# Improving patient flow through hospitals with machine learning based discharge prediction

Jiandong Zhou[1], Andrew J Brent[1,2], David A. Clifton[3], A. Sarah Walker[1,4,5,6], David W. Eyre[2,4,5,6,7]

[1] Nuffield Department of Medicine, University of Oxford, Oxford, UK

[2] Oxford University Hospitals NHS Foundation Trust, Oxford, UK

[3] Department of Engineering Science, University of Oxford, Oxford, UK

[4] The National Institute for Health Research Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at the University of Oxford, Oxford, UK

[5] National Institute for Health Research Oxford Biomedical Research Centre, Oxford, UK

[6] Oxford Biomedical Research Centre, University of Oxford, Oxford, UK

[7] Big Data Institute, Nuffield Department of Population Health, University of Oxford, Oxford, UK

Corresponding author:

Prof David Eyre, david.eyre@bdi.ox.ac.uk

## Abstract

**Background**: Advanced analytics, underpinned by large-scale electronic health record (EHR) data, have the potential to transform the efficiency of healthcare delivery.

**Methods**: We used data from 4 hospitals in Oxfordshire, UK (01-February-2017 to 31-January-2020), to develop machine learning models for predicting hospital discharge in the next 24 hours, conditional on the duration of hospitalisation to date. We fitted separate XGBoost models for patients in hospital after planned and emergency admissions and for patients with different lengths of stay since admission. Models were trained and tested using data from 126,054 emergency and 45,609 planned admissions.

**Findings**: In held-out test data (01-February-2019 to 31-January-2020), individual patient discharge was predicted with positive and negative predictive values of 87·5% and 94·1% respectively in emergency patients (sensitivity 82·3%, specificity 98·5%, AUPRC 87·6%, AUC 87·5%) and 95·6% and 97·8% after a planned admission (sensitivity 83·5%, specificity 98·1%, AUPRC 90·5%, AUC 90·1%). Combining individual discharge probabilities allowed accurate estimates of the total number of discharges from hospital, mean average error 5·8% for emergency and 3·1% for planned patients. Previous hospital exposure, procedures, antibiotic use, and comorbidities were the most predictive features, but the most important features differed between planned and emergency patients, and over the course of admission. Performance was best in short-stay and planned admission patients, but generally robust over time, across subgroups, and different model training strategies.

**Interpretation**: Our approach could help improve flow of patients through hospitals, resulting in faster care delivery and patient recovery, and better use of healthcare resources.



**Funding:** National Institute for Health Research.

## Research in Context

***Evidence before this study***: There is unprecedented demand for healthcare services, and therefore significant interest in improving the efficiency of healthcare delivery. Prediction of hospital discharge could potentially facilitate more efficient and rapid delivery of care. We searched PubMed and Google Scholar using the terms ("patient flow" OR "hospital discharge" OR "bed state") AND ("machine learning" OR "decision support") to 31 March 2023. Despite successful applications of machine learning elsewhere in healthcare, there are relatively few implementations of machine learning based tools in operational management of healthcare. Previous studies are often limited to a specific context, e.g., one specialty, and often fail to demonstrate adequate levels of performance to be operationally useful.

***Added value of this study***: We describe an approach for accurately predicting hospital discharge within the next 24 hours that performs well across the whole hospital. We use established machine learning methods, but a novel modelling framework considering emergency and planned admissions separately, and fitting separate models conditional on the prior length of stay in the current admission. We provide accurate predictions of discharge within 24 hours for emergency patients, positive and negative predictive values of 87·5% and 94·1%, respectively, and 95·6% and 97·8% for planned admissions, in held-out test data. Combining individual discharge probabilities, we achieve highly accurate estimation of the total number of discharges from hospital over the next 24 hours, with mean average error of 5·8% and 3·1% for emergency and planned patients, respectively. Our method outperforms previous studies. We also provide detailed analyses exploring the optimal quantity and recency of training data needed, variations in predictive power throughout the day, performance across subgroups, and details of the key factors driving our predictions and how they vary across patient groups and during a hospital admission.

**Implications of all the available evidence**: Adoption of our approach has the potential to improve the flow of patients through hospitals, leading to more efficient care delivery, better patient experiences and enhanced utilization of healthcare resources.

## Introduction

Increasing demand for healthcare, arising from changing population demographics, more long-term illness, societal changes, and technological advances, places considerable pressure on available hospital resources.[1] This has been exacerbated by the COVID-19 pandemic, with large backlogs developing in patients awaiting urgent and routine care.[2] Since healthcare resources are finite, there is considerable interest in how healthcare services can become more efficient. One focus is optimising patient flow through hospitals, to improve patient experience, to avoid delays in investigations, diagnosis, treatment, recovery, and discharge, and to reduce costs.

Predicting the timing of hospital discharge has multiple benefits: preparations for discharge can be made (e.g., support at home initiated, discharge medication made ready, rehabilitation therapy planned),[3] length of stay (LOS) reduced,[4] and the availability of an empty bed anticipated and planned for, and overcrowding reduced.[5] In many hospitals, discharge predictions are made by the clinical team based on the patient's diagnosis and current status, and updated throughout the hospital stay. However, the reliability and completeness of these potentially subjective assessments may vary, they may not always be well captured in electronic healthcare record (EHR) systems, and so they often do not contribute substantially to efficiently running hospitals.[6] Therefore, automated predictions of LOS and discharge, coupled to downstream action, could potentially bring substantial operational benefits.

Multiple applications of EHR-based machine learning algorithms to patient flow have been described,[7,8] including predicting hospital admissions and locations,[9,10] hospital discharge volume,[11] readmissions,[12] prolonged hospital stay after cancer surgery,[13] simulation of patient flow in multiple healthcare units,[14] and prediction of patient readiness for discharge.[15,16] However, improvements in accuracy, generalisability, and a better understanding of model resilience over time are required for discharge predictions to become widely used in practice.

Here we developed and comprehensively evaluated a machine learning algorithm to predict discharge from hospital within the next 24 hours. We demonstrate a high level of performance, that has the potential to substantially improve operational management of hospitals.

## Methods

We aimed to predict hospital discharges within 24 hours of an index date and time for all patients currently in the hospital at this time (individual-level prediction). Model prediction probabilities were also aggregated to predict the total number of patients discharged within the next 24 hours following the index date time (hospital-level prediction). We adopt a pragmatic, operationally focused approach, predicting actual discharge, rather than simply readiness for discharge, and including patients who died as well as those who were discharged alive.

