Pancreatology 23 (2023) 176–186

Contents lists available at ScienceDirect

Pancreatology

journal homepage: www.elsevier.com/locate/pan

Early prediction of the severe course, survival, and ICU requirements in acute pancreatitis by artificial intelligence☆

Ali Tüzün İnce [a,*], Gökhan Silahtaroğlu [b], Gülseren Seven [a], Koray Koçhan [a], Kemal Yıldız [c], Hakan Şentürk [a]

[a] Gastroenterology Clinic, Adnan Menderes Bulvarı, Vatan Cad, 34093, Fatih, Istanbul, Turkey
[b] Istanbul Medipol University, School of Business and Management Sciences, Kavacık Campus, Istanbul, Turkey
[c] Internal Medicine Clinic, Medicana International Hospital, Beylikdüzü Cd. No:3, 34520, Beylikdüzü, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 7 August 2022
Received in revised form
20 November 2022
Accepted 12 December 2022
Available online 30 December 2022

Keywords:
Artificial intelligence
Machine learning
Acute pancreatitis
Severity
Balthazar CTSI
Revised Atlanta Classification

ABSTRACT

Objective: To evaluate the success of artificial intelligence for early prediction of severe course, survival, and intensive care unit(ICU) requirement in patients with acute pancreatitis(AP).
Methods: Retrospectively, 1334 patients were included the study. Severity is determined according to the Revised Atlanta Classification(RAC). The success of machine learning(ML) method was evaluated by 13 simple demographic, clinical, etiologic, and laboratory features obtained on ER admission. Additionally, it was evaluated whether Balthazar-computerized tomography severity index(CTSI) at 48-h contributed to success. The dataset was split into two parts, 90% for ML(of which 70% for learning and 30% for testing) and 10% for validation and 5-fold stratified sampling has been utilized. Variable Importance was used in the selection of features during training phase of machine. The Gradient Boost Algorithm trained the machine by KNIME analytics platform. SMOTE has been applied to increase the minority classes for training. The combined effects of the measured features were examined by multivariate logistic regression analysis and reciever operating curve curves of the prediction and confidence of the target variables were obtained.
Results: Accuracy values for the early estimation of Atlanta severity score, ICU requirement, and survival were found as 88.20%, 98.25%, and 92.77% respectively. When Balthazar-CTSI score is used, results were found as 91.02%, 92.25%, and 98% respectively.
Conclusions: The ML method we used successfully predicted the severe course, ICU requirement and survival, with promising accuracy values of over 88%. If 48-h Balthazar-CTSI is included in the calculation, the severity score and survival rates increase even more.

1. Introduction

Acute pancreatitis(AP) is acute inflammation of the pancreas which may lead to results ranging from mild course to death. Most of the patients recover within 7–10 days, but 5–10% of them have a troubled course, intensive care unit(ICU) treatment may be required and may lose their lives. Overall mortality rate changes between 2 and 35% depending on whether presence of organ failure(OF) or necrotizing AP [1] and higher in the severe acute pancreatitis(SAP). Clinical predictors for poor course are older age [2], alcoholic etiology [3], obesity [4], OF that starting early and lasts longer than 48h [5]. Having a CRP level ≥150 mg/dl in 48-h, BUN level ≥20 mg/dl [6], a rising creatinine level, and a Balthazar-CTSI score ≥8 [7] are the indicators of severe course. A 48-h CRP≥150 mg/dl is an important mark for severity, but it has low Area Under Curve(AUC) values [8]. The Acute Physiology and Chronic Health Evaluation(APACHE)-II scoring is cumbersome to use and it has been shown that it does not accurately predict severity in the first 24h [9,10]. Bedside index for severity in AP(BISAP) and Harmless AP(HAPS) scorings cannot distinguish between moderately severe AP(MSAP) and SAP [11,12]. Despite its popularity today, the Revised Atlanta Classification(RAC) needs to be revised due to some deficiencies in the subjects of infected

☆ 'Artificial Intelligence in Severe Course of Acute Pancreatitis'.

* Corresponding author.
E-mail addresses: dralince@gmail.com (A.T. İnce), gsilahtaroglu@medipol.edu.tr (G. Silahtaroğlu), gulserenseven@yahoo.com (G. Seven), kochankoray@yahoo.com (K. Koçhan), kemalyildiz571@gmail.com (K. Yıldız), drhakansenturk@gmail.com (H. Şentürk).

https://doi.org/10.1016/j.pan.2022.12.005
1424-3903/

necrosis [13–16], extra-pancreatic infection [17] and necrosis and the dynamic nature of [18].

Early prediction of the topics of severe course, survival, and ICU requirement in patients carries utmost importance. Although presence of many laboratories, clinical, radiological predictors and scoring systems, none of them can accurately predict which patient will have severe disease course and most of them have shortcomings. For this reason, other searches have started in this regard [19]. Today, successfully usage of artificial intelligence(AI) in many areas of medicine has drawn attention to its usability in these topics. Machine learning(ML) is a branch of artificial intelligence and computer science that can precisely predict results by learning complex relationships between data using mathematical algorithms. AI applications have been used in diagnosis [20,21], drug development [22], gene editing skills [23,24], risk of malignancy in mammograms [25], response to treatment [26], in-hospital mortality in ICU [27]. Studies has also begun to appear on the early accurate prediction of the worsening course of AP [27–30].

In this study, success of AI method in early prediction of severe course, ICU requirement, and survival topics will be evaluated using some clinical, demographic findings, etiology, and laboratory tests during hospital admission time. It will also be investigated whether addition of 48-h Balthazar-CTSI will change the success rate.

2. Material and method

Retrospectively, 1550 AP patients those followed up at the Gastroenterology Clinic of Bezmialem Vakif University, Istanbul-Türkiye, during 10/2010-12/2020 period and diagnosed according to the RAC were screened. After exclusions due to lack of data, 1334 patients(aged≥18yo) were included. The study is approved by the local ethical council(14/285–25.08.2020) and registered in ClinicalTrials.gov(NCT04735055). Patients' data were obtained from hospital records. Patients with the presence of 2 out of 3 findings according to RAC were diagnosed as AP(severe girdle-style epigastric pain, amylase/lipase values ≥3ULN, and imaging studies compatible with AP) [10]. Patients with chronic pancreatitis, previous pancreatic surgery, pancreatic cancer, and contraindicated cases for contrast study were excluded from the study. The severity of the AP is classified as mild(no local complications and OF), moderate(local complications ± transient OF<48 h), and SAP(presence of permanent OF)]. OF(lung, kidney, gastrointestinal and cardiovascularly) was determined according to the Marshall scoring(present if ≥ 2) [31]. ICU admissions(yes/no) and survival(dead/alive) were also recorded. AI results were matched by the results of RAC scoring, ICU requirement and survival of the patients at the training test stage of the AI. Complications of AP were classified and numbered as 0:none, 1:local complications(defines the moderate AP; pseudocyst, abscess, necrosis, thrombosis, and mesenteric panniculitis), 2:systemic complications(presence of longer than 48th hour; defines the SAP) 3:mixed serious complications/co-morbidity situations, 4:infectious/septic complications. Infected pancreatic necrosis and sepsis developed during AP were accepted as SAP due to the inadequacy of some issues in Atlanta severity scoring.

2.1. Artificial intelligence method

In Artificial Intelligence, Decision Tree Models are widely used for supervised ML. They may depend on gini index, gain ratio/entropy, chi-square, regression and so on [32]. In AI they are preferred because they generate understandable rules for humans unlike other ML algorithms such as Artificial Neural Networks(ANN) and Support Vector Machines. On the other hand, they are weak learners. That means they are highly affected from noise and outliers existing in the data set [33]. To go around this handicap, models like Random Forest, Ensemble Trees, Gradient Boosting have been developed. The Gradient Boost Tree Model(GBTM) was used for the ML [34]. This model chooses a separate optimal value for each of the tree's parts rather than one for the whole tree. It is reported that the GBTM outperforms Random Forest and Regular Ensemble Trees [35] in many cases(Figure-1). Variable Importance was used for selection of features to understand what the machine finds important during training phase and how much benefit it provides from which parameter. When GBTM algorithm trains the machine, it uses each feature alone or with others to make predictions(Figure-1). That means, the algorithm calculates each algorithm's contribution to the correct predictions in various scenarios. This is done using the decrease in node impurity(Gini index) weighted by the probability of reaching that node. The node probability value is calculated by the number of samples to reach the node, divided by the total number of samples in all. In the analysis, Synthetic Minority Oversampling Technique(SMOTE) [36], a data augmentation technique, was used to avoid the disadvantage of class variable imbalance.

The dataset has been partitioned with 90%–10% ratio. 10% is for validation and 90% is for AI-ML. 90% ML part has also been divided into two parts as 70% for AI learning and 30% for testing the learning(Figure-2 and 3). For this purpose, 5-fold stratified sampling has been used. KNIME analytic platform has been used for AI-ML [37].

