American Journal of Medicine and Medical Sciences 2024, 14(5): 1229-1237
DOI: 10.5923/j.ajmms.20241405.19

Identifying Key Factors Impacting Hospitalization Outcomes in Variceal Upper Gastrointestinal Bleeding: A Machine Learning Approach

Ismati Amir Olimovich[1], Mamarajabov Sobirjon Ergashevich[1], Anosov Victor Davidovich[2]

[1]Samarkand State Medical University, Samarkand, Republic of Uzbekistan
[2]City Clinical Hospital No. 15 named after O.M. Filatov, Moscow, Russian Federation

Abstract **Background:** Managing patients with variceal bleeding from upper gastrointestinal (GI) tract is always a challenging task not only due to the various causes leading to this condition, but also because of the multitude of factors capable of impacting hospitalization outcomes. **Objective:** The aim of this study was to investigate the degree of influence of factors on 30-day hospitalization outcomes using machine learning (ML) methods. **Methods:** A retrospective dataset was collected from 105 patients and included clinical, anamnestic, laboratory, instrumental data. Subsequently, the entire database was divided into two samples with differing levels of data completeness. The obtained samples were processed using ML tools in two stages: imputation with usage model of multiple imputation by chained equations (MICE), and factor importance analysis using tuned random forest models. The primary outcome was death or successful discharge. **Results:** There were not only well-known predictors of mortality found among the most prognostically valuable indicators, but also factors that hold promise for the role of predictor in scientific community. The top-10 most prognostically significant factors were found to be: ferritin, blood urea level, arterial pressure, procalcitonin, creatinine, lactate, amylase, activated partial thromboplastin time (APTT), white blood cell count, aspartate aminotransferase (AST). **Conclusion:** Usage of advanced methods confirmed the significance of already known and validated predictors of mortality, contributed not only to the development of newly proposed predictors by scientific community in recent times, but also to those yet unexplored.

Keywords Variceal bleeding, Prognostically important, Outcome, Mortality, Predictor

1. Introduction

Upper gastrointestinal bleeding (UGIB) is a multifactorial acute pathology, still representing one of the common reasons of emergency admissions [1,2]. The highest number of fatal cases among patients with bleedings from upper gastrointestinal (GI) tract is traditionally encountered in group with variceal hemorrhage [3,4]. Mortality associated with this condition depends on a multitude of factors and increases with age and number of comorbidities [5]. In addition, there are specific factors that assist physicians in stratifying patients into risk groups. The most reliable of these factors are often mentioned as predictors in scientific literature [6].

There is a plethora of well-known predictors of mortality, frequently referenced in clinical guidelines [7,8,9]. The most effective predictors in terms of prognosis are already incorporated into simple scoring systems for predicting clinical outcomes [10,11,12,13,14,15].

Amidst the rapid advancement of artificial intelligence (AI) technologies, machine learning (ML) methods have become commonplace in the physician's toolkit. Over the past several years, there has been a growing body of researches utilizing ML techniques [16,17]. This fact indicates that scientists are increasingly favoring advanced statistical analysis methods. Implementation of predictive models emerges the possibility of obtaining valuable insights into the factors influencing clinical outcomes based on trained data.

Despite the existing number of well-known predictors of mortality, there remains a pertinent need for new researches aimed at potential identifying new predictors and re-evaluating the significance of already known ones. In this study, we focus our attention on analyzing a broad spectrum of patient data using advanced analytical methods based on ML. The aims of this research were to investigate factors in group of patients with variceal UGIB, identifying predictors of 30-day mortality among them, as well as to reassess already known predictors of 30-day mortality in order to use in clinical practice with high efficiency and assist physicians in more precise risk stratification with subsequent reducing mortality.

Received: Mar. 31, 2024; Accepted: Apr. 29, 2024; Published: May 11, 2024
Published online at http://journal.sapub.org/ajmms

2. Materials and Methods

2.1. Study Design

A single-center retrospective study was conducted at City Clinical Hospital No. 15, named after O.M. Filatov, affiliated with the Department of Healthcare of the City of Moscow, Russian Federation. The study database included patients with UGIB within a timeframe from 2020 to 2023.

The inclusion criteria for patients in the study were as follows: age ≥18 years, diagnosed variceal UGIB. The exclusion criteria were as follows: absence of performed endoscopic examination, refusal to sign provided informed consent for inclusion into the study, discharge from the hospital at the patient's request.

2.2. Data Collection

The study involved 105 patients with variceal UGIB. Patient information was gathered from the specialized electronic medical record system for the period from 2020 to 2023. All parameters in the database included 213 factors: anamnestic, clinical, laboratory, endoscopic, and some instrumental data.

Anamnestic data included information about: patient's gender, age, height, weight, known harmful habits, list of comorbidities, some heart procedures, outpatient medication use, initial manifestations of GIB and the duration since their onset, site of detection (pre-hospital or during hospitalization period), possible previous episodes of UGIB according to medical documentation or patient's report.

Laboratory data was collected on the first, third, and fifth days from the moment of confirmation of UGIB and included parameters from complete blood count (CBC), biochemical analysis (kidney and liver function tests, albumin, total protein, glucose, amylase), ferritin, lactate dehydrogenase (LDH), C-reactive protein (CRP), procalcitonin, D-dimer, venous or arterial blood pH, electrolyte levels (potassium, sodium, chloride), lactate level, arterial blood gas levels, coagulation profile (international normalised ratio, activated partial thromboplastin time, fibrinogen, prothrombin time).

