Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine Learning Models for Predicting Active Bleeding in Patients Presenting to the Emergency Department with Upper Gastrointestinal Bleeding: A Retrospective Cross-Sectional Study

Melih Uçan

ucan.melih@gmail.com

İstanbul Kanuni Sultan Süleyman Eğitim ve Araştırma Hastanesi

Serkan Doğan

İstanbul Kanuni Sultan Süleyman Eğitim ve Araştırma Hastanesi

Bilal Yeniyurt

İstanbul Kanuni Sultan Süleyman Eğitim ve Araştırma Hastanesi

Burçe Serra Koçkan Pamuk

Marmara Üniversitesi Eğitim ve Araştırma Hastanesi

Ahmed Edizer

Eyüp Sultan State Hospital

Süreyya Tuba Fettahoğlu

Haseki Eğitim ve Araştırma Hastanesi

Salih Fettahoğlu

İstanbul Kanuni Sultan Süleyman Eğitim ve Araştırma Hastanesi

Utku Murat Kalafat

İstanbul Kanuni Sultan Süleyman Eğitim ve Araştırma Hastanesi

Ramiz Yazıcı

İstanbul Kanuni Sultan Süleyman Eğitim ve Araştırma Hastanesi

---

Research Article

Keywords: upper gastrointestinal bleeding, machine learning, emergency department, endoscopy, Glasgow–Blatchford Score, Forrest classification, clinical prediction model, artificial intelligence

Posted Date: June 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9950485/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Emergency Medicine on September 11th, 2026. See the published version at https://doi.org/10.1186/s12873-026-01772-9.

Abstract

Background

Upper gastrointestinal (GI) bleeding is a common emergency department (ED) presentation with substantial morbidity and mortality. Endoscopy is the diagnostic and therapeutic cornerstone, yet pre-endoscopic risk-stratification tools such as the Glasgow–Blatchford Score (GBS) are highly sensitive but suffer from low specificity, leading to a high rate of negative endoscopies. We aimed to develop and internally validate supervised machine-learning (ML) models to predict endoscopically confirmed active bleeding (Forrest Ia/Ib) using routinely available admission data, with the goal of improving specificity and reducing unnecessary endoscopic procedures.

Methods

A retrospective, observational, cross-sectional, analytical study was conducted in the ED of a tertiary-care academic hospital after ethics committee approval. Patients aged ≥ 18 years who presented between 1 January 2015 and 1 August 2022 with a working diagnosis of upper GI bleeding and underwent endoscopy were screened. Demographics, vital signs, digital rectal examination findings, comorbidities, complete blood count, biochemistry, coagulation, venous blood gas parameters, GBS, and endoscopic Forrest classification were extracted. Active bleeding was defined as Forrest 1a or 1b. A priori power analysis indicated a target sample size of 1146; 1009 patients met inclusion criteria, of whom 666 (≈ 70%) were randomly allocated to the training set and 343 (≈ 30%) to the internal validation/test set using Python v3.9.0. Five ML algorithms — Gaussian Naive Bayes (GNB), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Random Forest, and Logistic Regression — were trained on admission features and evaluated on the held-out set. Performance was summarised by sensitivity, specificity, positive and negative likelihood ratios, positive and negative predictive values (PPV/NPV), and accuracy. Group comparisons in the test set were performed with Chi-square/Fisher's exact, independent-samples t and Mann–Whitney U tests in SPSS v24.0 (two-sided $p < 0.05$).

Results

Of 343 patients in the test set, 30 (8.7%) had endoscopically confirmed active bleeding. Compared with the no-active-bleeding group, patients with active bleeding had significantly higher GBS (13.07 ± 1.93 vs. 12.19 ± 2.36; $p = 0.039$) and lower platelet counts (224.37 ± 95.58 vs. 267.65 ± 118.78 × 10^3/μL; $p = 0.049$). Other admission parameters (age, sex, vital signs, hemoglobin, BUN, INR, pH, lactate) did not differ significantly. Forrest classification in the 153 patients with positive endoscopic findings was Forrest 3 in 58.2%, Forrest 1b in 17.0%, Forrest 1a in 2.6%, Forrest 2a in 8.5%, Forrest 2b in 8.5%, and Forrest 2c in 5.2%. Compared against the reference standard, the GNB, SVM, Random Forest, and Logistic Regression models showed a statistically significant association with active bleeding ($p = 0.017$, $p < 0.001$, $p < 0.001$, $p < 0.001$, respectively), whereas the KNN model did not ($p = 0.241$). Random Forest

achieved the highest sensitivity (80.0%) and NPV (97.0%), Logistic Regression sensitivity 70.0% / NPV 95.86%, and SVM sensitivity 63.33% / NPV 95.14%. KNN had the highest specificity (99.36%) and accuracy (90.96%) but very low sensitivity (3.33%).