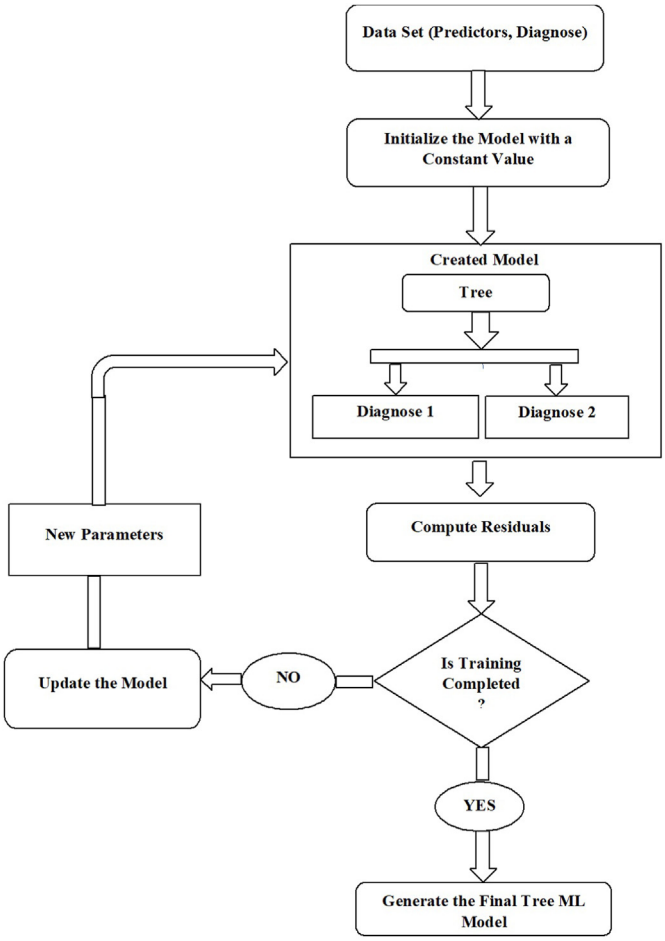

Fig. 1. General working system of Gradiest boost tree model.

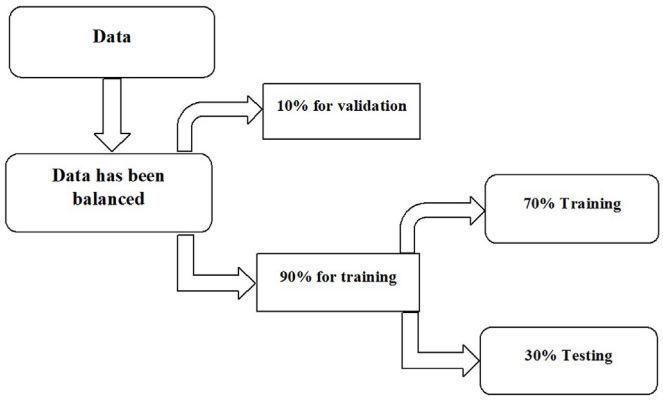

Fig. 2. Shows how Data have been used for Training,Testing and Validation.

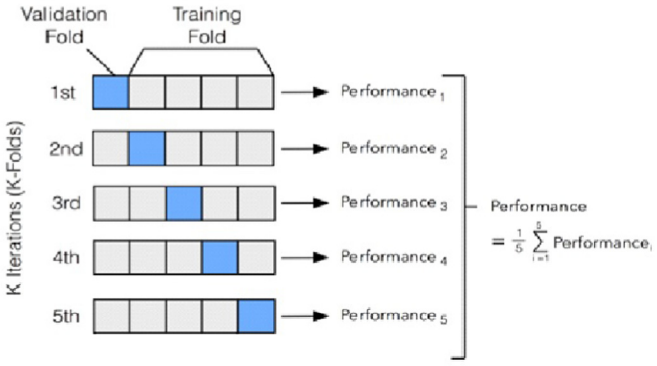

Fig. 4. Depicts the Schema of the 5-fold cross validation.

2.2. Selection of the training and validation groups

Stratified sampling was performed based on the target variable. Meanwhile, elections were held at random. However, there may be some problems with the train/test separation. We may not have been able to make the data set separation randomly. For this, 5-fold cross validation was applied. In k-folds cross validation, we divided our data into k different subsets. We used k-1 subsets to train our data and leave the final subset as test data(Figure-4).

Thirteen features those obtained from patients on hospital admission were used [; 1.gender(male/female); 2.Age(years); 3.Body Mass Index(BMI); group-1(≤25 kg/m^2), group-2(25–30 kg/m^2), and group-3(>30,1 kg/m^2), height(m); weight(kg); 4.cigarette(no/yes); 5.alcohol(no/yes); 6.diabetes mellitus(DM)(no/yes); 7.hypertension(no/yes); 8.etiology:[a; biliary, b; alcohol, c; hypertriglyceridemia, d; hypercalcemia, e; drug, f; hereditary, g; idiopathic, h; post endoscopic retrograde cholangiography(ERCP), ı; oddy sphincter dysfunction(OSD), i; other malignity, j; intra-ductal papillary mucinous neoplasia(IPMN), k:primary sclerosing cholangitis(PSC), l:autoimmune, m:multiple etiology]; 9.white blood cell number(WBC)(N:4,5-11x100/mm^3); 10.hematocrit(Hct)(N:35,5–48%); 11.C-reactive protein(CRP)(N:0–5 mg/dl); 12.blood urea nitrogen(BUN)(N:9,8–20,1 mg/dl); 13.creatinine(N:0,57-1,11 mg/dl)]. The results of AI study on the wanted topics were compared by the Baltazar-CTSI feature that obtained about 48-h [0; normal, 1; increase in pancreatic size, 2; inflammatory changes in pancreatic tissue and peripancreatic fatty tissue, 3; irregularly bordered, single fluid collection, 4; irregularly bordered 2 or more fluid collections, with various degrees of necrosis(between 5 and 10)].

The mathematical calculation of the Gradient Boost Algorithm and the underlying mathematical formula are summarized below.

2.3. Inputs

1) A training data set: $\{(x_i, y_i)\}\ i = 1\ to\ n$ with n dimension and a class variable
2) A differentiable loss function: $L(y, F(x))$
3) The number of iterations: M.

2.4. Algorithm

1. Initialize the model with a constant value:

$$F_0(x) = \arg\min \sum_{i=1}^{n} L(y_i, \gamma)$$

2. For $m = 1$ to M:

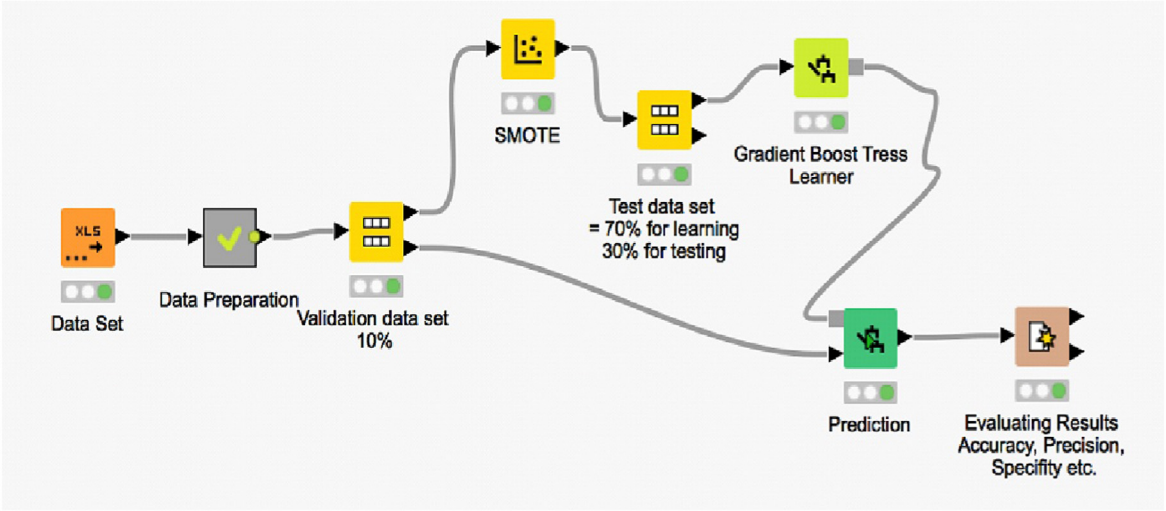

Fig. 3. Depicts the Overall Model used for AI Machine-Learning.