### Setting

We obtained deidentified electrical health record (EHR) data from Oxford University Hospitals (OUH) in Oxfordshire, United Kingdom, for adult patients (≥16 years) during the three-year period from 01-February-2017 to 31-January-2020. OUH comprises four teaching hospitals, with 1000 beds overall, and serves the population of Oxfordshire, United Kingdom (UK; 725,000 people) as well as providing specialist services to the surrounding region. Patients attending only as outpatients (e.g., those attending regular haemodialysis sessions), those <16 years at admission or under the care of

paediatrics, maternity and intensive care patients were excluded, the latter because discharge home within 24 hours was not expected in this group, however these patients were included once transferred to a 'step-down' ward.

## Model features

Domain knowledge and prior literature were used to determine which of the available features within the dataset were potentially informative for predicting patient discharge. Input features included patient demographics, comorbidities, details of prior hospital stays and readmissions in the last year, summaries of vital signs, laboratory test results, microbiology results, antibiotic prescriptions, radiology investigations, procedures, and the current main diagnosis (**Table S3**). Details on data pre-processing and feature selection are provided in the Supplement.

## Prediction task approach

We predicted hospital discharge events within 24 hours of an index date and time for all patients currently in the hospital as a supervised binary classification task. Each patient contributed once to the dataset per day they were in hospital. Given different features may be important in predicting discharge at different stages of a patient's stay, and may depend on the reason for admission, separate models were constructed for emergency and planned admissions, and for each of the following prior LOS in the current admission groups: <1 day, 2, 3, 4, 5, 6, 7-10, 11-13, 14-20, 21-27 and ≥28 days (2*11=22 models in total; **Fig. 1c**).

We used eXtreme gradient boosting (XGB) models to predict discharge within the next 24 hours (**Fig. 1d**;**Fig. S1**), after also testing classifiers based on CatBoost[17], TabNet[18], MLP[19], and logistic regression (**Fig. S2**;**Table S4**). Models were trained using data from the first two years of the study (01-February-2017 to 31-January-2019) and evaluated using data from the final year of the study (01-February-2019 to 31-January-2020). We used Bayesian hyperparameter optimization and SMOTETomek or SMOTEENN to account for class imbalance in training data (see Supplement).

## Performance assessment

We assessed individual patient-level model performance by comparing predicted discharge events within 24 hours from midnight with actual discharge events. We evaluated binary classification performance using area under the precision-recall curve (AUPRC), as this metric is robust to imbalanced data, as we had many more patient-days of data than discharge events. We also estimated area under receiver operating characteristic curve (AUROC) and sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) after selecting a model threshold for discharge probability that optimised F1-score. We focussed on PPV and NPV as these measures most closely relate to actual discharge decision making and are likely to be applicable to hospitals with similar daily discharge rates. We summarised the accuracy of predictions of the total number of patients discharged using mean absolute error (averaging over the 365 predictions in the test dataset), as a percentage of the mean number of discharges per day in the test data set (MAE%). We used the built-in feature importance with total gain approach in XGB to determine feature importance.

# Results

We used data from 01-February-2017 to 31-January-2020 on 126,054 emergency admissions (76,293 patients, 52% female, median (IQR) age 60·2 (37·1-75·8) years) with a median (IQR) LOS of 2·8 (1·2-7·0) days (**Table S1**)) and from 45,609 planned admissions (38,235 patients, 49% female, 62·0 (25·0-

83·6) years) with LOS 2·3 (1·3-5·1) days (**Table S2**)). The proportion of patients discharged from hospital within the next 24 hours decreased as prior LOS in the current admission increased (**Fig. 1a**), and varied between emergency and planned admissions, and by day of the week (**Fig. 1b**).

## Predicting individual patient discharge events within 24 hours

For emergency admissions, there were 537,917 patient-day observations (53,417 patients) and 82,334 (15·3%) discharge events in the training dataset, and 279,108 patient-day observations (31,112 patients) and 44,980 (16·1%) discharge events in the test dataset. Overall AUPRC for predicting discharge within 24 hours across the different prior LOS models was 87.6% (range by prior LOS 51·9-91·9%) and AUROC was 87·5% (**Fig. 2a-d**). Using a probability threshold for predicting discharge that optimised the F1-score in training data for each model, the overall PPV and NPV for emergency patients were 87·5% and 94·1% (sensitivity 82·3%, specificity 98·5%) (results were similar using a threshold that maximised accuracy).

Performance was better for planned admissions; there were 139,602 patient-day observations (26,912 patients) and 31,020 (22·2%) discharge events in the training dataset, and 63,464 patient-day observations (13,203 patients) and 14,682 (23·1%) discharge events in the test dataset. The overall AUPRC was 90·5% (range by prior LOS 56·9-95·9%) and AUROC 90·1%; PPV and NPV were 95·6% and 97·8% (sensitivity 83·5%, specificity 98·1%) (**Fig. S3a-b**).

Performance varied by LOS and was better for shorter prior LOS for both planned and emergency patients. For example, PPV and NPV for emergency admissions were 91·2% and 92·7% 1 day after admission, 76·5% and 95·1% 7 days after admission and 58·3% and 94·4% >28 days after admission.

## Predicting hospital-level daily discharge numbers

We calculated the total number of discharges expected in the next 24 hours amongst all emergency or planned admission patients in the hospital by summing the patient-level predicted discharge probabilities (**Fig. 2e**;**Fig. S3c-d;Fig. S4a-b**). The mean average error, as a percentage of the mean number of daily discharges across the whole test period (MAE%), was 5·8% and 3·1% for emergency and planned admissions respectively (percentage root mean square error [RMSE%] 6·6% and 3·6%). Similarly to individual patient-level predictions, hospital-wide predictions were more accurate for shorter prior LOS, and in planned compared to emergency admissions.

The predicted number of discharges accurately reflected weekly and calendar holiday fluctuations in discharge numbers for emergency patients (**Fig. 2f, g**) and planned patients (**Fig. 2h, i**), and performance was similar across the whole testing data period (**Fig. S5**).

## Feature importance

The most informative features for predicting discharge included those describing hospital exposure before the current admission and readmissions, comorbidities, procedures, antibiotic use, and prediction index date features (**Fig. 3;Fig. S6-7**). However, there were differences between emergency and planned admissions, with procedures and antibiotic use being more informative in emergency admissions and previous hospital exposure and index date features being more informative in planned admissions. Feature importance also varied as LOS increased. In emergency admissions initially, procedures, radiology procedures, readmissions, previous hospital exposure and index date features were the top five most important features, but comorbidities and antibiotic use became more important as LOS increased. Similarly, previous hospital exposure, index date features,

radiology procedures, and blood cultures were the four most important features in planned admissions with short LOS, whereas readmissions, index date features, comorbidities and antibiotics were the most important with longer LOS.

Interestingly, contingent on the other features in the models, values of vital signs and laboratory tests played a relatively minor role in predicting imminent discharge, despite a model structure that allowed extreme values of either to be considered.

