ORIGINAL PAPER

DOI: http://dx.doi.org/10.15403/jgld-5876

Machine Learning Models predicting Decompensation in Cirrhosis

Sophie Elisabeth Müller[1,2], Markus Casper[1], Cristina Ripoll[3,4], Alexander Zipprich[3,4], Paul Horn[3,5], Marcin Krawczyk[1], Frank Lammert[1,6], Matthias Christian Reichert[1]

1) Department of Medicine II, Saarland University Medical Center, Saarland University, Homburg;
2) Institute of Medical Microbiology and Hygiene, Center for Infectious Diseases, Saarland University, Homburg;
3) Department of Internal Medicine IV, University Hospital Jena, Jena;
4) Clinic for Internal Medicine I, University Hospital Halle, Halle;
5) Department of Hepatology and Gastroenterology, Charité – Universitätsmedizin Berlin, Berlin;
6) Health Sciences, Hannover Medical School (MHH), Hannover, Germany

Address for correspondence:
Sophie Elisabeth Müller
Department of Medicine II, Saarland University Medical Center
Kirrberger Straße 100, 66421 Homburg, Germany.
sophieelisabeth.mueller@uks.eu

Received: 04.09.2024
Accepted: 13.01.2025

ABSTRACT

Background & Aims: Decompensation of cirrhosis significantly decreases survival, thus, prevention of complications is paramount. We used machine learning techniques to identify parameters predicting decompensation.

Methods: Several machine learning techniques were applied to the INCA trial database containing pro- and retrospective data from 983 patients. Laboratory, clinical, and genetic data were analysed. After performing hierarchical clustering, Permutation Feature Importance was used to evaluate the impact of parameters on the prediction of decompensation.

Results: Achieving an accuracy of 81.6% on training and 70.5% on test data, Random Forests were best for retrospective prediction. In prospective assessment, Support Vector Machines performed best with an accuracy of 78.6% and 73.8%, respectively. Permutation Feature Importance demonstrated that baseline albumin and bilirubin levels and maximum bilirubin were the highest ranked parameters associated with former decompensation. In the prospective analysis, the maximum bilirubin value and the baseline values of sodium and albumin were ranked highest. In addition to the parameters of established scores, *NOD2* genotype and inflammatory markers were highly ranked.

Conclusions: Laboratory parameters, genetic variants and infections can help to predict the risk of cirrhosis decompensation. This proof-of-concept study adds data for the future development of advanced models to identify patients at risk.

Key words: artificial intelligence – cirrhosis – decompensation – liver - machine learning.

Abbreviations: ALBI: albumin-bilirubin; ALT: alanine aminotransferase; AST: asparatate aminotransferase; CRP: C-reactive protein; CSPH: clinically significant portal hypertension; DT: decision tree; EPOD: early prediction of decompensation; HVPG: hepatic venous pressure gradient; I*N*CA-trial: impact of *NOD2* genotype-guided antibiotic prevention on survival in patients with liver cirrhosis and ascites; Fib-4: fibrosis-4; HCC: hepatocellular carcinoma; HE: hepatic encephalopathy; LSM: liver stiffness measurement; MELD: model of end-stage liver disease; NOD2: nucleotide-binding oligomerization domain-containing protein 2; NN: neural network; PFI: permutation feature importance; RF: random forest; ROC: receiver operating curve; SVM: support vector machines.

INTRODUCTION

Cirrhosis is the common end-stage of chronic liver diseases. Chronic liver injury causes structural changes such as fibrosis with increased intrahepatic resistance. This together with raised splanchnic inflow leads to portal hypertension. Initially, most patients are in the compensated stage with long survival, but with the development of complications of portal hypertension or liver dysfunction the disease rapidly progresses to the decompensated stage with significantly increased mortality (median survival of two years compared to more than 12 years in the compensated stage [1]). Individual risk stratification is necessary to identify patients requiring intervention and closer monitoring, calling for the need to identify risk factors and markers for decompensation. In contrast to risk factors, risk markers are not causally associated with the disease, i.e., a direct change in risk markers does not necessarily influence the probability of developing the disease.

Although several scores are available and commonly used in clinical practice to assess the prognosis of the patients, the parameters predicting a decompensation still have not been fully defined. The Child-Pugh [2] and Model of End-Stage Liver Disease (MELD) scores [3, 4] have been used for a long time in clinical routine, but they only provide information about the short-term prognosis [5-7]. The albumin-bilirubin (ALBI) score, which combines serum albumin and bilirubin concentrations, has been proposed for the assessment of prognosis of patients with hepatocellular carcinoma (HCC) [8–11] or chronic liver diseases without HCC [12-14]. The Fibrosis-4 (Fib-4) index is a simple non-invasive score that estimates the extent of liver fibrosis based on age, alanine aminotransferase (ALT) and aspartate aminotransferase (AST) activities, and platelet count [15]. Guha et al. [16] combined the ALBI score with the Fib-4 index to develop a model that identifies those patients with compensated cirrhosis at highest risk of decompensation. In 2022, Schneider et al. [17] proposed the Early Prediction of Decompensation (EPOD) score that uses platelet count, albumin, and bilirubin concentration to predict decompensation during a 3-year follow-up.

