

Predicting Hospitalization following Psychiatric Crisis Care using Machine Learning

Matthijs Blankers (✉ mblankers@gmail.com)

Arkin <https://orcid.org/0000-0002-8821-3312>

Louk F. M. van der Post

Arkin

Jack J. M. Dekker

Arkin

Research article

Keywords: Psychiatric hospitalization, machine learning, acute psychiatry, prognostic modeling.

Posted Date: August 2nd, 2019

DOI: <https://doi.org/10.21203/rs.2.12338/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Medical Informatics and Decision Making on December 1st, 2020. See the published version at <https://doi.org/10.1186/s12911-020-01361-1>.

Abstract

Background: It is difficult to accurately predict whether a patient on the verge of a potential psychiatric crisis will need to be hospitalized. Machine learning may be helpful to improve the accuracy of psychiatric hospitalization prediction models. In this paper we evaluate and compare the accuracy of ten machine learning algorithms including the commonly used generalized linear model (GLM/logistic regression) to predict psychiatric hospitalization in the first 12 months after a psychiatric crisis care contact, and explore the most important predictor variables of hospitalization. **Methods:** Data from 2,084 patients with at least one reported psychiatric crisis care contact included in the longitudinal Amsterdam Study of Acute Psychiatry were used. The accuracy and area under the receiver operating characteristic curve (AUC) of the machine learning algorithms were compared. We also estimated the relative importance of each predictor variable. The best and least performing algorithms were compared with GLM/logistic regression using net reclassification improvement analysis. Target variable for the prediction models was whether or not the patient was hospitalized in the 12 months following inclusion in the study. The 39 predictor variables were related to patients' socio-demographics, clinical characteristics and previous mental health care contacts. **Results:** We found Gradient Boosting to perform the best (AUC=0.774) and K-Nearest Neighbors performing the least (AUC=0.702). The performance of GLM/logistic regression (AUC=0.76) was above average among the tested algorithms. Gradient Boosting outperformed GLM/logistic regression and K-Nearest Neighbors, and GLM outperformed K-Nearest Neighbors in a Net Reclassification Improvement analysis, although the differences between Gradient Boosting and GLM/logistic regression were small. Nine of the top-10 most important predictor variables were related to previous mental health care use. **Conclusions:** Gradient Boosting led to the highest predictive accuracy and AUC while GLM/logistic regression performed average among the tested algorithms. Although statistically significant, the magnitude of the differences between the machine learning algorithms was modest. Future studies may consider to combine multiple algorithms in an ensemble model for optimal performance and to mitigate the risk of choosing suboptimal performing algorithms.

Background

In this paper, we evaluate and compare the performance of ten different machine learning (ML) algorithms to predict psychiatric hospitalization in the first 12 months after a psychiatric crisis care contact. Hospitalization is traditionally a preferred care modality for patients with severe mental illnesses or for those experiencing acute psychiatric crisis¹. Recently, it has been debated whether hospitalization could be prescribed less often than has been done in the past, as in-patient acute mental health services are unpopular with service users^{2,3}. One of the reasons for this unpopularity is that hospitalization often fails to address individuals' needs or provide a safe and therapeutic environment^{3,4,5}. Acute psychiatric hospitalization is also hypothesized to be more expensive than out-patient alternatives, although research on cost-effectiveness of alternatives to acute psychiatric hospitalization is still in its infancy³.

Some patients however will still be hospitalized at some point during their illness and recovery. At this moment, it is difficult to know accurately from beforehand which patient will need to be hospitalized in the near future, as currently a valid prognostic model for hospitalization after a psychiatric crisis is lacking. Previous studies (e.g.^{6,7,8,9}) have done important work in identifying predictors (e.g. quality of life, psychiatric diagnosis, impact of symptoms, living situation) that can be relevant for a risk classification of the near-future need for psychiatric hospitalization. Results show that previous (involuntary) admission and level of previous psychiatric service use in general seem to be powerful predictors of readmission^{10,11,12}. Homelessness at admission discharge^{11,13}, being on benefits¹⁴, being unmarried or living alone and/or having a small social network^{11,15} and being of African and/or Caribbean origin¹⁴ also are known predictors of readmission.

One thing the previously mentioned studies have in common is that they used algorithms based on generalized linear modelling (GLM) – such as logistic regression - to find predictors for hospitalization and to construct risk classification models. Recently, papers have been published which have also used other, more recently developed ML algorithms to develop prediction models, for example to predict re-hospitalization after heart failure¹⁶, persistence of depression symptoms¹⁷, or prediction of suicides after psychiatric hospitalisation¹⁸. Kessler and colleagues¹⁷ found that ML outperformed logistic regression in terms of model accuracy while in the two other papers the results of the different algorithms were more similar. In the current paper, we want to compare the accuracy of a diverse set of ML algorithms. We will compare GLM and the alternative algorithms in their ability to exploit a set of predictors to construct a validated prognostic model for psychiatric hospitalization following psychiatric crisis care. We will use a routinely collected data set¹⁹, containing similar variables as in the previously discussed studies. We will address the following questions:

- Which of the evaluated ML algorithms has the best prognostic potential?
- Which variables are the most important predictors for psychiatric hospitalization among patients on the verge of psychiatric crisis?
- Do the alternative ML algorithms statistically outperform regression-based models?

