



OPEN Predicting onward care needs at admission to reduce discharge delay using explainable machine learning

Chris Duckworth^{1✉}, Dan Burns¹, Carlos Lamas Fernandez², Mark Wright³, Rachael Leyland³, Matthew Stammers⁴, Michael George⁴ & Michael Boniface¹

Early identification of patients who require onward referral to social care can prevent delays to discharge from hospital. We introduce an explainable machine learning (ML) model to identify potential social care needs at the first point of admission. This model was trained using routinely collected data on patient admissions, hospital spells and discharge at a large tertiary hospital in the UK between 2017 and 2023. The model performance (one-vs-rest AUROC = 0.915 [0.907 0.924] (95% confidence interval), is comparable to clinician's predictions of discharge care needs, despite working with only a subset of the information available to the clinician. We find that ML and clinicians perform better for identifying different types of care needs, highlighting the added value of a potential system supporting decision making. We also demonstrate the ability for ML to provide automated initial discharge need assessments, in the instance where initial clinical assessment is delayed and provide reasoning for the decision. Finally, we demonstrate that combining clinician and machine predictions, in a hybrid model, provides even more accurate early predictions of onward social care requirements (OVR AUROC = 0.936 [0.928 0.943]) and demonstrates the potential for human-in-the-loop decision support systems in clinical practice.

A critical task for hospital care teams is to promptly identify patients who require onward care support. For patients in need of onward care, whether at home or in care homes, any delays in assessment, choice, or access prevent discharge from hospital¹. Delays to discharge lead to poor patient outcomes, including increased risk of hospital-acquired infection, decreased mobility, physical deterioration, worsening mental health, and cognitive deconditioning^{2–5}. Operationally, medically fit patients occupying bedspace create substantial problems for patient flow; increasing pressure across ambulatory services and urgent care systems⁶. In the UK, the National Health Service (NHS) is under significant pressure, with a backlog of 7.6 million patients waiting for treatment and a median waiting time of 15 weeks⁷. Bedspace is at critical levels, with occupancy consistently exceeding 93% through last winter in NHS England⁸. As of January 2024, 14,436 patients a day (14% of bed capacity) remained in hospital while being ready to leave⁹.

Despite the majority of patients being discharged without the need for additional support (85% as of December 2022⁶), the driving reason for discharge delay is the need for onward care following an illness or injury. For patients in UK hospitals for 7 days or more, 65% of discharge delays are due to the arrangement of onward care. This includes: organisation of care at home (25% of all delays), short-term reablement (22%), and permanent care or a nursing home (18%). Figure 1 shows the four distinct 'Discharge to Assess Pathways' as defined by NHS England. The Pathways increase in severity from one to four, with NHS England targeting 95% of patients to be discharged home.

Community services face significant issues with both staffing and funding, ultimately leading to problems with capacity and response times^{1,10}. Despite this, internal processes at hospitals also cause delay, including delay in assessment or decision-making, misidentification of care requirements, transfer between clinical specialities¹¹, and lower discharge rates at weekends¹². In-line with NHS standard operating procedure, an initial discharge assessment should be made within the first 24 h of hospital admission¹³. In practice, this initial planning is provided for less than 50% of patients and a lack of planning is responsible for 20% of delays overall¹⁴.

¹IT Innovation Centre, Digital Health and Biomedical Engineering, University of Southampton, Southampton, UK.

²Southampton Business School, University of Southampton, Southampton, UK. ³University Hospital Southampton Foundation Trust, Southampton, UK. ⁴Southampton Emerging Therapies and Technologies Centre, University Hospital Southampton Foundation Trust, Southampton, UK. ✉email: C.J.Duckworth@soton.ac.uk

Pathway 0	Simple discharge home No additional support, return to usual residence	50%
Pathway 1	Supported discharge home Return to usual residence with additional care i.e., therapy, nursing, domiciliary care and new equipment	45%
Pathway 2	Short-term intensive care Transfer to a community bedded setting with dedicated health and/or social care i.e., rehabilitation, reablement and recovery	4%
Pathway 3	Care home placement or 24-hour care Transfer to a care home facility for complex and significant health and/or social needs. Follows long-term intermediate care in the community	1%

Fig. 1. Discharge pathways as defined by NHS England which broadly describe increasing severity of health and social care requirements. Percentages (grey box) represent NHS England targets for patients discharged on each pathway.

The aim of this paper is to demonstrate that machine learning can be used to support the early and accurate identification of onward care needs. Early identification creates the possibility to release a large proportion of capacity back into the system (both care hours wasted when a patient is too unwell to leave the hospital, and in social worker hours collating details regarding the patient's most current functional ability).

First, we introduce an explainable machine learning model to predict patients' potential discharge care needs at the point of admission to hospital. Learning from historical best practice, supervised machine learning has been validated in a variety of clinical settings¹⁵ including cancer diagnosis¹⁶, sepsis prediction¹⁷, and emergency admissions¹⁸. The prediction is made based on basic, digitally available patient information (e.g., age, arrival mode, admission route, chief complaint, history of prior spells, and prior discharge needs) at the first point of admission. Using machine learning explanations, we can also identify the most important factors to review when assessing a patient's onward care needs.

Second, we consider a human-machine hybrid model which updates the machine learning prediction based on the clinician's own initial assessment of discharge requirements. Machine learning algorithms naturally have capacity to capture more complex trends in data but are intrinsically limited by the inability to capture important contextual domain knowledge (e.g., visual cues of patient or conversation with patient or family). A hybrid approach combining human and machine predictions integrates both contextual domain knowledge and complex data mining to best identify potential discharge requirements to reduce likelihood of delay.

Methods

Data

We used a pseudonymised version of routinely collected data on patient admissions and hospital spells at UHS between 1st January 2017 and 1st January 2023. Cohort selection is described in Fig. 2, which we briefly summarise here. Patient history is computed from 757,421 spells (including 503,542 outside of study period). Discharge planning information is managed by the Application Express (APEX) system which was introduced into UHS working practice as of 19th August 2019 before being widely adopted at the Trust by 1st June 2020. APEX records discharge planning information such as expected discharge date, predicted discharge pathway, and additional notes about complex care needs. This accounted for 94,490 spells from which patient discharge history is computed. The main cohort used in modelling is 51,754 once short spells and those without planning information are removed.

We used the NHS England Discharge to Assess pathways as a multi-class target for our machine learning models (as shown in Fig. 1). Introduced as best practice in 2016, this framework ranges from reablement and rehabilitation to short-term care, with longer-term needs for care and support assessed at a point of optimal recovery^{19,20}. Other than in exceptional circumstances, no patient should be discharged directly into a permanent care home for the first time without first exhausting home first as a least restrictive discharge venue or discharging into a temporary placement. These pathways represent an ordinal mapping from care acuity. We took the last discharge pathway registered in APEX by the clinical team before discharge. In Fig. 3, we show the actual distribution of discharge pathways throughout the study period for training and validating our machine learning model. The overall distribution of pathways is significantly different to NHS guidelines as found in Fig. 1 (i.e., Pathway 0: 69.4%, Pathway 1: 12.8%, Pathway 2: 10.9%, Pathway 3: 7.0%). In the right panel of Fig. 2, we note a shift in the proportion of spells discharged on Pathways 1 and 3 around March 2022, likely reflective of discharge policy change.