Genetic risk factors might also play a role, such as the presence of certain *NOD2* (nucleotide-binding oligomerization domain-containing protein 2) variants (p.R702W, p.G908R, c.3020insC), which predict infections and increased risk of mortality in patients with cirrhosis [18, 19]. Further variants of the *NOD2* gene have been discovered (rs72796367 and rs5743271) [20] but have not been studied in the context of cirrhosis. *NOD2* is a pattern recognition receptor primarily expressed in leukocytes, but also in some intestinal epithelial cells. Risk variants are associated with an impaired mucosal barrier in Crohn's disease [21]. In carriers of these variants, inadequate activation of NF-κB may lead to deficient elimination and increased translocation of bacteria from the intestine [21, 22].

In recent years, machine learning tools [e.g. Neural Networks, Random Forests or Support Vector Machines (SVMs)] have increasingly been used for prediction purposes. They can be used to perform nonlinear statistical modelling and provide an alternative to logistic regression.

In the field of cirrhosis, machine learning has been used primarily for the analysis of radiological or histopathological images [23-27], but is also being tested for the prediction of mortality [28, 29]. For example, Guo et al. [30] analysed a large dataset with data from 34,575 patients using various machine learning techniques to predict mortality in patients with cirrhosis [30]. Recently, Reiniš et al. [31] have shown that machine learning can be used to assess the severity of portal hypertension as a risk factor for future decompensation. However, to our knowledge, machine learning has not yet been used to predict decompensation in an unselected group of patients.

In this proof-of-concept study, we aimed to evaluate whether machine learning algorithms can predict decompensation and identify associated parameters. We employed machine learning techniques because they offer advantages over conventional statistical methods, particularly in identifying non-linear relationships within datasets. Cirrhosis is a multifactorial disease, making it difficult to predict outcomes using traditional models that rely on a limited number of predefined parameters. Machine learning algorithms can analyze large datasets with numerous variables simultaneously and detect subtle interactions.

METHODS

We exploited the database of the I*N*CA-trial (Impact of *NOD2* genotype-guided antibiotic prevention on survival in patients with liver Cirrhosis and Ascites, EudraCT 2013-001626-26), which contains data of 1,415 patients from three German university hospitals (Homburg, Halle, Jena) and was conducted between 2014 and 2021. The study was performed in accordance with the guidelines of the Helsinki Declaration. Ethical approvals were obtained (271/11, 71/13, 3683-02/3, 2880-08/10), and all participants gave written, informed consent. The aim of the I*N*CA trial is to investigate whether the survival of patients with cirrhosis and *NOD2* variants as genetically defined high-risk group is improved by primary antibiotic prophylaxis of spontaneous bacterial peritonitis (SBP) with norfloxacin [32]. The database was built for screening purposes of the I*N*CA-trial and contains a large cohort of clinically well-characterized patients, including genetic data for *NOD2* variants. It contains both retrospective and prospective data. Retrospective data was retrieved from electronic medical records and included clinical data and medical history, including former decompensation events. Further information regarding medication use and laboratory parameters at the time of inclusion were also recorded. In addition, the minimal and maximal values of the laboratory parameters were included for patients from Saarland University Medical Center, Homburg. Cirrhosis was defined by (i) biopsy; (ii) a combination of clinical, laboratory, ultrasound, and endoscopy investigations; or (iii) liver stiffness measurement (LSM) >13.0kPa, as determined by transient elastography. In patients with LSM <19.7kPa, the diagnosis of cirrhosis was additionally confirmed by (i) or (ii) [32, 33]. Patients with severe comorbidities such as chronic heart failure (ejection fraction <20%) or non-HCC nonresectable cancers were excluded. As the database was built for screening purposes, the inclusion and exclusion criteria and the endpoints of the I*N*CA trial are not relevant for this study.

Decompensation was defined as the presence of ascites, development of hepatic encephalopathy (HE), jaundice, or variceal bleeding. Ascites was diagnosed by physical examination or ultrasound, HE according to the West Haven Criteria, and jaundice arbitrarily by serum bilirubin concentrations ≥3mg/dl or an increase of ≥3mg/dl from previous levels. Variceal bleeding was diagnosed by hematemesis/melena and endoscopic signs of variceal bleeding [32, 33].

For all patients, missing categorial variables were imputed with the value -1 (unknown). For patients from Homburg, missing numerical values were replaced with matching data from the electronic health records (if available); for the remaining patients, missing values were replaced with the mean values of the other patients. The appendix provides a table with the proportion of missing values for each parameter.

Since all patients from Jena were already decompensated at the index date and had many missing values, they were

excluded from this analysis. The resulting database contained data from 983 patients from Homburg and Halle.