Endoscopic data included information not only about the priority source of UGIB (variceal), but also about secondary less life-threatening sources (ulcers, erosions, neoplasms, angiodysplasias) in case of their presence: quantity, localization, depth of defect, dimensions when possible to measure, Forrest classification for ulcer bleedings, methods of performed endoscopic hemostasis (injection, electrocoagulation, argon plasma coagulation, clipping, band ligation), as well as calculation of the area of major bleeding defects.

Instrumental data included: results of computed tomography (CT) of lung parenchyma, echocardiography with determination of left ventricular ejection fraction (LVEF).

Clinical data included information on patient's condition during the first, third, and fifth days from the instrumental confirmation of UGIB, specifically: heart rate and rhythm, systolic and diastolic blood pressure, information about vasopressor support of hemodynamics, patient's level of consciousness, body temperature, information on 24-hour urine volume, presence of abdominal pain with clarification of its localization, respiratory rate during spontaneous breathing or mechanical ventilation of lungs. Additionally, the medications taken by the patient were recorded and classified into groups: anticoagulants, antiplatelet agents (including acetylsalicylic acid), proton pump inhibitors (PPI), antihypertensive drugs, antibiotics, analgesics, corticosteroids, non-steroidal anti-inflammatory drugs (excluding acetylsalicylic acid) (NSAID), biological therapy (tocilizumab, levilimab, olokizumab) in patients with confirmed COVID-19. Information on gastroprotective therapy and conservative hemostatic measures (hemostatic agents, Sengstaken–Blakemore tube) after the detection of UGIB was also registered. Hemotransfusion therapy (red blood cell, platelet and plasma transfusions) was recorded in milliliters. Clinical data also included information on recurrence of bleeding, information about surgery performed in case of failure of endoscopic and endovascular hemostasis and possible complications after surgical intervention, length of hospital stay (LOS), some indices and scoring systems calculated basing on collected information: body mass index, Charlson Comorbidity Index (CCI) in scores and percentage expression, average arterial pressure based on measured systolic and diastolic values of blood pressure, categorization of minimally registered level of consciousness during the first, third, and fifth days into three severity groups, namely 0-9 points according to the Glasgow Coma Scale (GCS), 10-12 points, 13-15 points, calculation of shock index, calculation of ASA score, total number of comorbidities. Clinical outcomes were formed in a binary format: survival and mortality.

2.3. Statistical Analysis

One of the weaknesses of this study lies in the fact that filling in over 200 clinical parameters for each patient appears challenging, particularly in the setting of emergency hospital. Due to this issue, associated with frequent inability to fill all factors in the database, a decision was made to resort to selecting a portion of patients from the most complete section of the database, followed by imputation of missing values using advanced analytical algorithms based on ML.

At the first stage of statistical analysis 48 patients with variceal bleeding (mean proportion of missing values - 20.33%) out of 105. Selected group was identified as final sample (Table 1).

Table 1. Main subgroups and final samples

Patients		Proportion of missing values		
Main subgroup, *n*	**Final sample, *n***	**Minimal (%)**	**Maximal (%)**	**Mean (%)**
105	48	9.48	26.72	20.33

At the second stage of analysis imputation (handling missing values in the database) was implemented. One of the most accurate methods of handling missing values to date is the statistical method based on ML called multiple imputation by chained equations (MICE). This multiple

imputation method allows to train a predictive model based on available data for further use with high accuracy. Thus, the ML model consists of two main steps: "training" or saturating the model with existing data, and "testing" or in simple words filling in missing values. In order to train the imputation model effectively considering possible interrelationships between factors, a decision was made to reunite final sample with main subgroup. Pre-processing of data included removing columns (parameters) with single unique value and encoding the target variable with clinical outcome values. Algorithm of MICE was constructed and executed by using Python programming language with specialized libraries such as scikit-learn, pandas and numpy. It involved dividing the data into categorical and numerical variables followed by their analysis in relation to each other. Numerical variables in MICE algorithm were processed using an iterative cyclic function "IterativeImputer" executed by a random forest regressor (RandomForestRegressor) with 50 iterations and a random value generator of 31. Categorical variables were processed using the same iterative function, but executed by a random forest classifier (RandomForest Classifier) with a random value generator also set to 31, a cross-validation parameter set to 5, in a specialized grid search mode (GridSearchCV) to select the best hyperparameters for the ML model out of entered options ("n_estimators": 100, 250, 500, 1000; "max_depth": 5, 10, 20, None; "min_samples_split": 2, 5, 10; "min_samples_leaf": 1, 2, 4; "bootstrap": True, False).

Upon completion of stage related to imputation an analysis of main subgroup and final sample was conducted using statistical analysis method based on ML, namely, ensemble algorithm "random forest" (Figure 1). The random forest analysis was performed in grid search mode (GridSearchCV) to automatically select the best hyperparameters out of presented: for hyperparameter "n_estimators" - options of 100, 500, 1000, 3000 trees; for "max_depth" - options for limiting the depth of trees' growth to 1, 2, 5, 10 and "None"; for "min_samples_split" - options of 2, 5, 10, 20; for "min_samples_leaf" - options of 1, 2, 5, 10; for "max_features" – options "sqrt" and "log2"; for "bootstrap" – "True" and "False"; for hyperparameter "criterion" - options "gini" and "entropy". The cross-validation parameter was set to 10 for main subgroup. Cross-validation parameter for final sample was set into "RepeatedStratifiedKFold" mode with parameter "n_splits" equal to 10, "n_repeats" - 3, and a random value generator of 31. The main subgroup of patients was randomly divided into training and testing sets with the "test_size" parameter set to 0.2, corresponding to 80% training and 20% testing data. In case of final sample, stratification based on target outcome was used, and the "test_size" parameter was set to optimal for small samples value of 0.3. The scoring parameter was always set to "accuracy".