1. Compute *pseudo-residuals* r_{im}

$$r_{im} = -\left[\frac{\partial L\left(y_i, F(x_i)\right)}{\partial F(x_i)}\right]$$

2. Train a base learner to pseudo-residuals, using the training set:

$\{(x_i, y_i)\}\ i = 1\ to\ n$

3. Compute multiplier γ

$$\gamma = argmin \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

4. Update the model:

$$F_m(x_i) = F_{m-1}(x_i) + \gamma_m h_m(x_i)$$

Output: $F_m(x_i)$

2.5. Target variables

1)Severity: described as mild, moderate, and severe; according to the RAC 2)ICU requirement; transferring the patient to the ICU where life support is needed in order to survive if patients have dyspnea(respiratory rate>25/min), hypotension(<90/60 mmHg), gastrointestinal bleeding(>2lt/day), BUN level(>20 mg and progressive increase)(yes/no)(time frame: within a week) 3)Survival:alive or dead(time frame:within a month).

2.6. Statistical analysis

All subjects who met the inclusion criteria were included in the study. Data obtained in the study were analyzed statistically using SPSS version 23.0 software. Descriptive statistics of the obtained data were calculated as mean ± standard deviation(SD), minimum, maximum values, percentiles(25th, median, and 75th), and frequencies(n,%) depending on the type of variables. Multivariate logistic regression analyzes were performed and ROC curves of the prediction and reliability of target variables were obtained. The level of statistical significance was accepted as $p \leq 0.05$.

3. Results

In the study, mild-AP(MAP) group including 57.1%, MSAP including 20.4%, and SAP including 22.5% of the patients. Rates of ICU admission and mortality were 4.6%, and 9.9%. Mean of age was 54.96 ± 17.48(18−101)(46%♀,54%♂). The rates of mild, moderate, and severe Balthazar-CTSI groups were in 69.4%, 24.4%, and 6.2% respectively. Biliary AP was noted as the most common etiology in our study. In fact, although 19% of the patients used alcohol, rate of alcohol-related AP, which is expected to be the second most common cause of AP in the world, was found to be 4.8% in our study, since the etiology of admission was recorded as the primary cause in patients with multiple etiologies(descriptive values for numerical and categorical characteristics of patients are given in Table-1). The accuracy rates of the early estimation of severity, ICU requirement, and survival in the study according to the presence of Balthazar-CTSI or not were 91.02% and 88.20%, 98.25% and 98.25%, 98% and 92.77% respectively(Table-2). Their AUC values were 0.896 and 0.914, 0.885 and 0.859, 0.91 and 0.978 respectively(Figures-5,-6,-7,-8,-9,-10). The variable importance tables for the feature selection procedures of the AI for the severity of AP, ICU requirement and survival are summarized in the Table-3, 4, and 5.

4. Discussion

In the study, we investigated whether early and accurate prediction could be made not only for the severe course of disease but also the need for ICU and survival. In addition to the accuracy of the algorithms and methods used, the contribution of the Balthazar-CTSI taken in 48-h to the accuracy value was also investigated by using demographics, clinical, etiological and laboratory tests that could be easily obtained in the emergency settings. The reason why Balthazar-CTSI is taken at 48-h is that pancreatic necrosis and other local complications do not appear before this time.

Today, the severity of AP is determined according to the RAC scoring. SAP is defined as presence of and required to stay in ICU. The rate of SAP is detected as 22.5% in the study. On the topic of early estimation of severity by ML method, the AUC value was 0.896 with an accuracy 88.2%, sensitivity 86.53% and specificity 93.27%. When Balthazar-CTSI is added, these values were found 0.914, 91.02%, 82.29% and 91.15% respectively(Table-2). Since the local complications such as abscess, pancreatic/peripancreatic necrosis, and vascularly thrombosis in Balthazar-CTSI are causes that can increase severity and mortality, adding Balthazar-CTSI to the evaluations is the expected result to increase the fertility values in the investigated subjects. In predicting the severity, variable significances that were more prominent features than others in the ML phase of the study were creatinine, etiology, age, Hct, BMI, BUN, and weight respectively if Balthazar-CTSI not used(Table-3). If used they were Balthazar-CTSI, creatinine, etiology, age, and Hct respectively. It has been shown that the ANN is increased the accuracy of severity in a review of 43 articles [38]. Zhou Y. et al. reviewed 24 eligible studies from 2913 studies and reported 10 studies estimation of SAP. The severity estimate included 13 models(7 of retrospective) and 3607 patients in these 10 studies(ANN models in 4, classification tree in 3, LR in 2, and SVM in 1 study). Performance of all ML-based studies was better than the single severity score estimation [39]. In the study, in which ML was applied on 139 patients, validated by 69 patients, and 6 of 23 features (pain, creatinine, Hb, ALT, heart rate and WBC) were selected as significant, the ANN method(AUC = 0.92) has been shown to be superior to the LR(AUC = 0.84) and APACHE-II(AUC = 0.63) [40]. In a large-scale study completed with 61,894 patients in the testing and training phase of ML to identify the inpatient SAP cases, the performance of the XGBoost model's AUC value(0.921) was observed to be more successful than the other models [LR(0.780), neoural network-based models(0.811), HAPS(0.533), and BISAP(0.682)] [41]. Pearce CB et al. showed that, APACHE-II and CRP data obtained at admission in determining the severity(AUC = 0.82) is increased the success more than the 24−48-h data(AUC = 0.74) by the ML method performed with Kernel LR [42]. Jin X et al. analyzed the disease severity using multiplayer perception-ANN(MPL-ANN) and partial least squares-discrimination(PLS-DA) methods using 15 blood parameters and 5 serum biochemistry indices on the 167 SAP(88 of them with MSAP), 133 MAP, and 69 healthy controls. As a result, MPL-ANN model was found more successful than PLS-DA with an AUC = 0.984(0.960−1.00), accuracy 93.0%, sensitivity 0,984, and specificity 0,933 [29]. While some of the ML studies predicting the disease severity directly, successes of different ML models were compared with each other, and different success rates were obtained. In these studies, ANN, LR and different ML algorithms were used, most of them were compared with scoring systems and it was observed that they were more successful. We did not use ANN or LR methods in our ML study. Most of the studies that have been done

Table-1
Descriptive values of the patients.

	Mean	SD	Minimum	Maximum	Percentiles		
					25	Median	75
Age (years)	54.96	17.48	18.00	101.00	42.00	55.00	68.00
Height (cm)	165.57	9.18	103.00	192.00	160.00	165.00	171.00
Weight (kg)	79.98	16.00	27.70	193.00	70.00	80.00	90.00
BMI (kg/m^2)	29.15	5,68	14.70	90.09	25.40	28.80	32.00
WBC (μL)	11.60	5.11	1.95	49.90	7.84	10.73	14,11
Hct (%)	38.92	6.70	18.9	68.20	35.30	39.40	43.06
CRP (mg/dL)	11.26	29.31	0.01	338.10	.56	2.11	9.39
BUN (mg/dL)	15.81	9.85	0.74	111.30	9.98	13.55	18.22
Creatinine(mg/dL)	0.96	0.92	0	21.70	0.69	0.80	0.96

		n(%)
Gender	Male	614(46%)
	Female	720(54.%)
BMI (kg/m^2)	<25(Normal weight)	302(22.6%)
	25-30(Overweight)	500(37.5%)
	>30(Obese)	532(39.9%)
Cigarette Smoking	No	1066(79.9%)
	Yes	268(20.1%)
Alcohol Use	No	1080(81%)
	Yes	254(19.0%)
Diabetes Mellitus	Absent	917(68.7%)
	Present	417(31.3%)
Hypertension	Absent	913(68.4%)
	Present	421(31.6%)
Etiology	Biliary	723(54.2%)
	Alcohol	64(4.8%)
	Hypertriglyceridemia	109(8.2%)
	Hypercalcemia	21(1.6%)
	Drug	44(3.3%)
	Congenital	32(2.4%)
	Idiopathic	223(16.7%)
	Post-ERCP	7(0.5%)
	Oddi sphincter dysfunction	12(0.9%)
	Malignancy	18(1.3%)
	Intrapapillary mucinous neoplasia	1(0.1%)
	Primary Sclerosing Cholangitis	15(1.1%)
	Autoimmunity	5(0.4%)
	Multiple etiology	60(4.5%)
Balthazar-CTSI Groups	Mild(0–3)	926(69.4%)
	Moderate [4–6]	326(24.4%)
	Severe [7–10]	82(6.2%)
Atlanta Severity Groups	Mild	762(57.1%)
	Moderate	272(20.4%)
	Severe	300(22.5%)
Survival	Alive	1202(90.1%)
	Died	132(9.9%)
ICU Admission	No	1273(95.4%)
	Yes	61(4.6%)

Table-2
Validation evaluation results for machine-learning during the presence or absence of the Balthazar-CTSI.