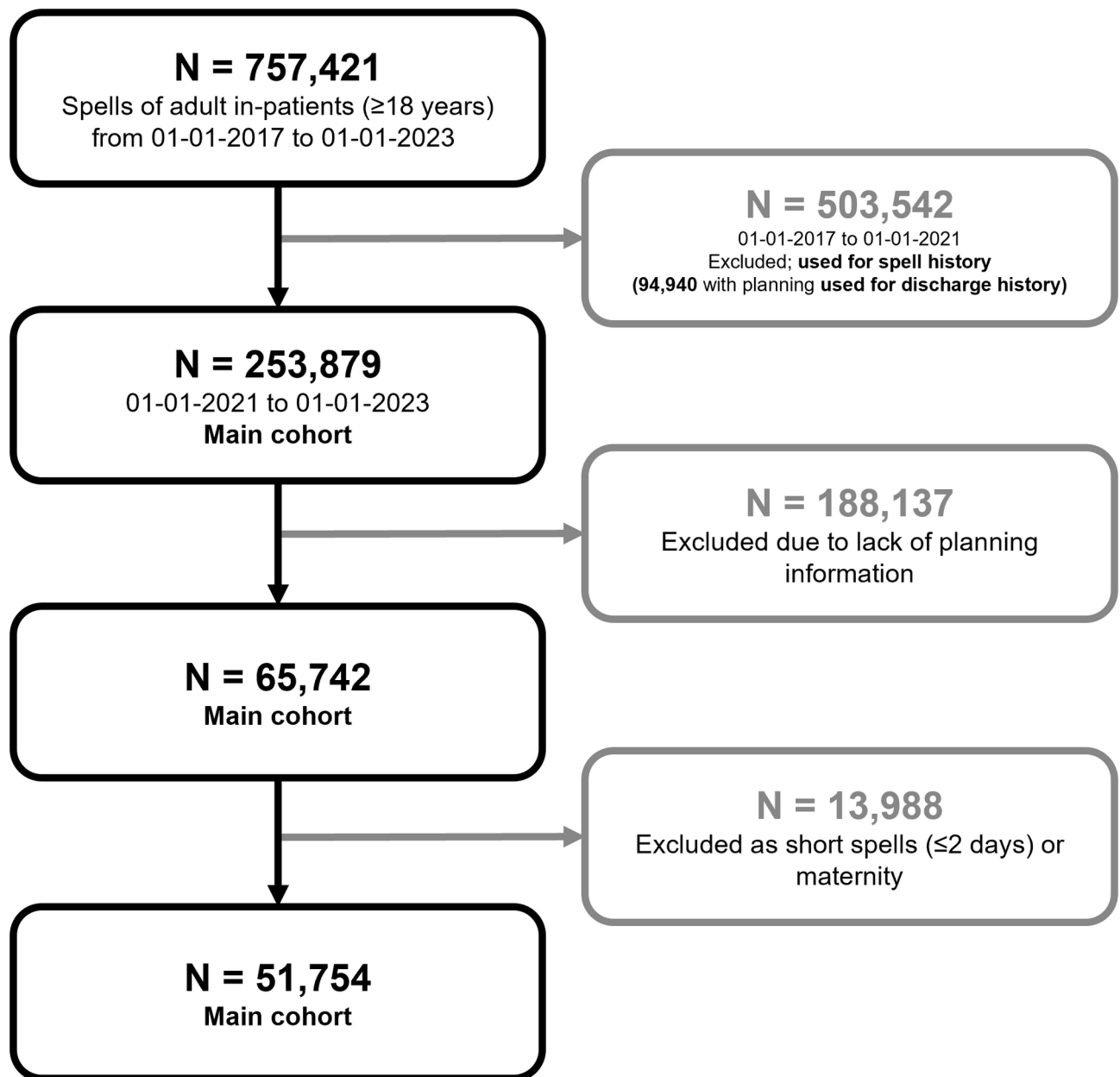


Fig. 2. Cohort selection diagram. Darker arrows and boxes correspond to selection of the core cohort, while lighter arrows and boxes (grey) correspond to excluded spells).

For our study cohort (i.e., the period in which we trained and validated our models), we selected spells that occurred between 1st January 2021 and 1st January 2023, with corresponding APEX discharge planning information. This ensures we collected patient history over a 4-year baseline and a discharge history over the prior six months. We removed short spells (i.e., ≤2 days) leaving 51,754 for modelling. Most short spells do not require discharge planning as they are generally elective procedures discharged on-the-day (84%) or held overnight for observation. These spells are typically not considered by discharge teams and hence we do not aim to make predictions. We note that including spells (≤2 days) would increase the proportion of patient spells discharged on Pathway 0, however, this does not significantly change the performance of our model. The general characteristics of our study cohort stratified by discharge pathway are given in Table 1. We describe our cohort further in Fig. 4, where the frequency of the most common occurring categories in prominent features is stratified by discharge pathway. We note that arrival by road ambulance and being previously treated by a geriatric specialty are skewed towards higher discharge needs.

Predictions of discharge pathways are made at the very first point of admission (i.e., before information relating to care speciality or planned treatment is made available), with each patient characterised by basic patient descriptors (e.g., age, ethnicity), admission method (e.g., if the patient was admitted as GP referral or through the emergency department; ED), information collected in the ED (if applicable), and histories from

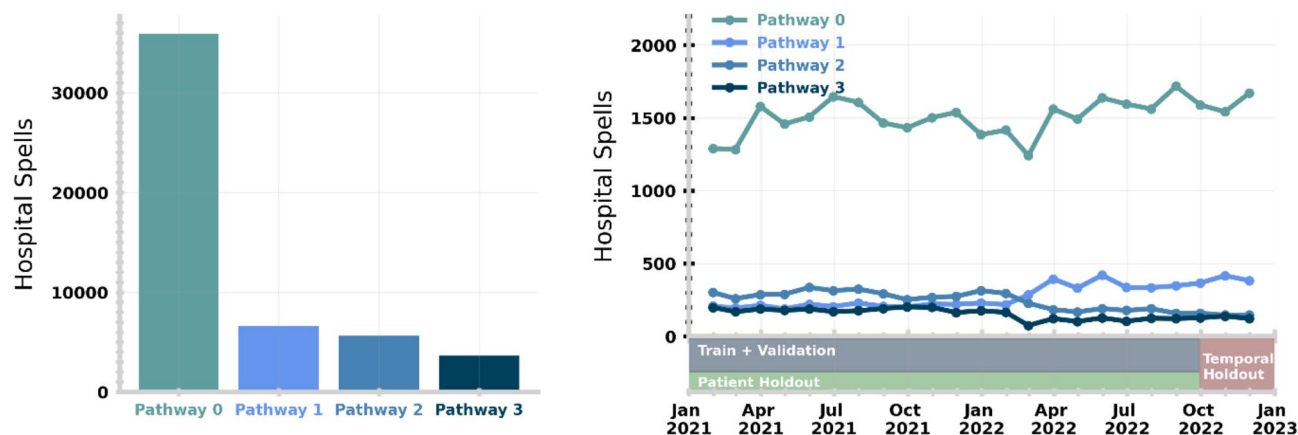


Fig. 3. (Left) Distribution of patient discharges across discharge pathways throughout the study period. (Right) Monthly average of patient discharged on each pathway. We demonstrate how we split our study data into training, patient test and temporal test sets below the monthly average.