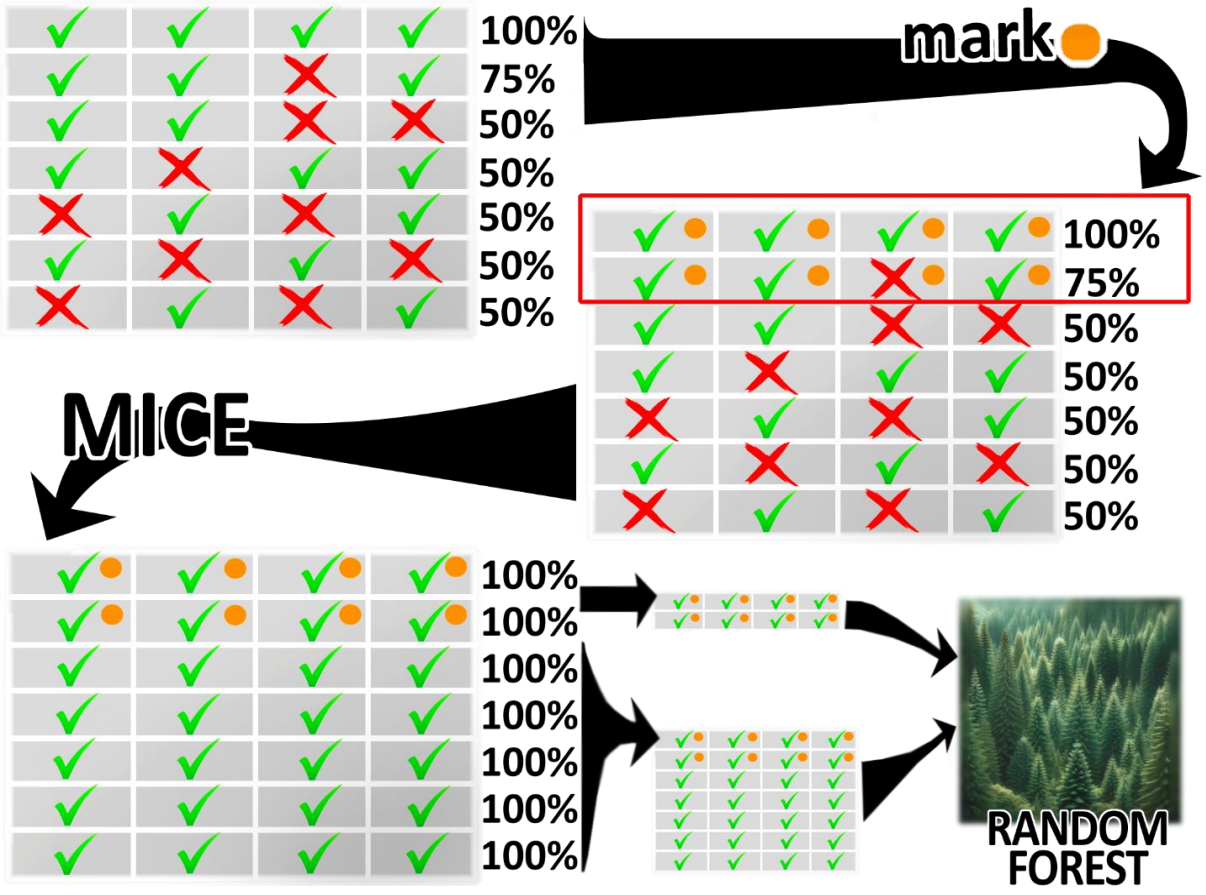

Figure 1. Schematic overview of implemented statistical analysis

Table 2. List of formed categories and their components forming sum of results for own categories

