Committee of the Canton Vaud (protocol #2016-00768) granted approval for this study.

2.3 Statistical analysis

The included sample (N=25584) was split into two subgroups based on legal status of the hospitalisation (voluntary hospitalisation (VH; n=15797) vs. involuntary hospitalisation (IH; n=9787)) and compared on all the available characteristics. Categorical variables were tested using Pearson’s Chi-square tests. Ordinal and continuous variables were analysed through non-parametric Mann-Whitney U tests. All statistical tests were
two-tailed. Because of the large sample size, significance level was set at p<.01 and effect sizes according to Cohen [52] were calculated.

In order to identify the main predictors of involuntary hospitalisation, we used two ML algorithms, Classification and Regression Tree (CART) [53] and Random Forests (RFs) [54], on observations with no missing data (N=14948; 63.3% VH vs 36.7% IH). Moreover, we used ordinary logistic regression to compare its predictive power with that obtained through CART and RFs.

First developed by Breiman et al. [53], CART is a flexible tool used to discover hierarchical and complex associations among variables. It is a non-parametric methodology, thus no assumptions are made on the distribution of the predicting variables and it can handle any type of data, numerical as well as categorical. Moreover, compared to traditional statistical methods, CART presents several advantages: 1. The best “splitting” variables are identified automatically at each step of the tree by searching among all available features; 2. It can easily deal with missing data; 3. It is relatively easy to perform, requiring little input from the user; 4. Its results are generally simple for clinicians to interpret [40]. CART presents a major improvement compared to previous tree-based algorithms, such as AID (Automatic Interaction Detection) and CHAID (Chi-square Automatic Interaction Detector). In CART, the tree can grow until further split is not possible due to the stop criteria, which can be defined a priori. For reasons of parsimony and generalization, we can also remove irrelevant branches by “pruning” in order to find the best potential CART [55].

RFs are a generalization of CART, in which a large number of trees are fitted on resampled versions of the original data and each single CART contributes to the final prediction by choosing the most important predictor.

Our predicted outcome was voluntary/involuntary hospitalisation (VH vs. IH). IH was defined as the positive class and VH as the negative one. Socio-demographic, clinical and procedural characteristics were the predictive variables (23 features; Table 1).

The three models were first fitted on the 70% of observations chosen randomly from the completely observed sample (training subsample; n=10464), and their predictive power was tested on the remaining 30%
Accepted manuscript: Authors' Copy

(test subsample; n=4484). For more details on the characteristics of training and test subsamples, please see the accompanying Supplementary Material. Default hyperparameters were used (maximum depth of the model = 30, minimum number of observations per node = 20, complexity parameter (cp) = 0.01, number of models for RFs =100). At each node, the split was performed on the variable with the lowest Gini index. Beside predictive accuracy, balanced accuracy and validity (AUC: Area Under Curve), other performance statistics were calculated: sensitivity, specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV).

Due to the large proportion of missing information, which varied between 0 and 23% depending on the variable, and in order to verify the robustness of our results, we performed sensitivity analyses. Missing data were imputed through Multiple Imputation using Chain Equations [56]. CART and RFs were retested on the imputed data (N=25584).

All statistical analyses were performed using IBM SPSS 26 and R 4.0.1 (mice, stat, rpart and randomForest libraries).

3. RESULTS

Between January 1, 2013 and December 31, 2017, 25790 hospitalisations were registered within the four hospitals. 206 cases were excluded from the analyses since the legal status of the hospitalisation was unknown. Of the included 25584 hospitalisations, 61.7% were voluntary and 38.3% involuntary. The total sample was composed of 51% women, with an average age of 45 years, single (50%), of Swiss nationality (66%) and with a primary diagnosis of schizophrenic (F20-F29) (24%) or depressive disorders (F32-F39) (19%).