Tables I and II summarize the baseline characteristics of the dataset.

Table I. Baseline characteristics of the dataset – Categorial parameters

Parameter		Total, n (%) (N = 983)	Homburg, n (%) (N = 771)	Halle, n (%) (N = 212)
Decompensation	0 (compensated at index date, never decompensated before)	313 (32)	267 (35)	46 (22)
	1 (compensated at index date, decompensated before)	354 (36)	188 (24)	166 (78)
	2 (decompensated at index date)	316 (32)	316 (41)	0 (0)
Decompensation during follow-up	-1 (unknown)	158 (16)	151 (20)	7 (3)
	0 (no)	441 (45)	392 (51)	49 (23)
	1 (yes)	384 (39)	228 (30)	156 (74)
Gender	0 (male)	647 (66)	510 (66)	137 (65)
	1 (female)	336 (34)	261 (34)	75 (35)
Aetiology	1 (alcoholic)	531 (54)	370 (48)	161 (76)
	2 (HCV)	134 (14)	130 (17)	4 (2)
	3 (HBV)	29 (3)	28 (4)	1 (0)
	4 (NAFLD)	75 (8)	57 (7)	18 (8)
	5 (Autoimmune hepatitis)	31 (3)	20 (3)	11 (5)
	6 (Primary biliary cholangitis)	10 (1)	7 (1)	3 (1)
	7 (Primary sclerosing cholangitis)	17 (2)	15 (2)	2 (1)
	8 (Hemochromatosis)	12 (1)	12 (2)	0 (0)
	9 (Wilson disease)	2 (0)	2 (0)	0 (0)
	10 (cryptogenic)	101 (10)	91 (12)	10 (5)
	11 (other)	40 (4)	38 (5)	2 (1)
ACLF at baseline	-1 (unknown)	1 (0)	0 (0)	1 (0)
	0 (no ACLF)	898 (91)	712 (92)	186 (88)
	1 (ACLF 1)	66 (7)	50 (6)	16 (8)
	2 (ACLF 2)	15 (2)	7 (1)	8 (4)
	3 (ACLF 3)	3 (0)	2 (0)	1 (0)
Diabetes mellitus	0 (no)	668 (68)	532 (69)	136 (64)
	1 (yes)	315 (32)	239 (31)	76 (36)
NOD2 rs5743271	-1 (unknown)	226 (23)	14 (2)	212 (100)
	0 (wildtype)	742 (76)	742 (96)	0 (0)
	1 (heterozygous)	15 (1)	15 (2)	0 (0)
NOD2 rs72796367	-1 (unknown)	227 (23)	15 (2)	212 (100)
	0 (wildtype)	720 (73)	720 (93)	0 (0)
	1 (heterozygous)	35 (4)	35 (5)	0 (0)
	2 (homozygous)	1 (0)	1 (0)	0 (0)
NOD2 p.R702W	-1 (unknown)	185 (19)	3 (0)	182 (86)
	0 (wildtype)	709 (72)	680 (88)	29 (14)
	1 (heterozygous)	89 (9)	88 (11)	1 (0)
NOD2 p.G908R	-1 (unknown)	209 (21)	3 (0)	206 (97)
	0 (wildtype)	749 (76)	743 (96)	6 (3)
	1 (heterozygous)	24 (3)	24 (3)	0 (0)
	2 (homozygous)	1 (0)	1 (0)	0 (0)
NOD2 c.3020insC	-1 (unknown)	198 (20)	3 (0)	195 (92)
	0 (wildtype)	744 (76)	728 (94)	16 (8)
	1 (heterozygous)	41 (4)	40 (5)	1 (0)

Table I (continued)				
Varices	-1 (unknown)	9 (1)	0 (0)	9 (4)
	0 (no)	547 (56)	408 (53)	139 (66)
	1 (yes)	427 (43)	363 (47)	64 (30)
PPI	-1 (unknown)	11 (1)	3 (0)	8 (4)
	0 (no)	290 (30)	234 (30)	56 (26)
	1 (yes)	682 (69)	534 (69)	148 (70)
Betablocker	-1 (unknown)	11 (1)	3 (0)	8 (4)
	0 (no)	543 (55)	419 (54)	12 (58)
	1 (yes)	429 (44)	349 (45)	80 (38)
Antibiotic long-term therapy	-1 (unknown)	12 (1)	5 (1)	7 (3)
	0 (no)	797 (81)	625 (81)	172 (81)
	1 (yes)	174 (18)	141 (18)	33 (16)

Baseline characteristics of the categorical parameters in the database. HCV: hepatitis C virus; HBV: hepatitis B virus; NAFLD: non-alcoholic fatty liver disease; ACLF: acute-on-chronic liver failure; PPI: proton-pump inhibitor.

