

# A machine learning approach for identification of gastrointestinal predictors for the risk of COVID-19 related hospitalization

Peter Lipták<sup>1</sup>, Peter Banovcin<sup>1</sup>, Róbert Rosolanka<sup>2</sup>, Michal Prokopič<sup>1</sup>, Ivan Kocan<sup>3</sup>, Ivana Žiačiková<sup>3</sup>, Peter Uhrík<sup>1</sup>, Marian Grendar<sup>4,5</sup> and Rudolf Hyrdel<sup>1</sup>

<sup>1</sup>Gastroenterology Clinic, University Hospital in Martin, Jessenius Faculty of Medicine in Martin, Comenius University in Bratislava, Martin, Slovak Republic

<sup>2</sup>Clinic of Infectology and Travel Medicine, University Hospital in Martin, Jessenius Faculty of Medicine in Martin, Comenius University in Bratislava, Martin, Slovak Republic

<sup>3</sup>Clinic of Pneumology and Phthysiology, University Hospital in Martin, Jessenius Faculty of Medicine in Martin, Comenius University in Bratislava, Martin, Slovak Republic

<sup>4</sup>Laboratory of Bioinformatics and Biostatistics, Biomedical Centre Martin, Jessenius Faculty of Medicine, Comenius University in Bratislava, Martin, Slovak Republic

<sup>5</sup>Laboratory of Theoretical Methods, Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia

## ABSTRACT

**Background and aim:** COVID-19 can be presented with various gastrointestinal symptoms. Shortly after the pandemic outbreak, several machine learning algorithms were implemented to assess new diagnostic and therapeutic methods for this disease. The aim of this study is to assess gastrointestinal and liver-related predictive factors for SARS-CoV-2 associated risk of hospitalization.

**Methods:** Data collection was based on a questionnaire from the COVID-19 outpatient test center and from the emergency department at the University Hospital in combination with the data from internal hospital information system and from a mobile application used for telemedicine follow-up of patients. For statistical analysis SARS-CoV-2 negative patients were considered as controls in three different SARS-CoV-2 positive patient groups (divided based on severity of the disease). The data were visualized and analyzed in R version 4.0.5. The Chi-squared or Fisher test was applied to test the null hypothesis of independence between the factors followed, where appropriate, by the multiple comparisons with the Benjamini Hochberg adjustment. The null hypothesis of the equality of the population medians of a continuous variable was tested by the Kruskal Wallis test, followed by the Dunn multiple comparisons test. In order to assess predictive power of the gastrointestinal parameters and other measured variables for predicting an outcome of the patient group the Random Forest machine learning algorithm was trained on the data. The predictive ability was quantified by the ROC curve, constructed from the Out-of-Bag data. Matthews correlation coefficient was used as a one-number summary of the quality of binary classification. The importance of the predictors was measured using the Variable Importance. A 2D representation of the data was obtained by means of

Submitted 1 September 2021

Accepted 24 February 2022

Published 21 March 2022

Corresponding author

Peter Lipták, peter.liptak@uniba.sk

Academic editor

Aslı Suner

Additional Information and  
Declarations can be found on  
page 15

DOI [10.7717/peerj.13124](https://doi.org/10.7717/peerj.13124)

© Copyright

2022 Lipták et al.

Distributed under

Creative Commons CC-BY 4.0

**OPEN ACCESS**

Principal Component Analysis for mixed type of data. Findings with the  $p$ -value below 0.05 were considered statistically significant.

**Results:** A total of 710 patients were enrolled in the study. The presence of diarrhea and nausea was significantly higher in the emergency department group than in the COVID-19 outpatient test center. Among liver enzymes only aspartate transaminase (AST) has been significantly elevated in the hospitalized group compared to patients discharged home. Based on the Random Forest algorithm, AST has been identified as the most important predictor followed by age or diabetes mellitus. Diarrhea and bloating have also predictive importance, although much lower than AST.

**Conclusion:** SARS-CoV-2 positivity is connected with isolated AST elevation and the level is linked with the severity of the disease. Furthermore, using the machine learning Random Forest algorithm, we have identified the elevated AST as the most important predictor for COVID-19 related hospitalizations.

**Subjects** Gastroenterology and Hepatology, Infectious Diseases, Data Mining and Machine Learning, COVID-19

**Keywords** COVID-19, SARS-CoV-2, Machine learning, Artificial intelligence, Random forest, Symptoms, Liver, Predictors, Hospitalization

## INTRODUCTION

Acute SARS-CoV-2 infection presents with variable symptoms associated with various organ systems. Typical symptoms of COVID-19 are fever, cough, and in the case of a more severe course of the disease, dyspnea with respiratory insufficiency occurs ([Guan et al., 2020](#)). In addition, COVID-19 may be presented with gastrointestinal symptoms, which include dominantly nausea, vomiting, diarrhea, anorexia and abdominal pain with relatively wide range of prevalence among different published studies ([Aziz et al., 2020](#); [Mao et al., 2020](#); [Sultan et al., 2020](#); [Patel et al., 2020](#); [D'Amico et al., 2020](#); [Jin et al., 2020](#); [Xiao et al., 2020](#)). Since COVID-19 pandemic is the cause of an immense world health crisis, new diagnostic and therapeutic methods are rapidly emerging ([Alimadadi et al., 2020](#)). The use of artificial intelligence is just one of them. Shortly after the COVID-19 outbreak, various machine learning algorithms have been implemented ([Randhawa et al., 2020](#); [Yan et al., 2020](#); [Ge et al., 2021](#); [Li et al., 2020](#)). Machine learning helps quickly identify patterns and trends of the large volume of data, that are difficult for humans to recognize ([Kushwaha et al., 2020](#)). The availability of objective stratification tools for the rapid assessment of a patient status and prognosis is of great use for the frontline health-care providers ([Bachtiger, Peters & Walsh, 2020](#)).

The primary aim of this study is to assess the possible predictive factors for SARS-CoV-2 outcome based on gastrointestinal symptoms and liver related laboratory results using machine learning algorithms of the Random Forest ([Guan et al., 2020](#); [Breiman, 2001](#)). The secondary aim is to determinate the prevalence of gastrointestinal symptoms among patients with COVID-19 within different groups based on the severity of the disease.

## MATERIALS AND METHODS

The study was performed from February through May 2021. Only subjects aged 18 years or older were included in the study. All patients enrolled in this study had signed the informed consent.

This study was approved by the Ethical Committee of the University Hospital in Martin, decision number: 14/2021.

Two distinct kinds of population were considered for this study. First group consists of patients who underwent nasopharyngeal swab in the outpatient hospital testing center for COVID-19 in order to determine whether they were SARS-CoV-2 positive. The method used for SARS-CoV-2 detection from nasopharyngeal swab was PCR (polymerase chain reaction). This group was then subdivided based on their positivity. The negative group was thereafter used as a control group for this study.

Second group consists of patients who attended COVID-19 emergency department (ED) in the University Hospital. These patients were confirmed positive from nasopharyngeal swab either by PCR or antigen method. Only patients with typical COVID-19 symptoms (fever, cough, dyspnoe) were included in this study. Patients who were SARS-CoV-2 positive but, didn't present with typical COVID-19 symptoms (*e.g.*, patients who came to emergency room because of other diagnoses, but simultaneously were SARS-CoV-2 positive) were excluded. Therefore, we considered for this study only patients who were both tested positive and had at least one typical COVID-19 symptom.

The second group was then divided based on further evaluation and course of the disease. First subgroup consists of patients that didn't require admission to the hospital and were referred to the outpatient care. Second subgroup of patients was admitted to the hospital. Consequently, this group was observed until the end of hospitalization either because of their death or resolution of the disease. This subgroup was also divided for analysis purposes to patients who required medical care in standard hospital ward and those who needed intensive care unit (ICU).

Data was collected by using a questionnaire in the group from COVID-19 outpatient test center at the University Hospital. Data from emergency room was obtained with the same questionnaire which was combined with information from medical examination by an attending physician and from the mobile application MEDAsistent used for telemedicine follow-up developed at the Clinic of Pneumology and Phthisiology in the University Hospital in Martin. Further information (including laboratory tests results, chest X-ray etc.) about patients who were hospitalized has been obtained from hospital information system.

The questionnaire consists of questions related to the present health complaints typical for COVID-19 and the spectrum of most common gastrointestinal symptoms which occurred in the last 5–7 days before examination. Patients were also allowed to write down other presented symptoms in the case they were not in the original list. In order to include only new or worsened gastrointestinal symptoms in the study the questionnaire also contained questions about chronic gastrointestinal symptoms and their possible worsening in the last 5–7 days before examination.

