# AI BASED PREDICTION AND DATABASE


## Abstract

### Objective

Gastrointestinal (GI) bleeding commonly requires intensive care unit (ICU) in cases of potentialhaemodynamiccompromise or likely urgent intervention. However, manypatientsadmitted to the ICU stop bleeding and do not require further intervention, including blood transfusion. The present work proposes an artificial intelligence (AI) solution for the prediction of rebleeding in patients with GI bleeding admitted to ICU.

### Methods

A machine learning algorithm was trained and tested using two publicly available ICU databases, the Medical Information Mart for Intensive Care V.1.4 database and eICU Collaborative Research Database using freedom from transfusion as a proxy for patients who potentially did not require ICU-level care. Multiple initial observation time frames were explored using readily available data including labs, demographics and clinical parameters for a total of 20 covariates.

### Results

The optimal model used a 5-hour observation period to achieve an area under the curve of the receiving operating curve (ROC-AUC) of greater than 0.80. The model was robust when tested against both ICU databases with a similar ROC-AUC for all.

### Conclusions

The potential disruptive impact of AI in healthcare innovation is acknowledge, but awareness of AI-related risk on healthcare applications and current limitations should be considered before implementation and deployment. The proposed algorithm is not

meant to replace but to inform clinical decision making. Prospective clinical trial validation as a triage tool is warranted.

**Keywords:** computer methodologies, BMJ health informatics

## Summary

### What is already known?

- Gastrointestinal bleeding is a severe event that requires admission to the ICU.
- Many patients in the ICU for gastrointestinal bleeding undergo only increased monitoring without intervention.
- ICU stay is associated with increased cost and morbidity.

### What does this paper add?

- An algorithmic approach using artificial intelligence on readily available electronic data can accurately predict ICU transfusion need.
- Using this approach to identify patients at low risk for ongoing bleeding and transfusion could be validated prospectively to identify patients who may not require ICU-level care.

Go to:

## Introduction

Gastrointestinal (GI) haemorrhage is a common condition that frequently requires hospitalisation, often in the intensive care unit (ICU)[1] with considerable associated morbidity. In particular, ICU admission is associated with increased costs and a greater rate of complications and poor outcomes compared with ward admission.[2–4] Some patients are initially admitted to the ICU for haemodynamic instability but stabilise without further intervention and are discharged to the ward the following day.

Previous instruments, such as the Rockall or the Blatchford score[5] have been applied to triage patients based on the likelihood of mortality, recurrent/ongoing bleeding, need for hospitalisation and requirement for endoscopic intervention. However, these models are validated only for upper GI bleeding with a focus on endoscopic

intervention and mortality and do not assist in informing level of monitoring for hospitalised patients. Currently, there is no model to assist in triaging patients with GI bleeding including those with an undifferentiated source to an appropriate acuity of care.

We identified the need for blood transfusion as a surrogate for persistent bleeding. Previous prospective studies have shown that up to half of patients with GI bleeding may not require transfusion.[6] We used an ICU database to train a prediction model but focused on the first few hours on arrival as a proxy of the patient's state in the emergency department.

The use of artificial intelligence (AI) represents an opportunity for more effective and efficient care delivery by predicting disease trajectory and complications.[7–12] Previous work in GI bleeding has used methods such as artificial neural networks,[13 14] support vector machines[13] to predict the need for intervention; and fuzzy models[15] to identify which lab test is likely to contribute information gain and influence clinical management of patients with GI bleeding in the ICU. This study focused on using machine learning to predict transfusion to better identify those patients who continue to bleed.

Go to:

## Methods

This study is reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology statement.[16]

### Database description

Data were collected from the Medical Information Mart for Intensive Care-III (MIMIC-III) V.1.4[17] and in the eICU Collaborative Research Database V.2.0 (eICU-CRD).[18] Both databases contain information from patients admitted to the ICU. The MIMIC-III database collects detailed haemodynamic and clinical parameters from all ICU patients admitted to a single major academic medical centre between 2008 and 2014, whereas the eICU-CRD is a multicentre database with high granularity

data for over 200 000 admissions to ICUs monitored by an eICU[19] across the USA.