Table II. Baseline characteristics of the dataset – Numerical parameters

Parameter	Mean value (standard deviation)
AFPMax (IU/ml)	1344.7 (10050.5)
AlbuminMin (g/l)	27.1 (7.9)
Albumin (g/l)	35.3 (7.3)
ALT (U/l)	63.3 (148.6)
ALTMax[b] (U/l)	278.8 (563.1)
AmmoniaMax (µg/dl)	136.8 (78.9)
APMax (U/l)	255.3 (209.3)
AST (U/l)	94.3 (155.5)
ASTMax (U/l)	564.2 (2094.3)
BilirubinMax (mg/dl)	5.8 (6.8)
Bilirubin (mg/dl)	2.5 (4.2)
Calcium (mmol/l)	2.3 (0.2)
CalciumMax (mmol/l)	2.5 (0.2)
CalciumMin (mmol/l)	2 (0.2)
CHEMin (kU/l)	3.3 (2.3)
CholesterolMax (mg/dl)	195.9 (58.4)
Creatinine (mg/dl)	1.1 (0.78)
CreatinineMax (mg/dl)	2.1 (1.5)
CRP (mg/l)	19.6 (31.3)
CRPMax[f](mg/l)	94.7 (86.2)
FerritinMax (ng/ml)	1099.2 (2576.2)
FerritinMin (ng/ml)	320.7 (658.6)
γGTMax (U/l)	544.7 (601.4)
GlucoseMax (mg/dl)	222.7 (115.3)
HB (g/dl)	12.1 (2.5)
HBMin (g/dl)	9.3 (2.5)
INR	1.3 (0.4)
INRMax	1.8 (0.7)
Potassium (mmol/l)	4.1 (0.5)
PotassiumMax (mmol/l)	5.1 (0.5)
PotassiumMin (mmol/l)	3.4 (0.4)
Sodium (mmol/l)	137.7 (4.7)
SodiumMin (mmol/l)	131.7 (4.3)
Platelets (x10^9/l)	154.3 (102.6)
PlateletsMin (x10^9/l)	88.5 (53)
PTT (s)	31.8 (7.8)
PTTMax(s)	53 (27.2)
TAGMax(mg/dl)	194 (172.4)
Urea (mg/dl)	43.6 (31.5)
UreaMax (mg/dl)	92.9 (58.9)
WBC (x10^9/l)	7.1 (4.0)
WBCMin (x10^9/l)	4.1 (1.9)

This table presents the baseline characteristics of the numerical parameters in the database. *Variable designation followed by Min or Max: minimum or maximum values of the parameter. AFP: alpha-fetoprotein; ALT: alanine aminotransferase; AP: alkaline phosphatase; AST: aspartate aminotransferase; CHE: cholinesterase; CRP: C-reactive protein; γGT: gamma-glutamyltransferase; HB: haemoglobin; INR: International Normalized Ratio; PTT: partial thromboplastin time; TAG: triacylglycerol; WBC: white blood cell count.

In order to determine the parameters indicating the development of decompensation, we tested multiple machine learning techniques, using Python 3.7.5, Keras 2.3.1, and Scikit-Learn 0.22.3 [34]. Keras and Scikit-learn are open-source machine learning software libraries for the Python programming language. Keras is commonly used for deep learning with neural networks, and Scikit-learn provides various classification, regression, and clustering algorithms. We tested tree-based approaches like decision trees (DT) and random forest classifiers (RF), neural networks (NN), and SVMs. As most machine learning algorithms are not scale-invariant, the numerical input features were scaled to a mean of 0 and a variance of 1.

For receiver operating curve (ROC) analysis, we used macro as weighting strategy for evaluating a multiclass classification, i.e., the metrics were calculated for each label and their unweighted mean was returned. We always trained

two models: one to retrospectively predict a decompensation prior to the index date (referred to as the retrospective dataset) and another to predict decompensation events in the follow-up period (referred to as the prospective dataset). For the retrospective dataset, we differentiated between patients who had never experienced a decompensation before, patients who presented with a decompensation earlier, but showed no signs of decompensation at the index date, and patients who were decompensated at the index date. For the prospective dataset, we only distinguished whether the patient had decompensation events during follow-up or not.

To evaluate the impact of features on the prediction of a decompensation, we used the permutation feature importance (PFI). This is a model inspection technique that can be used to determine the importance of features for a particular model. It is defined as the decrease in a model score (e.g. accuracy) when the value of a single feature is randomly permuted [35]. In this study, we applied the PFI on the test dataset to highlight the features that contribute the most to the generalization power.

As PFI can be biased by correlated features, we performed hierarchical clustering on the numerical variables based on the Spearman rank-order correlation using Ward's minimum variance method [36]. Ward's method is an agglomerative hierarchical clustering procedure in which the increase in variance of the clusters to be merged is decisive.