	Pathway 0	Pathway 1	Pathway 2	Pathway 3
Age group				
18–44	6333 (17.7%)	209 (3.17%)	99 (1.76%)	181 (5.02%)
45–64	11,385 (31.9%)	929 (14.1%)	558 (9.92%)	406 (1.13%)
65–84	15,955 (44.6%)	3302 (50.0%)	3039 (54.0%)	1638 (45.4%)
85+	2062 (5.77%)	2162 (32.7%)	1928 (34.3%)	1383 (38.3%)
Gender				
Female	16,260 (45.3%)	3751 (56.8%)	3212 (57.1%)	2029 (56.1%)
Male	19,644 (54.7%)	2855 (43.2%)	2414 (42.9%)	1585 (43.8%)
Elective spell				
Yes	9059 (25.2%)	251 (3.80%)	279 (4.96%)	78 (2.16%)
No	26,848 (74.8%)	6355 (96.2%)	5347 (95.0%)	3537 (97.8%)
Admission via the emergency department				
Yes	17,745 (49.4%)	5450 (82.5%)	4493 (79.9%)	3019 (83.5%)
No	18,162 (50.6%)	1156 (17.5%)	1133 (20.1%)	596 (16.4%)
Has spell history				
Yes	23,433 (65.2%)	5861 (88.7%)	4454 (79.2%)	2951 (81.6%)
No	12,474 (34.7%)	745 (11.2%)	1172 (20.8%)	664 (18.3%)
Has discharge history				
Yes	7863 (21.9%)	3562 (53.9%)	2183 (38.8%)	1706 (47.2%)
No	28,044 (78.1%)	3044 (46.0%)	3443 (61.1%)	1909 (52.8%)

Table 1. Descriptive statistics of the study cohort spells stratified by final discharge pathway. Values are spell count (percentage). Percentage is normalised by column-wise (i.e., by discharge pathway).

prior spells and discharge plans. Prediction at this stage helps early identification of elevated onward care needs, ensuring transfer of care teams can triage earlier and have additional time to organise social care, which has significant lead times.

All numerical variables were used directly as input to the model, whereas categorical variables are target-encoded^{21,22}. We provide a more detailed description of our encoding in Supplementary Methods Sect. 1.1 and the full list of features in Supplementary Table 1. In Fig. 3, we show the distribution of the most prevalent categories for a selection of features used in modelling, stratified by the final discharge pathway for the spell. The electronic health records used by our models are available to review by clinicians and are used in regular practice. Our model did not have access to any free text fields in the electronic health records. Analysis was performed in Python 3.9.5 using standard packages including numpy²³, matplotlib²⁴, and pandas²⁵.

Ethics and data governance

This study was approved by the University of Southampton's Ethics and Research Governance Committee (ERGO 78496) with project title "PROactive, Collaborative, and Efficient complex Discharge - UHS". Approval was obtained from the Health Research Authority. All methods were carried out following relevant guidelines

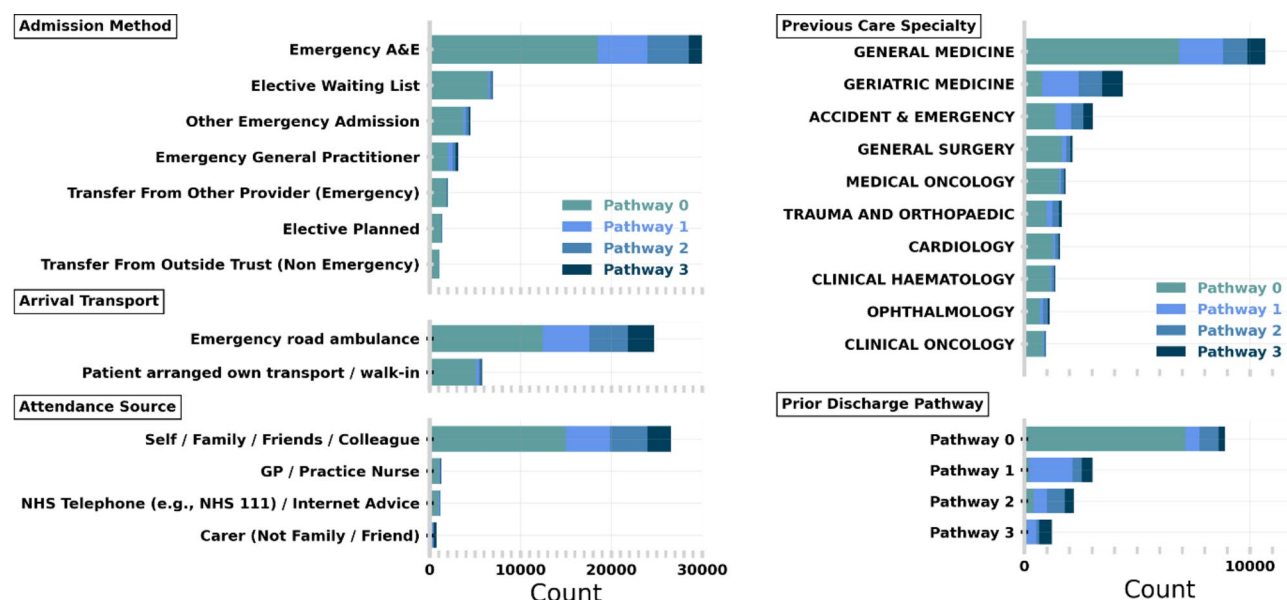


Fig. 4. Each panel shows the frequency of most common categories for a given feature. Admission Method (top left), Arrival Transport (middle left), Attendance Source (bottom left), Previous Care Speciality – i.e., from prior spell (top right) and Prior Discharge Pathway (bottom right). The stacked bar shows the total frequency of the category with each bar stratified by final pathway.

and regulations associated with the Health Research Authority, and NHS Digital. The research was limited to the use of previously collected, non-identifiable information. Informed consent was waived by the University of Southampton's Ethics and Research governance committee, University of Southampton, University Road, Southampton, SO17 1BJ, United Kingdom and the Health Research Authority, 2 Redman Place, Stratford, London, E20 1JQ, UK. Data was pseudonymised (and, where appropriate, linked) before being passed to the research team. The research team did not have access to the pseudonymisation key.

Modelling

We trained our classifier based on 29,084 unique spells (56.2% of total) and defined a validation set (7272; 14.0%) grouped at the patient level. To quantify the performance of our classifier, we defined two holdout test sets to consider how the model would operate in different circumstances. First, we defined a patient holdout test set (i.e., multiple spells relating to a patient cannot be split between holdout and training). This contains 9177 spells (17.7%) from 1st January 2021 to 30th September 2022. Second, we defined a 3-month temporal holdout (6221; 12.0%) which contains all spells from 1st October 2022 to 1st January 2023. This is designed to test the real-world performance of the model (i.e., where it is trained on historical hospital data and deployed in the same environment) and further validate how data drift may impact performance. While the temporal holdout best emulates how a model may be deployed, there is a risk of data leakage (i.e., if multiple spells of a given patient are correlated). A patient holdout tests the performance for patients whose prior spells were not included in the original training. For readability, we follow the results for our patient holdout test set in the main text, while also presenting results of the temporal holdout test set in the [Supplementary Materials](#).