## Ethical approval

Both databases are previously de-identified and have been reviewed by the institutional review boards (IRB) of their hosting organisations and determined to be exempt from subsequent IRB.

## Definition of outcome

The outcome of this study is ongoing GI bleeding after admission to the ICU. Since this outcome variable is not encoded, blood transfusions were used as surrogate marker.

## Software

Models were developed in Python V.3.7 using data science packages including pandas V.0.25.3 (data wrangling),[20] NumPy V.1.17.5 (computations),[21] SciPy V.1.4.1 (hypothesis testing),[22] Scikit-learn V.0.22.1 (modelling)[23] and Hyperopt V.0.2.3 (hyperparameter optimisation).[24]

## Data preparation

We included non-pregnant adult patients (≥18 years old) admitted to the ICU and diagnosed with GI bleeding based on the International Classification of Diseases (ICD-9) codes (see table A1, online supplemental digital content 1,). For patients with multiple ICU admissions within a single hospitalisation event, only the first ICU stay was considered. The inclusion criteria for each database are further detailed in figure 1.

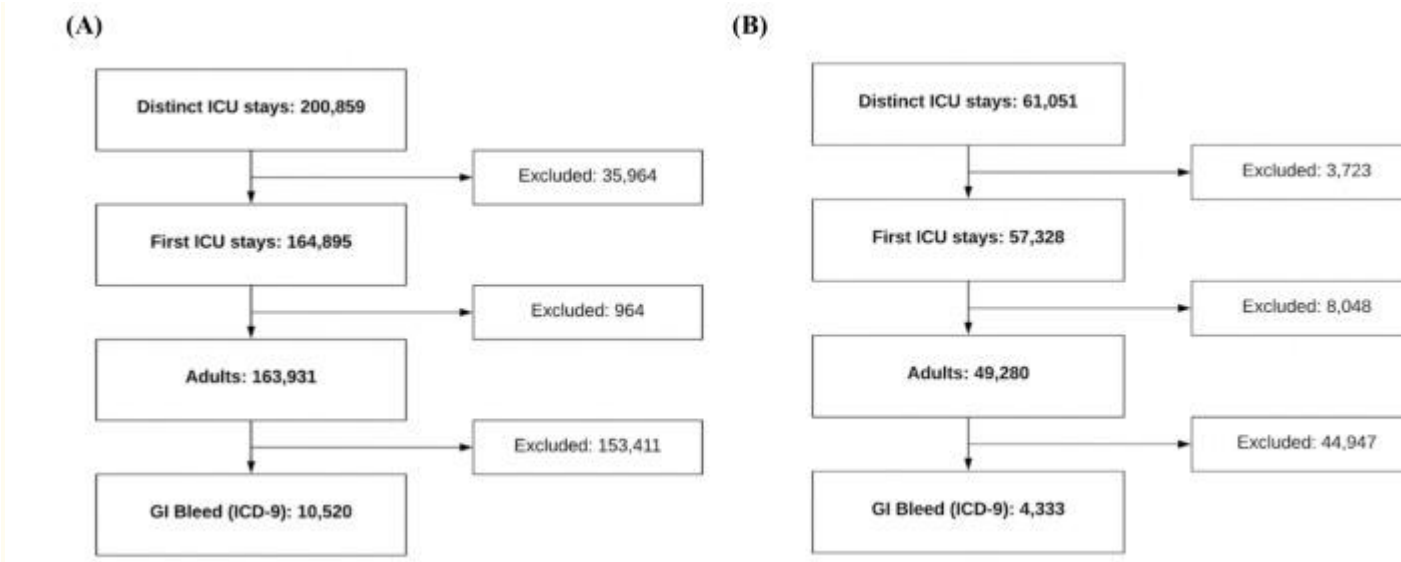


Figure 1
Inclusion criteria for the cohort extracted from the (A) eICU-CRD and (B) MIMIC-III. eICU-CRD, eICU Collaborative Research Database; ICD-9, International Classification of Diseases-9; ICU, intensive care unit; GI, gastrointestinal; MIMIC-III, Medical Information Mart for Intensive Care-III.