	BLTZ-CTSI		Sensitivity (%)	Specificity(%)	PLR	NLR	Disease Prevalance (*)(%)	PPV(*) (%)	NPV(*) (%)	Accuracy(*) (%)
Atlanta Severity	Present	Value	86.53	93.27	12.85	0.14	33.33	86.53	93.27	91.02
		95%CI	82.8–89.72	91.31–94.9	9.90–16.68	0.11–0.19	30.67–36.08	83.20–89.29	91.5–94.67	89.2–92.58
	Abscent	Value	82.29	91.15	9.3	0.19	33.33	82.29	91.15	88.20
		95%CI	78.2–85.90	88.96–93.02	7.41–11.66	0.16–0.24	30.67–36.08	78.75–85.36	89.28–92.72	86.24–89.97
ICU Requirement	Present	Value	99.22	77.78	4.46	0.01	95.51	98.96	82.35	98.25
		95%CI	97.7–99.84	52.36–93.59	1.88–10.60	0.00–0.03	93.00–97.32	97.56–99.56	59.55–93.67	96.44–99.30
	Abscent	Value	99.48	72.22	3.58	0.01	95.51	98.70	86.67	98.25
		95%CI	98.1–99.94	46.52–90.31	1.70–7.54	0.00–0.03	93.00–97.32	97.31–99.38	61.30–96.39	96.44–99.30
Survival	Present	Value	84.78	99.72	300.98	0.15	11.47	97.50	98.06	98.00
		95%CI	71.1–93.66	98.44–99.99	42.35–2139.0	0.08–0.30	8.52–15.00	84.59–99.64	96.24–99.01	96.11–99.13
	Abscent	Value	99.70	58.21	2.39	0.01	83.29	92.24	97.50	92.77
		95%CI	98.3–99.99	45.52–70.15	1.80–3.16	0.00–0.04	79.27–86.81	89.96–94.04	84.50–99.64	89.78–95.10

(*) These values are dependent on disease prevalance. BLTZ-CTSI: Balthazar-CTSI, Dis. Preval: disease prevalance, ICU: intensive care unit, PLR: Positive Likelihood Ratio, NLR: Negative Likelihood Ratio, PPV: Positive Predictive Value, NPV: Negative Predictive Value.

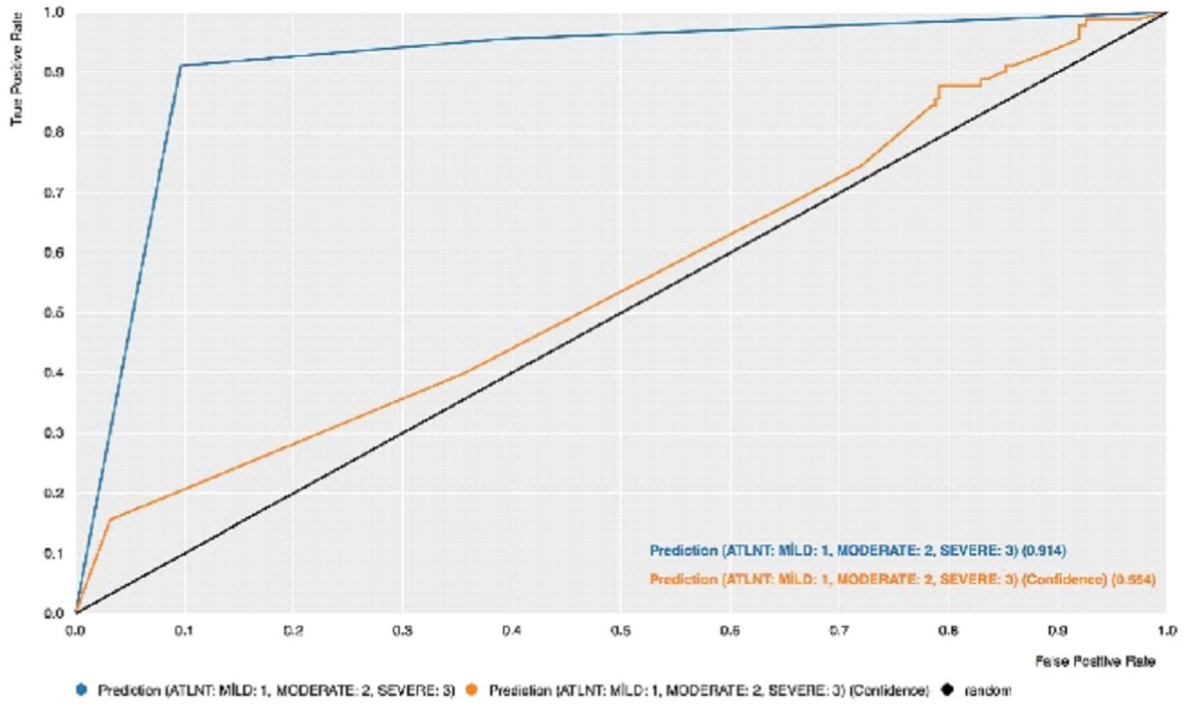

Fig. 5. Prediction of the Atlanta severity with Baltazar-CTSI.

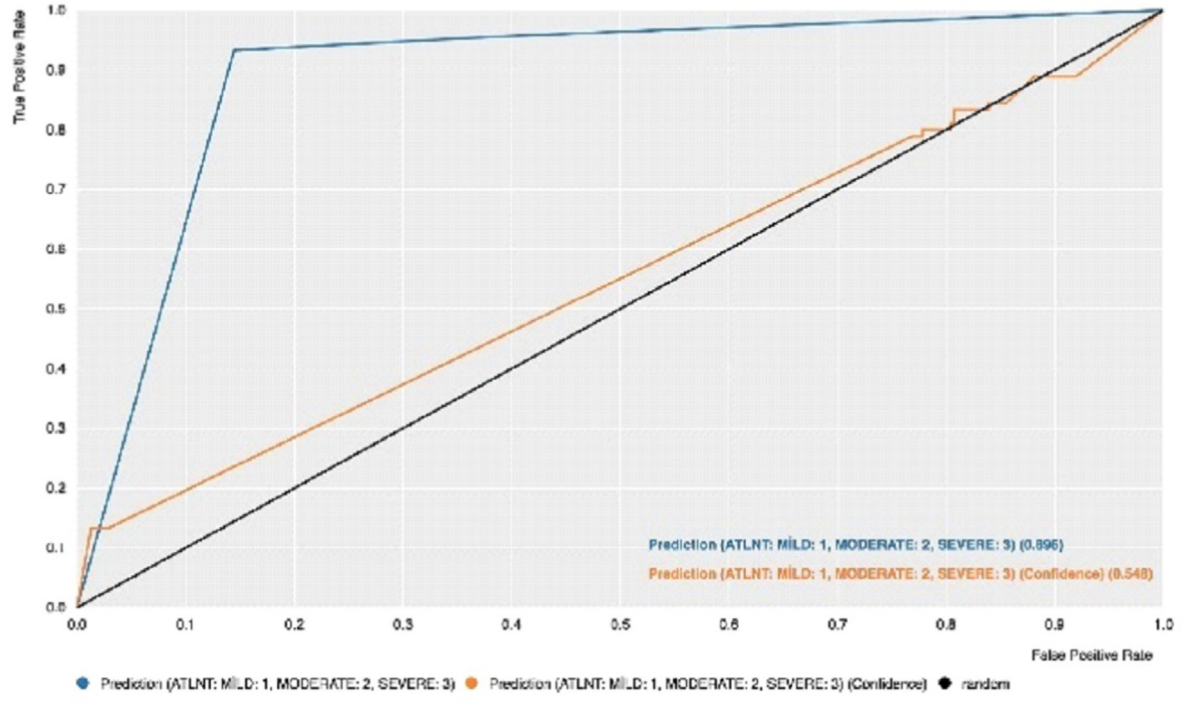

Fig. 6. Prediction of the Atlanta severity without Baltazar-CTSI.

are small-scale studies. Among them only few are large-scale studies, and the success of our results are quite close to the results of this studies by the help of SMOTE and variable significances. In another large-scale EASY-APP study, which 1184 patients were validated with 3543 cohorts, the Scikit-learn, XG Boost, and Catboost Python packages were used for machine learning predictions. ROC, AUROC curves and accuracy values were determined by evaluating the models with 4-fold cross validation. Eventually, the XG Boost model was found to be successful by an AUC score of 0.81 ± 0.033 and an accuracy of 89.1% [43] When compared with our study, instead of XG Boost, the Gradient Boosting algorithm trained the machine, and the values were evaluated with 5-fold cross validation instead of 4-fold cross-validation. The AUC score in our study was 0.89.6 and with an 88.2% accuracy rate and when

