

RESEARCH ARTICLE

# Prediction of Outcome in Acute Lower Gastrointestinal Bleeding Using Gradient Boosting

Lakshmana Ayaru<sup>1†\*</sup>, Petros-Pavlos Ypsilantis<sup>2‡</sup>, Abigail Nanapragasam<sup>1</sup>, Ryan Chang-Ho Choi<sup>1</sup>, Anish Thillanathan<sup>1</sup>, Lee Min-Ho<sup>1</sup>, Giovanni Montana<sup>2¶\*</sup>

**1** Department of Gastroenterology, Charing Cross and Hammersmith Hospitals, Imperial College Healthcare NHS Trust, London, United Kingdom, **2** Department of Biomedical Engineering, Kings College London, London, United Kingdom

‡ These two authors are joint first authors.

¶ These two authors are joint senior authors.

\* [Lakshmana.ayaru@imperial.nhs.uk](mailto:Lakshmana.ayaru@imperial.nhs.uk) (LA); [giovanni.montana@kcl.ac.uk](mailto:giovanni.montana@kcl.ac.uk) (GM)



## Abstract

### Background

There are no widely used models in clinical care to predict outcome in acute lower gastrointestinal bleeding (ALGIB). If available these could help triage patients at presentation to appropriate levels of care/intervention and improve medical resource utilisation. We aimed to apply a state-of-the-art machine learning classifier, gradient boosting (GB), to predict outcome in ALGIB using non-endoscopic measurements as predictors.

### Methods

Non-endoscopic variables from patients with ALGIB attending the emergency departments of two teaching hospitals were analysed retrospectively for training/internal validation ( $n=170$ ) and external validation ( $n=130$ ) of the GB model. The performance of the GB algorithm in predicting recurrent bleeding, clinical intervention and severe bleeding was compared to a multiple logic regression (MLR) model and two published MLR-based prediction algorithms (BLEED and Strate prediction rule).

### Results

The GB algorithm had the best negative predictive values for the chosen outcomes (>88%). On internal validation the accuracy of the GB algorithm for predicting recurrent bleeding, therapeutic intervention and severe bleeding were (88%, 88% and 78% respectively) and superior to the BLEED classification (64%, 68% and 63%), Strate prediction rule (78%, 78%, 67%) and conventional MLR (74%, 74% 62%). On external validation the accuracy was similar to conventional MLR for recurrent bleeding (88% vs. 83%) and therapeutic intervention (91% vs. 87%) but superior for severe bleeding (83% vs. 71%).

## OPEN ACCESS

**Citation:** Ayaru L, Ypsilantis P-P, Nanapragasam A, Choi RC-H, Thillanathan A, Min-Ho L, et al. (2015) Prediction of Outcome in Acute Lower Gastrointestinal Bleeding Using Gradient Boosting. PLoS ONE 10(7): e0132485. doi:10.1371/journal.pone.0132485

**Editor:** John Green, University Hospital Llandough, UNITED KINGDOM

**Received:** March 24, 2015

**Accepted:** June 15, 2015

**Published:** July 14, 2015

**Copyright:** © 2015 Ayaru et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper.

**Funding:** The authors received no specific funding for this work.

**Competing Interests:** The authors have declared that no competing interests exist.

## Conclusion

The gradient boosting algorithm accurately predicts outcome in patients with acute lower gastrointestinal bleeding and outperforms multiple logistic regression based models. These may be useful for risk stratification of patients on presentation to the emergency department.

## Introduction

Acute lower gastro-intestinal bleeding (ALGIB) is a common emergency increasing in incidence with age [1], and may be more common than acute upper gastrointestinal bleeding in the elderly [2]. The causes and severity are heterogeneous, e.g. large volume bleeding from diverticulosis or minor bleeding from colitis with the aetiology and outcome often obscure to the clinician at presentation.

The majority of ALGIB presentations (80–85%) resolve spontaneously with no adverse outcome and death is uncommon (2–4%) [3,4]. However a significant proportion (17–30%) undergo therapeutic intervention (angiographic embolisation, colonoscopic-based therapies or surgery) to control severe/recurrent bleeding [5,6]. Despite intervention re-bleeding occurs at rate of approximately 20% in the first year [7].

Due to concern about severe/recurrent bleeding or need for intervention routine clinical practice for the vast majority of patients with ALGIB who present to emergency department is admission to hospital for in-patient observation for variable number of days with a proportion undergoing endoscopy or radiological investigation. This strategy has the disadvantage of being invasive, expensive [8] and exposes patients to hospital acquired complications. If a reliable predictive model was available at the point of presentation to hospital low risk patients could be identified and triaged to outpatient management/a shorter inpatient stay. Resources could be then freed up for high-risk patients to be appropriately transferred to higher levels of care and undergo more aggressive investigation. Although multiple logistic regression (MLR) based scores have been developed for ALGIB [9,10] none have been recommended for routine clinical practice unlike the Glasgow Blatchford and Rockall scores for acute upper gastrointestinal bleeding [11]. Possible reasons include the lack of validation in diverse settings and modest accuracies in comparison with these scores. MLR based models may be limited in predicting outcome in ALGIB as they are based on the assumption that a linear combination of the observed features can be used to determine the probability of each particular outcome ignoring any variable interaction which may be the key for the accurate prediction.

In order to mitigate this potential limitation of MLR based scores and to accurately predict ALGIB outcome, we chose to implement and assess the performance of a non-parametric algorithm for classification, Gradient Boosting (GB) [12]. GB is a supervised machine learning algorithm, and is able to approximate the unknown functional mapping between the inputs, i.e. the non-endoscopic measurements, and the outputs, i.e. the ALGIB outcomes. Supervised learning algorithms are commonly trained on historical data consisting of examples of input-output pairs.