## Data analysis

The data was visualized and analyzed in R (*R Development Core Team, 2021*), version 4.0.5, with the aid of the libraries *gtsummary* (*Sjoberg et al., 2020*), *rstatix* (*Kassambara, 2021*), *DescTools* (*Signorell, 2021*), *randomForestSRC* (*Ishwaran & Kogalur, 2021*), *PCAmixdata* (*Chavent et al., 2017*) and *ggpubr* (*Kassambara, 2020*). The sample median and the lower and upper quartiles were used to summarize the data on continuous variables (e.g., age); counts and percentages were used to summarize factors (e.g., gender). The Chi-squared or Fisher test were applied to test the null hypothesis of independence between factors (gender vs group; fever vs group; cough vs group; diarrhea vs group; constipation vs group; bloating vs group; nausea vs group; heartburn vs group; abdominal pain vs group), followed, where appropriate, by multiple comparisons with the Benjamini Hochberg adjustment. Using a contingency table, an absence of trend was tested by Cochran Armitage test. The null hypothesis of the equality of the population medians of the continuous variable: age, Oxygen (O<sub>2</sub>) saturation, C-reactive protein (CRP), gamma glutamyltransferase (GGT), aspartate aminotransferase (AST), Bilirubin) was tested by the Kruskal Wallis test, followed by the Dunn multiple comparisons test with the Benjamini Hochberg correction of p-values. Two-way ANOVA was used to model the association between AST and group (discharged home, admitted to hospital) in interaction with the recent ATB use. Another two-way ANOVA model was utilized to quantify the association between AST and group (discharged home, admitted to hospital) in interaction with chronic liver disease (yes, no). The AST values were log-transformed to bring data to normality. Normality of residuals was assessed by the quantile-quantile plot with the 95% confidence band constructed by bootstrap. Assumption of homogeneity of variance was tested by the Levene test. In order to assess the predictive power of the gastrointestinal parameters and other measured variables (Gender, Age, No of Days of Symptoms, AST, alanine aminotransferase /ALT/, Bilirubin, Recent antibiotics /ATB/ Usage, Diabetes Mellitus, Arterial Hypertension, Chronic Liver Disease, Fever, Cough, Diarrhea, Constipation, Bloating, Nausea, Heartburn, Abdominal Pain) for predicting the outcome of the patient group the Random Forest machine learning algorithm was trained on the data. The predictive ability was quantified by the ROC curve, constructed from the Out-of-Bag data. The Matthews correlation coefficient was used as a one-number summary of the quality of binary classification. Importance of the predictors was measured by the Variable Importance. A 2D representation of the data (the predictors used in Random Forest; i.e., Gender, Age, Number of Days of Symptoms, AST, ALT, Bilirubin, Recent ATB Usage, Diabetes Mellitus, Arterial Hypertension, Chronic Liver Disease, Fever, Cough, Diarrhea, Constipation, Bloating, Nausea, Heartburn, Abdominal Pain) was obtained by Principal Component Analysis for a mixed type of data. Findings with the p-value below 0.05 were considered statistically significant.

## RESULTS

A total of 710 patients were enrolled in the study. Thirty (30) patients were excluded from the further analysis after primary screening. Participants ( $n = 352$ ) from the outpatient center who were tested PCR negative for SARS-CoV-2 virus were considered as the control

group. SARS-CoV-2 positive group from outpatient center included 166 participants. One hundred and sixty-two ( $n = 162$ ) patients from emergency department were enrolled. From this group 78 patient (48%) were discharged home, 57 (35.3%) admitted to the hospital for standard care until discharged from hospital. Twenty-seven (27) (16.7%) patients required intensive care unit. Based on age, the groups from outpatient center had almost similar median of 42 and 41 years of age respectively. Hospitalized patients were significantly older as shown in the [Table 1](#). The presence of typical COVID-19 symptoms such as fever and cough were significantly higher in the hospitalized groups as opposed to outpatient participants. There were no significant differences between groups in the men to women ratio.

### **Gastrointestinal symptoms occurrence and laboratory findings (Tables 1 and 2)**

The presence of diarrhea, constipation, bloating, nausea, heartburn and abdominal pain was considered in this study. Presence of diarrhea and nausea was significantly higher in SARS-CoV-2 positive patients than in SARS-CoV-2 negative controls. Comparing SARS-Cov-2 negative and SARS-CoV-2 positive participants the cumulative presence of diarrhea is 21.3% (70/328) in the positive group (combined outpatient center and emergency department) vs 6.2% (22/352) in the negative group and for nausea it is 13.1% (43/328) in the positive group vs 3.4% (12/352) in the negative group. This trend goes further considering ED patients and the severity of disease.

Among gastrointestinal symptoms, diarrhea and bloating were significantly more often manifested in patients who were admitted to the hospital compared to those discharged home (40% for diarrhea and 14% for bloating vs 18% and 2.6% respectively). Other symptoms such as abdominal pain, heart burn, nausea, vomitus, anorexia, and constipation were not presented differently in these groups in the meaning of statistical significance. C-reactive protein was also significantly higher in hospitalized group. In case of alanin transaminase (ALT), aspartate transaminase (AST) and bilirubin as markers of possible liver damage only AST ([Fig. 1](#)) was significantly higher in the hospitalized group. This difference is substantial. There is no statistically significant difference in the levels of ALT ([Fig. 1](#)) and Bilirubin when comparing different groups of patients.

### **Predictors of hospitalization based on machine learning**

Based on the Random Forest algorithm with the data on demographic characteristics, symptoms and gastrointestinal related laboratory findings in hospitalized and discharged patients, several predictors for risk of hospitalization were identified. AST was pinpointed as the most important predictor followed by age and diabetes mellitus. Diarrhea and bloating have also positive importance, although much lower than AST. Gastrointestinal symptoms such as nausea, abdominal pain or anorexia have none or negative predictive importance. The ROC curve for combined factors is shown in the [Fig. 2](#) with AUC 0.76. The Matthews correlation coefficient was 0.48.

When using only liver enzymes (AST, ALT), gastrointestinal symptoms (diarrhea and bloating), chronic liver disease, age and diabetes mellitus, the ROC curve ([Fig. 3](#)) for this

**Table 1** Patients characteristics.

	Outpatient test center (SARS-CoV-2 negative) ( <i>n</i> = 352) <sup>1</sup>	Outpatient test center (SARS-CoV-2 positive) ( <i>n</i> = 166) <sup>1</sup>	Discharged home ( <i>n</i> = 78) <sup>1</sup>	Admitted to hospital-standard care ( <i>n</i> = 57) <sup>1</sup>	Admitted to hospital-ICU ( <i>n</i> = 27) <sup>1</sup>	<i>p</i> value <sup>2</sup>
Age	42 (31, 51)	41 (32, 52)	48 (42, 58)	55 (46, 67)	65 (55, 72)	<0.001
Gender						0.07
female	206 (59%)	104 (63%)	41 (53%)	27 (47%)	11 (41%)	
male	141 (40%)	62 (37%)	37 (47%)	30 (53%)	16 (59%)	
NA	5 (1%)					
Fever						<0.001
Not-presented	340 (97%)	146 (88%)	48 (62%)	25 (44%)	15 (56%)	
Presented	12 (3.4%)	20 (12%)	30 (38%)	32 (56%)	12 (44%)	
Cough						<0.001
Not-presented	342 (97%)	144 (87%)	48 (62%)	24 (42%)	14 (52%)	
Presented	10 (2.8%)	22 (13%)	30 (38%)	33 (58%)	13 (48%)	
Diarrhea						<0.001
Not-presented	330 (94%)	141 (85%)	64 (82%)	34 (60%)	19 (70%)	
Presented	22 (6.2%)	25 (15%)	14 (18%)	23 (40%)	8 (30%)	
Constipation						0.37
Not-presented	347 (99%)	164 (99%)	76 (97%)	55 (96%)	26 (96%)	
Presented	5 (1.4%)	2 (1.2%)	2 (2.6%)	2 (3.5%)	1 (3.7%)	
Bloating						0.13
Not-presented	341 (97%)	161 (97%)	76 (97%)	51 (89%)	26 (96%)	
Presented	11 (3.1%)	5 (3.0%)	2 (2.6%)	6 (11%)	1 (3.7%)	
Nausea						<0.001
Not-presented	340 (97%)	147 (89%)	70 (90%)	45 (79%)	23 (85%)	
Presented	12 (3.4%)	19 (11%)	8 (10%)	12 (21%)	4 (15%)	
Heart burn						0.86
Not-presented	341 (97%)	163 (98%)	76 (97%)	56 (98%)	26 (96%)	
Presented	11 (3.1%)	3 (1.8%)	2 (2.6%)	1 (1.8%)	1 (3.7%)	
Abdominal pain						0.07
Not-presented	334 (95%)	150 (90%)	71 (91%)	49 (86%)	26 (96%)	
Presented	18 (5.1%)	16 (9.6%)	7 (9.0%)	8 (14%)	1 (3.7%)	

**Notes:**<sup>1</sup> Statistics presented: *n* (%); Median (IQR).<sup>2</sup> Statistical tests performed: Kruskal-Wallis test; Fisher's Exact Test for Count Data with simulated *p*-value (based on 2,000 replicates).

combination of factors attained AUC 0.799 with AST as the strongest predictor for hospitalization (Table 3). The Matthews correlation coefficient was 0.37.