## Subgroup performance

Relatively consistent performance was achieved across several subgroups, but performance varied across others (**Fig. 4**). PPV and NPV estimates were broadly similar by day of the week, sex, and ethnicity. However, there was greater variation by age, deprivation, and comorbidity in particular, with somewhat poorer performance amongst those >80 years, with more comorbidities and living in areas with greater deprivation. These patients were more likely to stay longer in hospital; in exploratory analyses restricting only to those with a final LOS ≤7 days, differences generally attenuated in the largest subgroups (**Fig. S8**).

## Training dataset size and recency impact prediction performance

Fixing the test dataset to the 12 months from 01-February-2019, we evaluated the performance of models trained with 1-24 months prior data, i.e., only January-2019 through to January-2019 back to February-2017 inclusive (**Fig. 5a**). Summary metrics for individual patient-level and hospital-wide performance all improved as the number of months of training data increased, largely plateauing after approximately 13 months. These saturation effects suggest that the 24 months' training data considered in the main analysis was more than sufficient to achieve optimal performance.

Mimicking implementation in the real-world, we also considered the impact of decreasing recency of training data (**Fig. 5b**). We used the same fixed test dataset, but only 12 months of training data varying from being as recent as possible to up to 11 months 'out of date'. Training data ending within the last six months produced the best performance, with performance decreasing as training data age increased, e.g., in emergency admissions PPV with the most recent training data was 86·0% versus 75·5% when it was 11 months out-of-date, NPV 90·9% versus 81·3%, and MAE% versus 6·9% and 11·4%, respectively.

## Index date decision time

Performance estimates to this point have used predictions made at midnight for the next 24 hours. However, in real-world use, predictions could be updated throughout the day as new patients are admitted and others discharged (**Fig. S9**). We investigated performance at different times of day using three training strategies: 1) fitting separate models for each time of day a prediction was required, e.g., 24 separate models for each hour of the day; 2) training the model with data from midnight only; 3) training the model with data randomly sampled from throughout the 24-hour period. Differences between the three strategies were relatively modest (**Fig. S10**), however index date-time specific models performed best, improving PPV, NPV and MAE% compared to models trained with data at midnight in absolute terms by around 2%, 1%, and ≤0·3% at most time points for emergency patients, and 2%, 2%, ≤0·2% for planned patients respectively. Performance gains need to be offset against greater computational (and storage) burdens for multiple time-of-day specific models. Unexpectedly models trained with data sampled equally throughout the day performed

216 somewhat worse, potentially as data from less informative times of the day contributed more to
217 training.

## 218 Impact of model features

219 We evaluated if using only the most informative features maintained performance (**Fig. S11**).
220 Feature categories (with multiple individual features included within each) were ranked by the mean
221 importance of their individual features. Including the top six categories increased model
222 performance markedly; smaller, but potentially important, improvements continued to be made as
223 the number of feature categories used increased up to 21; however, this was at the expense of
224 increased model training time.

225 When initially conceived, we included binary variables for the postcode districts containing each
226 patient's home address (typically approximately 25,000 people) to reflect variation in population
227 characteristics and delivery of social support services by area (**Table S5**). However, this limits
228 generalisability. Alternative models with summary features for LOS by location/postcode district
229 (including total number of episodes, and mean, standard deviation, median, 25% and 75% percentile
230 of LOS in the training dataset) (**Fig. S12;Table S6-7**) produced comparable performance (**Fig. S13-14**),
231 whereas performance without location data was somewhat worse (**Fig. S11**).

## 232 Performance by discharge destination

233 We used a pragmatic and operationally relevant endpoint, predicting discharge from hospital with
234 any outcome. When we analysed subgroups of patients discharge alive, those who died in hospital,
235 and those who were transferred to another hospital, performance was marginally better for patients
236 discharged alive (**Table S8**). For example, the MAE% values for total numbers of discharges per day
237 in emergency patients were 5·7%, 6·2%, and 6·4% for these three groups respectively, and 3·1%,
238 3·4% and 4·3% for planned patients.

# 239 Discussion

240 Machine learning underpinned by large-scale EHR data has the potential to transform how
241 healthcare is delivered; but applications to operational management of hospitals are largely
242 unexplored currently.[20] By careful feature curation and a novel model design, we were able to
243 accurately predict patient discharge events within the next 24 hours across a whole hospital and
244 wide range of patient subgroups. We predict the total number of discharges each day following a
245 planned admission with a mean average error of only 3·1% and following an emergency admission
246 with only 5·8% error. We were also able to make accurate predictions for individual patients: PPV
247 and NPV following a planned admission were 94·1% and 97·8%, respectively, and 87·5% and 95·6%
248 after an emergency admission.

249 Our models have potential to be applied to improve the efficiency of patient flows through hospital,
250 resulting in faster care delivery and patient recovery, and better use of healthcare resources.
251 Availability of beds and timely patient discharge are major factors underlying patient flow
252 problems,[21] but most current initiatives focus on static procedures, e.g., 'discharge by noon', or
253 respond only to critical levels of high demand,[22] and there is limited translation of automated
254 discharge prediction models into clinical practice.[23]

255 We build substantially on the prior studies, that have often focused only on discharge from specific
256 settings, e.g., surgical or medical inpatients. We exceed the performance of an earlier simpler model

using data from our hospital group.[24] Several factors underlie the performance we achieve: we fit different models by prior length of stay and for emergency and planned admissions; and we consider a wide range of input features, jointly selected and curated by experienced clinicians, statisticians, and machine learning experts. We account for individual patient factors, both immediate and longer-term and also consider hospital-wide factors including historical length of stay for specific conditions. We also conducted careful analyses of the optimal testing dataset size and composition, feature selection, data balancing, and hyper parameter choices. We aggregate patient-level probabilities into more precise predictions of daily discharge numbers at hospital-level, in contrast to other approaches modelling total discharges for the entire population[25] or combining binary patient-level predictions (e.g., admission/discharge yes/no).[26]

Explaining how our models work, and supporting their implementation, we describe which features contribute to our predictions and show that these differ by admission type and over a hospital admission, underlining why we fit separate models by LOS and for emergency and planned patients. This means that elective surgical admissions are largely modelled separately, however future work might consider if performance for emergency medical and emergency surgical patients could be improved by using separate models for these patient groups too.

We achieved the best performance for short-stay patients, where factors related to their active treatment and response are important in determining discharge and are relatively well captured by the data we use. In contrast, for longer-stay patients, and especially those >80 years, with more comorbidities, and with increasing socioeconomic deprivation, gaps in available data led to less accurate predictions. This highlights the importance of extending the data types available to improve performance, e.g., assessments of functional state (often documented only as free text and not available in our data) and external factors (such as availability of social care to support discharge).