3.1 Socio-demographic, clinical and referral process characteristics: subgroups comparisons

The associations between legal status of the hospitalisation and the socio-demographic, clinical and referral process characteristics are presented in Table 1. Globally, the two subgroups differed significantly on all the characteristics with the exclusion of sex, nationality and secondary diagnosis of addiction. However, only five variables reached an effect size of .2 (small to medium effect) and only one exceeded the threshold of .3 (medium effect). Indeed, the IH group showed a considerably higher percentage of people affected by
Accepted manuscript: Authors' Copy

Organic mental disorders (F00-F09) ($\chi^2(1)=1259.9; p<.001$). Moreover, involuntary patients scored higher on four HoNOS items: Overactive, aggressive, disruptive or agitated behaviour ($U=38316000; z=-48.5; p<.001$), Cognitive problems ($U=40715055; z=-31.5; p<.001$), Problems associated with hallucinations and delusions ($U=41784750; z=-35.8; p<.001$), and the Problems with psychotropic medication compliance additional item ($U=34761181; z=-32.4; p<.001$). Finally, a remarkably higher percentage of voluntary patients auto-referred themselves to the hospital ($\chi^2(1)=1663.5; p<.001$).

Table 1: Socio-demographic, clinical and referral process characteristics: subgroups comparisons (N=25584)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>VH (n=15797; 61.7%)</th>
<th>IH (n=9787; 38.3%)</th>
<th>p-value</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, Mean (SD) Mdn (IQR)</td>
<td>42.3 (16.4) 41.0 (22.0) 48.6 (21.4) 46.0 (34.0)</td>
<td>&lt;.001$^a$</td>
<td>$r=.12^a$</td>
<td></td>
</tr>
<tr>
<td>Sex, Male, % (n)</td>
<td>48.5 (7648) 49.3 (4829)</td>
<td>.217$^a$</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Marital status, % (n)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>50.7 (7947) 48.4 (4694)</td>
<td>&lt;.001$^a$</td>
<td>V=.11$^A$</td>
<td></td>
</tr>
<tr>
<td>Married/Registered partnership</td>
<td>21.5 (3367) 21.0 (2036)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Divorced/Separated</td>
<td>24.4 (3822) 21.7 (2102)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Widowed</td>
<td>3.5 (553) 8.9 (865)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nationality, Swiss, % (n)</td>
<td>66.7 (10479) 66.0 (6425)</td>
<td>&lt;.293$^a$</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Primary diagnosis at discharge (ICD-10), % (n)</td>
<td>Organic, including symptomatic, mental disorders (F00-F09)</td>
<td>2.7 (429) 14.7 (1433)</td>
<td>&lt;.001$^a$</td>
<td>V=.22$^{AA}$</td>
</tr>
<tr>
<td>Mental and behavioural disorders due to use of alcohol (F10)</td>
<td>10.4 (1632) 9.3 (908)</td>
<td>.003$^a$</td>
<td>V=.02$^A$</td>
<td></td>
</tr>
<tr>
<td>Mental and behavioural disorders due to psychoactive substance use (F11-F19)</td>
<td>9.4 (1469) 4.8 (468)</td>
<td>&lt;.001$^a$</td>
<td>V=.08$^A$</td>
<td></td>
</tr>
<tr>
<td>Schizophrenia, schizotypal and delusional disorders (F20-F29)</td>
<td>19.4 (3041) 30.8 (3007)</td>
<td>&lt;.001$^a$</td>
<td>V=.13$^A$</td>
<td></td>
</tr>
<tr>
<td>Manic and bipolar affective disorders (F30-F31)</td>
<td>6.0 (936) 7.7 (753)</td>
<td>&lt;.001$^a$</td>
<td>V=.03$^A$</td>