### Principal components visualization of data

Principal component analysis was used to get a two-dimensional visualization of the data, for patients discharged home after ED examination and patients admitted to hospital. Data used for the analysis consist of the data from Table 2, that means a combination of general patient characteristics, typical COVID-19 symptoms and gastrointestinal

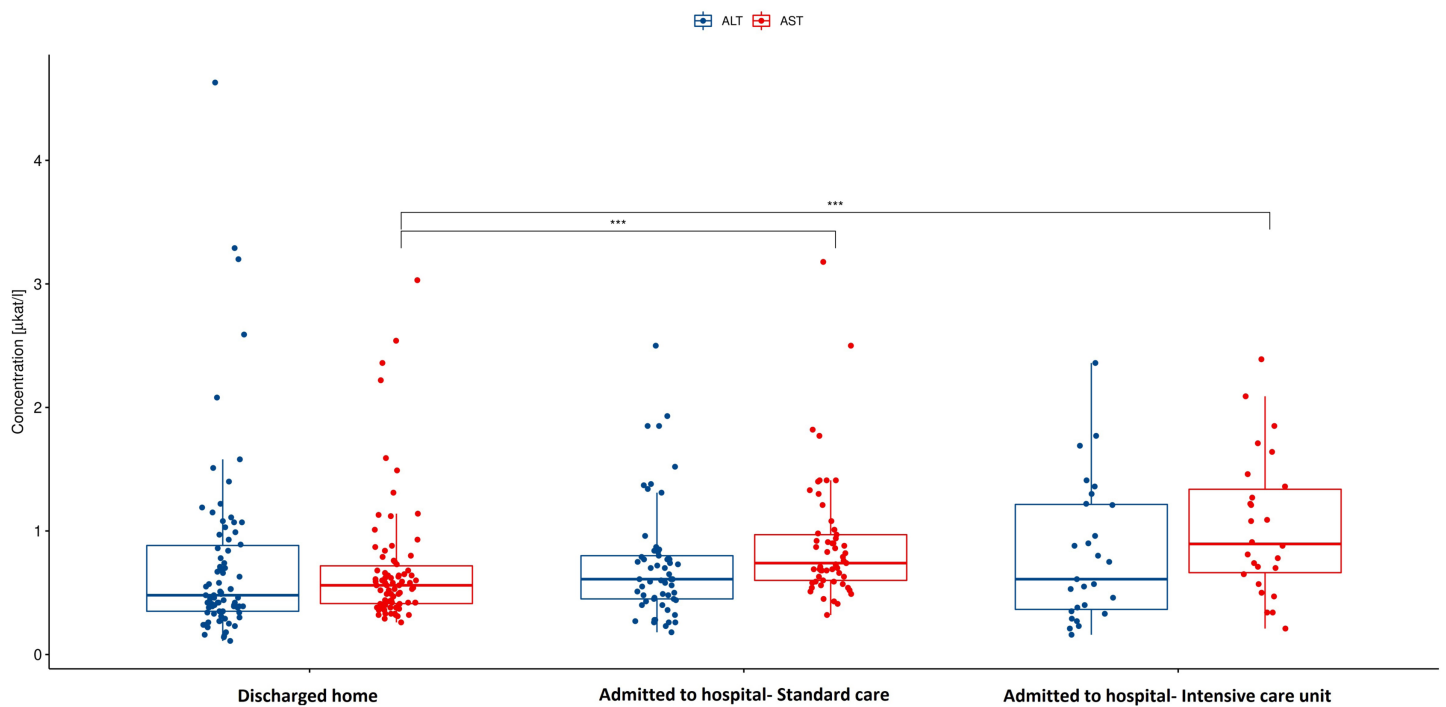


**Table 2** Gastrointestinal symptoms occurrence and laboratory findings.

		Discharged home ( <i>n</i> = 78) <sup>1</sup>	Admitted to hospital-standard care ( <i>n</i> = 57) <sup>1</sup>	Admitted to hospital-ICU ( <i>n</i> = 27) <sup>1</sup>	<i>p</i> value <sup>2</sup>
Gender					0.55
	female	41 (53%)	27 (47%)	11 (41%)	
	male	37 (47%)	30 (53%)	16 (59%)	
No of days with symptoms		7 (5,10)	7 (5,10)	5 (2,7)	0.03
Age		48 (42, 58)	55 (46, 67)	65 (55, 72)	<0.001
Diarrhea					0.02
	Not-presented	64 (82%)	34 (60%)	19 (70%)	
	Presented	14 (18%)	23 (40%)	8 (30%)	
Abdominal pain					0.23
	Not-presented	71 (91%)	48 (84%)	26 (96%)	
	Presented	7 (9.0%)	9 (16%)	1 (3.7%)	
Vomitus					0.24
	Not-presented	65 (83%)	53 (93%)	23 (85%)	
	Presented	13 (17%)	4 (7.0%)	4 (15%)	
Nausea					0.22
	Not-presented	70 (90%)	45 (79%)	23 (85%)	
	Presented	8 (10%)	12 (21%)	4 (15%)	
Heart burn					0.83
	Not-presented	76 (97%)	56 (98%)	26 (96%)	
	Presented	2 (2.6%)	1 (1.8%)	1 (3.7%)	
Bloating					0.03
	Not-presented	76 (97%)	49 (86%)	26 (96%)	
	Presented	2 (2.6%)	8 (14%)	1 (3.7%)	
Constipation					>0.9
	Not-presented	76 (97%)	55 (96%)	26 (96%)	
	Presented	2 (2.6%)	2 (3.5%)	1 (3.7%)	
O2 saturation (%)		96 (95, 98)	93 (89, 96)	87 (83, 91)	<0.001
CRP (norm < 5 mg/l)		29 (8, 74)	99 (65, 140)	102 (56, 166)	<0.001
AST (norm < 0.6 ukat/l)		0.56 (0.41, 0.72)	0.74 (0.60, 0.97)	0.91 (0.68, 1.41)	<0.001
ALT (norm < 0.6 ukat/l)		0.48 (0.35, 0.88)	0.61 (0.45, 0.80)	0.61 (0.36, 1.21)	0.25
Bilirubin (norm < 21 ukat/l)		8.8 (7.1, 11.1)	9.2 (8.3, 13.9)	9.5 (7.8, 12.5)	0.25
Recent ATB usage					0.09
	No	45 (58%)	22 (39%)	13 (48%)	
	Yes	33 (42%)	35 (61%)	14 (52%)	
Diabetes Mellitus					<0.001
	No	72 (92%)	45 (79%)	15 (56%)	
	Yes	6 (7.7%)	12 (21%)	12 (44%)	
Arterial Hypertension					0.003
	No	49 (63%)	26 (46%)	7 (26%)	
	Yes	29 (37%)	31 (54%)	20 (74%)	
Chronic liver disease					
	No	72 (92%)	47 (82%)	22 (81%)	
	Yes	6 (8%)	10 (18%)	5 (19%)	

**Notes:**<sup>1</sup> Statistics presented: *n* (%); Median (IQR).<sup>2</sup> Statistical tests performed: chi-square test of independence; Kruskal-Wallis test; Fisher's exact test.

CRP, C- reactive protein; AST, aspartate transaminase; ALT, alanine transaminase; ATB: antibiotics.