In line with other studies, we randomly split the resulting dataset (DS) into two datasets: DS1 (85%) was used for training and tuning, and DS2 (15%) was used for testing (holdout data). We always trained two machine learning models: one to predict a decompensation in the follow-up period and another to assess a decompensation retrospectively. Possible hyperparameter combinations (i.e., parameters that are not learned directly during the training process) were tested and evaluated with 5-fold cross-validation, and the best combination was selected [37].

After training and tuning the model on DS1, we evaluated its performance on DS2 and applied the PFI on this dataset. As we were interested in the feature importance rather independent of a specific model, we iterated the process of randomly splitting the dataset, training/tuning the model, and permuting features 50 times.

Fig. 1 illustrates the study design based on the examples of Random Forests and SVMs.

RESULTS

The resulting database contained data from 983 patients of whom 65.8% were male. The most common aetiology of cirrhosis was alcohol use (54%), followed by chronic hepatitis C virus infection (13.6%). At the index date, 313 patients (31.8%) had never experienced a decompensation before, 354 patients (36%) were decompensated before, but showed no current signs of decompensation, and 316 (32.1%) were decompensated. 825 patients (83.9%) attended follow up (median duration: 12 months, maximum: 55 months), of whom 384 (46.5%) experienced a decompensation event during follow-up (Table I). The proportion of variables with missing values is shown in the Appendix.

The dendrogram in Fig. 2 illustrates the results of the cluster analysis. In this dendrogram, the vertical axis shows the numerical parameters and the horizontal axis represents the distances or dissimilarities between the possible clusters. We selected a threshold value of 0.8 to group the features into clusters. For each cluster, we chose one feature and dropped the others to avoid biases induced by highly correlated variables.

As demonstrated in Table III, RF achieved the best results in terms of accuracy for predicting prior decompensation, SVM performed best for prospective assessment.

For predicting prior decompensation on the retrospective test data, DT achieved a mean accuracy of 65.5%, RF 70.5%, SVM 68.5% and NN 65%. For predicting decompensation in the follow-up period, the respective values were 71.1% for DT, 69.3% for RF, 73.8% for SVM, and 72.9% for NN.

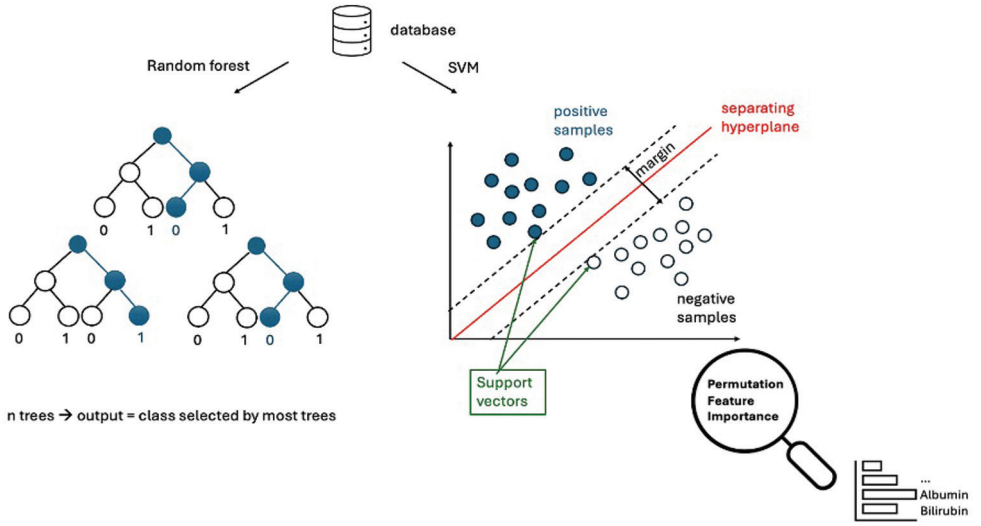

Fig. 1. Study design. Random Forests and Support Vector Machines were trained on a database containing clinical data and laboratory values of patients with cirrhosis and the permutation feature importance was used as a model inspection technique to determine the impact parameters on the prediction of decompensation.