Category name	Category components forming the sum of results
Age	Only age
Albumin	Albumin results for the first, third, fifth days
Alanine aminotransferase (ALT)	ALT results for the first, third, fifth days
Amylase	Blood amylase results for the first, third, fifth days
Activated partial thromboplastin time (APTT)	APTT results for the first, third, fifth days
Arrhythmia	Results for arrhythmia on the first, third, fifth days
ASA	Only ASA score
Aspartate aminotransferase (AST)	AST results for the first, third, fifth days
Bilirubin	Blood bilirubin results for the first, third, fifth days
Sengstaken–Blakemore tube	Only results concerning placing Sengstaken-Blackmore tube
Arterial pressure	Results of data related to arterial blood pressure: systolic, diastolic, average arterial pressure, fact of vasopressor support
State of consciousness	Results of all data related to the patients' level of consciousness: GCS on the first, third, fifth days
Central nervous system diseases	Presence or absence of a history of stroke, transient ischemic attack (TIA), signs of dementia
Respiratory rate	Results for respiratory rate on the first, third, fifth days
Arterial carbon dioxide partial pressure	Results on the level of partial pressure of carbon dioxide in arterial blood for the first, third, fifth days
Charlson comorbidity index (CCI)	CCI results in percentage and score expressions
Chloride	Results for blood chloride levels for the first, third, fifth days
Comorbidities	Results regarding comorbidities, namely: myocardial infarction (MI) in medical history, arterial hypertension, type 2 diabetes mellitus, chronic respiratory diseases, oncological pathologies with presence or absence of metastases, also leukemia and lymphoma, congestive heart failure (CHF), cirrhosis with or without signs of portal hypertension, autoimmune rheumatic diseases, chronic kidney diseases, COVID-19, peripheral vascular diseases, pressure (decubitus) ulcers, presence of injuries at the time of hospitalization, history of peptic ulcer disease (PUD)
Control Benchmark	Results of a control benchmark initially introduced into patient databases as a factor column consisting of random, non-repeating numbers for each patient as part of an experimental approach to separate potentially significant from potentially insignificant results after random forest analysis
Creatinine	Creatinine results for the first, third, fifth days
CRP	CRP results for the first, third, fifth days
Computed tomography (CT) of the lungs	Only results CT scans of lung parenchyma with assessing severity of pneumonia in patients with COVID-19
Length of stay (LOS)	Results regarding duration of patient in-hospital treatment (up to and including 30 days)
D-dimer	D-dimer results for the first, third, fifth days
Diuresis	Results regarding daily diuresis for the first, third and fifth days
Medications	Results regarding medications taken both on an outpatient and inpatient basis: analgesics, antibiotics, antihypertensive drugs, anticoagulants, antiplatelet agents (including acetylsalicylic acid), glucocorticoids, NSAIDs (except acetylsalicylic acid), hemostatics, gastroprotective agents, biological therapy medications
Endoscopic haemostasis	Results regarding various types of endoscopic hemostasis: injection, electrocoagulation, argon plasma coagulation, clipping, band ligation
Ferritin	Results for blood ferritin levels for the first, third, fifth days
Fibrinogen	Results on blood fibrinogen levels for the first, third, fifth days
Left ventricular ejection fraction (LVEF)	LVEF results only
Blood glucose	Blood glucose results for the first, third, fifth days
Bad habits	Results regarding bad habits, namely: smoking, consuming alcohol
Hemoglobin	Results for hemoglobin for the first, third, fifth days
Hemoglobin level's change	Results regarding the difference in hemoglobin levels between three observations during hospitalization: on the first, third, and fifth days
Bicarbonate (HCO3)	Results for blood HCO3 levels for the first, third, fifth days
Hematocrit	Results for hematocrit for the first, third, fifth days
International normalised ratio (INR)	INR results for the first, third, fifth days

Category name	Category components forming the sum of results
Lactate	Results for lactate levels for the first, third, fifth days
LDH	LDH level results for the first, third, fifth days
White blood cell count	Results on white blood cell levels for the first, third, fifth days
Onset of UGIB	Results regarding the onset of the disease: duration, outpatient or inpatient occurrence of UGIB
Surgical treatment	Results on surgical interventions in cases where minimally invasive hemostasis was not feasible, and postoperative complications if happened
Arterial oxygen partial pressure	Results on the level of partial pressure of oxygen in arterial blood for the first, third, fifth days
Oxygen saturation of blood	Results regarding the level of oxygen saturation on the first, third, fifth days
Abdominal pain	Results regarding the presence and localization of abdominal pain during the first, third and fifth days
pH of arterial blood	Results on the pH level of arterial blood for the first, third, fifth days
pH of venous blood	Results on the pH level of venous blood for the first, third, fifth days
Platelet count	Results on platelet count level for the first, third, fifth days
Potassium level in blood	Results on blood potassium levels for the first, third, fifth days
Procalcitonin	Results on procalcitonin levels for the first, third, fifth days
Total protein	Results on total protein levels for the first, third, fifth days
Prothrombin time (PT)	Results on PT for the first, third, fifth days
Heart rate	Heart rate results for the first, third, fifth days
Rebleedings	Results on rebleedings after endoscopic, endovascular, surgical hemostasis
Endoscopic data	Results of endoscopic findings
Gender	Only gender
Shock data	Results for the presence or absence of shock, as well as the value of the Allgower index
Sodium level in blood	Results on blood sodium levels for the first, third, fifth days
Symptoms of UGIB	Macroscopic signs of GIB at onset
Body temperature	Results on body temperature for the first, third, fifth days
Transfusions	Results of transfusion therapy: fact and volume of red blood cell transfusions, plasma transfusions and platelet transfusions
Blood urea level	Results for blood urea levels for the first, third, fifth days
Mechanical ventilation	Results regarding the fact of conducted mechanical ventilation during the first, third, and fifth days
Body mass index (BMI)	Results regarding growth and weight parameters, including BMI

Upon completion of the random forest analysis, the results of feature-parameter significance, commonly denoted in scientific literature as importance of features, were grouped based on their similarity with a further sum calculated for each such group to facilitate further review and analysis. For example, endoscopic features were categorized into one group, creatinine results for the first, third, and fifth days were grouped together into second category, the fact of any hemotransfusion was grouped into next category (Table 2). Weighted averages were calculated for results of main subgroup and final sample after getting sum for each category of features. To facilitate comprehension, the obtained weighted mean values were transformed into relative values by normalization with a sum equal to 100. Final results are presented as histogram with values sorted by magnitude (Figure 2).

3. Results

3.1. Study Population

There were more women registered in group than men. One in every six patients had a documented history of UGIB. Liver cirrhosis was present in 81% of the group, diabetes mellitus and CHF were encountered in one-fifth of the patients, hypertension was observed in 47% of the patients, and one-third of the group had alcohol abuse issues.

BMI, body mass index; UGIB, upper gastrointestinal bleeding; PUD, peptic ulcer disease; MI, myocardial infarction; HF, heart failure; NSAID, non-steroidal anti-inflammatory drugs; CRP, C-reactive protein; INR, international normalised ratio; ASA, American Society of Anesthesiologists physical status classification system; LOS, length of stay.

The hemoglobin level often fell within the range of moderate anemia, while renal indicators were frequently elevated. Up to three-quarters of the group required various forms of blood transfusion therapy. Initial manifestations of UGIB were most commonly observed in the form of hematemesis, rebleeding rates reached up to 8%, and mortality was at 41%.