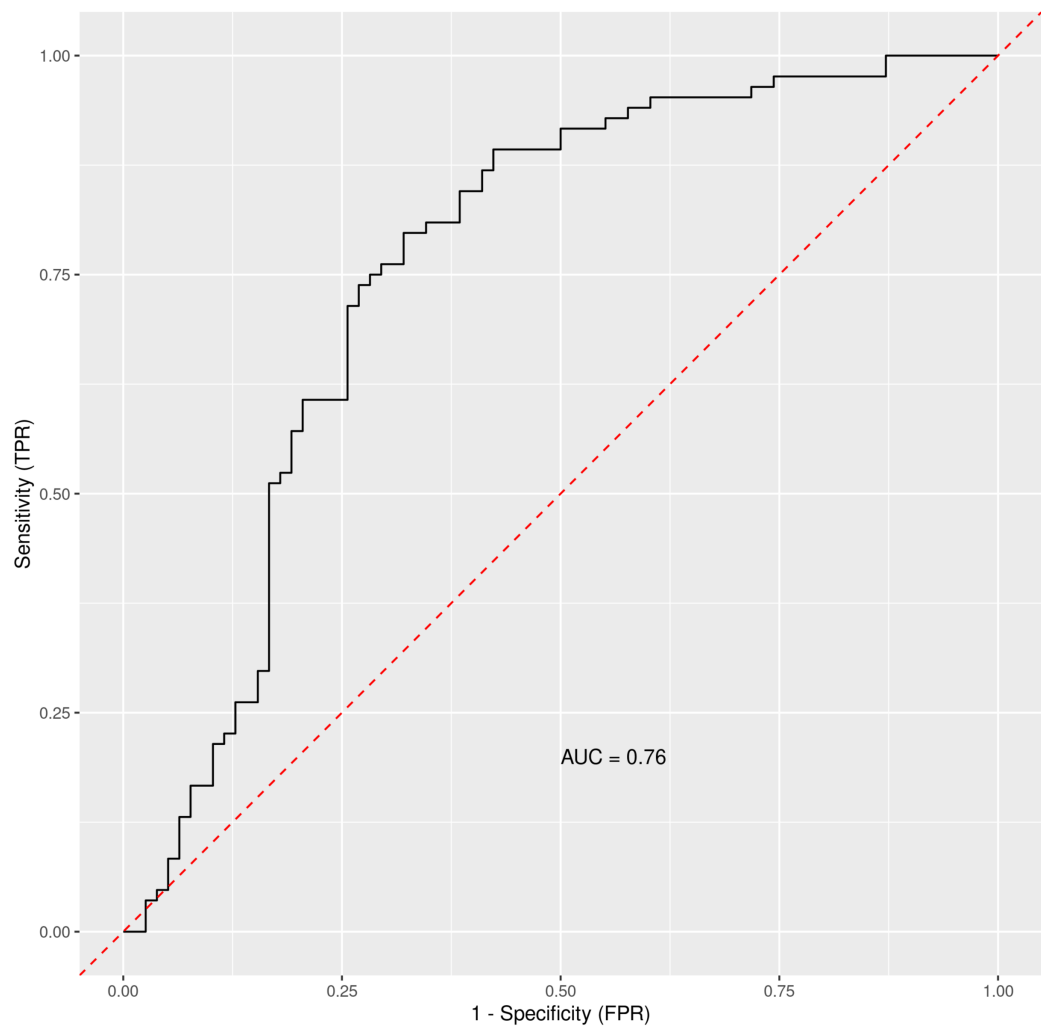
**Figure 1** Aspartate and alanine transaminase in hospitalized patients vs discharged home. AST, Aspartate transaminase; ALT, Alanine transaminase;  $p < 0.001$ . AST activity is significantly higher in the hospitalized group compared to patients discharged home after visit to emergency department. There are no significant differences in ALT activity between this groups of patients. [Full-size !\[\]\(1663bb69f307a960345edb0e712f8c02\_img.jpg\) DOI: 10.7717/peerj.13124/fig-1](https://doi.org/10.7717/peerj.13124/fig-1)

symptoms and liver related laboratory results. The PCA plot (Fig. 4) is showing two distinct clusters which are partially overlapping with tendencies to shift apart.

## DISCUSSION

Several studies and meta-analyses have pointed out the gastrointestinal involvement in the SARS-CoV-2 infection (Mao *et al.*, 2020; Sultan *et al.*, 2020; D'Amico *et al.*, 2020; Xiao *et al.*, 2020; Pan *et al.*, 2020; Villapol, 2020; Galanopoulos *et al.*, 2020). The data from the pooled prevalence of gastrointestinal symptoms are varying significantly from 10.5% to 53% between studies (Mao *et al.*, 2020; Sultan *et al.*, 2020; Pan *et al.*, 2020; Ashktorab *et al.*, 2021). Based on comprehensive meta-analysis by Sultan *et al.* (2020), the pooled prevalence of diarrhea is 7.7%, nausea and vomiting 7.8% and abdominal pain 3.6%. In the presented study we have focused on the presence of diarrhea, constipation, bloating, nausea, heart burn and abdominal pain. Statistically significant differences have been found in the case of diarrhea and nausea when comparing SARS-CoV-2 negative and positive patients. In the group of hospitalized patients (with standard care) the diarrhea was presented in 40% patients and nausea in 21%, which is higher compared to some meta-analysis mentioned, but consistent with the data considering general presence of gastrointestinal symptoms and gut involvement. When comparing only emergency department group the presence of bloating is significantly higher in the hospitalized group than in those who were discharged home. Interestingly, bloating has lower prevalence in the group of ICU patients than in patients with standard care management.



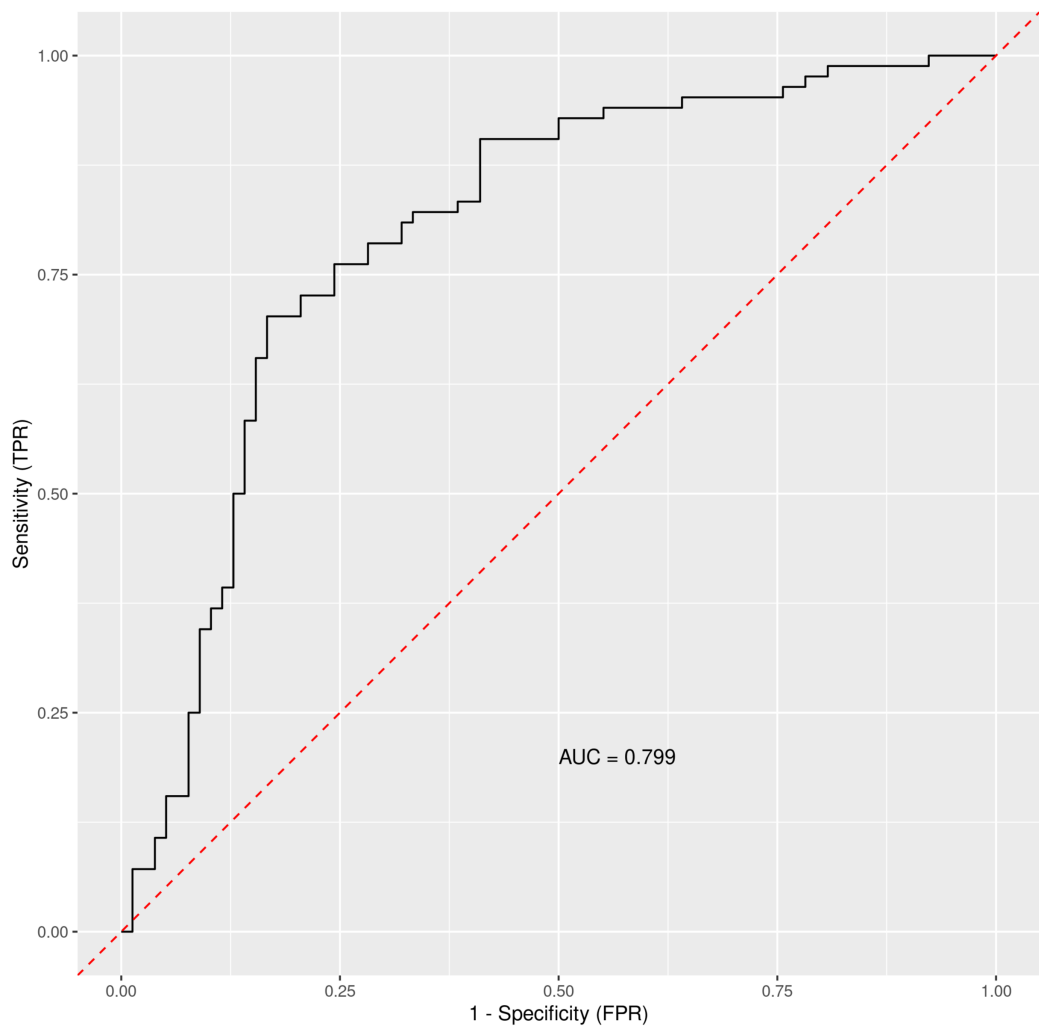


**Figure 2** The ROC curve for general COVID-19 and gastrointestinal symptoms and other measurable data in general clinical settings. Out-of-bag receiver operating characteristic curve with calculated area under the curve (AUC) = 0.76. The Matthews correlation coefficient was 0.48. For analysis were considered: general COVID-19 symptoms, gastrointestinal symptoms, age, sex, lasting of the symptoms and comorbidities (diabetes mellitus, arterial hypertension and chronic liver diseases).