Conclusions

In ED patients undergoing endoscopy for suspected upper GI bleeding, ML models trained on routinely collected admission data — particularly Random Forest, Logistic Regression and SVM — achieved high negative predictive values (≥ 95%) for endoscopically confirmed active bleeding. Given the 8.7% prevalence of active bleeding in our cohort, these models may help safely identify patients who can defer urgent endoscopy and complement, rather than replace, existing risk scores such as the GBS. Prospective multicentre validation is required before clinical deployment.

BACKGROUND

The annual incidence of upper gastrointestinal (GI) bleeding ranges from approximately 103 to 172 cases per 100,000 population, with reported mortality of 2–10% [1]. The condition is more common in men and incidence rises with age [2]. Morbidity varies with bleeding severity, baseline health status, and the timeliness of intervention; early diagnosis and effective haemostatic treatment are critical [3, 4].

Upper GI endoscopy is the cornerstone of both diagnosis and therapy. Early endoscopy identifies the bleeding source in 90–95% of acute cases and enables definitive intervention through thermal, mechanical and injection techniques, reducing length of stay and re-bleeding [5–7].

The Glasgow–Blatchford Score (GBS) is the most widely used pre-endoscopic risk-stratification tool. It integrates hemoglobin, systolic blood pressure, level of consciousness, comorbidities, and overt bleeding signs to classify patients [8]. Although highly sensitive for identifying low-risk patients suitable for outpatient management, its specificity is low, resulting in many endoscopies being performed on patients without endoscopically actionable bleeding [9].

Machine-learning (ML) algorithms are increasingly applied to large clinical datasets to derive prediction models that outperform traditional scoring systems in selected emergency presentations [10, 11]. Their ability to capture non-linear interactions among admission variables makes them attractive candidates for refining the indication for emergent endoscopy in upper GI bleeding.

Because the specificity of existing scores and guidelines is suboptimal, a non-trivial fraction of endoscopies performed for suspected upper GI bleeding are negative, with attendant resource and workforce cost [9]. We hypothesised that supervised ML models trained on routinely available admission data from our ED population could achieve higher specificity than the GBS while maintaining a high negative predictive value, thereby reducing unnecessary endoscopies without missing active bleeders.

METHODS

Study Design and Setting

This was a retrospective, observational, cross-sectional, analytical study conducted in the ED of Kanuni Sultan Süleyman Training and Research Hospital, a tertiary academic centre in Istanbul, Turkey. The study commenced on 1 October 2022 after approval by the institutional ethics committee (University of Health Sciences, Kanuni Sultan Süleyman TRH; approval no. KAEK/2022.08.188). The study was conducted in accordance with the Declaration of Helsinki. No external funding was received, and the authors declare no conflicts of interest.

Participants

Inclusion criteria

adults aged ≥ 18 years who presented to the ED between 1 January 2015 and 1 August 2022 with a working diagnosis of upper GI bleeding and underwent upper GI endoscopy after evaluation by the relevant specialty (gastroenterology, general surgery), and whose data were complete in the hospital information system.

Exclusion criteria

age < 18 years; lower GI bleeding; and patients with incomplete or missing records.

Patient Flow

A total of 1120 candidate cases were identified. Ninety-three were excluded because their archival records could not be accessed and a further 18 were excluded for missing information, yielding 1009 eligible patients. As described below, the dataset was then randomised into a model-training set of 666 patients (≈ 70%) and an internal validation/test set of 343 patients (≈ 30%) (Fig. 1).

Data Collection

Admission vital signs, presenting complaints, comorbidities and digital rectal examination findings were recorded on a structured case report form (CRF) by emergency-medicine residents and attending physicians. Laboratory parameters captured were hemoglobin (HGB), platelet count (PLT), neutrophil percentage (NEUTP), albumin, blood urea nitrogen (BUN), international normalised ratio (INR), pH, lactate and base excess (BE), all measured in the ED laboratory and the central hospital laboratory under standard calibration. Endoscopy reports authored by gastroenterology or general-surgery specialists were transcribed into the CRF, and findings were graded using the Forrest classification. Forrest 1a and

1b were pre-specified as the “actively bleeding” group; all remaining categories were classified as “no active bleeding.” The GBS was calculated at admission from the recorded variables.

Units and reference ranges used for all laboratory and vital-sign variables are presented in Table 1. Personal identifiers were removed before analysis and data were transferred to a secure computerised system.

Table 1
Units and normal reference ranges of laboratory tests and vital signs recorded at admission.