Table 3. Baseline characteristics of study groups

Demographics:	
Age, years, *median* (IQR)	54 (47-62)
Male, *n* (%)	45 (42.9)
Female, *n* (%)	60 (57.1)
BMI, *median* (IQR)	26.4 (23.4-29.9)
Time since onset of symptoms, hours, *median* (IQR)	4.5 (2-24)
Significant comorbid conditions:	
Previous UGIB, *n* (%)	17 (16.2)
PUD, *n* (%)	9 (8.6)
COVID-19, *n* (%)	9 (8.6)
Previous MI, *n* (%)	7 (6.7)
History of HF, *n* (%)	22 (20.9)
Previous heart procedures, *n* (%)	5 (4.8)
Arterial hypertension, *n* (%)	47 (44.8)
History of any peripheral artery diseases, *n* (%)	1 (1)
History of cerebrovascular accident, *n* (%)	7 (6.7)
Chronic respiratory disease, *n* (%)	6 (5.7)
Liver cirrhosis, *n* (%)	85 (81)
Diabetes mellitus, *n* (%)	21 (20)
Renal insufficiency, *n* (%)	6 (5.7)
History of cancer, *n* (%)	16 (15.2)
Current smoker, *n* (%)	11 (10.5)
Alcohol Consumption, *n* (%)	34 (32.4)
Medications intake:	
Antiplatelet agents (including aspirin), *n* (%)	3 (2.9)
Anticoagulants, *n* (%)	3 (2.9)
Glucocorticosteroids, *n* (%)	1 (1)
NSAIDs (except for acetylsalicylic acid), *n* (%)	2 (1.9)
Transfusions:	
Red cell transfusions, *n* (%)	79 (75.2)
Plasma transfusions, *n* (%)	64 (61)
Platelet transfusions, *n* (%)	14 (13.3)
Manifestation of symptoms:	
Melena, *n* (%)	20 (19)
Hematemesis, *n* (%)	34 (32.4)
Hematemesis + Melena, *n* (%)	28 (26.7)
Basic laboratory parameters:	
Hemoglobin, g/L, *median* (IQR)	77 (63-94)
Platelet count, 10^9/L, *median* (IQR)	159 (110-203)
Albumin, g/L, *median* (IQR)	26.6 (21.5-29.7)
Blood urea, mmol/L, *median* (IQR)	11.1 (7.4-17.5)
Creatinine, μmol/L, *median* (IQR)	99.6 (73.9-135.9)
CRP, mg/L, *median* (IQR)	18.5 (5.5-55.3)
INR, *median* (IQR)	1.8 (1.5-2.4)
ASA score, *median* (IQR)	4 (4-5)
LOS, days, *median* (IQR)	6 (3-8)
Rebleeding rate, *n* (%)	8 (7.6)
30-day mortality, *n* (%)	43 (41)

3.2. Predictors of Outcome

All the metrics obtained during validation stage on test groups, representing 20% of main subgroup and 30% of final sample, indicate that predictive model is to some extent more proficient in identifying factors influencing survival rather than mortality, likely due to the overwhelming majority of surviving patients in all groups, whose data were absorbed by random forest predictive model during its training phase (Table 4). Recall for mortality was low, however, recall for survival in both main subgroup and final sample, as well as precision for mortality in main subgroup showed high results.

The results of predictor significances were summed up across main subgroup and final sample yielding weighted average values, and the control benchmark parameter among them served as a threshold separating potentially important predictors from potentially insignificant indicators in histogram obtained from the results of random forest analysis (Figure 2).

Such well-known for their reliability predictors as creatinine level, albumin and INR were identified among the most significant factors (Figure 2). However, patients in these groups were most sensitive to changes in ferritin levels previously underestimated among patients with variceal hemorrhage.

The features, significance of which according to results of random forest analysis was at the lower end of the ranking, still played some role, however, they did not strongly influence clinical outcomes (Figure 2).

Table 4. Obtained metrics of random forest analysis according to implemented test stage

Group	Accuracy	Outcomes and weighted average	Precision	Recall	F1-score	Support
Main subgroup	0.62	Survival	0.53	0.89	0.67	9
		Mortality	0.83	0.42	0.56	12
		Weighted AVG	0.70	0.62	0.60	
Final sample	0.67	Survival	0.67	0.89	0.76	9
		Mortality	0.67	0.33	0.44	6
		Weighted AVG	0.67	0.67	0.63	