Full-size  DOI: [10.7717/peerj.13124/fig-2](https://doi.org/10.7717/peerj.13124/fig-2)

This could be explained by high subjectivity and interpersonal differences when reporting symptom such as bloating. Considering differences between these two groups of patients, those with more severe course of disease attach lower importance to less annoying symptoms such as heart burn, bloating and nausea when compared to more manifested symptoms such as diarrhea, abdominal pain or vomitus.

Focusing on the liver enzymes as markers of possible liver impairment resulting from SARS-CoV-2 infection the AST, ALT and bilirubin were considered for the evaluation. The results are showing that median level of liver enzymes was not elevated in the discharged group. Bilirubin and ALT were also within normal range in the hospitalized group with no statistically significant differences between these two groups. Only AST was elevated over the upper level of the reference value in the hospitalized group with



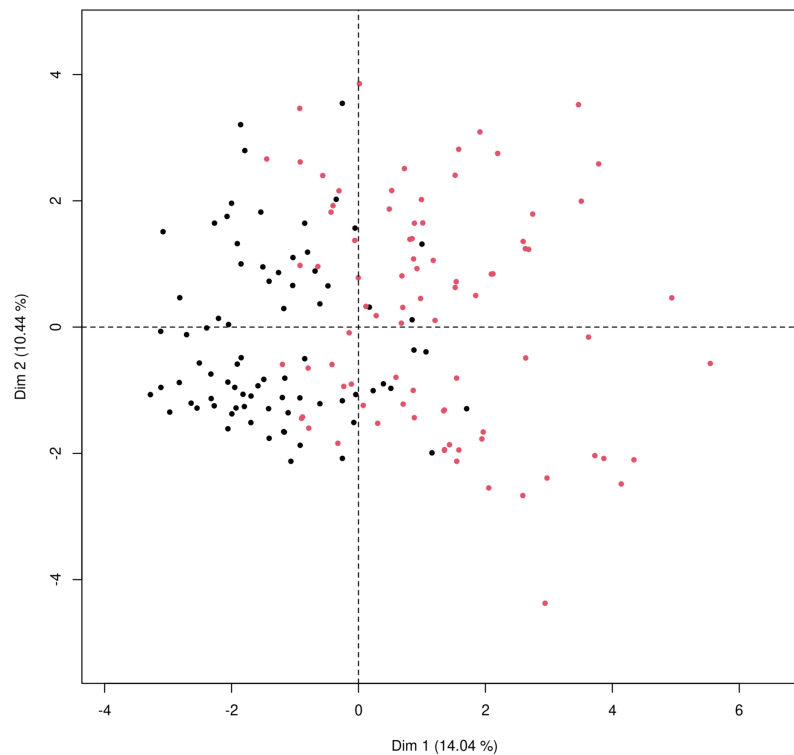
**Figure 3** The ROC curve for selected parameters. Out-of-bag receiver operating characteristic curve with calculated area under the curve (AUC) = 0.799. The Matthews correlation coefficient was 0.37. For analysis were considered selected parameters (clinically easily measurable): liver enzymes (AST, ALT), gastrointestinal symptoms (diarrhea and bloating), chronic liver disease, age and diabetes mellitus. [Full-size !\[\]\(1663bb69f307a960345edb0e712f8c02\_img.jpg\) DOI: 10.7717/peerj.13124/fig-3](https://doi.org/10.7717/peerj.13124/fig-3)

**Table 3** Relative predictors values.

	All	Variable importance	
		Discharged home	Admitted to hospital
AST	0.1451	0.5729	0.2217
Diabetes Mellitus	0.0248	0.1080	0.0288
Chronic liver disease	0.0169	0.0882	0.0061
ALT	0.0110	0.1025	-0.0379
Diarrhea	0.0068	0.0087	0.0272
Age	0.0139	0.0561	0.0197
Bloating	0.0011	0.0052	0.0006

**Note:**

AST, aspartate transaminase; ALT, alanine transaminase.



**Figure 4** Principal component analysis for mixed type of data to obtain two-dimensional representation of the data. Patients who were discharged home are marked as black dots and those who were admitted to the hospital marked as red dots. The first principal component (x axis) explains 14.04% of the variability; the second principal component (y axis) explains 10.44% of the variability in data. The two groups cannot be completely separated, as there is some overlap of the observations but there is a clear tendency to shift apart of the clusters. [Full-size](#) DOI: [10.7717/peerj.13124/fig-4](https://doi.org/10.7717/peerj.13124/fig-4)

progressively higher values in patients who required ICU. The differences between hospitalized and discharged patients are substantially significant.

Several previously published data have shown an elevation in both transaminases and bilirubin to a different extent ranging from 1% to 53% (mainly ALT and AST accompanied by slightly increased bilirubin concentrations) (Mao *et al.*, 2020). In most published data, severe liver alterations were uncommon (Marasco *et al.*, 2021) and the pooled prevalence of liver injury regarding severity was 12% based on the meta-analysis by Mao *et al.* (2020). More severe liver injury was also associated with worse outcomes, including intensive care unit admission and mortality (Phipps *et al.*, 2020).

The pathophysiology of liver involvement in COVID-19 is still not completely understood. The direct viral infection of the liver cells is proposed as one of potential causes of liver injury, but the comprehensive studies are scarce. A study with pathological analysis of liver tissues from dead victims of COVID-19 showed no viral inclusions in hepatocytes (Zhang, Shi & Wang, 2020). Another repeatedly proposed and generally accepted mechanism of liver impairment could be drug toxicity (Mao *et al.*, 2020). In order to determine the possible influence of recent ATB usage on the elevation of AST presented in this paper, a two-way analysis of variance (two-way ANOVA) was performed.

There are no significant differences between the groups with or without recent antibiotics usage. Therefore, we have concluded that ATB usage has no relevant influence on the elevated AST levels. The two-way ANOVA was also performed to assess the relationship between the presence of chronic liver disease and AST. There is no statistically relevant difference in AST levels in hospitalized patients with and without chronic liver disease.

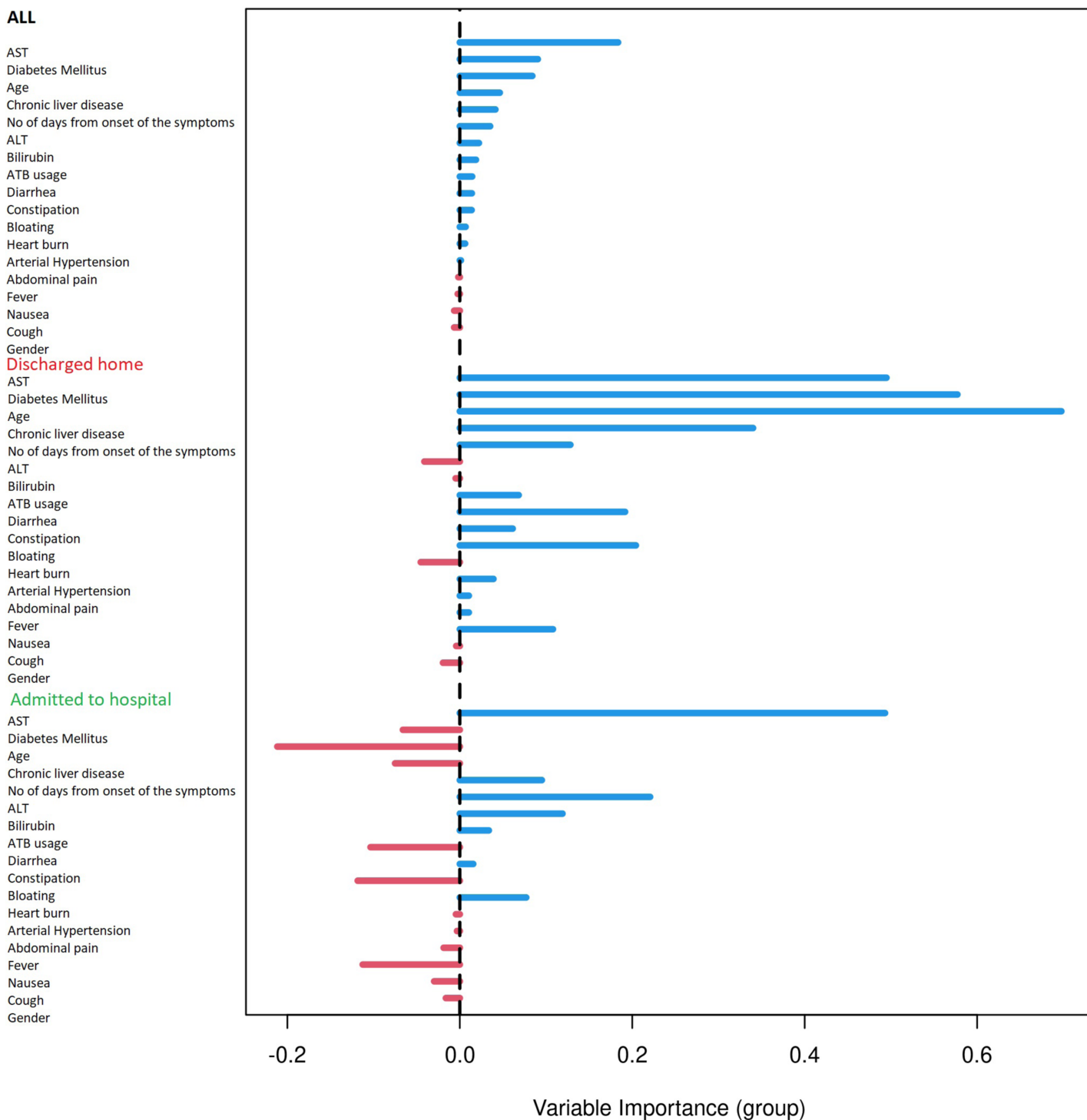
Another possible explanation of elevated transaminases is that it could be the result of a systemic inflammation. ALT is an enzyme most commonly found in liver, with small levels in striated muscle tissue and myocardium. On the other hand, AST could be found in liver, but also in striated and myocardial muscle, kidneys, brain and red blood cells. AST had been used as a marker for myocardial infarction for a long time before more sensitive markers were identified and implemented into the routine clinical practice (*Ndrepepa, 2021*). Based on the results of this study and current knowledge of SARS-CoV-2 interaction in human organism it is possible that elevated levels of AST in COVID-19 patients could be the result of a systemic inflammation with general tissue hypoperfusion rather than a result of a direct influence of SARS-CoV-2 virus on the hepatocytes or hepatotoxic drug use.