Methods

Patient data source

We will train and test the performance of GLM and nine other ML algorithms using historical data from the Amsterdam Study of Acute Psychiatry (ASAP). The aim of ASAP was to study the relationship between the incidence of (involuntary) psychiatric hospitalizations on the one hand and prior psychiatric history, the course of the psychiatric disorder, the patient's social circumstances, and patient opinions and experiences on the other^{19,20}. The dataset used in the current analysis contains data from a cohort of patients who had an emergency consultation either by the Psychiatric Emergency Service Amsterdam or the Acute Treatment Unit in Amsterdam between 15 September 2004 and 15 September 2006 (the “index”

contact), with a follow-up period of 12 months. Although some years old, it is still the largest and most extensive and complete dataset on long term hospitalization outcomes after psychiatric crisis care in the Netherlands.

The data collected at baseline during the emergency consultation were: age, gender, domestic situation and the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision axis I diagnostic category. To determine the severity of the current psychopathology, the Severity of Psychiatric Illness rating scale (SPI) ²¹ was used. The SPI contains 14 items rated using a four-point scale: no risk, low risk, moderate risk and high risk – or no information present²².

All variables related to health care consumption, and the number of care contacts in the five years before and the 12 months after the index contact were extracted from the patient health records kept by the three largest mental health institutions in Amsterdam: JellinekMentrum (now Arkin), AMC de Meren (now Arkin), and GGZ inGeest.

All analyses were performed on routinely collected anonymized data from the participating institutions. Therefore, this study was exempted from medical ethics review and opt-in informed consent from participants was not necessary according to article 9 of the General Data Protection Regulation²³.

Dependent variable

The dependent variable in our prediction model was hospitalization, operationalized as any psychiatric hospitalization in any of the three participating psychiatric hospitals in the 12 months after the index psychiatric crisis care contact.

Predictor variables

The following 39 variables have been collected in the ASAP study and were used to train our prediction model.

Socio-demographic (5): Gender, age, living situation (alone, with partner, with parents, institutionalized, other), marital status, cultural background (Dutch, Moroccan, Turkish, Surinamese, Netherlands Antilles, North Africa excl. Morocco, Sub Saharan Africa, other western minorities, other non-western minorities)

SPI items (14): Suicide risk, danger to others, severity of psychiatric symptoms, problems with self-care, substance misuse, medical condition(s), disturbances in patients' family connectedness, professional functioning, stability of patients' living situation, patient is motivated to receive treatment, prescription medication compliance, anosognosia, patients' family involvement in informal care, and symptom persistence.

Clinical (2): Psychiatric diagnosis (Depression, psychotic disorder, mania/bipolar disorder, alcohol/substance use disorder, all other disorders, no diagnosis) and Global Assessment of Functioning (GAF) score (0-100).

Psychiatric intake and care register data (18): Patients' informal social support system involved (Yes/No), patient referrer (general practitioner, first aid station, mental health care, police, other), number of previous face-to-face treatment contacts up to 2 weeks / 1 month / 3 months / 6 months / 12 months before the index crisis care contact, number of previous psychiatric hospitalizations (last 12 months and last 5 years), number of previous psychiatric day care treatments (last 12 months and last 5 years), number of involuntary treatments/hospitalizations (last 12 months and last 5 years), days of psychiatric hospitalization (last 12 months), any earlier psychiatric care referrals (> 1 year and >5 years before current contact).

Statistical procedures

First, a dataset was created consisting of the 39 predictor variables and the dependent variable. Patients with missing hospitalization data, missing SPI data, or who died during the study were removed. As some of the used statistical techniques cannot adequately handle missing observations, the remaining missing data were imputed using the mice package²⁴ in R.

The ML techniques were first ran on training data to create a model, and then evaluated on independent test data. We used K-fold cross-validation (with K=10) to validate the model parameters. For K-fold cross-validation, K successive mutually exclusive test sets are created. Algorithm fitting is iteratively done on the training datasets. Predicted classifications are then calculated for the test set. With K=10, at each iteration another 10% of the data is set aside from the original dataset for validation purposes. In the end, each observation in the original data set has a predicted classification that was obtained when it was part of the test set²⁵. We chose K=10 as a simulation study by Kohavi²⁶ indicated that for real word datasets the best method to use for model selection is 10-fold stratified cross-validation.