<td></td>
</tr>
<tr>
<td>Mood [affective] disorders (manic and bipolar affective disorders excluded) (F32-F39)</td>
<td>24.0 (3764) 11.0 (1075)</td>
<td>&lt;.001$^a$</td>
<td>V=.16$^A$</td>
<td></td>
</tr>
<tr>
<td>Neurotic, stress-related and somatoform disorders (F40-F48)</td>
<td>13.0 (2029) 7.8 (763)</td>
<td>&lt;.001$^a$</td>
<td>V=.08$^A$</td>
<td></td>
</tr>
<tr>
<td>Disorders of adult personality and behaviour (F60-F69)</td>
<td>10.4 (1632) 7.3 (714)</td>
<td>&lt;.001$^a$</td>
<td>V=.05$^A$</td>
<td></td>
</tr>
<tr>
<td>Other mental disorders</td>
<td>3.2 (497) 3.9 (384)</td>
<td>.001$^a$</td>
<td>V=.02$^A$</td>
<td></td>
</tr>
<tr>
<td>Other non-mental disorders</td>
<td>1.4 (223) 2.7 (268)</td>
<td>&lt;.001$^a$</td>
<td>V=.05$^A$</td>
<td></td>
</tr>
<tr>
<td>Comorbidity F10-F19, % (n)</td>
<td>11.2 (1758) 10.8 (1056)</td>
<td>.292$^a$</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Comorbidity F60-F69, % (n)</td>
<td>15.3 (2394) 10.6 (1033)</td>
<td>&lt;.001$^a$</td>
<td>V=.07$^A$</td>
<td></td>
</tr>
<tr>
<td>HoNOS scores at admission, Mean (SD) Mdn (IQR)</td>
<td>Overactive, aggressive, disruptive or agitated behaviour</td>
<td>0.8 (1.2) 0 (2) 1.7 (1.5) 2 (3)</td>
<td>&lt;.001$^b$</td>
<td>r=.32$^{AAA}$</td>
</tr>
<tr>
<td>Non-accidental self-injury</td>
<td>0.7 (1.3) 0 (1) 0.6 (1.2) 0 (0)</td>
<td>&lt;.001$^b$</td>
<td>r=.04$^A$</td>
<td></td>
</tr>
<tr>
<td>Problem drinking or drug-taking</td>
<td>1.3 (1.6) 0 (3) 1.2 (1.6) 0 (3)</td>
<td>&lt;.001$^b$</td>
<td>r=.04$^A$</td>
<td></td>
</tr>
<tr>
<td>Cognitive problems</td>
<td>0.7 (1.1) 0 (1) 1.4 (1.5) 1 (3)</td>
<td>&lt;.001$^b$</td>
<td>r=.22$^{AA}$</td>
<td></td>
</tr>
<tr>
<td>Physical illness or disability problems</td>
<td>0.7 (1.2) 0 (1) 0.9 (1.4) 0 (2)</td>
<td>&lt;.001$^b$</td>
<td>r=.08$^A$</td>
<td></td>
</tr>
<tr>
<td>Problems associated with hallucinations and delusions</td>
<td>0.8 (1.3) 0 (1) 1.5 (1.6) 1 (3)</td>
<td>&lt;.001$^b$</td>
<td>r=.24$^{AA}$</td>
<td></td>
</tr>
<tr>
<td>Problems with depressed mood</td>
<td>2.3 (1.3) 3 (1) 1.8 (1.4) 2 (3)</td>
<td>&lt;.001$^b$</td>
<td>r=.18$^A$</td>
<td></td>
</tr>
<tr>
<td>Other mental and behavioural problems</td>
<td>2.2 (1.4) 2 (2) 2.1 (1.5) 2 (3)</td>
<td>.003$^b$</td>
<td>r=.02$^A$</td>
<td></td>
</tr>
<tr>
<td>Problems with relationships</td>
<td>1.6 (1.3) 2 (3) 1.9 (1.4) 2 (3)</td>
<td>&lt;.001$^b$</td>
<td>r=.12$^A$</td>
<td></td>
</tr>
<tr>
<td>Problems with activities of daily living</td>
<td>1.1 (1.3) 1 (2) 1.6 (1.4) 2 (3)</td>
<td>&lt;.001$^b$</td>
<td>r=.16$^A$</td>
<td></td>
</tr>
<tr>
<td>Problems with living conditions</td>
<td>1.2 (1.4) 1 (2) 1.4 (1.5) 1 (3)</td>
<td>&lt;.001$^b$</td>
<td>r=.06$^A$</td>
<td></td>
</tr>
<tr>
<td>Problems with occupation and activities</td>
<td>1.7 (1.4) 2 (3) 1.8 (1.4) 2 (3)</td>
<td>&lt;.001$^b$</td>
<td>r=.06$^A$</td>
<td></td>
</tr>
<tr>
<td>Problems with psychotropic medication compliance (additional item)</td>
<td>0.9 (1.3) 0 (2) 1.6 (1.6) 1 (3)</td>
<td>&lt;.001$^b$</td>
<td>r=.23$^{AA}$</td>
<td></td>
</tr>
<tr>
<td>Referral from, % (n)</td>
<td>Patient</td>
<td>20.5 (3122) 2.3 (220)</td>
<td>&lt;.001$^a$</td>
<td>V=.26$^{AA}$</td>
</tr>
</tbody>
</table>
3.2 Predicting factors of involuntary hospitalisation