Further, we focused on identifying the possible predicting factors for hospitalization in COVID-19 patients using the Random Forest (RF) machine learning algorithm.

Different types of machine learning are being used in an increased rate to determine the predictors of outcome in various areas of clinical practice from brain trauma injuries (*Hanko et al., 2021*), radiology (*Choy et al., 2018*), oncology (*Cruz & Wishart, 2017*) to dermatology (*Rajkomar, Dean & Kohane, 2019*). Since COVID-19 pandemic has been affecting the global population for more than two years now and it is the cause of an immense health crisis in most world countries new diagnostic tools—machine learning being one of them—and therapeutic methods have been rapidly emerging (*Alimadadi et al., 2020*). Shortly after the COVID-19 outbreak various machine learning techniques were used, including taxonomic classification of COVID-19 genomes (*Randhawa et al., 2020*), determining the predictors of severe COVID-19 (*Yan et al., 2020*) and searching for new potential drug candidates against SARS-CoV-2 viral infection (*Ge et al., 2021*). Another example of a successful implementation of artificial intelligence in COVID-19 diagnosis is the evaluation of the CT scans detecting SARS-CoV-2 associated pneumonia and their differentiation from the community acquired pneumonia and other similar conditions with specificity and sensitivity higher than 90% (*Li et al., 2020*).

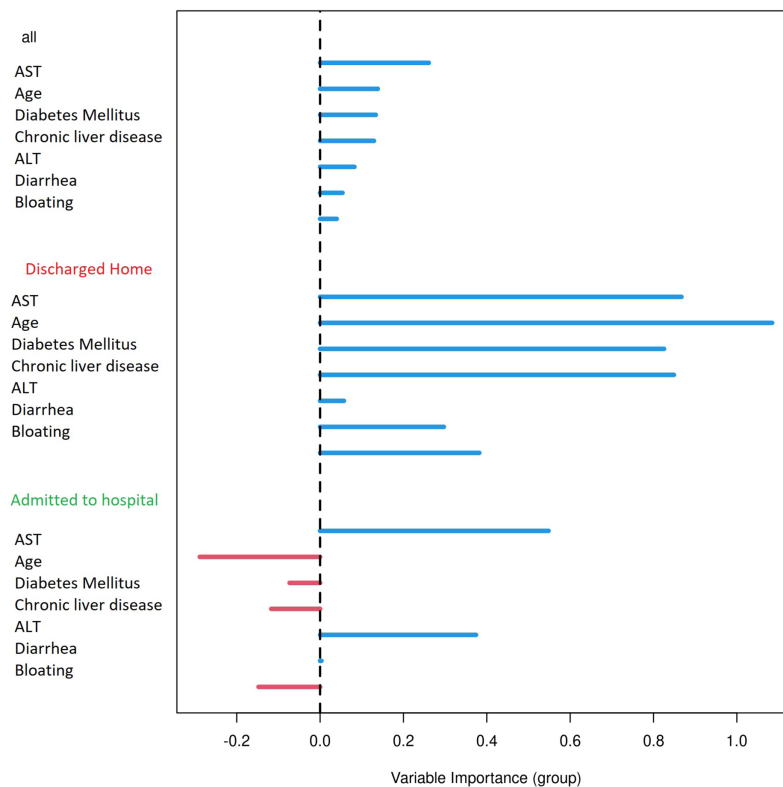
So far, several studies have been published using Random Forest Machine Algorithm for identifying the predictors for COVID-19 outcome from a wide variety of symptoms, socioeconomical factors (*Wollenstein-Betech et al., 2020*) and laboratory results with various results (*Iwendi et al., 2020; Jie et al., 2020*). To our current knowledge there are no studies specifically focused on the gastrointestinal symptoms and gut related laboratory findings to this date.

In order to assess the predictive power of the gastrointestinal parameters and other measured variables for predicting the need for hospitalization the Random Forest machine learning algorithm was trained on the data from our study. Random Forest has become the Machine Learning method of choice for several reasons: (a) it usually appears among



**Figure 5** Variable importance plot for all measured factors. Variable importance plot for all measured factors. The positive value of importance of a predictor represents a positive factor for the predictive accuracy of the Random Forest algorithm. The negative value of importance of a predictor indicates that omitting the predictor increases the predictive accuracy of the Random Forest algorithm.

Full-size DOI: 10.7717/peerj.13124/fig-5



**Figure 6 Variable importance plot for selected factors.** Variable importance plot for selected factors that are fast and easy to measure in the emergency department setting (liver enzymes: AST and ALT, gastrointestinal symptoms /diarrhea and bloating/, age and presence of chronic liver disease and diabetes mellitus). The positive value of importance of a predictor represents a positive factor for the predictive accuracy of the Random Forest algorithm. The negative value of importance of a predictor indicates that omitting the predictor increases the predictive accuracy of the Random Forest algorithm.

Full-size DOI: [10.7717/peerj.13124/fig-6](https://doi.org/10.7717/peerj.13124/fig-6)

the top performing classification algorithms; (b) it has small number of tuning parameters; (c) it does not overfit; (d) and last but not least, by its construction it provides a realistic estimate of the performance on a future data *via* the Out-Of-Bag data. Moreover, Random Forest, at least as implemented in the R library randomForestSRC, provides two different measures of importance of predictors. For these reasons, we have selected RF algorithm to assess predictive power of the studied variables, and to obtain their ranking.

Results were plotted as a ROC curve obtained from the Out-Of-Bag data. When considering the general COVID-19 symptoms, gastrointestinal symptoms, age, sex, lasting of the symptoms and comorbidities (diabetes mellitus, arterial hypertension and chronic liver diseases) the AUC is 0.76. The variable importance plot is shown in Fig. 5. When measuring the variable importance, the most important predictor is AST followed by age and diabetes mellitus, which are substantially less important. When using only liver enzymes (AST, ALT), gastrointestinal symptoms (diarrhea and bloating), age and presence of chronic liver disease and diabetes mellitus the AUC is 0.799 with AST as the strongest predictor for hospitalization. The variable importance plot is shown in Fig. 6. Previously published studies, which used mostly the methods of classical statistics, have



identified the presence of gastrointestinal symptoms (*Sun et al., 2020*), predominantly diarrhea (*Aumpan, Nunanan & Vilaichone, 2020; Ghoshal et al., 2020*) and elevated liver enzymes (*Aziz et al., 2020*) as predictors of hospitalization associated with COVID-19. In our data, we have singled out aspartate transaminase (AST) as not only the statistically significantly elevated liver enzyme in patients requiring hospitalization, but using the artificial intelligence with the Random Forest algorithm the AST proved to be the most important predictor of hospitalization. Finally, we performed the principal component analysis for mixed type of data in order to obtain a two-dimensional representation of the data on patients who were discharged home and those who were admitted to hospital. As could be seen on the Plot 4 these two groups are partially overlapping, but with clear tendencies to shift apart, which is in accordance with the predictive performance of the studied variables in the Random Forest algorithm.