The GB algorithm embraces the notion of “ensemble learning”, whereby multiple simple learning algorithms are used jointly in order to obtain better predictive performance than could be achieved from any of the constituent learning algorithms [13]. Of clinical relevance are several reports demonstrating that ensemble learning classification models are accurate in predicting outcome in a variety of clinical settings [14–16].

In particular, the GB algorithm relies on decision trees as constituent or “base” predictive algorithms. Decision trees are statistical models that recursively partition the input space in order to find rules that are predictive of the output. The classical CART (Classification and Regression Tree) algorithm was popularized in the 1980s by Breiman et al. [17]. Compared to other machine learning methods, GB possesses several strengths: 1) it is less prone to overfitting [13] 2) it is robust to noise [18] 3) it has an internal mechanism to estimate error rates, 4) it provides indices of variable importance and 5) it can be used when the predictors are both continuous and categorical.

The aim of this study was to test whether the GB algorithm was able to accurately predict clinical outcomes in patients presenting to emergency departments with ALGIB using non-endoscopic variables available to clinicians at that time. We also set out to compare the performance of the GB approach with conventional MLR and two previously published multivariate logistic regression models [9,10]. We show in this study that the gradient boosting algorithm accurately predicts rebleeding, severe bleeding and clinical intervention in patients with acute lower gastrointestinal bleeding and outperforms multiple logistic regression based models.

## Methods

### Ethics statement

This study was performed by analysing an existing anonymised database of patients presenting with ALGIB collated for the purpose of audit to the emergency departments of Charing Cross and Hammersmith hospitals, London, UK (1<sup>st</sup> January 2007 to 31<sup>st</sup> December 2011). The study was approved by the Joint Research Compliance office at Imperial College Healthcare NHS Trust (ref 125HH25060). The office confirmed that no formal ethical review or informed consent was required as the study involved existing anonymised routinely collected data, no new data was being collected and there was no clinical intervention.

### Study design

The database had been generated retrospectively by identifying consecutive patients from electronic records who were aged 18 or over and presented to the emergency department with a primary diagnosis of ALGIB defined as PR bleeding (bright red or maroon coloured blood passed per rectum) within the previous three days. For this study we excluded patients with: 1) presentation most indicative/final diagnosis of an upper GI bleed (haematemesis, melaena, upper GI bleeding source detected at endoscopy or angiography), 2) inpatient ALGIB bleed, 3) ALGIB as a secondary admission symptom, 4) patients transferred from other hospitals and 5) incomplete patient data. Patients with haemodynamic instability and PR bleeding were required to have an upper GI endoscopy before inclusion into the study unless a definitive colonic source was identified.

Patients admitted to Charing Cross Hospital with ALGIB were analysed for training and internal validation of the GB algorithm. The external validation cohort consisted of patients admitted to Hammersmith Hospital over the same time period. Charing Cross and Hammersmith hospitals are large, busy general teaching hospitals with separate emergency departments and surgical teams.

### Definitions of variables and final diagnoses

Data on 39 previously published variables associated with need for intervention or adverse outcome in ALGIB (Table 1) were identified from the literature and collected from prospectively generated databases of presenting diagnoses, laboratory results, endoscopic, discharge and

**Table 1. Baseline Variables at initial ALGIB presentation.**

<b>Demographic</b>
Age
Gender
<b>History</b>
Use of omeprazole/lansoprazole
Use of NSAID drugs/antocoagulants
Alcoholism
Smoking
Nursing home resident
Colorectal polyp
Haemorrhoids
Diverticular disease history
Colonic AVM
Syncope
<b>Comorbidities</b>
Cardiovascular disease
Hypertension
Stroke history
COPD
Chronic renal failure
Diabetes mellitus
Dementia
Cancer
Chronic liver disease
Previous GI bleed history
Unstable comorbidities
<b>Initial assessment</b>
Heart rate
Systolic BP
Diastolic BP
Erratic mental status
Abdominal pain
Tender abdominal exam
Ongoing bleed in ED
Gross blood on DRE
<b>Baseline bloods</b>
Haemoglobin
Haematocrit
White blood cell count
Platelet
APTT
Prothrombin time
Urea
Creatinine

APTT, activated partial prothrombin time; AVM, arteriovenous malformation; DRE, digital rectal examination; ED emergency department

doi:10.1371/journal.pone.0132485.t001

coding databases. The variables were defined as follows: Unstable co-morbidity-any organ system abnormality that usually requires ICU admission, erratic mental status-clouding of consciousness due to any cause or presence of syncope confusion or coma, cardiovascular disease-history of angina, myocardial infarction, cardiomyopathy or heart failure, respiratory disease-current or past history of copd, liver disease-history or presence of jaundice, cirrhosis or portal hypertension and renal failure-creatinine >125micromol/l.

Definite colonic bleeding was defined as any signs of active or recent bleeding on endoscopy or angiography (stigmata of recent haemorrhage: active bleeding, a non bleeding visible vessel, or an adherent clot). Presumptive was defined haematochezia or blood per rectum and no suspicion of upper GI bleeding, with one or more potential bleeding sources below the ligament of Treitz.