The original data contained a large number of missing values, and only 58% (14948 out of 25584) of the observations were complete.

Table 2 shows the predictive statistics produced for the three models on the completely observed test subsample (n=4484). The three methods achieved very similar results, with RFs and logistic regression showing slightly better performances compared to CART. Predictive balanced accuracy ranged from 68% for CART to 72% for logistic regression. The three models showed better specificity than sensitivity. Indeed, on the test subsample they were able to correctly classify between 81% and 86% of actual cases of voluntary hospitalisation and between 54% and 58% of actual cases of involuntary hospitalisation.

Table 2: Comparison of models’ validity

<table>
<thead>
<tr>
<th></th>
<th>CART</th>
<th>Random Forests</th>
<th>Logistic regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balanced accuracy (%)</td>
<td>68.0</td>
<td>71.0</td>
<td>72.0</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>71.3</td>
<td>75.6</td>
<td>75.6</td>
</tr>
<tr>
<td>AUC</td>
<td>0.68</td>
<td>0.71</td>
<td>0.72</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.54</td>
<td>0.57</td>
<td>0.58</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.81</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>PPV</td>
<td>0.61</td>
<td>0.69</td>
<td>0.69</td>
</tr>
<tr>
<td>NPV</td>
<td>0.76</td>
<td>0.78</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Note. AUC = Area Under Curve; PPV = Positive Predictive Value; NPV = Negative Predictive Value.
When we fitted a CART on the training subsample (n=10464), the resulting tree showed that only three parameters were essential to estimate the probability of being involuntarily admitted to hospital: the HoNOS score on the “Overactive, aggressive, disruptive or agitated behaviour” item, who referred the patient to hospital and primary diagnosis (Figure 1).

Fig. 1: Classification and Regression Tree to predict involuntary hospitalisations on 70% of observations chosen randomly from completely observed sample (N=10464)

- VH Voluntary Hospitalisation, IH Involuntary Hospitalisation. HoNOS scores: 0 = no problem; 1 = minor problem requiring no action; 2 = mild problem but definitely present; 3 = moderately severe problem; 4 = severe to very severe problem. Referral from: 1 = Patient; 2 = Family/relatives; 3 = General practitioner; 4 = General hospital; 5 = Outpatient psychiatrist; 6 = Psychiatric hospital; 7 = Civil justice/other justice authority; 8 = Other. Primary diagnosis: 1 = F00-F09; 2 = F10; 3 = F11-F19; 4 = F20-F29; 5 = F30-F31; 6 = F32-F39; 7 = F40-F48; 8 = F40-F48; 9 = Other mental disorders; 10 = Other non-mental disorders.

The likelihood of coercion was higher for people with moderate to very severe aggressive behaviours (65%), especially if not auto-referred to hospital (68%). This percentage dropped to 21% when they referred themselves.