GLM and nine other ML algorithms were selected in order to achieve maximum variation among the used approaches. The nine other ML algorithms were DeepBoost (R package deepboost), Keras/TensorFlow (R package keras and the TensorFlow and Keras libraries for Python), k-nearest neighbors (R package class), naive Bayes (R package klaR), neural network (R package nnet), oblique random forest (R package obliqueRF), random forest (R package randomForest), stochastic gradient boosting (R package gbm), and (model averaged) support vector machines (with class weights) (R package kernlab). All algorithms had implementations in R and/or Python. A detailed description of all ten algorithms is presented in Additional File 1.

All numeric predictor variables were centered and scaled in the pre-processing phase. Categorical variables were recoded into dummy variables. In the presented base case analysis, we have not applied balancing of the two levels of the dependent variable (hospitalized/not hospitalized); in a sensitivity analysis, all results were validated under a balanced scenario which were created by under-sampling the most prevalent outcome. Confusion matrices, accuracy scores, sensitivities, specificities and the Area under the Receiver Operating Characteristic (ROC) curves (AUC, or c-statistic) were calculated for each model. The AUC measures the area underneath the plot of the ROC curve and is an aggregate measure of

the performance of the model²⁷. Theoretically the AUC can have any value between 0 and 1, with 0 corresponding with 100% wrong predictions, and 1 corresponding with 100% correct predictions.

We also estimated the relative unique importance of each individual predictor variable for the overall AUC score using the `filterVarImp` function in the R package `caret`²⁸. We standardized the AUC associated with each variable by dividing the absolute deviation for each variable by the absolute AUC deviation associated with the most impactful variable.

In order to evaluate the predictive accuracy of the best performing model against the GLM-based model and against the least performing model, we calculated the Net Reclassification Improvement (NRI) of the best performing model in comparison to the GLM-based model, and against the least performing model. The NRI is an index that provides an estimate (with a confidence interval) of how well a model classifies subjects compared to another model²⁹.

Results

The original dataset contained data from 2,707 patients. After removal of patients who had missing hospitalization data, completely missing SPI data, or who diseased during the follow-up period, data from 2,084 patients remained. The completeness rate of this data set was high with only 4.2% missing data.

Table 1 presents some key characteristics of the full study sample (n=2,084) and for those hospitalized and not hospitalized in the year following the index contact separately. Male participants have a higher probability of becoming hospitalized than female participants (37% vs. 31%, $p = 0.001$) and diagnosis ($X^2=120.1$, $df=5$; $p<0.0001$), cultural background ($X^2=12.16$, $df=5$; $p=0.033$) and living situation ($X^2=30.33$; $df=5$; $p<0.0001$) are also associated with future hospitalization, while age ($p=0.33$) is not (Table 1).

Figure 1 presents the AUC statistics for the models using the ML algorithms based on the 10-fold cross-validation tests using all 39 predictor variables. What can be observed foremost from Figure 1 is that most confidence intervals of the models overlap. The Gradient Boosting-based model shows the best prognostic performance (AUC=0.77), and K-Nearest Neighbors model has the least prognostic performance (AUC=0.70). The performance of the GLM-based model is slightly above average (AUC=0.76). The Gradient Boosting model also has the highest accuracy (0.744, see also Table 2). All models have an accuracy which is significantly above the 'no information rate' of 0.659, which is the proportion of not hospitalized patients.

Figure 2 presents data on the relative importance of each variable for the AUC. Results are averaged over the ten models; in Additional File 2 we have presented the variable importance data for each model separately. Overall, it can be observed that the number of earlier psychiatric hospitalizations in the five

years before the index contact and the number of face to face contacts the patient has had with professionals working for the participating mental health care center in the 12 months before have the strongest association with hospitalization in the year after the index contact.

In the NRI analysis the Gradient Boosting model led to 9.9% more correct classifications of hospitalized patients ($z=5.42$, $p<0.0001$) than the K-nearest Neighbors model, and 1.5% more correct classifications of non-hospitalized patients, which was a non-significant improvement ($z=1.56$, $p=0.12$). Compared to K-nearest Neighbors, Gradient Boosting led to an 11.3% increase in correctly classified patients overall ($z=5.53$, $p<0.0001$). Also the GLM/logistic regression model outperformed the K-nearest Neighbors model, and led to 8.7% more correct classifications of hospitalized patients ($z=4.64$, $p<0.0001$). The classification of not-hospitalized patients did not differ significantly between GLM/logistic regression and K-nearest Neighbors (-0.3% , $z=0.31$, $p=0.76$). Compared to K-nearest Neighbors, GLM/logistic regression led to an 8.4% increase in correctly classified patients overall ($z=4.00$, $p<0.0001$). Compared to GLM/Logistic regression, Gradient Boosting led to 1.1% more correct classifications of hospitalized patients, a non-significant improvement ($z=0.88$, $p=0.377$), and 1.8% more correct classifications non-hospitalized patients ($z=2.57$, $p=0.010$). Compared to GLM/Logistic regression, Gradient Boosting led to a 2.9% increase in correctly classified patients overall ($z=1.99$, $p=0.046$).