## Supplementary data

bmjhci-2020-100245supp001.pdf

Missing records were imputed with the last observation available carried forward. Patients missing their first value were imputed with the intra-subject median. In order to take into account the dynamics of the observed features within the training window (eg, increasing, decreasing trends), we adopted a feature engineering approach (see text, online supplemental digital content 2). Also, non-normally distributed features (skewness >3) were log-transformed[25] in order to obtain a normal distribution for improved model performance.

## Supplementary data

bmjhci-2020-100245supp002.pdf

Feature selection has been performed by recursively discarding features that do not reduce accuracy performance when eliminated. This procedure is called recursive feature elimination (RFE), a method used to remove non-predictive covariates with a greedy approach[26] (see text, online supplemental digital content 3). Final

input datasets gather 4333 first ICU admissions from the MIMIC-III database and 10520 first ICU admissions from the eICU-CRD along with 20 covariates. Input variables include several laboratory analyses and demographic information that are available in each database. Detailed information of these features is described in table 1.

## Table 1

List of covariates, the output variable and demographic information for each cohort. Continuous variables are stated as mean (IQR), otherwise are the number of occurrences. only a subset of these variables (selected by recursive feature elimination procedure) enters in the final models.

| | MIMIC-III (n=4314) | eICU-CRD (n=10 306) |
|---|---|---|
| *Demographics* | | |
| Age at admission (years) | 83.5 (56–81) | 76.7 (56–79) |
| Gender (n) | | |
| Male | 2491 | 5927 |
| Female | 1823 | 4379 |
| *Output variable (transfusion)* | | |
| Transfused patients (n, % wrt total number of patients) | 2077 (48.15%) | 2712 (26.31%) |
| *Covariates* | | |
| Heart rate (bpm) | 92.9 (79.0–105.7) | 94.0 (79.9–106.5) |
| Mean blood pressure (mm Hg) | 78.9 (68.5–87.8) | 78.4 (67.6–87.5) |
| Systolic blood pressure (mm Hg) | 114.5 (99.0–129.0) | 108.1 (93–121) |
| Diastolic blood pressure (mm Hg) | 60.3 (54.7–65.2) | 62.6 (56.0–68.2) |
| Respiratory rate (breaths/min) | 21.2 (18.0–24.0) | 21.9 (17.8–24.4) |
| Haematocrit (%) | 28.4 (23.8–32.6) | 26.5 (20.7–31.6) |
| Haemoglobin (g/L) | 97 (80–112) | 87 (67–104) |
| White blood cell ($\times 10^9$/L) | 11.8 (7.2–14.1) | 11.7 (7.4–14.4) |
| Platelet ($\times 10^9$/L) | 227.5 (137.0–286.0) | 207 (129.0–263.0) |
| Creatinine (mg/dL) | 1.79 (0.85–1.88) | 1.73 (0.80–1.90) |
| Blood urea nitrogen (mg/dL) | 39.5 (19.0–51.0) | 39.2 (19.0–51.0) |
| Potassium (mEq/L) | 4.34 (3.80–4.70) | 4.38 (3.80–4.80) |
| Bicarbonate (mEq/L) | 22.6 (20.0–26.0) | 22.7 (20.0–26.0) |

| | MIMIC-III (n=4314) | eICU-CRD (n=10 306) |
|---|---|---|
| Amount blood transfused (mL) | 601.0 (375.0–750.0) | 571.9 (324.0–700.0) |
| Glucose (mg/dL) | 160.2 (106.0–174.0) | 153.2 (105.0–176.0) |
| Albumin (g/dL) | 3.17 (3.2–3.2) | 2.96 (2.8–3.1) |
| Temperature (°C) | 36.3 (36.0–36.7) | 36.4 (36.4–36.5) |
| Partial thromboplastin time (s) | 37.3 (26.1–37.9) | 35.3 (26.0–37.0) |

Open in a separate window

eICU-CRD, eICU Collaborative Research Database; ICU, intensive care unit; MIMIC-III, Medical Information Mart for Intensive Care-III.