For our machine learning model, we used XGBoost with a 'multi-softmax' objective^{26,27}. We choose XGBoost since gradient-boosted trees are a popular choice for large tabular data, generally out-performing other tree-based models such as random forests²⁸. XGBoost can also naturally deal with missing data consistently (as opposed to neural networks), which is important for routinely collected clinical data, as here. For hyperparameter tuning, we used five-fold cross-validation and on training defined 50 early stopping rounds (i.e., when the train set performance metrics become better than the validation set) to prevent overfitting, which gradient-boosted trees are prone to. We provide more detailed information on our choice of XGBoost and hyperparameter tuning²⁹ in [Supplementary Materials Sect. 1.2](#), with [Supplementary Table 2](#) giving the final hyperparameters.

Model performance was evaluated independently for both temporal and patient holdout test sets using Micro F1 Score, One-Vs-Rest (OVR) Area Under the Receiver Operating Curve (AUROC), and Average Precision (AP) for each target class.

Model explanations

An important part of machine learning in a clinical setting is the ability to attribute why a given recommendation has been made to the user. Explanations enable users to understand the machine learning decision making process and helps identify data which should be reviewed before making the final decision. While machine learning models are often described as 'black boxes', there are various methods to interpret and explain the predictive process.

We attributed the fractional contribution of each input feature to individual pathway predictions using Shapley values (i.e., Shapley Additive exPlanations; SHAP^{30,31}). One strength of SHAP is that explanations are: (1) locally accurate (i.e., each pathway prediction can be explained), and (2) globally consistent (i.e., explanations can be consistently aggregated across multiple predictions). This flexibility is critical in a medical environment where explanations are important on a local (i.e., for each patient) and global (i.e., to understand the operating circumstances of the model) level. We used the TreeExplainer^{32,33} implementation of SHAP designed for tree-based algorithms.

Clinical pathway updates

As part of the standard operating procedure at UHS, an initial discharge assessment should be made within the first 24 h of hospital admission, usually as part of the first consultant review¹³. In order to compare how the proposed machine learning model performs in relation to current practice, we considered all pathway assessments made within the first 24 h post-admission by clinical care teams. In practice, this corresponds to all pathway updates registered on the APEX system. Multiple pathway updates within 30 min of each other are combined (retaining the most recent) as they are more likely to reflect administrative updates (e.g., opening the patient record with incomplete information) rather than distinct assessments. For the patient holdout test set, this corresponds to 4283 updates across 4076 spells (46.67%). For the temporal holdout test set (Results in [Supplementary Materials](#)), this corresponds to 2845 total updates for 2757 spells (45.73% of test set).

Hybrid model

Machine learning is limited relative to humans by not being able to capture important contextual domain knowledge that is not easily captured in data (e.g., visual cues of a patient's current state, conversations with family or carers). Despite this, machine learning models capture more complex trends in data than humans can retain, offering consistency and nuance in predictions. A hybrid approach, combining human and machine predictions integrates domain knowledge, and complex data mining to deliver a more powerful recommendation.

For spells with an initial pathway prediction, a 'hybrid' recommendation based on this input and the model probabilities can be made. For this model, we made use of the Scikit-Learn implementation of a Logistic Regression, adapted to make multi-class predictions²². Logistic regression is a linear model which aims to classify outcomes through probabilities estimated by the logistic function. We implement optimised for ≤ 100 iterations using an L2 penalty with class weights inversely proportional to class frequencies. For both the XGBoost and 'hybrid' logistic regression models, we defined class weights inversely proportional to the number of the samples in a target class. This ensures our classifiers are tuned to be robust against class imbalance and are calibrated to better identify higher acuity pathways.

Results

Machine learning model performance

In Fig. 5, we consider the performance of our XGBoost model for each target class. Performance is evaluated using the Receiver Operating Curve (ROC; left) and the Precision-Recall (PR; right) Curve. Performance is given using a One-Vs-Rest strategy (i.e., each class is considered as binary with the rest of the predictions of the negative outcome). The overall metric score for each pathway is given in the bottom right of the panel with a 95% confidence interval. Each value quantifies the performance of the model correctly identifying the given pathway in isolation (0–1 where 1 is perfect performance). AUROC values show how well the model can distinguish between positive and negative values, whereas Average Precision is more sensitive to imbalanced class sizes. For

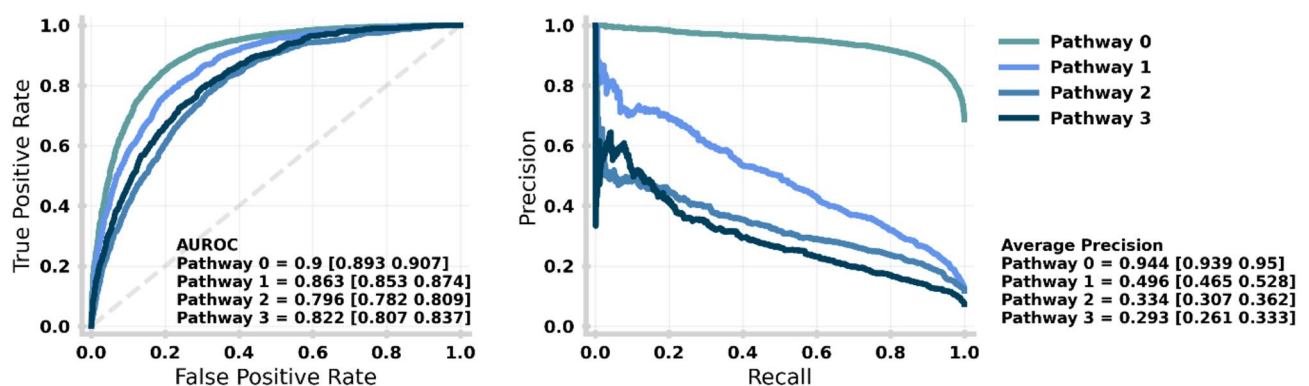


Fig. 5. Performance of the machine learning model for each discharge pathway. The performance is given by the receiving operator curve (left) and precision-recall curve (right) with the Area Under the ROC and Average Precision (over all recall thresholds) annotated respectively with 95% confidence intervals given in square brackets. Each value quantifies the performance of the model correctly identifying the given pathway in isolation (0–1 where 1 is perfect performance). AUROC values show how well the model can distinguish between positive and negative values, whereas Average Precision is more sensitive to imbalanced class sizes. An equivalent plot for the temporal holdout test set is given in [Supplementary Fig. 2](#).

both patient (top) and temporal (bottom) holdout test sets, performance is generally better for lower acuity care pathways (i.e., no care needs easier to identify than elevated care needs).

Figure 6 shows the confusion matrix for the patient holdout test set, providing more information on performance stratified by pathway. The matrix shows the distribution of model predictions (columns) against the actual resulting pathway (rows). In each square, the total counts and row normalised (i.e., by actual pathway) proportions are shown. An accurate model will show much higher counts along the diagonal. Despite implementing sample weights (to balance the importance of predictions across pathways), the XGBoost model is best at identifying patients to be discharged on Pathway 0. Regardless, the model performs well at identifying patients in need of needing any elevated care needs (i.e., predicting pathways 1–3) as denoted by the PR curve of Pathway 0 in Fig. 5.