As a sensitivity analysis, we have performed all analyses on a balanced dataset as well, in which the prevalence of hospitalized and non-hospitalized patients was fixed to 0.5 and 0.5 respectively by under-sampling of the non-hospitalized patients. The results of the balanced dataset were very similar to those of the presented unbalanced dataset, including the differences between the most accurate algorithm (Gradient Boosting), GLM/Logistic regression and the least accurate algorithm (K-nearest Neighbors). Therefore, these results are not included in the main text but we included these as Additional File 3.

Discussion

In this paper, we evaluated and compared the performance of prognostic models based on ten ML algorithms. We tested which model predicted hospitalization the most accurately, and which variables are the most important for a psychiatric hospitalization prognostic model.

All ten ML based models had AUC scores >0.7 and only three models had an AUC <0.75 . This result could be considered to be good in the field of hospitalization prediction using clinical registry data. Artetxe and colleagues³⁰ for example found that over 80% of the hospital readmission models in their review had an AUC score below 0.75 – a finding in line with an earlier review by Kansagara and colleagues³¹.

We found differences in accuracy between the ML algorithms in this study, mostly of modest size. The Gradient Boosting model performed the best and the model based on K-Nearest Neighbors had the lowest predictive accuracy and AUC. The GLM/logistic regression model performed average, although the net

classification improvement of Gradient Boosting over GLM/logistic regression was only 2.9%, while the improvement of Gradient Boosting compared to K-Nearest Neighbors was 11.3% in our models.

With regard to the importance of individual variables generalized over the different models, we found that the number of earlier psychiatric hospitalizations in the last five years, the number of face to face therapy sessions in the last 12 months, and the number of earlier psychiatric care referrals > 12 months year before the initial crisis care contact are the most important variables for future psychiatric hospitalization after a crisis care contact. Nine out of the ten most important predictors generalized over the different models are associated to earlier mental health care consumption. Over the different models there is some variation in which variables have the highest impact on classification (see Additional File 2), although for all models the number of earlier hospitalizations is among the three most impactful predictors, and the ten most impactful predictors are predominantly related to earlier mental health care consumption for each model individually.

We have compared the results of the best performing ML algorithm (Gradient Boosting) with logistic regression (GLM) and the least performing ML algorithm (K-Nearest Neighbors). We found that the Gradient Boosting model did outperform the GLM/logistic regression model and K-Nearest Neighbors model, and that the GLM/logistic regression model outperformed the K-Nearest Neighbors model in terms of classification accuracy. The reported differences in accuracy between the tested models are statistically significant but modest in size for Gradient Boosting vs. GLM/logistic regression. This finding echoes the conclusion of the review by Artetxe and colleagues³⁰ that although promising, the real impact of recent ML algorithms in the domain of readmission risk prediction needs further study.

Strengths and limitations

The findings of this study should be interpreted in the light of its strengths and limitations. A strength of this study is the relatively large dataset of 2,084 patients from which 710 patients were hospitalized during the follow-up period, and the availability of 39 clinically relevant potential predictors of hospitalization. Although for ML purposes this may not be considered a notably large dataset, for psychiatry crisis care research projects it is rare to achieve such sample sizes. Another strength is that the dataset consists of routinely collected data with high ecological validity, while its completeness rate is high. Methodological strengths of this study include the direct comparison of ten different ML algorithms, and the use of K-fold cross-validation to optimally use the available data to train and test the models^{25,26}.

Limitations regarding our study are the fact that although the average missing data rate after data selection was quite low (4.2%) we still had to address data missingness, as most ML algorithms are not capable of working with data with missing observations. Another limitation is that although we have made a diverse selection of ML algorithms, it is a matter of debate to what extent findings regarding the selected algorithms generalize to slightly different algorithms. There is a possibility that better results could have been reached with other ML algorithms. A third limitation is that the number of variables available in the dataset was – for ML purposes – still somewhat limited. A forth limitation is that

regarding the impact of the individual variables, only the unique variance explained by each variable could be assessed. This may have led to an underestimation of the importance of some variables when algorithms which are less well able to handle correlated predictors such as GLM were applied. Fifth, we do not know to what extent our findings related to the superiority of some algorithms over others generalize well beyond the context in which we evaluated them.

Implications

One of the findings of this study is that there may be more accurate algorithms than GLM /logistic regression to develop a prognostic model for future psychiatric hospitalization. We found Gradient Boosting to outperform the other algorithms in this analysis, but we do not know whether finding generalizes beyond our study. As long as a more definitive and validated answer is lacking to what the most accurate algorithm is, one could consider to train ensemble models: multiple algorithms are used to solve a prediction or classification problem, and the results of these algorithms are then combined in one prognostic model. In this way, the risk of relying on a poorly performing algorithm is mitigated. More research is needed to evaluate which set of ML algorithms performs well when combined in an ensemble model.