## Supplementary data

bmjhci-2020-100245supp003.pdf

## Prediction time windows

Several time windows were assessed for data extraction of the training/testing data and the data for the output variable (blood transfusion) that was predicted. Four different time windows starting from ICU admission (hour 0) were evaluated: training time from 0 to 3 hours/prediction time 4–24 hours, training time 0–4 hours/prediction time 5–24 hours, training time 0–5 hours/prediction time 6–24 hours, training time 0–6 hours/prediction time 7–24 hours. The training timeframe contains the covariates recorded during that time frame for each ICU stay. All training time windows include information recorded prior to the ICU admission (up to −1 hour). The prediction time window is when the surrogate variable (blood transfusion) was recorded (see figure 2).

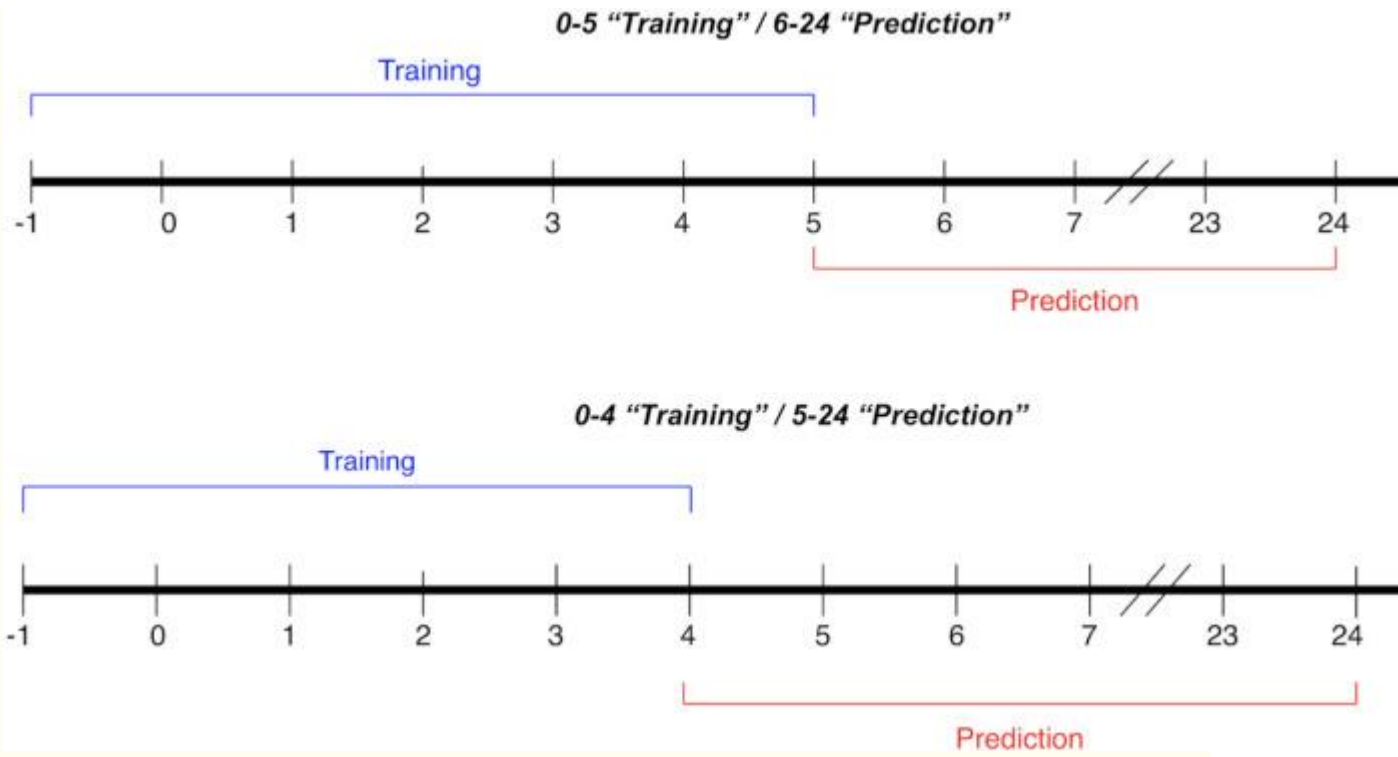


Figure 2
Graphical schema of the time windows.