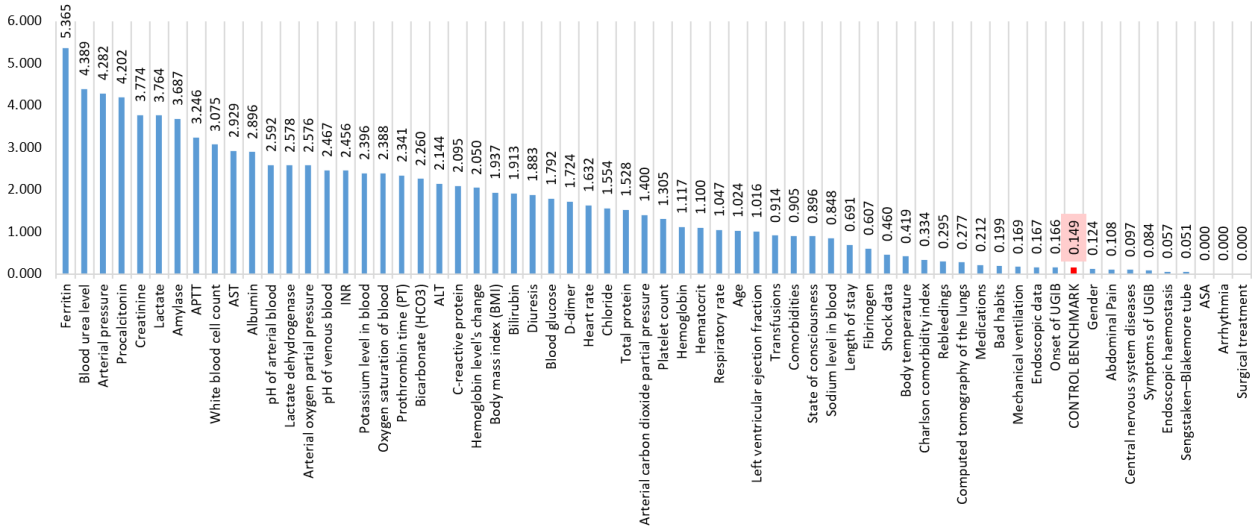

Figure 2. Significance of factors according to random forest analysis of factors associated with variceal bleedings

4. Discussion

A comprehensive approach based on usage of ML tools was utilized within this study. Patient database containing all relevant information was preprocessed using MICE algorithms and subsequently subjected to processing using random forest models (Figure 1). It is assumed that predictors exceeding the control benchmark parameter in histogram has greater prognostic influence, than those which are below control parameter. However, factors with relatively low significance values are not useless but require further analysis to identify conditions allowing predictors to reliably correlate with hospitalization outcomes.

The majority of results regarding the significance of factors did not raise doubts, as they logically fit into the traditional clinical picture of the pathological condition. However, there were outliers, such as the peak indicator of ferritin's significance in patients with variceal bleeding, number of researches on which is still insufficient [18,19]. It appears that depletion of iron reserves may serve as a rather reliable predictor of mortality among patients with variceal UGIB (Figure 2).

Urea level of blood and creatinine are renowned for their predictive significance and have been utilized in prognostic scoring systems for quite some time: the latest notable study was the development of ABC scoring system, which includes creatinine levels in the algorithm for calculating mortality prognosis [13,15]. The analysis conducted in this study also confirms high efficacy of these predictors in forecasting clinical outcomes.

Arterial pressure, identified in this research as one of the strongest influencing factors on clinical outcomes, has long been a reliable predictor and is still utilized in well-known prognostic scoring systems used in cases of UGIB such as the Rockall Score (RS), Glasgow-Blatchford Score (GBS), AIMS65, Cedars-Sinai Medical Center Predictive Index (CSMCPI), Progetto Nazionale Emorragia Digestive Score (PNED) [10,11,12,13,14].

Procalcitonin has previously already been discussed among scientists in context of increased frequency of variceal bleedings among patients with cirrhosis [20], and studies have emerged associating higher levels of procalcitonin to increased frequency of GIB [21]. Sepsis is a pathological condition with high mortality, and this study can join to research pool confirming the prognostic value of procalcitonin in patients with variceal UGIB: procalcitonin's significance was among especially noticeable ones in the results.

According to the correlation potential of LDH with clinical outcomes in patients with UGIB, compelling scientific literature has not yet been accumulated. However, in this study, unexpected elevated levels were found not only for LDH but also for participant of the chemical reaction in which LDH takes part. This reaction participant is called lactate. Blood lactate has been noted in scientific literature not only as a useful predictor in patients with UGIB but also as a cost-effective option in clinical practice applicable in forecasting various scenarios: mortality, recurrent bleeding, the need for transfer to the intensive care unit, and the need for transfusion therapy [6,22,23].

The level of white blood cells has previously been repeatedly noted as an independent predictor of mortality in patients with UGIB [24,25]. In this study, we also confirmed the high significance of this predictor in managing patients with UGIB of variceal etiology.

Factors such as blood amylase, APTT and AST, according to the obtained results, were also among ten the most important prognostic indicators. They have been noted in studies defining their significance in predicting the occurrence of variceal UGIB in at-risk patient groups [26,27]. However, they were not identified in compelling scientific researches determining the value of these predictors in predicting 30-day mortality.

Until recently, the ability to predict clinical outcomes in patients with UGIB has been limited to a small number of factors upon which subsequent treatment strategies were structured. However, the rapid proliferation of AI tools has

transformed the exploitation of its capabilities into commonplace practice. ML, which is a branch of AI, enables the utilization of advanced analytical methods and the extraction of valuable datasets as output. The analysis conducted in this scientific work, utilizing various ML models, not only allows for a reassessment of acknowledged predictors but also enables a deeper exploration of previously unknown factors requiring validation.

However, our study has several limitations. Firstly, the work is single-center and retrospective. Secondly, patient data collection was conducted at an emergency center, which made it difficult to completely fill out more than 200 parameters for each patient, leading to the adoption of MICE method. Thirdly, small patient samples with an unbalanced ratio in terms of target indicator (hospitalization outcome: survival or mortality) were utilized, which could to some extent affect the inability of random forest predictive model to adequately perform testing phase of analysis on patients with lethal outcomes, potentially leading to the lack of confident manifestation of certain predictors of lethality in results of final samples. Thus, some factors that appeared insignificant in this study may demonstrate statistical significance in larger scientific works. Fourthly, the random forest method is highly accurate but complex to interpret regarding its analysis results, necessitating further research into the identified predictors. Lastly, the ML methods used are among the most reliable modern imputation and statistical analysis techniques, however, achieving the highest accuracy requires substantial computational power—supercomputers—which is currently a tool difficult to access.