### Gradient boosting

We deployed the GB algorithm, as originally proposed by Friedman [13], which has been successfully applied in a number of clinical applications [19–23]. GB is a non-parametric algorithm for supervised machine learning. It approximates the unknown functional mapping from input explanatory variables to corresponding output variables. The non-parametric nature of the GB algorithm enables the estimation of a functional mapping from non-endoscopic measurements to ALGIB outcomes without the need to decide a priori the parametric form of this function. By contrast, parametric models like logistic regression assume that the log odds depend linearly on the covariates, and this linearity may be insufficient to capture the complexity of relationship between inputs and outputs [24].

For the training of the GB algorithm we consider  $m$  patients with non-endoscopic input measurements and their corresponding ALGIB outcome in the form of  $(x_1, y_1), \dots, (x_m, y_m)$  pairs, where each  $\{x_i\}_{i=1}^m$  is a vector containing the non-endoscopic measurements for patient  $i$ . We seek to approximate the unknown function  $y = F(x^*)$  so that predictions can be made on a new patient for which we have observed the non-endoscopic input measurements  $x^*$ .

The GB algorithm relies on an iterative model fitting procedure making use of many simple predictive algorithms or “base” learners [13], and combine them to form more complex decision rules. The unknown function  $F$  is estimated by minimizing a loss function  $\mathcal{L}$  defined over the training set:

$$F = \operatorname{argmin}_F \sum_{i=1}^m \mathcal{L}(y_i, F(x_i)) \tag{1}$$

GB constructs an approximation  $F_N$  of  $F$  as a sum of  $N+1$  “base” learners constructed through  $N$  boosting iterations,  $F_N = \sum_{n=0}^N f_n$ . In our implementation, the “base” learners are regression trees<sup>12</sup>, which are particularly useful in clinical applications as they provide easy to interpret decision rules. GB starts with an initial base learner  $F_0$  and then applies a steepest descent step for the minimization of the loss function with respect to  $F_0$ . These two steps are repeated sequentially and each time a new learner is constructed to follow the direction along which the loss of the previous learner is minimized. The steepest descent method takes steps proportional to the negative gradient of the loss function in order to find the local minimum. More explicitly, the gradient of the loss function  $\mathcal{L}$  for each training point  $x_i$  at the iteration step  $n$  is given by

$$g_{i,n}(x_i) = \nabla_{F_{n-1}(x_i)} \mathcal{L}(y_i, F_{n-1}(x_i)), \quad 1 \leq i \leq m. \tag{2}$$

The gradient is defined only at the data points  $\{x_i\}_{i=1}^m$  and cannot be generalized to other  $x$  values. One way to enable generalization is to choose a regression tree  $h(x, a_n)$  that produces

$h_n = \{h(x_i, a_n)\}_1^m$  most parallel to  $-g_n \in \mathcal{R}^m$ . This regression tree can be obtained from the solution

$$a_n = \operatorname{argmin}_{a,\beta} \sum_{i=1}^m [-g_n(x_i) - \beta h(x_i, a)]^2, \tag{3}$$

where  $a_n$  are the parameters of the regression tree  $h(x, a_n)$  and  $\beta$  is the learning rate, which determines the contribution of each tree to the approximation. Having estimated the regression tree that is most highly correlated with  $-g_n(x)$  over the data distribution, the next update of the approximation  $F_N$  is given by

$$F_n(x) = F_{n-1}(x) + \gamma_n h(x, a_n), \tag{4}$$

which uses the optimal length,

$$\gamma_n = \operatorname{argmin}_{\gamma} \sum_{i=1}^m \mathcal{L}(y_i, F_{n-1}(x_i) + \gamma h(x_i, a_n)). \tag{5}$$

The GB algorithm is summarized in the following pseudo-code:

**Gradient Boosting Algorithm**

```

1  $F_0(x) = \operatorname{argmin}_p \sum_{i=1}^m \mathcal{L}(y_i, p)$ , where  $p$  is the response of the regression tree to
  the training data.
2 for  $n = 1 \rightarrow N$  do
3    $g_n = \nabla_{F_{n-1}(x)} \mathcal{L}(y, F_{n-1}(x))$  gradient at the training data  $x$ .
4    $a_n = \operatorname{argmin}_{a,\beta} \sum_{i=1}^m [-g_n(x_i) - \beta h(x_i, a)]^2$ , fit a regression tree  $h_n(x, a_n)$ 
5    $F_n(x) = F_{n-1}(x) + \gamma_n h(x, a_n)$ , for
6    $\gamma_n = \operatorname{argmin}_{\gamma} \sum_{i=1}^m \mathcal{L}(y_i, F_{n-1}(x_i) + \gamma h(x_i, a_n))$ 
7 end for
8 end Algorithm

```

After training, the parameters of the learned regression trees encode rules capturing the (possibly non-linear) relationship between non-endoscopic variables and the ALGIB outcome. A new patient with non-endoscopic measurements  $x^*$  is then assigned to a specific outcome class  $y$  simply by following the decision rules associated to that class.

To enable an optimal training of the GB algorithm, first we randomly divided our cohort of 170 patients collected at Charing Cross Hospital into training and validation datasets. Specifically, 70% of patients were assigned to the training set and the remaining 30% were utilized for internal validation. Internal validation was carried out with the objective of optimally tuning the GB hyperparameters (i.e. the number of boosting iterations, the depth of the regression trees and the learning rate) before assessing the performance of the algorithm externally on a cohort of patients admitted to Hammersmith Hospital (completely independent dataset).

For each one of the clinical outcomes in our studies we ranked the covariates using the internal GB mechanism for variable ranking, and selected the 10 best predictive variables for each outcome. We then investigated whether re-fitting the GB algorithm using only this reduced set of covariates would yield comparable performance.