We find that values of vital signs and laboratory tests played a relatively minor role in predicting imminent discharge, contingent on the other model features. This potentially reflects that we trained our models to pragmatically predict actual discharge, rather than readiness for discharge.

Performance became better with increasing training data size, but saturated with around 13 months' training data, finishing no more than 6 months before the test dataset, suggesting that training could be undertaken without excessive historical data and updated two or three times a year in a real-world application. Additionally, differences between the time of day that models were trained and tested on had relatively modest impacts, however where computational resources allow optimal performance could be achieved by using models tuned to specific times of day.

Sequentially including more of the most predictive features (e.g., previous hospital exposure, readmissions, comorbidities, and procedures, etc) significantly increased model performance; however, improvements in model performance were smaller when further including more features that were less important (e.g., vital signs, admission specialty, etc). Trade-offs between performance and computational time may dictate the number of features hospitals choose to consider. This may also motivate hospitals to pay more attention to the data collection quality of the subset of key features that are most predictive to achieve accurate predictions most efficiently.

Limitations of our study include that we focus on a relatively short prediction horizon of 24 hours, and so our models as described could only support planning and interventions over the same timescale. However, our approach could be adapted to make longer-range predictions. We also used

diagnosis categories derived from ICD10 codes for training and test predictions, which are only recorded at discharge, however in reality the primary working diagnosis is known in real-time to clinicians, and could be used if documented electronically. We only had specific structured EHR data available and not data recorded as unstructured free text or in other forms. We therefore lacked data on frailty and functional assessments, therapist input, social care provision, and discharge planning that would almost certainly have improved prediction further, particularly for longer stay patients. We also do not capture directly some input features on the status of the hospital which could have improved performance, e.g., percentage of occupied beds (as numbers of available beds were not available in our dataset). We also did not evaluate how model performance changed with the degree of operational pressure the hospital was under, which could be considered in future as accurate predictions are arguably most helpful when resources are most constrained.

We adopted a relatively simple, pragmatic approach to feature engineering, using summaries of time series, such as vital signs, but potentially performance could be improved by better representing these dynamic data in future work. Similarly, alternatives to our feature selection approach could be tested, where multiple complex input features could be summarised through latent representations such as *t*-SNE[27] or autoencoders[28]. Deep learning architectures were not investigated, but could also potentially improve performance and allow more flexible updating, e.g., as new data is obtained over time or in different settings. Another extension would be to generate estimates of the uncertainty in our predictions, rather than just point estimates. Finally, we used temporal external validation rather than data from a completely different hospital group. Future work should include validation with data from other settings, including prospective *in silico* validations and real-world clinical evaluations of the impact that deploying a model would have on operational and patient-level outcomes.

In conclusion, we show using a novel model design and feature set how combining machine learning modelling approaches with EHR data could be used to inform real-time operational management of hospitals, with realistic requirements for training data and model updating. Our findings have the potential to be deployed to help improve efficiency of hospitals and to help patients recover and get home faster.

## Data availability

The datasets analysed during the current study are not publicly available as they contain personal data but are available from the Infections in Oxfordshire Research Database (https://oxfordbrc.nihr.ac.uk/ research-themes-overview/antimicrobial- resistance-and-modernising-microbiology/ infections-in-oxfordshire-research-database-iord/), subject to an application and research proposal meeting the ethical and governance requirements of the Database. For further details on how to apply for access to the data and for a research proposal template please email iord@ndm.ox.ac.uk.

## Code availability

A copy of the analysis code is available at https://github.com/jadonzhou/PatientDischargePrediction

## Ethics Committee Approval

Deidentified data were obtained from the Infections in Oxfordshire Research Database which has ethical approvals from the National Research Ethics Service South Central – Oxford C Research Ethics Committee (19/SC/0403), the Health Research Authority and the national Confidentiality Advisory Group (19/CAG/0144), including provision for use of pseudonymised routinely collected data without individual patient consent. Patients who choose to opt out of their data being used in research are not included in the study. The study was carried out in accordance with all relevant guidelines and regulations.

## Funding

This study was funded by the National Institute for Health Research (NIHR) Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at Oxford University in partnership with the UK Health Security Agency (UKHSA) and the NIHR Biomedical Research Centre, Oxford. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, the Department of Health and Social Care or the UKHSA. DAC was supported in part by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC); the InnoHK Hong Kong Centre for Cerebro-cardiovascular Health Engineering; the Pandemic Sciences Institute at the University of Oxford; an NIHR Research Professorship and a Royal Academy of Engineering Research Chair. DWE is a Robertson Foundation Fellow. The funders had no role in in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

## Acknowledgements

We thank all the people of Oxfordshire who contribute to the Infections in Oxfordshire Research Database. Research Database Team: L Butcher, H Boseley, C Crichton, DW Crook, DW Eyre, O Freeman, J Gearing (community), R Harrington, K Jeffery, M Landray, A Pal, TEA Peto, TP Quan, J Robinson (community), J Sellors, B Shine, AS Walker, D Waller. Patient and Public Panel: G Blower, C Mancey, P McLoughlin, B Nichols. We would also like to thank Lisa Glynn and colleagues at Oxfordshire University Hospitals NHS Trust for helpful feedback and insights into early versions of the data analyses presented.

## Author Contributions

The study was designed and conceived by DWE, ASW, DAC and AJB. DWE and JZ curated the data. JZ analysed the data and created the visualisations. JZ wrote the first draft of the manuscript. All authors contributed to editing and revising the manuscript.

## Competing Interests Statement

DWE declares lecture fees from Gilead, outside the submitted work. DAC reports personal fees from Oxford University Innovation, personal fees from BioBeats, personal fees from Sensyne Health, outside the submitted work. No other author has a conflict of interest to declare.