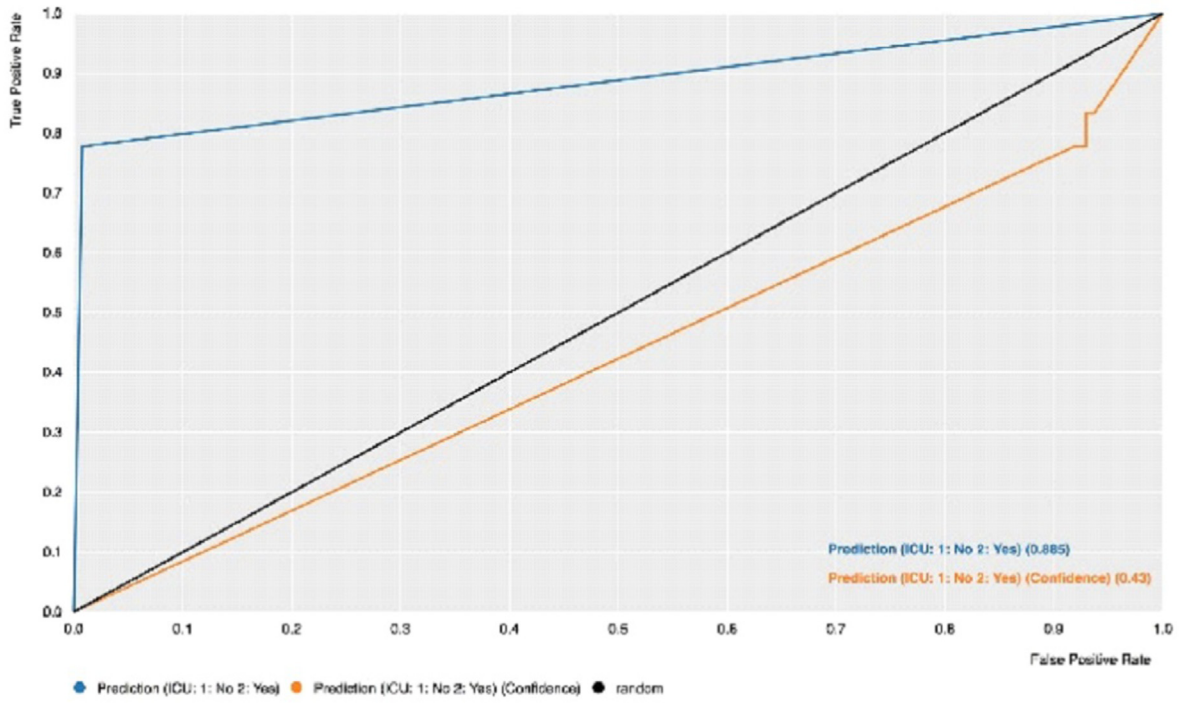

Fig. 7. Prediction of the intensive care requirement with Balthazar-CTSI.

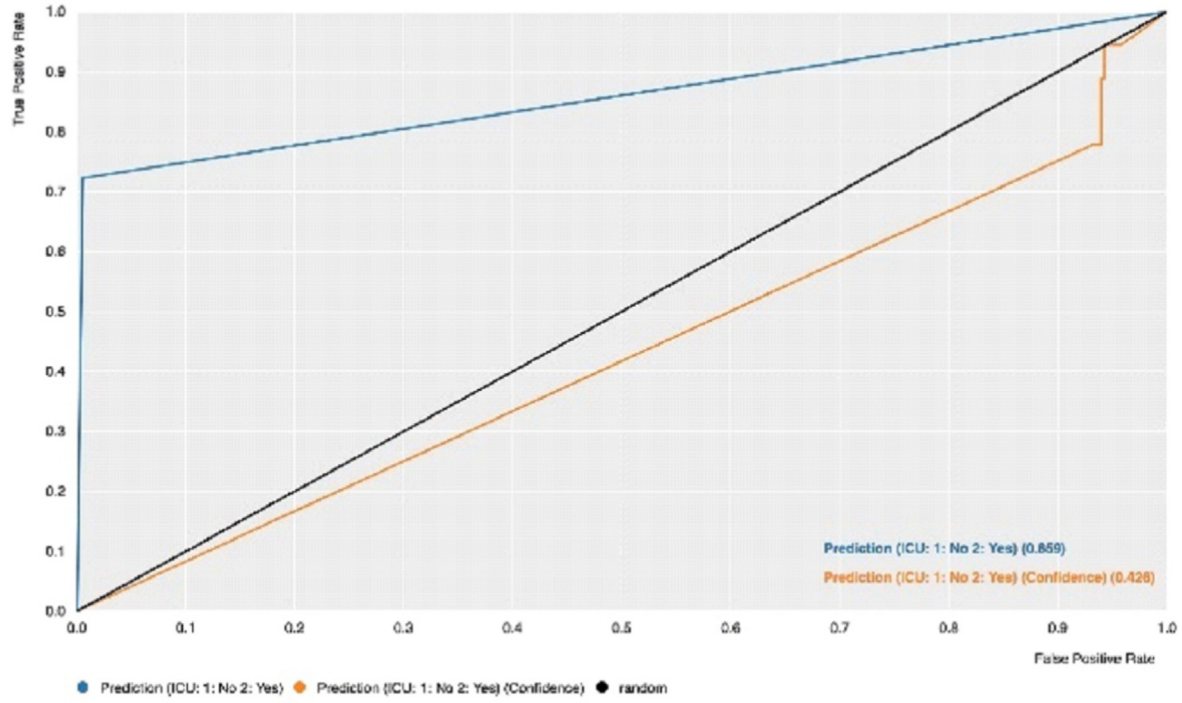

Fig. 8. Prediction of the intensive care requirement without Balthazar-CTSI.

the Balthazar-CTSI score used, the accuracy rate was increased to 91.02%. Gradient Boost and XGBoost work on the same principle. The differences between them are in the details. XGBoost shows higher prediction success using different techniques and is optimized to work on large datasets.

ICU admission rate of in our study was 4.6%. By ML method we applied, AUC value was 0.885 with an accuracy rate 98.25%, sensitivity 99.48%, and specificity 72.22% during validation evaluation by ML. If Balthazar-CTSI feature used, these were 0.859, 92.8%, 99.22%, and 77.78% respectively(Table-2). The fact that the accuracy value of the ICU requirement is lower in the presence of 48-h Balthazar-CTSI than in its absence is due to the Balthazar-CTSI being able to evaluate only the pancreas and its surroundings and not detecting OF. The prominent variable importances were weight,

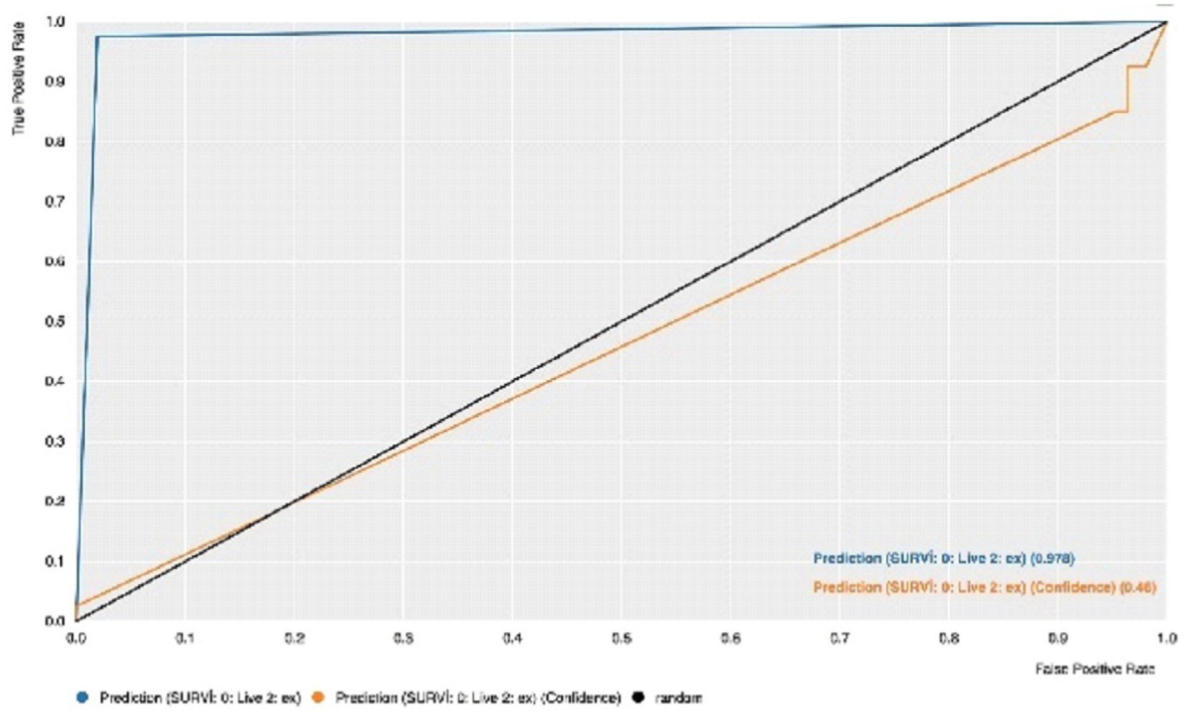

Fig. 9. Prediction of the survival with Balthazar-CTSI.