**MLR and published MLR based models for ALGIB**

Conventional multiple logistic regression was applied to the Charing Cross and Hammersmith cohorts using the same 39 non-endoscopic as in GB. Moreover, two published MLR based models for ALGIB, the BLEED score [10] (persistent bleeding, low systolic blood pressure, elevated prothrombin time, erratic mental status and unstable comorbid disease) and Strate prediction rule [9] (heart rate  $\geq 100$ , systolic blood pressure  $\leq 115$ , syncope, non tender abdominal examination, rectal bleeding within first 4hrs of evaluation, aspirin use and  $> 2$  comorbid conditions) were calculated for the both cohorts.

## Outcomes

The outcomes measured were therapeutic intervention (endoscopic, angiographic, surgical), severe bleeding (defined as ongoing or recurrent bleeding) and recurrent bleeding. These outcomes were chosen as they indicate the need for inpatient care. Therapeutic intervention to stop the source of a bleed was included as this suggested the presence of an ongoing bleed that was not resolving spontaneously. Definitions for the three outcomes were taken from the published literature [5,9,25,26]. Severe bleeding was defined as the following: continued bleeding in the first 24 hours of hospitalisation (defined as a RBC transfusion of  $\geq 2$  units, and/or a haematocrit decrease of  $\geq 20\%$ ), or recurrent bleeding after 24 hours of stability (defined as more than one transfusion of RBCs, a further haematocrit decrease of  $\geq 20\%$ , or readmission for ALGIB within 1 week of discharge). Recurrent bleeding was defined as recurrent haematochezia after 24 hours of stabilisation during which no active bleeding was observed, associated with any of the following as a new finding: decrease in haemoglobin of  $\geq 2\text{g/dl}$ , decrease in haematocrit of  $\geq 5\%$ , haemodynamic instability, or having an additional RBC transfusion ( $\geq 2$  units received in total).

## Statistical Analysis

The following statistical figures to predict severe bleeding, recurrent bleeding and therapeutic intervention were derived for all models in the internal and external validation cohorts: Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV) and Accuracy (sum of correct predictions over total predictions) using HDS Epimax, 2004 and Graph Pad Prism software. Comparison of continuous and categorical data (between the internal and external validation cohorts) was carried out using Mann-Whitney U and Fisher exact tests respectively. A two-tailed significance of 5% was used in all comparisons.

## Results

### Characteristics and clinical outcomes of cohorts

For the Charing Cross cohort (CXC) following the initial search through emergency and endoscopy databases at Charing Cross hospital 174 patients were identified as having had a history of ALGIB and presentation to the emergency department. Four patients were excluded due to incomplete information for risk scoring. The 170 remaining patients made up the training/internal validation cohort. The same search process was used to compile the Hammersmith cohort (HC). 133 patients were attended the emergency department of Hammersmith hospital with a primary diagnosis of ALGIB. Of these patients three were excluded due to incomplete data for risk scoring. The remaining 130 patients made up the external validation cohort. The accuracy of data collection was by random re-assessment of 10% of notes by another author (LA).

The demographic characteristic, clinical features and final diagnoses of the internal and external validation cohorts are shown in Tables 2 and 3. The patients in the internal Charing Cross cohort were similar to the Hammersmith cohort with regard to sex ratio, median age and length of stay. Nearly all patients (98–100%) in both cohorts were admitted to hospital for management of ALGIB which lasted a median of 4 days. Upper GI endoscopy to rule out an upper GI bleed was carried out in 20% of cases in the internal cohort and 14% of cases in the external cohort. Patients in both cohorts were similar in terms of undergoing a lower endoscopy procedure (74% vs. 81%) but in-patient colonoscopy was more common in the Charing Cross cohort (76% vs. 46%). This consisted of colonoscopy (56% Charing cross cohort, 45% Hammersmith cohort) with the remainder having flexible sigmoidoscopy, rigid sigmoidoscopy

**Table 2. Characteristics and outcomes of cohorts.**

Characteristic	Charing Cross (n = 170)	Hammersmith (n = 130)	P value
Male	90 (53%)	69 (53%)	ns
Median Age (range)	70 (16–99)	70 (17–101)	ns
Discharge within 24 hrs	4 (2%)	0 (0%)	ns
Median, mean hospital stay (range)	4, 7 (0–102)	4, 7 (0–182)	ns
Blood cell transfusion	58 (34%)	23 (18%)	0.001
Had lower GI Endoscopy	125 (74%)	105 (81%)	ns
<b>Outcomes</b>			
Severe bleeding	60 (35%)	28 (22%)	0.01
Recurrent bleeding	34 (20%)	19 (14%)	ns
Therapeutic intervention	26 (15%)	9 (7%)	0.02
endoscopic	9 (5%)	4 (3%)	ns
angiographic	9 (5%)	4 (3%)	ns
surgery	8 (4.7%)	1 (0.7%)	ns
Death	4 (2.3%)	3 (2.3%)	ns

ns-not significant

doi:10.1371/journal.pone.0132485.t002

or proctoscopy. A CT or mesenteric angiogram was carried out in 8% of patients in the Charing Cross cohort and 5% in the Hammersmith cohort.