## Figures

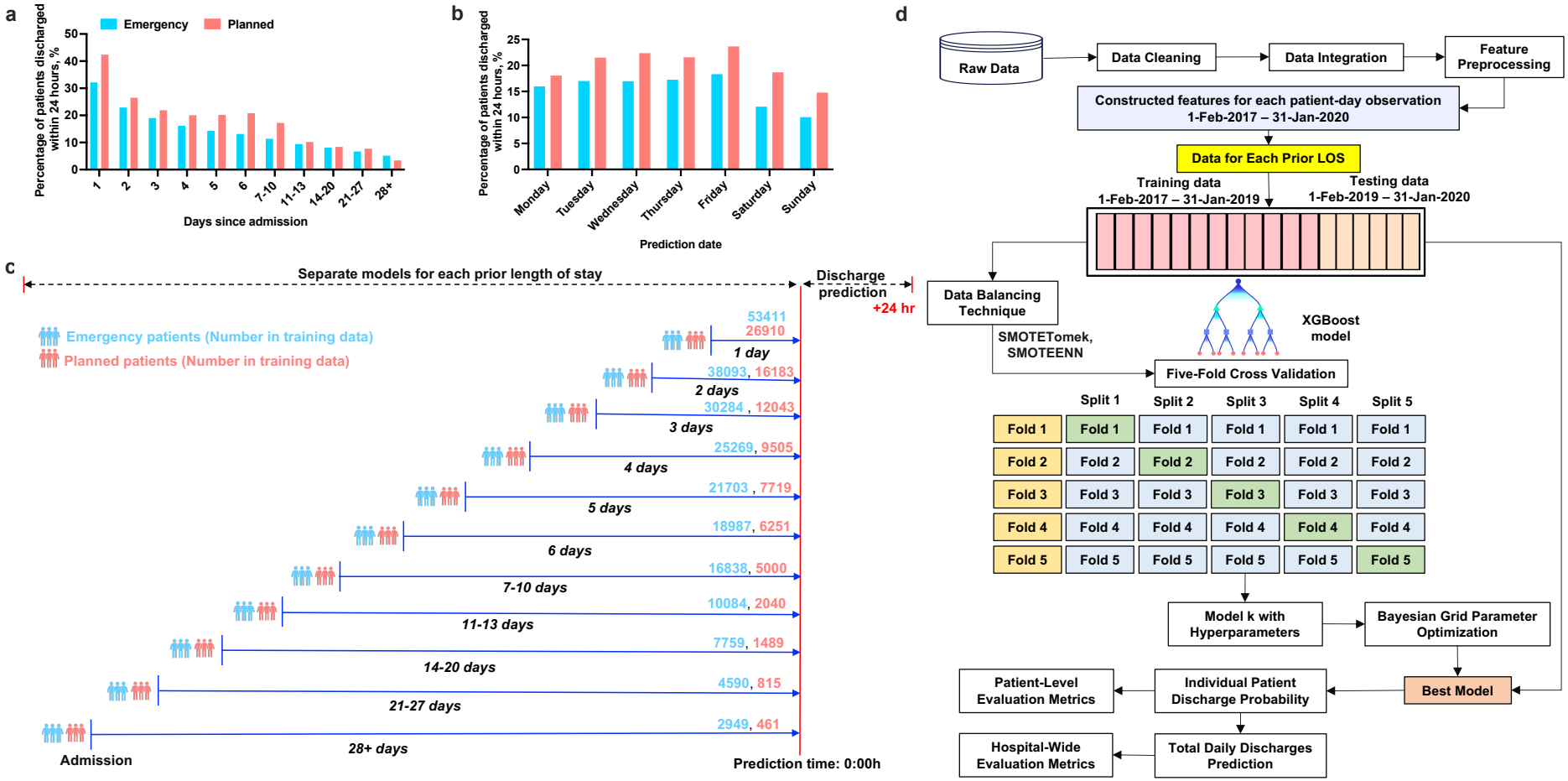


**Fig. 1. Overview of model development**. **a**, Percentage of patients discharged within the next 24 hours in emergency and planned patients by duration of previous hospital stay. **b**, Percentage of patients discharged by weekday. **c**, The binary prediction problem was defined by classifying the outcome as 'positive' (discharge occurred within the next 24 h) or 'negative' (discharge did not occur within the next 24h) separately for emergency patients and

planned patients with different durations of previous hospital stay. **d**, A binary classifier to predict imminent discharge within the next 24 hours was trained on the extracted labels and features from episodes with admission time between 1-Feb-2017 and 31-Jan-2019, and was tested on data between 1-Feb-2019 and 31-Jan-2020. XGBoost model was chosen as the classifier after comparison of different machine learning algorithms. The proposed prediction model consists of the trained binary classifier, five-fold cross validation, a data balance technique, Bayesian grid parameter search, and a decision probability threshold optimizing both precision and recall.