Figure 6 demonstrates our model to be well tuned to identify higher acuity pathways. Despite this, we note distinct differences in class mixing between the patient holdout and the temporal holdout (Supplementary Fig. 2). We find the temporal holdout has a slight bias towards Pathway 1, with potential data drift towards the end of the study period (containing the temporal holdout). This is further justified in Fig. 2 (right), where a shift around March 2022 leads to a higher proportion of patients discharged on Pathway 1 and a lower proportion of those discharged on Pathway 3. This highlights the need for data drift detection and frequent re-evaluation of the machine learning model if deployed in practice.

Machine learning explanations

Explainable machine learning enables the attribution of the most important features for prediction. Since SHAP is globally consistent, the relative contribution to prediction can be stratified by target class (by grouping all predictions of a given discharge pathway). In Fig. 7, we show the relative importance of admission characteristics (i.e., features) in predicting discharge needs. Each bar is stacked so the importance for predicting individual discharge pathway can be identified (e.g., Arrival Transport is far more important for predicting pathway 0 than 1–3). Features are grouped by types of information (e.g., Demographics, Patient History) and combined when encoding similar raw data (e.g., Previous Specialties is the sum of individual features ‘Last Spell Specialty’ and ‘Dominant Specialty’). Supplementary Table 1 lists all individual features, and Supplementary Fig. 1 shows the importance of individual features without combination.

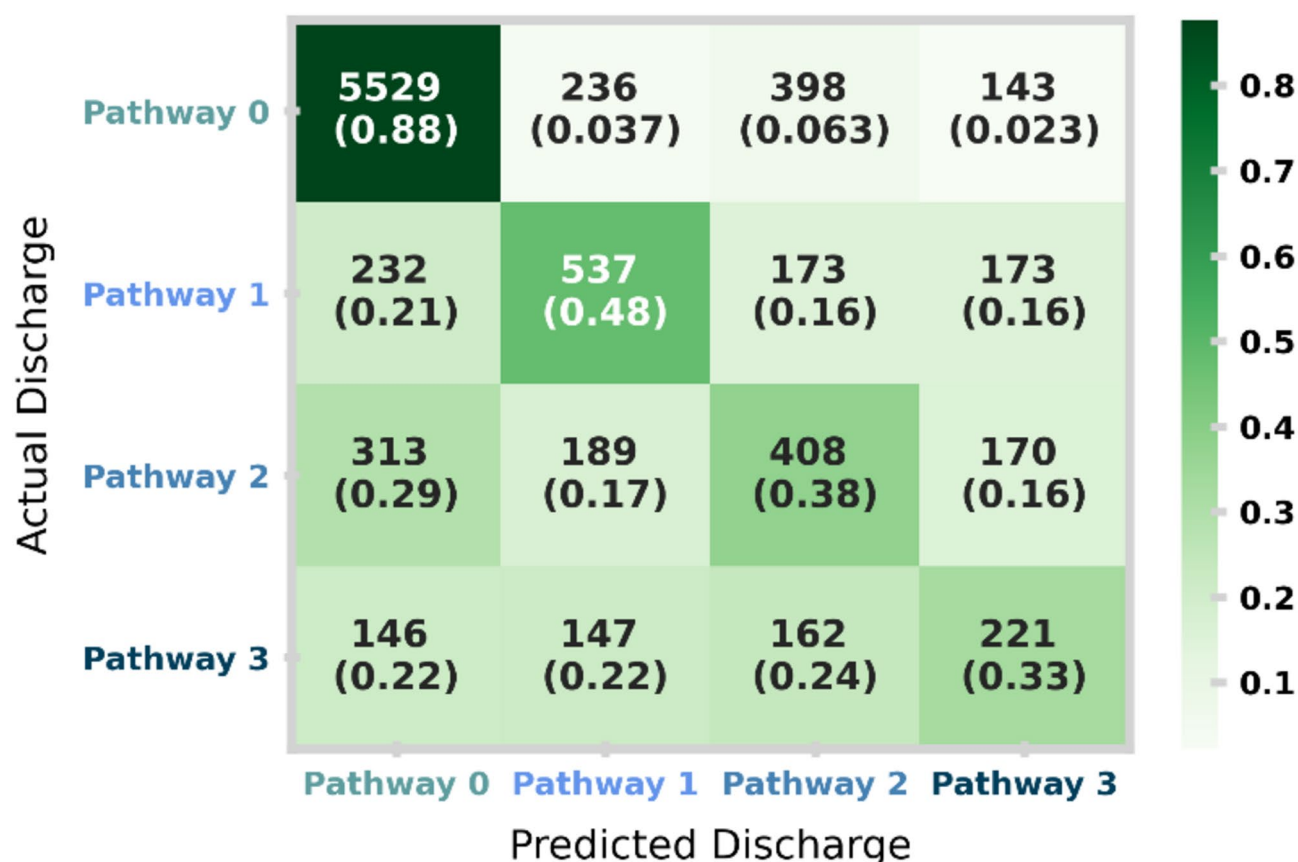


Fig. 6. Confusion matrix for the patient holdout test set. The matrix shows the distribution of model predictions (columns) against the actual resulting pathway (rows). In each square, the total counts and row normalised (i.e., by actual pathway) proportions are shown. An accurate model will show much higher counts along the diagonal. An equivalent plot for the temporal holdout test set is given in Supplementary Fig. 3.