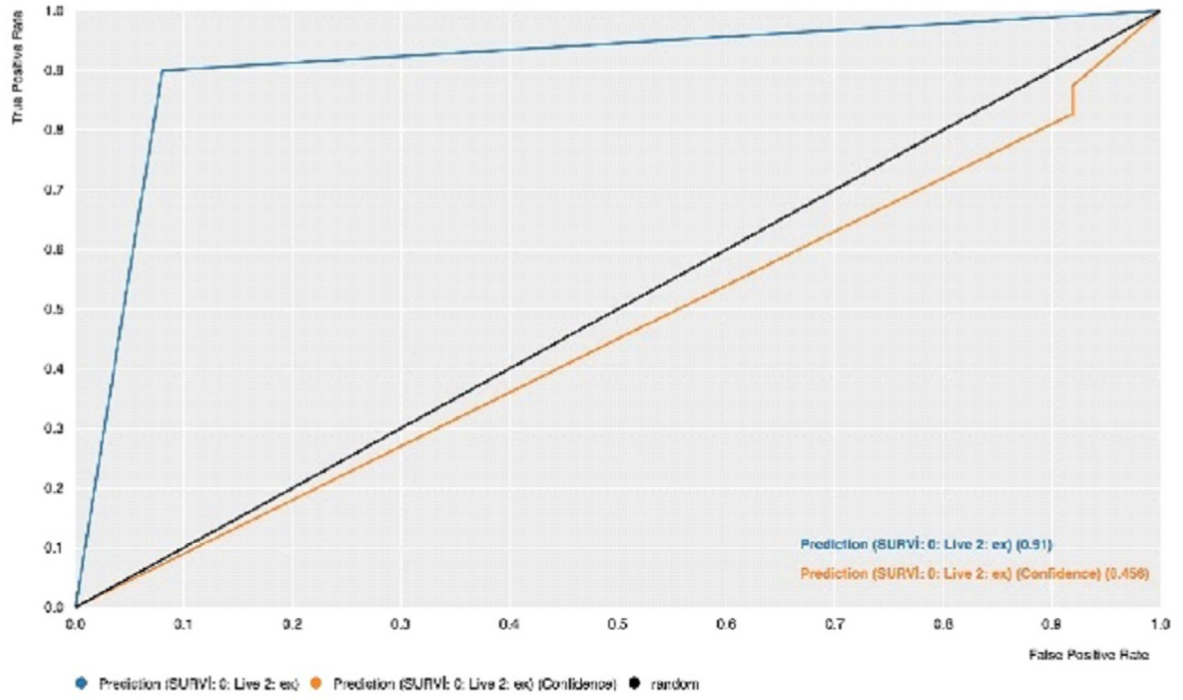

Fig. 10. Prediction of the survival without Balthazar-CTSI.

Balthazar-CTSI, age, etiology, and height respectively when Balthazar-CTSI used and creatinine, height, age, WBC, and BUN were respectively when not used(Table-4). In Fei Y. et al.'s study, ANN and LR were compared in 217 patients at risk of developing acute respiratory distress syndrome(ARDS) after SAP, and while the AUC and accuracy of ANN were found to be 0.859 ± 0.048 and 84.4%, these values were 0.701 ± 0.041 and 71.9% with LR [44]. Qiu Q. et al. used 5-fold cross-validation to train SVM, LR, and ANN models, using blood volume, inflammatory, coagulation, and renal functions in 16 features by univariate analysis for estimation of multiple organ failure(MOF) in 263 patients admitted with MSAP and SAP. The AUC of the models were 0.840, 0.832, and 0.834

Table-3
Variable importance when learning Atlanta severity status.

	Using Balthazar-CTSI				Not using Balthazar-CTSI		
	Relative Importance	Scaled Importance	%		Relative Importance	Scaled Importance	%
Balthazar	300.46	1.00	0.30	**CREA**	157.613113	1	0.16251295
CREA	118.20	0.39	0.12	**Etiology**	109.27	0.69	0.11
Etiology	93.47	0.31	0.09	**Age**	103.96	0.66	0.11
Age	86.31	0.29	0.09	**Hct**	99.95	0.63	0.10
Hct	83.45	0.28	0.08	**BMI**	94.78	0.60	0.10
BUN	56.84	0.19	0.06	**BUN**	88.10	0.56	0.09
WBC	51.06	0.17	0.05	**Weight**	84.67	0.54	0.09
CRP	48.55	0.16	0.05	**WBC**	76.06	0.48	0.08
BMI	44.67	0.15	0.04	**Height**	70.90	0.45	0.07
Height	42.63	0.14	0.04	**CRP**	36.06	0.23	0.04
Weight	37.99	0.13	0.04	**Gender**	13.02	0.08	0.01
Alcohol	7.51	0.03	0.01	**BMI**	12.98	0.08	0.01
DM	6.71	0.02	0.01	**DM**	10.27	0.07	0.01
BMI	5.73	0.02	0.01	**HT**	5.91	0.04	0.01
HT	4.59	0.02	0.00	**Smoking**	3.93	0.02	0.00
Smoking	3.83	0.01	0.00	**Alcohol**	2.37	0.02	0.00
Gender	2.85	0.01	0.00				

CRP: C-reactive protein, BMI: Body Mass Index, Hct: Hematocrite, WBC: White Blood Cell, BUN: Blood urea nitrogene, DM: Diabetes Mellitus, HT: Hypertension, CREA: creatinine.

Table-4
Variable importance when learning intensive care unite admission status.

	Using Balthazar-CTSI				Not Using Balthazar-CTSI		
	Relative Importance	Scaled Importance	%		Relative Importance	Scaled Importance	%
Weight	8.58	1.00	0.19	**CREA**	8.12	1.00	0.19
Balthazar	7.88	0.92	0.17	**Height**	5.89	0.73	0.14
Age	7.41	0.86	0.16	**Age**	4.87	0.60	0.12
Etiology	3.88	0.45	0.09	**WBC**	4.72	0.58	0.11
Height	3.08	0.36	0.07	**BUN**	4.66	0.57	0.11
CRP	2.30	0.27	0.05	**Hct**	4.16	0.51	0.10
BMI	2.26	0.26	0.05	**Etiology**	2.68	0.33	0.06
BMI	2.21	0.26	0.05	**Weight**	1.99	0.25	0.05
Hct	2.00	0.23	0.04	**BMI**	1.62	0.20	0.04
CREA	1.38	0.16	0.03	**BMI**	0.91	0.11	0.02
WBC	1.31	0.15	0.03	**Gender**	0.81	0.10	0.02
BUN	1.27	0.15	0.03	**Alcohol**	0.62	0.08	0.01
Gender	0.94	0.11	0.02	**CRP**	0.51	0.06	0.01
DM	0.65	0.08	0.01	**DM**	0.21	0.03	0.01
Alcohol	0.26	0.03	0.01	**Smoking**	0.06	0.01	0.00
Smoking	0.20	0.02	0.00	**HT**	0.00	0.00	0.00
HT	0.03	0.00	0.00				

CRP: C-reactive protein, BMI: Body Mass Index, Hct: Hematocrite, WBC: White Blood Cell, BUN: Blood urea nitrogene, DM: Diabetes Mellitus, HT: Hypertension, CREA: creatinine.

respectively while AUC of the APACHE-II was 0.814. $p > 0.05$. The accuracies of the models were 79.85%, 77.95%, 71.10%, and 69.58% respectively. The common important predictive factors were Hct, K-time, IL-6, and creatinine in three models [30]. In the back-propagation study of Hong W-D et al. using the three-layer multi-player perceptron ANN model, permanent OF was investigated by 13 variables in 312 patients. SIRS and APACHE-II scores were calculated. AUC is used to demonstrate the performance of ANN, LR, and APACHE-II. The SIRS, BUN, and serum calcium were determined as 3 independent variables by MLR(multivariate logistic regression). Age, Hct, BUN, serum calcium and glucose were found to be the most important predictors with ANN analysis. The sensitivity, specificity, positive likelihood ratio(PLR), negative likelihood ratio(NLR), and diagnostic accuracy in the ANN model were found as 81.3%, 98.9%, 71.5, 0.19, and 96.2%, respectively. The AUC of ANN model(0.96 ± 0.02) was statistically higher than the LR(0.88 ± 0.03) and APACHE-II models(0.83 ± 0.03). The sensitivity and specificity, PLR, NLR and accuracy of the ANN model found as 81.3%, 98.9%, 71.5, 0.19, and 96.2% [45]. In another study the prediction models were developed, with features selected by 6 ML methods by scanning with univariate analysis, and they were validated with 5-fold cross-validation and the prediction performance of the methods was evaluated. In the training and validation set for modelling, interleukin-6, creatinine levels and kinetic time were found to be significant among 18 features that showed significant differences between groups with and without MOF. MOF developed in 305 out of 455 SAP or MSAP patients. The adaptive boosting algorithm's(AdaBoost) predictive performance in validation were found with 80.2% sensitivity, 80.4% specificity, and 80.4% accuracy($AUC = 0.863$). The sensitivity of the AdaBoost model and specificity of the LR models were the best scores among the 6 models in the test set [46]. When compared with the results of the studies mentioned above, it is seen that the accuracy result obtained by ML method are much more successful than others in the early estimation of ICU requirement.