This analysis helped us to find the optimal training/prediction time windows. The selected time windows were those that achieved the best predictive performance. In addition to that, the best training time window is the one that gathered the highest amount of data before a blood transfusion. Except from that, there is no other contextual detail that was considered during this analysis.

## Training and testing partitions

Several training/testing partitions and strategies were designed in order to fully exploit the information contained in both datasets. Specifically, both datasets are randomly divided into a test (25% of records) and training set (75% of records). A model is fitted on each of the training sets and on a combination of the two. All training subsets were split to perform 10-fold cross validation and to optimise model's hyperparameters. The testing subsets had data that were not used for training/validation.

Three different training sets were considered: (1) including MIMIC-III data only (n=3235); (2) including eICU-CRD data only (n=7729) and (3) a training set composed by 29.17% of MIMIC-III and 70.83% of

eICU-CRD (n=10 964). The performance of the models is then gauged on both the test sets, allowing for an external validation of the classifiers for a total of three models per each considered time window:

1. Train on MIMIC-III, internal validation on MIMIC, external validation on eICU-CRD.
2. Train on eICU-CRD, internal validation on eICU-CRD, external validation on MIMIC-III.
3. Train on MIMIC-III and eICU-CRD, internal validation on MIMIC-III and eICU-CRD.

## Predictive models

In order to improve the performance of individual machine learning models, the final classifier is determined as an ensemble of machine learning models combined together. To select the models for this ensemble, we assessed several classifiers. Hyperparameter tuning was performed through Bayesian optimisation[27] with a stratified 10-fold cross validation, where class imbalance is taken into account in the parameters of the models. This tuning is carried out with a customised loss function that takes into account accuracy and F1 score (see text, online supplemental digital content 4). This delivers a model based (and hence non arbitrary) procedure to find cut-off-thresholds that optimise jointly the accuracy, specificity and sensitivity of the model. By specifying the weights of F1 score and accuracy inside the custom loss function the model could be oriented to avoid false negative predictions (higher F1 score and recall) with a high accuracy. However, since the model also provides the probability that a patient will bleed the physician could in principle perform standard sensitivity–specificity trade-off decisions.

## Supplementary data

bmjhci-2020-100245supp004.pdf

Given that eICU-CRD exhibits target imbalance (26% transfused patients against 74% non-transfused patients) classifiers trained on this dataset are imbalance-aware in order not to skew predictions

towards the majority class (ie, predicting all patients as low risk patients, which is not desirable).

Permutation feature importance[28] of the five most important covariates is estimated for each model. Moreover, the partial dependence function[29] function of the outcome with respect to the most important variable is estimated (see text, online supplemental digital content 5).

Supplementary data

bmjhci-2020-100245supp005.pdf

In order to assess the goodness of the classifier during testing, we estimated the model's accuracy, sensitivity (recall or true classification positive rate), specificity (true negative classification rate) and area under the curve of the receiving operating curve (ROC-AUC).

To conclude, models are calibrated through Platt's scaling[30 31] to obtain reliable probability estimates. The effects of the calibration can be diagnosed visually with the calibration curves (see text, online supplemental digital content 6).

Supplementary data

bmjhci-2020-100245supp006.pdf

## Results

The best results are achieved when the models are trained on the MIMIC-III dataset (see table A2, online supplemental digital content 7), and the lowest values are observed in the models trained on the eICU-CRD data (see table A3, online supplemental digital content 8). When both datasets are merged (see table A4, online supplemental digital content 9), the performance does not improve considerably, but we can observe a significant improvement in terms of sensitivity.

Of note, the sensitivity obtained in the models trained with MIMIC-III is the highest among all other models; which indicates that it is better to detect true positive cases or patients that would require transfusion.

Supplementary data

bmjhci-2020-100245supp007.pdf

Supplementary data

bmjhci-2020-100245supp008.pdf

Supplementary data

bmjhci-2020-100245supp

proceed by

S.Praveenkumar

S.Abishek

V.Jayaprakash

M.Jayaseelan