Test/Vital	Unit	Normal Range
Systolic Blood Pressure	mmHg	120–139
Diastolic Blood Pressure	mmHg	80–89
Pulse	/min	60–100
Hemoglobin (HGB)	g/dL	13–17.5 (Male); 11.5–15.5 (Female)
Platelet (PLT)	$\times 10^3/\mu L$	100–400
Neutrophil percentage (NEUTP)	%	45–78
Albumin	g/L	39–49
Blood urea nitrogen (BUN)	mg/dL	6–21
INR	—	0.8–1.2
pH	—	7.35–7.45
Base excess (BE)	mmol/L	—
Lactate	mmol/L	0.5–2.0

HGB: hemoglobin; PLT: platelet; NEUTP: neutrophil percentage; BUN: blood urea nitrogen; INR: international normalised ratio; pH: potential of hydrogen; BE: base excess.

Sample Size

A priori power analysis was performed in reference to Wang et al. [12], assuming $\alpha = 0.05$, power $1 - \beta = 0.8$, an AUC difference of 0.026, and a one-tailed alternative hypothesis. The minimum required sample size was 1146 patients, with 343 in the validation set and 803 in the training set. Following that template, 70% of eligible patients (666) were randomly assigned to model training and 30% (343) to internal validation/testing using Python v3.9.0.

Machine-Learning Pipeline

All preprocessing, training and evaluation were performed in Python v3.9.0 using the Spyder IDE and Kaggle Notebook. Continuous predictors were standardised; categorical variables were one-hot encoded. Random allocation into training and test sets was carried out using stratified random sampling on outcome status. Five supervised classifiers were trained and compared: Gaussian Naive Bayes (GNB), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Random Forest, and Logistic Regression. Default hyperparameters of scikit-learn were used as the baseline; the threshold for binary classification was set at 0.5. Models were trained on the 666-patient training set and evaluated on the 343-patient held-out test set against the endoscopic reference standard. For each model, sensitivity, specificity, positive and negative likelihood ratios (PLR/NLR), positive and negative predictive values (PPV/NPV), and overall accuracy were computed.

Statistical Analysis

Categorical data are reported as counts and percentages. Normality of continuous variables was assessed using the Kolmogorov–Smirnov and Shapiro–Wilk tests together with skewness and kurtosis. Normally distributed variables are reported as mean ± standard deviation (SD); non-normally distributed variables are reported as median with interquartile range (IQR). Categorical comparisons used the Chi-square or Fisher's exact test, as appropriate; continuous comparisons used the independent-samples t-test or the Mann–Whitney U test. Model–reference-standard concordance was tested with Chi-square or Fisher's exact tests on 2×2 tables. Analyses were performed in IBM SPSS Statistics v24.0 (IBM Corp., Armonk, NY, USA), with a two-sided α of 0.05.

Study Termination Criterion

Data collection ended upon reaching the pre-specified target number of cases.

RESULTS

Study Population

Of the 1009 eligible patients, the 343 internally validated/test cases included 313 patients without active bleeding and 30 (8.7%) with endoscopically confirmed active bleeding (Forrest 1a/1b). Mean age was 56.0 ± 19.26 years in active bleeders versus 51.07 ± 20.18 years in non-bleeders (p = 0.202). Male sex predominated (218 patients, 63.6%); the sex distribution did not differ between groups (p = 0.244) (Table 2).

Table 2
Comparison of demographic, clinical and laboratory variables according to endoscopically confirmed active bleeding (Forrest 1a/1b) in the internal validation/test set (n = 343).

Variable	No Active Bleeding (n = 313)	Active Bleeding (n = 30)	p-value
Continuous variables — Mean ± SD (Median [IQR])			
Age (years)	56.0 ± 19.26 (61 [27.5])	51.07 ± 20.18 (46 [32.5])	0.202
Systolic BP (mmHg)	119.25 ± 20.34 (115 [25])	117.23 ± 21.20 (120 [15])	0.687
Diastolic BP (mmHg)	71.50 ± 12.07 (70 [20])	69.80 ± 13.20 (70 [12])	0.829
Pulse (beats/min)	88.10 ± 16.18 (85 [21])	91.40 ± 19.02 (87 [37.5])	0.494
Glasgow–Blatchford Score	12.19 ± 2.36 (13 [3])	13.07 ± 1.93 (13 [3])	**0.039***
Hemoglobin (g/dL)	9.69 ± 3.41 (8.5 [4.8])	9.61 ± 3.17 (10.2 [5.25])	0.919
Platelet (×10³/μL)	267.65 ± 118.78 (240 [145])	224.37 ± 95.58 (214 [99.5])	**0.049***
Neutrophil (%)	70.68 ± 12.89 (72.2 [19.35])	71.75 ± 11.14 (72.8 [16.65])	0.738
Albumin (g/L)	34.04 ± 6.44 (34.1 [8.7])	34.27 ± 5.85 (34.5 [8.85])	0.747
BUN (mg/dL)	28.92 ± 21.72 (25.23 [20.02])	37.81 ± 31.27 (25.23 [24.07])	0.071
INR	1.41 ± 1.30 (1.16 [0.31])	1.63 ± 1.11 (1.18 [0.57])	0.296
pH	7.37 ± 0.06 (7.37 [0.07])	7.38 ± 0.06 (7.37 [0.05])	0.708
Base excess (mmol/L)	−0.47 ± 4.15 (− 0.2 [4.6])	0.46 ± 4.42 (1.3 [5.1])	0.423
Lactate (mmol/L)	2.28 ± 1.60 (1.95 [1.48])	1.88 ± 0.92 (1.68 [1.06])	0.215
Categorical variables — n (%)			
Male sex	196 (62.6)	22 (73.3)	0.244