In our study, AUC value was 0.91 with an accuracy rate 92.8%, sensitivity 99.70%, and specificity 58.21% during validation evaluation. If Balthazar-CTSI feature used, these values were 0.978, 98%,

Table-5
Variable importance when learning survival status.

	Using Balthazar-CTSI				Not Using Balthazar-CTSI		
	Relative Importance	Scaled Importance	%		Relative Importance	Scaled Importance	%
Age	24.32	1.00	0.19	**Age**	38.62	1.00	0.30
Height	19.27	0.79	0.15	**WBC**	15.51	0.40	0.12
Balthazar	17.49	0.72	0.13	**Etiology**	11.28	0.29	0.09
BUN	15.53	0.64	0.12	**Weight**	9.78	0.25	0.08
Etiology	14.79	0.61	0.11	**Hct**	8.15	0.21	0.06
CREA	9.11	0.37	0.07	**BUN**	7.78	0.20	0.06
WBC	7.48	0.31	0.06	**BMI**	7.58	0.20	0.06
Weight	6.35	0.26	0.05	**CREA**	6.87	0.18	0.05
Hct	5.93	0.24	0.05	**Gender**	5.83	0.15	0.05
BMI	3.72	0.15	0.03	**CRP**	4.79	0.12	0.04
CRP	2.41	0.10	0.02	**Height**	4.04	0.10	0.03
Alcohol	2.19	0.09	0.02	**Alcohol**	3.56	0.09	0.03
Smoking	0.94	0.04	0.01	**BMİ**	1.39	0.04	0.01
BMİ	0.47	0.02	0.00	**Smoking**	1.03	0.03	0.01
Gender	0.42	0.02	0.00	**HT**	0.73	0.02	0.01
HT	0.10	0.00	0.00	**DM**	0.38	0.01	0.00
DM	0.02	0.00	0.00				

CRP: C-reactive protein, BMI: Body Mass Index, Hct: Hematocrite, WBC: White Blood Cell, BUN: Blood urea nitrogene, DM: Diabetes Mellitus, HT: Hypertension, CREA: creatinine.

84.78%, and 99.72% respectively(Table-2). Selected variable importances were the age, height, Balthazar-CTSI, and BUN respectively if Balthazar-CTSI used. If it used, they were age, WBC, etiologies, weight, and Hct respectively(Table-5). The Balthazar-CTSI [39], creatinine [40,42,46], age [42,43,48], Hct [41,48], BUN [48], WBC [29,39,42,48], CRP [42], BMI [41], abdominal pain [40,41], LDH [39], glucose [43], and respiratory rate [42,43] were found as important ML features related to AP severity. In a study, the overall mortality rate has been found as 2.83% and it was higher in patients with SAP(28.3%; $p < 0.001$) versus MSAP(0.6%) and MAP(0.3%) [47]. Moidifi et al.'s study with 664 patients for the early prediction of mortality and SAP by comparing the accuracy values of ANN, APACHE-II, and Glasgow severity(GS) scores found that ANN was much more successful. While the accuracy values of ANN, APACHE-II and GS were found to be 92.5%, 82.6% and 72.4%, respectively, in estimating disease severity, they were 97.5%, 82.4% and 80.7% in mortality estimation [48]. Another study is conducted on the early estimation of in-hospital mortality in AP patients(n:337) using three-layer ANN back propagation method. In the MLR analysis using 12 variables in the deceased and surviving groups, ALT, WBC, and calcium were determined as independent variables, and the AUC value and accuracy of the ANN model was found to be 0.769 and 0.662 what found superior to the LR, Ranson, and SOFA score [28]. The accuracy values of our study's were found more successful than both studies either using with or without Balthazar-CTSI.

ROC curves have been calculated for confidence and predictions in the study; prediction confidences for Atlanta severity and ICU requirement yield AUC values higher than 50%. This means, in overall, prediction confidences are good enough although they are not great. On the other hand, AUC values for survival prediction confidences(with or without using Balthazar-CTSI) are lower than 50%. In this case, overall confidence of survive predictions are low. Surprisingly enough, using Balthazar-CTSI does not affect the AUC values very much. As we see from the ROC curves and AUC for predictions, AUC values are higher than 85%, especially ICU prediction using Balthazar-CTSI is 97.8%(they are very close to 100%). While prediction ROC shows the machine's predictive power unbiasedly, confidence ROC shows the machine's own confidence in its predictions. Example: If the machine makes a 100% correct guess, but the confidence ROC is 90%, then we say that it knows everything, but it is not fully confident. On the other hand, we say that we cannot guess well, but we think we know. Machine predicts with a high confidence in most cases. However, this does not mean that the prediction power is low, but machine self-confidence is low. Suffice it to say, continuing with reinforcement ML or transfer learning this weakness may be eliminated in the future. For the time being, in clinical usage a confidence level over %50 is high enough to rely on this model's predictions.

5. Conclusion

By the ML method we applied, we were able to make more successful predictions about the course of AP, whether the patient will require ICU or not, and survival status by using 13 simple features including some lab, clinical and demographic parameters those easily obtained during the admission. Besides, it is an expected situation that the success of the estimations will increase with the addition of Balthazar-CTSI, which is quoted 48-h, to the evaluation.

Author contributions

ATI: Conception, design, data collection, data analysis, and interpretation. GSilahtaroğlu: Artificial intelligence, machine learning, data analysis and interpretation. GSeven: interpretation and review of the manuscript. KK, KY: data collection and manuscript writing, HŞ: reviewed and revised the article.

Declaration of competing interest

The authors have no conflict of interests to declare.

References

[1] Schepers NJ, Bakker OJ, Besselink MG, Ahmed Ali U, Bollen TL, Gooszen HG, et al. Dutch Pancreatitis Study Group. Impact of characteristics of organ failure and infected necrosis on mortality in necrotising pancreatitis. Gut 2019;68:1044–51. https://doi.org/10.1136/gutjnl-2017-314657.
[2] Frey CF, Zhou H, Harvey DJ, White RH. The incidence and case-fatality rates of acute biliary, alcoholic, and idiopathic pancreatitis in California. Pancreas 2006;33:336–44. https://doi.org/10.1097/01.mpa.0000236727.16370.99. 1994-2001.
[3] Papachristou GI, Papachristou DJ, Morinville VD, Slivka A, Whitcomb DC. Chronic alcohol consumption is a major risk factor for pancreatic necrosis in acute pancreatitis. Am J Gastroenterol 2006;101:2605–10. https://doi.org/10.1111/j.1572-0241.2006.00795.x.
[4] Martínez J, Johnson CD, Sánchez PayáJ, de Madaria E, Robles-Díaz G, Pérez-