5. Conclusions

Through ML not only a revision of already recommended predictors was performed, but also the identification of factors with high prognostic significance, previously overlooked. It was revealed based on advanced analysis methods utilizing ML that top-10 the most important factors were: ferritin, blood urea level, arterial blood pressure, procalcitonin, creatinine, lactate, amylase, APTT, white blood cell count, AST.

However, conducted research may be insufficient for practical application of identified predictors in stratifying patients due to limitations of research. Therefore, we recommend multicenter studies with a larger number of participating patients to effectively identify those at high risk of adverse clinical outcomes.

REFERENCES

[1] Kamboj AK, Hoversten P, Leggett CL. Upper Gastrointestinal Bleeding: Etiologies and Management. Mayo Clin Proc. 2019 Apr; 94(4): 697-703. doi: 10.1016/j.mayocp.2019.01.022. PMID: 30947833.

[2] Antunes C, Copelin II EL. Upper Gastrointestinal Bleeding. 2023 Apr 7. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan–. PMID: 29262121.

[3] Chason, Rebecca BA1; Singal, Amit MD, MS1; Rockey, Don MD2. Mortality in Acute Upper Gastrointestinal Bleeding Is Uncommonly due to Persistent Hemorrhage: 1634. American Journal of Gastroenterology 108(): p S490-S491, October 2013.

[4] Radadiya D, Devani K, Rockey DC. The impact of red blood cell transfusion practices on inpatient mortality in variceal and non-variceal gastrointestinal bleeding patients: a 20-year US nationwide retrospective analysis. Aliment Pharmacol Ther. 2022 Jul; 56(1): 41-55. doi: 10.1111/apt.16965. Epub 2022 May 19. PMID: 35591774; PMCID: PMC10829766.

[5] Elsebaey MA, Elashry H, Elbedewy TA, Elhadidy AA, Esheba NE, Ezat S, Negm MS, Abo-Amer YE, Abgeegy ME, Elsergany HF, Mansour L, Abd-Elsalam S. Predictors of in-hospital mortality in a cohort of elderly Egyptian patients with acute upper gastrointestinal bleeding. Medicine (Baltimore). 2018 Apr; 97(16): e0403. doi: 10.1097/MD.0000000000010403. PMID: 29668596; PMCID: PMC5916675.

[6] Zeng F, Du L, Ling L. Lactate level as a predictor of outcomes in patients with acute upper gastrointestinal bleeding: A systematic review and meta‑analysis. Exp Ther Med. 2024 Jan 24; 27(3): 113. doi: 10.3892/etm.2024.12401. PMID: 38361514; PMCID: PMC10867736.

[7] Laine L, Barkun AN, Saltzman JR, Martel M, Leontiadis GI. ACG Clinical Guideline: Upper Gastrointestinal and Ulcer Bleeding. Am J Gastroenterol. 2021 May 1; 116(5): 899-917. doi: 10.14309/ajg.0000000000001245. Erratum in: Am J Gastroenterol. 2021 Nov 1; 116(11): 2309. PMID: 33929377.

[8] Tripathi D, Stanley AJ, Hayes PC, et alUK guidelines on the management of variceal haemorrhage in cirrhotic patientsGut 2015; 64: 1680-1704.

[9] Gralnek IM, Camus Duboc M, Garcia-Pagan JC, Fuccio L, Karstensen JG, Hucl T, Jovanovic I, Awadie H, Hernandez-Gea V, Tantau M, Ebigbo A, Ibrahim M, Vlachogiannakos J, Burgmans MC, Rosasco R, Triantafyllou K. Endoscopic diagnosis and management of esophagogastric variceal hemorrhage: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. Endoscopy. 2022 Nov; 54(11): 1094-1120. doi: 10.1055/a-1939-4887. Epub 2022 Sep 29. PMID: 36174643.

[10] Saltzman JR, Tabak YP, Hyett BH, Sun X, Travis AC, Johannes RS. A simple risk score accurately predicts in-hospital mortality, length of stay, and cost in acute upper GI bleeding. Gastrointest Endosc. 2011 Dec; 74(6): 1215-24. doi: 10.1016/j.gie.2011.06.024. Epub 2011 Sep 10. PMID: 21907980.

[11] Rockall TA, Logan RF, Devlin HB, Northfield TC. Risk assessment after acute upper gastrointestinal haemorrhage. Gut. 1996 Mar; 38(3): 316-21. doi: 10.1136/gut.38.3.316. PMID: 8675081; PMCID: PMC1383057.

[12] Hay JA, Lyubashevsky E, Elashoff J, Maldonado L, Weingarten SR, Ellrodt AG. Upper gastrointestinal hemorrhage clinical--guideline determining the optimal hospital length of stay. Am J Med. 1996 Mar; 100(3): 313-22. doi: 10.1016/s0002-9343(97)89490-9. PMID: 8629677.

[13] Blatchford O, Murray WR, Blatchford M. A risk score to predict need for treatment for upper-gastrointestinal haemorrhage. Lancet. 2000 Oct 14; 356(9238): 1318-21. doi: 10.1016/S0140-6736(00)02816-6. PMID: 11073021.