The three most common diagnoses were diverticulosis, colitis and anorectal disorders such as haemorrhoids. Final diagnoses were similar in both cohorts apart from a presumptive

**Table 3. Final Diagnoses in Cohorts.**

	Charing Cross (n = 170)	Hammersmith (n = 130)
<b>Definite</b>		
Diverticulosis and its complications	13	16
Colitis	14	5
Anorectal (including varices)	6	12
Neoplasia and post-neoplasia therapy	14	5
Angiodysplasia	3	4
Isolated large bowel ulcers	3	1
Coagulation disorders	1	0
Small bowel bleeding	1	1
<b>Presumptive</b>		
Diverticulosis and its complications*	44	13
Colitis	11	11
Anorectal (including varices)	11	9
Neoplasia and post-neoplasia therapy	2	4
Angiodysplasia	3	2
Coagulation disorders	0	1
solated large bowel ulcers	4	0
Small bowel bleed	1	0
<b>No diagnosis made*</b>	39	46

\* significant difference between cohorts p<0.05

doi:10.1371/journal.pone.0132485.t003



diagnosis of a diverticular bleed which was more common in the Charing Cross cohort and no diagnosis made which was more common in the Hammersmith cohort.

Therapeutic intervention for bleeding (endoscopic therapy, angiographic embolisation or surgery) was more common in the Charing Cross cohort. Endoscopic intervention in both cohorts consisted of clipping, APC and banding. Angiographic embolisation of colonic vessels was carried out in nine patients in the internal cohort and four in the external cohort. Blood transfusion was significantly more common in the Charing Cross cohort (mean 1.6 units vs. 0.9 units) as was severe bleeding. There was no significant difference between the cohorts in re-bleeding or death. All patients who died were  $\geq 65$  years of age and the causes of death were cancer (n = 2) cardiac failure (n = 1), pneumonia (n = 1), colonic ischemia (n = 1), pulmonary hypertension (n = 1) and unknown (n = 1). No patient died because of uncontrolled bleeding.

### Predictive performance of gradient boosting and multiple logistic regression

The best GB algorithms using the 39 variables had predictive accuracies of 88%, 91% and 83% (Table 4) for recurrent bleeding, therapeutic intervention and severe bleeding respectively. The accuracies were similar in the Charing Cross and Hammersmith cohorts. The positive predictive value of all GB algorithms were not high in either cohort although importantly for clinical decision making the negative predictive values were high (88–98%) for the three outcomes.

**Table 4. Predictive performance of models in Charing Cross (CXC) and Hammersmith (HC) cohorts.**

Outcome variable	Accuracy	Sensitivity	Specificity	PPV	NPV
<b>GB model: Recurrent bleeding</b>					
CXC	88	67	91	50	95
HC	88	57	91	50	94
<b>GB model: Therapeutic intervention</b>					
CXH	88	80	89	44	98
HC	91	60	92	27	98
<b>GB model: Severe bleeding</b>					
CXH	78	73	80	61	88
HC	83	57	89	58	90
<b>MLR model: Recurrent bleeding</b>					
CXC	74	22	85	25	83
HC	83	20	85	6	95
<b>MLR model: Therapeutic intervention</b>					
CXC	74	16	82	11	97
HC	87	20	90	9	96
<b>MLR model: Severe Bleeding</b>					
CXC	62	46	69	39	75
HC	71	35	83	42	80
<b>BLEED classification (CXC only)</b>					
Recurrent bleeding	64	24	75	21	77
Therapeutic intervention	68	27.5	76	19	84
Severe bleeding	63	33	79	44	69
<b>Strate prediction rule (CXC only)cut off &gt;3</b>					
Recurrent bleeding	78	16	94	33	81
Therapeutic intervention	78	4	92	8	84
Severe bleeding	67	13	96	66	66

doi:10.1371/journal.pone.0132485.t004

**Table 5. Top ten variable importance using gradient boosting models.**

	Contribution %
<b>Severe bleeding</b>	
Platelet count	13.4
APTT	13.0
Haematocrit	12.4
Urea	10.9
Creatinine	9.7
Prothrombin time	8.9
Diastolic blood pressure	6.8
Heart rate	4.1
Systolic blood pressure	3.9
Alcohol abuse	3.9
<b>Therapeutic Intervention</b>	
Haemoglobin	15.7
Diastolic blood pressure	13.9
haematocrit	9.5
APTT	9.0
Creatinine	8.2
Fresh blood on PR	7.1
Prothrombin time	6.7
Heart rate	5.0
Past medical history of colorectal polyp	3.4
Use of NSAIDs or anticoagulants	3.4
<b>Recurrent bleeding</b>	
Creatinine	19.1
Haemoglobin	18.8
Age	17.9
Diabetes	13.2
APTT	11.5
Diastolic blood pressure	6.8
Heart Rate	4.6
Urea	4.4
Alcoholism	2.4
Total number of co-morbidities	1.3

doi:10.1371/journal.pone.0132485.t005

The top ten contributing predictors used in the GB algorithms for each outcome are listed on [Table 5](#) in descending order. Four variables (heart rate, diastolic blood pressure, creatinine and APTT) were in the top ten most frequently used variables for all outcomes. On internal validation the accuracy of the GB models for predicting recurrent bleeding, therapeutic intervention and severe bleeding was (88%, 88% and 78% respectively) and superior to the BLEED classification (64%, 68% and 63%) ([Table 4](#)), Strate prediction rule (78, 78, 67%) and conventional MLR (74%, 74% 62%). On external validation the accuracy of the GB algorithm was similar to conventional MLR for recurrent bleeding (88% vs. 83%), and therapeutic intervention (91% vs. 87%) but superior for severe bleeding (83 vs 71%). GB models using just the top ten predictive variables were less accurate in predicting rebleeding, severe bleeding and therapeutic intervention than those with the full set of 39 variables by 8–10% on average.