Among people showing none up to mild aggressiveness, there was a higher risk of involuntary admission if they were referred to hospital by a general practitioner, a general hospital, a psychiatric hospital or the civil justice (46%). A further increase was recorded if patients were affected by organic disorders (F00-F09), disorders due to alcohol use (F10), schizophrenia (F20-F29) and other non-mental disorders (58%). On the contrary, the odds of being involuntarily hospitalised decreased when the HoNOS score for aggressive behaviour was lower than 3 (30%), and the patients referred themselves, or were referred by a family member/relative, an outpatient psychiatrist or others (20%).

When RFs were fitted on the training subsample, the importance of each predictor was estimated based on its mean decrease in accuracy and its mean decrease in Gini coefficients (Figure 2). Based on these results, four factors emerged as the most important in both classifications: who referred the patient, the HoNOS score for
“Overactive, aggressive, disruptive or agitated behaviour”, primary diagnosis and age. The exclusion of one of these four variables would reduce either the accuracy of the model or its homogeneity.

Finally, of the 23 factors included in the logistic regression model, 15 were found to affect significantly the likelihood of being coerced to hospital (Table 3). However, in order to calculate the probability of an individual belonging to the IH group using this model, all the 23 characteristics must be available.

### Table 3: Predictors of involuntary hospitalisations: logistic regression analysis on 70% of observations chosen randomly from completely observed sample (N=10464)

<table>
<thead>
<tr>
<th>Predicting factors</th>
<th>OR</th>
<th>95% C.I.</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1.01</td>
<td>1.00</td>
<td>1.01</td>
</tr>
<tr>
<td>Sex (ref. Male)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1.05</td>
<td>0.95</td>
<td>1.15</td>
</tr>
<tr>
<td>Marital status (ref. Single)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married/Registered partnership</td>
<td>0.92</td>
<td>0.80</td>
<td>1.07</td>
</tr>
<tr>
<td>Divorced/Separated</td>
<td>0.99</td>
<td>0.86</td>
<td>1.13</td>
</tr>
<tr>
<td>Widowed</td>
<td>1.12</td>
<td>0.87</td>
<td>1.44</td>
</tr>
<tr>
<td>Nationality (ref. Swiss)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1.28</td>
<td>1.15</td>
<td>1.42</td>
</tr>
<tr>
<td>Primary diagnosis at discharge (ref. F00-F09)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental and behavioural disorders due to use of alcohol (F10)</td>
<td>0.48</td>
<td>0.36</td>
<td>0.64</td>
</tr>
<tr>
<td>Mental and behavioural disorders due to psychoactive substance use (F11-F19)</td>
<td>0.21</td>
<td>0.15</td>
<td>0.29</td>
</tr>
<tr>
<td>Schizophrenia, schizotypal and delusional disorders (F20-F29)</td>
<td>0.50</td>
<td>0.38</td>
<td>0.64</td>
</tr>
<tr>
<td>Manic and bipolar affective disorders (F30-F31)</td>
<td>0.46</td>
<td>0.34</td>
<td>0.61</td>
</tr>
<tr>
<td>Mood [affective] disorders (manic and bipolar affective disorders excluded) (F32-F39)</td>
<td>0.36</td>
<td>0.28</td>
<td>0.47</td>
</tr>
<tr>
<td>Neurotic, stress-related and somatoform disorders (F40-F48)</td>
<td>0.44</td>
<td>0.33</td>
<td>0.57</td>
</tr>
<tr>
<td>Disorders of adult personality and behaviour (F60-F69)</td>
<td>0.38</td>
<td>0.29</td>
<td>0.51</td>
</tr>
<tr>
<td>Other mental disorders</td>
<td>0.41</td>
<td>0.28</td>
<td>0.58</td>
</tr>
<tr>
<td>Comorbidity F10-F19</td>
<td>0.80</td>
<td>0.68</td>
<td>0.94</td>
</tr>
<tr>
<td>Comorbidity F60-F69</td>
<td>1.13</td>
<td>0.98</td>
<td>1.31</td>
</tr>
<tr>
<td>HoNOS:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overactive, aggressive, disruptive or agitated behaviour</td>
<td>1.33</td>
<td>1.28</td>
<td>1.39</td>
</tr>
<tr>
<td>Non-accidental self-injury</td>
<td>1.07</td>
<td>1.02</td>
<td>1.11</td>
</tr>
<tr>
<td>Problem drinking or drug-taking</td>
<td>0.99</td>
<td>0.95</td>
<td>1.04</td>
</tr>
<tr>
<td>Cognitive problems</td>
<td>1.13</td>
<td>1.07</td>
<td>1.18</td>
</tr>
<tr>
<td>Physical illness or disability problems</td>
<td>0.97</td>
<td>0.93</td>
<td>1.01</td>
</tr>
</tbody>
</table>
Problems associated with hallucinations and delusions    1.14  1.10  1.19  <.001
Problems with depressed mood                     0.80  0.76  0.83  <.001
Other mental and behavioural problems            0.93  0.90  0.96  <.001
Problems with relationships                       1.04  0.99  1.08  .102
Problems with activities of daily living          1.01  0.96  1.06  .706
Problems with living conditions                   1.06  1.02  1.10  .005
Problems with occupation and activities           0.96  0.92  1.01  .102
Problems with psychotropic medication compliance (additional item) 1.18  1.14  1.23  <.001

