Predicting the Future of Mental Health Care: A Data-Driven Tool for Spotting At-Risk Patients Earlier

Using Electronic Health Records (EHRs) to Forecast High-Intensity Service Use in the NHS

Dr Bharadwaj V. Chada, Professor Robert Stewart, Dr James Lai

bharadwaj.chada1@nhs.net

Introduction

Mental health services represent a significant and growing component of healthcare systems, with approximately 1 in 6 adults affected by mental health conditions, and 14% of healthcare expenditure attributed to mental health care. Early identification and risk stratification are crucial for improving outcomes, yet predictive modelling in routine mental health care remains underutilised. Colling et al. (2020) explored three potential outcomes for predictive modelling in this context, ultimately finding that high-intensity service use was the most effective target for prediction, outperforming the other outcomes of length of stay and readmission risk. Despite the growing availability of EHR data, its potential for predictive modelling in mental health has yet to be fully realised, highlighting a critical opportunity for further research and implementation.

Building on the earlier work of Colling et al, this project aimed to evaluate the performance of a predictive algorithm targeted at adults 3 months after their first presentation to mental health services, and sought to identify those at risk of requiring high-intensity mental health care over the subsequent 12-month period.

Methods

Setting

This analysis was conducted at the **South London and Maudsley (SLaM) NHS Foundation Trust**, a specialist mental health care provider serving a catchment population of 1.3 million residents in the South London Boroughs of Lambeth, Southwark, Lewisham, and Croydon. Data were extracted using SLaM's **Clinical Record Interactive Search (CRIS) platform**, which was first developed in 2008, and permits the real-time retrieval of anonymised patient information from the mental health EHR.

Index date, predictors, and outcome

The index date was defined as 3 months after a patient's initial assessment at SLaM. Predictions were made on the index date, with the aim of forecasting high-intensity service use over the subsequent 12 months.

The predictor variables examined in this project included:

- **Demographic factors** (e.g. a patient's ethnicity, age at first assessment, gender, and marital status)
- **Service user factors** (e.g. the number of teams a patient was under, the number of inpatient admissions in the 3 months after first assessment, as well as their total number of inpatient and community contact days during this time)
- **Clinical factors** (e.g. a patient's diagnosis, and recorded risk of suicide, violence, and neglect).
- **Meta-data from pre-existing natural language processing (NLP) algorithms were considered**, including routinely extracted mentions of over fifty symptoms from the EHR.

High-intensity service use was defined as the top decile of service costs, calculated using the University of Kent compendium of unit costs, and a patient's total number of service use days. The primary outcome is binary, classifying individuals as either high-cost (falling within the top decile of healthcare expenditure) or not.

Statistical analysis and predictive model building

- The prediction model was **trained and tested on patients first assessed between 2007-11** (n = 18,869 patients). **Subsequent validations took place in the time periods 2012-17, and 2018-2023**.
- **The default preferred prediction approach was a logistic regression model**. The rationale for this choice was based on its interpretability and efficiency. Alternative models would only be considered if they showed demonstrably superior performance.
- **NLP-derived variables were included in initial models but later omitted after performance comparisons**. Despite moderate gains in sensitivity and AUROC, the limited availability of text mining capabilities may limit the use of the model when attempting to deploy elsewhere.
- **The final model was further compared against other predictive modelling techniques**, namely XGBoost, Random Forest, and Decision Tree, developed using the same feature set on data between 2007 – 2011.

Patient and Public Involvement and Engagement (PPIE)

As part of this work, a PPIE session was conducted in partnership with the **National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre (BRC)**. A committee of service users and expert lay members were provided with an overview of this project, and their feedback was obtained regarding the perceived clinical utility of this work, foreseeable barriers to implementation, as well as suggestions for the direction of travel for this study.

Results

Model	AUROC	Sensitivity	Specificity
Model - NLP variables retained	0.794	0.828	0.581
Model - NLP variables omitted	0.786	0.820	0.536