## Discussion

There is a need for non-endoscopic risk scores to help risk stratify patients with ALGIB for early discharge/outpatient management or higher levels of care and thereby utilise resources efficiently. Current clinical practice includes colonoscopic-based triage which is invasive, adequate preparation difficult to achieve and treatable stigmata of haemorrhage infrequent [1]. This study has shown that a GB algorithm based on clinical and laboratory variables was accurate (>80%) in predicting the clinical outcomes of recurrent bleeding, therapeutic intervention and severe bleeding.

GB had high negative predictive values (88–98%) in both Charing Cross and Hammersmith Cohorts. This suggests that these models may be useful to triage patients into a low-risk group who could be managed with an abbreviated stay in hospital avoiding high levels of care or as outpatients. The median and mean inpatient stay in the Charing Cross and Hammersmith cohorts was four and seven days respectively and therefore a reduction in this would allow for significant cost-savings and decrease exposure of patient to hospital associated hazards such as infections.

A particular strength of this study is the validation and good performance in an external cohort with a lower incidence of severe bleeding indicating the algorithm can maintain accuracy in a different setting. Our GB algorithm used only non-endoscopic variables available to the clinician in the emergency department and therefore has clinical applicability for decision-making. We would however emphasise that such a model is not aimed at replacing experienced decision making but rather aiding the process. This is the case with recommended risk scores for acute upper gastrointestinal bleeding such as the Rockall and Glasgow Blatchford scores [11].

One study to our knowledge has examined an ensemble machine learning model in acute gastrointestinal bleeding and was developed to identify the bleeding source, need for resuscitation and those who require urgent endoscopy [27]. This study differs substantially from ours in that a random forests algorithm was used by building classification trees ignoring the error of the previous tree in the sequence [28]. Also a mixed population of patients with both acute upper, middle and lower gastrointestinal bleeding was studied (no numbers given for ALGIB), there was no comparison with previously published scores and the model was not tested in an external cohort. Nevertheless accuracies of >75–80% were found for the studied outcomes providing evidence of the utility of the ensemble machine learning techniques.

The performance of the GB algorithm in our study was superior to MLR and two published MLR based models. Ensemble machine learning models have been shown to be more accurate than conventional logistic regression to classify disease or predict outcome in a variety of clinical settings [29–31]. Theoretical reasons for this are that logistic regression predicts outcomes based on linear combinations of independent variables by fitting a single model that best explain the relationship between observed values and outcome. On the other hand, the rationale of the GB algorithm to fit many simple models whose predictions are then combined can produce a good fit of the predicted outcome values to the observed values, even if the specific nature of the relationship between the predictor variables and the corresponding outcome is complex (e.g. nonlinear, interacted or noisy with outliers). Also, unlike multiple logistic regression, GB method can handle a large number of input variables and generate an internal unbiased estimate of the generalization error as the simple classification tree estimation progresses. Finally, the stage-wise model fitting procedure of the GB algorithm allows to automatically assess the influence of each non-endoscopic variable in the construction of a robust classification rule [12].

Other explanations are that the BLEED score [10] was designed to predict a composite endpoint of in-hospital complication (recurrent haemorrhage, surgery to control haemorrhage

and hospital mortality) rather than the end-points we examined. The Strate prediction rule [9] was designed to predict severe bleeding which we studied but in our cohort performed least well for this outcome and better for recurrent bleeding and need for clinical intervention. Artificial neural networks, another machine learning classifier, have also been shown to be accurate in predicting re-bleeding and clinical intervention in ALGIB [5]. However for our cohort, the classification performance of neural networks was found inferior to GB in ALGIB (data not shown).

Our study has a number of potential limitations: First the database of patients was collated retrospectively and relied on the inherent accuracy of patient records. The majority of data collected was however quantitative and also available from prospectively generated electronic laboratory, endoscopic and patient records. Second the GB requires the input of many more variables than BLEED or Strate prediction rule (39 vs. 5 vs. 8) which increases complexity. Reduction in the number of variables used in the GB model to 10 led to decreased accuracy of 8–10% for the studied outcomes which would compromise clinical utility of the GB algorithm. Importantly however in our experience input of data for the 39 variables into the programme takes less than 5 minutes and therefore would be suitable for use in an emergency department/ward particularly given the explosion of smart phone apps which allow for quick data entry with drop down menus where only positive inputs are required. This would be similar to endoscopy reports which are generated electronically, typically require >50 pieces of data and are used in routine clinical practice. Third the decision to give blood transfusion and apply endoscopic therapy was not protocol based which could have led to bias. In mitigation our analysis showed that endoscopic therapy was consistently applied in this study according to consensus guidelines (data not shown) and therefore limits this as a potential source of bias. Finally death was not examined as an outcome due its infrequent nature in our cohort. Death is rare in ALGIB occurring in <4% in large series and generally occurs in those with co-morbid conditions [6] and after an in-patient bleed [32] that latter which was an exclusion criteria for our study. Future work to examine the GB model in a larger cohort could be undertaken to examine its utility in predicting death. To determine the impact of the model in predicting outcomes in ALGIB a randomised study of the GB model plus routine clinical decision-making versus routine clinical decision-making could be performed.

In summary, gradient boosting accurately predicts outcome in patients with acute lower gastrointestinal bleeding. This machine learning approach has the potential to aid in the risk stratification of patients with ALGIB on presentation to the emergency department.

## Author Contributions

Conceived and designed the experiments: LA PY GM. Performed the experiments: LA PY RC AT LM GM. Analyzed the data: LA PY AN RC GM. Contributed reagents/materials/analysis tools: LA PY GM. Wrote the paper: LA PY GM.