Referral from (ref. Patient)
Family/relatives                                   4.65  3.28  6.58  <.001
General practitioner                              13.48 10.60 17.29  <.001
General hospital                                  11.12  8.77 14.22  <.001
Outpatient psychiatrist                           5.70  4.56  7.19  <.001
Psychiatric hospital                              8.86  6.73 11.73  <.001
Civil justice/ other justice authority             32.03 22.23 46.73  <.001
Other                                              5.55  3.99  7.72  <.001

Hospital (ref. Hospital 1)
Hospital 2                                          0.65  0.56  0.74  <.001
Hospital 3                                          0.91  0.80  1.03  .133
Hospital 4                                          0.76  0.65  0.89  .001

Time of admission (ref. Regular service hours)
Outside regular service hours                     1.51  1.37  1.67  <.001
Intercept                                          0.65  0.45  0.92  .015

Note. OR = Odds Ratio; C.I = Confidence Interval.

3.3 Sensitivity analyses
Due to the large proportion of missing information and in order to test the robustness of our results, we fitted both CART and RFs on 100 copies of data completed via multiple imputation. For CART the predictive accuracy on all scenarios ranged from 70% to 72%. The same CART shown in Figure 1 was obtained in 96 out of 100 replications of data completed via multiple imputation. For RFs the predictive accuracy for the 100 multiple imputation data versions varied between 73% and 75%.

4. DISCUSSION
The main aim of this study was to identify predicting factors of involuntary hospitalization using machine-learning models. For this purpose, we analyzed our data using CART and RFs, and compared their predictive power with traditional logistic regression.

Our results showed that the performances of the three methods were very similar. CART showed the lowest predictive balanced accuracy (68%) but the most parsimonious model, allowing to estimate the probability of
being involuntarily admitted to hospital with only three checks (aggressive behaviours, who referred the patient to hospital and primary diagnosis). Thanks to his binary structure, CART provided a prompt view of the most relevant predictors of involuntary admission and of their non-linear interactions. RFs and logistic regression achieved slightly higher results (71% and 72%), but at the cost of an increased complexity and a depleted practical feasibility of their models. In RFs, the mean decrease in accuracy and the mean decrease in Gini coefficients provided information about the importance of each variable and its impact on the accuracy and homogeneity of the model. However, nothing can be disclosed on how these factors interacted and affected each other. Logistic regression also made statistical interplay between risk factors difficult to model and interpret [13, 39, 40]. Moreover, more than 20 parameters had to be taken into account in order to estimate the probability of an individual to belong to the IH group.

The three models showed a greater ability to correctly classify real cases of voluntary hospitalisation than to detect real cases of involuntary hospitalisation. This result calls into question their actual use in real life and what margin of error should be considered ethically acceptable when dealing with a topic as sensitive as the use of coercion. Further research is mandatory to improve the sensitivity of the models.