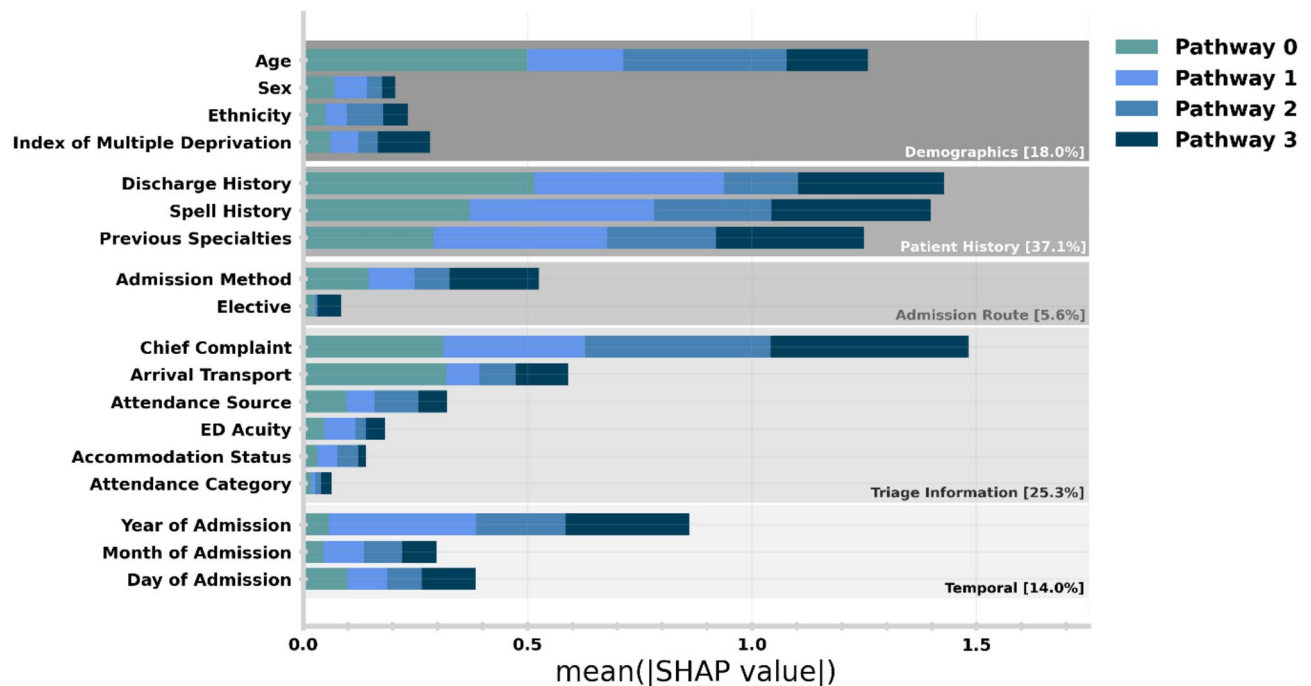


Fig. 7. Relative importance of admission characteristics for predicting discharge pathway in the patient holdout test set. Importance is computed by normalised absolute SHAP values (i.e., larger values mean higher relative importance). The size of each complete bar shows how relatively important this information was for making pathway predictions. Each bar is stacked so the importance for predicting individual discharge pathway can be identified. An equivalent plot for the temporal holdout test set is given in Supplementary Fig. 3.

We found that having patient history (both in terms of spells and discharges), is the most significant factor in prediction, followed by the information collected at Triage in the ED. The importance of the Year of Admission reflects the importance of operational circumstances and policy at the given timeframe. Given that our study cohort contains periods still operationally disrupted by the COVID-19 pandemic, this is likely to contribute to this dependence.

Practically, feature importance can be used as guidance for which information should be reviewed when a pathway prediction does not align with expectations. For example, an explanation of an unexpected prediction of Pathway 3 may identify discharge history being a particularly important feature. The discharge history may then identify the patient has previously been discharged on Pathway 1/2 and now needs escalation in care.

Having identified the most important features for prediction, it is important to ensure our model is well generalised in the case that any of this information is not available for a particular prediction. Performing a subgroup analysis, we found our model is robust for patients without discharge history (OVR AUROC: 0.834 [0.824 0.843]), spell history (0.843 [0.827 0.859]), and for those which are not admitted via ED (0.849 [0.833 0.867]) compared with the overall model performance of 0.845 [0.838 0.852].

Comparison to clinical predictions and hybrid model

We now compare our model to pathway predictions made by clinicians in-line with UHS standard operating procedure¹³. This corresponds to all pathway updates registered in APEX by clinicians within 24 h of admission, usually at the first consultant review. It is important to note that only 46.7% of spells (relative to Fig. 6) have clinical assessments within 24 h of admission.

In Fig. 8, we compared the performance between the machine learning model (solid), clinical updates within 24 h of admission (dot-dashed), and the hybrid model (dashed) as described by the ROC (left) and PR (right) curves. These show the discriminative power for Pathway 0 (i.e., the ability to recognise onward discharge needs or not) through an OVR strategy (see “Methods” for explanation). The hybrid model aims to combine the machine learning model class probabilities with the clinical prediction to derive a more powerful recommendation. We note that the AUROC and Average Precision are computed using probabilities for machine learning and hybrid models, whereas clinical pathway updates are absolute (i.e., only probability 1 for a single class), which likely understates their skill.

Figure 9 shows the confusion matrices for each of the machine learning models (left), clinical predictions made within the first 24 h of the spell (middle), and the hybrid model (right). Clinical predictions have significantly higher recall for Pathway 0, meaning clinicians are far more effective (correct 98% of the time) at identifying no onward care in comparison to the ML (86%) and hybrid (88%) models. This is a natural reflection of the tuning of the machine learning and hybrid models being weighted to improve the recall of higher acuity onward care needs coming at the cost of a higher false alarm rate. As a result, our hybrid (and machine learning

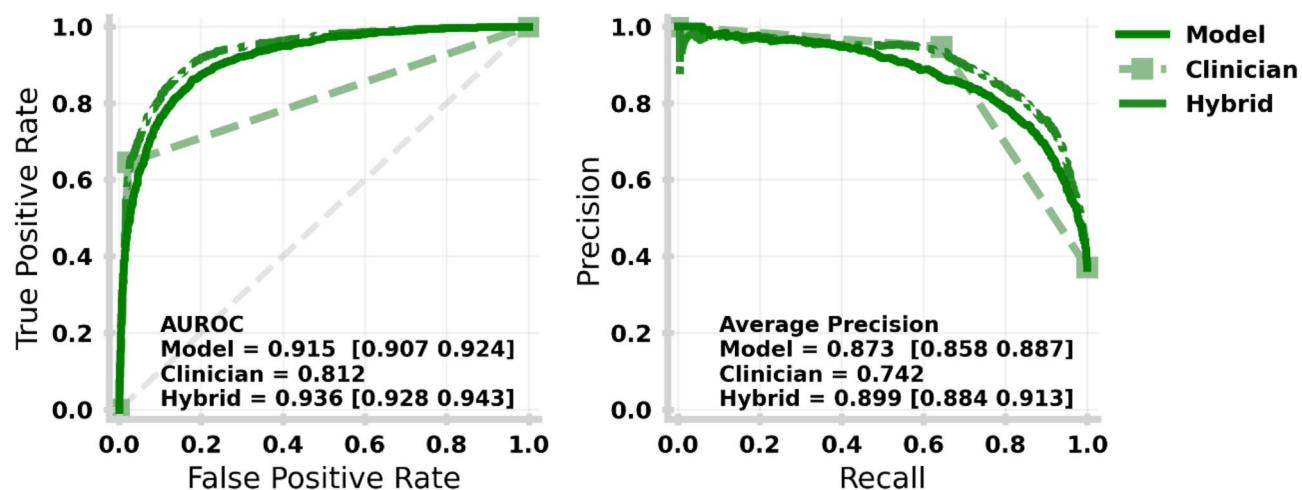


Fig. 8. Receiving Operating Characteristic (left) and Precision-Recall (right) curves for each of the machine learning model (solid dark green), clinician pathway updates (dashed light green) and hybrid model (dot-dashed mid green) for the patient holdout. AUROC values show how well the model can distinguish between positive and negative values, whereas Average Precision is more sensitive to imbalanced class sizes. An equivalent plot for the temporal holdout test set is given in Supplementary Fig. 6.

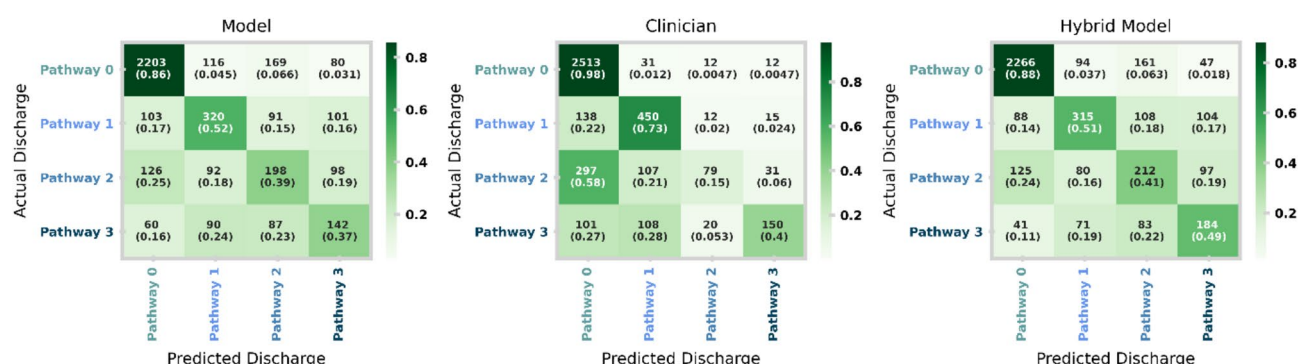


Fig. 9. Confusion matrices for the machine learning model (left), clinical pathway updates made within first 24 h of spell (middle) and the hybrid model (right) for the patient holdout test set. An equivalent plot for the temporal holdout test set is given in Supplementary Fig. 7.

only) model offers a marked improvement on identifying Pathway 2 (41% vs. 15%) and Pathway 3 (49% vs. 40%). Independent of tuning, we also note intrinsic differences in skill with clinicians being more effective at identifying Pathway 1, while generally struggling with Pathway 2.