Table 1: Comparison of model performance after omission of NLP-derived variables

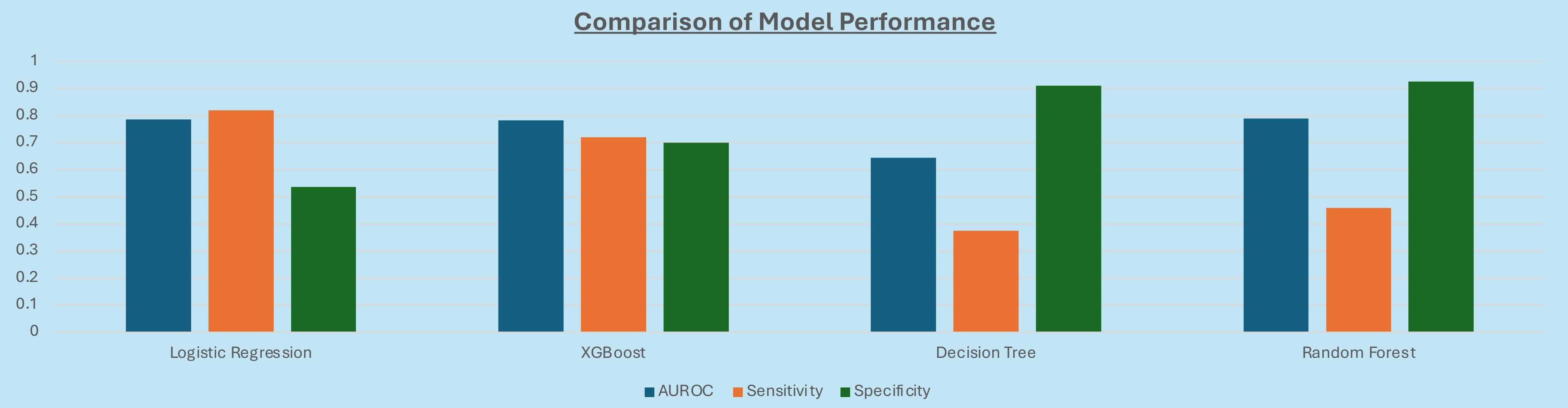

Figure 1: Bar chart comparing model performance for Logistic Regression, XGBoost, Decision Tree, and Random Forest Models.

Model	AUROC	Sensitivity	Specificity
Development Data (2007 - 2011)	0.786	0.820	0.536
Validation Data (2012 - 2017)	0.829	0.867	0.572
Validation Data (2018 - 2023)	0.815	0.795	0.669
Validation Data (2018 - 2023) – Covid-19 period omitted	0.815	0.784	0.679
Validation Data (2020 - 2021) – Covid-19 pandemic	0.813	0.819	0.646

Table 2: Comparison of model performance in the initial development data, as well as subsequent validations (including during the Covid-19 pandemic)

Discussion

This study aimed to evaluate the predictive performance of a logistic regression model for identifying high-intensity service users among mental health patients within SLaM. Our findings underscore the clinical utility of this approach, whilst also suggesting areas for further research:

Interpreting the Predictor Variables

Interpretation of predictor variables revealed several clinically relevant associations. As expected, a high volume of service use in the initial three months appeared to strongly predict continued high-intensity use over the following year. Additionally, the finding that conditions such as schizophrenia and eating disorders are associated with high service costs is consistent with existing literature on the management challenges and resource demands of these diagnoses. Conversely, conditions such as attention deficit hyperactivity disorder (ADHD) or autism spectrum disorders (ASD) were linked with relatively lower service costs.

Model Choice and Suitability of Logistic Regression

Logistic regression was selected as the primary model due to its balance of interpretability, efficiency, and robustness when handling many predictor variables. Compared to alternative techniques, logistic regression provided a clear understanding of the relationships between predictor variables and outcomes. In a clinical setting, where transparent decision-making is critical, the ability to directly interpret coefficients (e.g. quantifying the impact of high service use in the first three months) makes logistic regression particularly appealing.

The Role and Limitations of NLP-Derived Variables

Our analysis initially incorporated a range of NLP-derived variables. However, the additive value of these variables to the model's overall sensitivity and AUROC was limited. It is important to note, however, that despite their marginal contribution to the prediction metrics, NLP-derived features may have clinical utility – particularly in explaining misclassified cases.

Next Steps: Analysing Misclassified Cases

The analysis of false positive and false negative cases presents a valuable opportunity for refining the predictive model and clinical workflows. Future research should focus on a detailed review of false positive cases – patients flagged as high-intensity users at three months who do not sustain this level of service use at 12 months. This review may shed light on whether early interventions, such as care coordination or proactive clinician engagement, played a role in altering the clinical trajectory. Additionally, examining patients who were misclassified as low risk (false negatives) may help to identify factors or post-prediction events that contributed to their eventual high-intensity status.

Conclusion

In summary, this study demonstrates that a logistic regression model, leveraging a range of demographic, clinical, and service use predictors, can effectively identify mental health patients at risk of high-intensity service use. The model, validated across multiple time periods, including during the Covid-19 pandemic, achieved robust performance metrics (AUROC ranging from 0.79 to 0.83, with high sensitivity throughout), highlighting its potential clinical utility. These findings highlight critical areas for early intervention and suggest opportunities for further refinement and validation in diverse healthcare settings.