## References

1. Strate LL, Naumann CR. The role of colonoscopy and radiological procedures in the management of acute lower intestinal bleeding. *Clin Gastroenterol Hepatol*. 2010; 8: 333–343. doi: [10.1016/j.cgh.2009.12.017](https://doi.org/10.1016/j.cgh.2009.12.017) PMID: [20036757](https://pubmed.ncbi.nlm.nih.gov/20036757/)
2. Lanás A, García-Rodríguez LA, Polo-Tomás M, Ponce M, Alonso-Abreu I, Perez-Aisa MA, et al. Time trends and impact of upper and lower gastrointestinal bleeding and perforation in clinical practice. *Am J Gastroenterol*. 2009; 104: 1633–1641. doi: [10.1038/ajg.2009.164](https://doi.org/10.1038/ajg.2009.164) PMID: [19574968](https://pubmed.ncbi.nlm.nih.gov/19574968/)
3. Barnert J, Messmann H. Diagnosis and management of lower gastrointestinal bleeding. *Nat Rev Gastroenterol Hepatol*. 2009; 6: 637–646. doi: [10.1038/nrgastro.2009.167](https://doi.org/10.1038/nrgastro.2009.167) PMID: [19881516](https://pubmed.ncbi.nlm.nih.gov/19881516/)
4. Farrell JJ, Friedman LS. Review article: the management of lower gastrointestinal bleeding. *Aliment Pharmacol Ther*. 2005; 21: 1281–1298. doi: [10.1111/j.1365-2036.2005.02485.x](https://doi.org/10.1111/j.1365-2036.2005.02485.x) PMID: [15932359](https://pubmed.ncbi.nlm.nih.gov/15932359/)

5. Das A, Ben-Menachem T, Cooper GS, Chak A, Sivak MV, Gonet JA, et al. Prediction of outcome in acute lower-gastrointestinal haemorrhage based on an artificial neural network: internal and external validation of a predictive model. *Lancet*. 2003; 362: 1261–1266. doi: [10.1016/S0140-6736\(03\)14568-0](https://doi.org/10.1016/S0140-6736(03)14568-0) PMID: [14575969](https://pubmed.ncbi.nlm.nih.gov/14575969/)
6. Venkatesh PGK, Njei B, Sanaka MR, Navaneethan U. Risk of comorbidities and outcomes in patients with lower gastrointestinal bleeding—a nationwide study. *Int J Colorectal Dis*. 2014; 29: 953–960. doi: [10.1007/s00384-014-1915-x](https://doi.org/10.1007/s00384-014-1915-x) PMID: [24913253](https://pubmed.ncbi.nlm.nih.gov/24913253/)
7. Aoki T, Nagata N, Niikura R, Shimbo T, Tanaka S, Sekine K, et al. Recurrence and Mortality Among Patients Hospitalized for Acute Lower Gastrointestinal Bleeding. *Clin Gastroenterol Hepatol*. 2014; doi: [10.1016/j.cgh.2014.06.023](https://doi.org/10.1016/j.cgh.2014.06.023)
8. Parker DR, Luo X, Jalbert JJ, Assaf AR. Impact of upper and lower gastrointestinal blood loss on health-care utilization and costs: a systematic review. *J Med Econ*. 2011; 14: 279–287. doi: [10.3111/13696998.2011.571328](https://doi.org/10.3111/13696998.2011.571328) PMID: [21456948](https://pubmed.ncbi.nlm.nih.gov/21456948/)
9. Strate LL, Saltzman JR, Ookubo R, Muting ML, Syngal S. Validation of a clinical prediction rule for severe acute lower intestinal bleeding. *Am J Gastroenterol*. 2005; 100: 1821–1827. doi: [10.1111/j.1572-0241.2005.41755.x](https://doi.org/10.1111/j.1572-0241.2005.41755.x) PMID: [16086720](https://pubmed.ncbi.nlm.nih.gov/16086720/)
10. Kollef MH, O'Brien JD, Zuckerman GR, Shannon W. BLEED: a classification tool to predict outcomes in patients with acute upper and lower gastrointestinal hemorrhage. *Crit Care Med*. 1997; 25: 1125–1132. PMID: [9233736](https://pubmed.ncbi.nlm.nih.gov/9233736/)
11. Dworzynski K, Pollit V, Kelsey A, Higgins B, Palmer K, Guideline Development Group. Management of acute upper gastrointestinal bleeding: summary of NICE guidance. *BMJ*. 2012; 344: e3412. doi: [10.1136/bmj.e3412](https://doi.org/10.1136/bmj.e3412) PMID: [22695897](https://pubmed.ncbi.nlm.nih.gov/22695897/)
12. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. Springer Series in Statistics [Internet]. Springer. New York, NY: Springer; 2009. Available: <http://uzpdf.namiofpreblecounty.com/the-elements-of-trevor-68729210.pdf>
13. Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Stat*. 2001; 29: 1189–1232. doi: [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451)
14. Chen H-Y, Yu S-L, Chen C-H, Chang G-C, Chen C-Y, Yuan A, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med*. 2007; 356: 11–20. doi: [10.1056/NEJMoa060096](https://doi.org/10.1056/NEJMoa060096) PMID: [17202451](https://pubmed.ncbi.nlm.nih.gov/17202451/)
15. Gurm HS, Kooiman J, LaLonde T, Grines C, Share D, Seth M. A random forest based risk model for reliable and accurate prediction of receipt of transfusion in patients undergoing percutaneous coronary intervention. *PLoS One*. 2014; 9: e96385. doi: [10.1371/journal.pone.0096385](https://doi.org/10.1371/journal.pone.0096385) PMID: [24816645](https://pubmed.ncbi.nlm.nih.gov/24816645/)
16. Peng S-Y, Chuang Y-C, Kang T-W, Tseng K-H. Random forest can predict 30-day mortality of spontaneous intracerebral hemorrhage with remarkable discrimination. *Eur J Neurol Off J Eur Fed Neurol Soc*. 2010; 17: 945–950. doi: [10.1111/j.1468-1331.2010.02955.x](https://doi.org/10.1111/j.1468-1331.2010.02955.x)
17. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. CRC press; 1984.
18. Kalai A, Servedio RA. Boosting in the Presence of Noise. Proceedings of the Thirty-fifth Annual ACM Symposium on Theory of Computing. New York, NY, USA: ACM; 2003. pp. 195–205. doi: [10.1145/780542.780573](https://doi.org/10.1145/780542.780573)
19. Dodd S, Berk M, Kelin K, Zhang Q, Eriksson E, Deberdt W, et al. Application of the Gradient Boosted method in randomised clinical trials: Participant variables that contribute to depression treatment efficacy of duloxetine, SSRIs or placebo. *J Affect Disord*. 2014; 168: 284–293. doi: [10.1016/j.jad.2014.05.014](https://doi.org/10.1016/j.jad.2014.05.014) PMID: [25080392](https://pubmed.ncbi.nlm.nih.gov/25080392/)
20. Xie J, Coggeshall S. Prediction of transfers to tertiary care and hospital mortality: A gradient boosting decision tree approach. *Stat Anal Data Min*. 2010; 3: 253–258. doi: [10.1002/sam.10079](https://doi.org/10.1002/sam.10079)
21. Weiss JC, Page D, Peissig PL, Natarajan S, McCarty C. Statistical Relational Learning to Predict Primary Myocardial Infarction from Electronic Health Records. Proc Innov Appl Artif Intell Conf Spons Am Assoc Artif Intell Innov Appl Artif Intell Conf. 2012;2012.
22. Chen Y, Jia Z, Mercola D, Xie X. A gradient boosting algorithm for survival analysis via direct optimization of concordance index. *Comput Math Methods Med*. 2013;2013. doi: [10.1155/2013/873595](https://doi.org/10.1155/2013/873595)
23. Martin R, Rose D, Yu K, Barros S. Toxicogenomics strategies for predicting drug toxicity. *Pharmacogenomics*. 2006; 7: 1003–1016. doi: [10.2217/14622416.7.7.1003](https://doi.org/10.2217/14622416.7.7.1003) PMID: [17054411](https://pubmed.ncbi.nlm.nih.gov/17054411/)
24. Breiman L. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Stat Sci*. 2001; 199–231. doi: [10.1214/ss/1009213726](https://doi.org/10.1214/ss/1009213726)
25. Strate LL, Orav EJ, Syngal S. Early predictors of severity in acute lower intestinal tract bleeding. *Arch Intern Med*. 2003; 163: 838–843. doi: [10.1001/archinte.163.7.838](https://doi.org/10.1001/archinte.163.7.838) PMID: [12695275](https://pubmed.ncbi.nlm.nih.gov/12695275/)