The CART identified aggressive behaviours as the strongest predictor of coercion. People with moderate to very severe level of aggressiveness, especially if not auto-referred, showed the highest risk of involuntary hospitalisation. Interestingly, the few aggressive patients who referred themselves had a low probability of compulsory admission. This suggests that, beside aggressiveness and agitation, uncooperativeness played a key role during the hospitalization process. RFs and logistic regression also identified aggression as one of the most important factor. Considering that in Switzerland dangerousness is not a prerequisite for involuntary admission but is only required to keep in hospital a voluntarily admitted person who wants to leave the institution, this finding cannot be considered as a mere consequence of the legal requirements in force.

Several previous studies have already highlighted the key role of aggressive behaviours as a risk factor of coercion [15, 17, 18, 20, 21, 23, 32, 41, 44, 57]. Aggressiveness may also affect the likelihood of being involuntarily admitted because it often requires the involvement of the police, which has been found to be another important risk factor of involuntary admission [17, 20, 44]. Perceived risk of danger to self or others has been found to be crucial in professionals’ decisions on involuntary admissions [58, 59]. However, an
accurate prediction of serious violence is very complex [60-62]. The use of a structured risk assessment during the admission process and the first days of treatment should be fostered [63-65]. Furthermore, the implementation of interventions and tools able to improve patients’ empowerment and participation in decision-making during psychiatric crisis, such as Joint Crisis Plan [66], should also be promoted in order to enhance cooperation, reduce the risk of aggressive behaviours and consequently reduce the use of coercion. Finally, dealing with aggression can be extremely challenging and stressful for staff [67]. A large number of referring physicians felt threatened by their patients when facing psychiatric emergency situations [68]. Thus, specific training programmes on communication skills, management of aggressiveness and de-escalation techniques should be provided to all professionals involved in the referral process [63, 64, 69-71].

Indeed, referral process was identified as the second most important predictive factor in our decision tree. For people with lower levels of aggression, there was an increased risk of compulsion if they were referred to hospital by a general practitioner, a general hospital, a psychiatric hospital or the civil justice. On the contrary, non-aggressive patients referred by themselves, by their relatives or by a psychiatrist were less likely to be involuntarily admitted regardless of diagnosis. Referral was also recognised as a strong predictor of coercion in both RFs and logistic regression. A previous study had already highlighted the association between physicians’ qualifications and coercion rates, showing that limiting the right to require involuntary admission to psychiatrists could reduce the use of compulsion [16]. Another study in Zurich, Switzerland, showed that referring general practitioners found it more difficult to apply legal criteria for involuntary admission and to assess its necessity [72]. GPs also provided the lowest quality commitment certificates [73, 74]. In Cologne, Germany, Schmitz-Buhl et al. (2019) found that the risk of involuntary treatment of patients with suicidal and self-harm behaviour was higher in the two psychiatric hospitals that also showed a higher proportion of referrals from general emergency units [13]. This result should call into question the organisation of the on-call system in the canton of Vaud, which, despite the local high provision of psychiatrists (0.73/1’000 inhabitants), has been so far carried on solely by general practitioners. It is also worth to mention that psychiatric emergency services are not fully developed in the canton. More efforts should be made to enhance the psychiatric training of primary care physicians, who are often called outside
regular service hours and in general emergency services to provide support and assess patients with mental health crisis.

The paramount need to develop trainings that improve the clinical knowhow and the aggressiveness management skills of all professionals involved in the admission process was also one of the recommendations that emerged from the EUNOMIA study for good clinical practice on involuntary hospitalisation [75].