Data are mean ± SD (median [IQR]) for continuous variables or n (%) for categorical variables. *p < 0.05. GBS: Glasgow–Blatchford Score; BP: blood pressure; BUN: blood urea nitrogen; INR: international normalised ratio; BE: base excess.

Variable	No Active Bleeding (n = 313)	Active Bleeding (n = 30)	p-value
Continuous variables — Mean ± SD (Median [IQR])			
Digital rectal examination — Melena	151 (48.2)	21 (70.0)	0.123
Digital rectal examination — Hematochezia	13 (4.2)	0 (0.0)	0.123
Digital rectal examination — Normal stool	76 (24.3)	4 (13.3)	0.123
Digital rectal examination — Empty ampulla	73 (23.3)	5 (16.7)	0.123
Syncope	9 (2.9)	0 (0.0)	0.434
Liver failure	17 (5.4)	4 (13.3)	0.099
Congestive heart failure	54 (17.3)	4 (13.3)	0.584

Data are mean ± SD (median [IQR]) for continuous variables or n (%) for categorical variables. *p < 0.05. GBS: Glasgow–Blatchford Score; BP: blood pressure; BUN: blood urea nitrogen; INR: international normalised ratio; BE: base excess.

Clinical and Laboratory Features

Melena was the most common digital rectal examination finding (172/343, 50.1%), followed by normal stool (80/343, 23.3%), empty ampulla (78/343, 22.7%) and hematochezia (13/343, 3.8%); the distribution did not differ between bleeding groups (p = 0.123) (Table 2). Syncope was rare (n = 9, 2.9%; p = 0.434). Liver failure (5.4% vs. 13.3%; p = 0.099) and congestive heart failure (17.3% vs. 13.3%; p = 0.584) were not significantly more frequent in either group. Vital signs at admission did not differ between groups (systolic BP p = 0.687; diastolic BP p = 0.829; heart rate p = 0.494). Among laboratory parameters, only platelet count was significantly lower in the active-bleeding group (224.37 ± 95.58 vs. 267.65 ± 118.78 × 10^3/μL; p = 0.049). Hemoglobin, neutrophil percentage, albumin, BUN, INR, pH, BE and lactate did not differ significantly (all p > 0.05). The GBS was significantly higher in patients with active bleeding (13.07 ± 1.93 vs. 12.19 ± 2.36; p = 0.039) (Table 2).

Endoscopic Findings

Endoscopic findings were present in 153 of 343 patients (44.6%). The Forrest distribution was 4 (2.6%) Forrest 1a, 26 (17.0%) Forrest 1b, 13 (8.5%) Forrest 2a, 13 (8.5%) Forrest 2b, 8 (5.2%) Forrest 2c, and 89 (58.2%) Forrest 3; no bleeding or lesions were found in 190 patients (55.4%) (Table 3).

Table 3
Distribution of Forrest classification among patients with positive endoscopic findings (n = 153).

Forrest Classification	Number (n)	Percentage (%)
Forrest 1a	4	2.6
Forrest 1b	26	17.0
Forrest 2a	13	8.5
Forrest 2b	13	8.5
Forrest 2c	8	5.2
Forrest 3	89	58.2
Total	**153**	**100.0**

Machine-Learning Model Performance

Compared against the endoscopic reference standard in the held-out 343-patient test set, the GNB, SVM, Random Forest and Logistic Regression models all showed a statistically significant association with active bleeding ($p = 0.017$, $p < 0.001$, $p < 0.001$ and $p < 0.001$, respectively); the KNN model did not ($p = 0.241$) (Table 4).

Table 4
Cross-tabulation of machine-learning model predictions against the endoscopic reference standard for active bleeding (test set, n = 343).