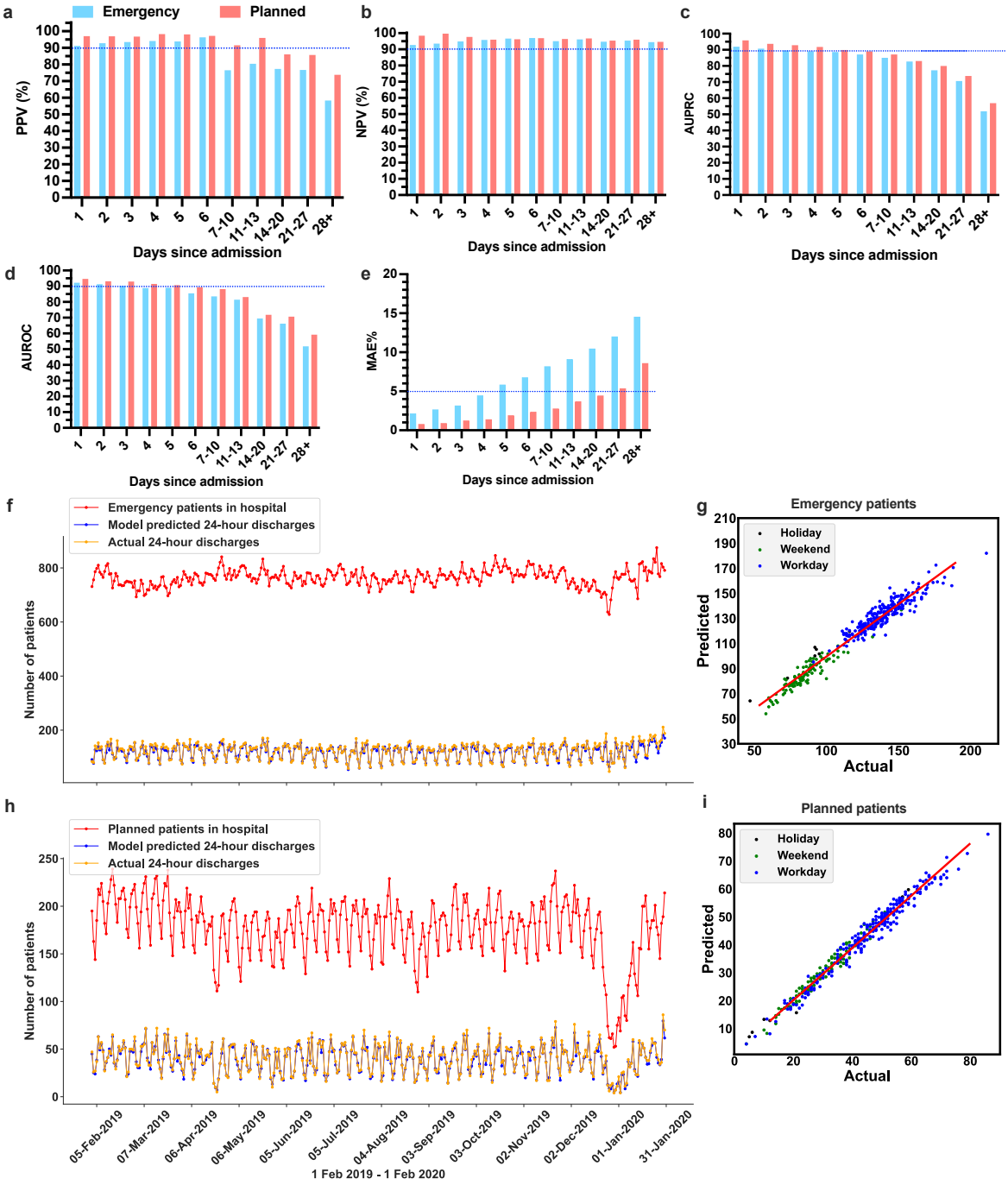


**Fig. 2. Summary of model performance in the test dataset (01-Feb-2019 to 31-Jan-2020)**. At the individual patient admission-level, (**a**) Positive predictive value, PPV (**b**) negative predictive value, NPV, (**c**) area under the precision recall curve, AUPRC, (**d**) area under the receiver operating characteristic curve, AUROC. At the hospital-level, (**e**) mean average error (relative) (MAE%) (**f**) (**g**) predicted and actual discharges within 24 hours, and number of patients in hospital, emergency admissions (**h**) (**i**) predicted and actual discharges, and number of patients in hospital, planned admissions. Horizontal lines are provided in panels **a-e** to aid visualisation.

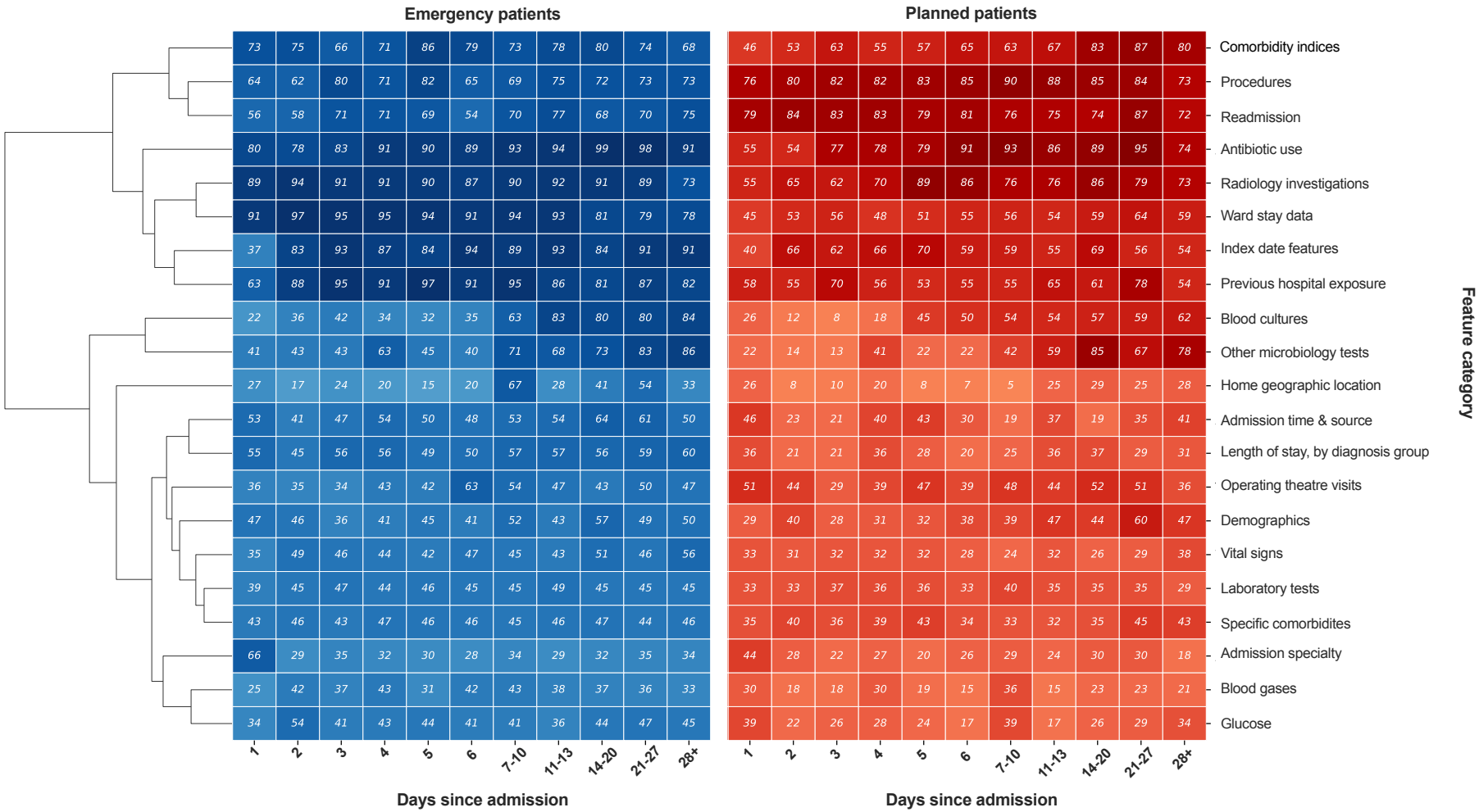


**Fig. 3. Feature importance**. Numbers show mean importance of the individual features within each feature category (higher numbers, with darker shading, represent greater importance), for each prior length of stay and emergency/planned admission model. Importance data was clustered using hierarchical agglomerative cluster analysis with equilibrious between-group average linkage method.