We note that inaccurate initial assessments are most impactful (i.e., largest discharge delays) for patients eventually discharged on Pathway 2 and 3¹¹. In particular, for patients eventually discharged on Pathway 3, an inaccurate initial assessment increases the odds of discharge delay almost five-fold¹¹. A hybrid model which markedly improves the accuracy of initial assessments for patients discharged on Pathways 2 and 3 will directly reduce the likelihood of delay.

Discussion

A significant bottleneck for patient flow through the NHS is the transfer of patients from hospital beds to social care services. Delays in discharge for medically fit patients are driven by the mismatch between bedspace pressures in hospitals, and the ability of understaffed social care services to continue to accept further patients. In this work, we have demonstrated the potential for explainable machine learning to support clinical care teams with reliable information on the initial assessment of onward care needs (AUROC 0.915 [0.907–0.924], AP 0.873 [0.858–0.887] for identifying onward needs in general).

In-line with University Hospitals Southampton (UHS) NHS standard operating procedure, an initial discharge assessment should be made within the first 24 h of hospital admission¹³. In practice, this initial planning is provided for less than 50% of patients. Awaiting completion of assessment was listed as the primary reason for discharge delay for 20% of NHS patients¹⁴. We find that model performance is comparable to an initial clinical assessment and can provide initial planning suggestions for spells (>50%), when clinical assessment

does not occur within 24 h of admission. Capacity pressures in hospitals compound delays by reducing the time and bandwidth of clinical teams to review and assess potential onward care requirements promptly. These predictions and subsequent model explanations could help clinical teams ensure that elevated onward care needs are identified promptly, especially as time and resource pressures prevent review.

In addition to delays due to pressures on clinical teams, pressures on social care create significant lead times in finding space for a new care patient. As a result, it is critical that onward care needs can be identified as early as possible post admission to the hospital, with the highest acuity (i.e., Pathways 2 and 3: short- and long-term intensive care respectively) most important to identify early. We identify that leveraging the clinician's initial assessment along with the ML prediction (i.e., hybrid model) offers a significant improvement (AUROC 0.936 [0.928–0.943], AP 0.899 [0.884–0.913]) on the ability to recognise acute care needs (i.e., Pathway 2 and 3) from clinicians or ML alone (Fig. 8 bottom panel).

To adapt machine learning predictions and explanations into clinical practice, the implementation of a clinical decision support (CDS) system must be considered. Clinical decision support (CDS) systems are digital tools which can support clinicians to make more equitable, evidence-based decisions^{34–36}. CDS systems have been demonstrated to improve the quality of decision-making, in turn reducing errors in addition to supporting shared decision-making practice³⁵. They have been applied across a variety of care and decision settings using a combination of clinical tools, technologies, information resources, and guidance^{34,36}. Specifically, the implementation of knowledge-based systems has been further expanded through the use of data-driven machine learning-based CDS systems^{37,38} and medical devices^{39,40}. In this context, it is important to note the intrinsic differences in skill between our proposed machine learning and hybrid models in comparison to the clinical assessments. While our models are tailored to better identify high acuity Pathways (2 and 3), they have a higher error rate for Pathways 0 and 1 (generally predicting higher acuity than required). A higher false-alarm rate would 'over-prescribe' which in turn could exacerbate the arrangement of care placement spaces and delay arrangement of care for others.

A higher 'false-alarm' rate (i.e., over predicting Pathways 2 and 3) is also one example of potential barriers to clinical integration. This could degrade trust in the system and probability thresholds will have to be tuned to balance the trade-off between precision and a higher false-alarm rate. Other potential barriers include how and when predictions (and explanations) should be presented through the CDS. It is critical that any CDS provides information (or predictions) that is interpretable and does not disrupt clinical workflows. An open question is the operational circumstances on when model predictions may be used. While predictions could be integrated into a CDS used by clinicians to improve the accuracy of initial assessments, we also highlight that the model can make a prediction in lieu of an initial assessment not being completed within 24 h. These are distinct workflow integrations (i.e., clinical workflow and discharge team workflow). In future work, we look to integrate our explainable machine learning model into a CDS system designed to assist complex discharge throughout a patient spell. This will involve adapting our existing supervised ML model to accept additional information as it becomes available throughout treatment (e.g., blood test results and procedures), which in turn will improve predictive performance as the spell progresses, reducing the false-alarm rate for those who only need lower acuity care. These predictions and explanations will be integrated into a clinical interface through co-design with clinical care teams and the acute discharge bureau at UHS.

Another barrier to real-world implementation is the notion of data drift (i.e., a systematic shift in the underlying distribution of input features or target)^{41,42} or concept drift (i.e., the fundamental connection between features and target)⁴³. For example, this could be due to changes in patient populations over time or operational policy on discharge¹⁸. We validate the performance of our model over time with the temporal holdout test set (Supplementary Results S2), finding changes in the distribution of Pathway predictions (Supplementary Fig. 3) and actual discharges (Main text Fig. 3). This highlights that the need for monitoring of ML models and periodic retraining. This also has implications for the deployment of this model at other hospital sites. In future work, we will validate our model on another hospital site in the UK to understand its generalisation and robustness to data drift.

Strengths and limitations

Strengths of this research include the large and robust data collection for discharge planning, which itself was used by complex discharge teams at UHSFT; the large hospital (in terms of bed number) is NHS England. The routinely collected and digitised data is used directly for national reporting as part of the statistical snapshot for acute discharge delays aggregated by NHS England. As a result, the discharge pathway updates are accurately recorded for the cohort spells. This study also benefits from direct collaboration between hospital care teams, discharge teams, and researchers to provide different perspectives.

Limitations of this research include that this study is single site (i.e., no other site validation) and performed retrospectively. There is also an absence of planning information for most hospital spells at UHSFT which are not considered in this study. Discharge planning is only created for more complex cases, and, as a result, our patient cohort is more complex than hospital averages. However, most hospital spells which were removed were short spells with 84% having a length of stay of one day or less. Consequently, these spells are unlikely to contribute to significant delay or require onward planning.

Data availability

'The data that support the findings of this study are available from the University Hospital Southampton NHS Trust, but restrictions apply to the availability of these data as they relate to patient records and personally identifiable information. Data were used under strict governance for the current study and presented only through aggregation and so are not publicly available. Data are however available from the authors upon reasonable re-

quest with appropriate data protection, governance, and ethical review in place. Please contact C.J.Duckworth@soton.ac.uk for enquiries.