Conclusions

In this paper, we showed it is feasible to construct a classification model with acceptable AUC and accuracy using the 39 predictors we evaluated. Combined, the ten evaluated ML based models found variables on previous mental health care consumption to be the most impactful predictors of future psychiatric hospitalization in the year after the index crisis care contact. We found that Gradient Boosting led to the highest predictive accuracy and AUC, and that GLM/logistic regression performed average compared to the other algorithms. Although statistically significant, we conclude that the improvement of the best performing algorithm over GLM/logistic regression is modest in size, and that not all ML algorithms outperform GLM/logistic regression. We however also conclude that the difference in predictive performance between the best and least performing model is considerable. Therefore, future studies may consider to combine multiple ML algorithms in an ensemble model to mitigate the risk of constructing poorly performing prognostic models.

Abbreviations

ASAP	Amsterdam Study of Acute Psychiatry
AUC	Area under the Receiver Operating Characteristic Curve
GAF	Global Assessment of Functioning
GLM	Generalized Linear Modelling

ML	Machine Learning
NRI	Net Reclassification Improvement
ROC	Receiver Operating Characteristic
SPI	Severity of Psychiatric Illness rating scale

Declarations

Ethics approval and consent to participate

All analyses were performed on routinely collected anonymized data from the participating institutions. Therefore, this study was exempted from medical ethics review and opt-in informed consent from participants was not necessary according to article 9 of the General Data Protection Regulation²³.

Consent for publication

Not applicable.

Availability of data and material

The data that support the findings of this study are available from Arkin Mental Health Care and GGZ inGeest but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Arkin Mental Health Care and GGZ inGeest.

Competing interests

The authors declare that they have no competing interests.

Funding

No external funding was received for the analyses and preparation of the manuscript. Time to work on this manuscript was kindly made available by Arkin Mental Health Care for the authors MB and JJMD.

Authors' contributions

MB JJMD and LFMvdP conceived the study and outline of the manuscript. JJMD and LFMvdP were involved in the data collection for ASAP. MB took the lead in performing the analyses. All authors contributed to the interpretation of the data and the results. MB drafted the first version of the manuscript, which was substantially revised by JJMD and LFMvdP. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Sharfstein SS. Goals of inpatient treatment for psychiatric disorders. *Annu Rev Med.* 2009;60:393-403.
2. Quirk A, Lelliott P. What do we know about life on acute psychiatric wards in the UK? A review of the research evidence. *Soc Sci Med.* 2001;53:1565–74.
3. Lloyd-Evans B, Slade M, Jagielska D, Johnson S. Residential alternatives to acute psychiatric hospital admission: systematic review. *Br J Psychiatry.* 2009;195(2):109-17.
4. Muijen M. Acute hospital care: ineffective, inefficient and poorly organised. *Psychiatr Bull.* 1999;23:257–9.
5. Barker S. Environmentally Unfriendly: Patients' Views of Conditions on Psychiatric Wards. *MIND,* 2000.
6. Shadmi E, Gelkopf M, Garber-Epstein P, Baloush-Kleinman V, Doudai R, Roe D. Routine patient reported outcomes as predictors of psychiatric rehospitalization. *Schizophr Res.* 2018;192:119-123.
7. Barker LC, Gruneir A, Fung K, et al. Predicting psychiatric readmission: sex-specific models to predict 30-day readmission following acute psychiatric hospitalization. *Soc Psychiatry Psychiatr Epidemiol.* 2018;53(2):139-149.
8. Righi G, Benevides J, Mazefsky C, et al. Predictors of Inpatient Psychiatric Hospitalization for Children and Adolescents with Autism Spectrum Disorder. *J Autism Dev Disord.* 2018;48(11):3647-3657.
9. Fornaro M, Iasevoli F, Novello S, et al. Predictors of hospitalization length of stay among re-admitted treatment-resistant Bipolar Disorder inpatients. *J Affect Disord.* 2018;228:118-124.
10. Hamilton, JE, Passos IC, de Azevedo Cardoso T, et al. Predictors of psychiatric readmission among patients with bipolar disorder at an academic safety-net hospital. *Aust N Z J Psychiatry.* 2016;50(6):584-93.
11. Hung YY, Chan HY, Pan YJ. Risk factors for readmission in schizophrenia patients following involuntary admission. *PLoS One.* 2017;12(10):e0186768.
12. Kallert TW, Glöckner M, Schützwohl M. Involuntary vs. voluntary hospital admission. A systematic literature review on outcome diversity. *Eur Arch Psychiatry Clin Neurosci.* 2008;258(4):195-209.
13. Laliberté V, Stergiopoulos V, Jacob B, Kurdyak P. Homelessness at discharge and its impact on psychiatric readmission and physician follow-up: a population-based cohort study. *Epidemiol Psychiatr Sci.* 2019:1-8.
14. Priebe S, Katsakou C, Amos T, et al. Patients' views and readmissions 1 year after involuntary hospitalisation. *Br J Psychiatry.* 2009;194(1):49-54.
15. Webber M, Huxley P. Social exclusion and risk of emergency compulsory admission. A case-control study. *Soc Psychiatry Psychiatr Epidemiol.* 2004;39(12):1000-9.