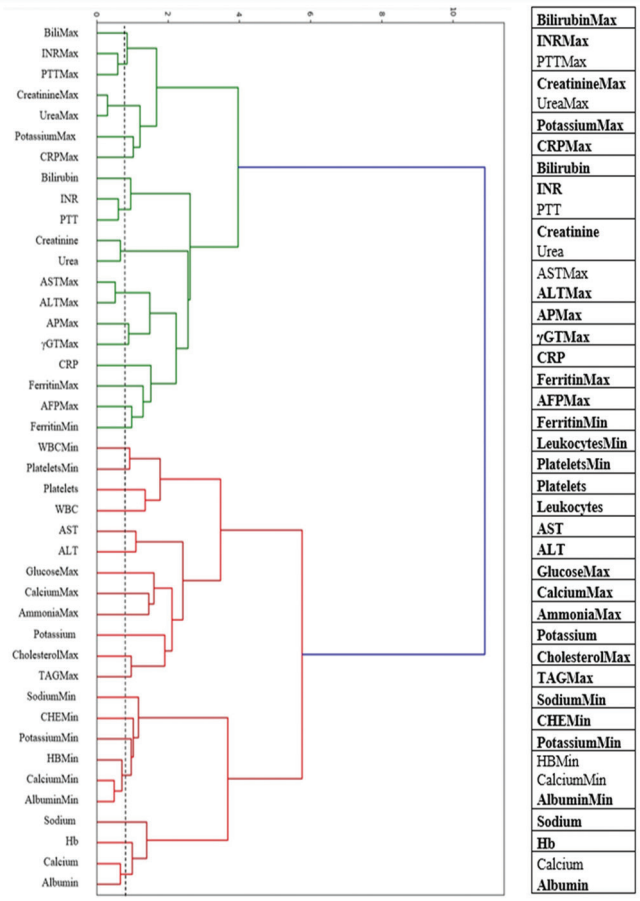

Fig. 2. Cluster analysis: The dashed line in the dendrogram represents the threshold value; this table shows the resulting clusters, with the selected parameters marked in bold. As the variables correlate strongly within a cluster and can influence the results of the PFI, only one variable per cluster was used for further analysis. For abbreviations see Table II. Variable designation followed by Min or Max: minimum or maximum values of the parameter.

Table III. Comparison of the machine-learning methods

Machine learning technique		Accuracy Training data	Accuracy Test data
Decision Trees	Retrospective assessment	0.7189	0.6551
	Prospective assessment	0.7862	0.7113
Random Forests	Retrospective assessment	0.8155	0.7047
	Prospective assessment	0.8264	0.6928
SVM	Retrospective assessment	0.7841	0.6851
	Prospective assessment	0.7859	0.7384
Neural Networks	Retrospective assessment	0.9164	0.6502
	Prospective assessment	0.8225	0.7295

Accuracy values of the tested machine learning techniques for the retrospective assessment of a former decompensation as well as for the prospective prediction of a decompensation during the follow-up period. SVM: support vector machine.

Table IV summarizes the performance metrics of the best method in each case. As shown in these tables, the Random Forest achieved a mean accuracy of 81.6% on the training data and 70.5% on the test data for predicting prior decompensation in the retrospective dataset. For the test data, the mean ROC area under the curve (AUC) was 0.87 and the F1 value was 0.7.

For predicting a decompensation in the prospective dataset, the SVM achieved a mean accuracy of 78.6% on the training and 73.8% on the test data. ROC AUC was 0.81 and F1-score 0.74 for the test data.

Table IVa. Performance Metrics of the Random Forest (retrospective assessment)

Metric	Training data	Test data
Accuracy	0.8156	0.7048
ROC AUC (macro)	0.9549	0.8694
F1 Score (macro)	0.8147	0.7008

Table IVb. Performance Metrics of the SVM (prospective assessment)

Metric	Training data	Test data
Accuracy	0.7859	0.7384
ROC AUC	0.8567	0.8166
F1 Score	0.7851	0.7372

Performance metrics of the Random Forest for retrospective and of the Support Vector Machine for prospective assessment of decompensation. ROC: receiver operating characteristic; AUC: area under the curve.

As shown in panel A of Fig. 3, the analysis of the PFI revealed that the baseline levels of albumin and bilirubin, the maximum bilirubin level and the baseline level of C-reactive protein (CRP) were the highest ranked parameters for assessing a former decompensation in the retrospective dataset.

For predicting a decompensation event in the prospective dataset, the maximum value of bilirubin, the baseline values of sodium and albumin and the maximum value of ammonium as well as the baseline value of CRP were the highest ranked parameters (Fig. 3B).

DISCUSSION

This is the first proof-of-concept study that aimed to examine whether parameters predicting decompensation can be identified using machine learning techniques in an unselected group of patients with cirrhosis. In the field of hepatology, machine learning has been tested for the diagnosis and staging of fibrosis, the detection of hepatocellular carcinoma in radiological imaging and histopathological analysis [23-26]. Recently, Bosch et al. [27] analysed liver histology with neural networks to extrapolate the hepatic venous pressure gradient. Several studies investigated disease progression in patients with chronic hepatitis B or C virus infections [38-40]. Furthermore, few studies have used machine learning for the prediction of mortality in patients with cirrhosis [28-30]. However, we could not find a study that investigated the prediction of decompensation with machine learning, and genetic data (such as *NOD2* genotype) have not been examined in previous studies.