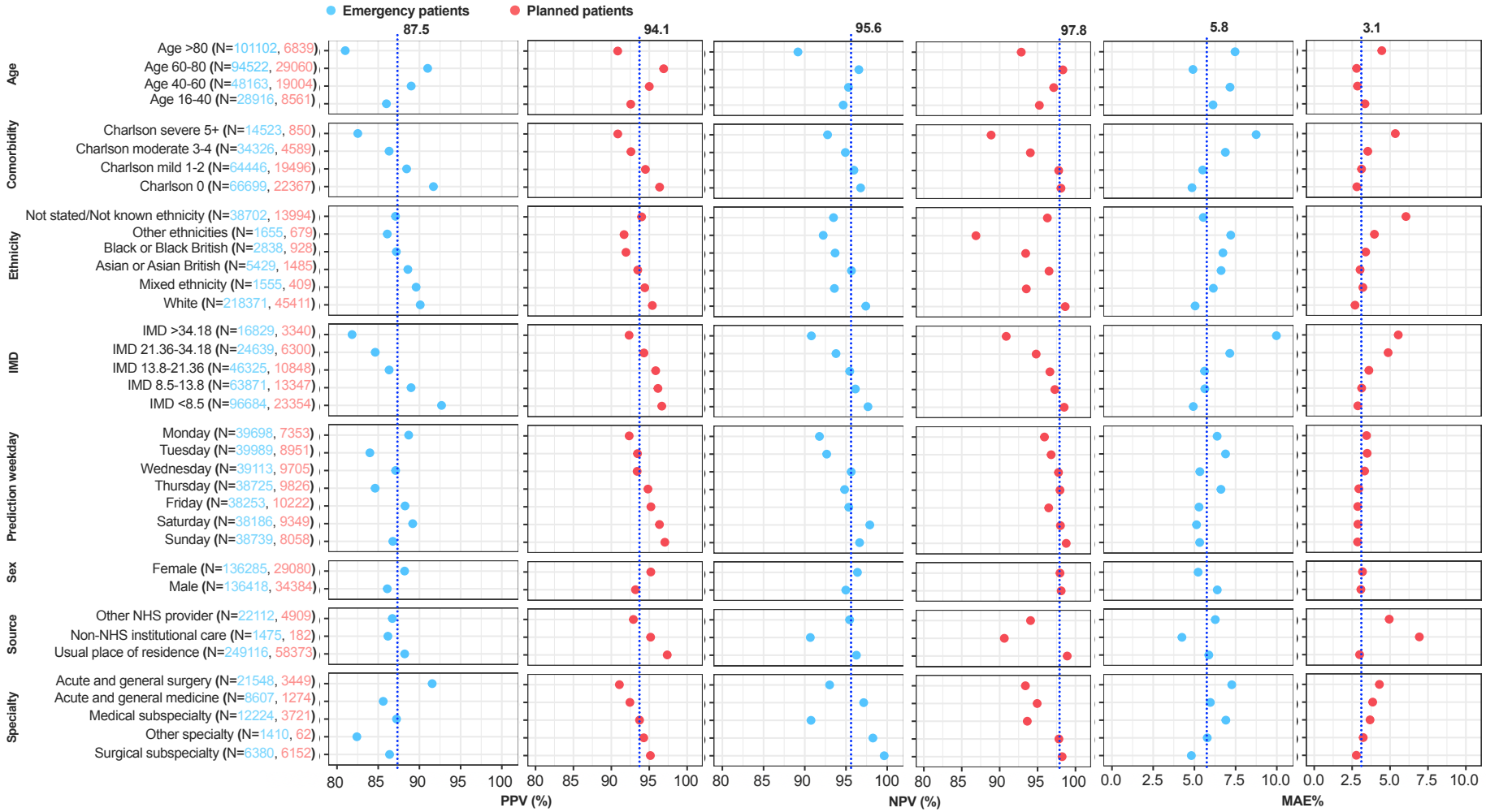


**Fig. 4. Model performance in various subgroups within the test dataset (01-Feb-2019 to 31-Jan-2020).** Positive predictive value (PPV) and negative predictive value (NPV) were used for patient-level discharge prediction performance evaluation, while mean absolute error (MAE%) was used for performance analysis of predicting discharge number at hospital-level. IMD=index of multiple deprivation score (higher scores indicate greater deprivation). Overall performance is shown by the dashed line in each plot.