Model	Model prediction	Active Bleeding (reference)		Total	p-value
		No	Yes		
GNB	**No**	231	16	247	**0.017***
	Yes	82	14	96	
KNN	**No**	311	29	340	0.241
	Yes	2	1	3	
SVM	**No**	217	11	228	**< 0.001***
	Yes	96	19	115	
Random Forest	**No**	195	6	201	**< 0.001***
	Yes	118	24	142	
Logistic Regression	**No**	210	9	219	**< 0.001***
	Yes	103	21	124	

*p < 0.05 (Chi-square / Fisher's exact). GNB: Gaussian Naive Bayes; KNN: K-Nearest Neighbors; SVM: Support Vector Machines.

Per-model operating characteristics were: GNB — sensitivity 46.67%, specificity 73.80%, PLR 1.78, NLR 0.72, PPV 14.67%, NPV 93.48%, accuracy 71.43%; KNN — sensitivity 3.33%, specificity 99.36%, PLR 5.22, NLR 0.97, PPV 33.48%, NPV 91.42%, accuracy 90.96%; SVM — sensitivity 63.33%, specificity 69.33%, PLR 2.07, NLR 0.53, PPV 16.61%, NPV 95.14%, accuracy 68.80%; Random Forest — sensitivity 80.00%, specificity 62.30%, PLR 2.12, NLR 0.32, PPV 17.00%, NPV 97.00%, accuracy 63.85%; Logistic Regression — sensitivity 70.00%, specificity 67.09%, PLR 2.13, NLR 0.45, PPV 17.03%, NPV 95.86%, accuracy 67.35% (Table 5).

Table 5
Operating characteristics of machine-learning models for detecting active bleeding (test set, n = 343).

Model	Sensitivity (%)	Specificity (%)	PLR	NLR	PPV (%)	NPV (%)	Accuracy (%)
GNB	46.67	73.80	1.78	0.72	14.67	93.48	71.43
KNN	3.33	99.36	5.22	0.97	33.48	91.42	90.96
SVM	63.33	69.33	2.07	0.53	16.61	95.14	68.80
Random Forest	80.00	62.30	2.12	0.32	17.00	97.00	63.85
Logistic Regression	70.00	67.09	2.13	0.45	17.03	95.86	67.35

PLR: positive likelihood ratio; NLR: negative likelihood ratio; PPV: positive predictive value; NPV: negative predictive value; GNB: Gaussian Naive Bayes; KNN: K-Nearest Neighbors; SVM: Support Vector Machines.

DISCUSSION

In this retrospective single-centre study of patients undergoing endoscopy for suspected upper GI bleeding, four of five supervised ML models — GNB, SVM, Random Forest and Logistic Regression — predicted endoscopically confirmed active bleeding with statistical significance, while KNN did not. The Random Forest model showed the highest sensitivity (80%) and NPV (97%), and SVM and Logistic Regression also achieved NPVs above 95%, suggesting that these algorithms can robustly identify patients without active bleeding from routinely available admission data.

The 16.33% rate of active bleeding among patients selected for endoscopy in our overall cohort is close to the 22.4% reported by Koruk et al. in 951 endoscoped patients [13]; the modest difference may reflect operator-dependent variability in endoscopy practice across centres.

Our GBS values (12.19 ± 2.36 in non-bleeders versus 13.07 ± 1.93 in active bleeders) are higher than the 10.19 ± 3.78 reported by Gökçek et al. in a mixed cohort comparing GBS and Rockall scores in 423 patients [14]. This likely reflects the fact that we included only patients who proceeded to endoscopy, enriching the cohort for moderate-to-high GBS scores.

The distribution of Forrest classes in our cohort (Forrest 3 most common at 58.2%, Forrest 1a 2.6%, 1b 17.0%, 2a 8.5%, 2b 8.5%, 2c 5.2%) is broadly consistent with Sağlık et al., who reported Forrest 3 in 47.8% of endoscoped patients [15]. Across the published literature Forrest 3 remains the most prevalent finding, in keeping with our results.

Demographically, our cohort mean age of 56.0 ± 19.26 years sits between the younger 39.64-year mean reported by Keçeci et al. [16] and the older 59.2-year mean reported by Yenigün et al. [17]; the 63.6% male predominance is comparable to the 55.1–74.5% reported elsewhere [16, 18]. Melena was the most

common digital rectal examination finding (50.1%), again consistent with prior series that have reported melena rates from 68.9% to 91.2% [17, 19].

Syncope occurred in only 2.9% of our patients versus 12.1% in the cohort of Gökçek et al. [14], a discrepancy plausibly explained by differences in patient reporting and timing of ED presentation. Mean admission vital signs and the prevalence of chronic heart and liver disease in our cohort were similar to published series [17, 20]. Laboratory values (hemoglobin, platelets, albumin, BUN, neutrophil percentage, INR, lactate) were in the ranges reported by Tomizawa et al., Zhuang et al., Shingina et al. and Gulen et al. [19, 21–23].