16. Frizzell JD, Liang L, Schulte PJ, et al. Prediction of 30-Day All-Cause Readmissions in Patients Hospitalized for Heart Failure: Comparison of Machine Learning and Other Statistical Approaches. *JAMA Cardiol.* 2017;2(2):204-209.
17. Kessler RC, van Loo HM, Wardenaar KJ, et al. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Mol Psychiatry.* 2016;21(10):1366-71.
18. Kessler RC, Warner CH, Ivany C, et al. Predicting suicides after psychiatric hospitalization in US Army soldiers: the Army Study To Assess Risk and Resilience in Service members (Army STARRS). *JAMA Psychiatry.* 2015;72(1):49-57.
19. van der Post LF, Peen J, Dekker JJ. A prediction model for the incidence of civil detention for crisis patients with psychiatric illnesses; the Amsterdam study of acute psychiatry VII. *Soc Psychiatry Psychiatr Epidemiol.* 2014;49(2):283-90.
20. van der Post LF, Schoevers R, Koppelmans V, et al. The Amsterdam Studies of Acute Psychiatry I (ASAP-I); a prospective cohort study of determinants and outcome of coercive versus voluntary treatment interventions in a metropolitan area. *BMC Psychiatry.* 2008;8:35.
21. Lyons JS, Colletta J, Devens M, Finkel SI. Validity of the Severity of Psychiatric Illness rating scale in a sample of inpatients on a psychogeriatric unit. *Int Psychogeriatr.* 1995;7(3):407-16.
22. van der Post LF, Mulder CL, Bernardt CM, Schoevers RA, Beekman AT, Dekker J. Involuntary admission of emergency psychiatric patients: report from the Amsterdam Study of Acute Psychiatry. *Psychiatr Serv.* 2009;60(11):1543-6.
23. European Parliament. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). URL: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>. Accessed April 29, 2019.
24. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw.* 2011;45(3):1–67.
25. Rose S. Machine Learning for Prediction in Electronic Health Data. *JAMA Netw Open.* 2018;1(4):e181404.
26. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proc Int JointConf Artif Intell.* 1995;14(2):1137–1145.
27. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett.* 2006;27:861–874.
28. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Softw.* 2008;28(5):1-26.
29. Leening MJG, Vedder MM, Witteman JCM, Pencina MJ, Steyerberg EW. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. *Ann Intern Med.* 2014;160(2):122-131.
30. Artetxe A, Beristain A, Graña M. Predictive models for hospital readmission risk: A systematic review of methods. *Comput Methods Programs Biomed.* 2018;164:49-64.

31. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. JAMA. 2011;306(15):1688-98.

Tables

Table 1. Descriptive statistics for the 2084 patients in the first year after a psychiatric crisis care contact

		All participants (n=2084)	Hospitalized (n=710)	Not Hospitalized (n=1374)		
Variable		M (SD) n (%)	M (SD) n (%)	M (SD) n (%)	X ² (df)	p
Age	Years	40.8 (15.1)	41.0 (13.8)	40.7 (15.7)	0.94 (1)	0.33
Sex	Male	1083 (52.0%)	405 (57.0%)	678 (49.3%)	10.81 (1)	0.001
	Female	1001 (48.0%)	305 (43.0%)	696 (50.7%)		
Diagnosis					120.2 (5)	<0.0001
	Psychotic	807 (38.7%)	373 (52.5%)	434 (31.6%)		
	Depressive	285 (13.7%)	98 (13.8%)	187 (13.6%)		
	Substance related	239 (11.5%)	84 (11.8%)	155 (11.3%)		
	Manic/Bipolar	34 (1.6%)	12 (1.7%)	22 (1.6%)		
	Other	561 (26.9%)	103 (14.5%)	458 (33.3%)		
	No or deferred	158 (7.6%)	40 (5.6%)	118 (8.6%)		
Living situation					35.35 (5)	<0.0001
	Alone	1018 (48.8%)	385 (54.2%)	633 (46.1%)		
	With partner/other(s)	564 (27.1%)	142 (20.0%)	422 (30.7%)		
	With parents	235 (11.3%)	73 (10.3%)	162 (11.8%)		
	Homeless	96 (4.6%)	42 (5.9%)	54 (3.9%)		
	Institutionalized	68 (3.3%)	31 (4.4%)	37 (2.7%)		
	Other	103 (4.9%)	37 (5.2%)	66 (4.8%)		
Cultural background					12.16 (5)	0.033
	Dutch	1151 (55.2%)	409 (57.6%)	742 (54.0%)		
	Surinamese / Antilles	303 (14.5%)	124 (17.5%)	189 (13.8%)		
	Moroccan	145 (7.0%)	44 (6.2%)	101 (7.4%)		
	Turkish	82 (4.0%)	22 (3.1%)	60 (4.4%)		
	Other non- western	243 (11.7%)	78 (11.0%)	165 (12.0%)		
	Other western	160 (7.7%)	43 (6.1%)	117 (8.5%)		