Received: 7 October 2024; Accepted: 30 April 2025

Published online: 08 May 2025

References

1. Oliver, D. *David Oliver: Delayed Discharges Harm Patients, Staff, and Health Systems Alike* (British Medical Journal Publishing Group, 2023).
2. Friebe, R. et al. The implications of high bed occupancy rates on readmission rates in England: a longitudinal study. *Health Policy*. **123** (8), 765–772 (2019).
3. Kulski, K., Im, J. & McGeown, M. It's a waiting game: a qualitative study of the experience of carers of patients who require an alternate level of care. *BMC Health Serv. Res.* **17**, 1–10 (2017).
4. Limb, M. *Delayed Discharge: How Are Services and Patients Being Affected??* (British Medical Journal Publishing Group, 2022).
5. Rojas-García, A. et al. Impact and experiences of delayed discharge: a mixed-studies systematic review. *Health Expect.* **21** (1), 41–56 (2018).
6. Schlepper, L., Dodsworth, E. & Scobie, S. *Understanding delays in hospital discharge*. <https://www.nuffieldtrust.org.uk/news-item/understanding-delays-in-hospital-discharge> (2023).
7. NHS-England. *Consultant-led Referral to Treatment Waiting Times Data 2023–24*. <https://www.england.nhs.uk/statistics/statistical-work-areas/rtt-waiting-times/rtt-data-2023-24/#Jan24> (2023).
8. NHS-England. *NHS weekly winter operational update*. <https://www.england.nhs.uk/2023/02/nhs-weekly-winter-operational-update-for-the-week-ending-29-january-2023/> (2023).
9. NHS-England. *NHS Delivers on Ambition of 5000 More Permanent Beds*. <https://www.england.nhs.uk/2024/01/nhs-delivers-on-a-ambition-of-5000-more-permanent-beds/#:~:text=The%205%2C000%20extra%20beds%20ambition,the%20same%20week%20last%20year> (2023).
10. Discombe, M. *Delayed discharges rise in 17 ICSs despite Barclay's £250m fund*. <https://www.hsj.co.uk/quality-and-performance/delayed-discharges-rise-in-17-icss-despite-barclays-250m-fund/7034213.article> (2023).
11. Burns, D. K. et al. Impact of accurate initial discharge planning and in-patient transfers of care on discharge delays: a retrospective cohort study *medRxiv preprint* (2024).
12. Micallef, A., Buttigieg, S. C., Tomaselli, G. & Garg, L. Defining delayed discharges of inpatients and their impact in acute hospital care: a scoping review. *Int. J. Health Policy Manage.* **11** (2), 103 (2022).
13. NHS UHSFT. *Expected Date of Discharge (EDD) & Clinically Optimised for Discharge (COFD) Standard Operating Procedure* (2021).
14. Fund, K. *Delayed transfers of care: a quick guide*. <https://www.kingsfund.org.uk/insight-and-analysis/articles/delayed-transfers-care-quick-guide> (2023).
15. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25** (1), 44–56 (2019).
16. Dong, J. et al. Clinical trials for artificial intelligence in cancer diagnosis: a cross-sectional study of registered trials in clinicaltrials.gov. *Front. Oncol.* **10**, 1629 (2020).
17. Rahmani, K. et al. Assessing the effects of data drift on the performance of machine learning models used in clinical sepsis prediction. *Int. J. Med. Informatics*. **173**, 104930 (2023).
18. Duckworth, C. et al. Using explainable machine learning to characterise data drift and detect emergent health risks for emergency department admissions during COVID-19. *Sci. Rep.* **11** (1), 23017 (2021).
19. Department of Health & Social Care N-E. *Hospital discharge and community support guidance*. <https://www.gov.uk/government/publications/hospital-discharge-and-community-support-guidance/hospital-discharge-and-community-support-guidance> (2023).
20. Vernon, M. J. *When it comes to discharge, timing is everything*. Weblog. <https://www.england.nhs.uk/blog/martin-vernon-3/> (2023).
21. Micci-Barreca, D. A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems. *ACM SIGKDD Explorations Newsl.* **3** (1), 27–32 (2001).
22. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
23. Harris, C. R. et al. Array programming with numpy. *Nature* **585** (7825), 357–362 (2020).
24. Hunter, J. D. & Matplotlib, A. 2D graphics environment. *Comput. Sci. Eng.* **9** (03), 90–95 (2007).
25. Reback, J. et al. pandas-dev/pandas: Pandas 1.0. 5. *Zenodo* (2020).
26. Chen, T., Guestrin, C. & Xgboost A scalable tree boosting system. In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining* 785–794 (2016).
27. Breiman, L. *Bias, Variance, and Arcing Classifiers* (1996).
28. In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining* (2023).
29. Akiba, T. et al. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 2623–31 (2019).
30. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural. Inf. Process. Syst.* **30**, 89 (2017).
31. Lundberg, S. M. et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomedical Eng.* **2** (10), 749–760 (2018).
32. Lundberg, S. M. et al. From local explanations to global Understanding with explainable AI for trees. *Nat. Mach. Intell.* **2** (1), 56–67 (2020).
33. Lundberg, S. M., Erion, G. G. & Lee, S.-I. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888* (2018).
34. NHS-England. *Supporting clinical decisions with health information technology*. <https://www.england.nhs.uk/long-read/supporting-clinical-decisions-with-health-information-technology/> (2023).
35. Kwan, J. L. et al. Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. *Bmj* **370**, 56 (2020).
36. Greenes, R. *Clinical Decision Support: The Road ahead* (Elsevier, 2011).
37. Moja, L. et al. Effectiveness of a hospital-based computerized decision support system on clinician recommendations and patient outcomes: a randomized clinical trial. *JAMA Netw. Open.* **2** (12), e1917094–e94 (2019).
38. Sutton, R. T. et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit. Med.* **3** (1), 17 (2020).
39. Benjamens, S., Dhunoo, P. & Meskó, B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit. Med.* **3** (1), 118 (2020).
40. Muehlmatter, U. J., Daniore, P. & Vokinger, K. N. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digit. Health* **3** (3), e195–e203 (2021).
41. Quiñero-Candela, J., Sugiyama, M., Schwaighofer, A. & Lawrence, N. D. *Dataset Shift in Machine Learning* (MIT Press, 2022).
42. Moreno-Torres, J. G. et al. A unifying view on dataset shift in classification. *Pattern Recogn.* **45** (1), 521–530 (2012).

43. Widmer, G. & Kubat, M. Learning in the presence of concept drift and hidden contexts. *Mach. Learn.* **23**, 69–101 (1996).

Acknowledgements

This report includes independent research funded by the National Institute for Health Research Applied Research Collaboration Wessex. The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care. We thank the Southampton Emerging Therapies and Technologies (SETT) Centre at UHS, in particular, Cai Davis for support with data access and insight.

Author contributions

CD and DB performed the data analysis and modelling. CF, MS, MG and MB discussed technical analysis with CD and DB. CD, DB, CF and MB designed the study with MW and RL. MW and RL provided clinical guidance and insight. CF and MB obtained governance and ethical approval and managed the study at UoS. MS and MG managed data curation and access throughout the study. CD wrote the first draft of the manuscript with all authors contributing to this and future iterations.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-00825-6>.

Correspondence and requests for materials should be addressed to C.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025