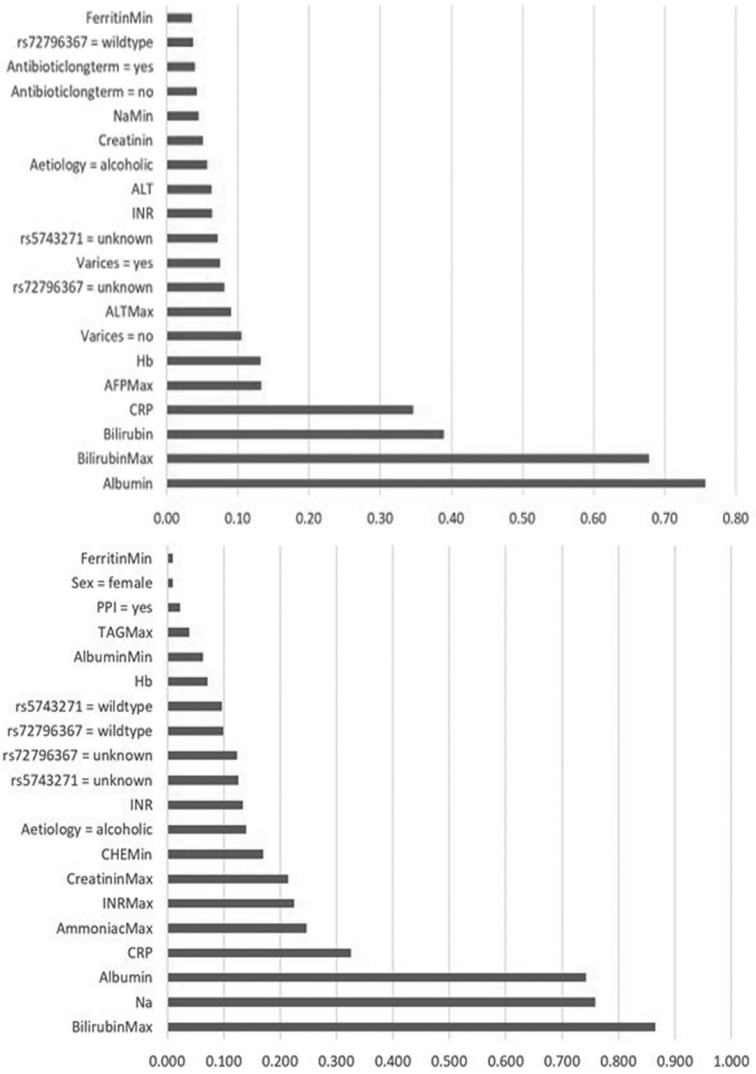

Fig. 3. The 20 highest ranked parameters of the Permutation Feature Importance. A) Permutation Feature Importances for the retrospective assessment of decompensation. B) Permutation Feature Importances for predicting decompensation during follow-up. These figures show the 20 highest ranked parameters for the prediction of decompensation. For abbreviations see Table II.

Random Forests and SVMs have some advantages compared with "classical" statistical analysis such as logistic regression or other machine learning techniques: They are supervised learning models, which can be used for nonlinear classification. They are easily scalable and can be trained on large datasets with many variables. In contrast to neural networks, SVMs are still effective in high-dimensional spaces. Once trained, Random Forests or SVMs could easily be implemented in the clinical routine and might help to stratify the treatment of patients.

In this study, machine learning techniques were applied to predict the decompensation of cirrhosis and determine parameters indicating decompensation. For the prediction of a former decompensation, the baseline values of albumin and bilirubin and the maximum bilirubin level were the highest ranked parameters. The maximum bilirubin level and the baseline values of sodium and albumin were the highest ranked for predicting a decompensation in the follow-up period.

The results of this study partly match the established scores including MELD, Child-Pugh, ALBI, and FIB-4. The ALBI score combines serum albumin and total bilirubin to assess liver function, these two parameters were also highly ranked by the PFI [8]. The FIB-4 index estimates the extent of fibrosis based on age, AST, ALT, and platelet count [15]. It has been extensively validated and shown to have prognostic ability for patients with chronic liver diseases [41-43]. In our study, the importance of platelet count was relatively low, and serum aminotransferase activities were ranked among the 20 most important parameters only in retrospective assessment.

The MELD score is an established model to predict the survival probability of patients with end stage liver disease [3, 4]. Over the time, several modifications of the MELD score were proposed, but the values of bilirubin, creatinine, and INR were consistently included. Initially, the aetiology of the disease was also considered in the MELD score [3], but later it was removed as it was found to be less important for survival estimation [4]. In our analysis, only alcoholic liver disease was

relevant. Several investigators have reported that incorporating sodium into the MELD score increases its predictive accuracy [44, 45]. In line with these reports, serum sodium levels were also highly ranked in our analysis.

Our results are only partially consistent with those of Guo et al. [30] who analysed a dataset comprising 34,575 patients using a deep neural network, random forests, and logistic regression to predict 30-, 90-, and 180-day mortality in patients with cirrhosis. They demonstrated that all models outperformed the MELD score, with NNs performing best. In addition to MELD score variables, alkaline phosphatase, ALT, and haemoglobin were important parameters in their analysis. However, we could not confirm alkaline phosphatase as an important parameter for the prediction of decompensation.