Mateo M. Obesity is a definitive risk factor for severity and mortality in acute pancreatitis: an updated meta-analysis. Pancreatology 2006;6:206–9. https://doi.org/10.1159/000092104.
[5] Lankisch PG, Blum T, Maisonneuve P, Lowenfels AB. Severe acute pancreatitis: when to be concerned? Pancreatology 2003;3:102–10. https://doi.org/10.1159/000070077.
[6] Wu BU, Johannes RS, Sun X, Conwell DL, Banks PA. Early changes in blood urea nitrogen predict mortality in acute pancreatitis. Gastroenterology 2009;137:129–35. https://doi.org/10.1053/j.gastro.2009.03.056.
[7] Balthazar EJ. Acute pancreatitis: assessment of severity with clinical and CT evaluation. Radiology 2002;223:603–13. https://doi.org/10.1148/radiol.2233010680.
[8] Farkas N. A multicenter, international cohort analysis of 1435 cases to support clinical trial design in acute pancreatitis. Front Physiol 2019;10:1092. https://doi.org/10.3389/fphys.2019.01092.
[9] Larvin M. Assessment of clinical severity and prognosis. In: Beger HG, Warshaw AL, Buchler MW, et al., editors. The pancreas. Oxford: Blackwell Science; 1998. p. 489.
[10] Banks PA. Freeman ML, practice parameters committee of the American college of Gastroenterology. Practice guidelines in acute pancreatitis. Am J Gastroenterol 2006;101:2379–400. https://doi.org/10.1111/j.1572-0241.2006.00856.x.
[11] Wu BU, Johannes RS, Sun X, Tabak Y, Conwell DL, Banks PA. The early prediction of mortality in acute pancreatitis: a large population-based study. Gut 2008;57:1698–703. https://doi.org/10.1136/gut.2008.152702.
[12] Lankisch PG, Weber-Dany B, Hebel K, Maisonneuve P, Lowenfels AB. The harmless acute pancreatitis scoe: a clinical algorithm for rapid initial stratification of nonsevere disease. Clin Gastroenterol Hepatol 2009;7:702–5. https://doi.org/10.1016/j.cgh.2009.02.020. quiz 607.
[13] Bradley III EL, Allen K. A prospective longitudinal study of observation versus surgical intervention in the management of necrotizing pancreatitis. Am J Surg 1991;161:19–24. https://doi.org/10.1016/0002-9610(91)90355-h.
[14] Fedorak IJ, Ko TC, Djuricin G, McMahon M, Thompson K, Prinz RA, et al. Secondary pancreatic infections: are they distinct clinical entities? Surgery 1992;112:824–30.
[15] Isenmann R, Rau B, Beger HG. Bacterial infection and extent of necrosis are determinants of organ failure in patients with acute necrotizing pancreatitis. Br J Surg 1999;86:1020–4. https://doi.org/10.1046/j.1365-2168.1999.01176.x.
[16] Petrov MS, Shanbhag S, Chakraborty M, Phillips AR, Windsor JA. Organ failure and infection of pancreatic necrosis as determinants of mortality in patients with acute pancreatitis. Gastroenterology 2010;139:813–20. https://doi.org/10.1053/j.gastro.2010.06.010.
[17] Garg PK, Khanna S, Bohidar NP, Kapil A, Tandon RK. Incidence, spectrum and antibiotic sensitivity pattern of bacterial infections among patients with acute pancreatitis. J Gastroenterol Hepatol 2001;16:1055–9. https://doi.org/10.1046/j.1440-1746.2001.02589.x.
[18] Wig JD, Bharathy KG, Kochhar R, Yadav TD, Kudari AK, Doley RP. Correlates of organ failure in severe acute pancreatitis. JOP 2009;10:271–5.
[19] Windsor JA, Petrov MS. Acute pancreatitis reclassified. Gut 2013;62:4–5. https://doi.org/10.1136/gutjnl-2012-303725.
[20] Hirasawa T, Aoyama K, Tanimoto T, Ishihara S, Shichijo S, Ozawa T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. Gastric Cancer 2018;21:653–60. https://doi.org/10.1007/s10120-018-0793-2.
[21] Lee JH, Kim YJ, Kim YW, Park S, Choi Y-I, Kim YJ, et al. Spotting malignancies from gastric endoscopic images using deep learning. Surg Endosc 2019;33:3790–7. https://doi.org/10.1007/s00464-019-06677-2.
[22] Zhavoronkov A, Vanhaelen Q, Oprea TI. Will Artificial Intelligence for Drug Discovery Impact Clinical Pharmacology? Clin Pharmacol Ther 2020;107:780–5. https://doi.org/10.1002/cpt.1795.
[23] Koromina M, Pandi MT, Patrinos GP. Rethinking drug repositioning and development with artificial intelligence, machine learning, and omics. OMICS 2019;23:539–48. https://doi.org/10.1089/omi.2019.0151.
[24] Kumar P, Sinha R, Shukla P. Artificial intelligence and synthetic biology approaches for human gut microbiome. Crit Rev Food Sci Nutr 2022;63:2103–21. https://doi.org/10.1080/10408398.2020.1850415.
[25] Orr RK. Use of an artificial neural network to quantitate risk of malignancy for abnormal mammograms. Surgery 2001;129:459–66. https://doi.org/10.1067/msy.2001.112069.
[26] Waljee AK, Liu B, Sauder K, J Zhu J, Govani SM, Stidham RW, et al. Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis. Aliment Pharmacol Ther 2018;47:763–72. https://doi.org/10.1111/apt.14510.
[27] Clermont G, Angus DC, DiRusso SM, Griffin M, Linde-Zwirble WT. Predicting hospital mortality for patients in the intensive care unit: a comparison of artificial neural networks with logistic regression models. Crit Care Med 2001;29:291–6. https://doi.org/10.1097/00003246-200102000-00012.
[28] Ding N, Guo C, Li C, Zhou Y, Chai X. An artificial neural networks model for early predicting in-hospital mortality in acute pancreatitis in MIMIC-III. Bio-Med Res Int 2021;2021:6638919. https://doi.org/10.1155/2021/6638919.
[29] Jin X, Ding Z, Li T, Xiong J, Tian G, Liu J. Comparison of MPL-ANN, and PLS-DA models for predicting the severity of patients with acute pancreatitis: an exploratory study. Am J Emerg Med 2021;44:85–91. https://doi.org/10.1016/j.ajem.2021.01.044.
[30] Qiu Q, Nian YJ, Guo Y, Tang L, Lu N, Wen LZ, et al. BMC Development and validation of three machine-learning models for predicting multiple organ failure in moderately severe and severe acute pancreatitis. BMC Gastroenterol 2019;19:118. https://doi.org/10.1186/s12876-019-1016-y.
[31] Marshall JC, Cook DJ, Christou NV, Bernard GR, Sprung CL, Sibbald WJ. Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome. Crit Care Med 1995;23:1638–52. https://doi.org/10.1097/00003246-199510000-00007.
[32] Clustering A. Clustering categorical data using hierarchies. Eng Technol 2009;1(2):334–9. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.308.1345&rep=rep1&type=pdf.
[33] Silahtaroğlu G. An attribute-centre based decision tree classification algorithm. Engineering andTechnology 2009:302. – 6, https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.193.6458&rep=rep1&type=pdf.
[34] Friedman JH. Greedy Function Approximation: a gradient boosting machine. Ann Stat 2001;29:1189–232. https://doi.org/10.1214/aos/1013203451.
[35] Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. Artif Intell Rev 2021;54:1937–67. https://doi.org/10.1007/s10462-020-09896-5.
[36] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over sampling technique. J Artif Intell Res 2002;16:321–57. https://doi.org/10.1613/jair.953.
[37] Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, et al. KNIME—the konstanz information miner. ACM SIGKDD Explorations Newsletter 2009;11(1):26–31. https://doi.org/10.1145/1656274.1656280.
[38] van den Heever M, Mittal A, Haydock M, Windsor J. The use of intelligent database systems in acute pancreatitis–a systematic review. Pancreatology 2014;14:9–16. https://doi.org/10.1016/j.pan.2013.11.010.
[39] Zhou Y, Ge YT, Shi XL, Wu KY, Chen WW, Ding YB, et al. Machine learning predictive models for acute pancreatitis: a systematic review. Int J Med Inf 2022;157:104641. https://doi.org/10.1016/j.ijmedinf.2021.104641.
[40] Andersson B, Andersson R, Ohlsson M, Nilsson J. Prediction of severe acute pancreatitis at admission to hospital using artificial neural networks. Pancreatology 2011;11:328–35. https://doi.org/10.1159/000327903.
[41] Thapa R, Iqbal Z, Garikipati A, Siefkas A, Hoffman J, Mao Q, et al. Early prediction of severe acute pancreatitis using machine learning. Pancreatology 2022;22:43–50. https://doi.org/10.1016/j.pan.2021.10.003.
[42] Pearce CB, Gunn SR, Ahmed A, Johnson CD. Machine learning can improve prediction of severity in acute pancreatitis using admission values of Apache II score and C-reactive protein. Pancreatology 2006;6:123–31. https://doi.org/10.1159/000090032.
[43] Kui B, Pintér J, Molontay R, Nagy M, Farkas N, Gede N, et al. EASY-APP: an artificial intelligence model and application for early and easy prediction of severity in acute pancreatitis. Hungarian Pancreatic Study Group. Clin Transl Med 2022;12:e842. https://doi.org/10.1002/ctm2.842.
[44] Fei Y, Gao k, Li WQ. Prediction and evaluation of the severity of acute respiratory distress syndrome following severe acute pancreatitis using an artificial neural network algorithm model. HPB 2019;21:891–7. https://doi.org/10.1016/j.hpb.2018.11.009.
[45] Hong W-D, Chen X-R, Jin S-Q, Huang Q-K, Zhu Q-H, Pan J-Y. Use of an artificial neural network to predict persistent organ failure in patients with acute pancreatitis. Clinics 2013;68:27–31. https://doi.org/10.6061/clinics/2013(01)rc01.
[46] Xu F, Chen X, Li C, Liu J, Qiu Q, He M, et al. Prediction of multiple organ failure complicated by moderately severe or severe acute pancreatitis based on machine learning: a multicenter cohort study. Mediat Inflamm 2021;2021:525118. https://doi.org/10.1155/2021/5525118.
[47] Párniczky A, Kui B, Szentesi A, Balázs A, Szűcs A, Mosztbacher D. Prospective, multicentre, nationwide clinical data from 600 cases of acute pancreatitis. PLoS One 2016;11(10):e0165309. https://doi.org/10.1371/journal.pone.0165309.
[48] Mofidi R, Duff MD, Madhavan KK, Garden OJ, Parks RW. Identification of severe acute pancreatitis using an artificial neural network. Surgery 2007;141:59–66. https://doi.org/10.1016/j.surg.2006.07.022.