From a methodological perspective, the principal contribution of our study is the head-to-head comparison of five supervised ML algorithms on a uniformly processed dataset derived from a single ED. The high NPVs of the Random Forest, Logistic Regression and SVM models (97%, 95.86% and 95.14%, respectively) are particularly relevant in the ED setting, where the most clinically consequential question is whether endoscopy can be safely deferred. Given the 8.7% prevalence of active bleeding in our cohort, these high NPVs suggest that an ML-based triage tool could meaningfully reduce the burden of negative endoscopies while preserving safety. KNN, in contrast, showed unacceptably low sensitivity (3.33%) despite high specificity and accuracy, illustrating that overall accuracy is misleading in the presence of low outcome prevalence — a well-known limitation of nearest-neighbour approaches in imbalanced settings.

Limitations

This study has several limitations. First, it is a single-centre, retrospective analysis with a relatively limited number of outcome events (n = 30 active bleeders in the test set); multicentre prospective validation is necessary to generate robust performance estimates. Second, our dataset did not include detailed data on chronic medications (including antiplatelet and anticoagulant therapy), substance-use habits, or pre-ED treatments, which are known to influence both presentation and bleeding risk. Third, endoscopy was performed by multiple gastroenterology and general-surgery specialists, and the threshold for proceeding to endoscopy may have varied between ED attendings and consultants, introducing potential ascertainment and operator-related variability. Fourth, model calibration was not formally assessed; future work should incorporate calibration plots and net-benefit analyses. Fifth, the prevalence of active bleeding observed here (8.7%) is population-dependent and may not generalise to other settings.

CONCLUSIONS

Supervised machine-learning models trained on routinely available ED admission data — particularly Random Forest, Logistic Regression and SVM — predicted endoscopically confirmed active upper GI bleeding with high negative predictive value (≥ 95%) in our cohort. These results support the potential of ML-based decision-support tools to complement existing risk scores such as the GBS in identifying

patients in whom emergent endoscopy may safely be deferred. Hospitals will likely benefit from training such models on their own patient populations; prospective multicentre validation, formal calibration analysis, and assessment of clinical impact and workforce/cost outcomes are the necessary next steps before clinical implementation.

ABBREVIATIONS

AUC
Area under the curve
BE
Base excess
BUN
Blood urea nitrogen
CRF
Case report form
ED
Emergency department
GBS
Glasgow–Blatchford Score
GI
Gastrointestinal
GNB
Gaussian Naive Bayes
HGB
Hemoglobin
INR
International normalised ratio
IQR
Interquartile range
KNN
K-Nearest Neighbors
ML
Machine learning
NEUTP
Neutrophil percentage
NLR
Negative likelihood ratio
NPV
Negative predictive value
PLR

Positive likelihood ratio
PLT
Platelet
PPV
Positive predictive value
ROC
Receiver operating characteristic
SD
Standard deviation
SVM
Support Vector Machines.

DECLARATIONS

Ethics Approval:

Kanuni Sultan Süleyman Training and Research Hospital Ethics Committee, University of Health Sciences (Approval No: KAEK/2022.08.188).

Ethics Approval and Consent to Participate

The study was approved by the Ethics Committee of Kanuni Sultan Süleyman Training and Research Hospital, University of Health Sciences (approval number KAEK/2022.08.188). Because of the retrospective design and use of de-identified data, the requirement for individual written informed consent was waived by the committee. The study was conducted in accordance with the Declaration of Helsinki.

Consent for Publication

Not applicable.

Competing Interests

The authors declare that they have no competing interests.

Funding

No funding was received for the conduct of this study or for the preparation of this manuscript.

Author Contribution

MU conceptualised and designed the study, supervised data acquisition, performed the machine-learning analysis and drafted the manuscript. BY contributed to study design, data curation, statistical analysis and manuscript revision. SD and UMK supervised the methodology and critically revised the manuscript for important intellectual content. BSKP contributed to clinical-data interpretation and manuscript revision. AE and STF contributed to data collection and interpretation of endoscopic data. SF contributed to data acquisition and statistical analysis. RY contributed to study coordination and manuscript revision. All authors read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Availability of Data and Materials

The de-identified dataset analysed during the current study is available from the corresponding author (MU, ucan.melih@gmail.com) on reasonable request, subject to institutional approval and a data-use agreement.