[14] Marmo R, Koch M, Cipolletta L, Capurso L, Pera A, Bianco

MA, Rocca R, Dezi A, Fasoli R, Brunati S, Lorenzini I, Germani U, Di Matteo G, Giorgio P, Imperiali G, Minoli G, Barberani F, Boschetto S, Martorano M, Gatto G, Amuso M, Pastorelli A, Torre ES, Triossi O, Buzzi A, Cestari R, Della Casa D, Proietti M, Tanzilli A, Aragona G, Giangregorio F, Allegretta L, Tronci S, Michetti P, Romagnoli P, Nucci A, Rogai F, Piubello W, Tebaldi M, Bonfante F, Casadei A, Cortini C, Chiozzini G, Girardi L, Leoci C, Bagnalasta G, Segato S, Chianese G, Salvagnini M, Rotondano G. Predictive factors of mortality from nonvariceal upper gastrointestinal hemorrhage: a multicenter study. Am J Gastroenterol. 2008 Jul; 103(7): 1639-47; quiz 1648. doi: 10.1111/j.1572-0241.2008.01865.x. PMID: 18564127.

[15] Laursen SB, Oakland K, Laine L, Bieber V, Marmo R, Redondo-Cerezo E, Dalton HR, Ngu J, Schultz M, Soncini M, Gralnek I, Jairath V, Murray IA, Stanley AJ. ABC score: a new risk score that accurately predicts mortality in acute upper and lower gastrointestinal bleeding: an international multicentre study. Gut. 2021 Apr; 70(4): 707-716. doi: 10.1136 /gutjnl-2019-320002. Epub 2020 Jul 28. PMID: 32723845.

[16] Ungureanu BS, Gheonea DI, Florescu DN, Iordache S, Cazacu SM, Iovanescu VF, Rogoveanu I, Turcu-Stiolica A. Predicting mortality in patients with nonvariceal upper gastrointestinal bleeding using machine-learning. Front Med (Lausanne). 2023 Feb 17; 10: 1134835. doi: 10.3389/fmed. 2023.1134835. PMID: 36873879; PMCID: PMC9982090.

[17] Udriştoiu AL, Cazacu IM, Gruionu LG, Gruionu G, Iacob AV, Burtea DE, Ungureanu BS, Costache MI, Constantin A, Popescu CF, Udriştoiu Ş, Săftoiu A. Real-time computer-aided diagnosis of focal pancreatic masses from endoscopic ultrasound imaging based on a hybrid convolutional and long short-term memory neural network model. PLoS One. 2021 Jun 28; 16(6): e0251701. doi: 10.1371/journal.pone.0251701. PMID: 34181680; PMCID: PMC8238220.

[18] Abbas OM, Khalifa KAE, Makhlouf MM, Osman NF, Abdel Razek WM, Atta AS. Influence of esophageal variceal bleeding on iron status in chronic hepatitis C patients. Eur J Gastroenterol Hepatol. 2020 May; 32(5): 616-622. doi: 10.1097/MEG.0000000000001547. PMID: 31567713.

[19] Oikonomou T, Goulis I, Soulaidopoulos S, Karasmani A, Doumtsis P, Tsioni K, Mandala E, Akriviadis E, Cholongitas E. High serum ferritin is associated with worse outcome of patients with decompensated cirrhosis. Ann Gastroenterol. 2017; 30(2): 217-224. doi: 10.20524/aog.2016.0112. Epub 2016 Dec 8. PMID: 28243043; PMCID: PMC5320035.

[20] Zidan, M.H.S., Zaghloul, S.G., Seleem, W.M. et al. Bacteremia as a risk factor for variceal upper gastrointestinal tract bleeding in cirrhotic patients: a hospital-based study. Egypt Liver Journal 11, 8 (2021). https://doi.org/10.1186/s43066 -021-00078-8.

[21] Yang QY, Ouyang J, Yang JD. Sepsis as an important risk factor for gastrointestinal bleeding in acute coronary syndrome patients: Two case reports. Medicine (Baltimore). 2018 Sep; 97(36): e12273. doi: 10.1097/MD.0000000000012273. PMID: 30200168; PMCID: PMC6133616.

[22] 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 Oct 24; 2019: 5048078. doi: 10.1155/2019/5048078. PMID: 31781189; PMCID: PMC6855015.

[23] Strzałka M, Winiarski M, Dembiński M, Pędziwiatr M, Matyja A, Kukla M. Predictive Role of Admission Venous Lactate Level in Patients with Upper Gastrointestinal Bleeding: A Prospective Observational Study. J Clin Med. 2022 Jan 11; 11(2): 335. doi: 10.3390/jcm11020335. PMID: 35054029; PMCID: PMC8780414.

[24] He L, Zhang J, Zhang S. Risk factors of in-hospital mortality among patients with upper gastrointestinal bleeding and acute myocardial infarction. Saudi J Gastroenterol. 2018 May-Jun; 24(3): 177-182. doi: 10.4103/sjg.SJG_492_17. PMID: 29652028; PMCID: PMC5985637.

[25] Moledina SM, Komba E. Risk factors for mortality among patients admitted with upper gastrointestinal bleeding at a tertiary hospital: a prospective cohort study. BMC Gastroenterol. 2017 Dec 20; 17(1): 165. doi: 10.1186/s12876-017-0712-8. PMID: 29262794; PMCID: PMC5738843.

[26] Rockey DC, Elliott A, Lyles T. Prediction of esophageal varices and variceal hemorrhage in patients with acute upper gastrointestinal bleeding. J Investig Med. 2016 Mar; 64(3): 745-51. doi: 10.1136/jim-2015-000047. Epub 2016 Feb 12. PMID: 26912006.

[27] Civan JM, Lindenmeyer CC, Whitsett M, Herrine SK. A Clinical Decision Rule Based on the AST-to-Platelet Ratio Index Improves Adherence to Published Guidelines on the Management of Acute Variceal Bleeding. J Clin Gastroenterol. 2015 Aug; 49(7): 599-606. doi: 10.1097/MCG.00000000000 00173. PMID: 26167719.