Table 2. Key performance statistics of the machine learning based models

ML algorithm	AUC	Sensitivity	Specificity	Accuracy	P-value Acc.
Gradient Boosting	0.774	0.455	0.894	0.744	<0.0001
Oblique Random Forest	0.762	0.509	0.847	0.732	<0.0001
DeepBoost	0.760	0.461	0.871	0.731	<0.0001
Random Forest	0.757	0.478	0.864	0.732	<0.0001
GLM (Logistic Regression)	0.756	0.444	0.876	0.729	<0.0001
Support Vector Machines	0.751	0.370	0.917	0.731	<0.0001
Naive Bayes	0.751	0.455	0.861	0.723	<0.0001
Neural Network	0.749	0.528	0.828	0.726	<0.0001
Keras / TensorFlow	0.741	0.465	0.850	0.719	<0.0001
K-nearest Neighbors	0.702	0.356	0.879	0.701	<0.0001

Figures

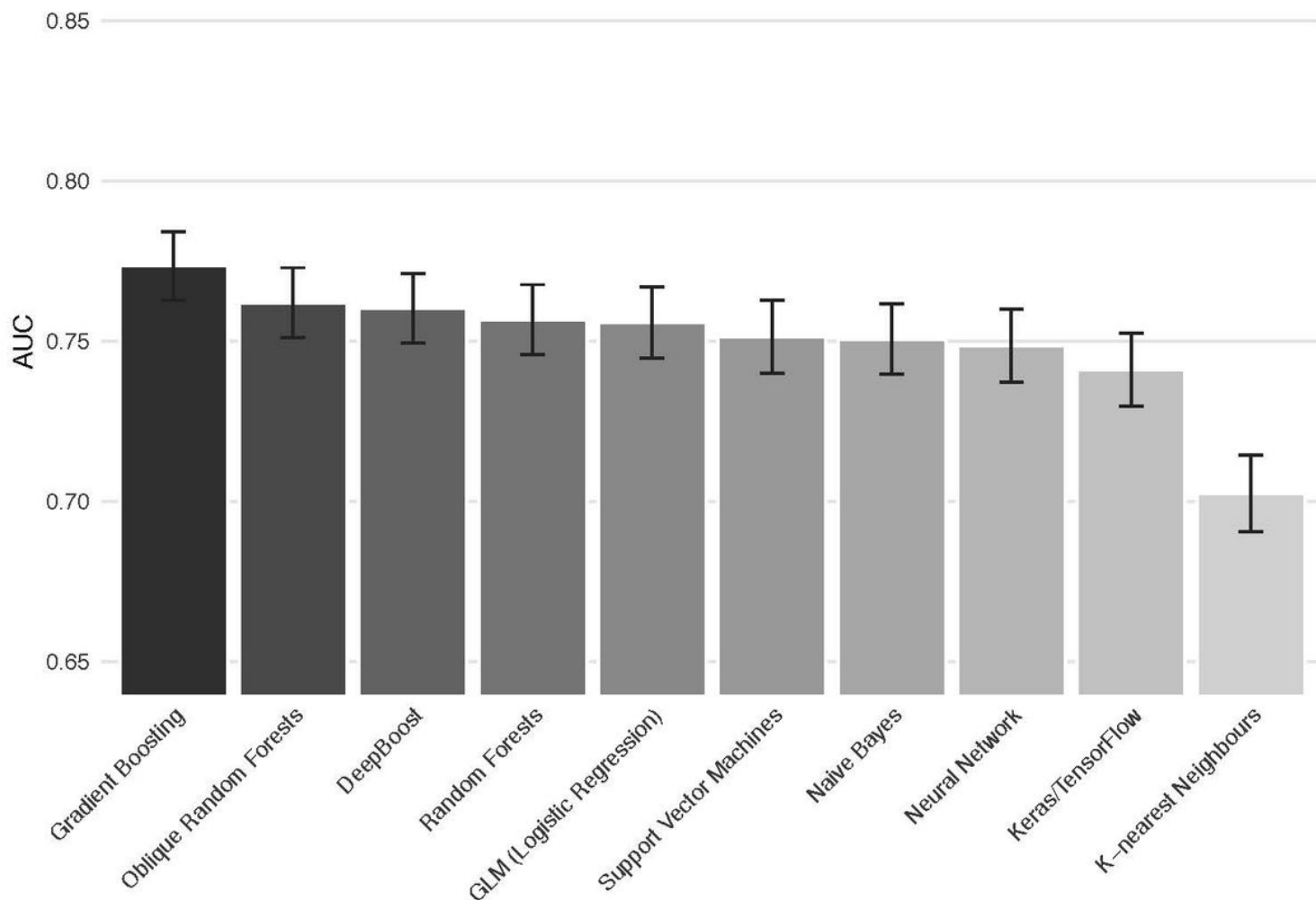


Figure 1

Comparison of AUC scores for the ten machine learning based models Note. AUC (or c-statistic) indicates the performance of the different machine learning based models. The error bars indicate +/- 1 standard error intervals.