The last predictive factor identified by the CART was primary diagnosis. Among people with lower levels of aggressive behaviours and referred to hospital by a general practitioner, a general hospital, a psychiatric hospital or the civil justice, a higher risk for involuntary admission was registered if they were affected by organic disorders (F00-F09), disorders due to alcohol use (F10), schizophrenia (F20-F29) and other non-mental disorders. The logistic regression model upheld the strong predictive power of organic disorders, showing that they significantly increased the risk of being involuntarily admitted compared to all other diagnoses. New alternative treatments should be developed to better meet the special needs of this particular population. Indeed, several studies have already confirmed the relevance of this diagnosis as well as of the diagnosis of schizophrenia in predicting the risk of coercion [3, 13-15, 19, 23, 25, 26, 28, 30, 33, 43, 76]. Less conclusive are instead the findings concerning disorders due to alcohol use. Some earlier studies had found that being affected by substance abuse disorders, included alcohol, increased the odds of involuntary admission [3, 28] while others showed a diminished risk for this population [13]. Finally, the increased risk associated with other non-mental disorders was not confirmed in the existing literature, according to which this diagnosis would have no impact on the likelihood of being involuntarily admitted [15]. The heterogeneity of the diagnostic categories included in this node makes it difficult to draw conclusions on which preventive measures should be fostered for this subgroup of patients. More homogeneous and therefore more informative subgroups could be obtained by including other parameters in the model.

Even though our main findings are mainly in line with previous studies, we cannot exclude that their representativeness might be affected by specific area-related characteristics, such as the local rate of involuntary admission, the specific regulation in place and, the local mental health organisation and
availability. According to the Swiss Health Observatory data, in 2016 the canton of Vaud was the canton with the highest rate of involuntary hospitalisation in Switzerland, followed by Zurich and Geneva [48]. With 2.3 involuntary admission per 1’000 inhabitants, the canton of Vaud also ranked above most European countries, along with Germany and Austria [6]. Moreover, contrary to the majority of other Swiss cantons, in the canton of Vaud only some specific categories of medical doctors (general practitioners, psychiatrists, paediatricians and on-call doctors) are allowed to order an involuntary inpatient treatment. Furthermore, as mentioned above, lack of capacity to consent to treatment and imminent danger, which are required in most countries, are not essential in Switzerland in order to admit someone involuntarily. Finally, model results could also be affected by the important variations existing among cantons and countries in term of psychiatric beds provision and community service development. Therefore, further national and international studies are needed to confirm the generalizability of our model.

4.1 Strengths and limitations

The main strength of this study is the large sample size. Thanks to the available routinely collected data, we were able to include in the study all psychiatric admissions occurred between 2013 and 2017 in the canton of Vaud, Switzerland. The second main strength lies in the innovative methodologies deployed to analyse a large amount of data, without a priori selection of variables. The use of CART allowed us to overcome some of the main limitations of traditional statistical methods [13, 39, 40], providing a parsimonious and feasible hierarchical model with comparable predictive power.

Two main limitations should be reported. First, because of the retrospective nature of the study, only available routinely collected data could be included in our predictive models. Several important factors could not be taken into account. The introduction of these parameters into the models could have further improved their performances, especially their sensitivity. The second shortcoming was the large amount of missing values in the original data. To overcome this limitation, sensitivity analyses were performed on 100 copies of data completed via multiple imputation. The results of both CART and RFs on the imputed data were almost identical to those obtained on the original data, confirming the robustness of our models.
4.2 Conclusion

This study confirmed the essential role of aggressiveness, referral and diagnosis as predictors of coercion. Identifying risk factors of involuntary admission is essential in order to efficiently target the development of professional training, preventive strategies and alternative interventions. Machine-learning methodologies offer new effective tools to achieve this goal, providing accurate yet simple models that could be used in clinical practice.
Acknowledgements:

The authors would like to thank the hospitals’ management and data managers for their collaboration.

Financial support:

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Conflicts of interest:

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Supplementary Material

For supplementary material accompanying this paper, visit cambridge.org/EPA.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.
REFERENCES

Accepted manuscript: Authors' Copy


Accepted manuscript: Authors' Copy


[40] Lewis RJ, editor An introduction to classification and regression tree (CART) analysis. Annual meeting of the society for academic emergency medicine in San Francisco, California; 2000.


Accepted manuscript: Authors' Copy


