Predicting high intensity users in mental healthcare

Fellow: James Lai[1, 2, 3]

Supervisor: Prof Rob Stewart[1, 2]

[1]South London and Maudsley Hospital; [2]Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London; [3]Emergency Department, Royal Free London NHS Trust

AI CENTRE for Value Based Healthcare

NHS South London and Maudsley NHS Foundation Trust

Background

Mental healthcare is costly with estimates around £22.5 billion per year.

Previous work used routinely recorded data captured in electronic healthcare records (EHR) to predict service outcomes. These models were trained on a single year of patient data and tested on the subsequent year. The best performing model was able to predict high intensity service use following first referral to mental healthcare.

This project aims to expand on the previous work, to develop classification models using supervised machine learning to predict high service use at 12 months following first presentation to mental health services and evaluate if model performance remains accurate over time.

Methods

This project was carried out in the South London & Maudsley NHS Foundation Trust (SLaM). Electronic records have been in use since 2006 and anonymised patient data were extracted using Clinical Record Interactive Search (CRIS). We selected patients aged 18 years and over following their first presentation to SLaM services

Patient demographics, 90-day risk assessment, service use and symptom scores generated using natural language processing were used as predictor variables to determine high cost care, defined as cases receiving the top decile of service use.

Training data from 2007 – 2011 were used to develop logistic regression, decision tree, random forest and boosted ensemble models using cross validation. Model development was performed using the tidymodels package in R. Predictive models underwent hyperparameter tuning with the best performing specification used for the final model.

The model performance was evaluated using receiver operating characteristic (ROC) curves, calculating the area under the curve (AUC). AUCs were obtained for the training set, a test set from the same year and data from 2012 – 2017 for external validation.

Findings

36,300 patients were used to train, test and validate four predictive binary classification models (Table 1). These models were developed using logistic regression, decision tree, random forest and ensemble boosted supervised machine learning. Model performance was evaluated using AUC (Fig. 1, Table. 2), with similar performance across model types and between test and future validation cohorts.

	Train (2007 – 2011) N = 13540	Test (2007 – 2011) N = 5803	Future (2012 – 2017) N = 16957
Age (IQR)	42 (29-65)	42 (29-67)	41 (28-65)
Female (%)	5845 (43.1)	2491 (42.9)	7262 (42.8)
Lives alone (%)	2044 (15.1)	889 (15.3)	2399 (14.1)
White (%)	9665 (71.3)	4114 (70.8)	9749 (57.4)
Black (%)	1720 (12.7)	719 (12.4)	1980 (11.6)
Asian (%)	816 (6.0)	389 (6.7)	1059 (6.2)
Risk (%)	730 (5.4)	309 (5.3)	2770 (16.3)
High intensity user (%)	1345 (9.9)	604 (10.4)	1723 (10.1)

Table 1. Patient demographics across the training, test and future validation cohorts. Data represented as median values with interquartile range (IQR) or as absolute counts and percentages.

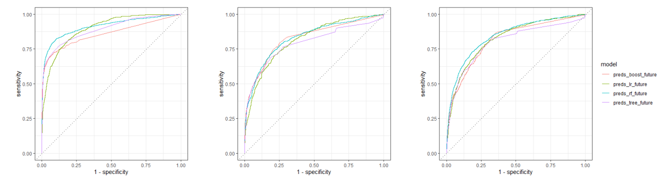

Figure 1. Receiver operating characteristic (ROC) curves for model performance in training (left), test (middle) and future validation (right) cohorts. Models described include: Logistic regression (green); decision tree (purple); random forest (blue); and boosted ensemble (red)

	Train (2007 – 2011) N = 13540	Test (2007 – 2011) N = 5803	Future (2012 – 2017) N = 16957
Logistic Regression	0.89	0.81	0.82
Decision Tree	0.89	0.79	0.79
Random Forest	0.91	0.82	0.84
Boosted Ensemble	0.85	0.82	0.81

Table 2. Binary classification models were developed using logistic regression, decision tree, random forest and boosted ensemble methods. In- and out-of sample performance are measured using AUC on the 2007-2011 cohort, along with predictive performance on a later validation cohort.

Conclusion

- We demonstrate that routinely collected patient data held in EHRs can be used to train models to predict high intensity users of mental healthcare with similar performance to previous work.
- Four models were trained to identify high service use demonstrating similar performance across model types.
- Discriminatory performance remains constant between the test and future validation cohorts, suggesting good longitudinal model performance
- Future work is needed to integrate the use of predictive algorithms into clinical workflows

KING'S College LONDON

NHS Digital Academy

NHS Guy's and St Thomas' NHS Foundation Trust

References:

Colling et al. Predicting high-cost care in a mental health setting. BJPsych Open. 2020