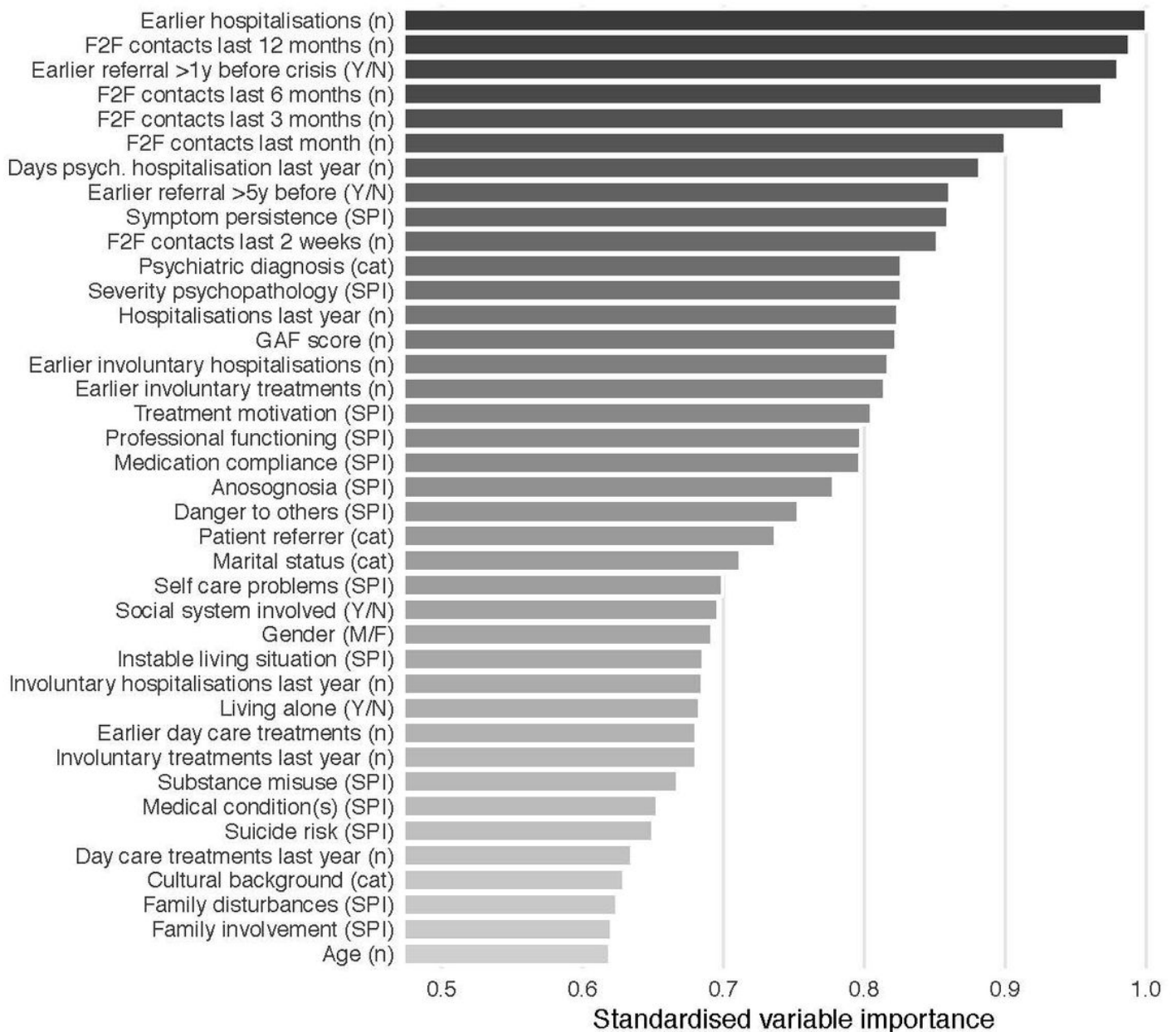


Figure 2

Overall variable importance plot for the machine learning based models Note. This plot presents the 39 predictors (before dummy-recoding) in descending order of unique predictive value. (n) indicates a numeric variable, (cat) indicates the variable is categorical, (SPI) indicates the variable is part of the SPI instrument, (M/F) and (Y/N) variables are dichotomous. Psychiatric care register data have a 5-year time horizon unless otherwise indicated.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [190719Additionalfile3.docx](#)

- [190719Additionalfile2.xlsx](#)
- [190719Additionalfile1.docx](#)