## CONCLUSIONS

This study has identified elevated AST for being the most important predictor for COVID-19 related hospitalizations using the machine learning Random Forest algorithm. We have also shown that SARS-CoV-2 positivity is connected with isolated AST elevation and the level is linked with the severity of the disease. Furthermore, the prevalence of diarrhea and nausea among SARS-CoV-2 positive patients is significantly higher compared to SARS-CoV-2 negative controls. Bloating is occurring significantly more frequently in COVID-19 patients who require hospitalization than those who could be discharged to outpatient care.

## ABBREVIATIONS

PCR	polymerase chain reaction
ALT	alanin transaminase
AST	aspartate transaminase
GMT	gamma glutamyltransferase
ICU	intensive care unit
ED	emergency department
ROC	receiver operating characteristic
PCA	principal component analysis
RF	Random Forest
ATB	antibiotics

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This publication has been produced with the support of: The Integrated Infrastructure Operational Program for the project: Research and development of telemedicine solutions to support the fight against pandemic diseases induced COVID-19 and reducing its negative consequences by monitoring the health status of people in order to eliminate the risk of infection in at-risk populations, ITMS: 313011ASY8, co-financed by the European

Regional Development Fund, by the Integrated Infrastructure Operational Program for the project: New possibilities for laboratory diagnostics and massive screening of SARS-Cov-2 and identification of mechanisms of virus behavior in human body, ITMS: 313011AUA4, co-financed by the European Regional Development Fund; and by Ministry of Health of the Slovak Republic under the project registration number 2019/44-UKMT-7. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### **Grant Disclosures**

The following grant information was disclosed by the authors:

Research and Development of Telemedicine Solutions to Support the Fight against Pandemic Diseases Induced COVID-19 and Reducing its Negative Consequences by Monitoring the Health Status of People in Order to Eliminate the Risk of Infection in at-risk Populations, ITMS: 313011ASY8.

New Possibilities for Laboratory Diagnostics and Massive Screening of SARS-Cov-2 and Identification of Mechanisms of Virus Behavior in Human Body, ITMS: 313011AUA4. Ministry of Health of the Slovak Republic: 2019/44-UKMT-7.

### **Competing Interests**

The authors declare that they have no competing interests.

### **Author Contributions**

- Peter Lipták conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Peter Banovcin conceived and designed the experiments, prepared figures and/or tables, and approved the final draft.
- Róbert Rosolanka performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Michal Prokopič performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Ivan Kocan performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Ivana Žiačiková performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Peter Uhrík performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Marian Grendar analyzed the data, prepared figures and/or tables, and approved the final draft.
- Rudolf Hyrdel conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

## Human Ethics

The following information was supplied relating to ethical approvals (*i.e.*, approving body and any reference numbers):

This study was approved by the Ethical committee of the University hospital in Martin, decision number: 14/2021.

## Ethics

The following information was supplied relating to ethical approvals (*i.e.*, approving body and any reference numbers):

This study was approved by the Ethical committee of the University hospital in Martin, decision number: 14/2021.

## Data Availability

The following information was supplied regarding data availability:

The raw data and the R code are available in the [Supplemental Files](#).

## Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.13124#supplemental-information>.

## REFERENCES

- Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X. 2020. Artificial intelligence and machine learning to fight COVID-19. *Physiological Genomics* 52(4):200–202 DOI 10.1152/physiolgenomics.00029.
- Ashktorab H, Pizuorno A, Oskroch G, Fierro NA, Sherif ZA, Brim H. 2021. COVID-19 in latin America: symptoms, morbidities, and gastrointestinal manifestations. *Gastroenterology* 160:938–940 W.B. Saunders DOI 10.1053/j.gastro.2020.10.033.
- Aumpan N, Nunanan P, Vilaichone R. 2020. Gastrointestinal manifestation as clinical predictor of severe COVID-19: a retrospective experience and literature review of COVID-19 in Association of Southeast Asian Nations (ASEAN). *JGH Open* 4(6):1096–1101 DOI 10.1002/jgh3.12394.
- Aziz M, Haghbin H, Lee-Smith W, Goyal H, Nawras A, Adler DG. 2020. Gastrointestinal predictors of severe COVID-19: systematic review and meta-analysis. *Annals of Gastroenterology* 33(6):615 DOI 10.20524/aog.2020.0527.
- Bachtiger P, Peters NS, Walsh SL. 2020. Machine learning for COVID-19—asking the right questions. *The Lancet Digital Health* 2(8):e391–e392 DOI 10.1016/S2589-7500(20)30162-X.
- Breiman L. 2001. Random forests. *Machine Learning* 45(1):5–32 DOI 10.1023/A:1010933404324.
- Chavent M, Kuentz V, Labenne A, Liquet B, Saracco J. 2017. PCAmixdata: multivariate analysis of mixed data. R package version 3.1. Available at <https://cran.r-project.org/web/packages/PCAmixdata/PCAmixdata.pdf>.
- Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pinykh OS, Geis JR, Pandharipande PV, Brink JA, Dreyer KJ. 2018. Current applications and future impact of machine learning in radiology. *Radiology* 288(2):318–328 DOI 10.1148/radiol.2018171820.
- Cruz JA, Wishart DS. 2017. Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics* 2:59–77 DOI 10.1177/117693510600200030.

- D'Amico F, Baumgart DC, Danese S, Peyrin-Biroulet L. 2020. Diarrhea during COVID-19 infection: pathogenesis, epidemiology, prevention, and management. *Clinical Gastroenterology and Hepatology* 18(8):1663–1672 W.B. Saunders DOI 10.1016/j.cgh.2020.04.001.
- Galanopoulos M, Gkeros F, Doukatas A, Karianakis G, Pontas C, Tsoukalas N, Viazis N, Liatsos C, Mantzaris GJ. 2020. COVID-19 pandemic: pathophysiology and manifestations from the gastrointestinal tract. *World Journal of Gastroenterology* 26:4579–4588 Baishideng Publishing Group Co DOI 10.3748/wjg.v26.i31.4579.
- Ge Y, Tian T, Huang S, Wan F, Li J, Li S, Wang X, Yang H, Hong L, Wu N, Yuan E, Luo Y, Cheng L, Hu C, Lei Y, Shu H, Feng X, Jiang Z, Wu Y, Chi Y, Guo X, Cui L, Xiao L, Li Z, Yang C, Miao Z, Chen L, Li H, Zeng H, Zhao D, Zhu F, Shen X, Zeng J. 2021. An integrative drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. *Signal Transduction and Targeted Therapy* 6(1):1–16 DOI 10.1038/s41392-021-00568-6.
- Ghoshal UC, Ghoshal U, Mathur A, Singh RK, Nath A, Garg A, Singh D, Singh S, Singh J, Pandey A, Rai S, Vasanth S, Dhiman RK. 2020. The spectrum of gastrointestinal symptoms in patients with coronavirus disease-19: predictors, relationship with disease severity, and outcome. *Clinical and Translational Gastroenterology* 11(12):e00259 DOI 10.14309/ctg.0000000000000259.
- Guan W-J, Ni Z-Y, Hu Y, Liang W-H, Ou C-Q, He J-X, Liu L, Shan H, Lei C-L, Hui DSC, Du B, Li L-J, Zeng G, Yuen K-Y, Chen R-C, Tang C-L, Wang T, Chen P-Y, Xiang J, Li S-Y, Wang J-L, Liang Z-J, Peng Y-X, Wei L, Liu Y, Hu Y-H, Peng P, Wang J-M, Liu J-Y, Chen Z, Li G, Zheng Z-J, Qiu S-Q, Luo J, Ye C-J, Zhu S-Y, Zhong N-S. 2020. Clinical characteristics of coronavirus disease 2019 in China. *New England Journal of Medicine* 382(18):1708–1720 DOI 10.1056/NEJMoa2002032.
- Hanko M, Mán G, Snopko P, Ré O, Šutovský J, Benčo M, Soršák J, Zelenák K, Kolarovszki B. 2021. Random forest-based prediction of outcome and mortality in patients with traumatic brain injury undergoing primary decompressive craniectomy. *World Neurosurgery* 148:e450–e458 DOI 10.1016/j.wneu.2021.01.002.
- Ishwaran H, Kogalur UB. 2021. Fast unified random forests for survival, regression, and classification (RF-SRC), R package version 2.11.0. Available at <https://cran.r-project.org/web/packages/randomForestSRC/randomForestSRC.pdf>.
- Iwendi C, Bashir AK, Atharva Peshkar RS, Chatterjee JM, Pasupuleti S, Mishra R, Pillai S, Jo O. 2020. COVID-19 patient health prediction using boosted random forest algorithm. *Frontiers in Public Health* 8:357 DOI 10.3389/fpubh.2020.00357.
- Jie W, Heping Y, Qingquan H, Shuili J, Zhifen L, Xiang P, Cheng'an C, Yongwen L. 2020. A descriptive study of random forest algorithm for predicting COVID-19 patients outcome. *PeerJ* 8:e9945 DOI 10.7717/peerj.9945.
- Jin X, Lian J-S, Hu J-H, Gao J, Zheng L, Zhang Y-M, Hao S-R, Jia H-Y, Cai H, Zhang X-L, Yu G-D, Xu K-J, Wang X-Y, Gu J-Q, Zhang S-Y, Ye C-Y, Jin C-L, Lu Y-F, Yu X, Yu X-P, Huang J-R, Xu K-L, Ni Q, Yu C-B, Zhu B, Li Y-T, Liu J, Zhao H, Zhang X, Yu L, Guo Y-Z, Su J-W, Tao J-J, Lang G-J, Wu X-X, Wu W-R, Qv T-T, Xiang D-R, Yi P, Shi D, Chen Y, Ren Y, Qiu Y-Q, Li L-J, Sheng J, Yang Y. 2020. Epidemiological, clinical and virological characteristics of 74 cases of coronavirus-infected disease 2019 (COVID-19) with gastrointestinal symptoms. *Gut* 69(6):1002–1009 DOI 10.1136/gutjnl-2020-320926.
- Kassambara A. 2020. ggpubr: 'ggplot2' based publication ready plots. R package version 0.4.0. Available at <https://cran.r-project.org/web/packages/ggpubr/index.html>.
- Kassambara A. 2021. Rstatix: pipe-friendly framework for basic statistical tests. R package version 0.7.0. Available at <https://cran.r-project.org/web/packages/rstatix/rstatix.pdf>.