26. Velayos FS, Williamson A, Sousa KH, Lung E, Bostrom A, Weber EJ, et al. Early predictors of severe lower gastrointestinal bleeding and adverse outcomes: a prospective study. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc.* 2004; 2: 485–490.
27. Chu A, Ahn H, Halwan B, Kalmin B, Artifon ELA, Barkun A, et al. A decision support system to facilitate management of patients with acute gastrointestinal bleeding. *Artif Intell Med.* 2008; 42: 247–259. doi: [10.1016/j.artmed.2007.10.003](https://doi.org/10.1016/j.artmed.2007.10.003) PMID: [18063351](https://pubmed.ncbi.nlm.nih.gov/18063351/)
28. Breiman L. Random Forests. *Mach Learn.* 2001; 45: 5–32.
29. Casanova R, Saldana S, Chew EY, Danis RP, Greven CM, Ambrosius WT. Application of random forests methods to diabetic retinopathy classification analyses. *PLoS One.* 2014; 9: e98587. doi: [10.1371/journal.pone.0098587](https://doi.org/10.1371/journal.pone.0098587) PMID: [24940623](https://pubmed.ncbi.nlm.nih.gov/24940623/)
30. Maroco J, Silva D, Rodrigues A, Guerreiro M, Santana I, de Mendonça A. Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Res Notes.* 2011; 4: 299. doi: [10.1186/1756-0500-4-299](https://doi.org/10.1186/1756-0500-4-299) PMID: [21849043](https://pubmed.ncbi.nlm.nih.gov/21849043/)
31. Austin PC, Lee DS, Steyerberg EW, Tu JV. Regression trees for predicting mortality in patients with cardiovascular disease: what improvement is achieved by using ensemble-based methods? *Biom J Biom Z.* 2012; 54: 657–673. doi: [10.1002/bimj.201100251](https://doi.org/10.1002/bimj.201100251)
32. Strate LL, Ayanian JZ, Kotler G, Syngal S. Risk factors for mortality in lower intestinal bleeding. *Clin Gastroenterol Hepatol.* 2008; 6: 1004–1010. doi: [10.1016/j.cgh.2008.03.021](https://doi.org/10.1016/j.cgh.2008.03.021) PMID: [18558513](https://pubmed.ncbi.nlm.nih.gov/18558513/)