REFERENCES

1. Tham J, Stanley A. Clinical utility of pre-endoscopy risk scores in upper gastrointestinal bleeding. Expert Rev Gastroenterol Hepatol. 2019;13(12). 10.1080/17474124.2019.1698292.
2. Hearnshaw SA, Logan RFA, Lowe D, Travis SPL, Murphy MF, Palmer KR. Acute upper gastrointestinal bleeding in the UK: patient characteristics, diagnoses and outcomes in the 2007 UK audit. Gut. 2011;60(10). 10.1136/gut.2010.228437.
3. Bhala N, Lanas A. Low-dose acetylsalicylic acid and the use of gastroprotectors among older adults. Geriatr Aging. 2008;11(1).
4. McPherson C. Vitamin K deficiency bleeding: an ounce of prevention. Neonatal Netw. 2020;39(6). 10.1891/0730-0832/11-T-630.
5. Samuel R, Bilal M, Tayyem O, Guturu P. Evaluation and management of non-variceal upper gastrointestinal bleeding. Dis Mon. 2018;64(7). 10.1016/j.disamonth.2018.02.003.
6. Mujtaba S, Chawla S, Massaad JF. Diagnosis and management of non-variceal gastrointestinal hemorrhage: a review of current guidelines and future perspectives. J Clin Med. 2020;9(2). 10.3390/jcm9020402.
7. Laine L, Jensen DM. Management of patients with ulcer bleeding. Am J Gastroenterol. 2012;107(3). 10.1038/ajg.2011.480.
8. Tielleman T, Bujanda D, Cryer B. Epidemiology and risk factors for upper gastrointestinal bleeding. Gastrointest Endosc Clin N Am. 2015;25(3). 10.1016/j.giec.2015.02.010.
9. Stanley AJ, Laine L, Dalton HR, et al. Comparison of risk scoring systems for patients presenting with upper gastrointestinal bleeding: international multicentre prospective study. BMJ. 2017;356. 10.1136/bmj.i6432.

10. Shouval R, Bondi O, Mishan H, Shimoni A, Unger R, Nagler A. Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT. Bone Marrow Transpl. 2014;49(3). 10.1038/bmt.2013.146.
11. Rao VL, Gupta N, Swei E, et al. Predictors of mortality and endoscopic intervention in patients with upper gastrointestinal bleeding in the intensive care unit. Gastroenterol Rep (Oxf). 2020;8(4). 10.1093/gastro/goaa009.
12. Wang H, Zhao QY, Luo JC, et al. Early prediction of noninvasive ventilation failure after extubation: development and validation of a machine-learning model. BMC Pulm Med. 2022;22(1). 10.1186/s12890-022-02096-7.
13. Koruk İ, Çam H, Aydınlı M, et al. Üst gastrointestinal sistem kanaması ön tanısı ile refere edilen hastalarda endoskopik aktif kanama oranları. Endoskopi Gastrointest. 2015;21(1). 10.17940/endoskopi.74813.
14. Gökçek K, Ersel M, Altuncı YA, Karbek Akarca F, Kıyan S. Retrospective analyses of the utility of Glasgow–Blatchford and Rockall and pre-Rockall scoring systems in patients admitted to the emergency department with upper gastrointestinal system bleeding. Forbes J Med. 2022;3(3). 10.4274/forbes.galenos.2022.36450.
15. Sağlık A, Özüçelik DN, Kalafat UM, et al. Evaluation of the relationship between D-dimer levels and endoscopic findings of patients with upper gastrointestinal hemorrhage. Ann Clin Anal Med. 2021;12(Suppl 1):86–90.
16. Keçeci M, Sevinç C, Kayataş K, et al. Varis dışı akut üst gastrointestinal sistem kanamalı hastalarda lökositoz ile kanama ciddiyeti arasındaki ilişki. Akademik Gastroenteroloji Dergisi. 2014;13(1):7–12.
17. Coşkun Yenigün E, Pirpir A, Aytan P, Ulusal G, Yıldırım İS. Üst gastrointestinal sistem kanamalı hastaların özelliklerinin değerlendirilmesi. Akademik Gastroenteroloji Dergisi. 2006;5(2):116–22.
18. Shrestha MP, Borgstrom M, Trowers E. Digital rectal examination reduces hospital admissions, endoscopies and medical therapy in patients with acute gastrointestinal bleeding. Am J Med. 2017;130(7). 10.1016/j.amjmed.2017.01.036.
19. Shingina A, Barkun AN, Razzaghi A, Martel M, Bardou M, Gralnek I. Systematic review: the presenting international normalised ratio (INR) as a predictor of outcome in patients with upper non-variceal gastrointestinal bleeding. Aliment Pharmacol Ther. 2011;33(9). 10.1111/j.1365-2036.2011.04618.x.
20. Sağıroğlu MF, Çalapkulu M, Gülten M. Gastroenteroloji kliniğine başvuran akut üst gastrointestinal sistem kanamalı hastaların retrospektif değerlendirilmesi. Uludağ Üniv Tıp. Fak Derg. 2020;46(3). 10.32708/uutfd.747285.
21. Gulen M, Satar S, Tas A, Avci A, Nazik H, Toptas Firat B. Lactate level predicts mortality in patients with upper gastrointestinal bleeding. Gastroenterol Res Pract. 2019;2019. 10.1155/2019/5048078
22. Tomizawa M, Shinozaki F, Hasegawa R, et al. Low hemoglobin levels are associated with upper gastrointestinal bleeding. Biomed Rep. 2016;5(3). 10.3892/br.2016.727.
23. Zhuang Y, Xia S, Chen J, et al. Construction of a prediction model for rebleeding in patients with acute upper gastrointestinal bleeding. Eur J Med Res. 2023;28(1). 10.1186/s40001-023-01349-3.