- Kushwaha S, Bahl S, Bagha AK, Parmar KS, Javaid M, Haleem A, Singh RP. 2020.** Significant applications of machine learning for COVID-19 pandemic. *Journal of Industrial Integration and Management* 5(4):453–479 DOI 10.1142/S2424862220500268.
- Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q, Cao K, Liu D, Wang G, Xu Q, Fang X, Zhang S, Xia J, Xia J. 2020.** Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. *Radiology* 296(2):E65–E71 DOI 10.1148/radiol.2020200905.
- Mao R, Qiu Y, He J-S, Tan J-Y, Li X-H, Liang J, Shen J, Zhu L-R, Chen Y, Iacucci M, Ng SC, Ghosh S, Chen M-H. 2020.** Manifestations and prognosis of gastrointestinal and liver involvement in patients with COVID-19: a systematic review and meta-analysis. *The Lancet Gastroenterology and Hepatology* 5(7):667–678 DOI 10.1016/S2468-1253(20)30126-6.
- Marasco G, Lenti MV, Cremon C, Barbaro MR, Stanghellini V, Sabatino AD, Barbara G. 2021.** Implications of SARS-CoV-2 infection for neurogastroenterology. *Neurogastroenterology and Motility* 33(3):e14104 Blackwell Publishing Ltd DOI 10.1111/nmo.14104.
- Ndrepepa G. 2021.** Aspartate aminotransferase and cardiovascular disease—a narrative review. *Journal of Laboratory and Precision Medicine* 6:6 DOI 10.21037/jlpm-20-93.
- Pan L, Mu M, Yang P, Sun Y, Wang R, Yan J, Li P, Hu B, Wang J, Hu C, Jin Y, Niu X, Ping R, Du Y, Li T, Xu G, Hu Q, Tu L. 2020.** Clinical characteristics of COVID-19 patients with digestive symptoms in Hubei, China: a descriptive, cross-sectional, multicenter study. *American Journal of Gastroenterology* 115(5):766–773 DOI 10.14309/ajg.0000000000000620.
- Patel KP, Patel PA, Vunnam RR, Hewlett AT, Jain R, Jing R, Vunnam SR. 2020.** Gastrointestinal, hepatobiliary, and pancreatic manifestations of COVID-19. *Journal of Clinical Virology* 128:104386 DOI 10.1016/j.jcv.2020.104386.
- Phipps MM, Barraza LH, LaSota ED, Sobieszczyk ME, Pereira MR, Zheng EX, Fox AN, Zucker J, Verna EC. 2020.** Acute liver injury in COVID-19: prevalence and association with clinical outcomes in a large U.S. Cohort. *Hepatology* 72(3):807–817 DOI 10.1002/hep.v72.3.
- R Development Core Team. 2021.** A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <http://www.R-project.org/>.
- Rajkomar A, Dean J, Kohane I. 2019.** Machine learning in medicine. *The New England Journal of Medicine* 380(14):1347–1358 DOI 10.1056/NEJMra1814259.
- Randhawa GS, Soltysiak MPM, Roz HEI, de Souza CPE, Hill KA, Kari L. 2020.** Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. *PLOS ONE* 15(4):e0232391 DOI 10.1371/journal.pone.0232391.
- Signorell A. 2021.** DescTools: tools for descriptive statistics. R package version 0.99.41. Available at <https://cran.r-project.org/web/packages/DescTools/index.html>.
- Sjoberg DD, Curry M, Hannum M, Whiting K, Zabor EC. 2020.** Gtsummary: presentation-ready data summary and analytic result tables. R package version 1.3.5. Available at <https://cran.r-project.org/web/packages/gtsummary/gtsummary.pdf>.
- Sultan S, Altayar O, Siddique SM, Davitkov P, Feuerstein JD, Lim JK, Falck-Ytter Y, El-Serag HB. 2020.** AGA institute rapid review of the gastrointestinal and liver manifestations of COVID-19, meta-analysis of international data, and recommendations for the consultative management of patients with COVID-19. *Gastroenterology* 159(1):320–334.e27 DOI 10.1053/j.gastro.2020.05.001.
- Sun Y, Koh V, Marimuthu K, Ng OT, Young B, Vasoo S, Chan M, Lee VJM, De PP, Barkham T, Lin RTP, Cook AR, Leo YS, Ang B, Chuan Lee C, Lye DCB, Ling LM, Lee LS-U, Sadarangani S, Seong Wong C, Lee TH, Junhao Lin R, Chia PY, Sadasiv MS, Ng DHL, Choy CY, Yeo TW, Tan GSE, Chan YK, Tay JY, Lee PH, Ong SWX, Sutjipto S, Wee ILE,**

- Frederico D, Go CJ, Isais FS. 2020.** Epidemiological and clinical predictors of COVID-19. *Clinical Infectious Diseases* 71(15):786–792 DOI [10.1093/cid/ciaa322](https://doi.org/10.1093/cid/ciaa322).
- Villapol S. 2020.** Gastrointestinal symptoms associated with COVID-19: impact on the gut microbiome. *Translational Research* 226:57–69 Mosby Inc DOI [10.1016/j.trsl.2020.08.004](https://doi.org/10.1016/j.trsl.2020.08.004).
- Wollenstein-Betech S, Silva AAB, Fleck JL, Cassandras CG, Paschalidis IC. 2020.** Physiological and socioeconomic characteristics predict COVID-19 mortality and resource utilization in Brazil. *PLOS ONE* 15(10):e0240346 DOI [10.1371/journal.pone.0240346](https://doi.org/10.1371/journal.pone.0240346).
- Xiao F, Tang M, Zheng X, Liu Y, Li X, Shan H. 2020.** Evidence for gastrointestinal infection of SARS-CoV-2. *Gastroenterology* 158:1831–1833 DOI [10.1053/j.gastro.2020.02.055](https://doi.org/10.1053/j.gastro.2020.02.055).
- Yan L, Zhang H-T, Goncalves J, Xiao Y, Wang M, Guo Y, Sun C, Tang X, Jin L, Zhang M, Huang X, Xiao Y, Cao H, Chen Y, Ren T, Wang F, Xiao Y, Huang S, Tan X, Huang N, Jiao B, Zhang Y, Luo A, Mombaerts L, Jin J, Cao Z, Li S, Xu H, Yuan Y. 2020.** A machine learning-based model for survival prediction in patients with severe COVID-19 infection. *medRxiv* DOI [10.1101/2020.02.27.20028027](https://doi.org/10.1101/2020.02.27.20028027).
- Zhang C, Shi L, Wang F-S. 2020.** Liver injury in COVID-19: management and challenges. *The Lancet Gastroenterology and Hepatology* 5(5):428–430 DOI [10.1016/S2468-1253\(20\)30057-1](https://doi.org/10.1016/S2468-1253(20)30057-1).