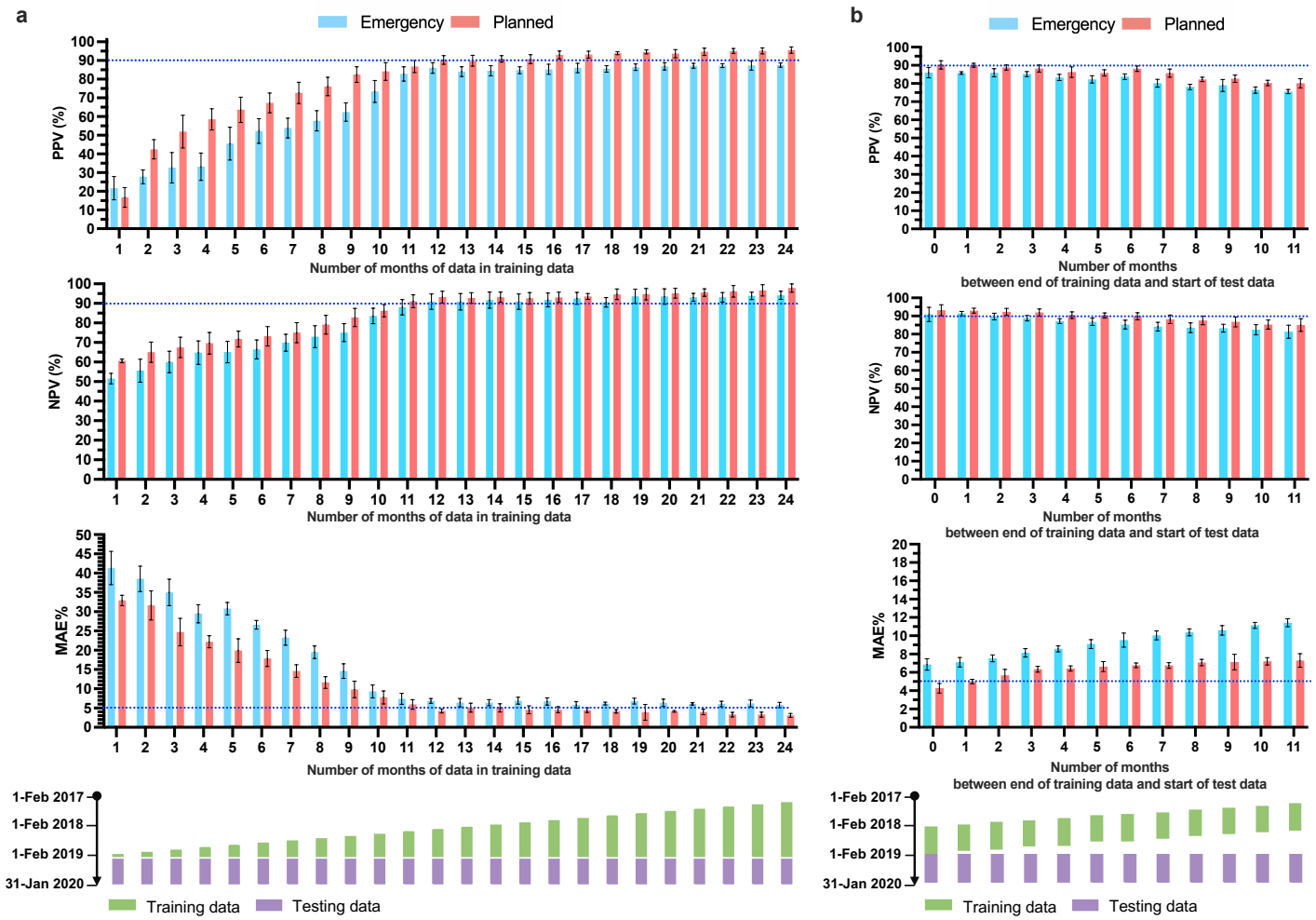


**Fig. 5. Impact of training dataset size and recency on discharge prediction performance on discharge prediction performance within the test dataset (01-Feb-2019 to 31-Jan-2020**). **a**, increasing training data size. **b**, decreasing recency of same-size training data. PPV=positive predictive value, NPV=negative predictive value, MAE%=mean absolute error (relative). Dashed blue lines indicate references of PPV=90%, NPV=90%, and MAE%=5%, respectively, to aid visualization.

## References


1. Kruk, M. E. *et al.* High-quality health systems in the Sustainable Development Goals era: time for a revolution. *Lancet Glob Health* **6**, e1196–e1252 (2018).

2. Arsenault, C. *et al.* COVID-19 and resilience of healthcare systems in ten countries. *Nat Med* **28**, 1314–1324 (2022).

3. Rodi, S. W., Grau, M. v & Orsini, C. M. Evaluation of a fast track unit: Alignment of resources and demand results in improved satisfaction and decreased length of stay for emergency department patients. *Quality Management in Healthcare* **15**, 163–170 (2006).

4. Li, C. *et al.* An enhanced recovery pathway decreases duration of stay after esophagectomy. *Surgery (United States)* **152**, 606–616 (2012).

5. Carratalà, J. *et al.* Effect of a 3-step critical pathway to reduce duration of intravenous antibiotic therapy and length of stay in community-acquired pneumonia: A randomized controlled trial. *Arch Intern Med* **172**, 922–928 (2012).

6. Levin, S. *et al.* Machine-learning-based hospital discharge predictions can support multidisciplinary rounds and decrease hospital length-of-stay. *BMJ Innov* **7**, 414–421 (2021).

7. Rajpurkar, P., Chen, E., Banerjee, O. & Topol, E. J. AI in health and medicine. *Nat Med* **28**, 31–38 (2022).

8. Bertsimas, D., Pauphilet, J., Stevens, J. & Tandon, M. Predicting inpatient flow at a major hospital using interpretable analytics. *Manufacturing & Service Operations Management* **24**, 2809–2824 (2022).

9. Ryu, A. J. *et al.* A clinician's guide to running custom machine-learning models in an electronic health record environment. *Mayo Clin Proc* **98**, 445–450 (2023).

10. Wellner, B. *et al.* Predicting unplanned transfers to the intensive care unit: A machine learning approach leveraging diverse clinical elements. *JMIR Med Inform* **5**, e45 (2017).

11. Feng, G., Yu, K., Wang, Y., Yuan, Y. & Djuric, P. M. Exploiting causality for improved prediction of patient volumes by Gaussian processes. *IEEE J Biomed Health Inform* **25**, 2487–2496 (2021).

12. Mišić, V. V., Rajaram, K. & Gabel, E. A simulation-based evaluation of machine learning models for clinical decision support: application and analysis using hospital readmission. *NPJ Digit Med* **4**, 98 (2021).

13. Jo, Y.-Y. *et al.* Prediction of prolonged length of hospital stay after cancer surgery using machine learning on electronic health records: Retrospective cross-sectional study. *JMIR Med Inform* **9**, e23147 (2021).

14. Kovalchuk, S. V., Funkner, A. A., Metsker, O. G. & Yakovlev, A. N. Simulation of patient flow in multiple healthcare units using process and data mining techniques for model identification. *J Biomed Inform* **82**, 128–142 (2018).

15. Safavi, K. C. *et al.* Development and Validation of a Machine Learning Model to Aid Discharge Processes for Inpatient Surgical Care. *JAMA Netw Open* **2**, e1917221 (2019).

16. Zhang, X., Yan, C., Malin, B. A., Patel, M. B. & Chen, Y. Predicting next-day discharge via electronic health record access logs. *Journal of the American Medical Informatics Association* **28**, 2670–2680 (2021).

17. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. & Gulin, A. CatBoost: unbiased boosting with categorical features. in *Advances in neural information processing systems* (2018).

18. Arik, S. Ö. & Pfister, T. TabNet: Attentive Interpretable Tabular Learning. *Proceedings of the AAAI Conference on Artificial Intelligence* **35**, 6679–6687 (2021).

19. Windeatt, T. Accuracy/Diversity and Ensemble MLP Classifier Design. *IEEE Trans Neural Netw* **17**, 1194–1211 (2006).

20. Pianykh, O. S. *et al.* Improving healthcare operations management with machine learning. *Nat Mach Intell* **2**, 266–273 (2020).

21. Hoot, N. R. & Aronsky, D. Systematic Review of Emergency Department Crowding: Causes, Effects, and Solutions. *Ann Emerg Med* **52**, 126-136.e1 (2008).

22. Viccellio, A., Santora, C., Singer, A. J., Thode, H. C. & Henry, M. C. The Association Between Transfer of Emergency Department Boarders to Inpatient Hallways and Mortality: A 4-Year Experience. *Ann Emerg Med* **54**, 487–491 (2009).

23. Seneviratne, M. G., Shah, N. H. & Chu, L. Bridging the implementation gap of machine learning in healthcare. *BMJ Innov* **6**, 45–47 (2020).

24. Bishop, J. A. *et al.* Improving patient flow during infectious disease outbreaks using machine learning for real-time prediction of patient readiness for discharge. *PLoS One* **16**, (2021).

25. van Walraven, C. & Forster, A. J. The TEND (Tomorrow's Expected Number of Discharges) model accurately predicted the number of patients who were discharged from the hospital the next day. *J Hosp Med* **13**, 158–163 (2018).

26. Sun, Y., Heng, B. H., Tay, S. Y. & Seow, E. Predicting Hospital Admissions at Emergency Department Triage Using Routine Administrative Data. *Academic Emergency Medicine* **18**, 844–850 (2011).

27. Laurens van der Maaten & Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research* **9**, 2579–2605 (2008).

28. Zhang, L., Qi, G.-J., Wang, L. & Luo, J. AET vs. AED: Unsupervised Representation Learning by Auto-Encoding Transformations Rather Than Data. in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 2542–2550 (IEEE, 2019). doi:10.1109/CVPR.2019.00265.