Figures

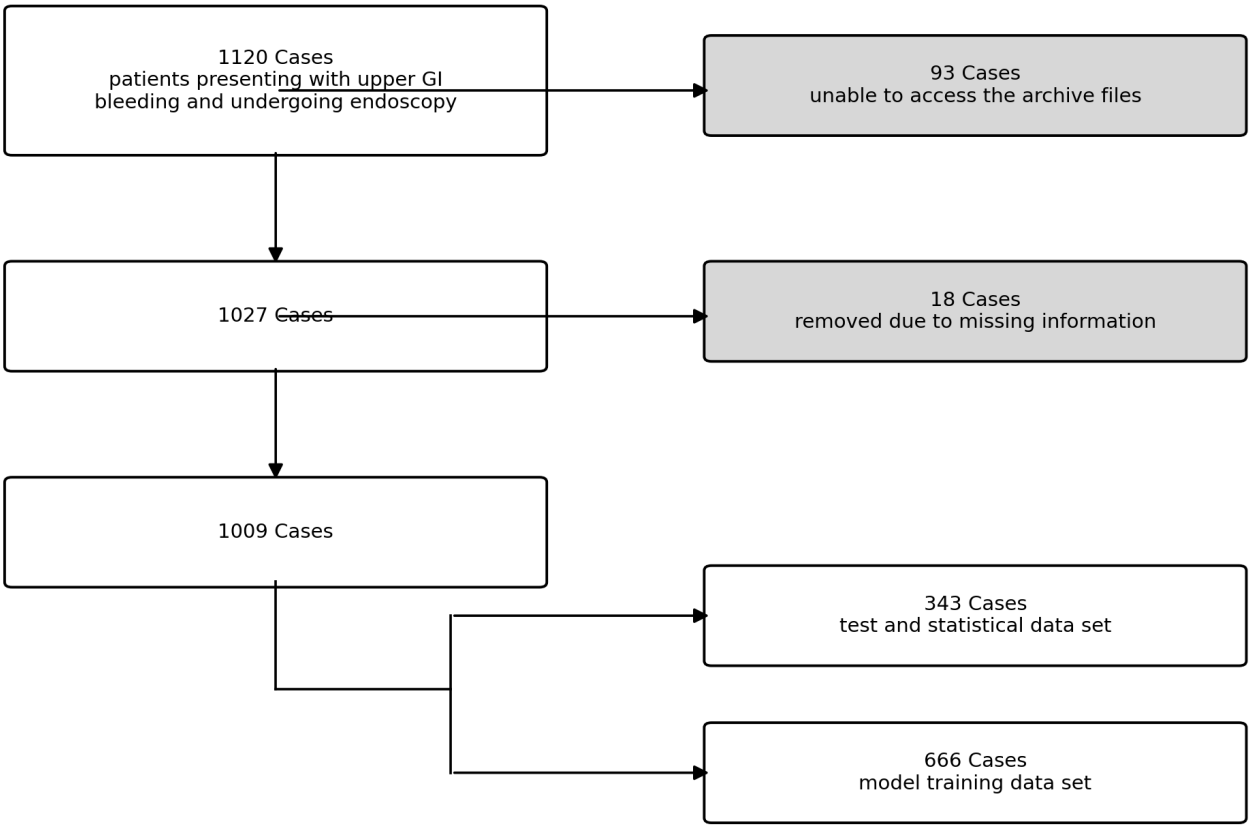


Figure 1

Patient flow diagram. From 1120 candidate cases, 93 were excluded because their archival records could not be accessed and 18 were excluded for missing information, yielding 1009 eligible patients. Of these, 666 patients (≈70%) were randomly allocated to the model-training data set and 343 patients (≈30%) to the test/statistical data set.