The newly introduced EPOD score is the only score that estimates decompensation rather than mortality. It uses platelet count, albumin and bilirubin concentration [17]. Compared with this score, our analysis also revealed a high importance of albumin and bilirubin, whereas platelets were surprisingly not ranked among the top 20 parameters.

Studies have shown that clinically significant portal hypertension (CSPH) predicts decompensation, as assessed by the hepatic venous pressure gradient (HVPG) [46]. However, because measurement of HVPG is an invasive procedure, its use as a screening parameter for decompensation is difficult, especially in compensated patients. According to the Baveno VII consensus, clinically significant portal hypertension can also be diagnosed in a non-invasive fashion using LSM +/- platelet count [47]. Furthermore, Reiniš et al. [31] demonstrated that the severity of portal hypertension can be assessed by machine learning. Therefore, they tested different machine learning methods and determined the most important parameters via recursive feature elimination. They found that a 3-parameter model with platelet count, bilirubin and INR or a 5-parameter model with platelet count, bilirubin, activated partial thromboplastin time, γGT and cholinesterase in combination with LSM achieved the best results [31].

Further studies identified anaemia and decreased vitamin D levels as predictors of decompensation [48, 49]. Haemoglobin was ranked among the top 20 parameters in both our retrospective and prospective analysis; however, we cannot make any statement about the vitamin D level, as we have no data available.

The importance of CRP and antibiotics intake confirm a prominent role of infections as triggers of decompensation. Consistent with this, another study has already identified inflammation markers like IL-6 as predictors of decompensation [50]. This is supported by our finding that the *NOD2* genotype, with variants p.R702W, p.G908R and c.3020insC is described as being associated with early death, spontaneous bacterial peritonitis and other infections [18, 21], was also considered as an important factor based on the PFI. Interestingly, variants rs5743271 und rs72796367 - whose prognostic value in liver cirrhosis has not been studied so far - ranked higher based on PFI than the better characterised variants p.R702W, p.G908R and c.3020insC. However, genotypic data were missing for a relevant number of patients, limiting the robustness of our model. Future independent studies will need to verify the prognostic relevance of these variants in liver cirrhosis.

We are aware that our analysis has several limitations. First of all, the application of Random Forests and SVMs is associated with several limitations/shortcomings. Most machine learning techniques cannot deal with missing values, but medical patient data is often incomplete and difficult to retrieve. In the I*N*CA trial database, a large part of the numerical and genotypic data was missing due to the screening character of data collection. These had to be imputed for our analysis, which could have biased the results. Although there are several approaches to make machine learning more explainable, it is still difficult to identify parameters with good predictive ability. PFI is one approach to determine the importance of features but is strongly biased by correlated features. Secondly, our study lacks an independent external validation cohort, and although we used a holdout test dataset, further external validation is necessary. Furthermore, a part of the study had a retrospective design, and the mean follow-up time was rather short for the prospective part. Besides that, the size of the dataset was relatively small, particularly for the prospective part, and all patients were recruited at tertiary referral centres, limiting the generalisability to other healthcare settings. In addition, our database did not contain data regarding hospitalization and for the patients from Halle laboratory values were only collected at the time of inclusion. Data on the presence of recompensation were extracted from electronic health records, but further information (e.g. regarding therapy) could not be retrieved. Furthermore, a third of the patients were decompensated at the index date, which might have biased the results, as current decompensation is linked to abnormal liver test results. A direct statistical comparison of our model with established scores was not possible due to the different endpoints.

Further research is required with larger sample sizes, extended follow-up periods, and more comprehensive patient characterization to enhance the predictive accuracy and generalizability of machine learning models. Independent external validation in diverse patient populations and across different healthcare settings is crucial to confirm the robustness and reliability of these models. Furthermore, the development and application of advanced model inspection techniques could help identify clinically relevant predictors and address challenges such as bias by correlated parameters in permutation feature importance analyses. Finally, the role of *NOD2* genotype variants (e.g., rs5743271 and rs72796367) warrants further investigation.

CONCLUSIONS

In this proof-of-concept study, we demonstrated the potential of machine learning techniques to predict decompensation and identify key predictive parameters in patients with cirrhosis. Our findings suggest that machine learning could play a valuable role in improving risk assessment and guiding early interventions, although further research is required to validate these results in prospective studies in larger cohorts.

Conflicts of interest: None to declare.

Authors' contributions: F.L. and M.C.R. conceived the study. F.L., M.C.R. and S.E.M designed the methodology and analysed the data.

S.E.M was responsible for the spftware and drafted the manuscript. All the authors revised and edited the manuscript. F.L. and M.C.R. supervised the study. All the authors read and approved the final version of the manuscript.

REFERENCES

1. D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. J Hepatol 2006;44:217–231. doi:10.1016/j.jhep.2005.10.013
2. Pugh RNH, Murray-Lyon IM, Dawson JL, Pietroni MC, Williams R. Transection of the oesophagus for bleeding oesophageal varices. Br J Surg 1973;60:646–649. doi:10.1002/bjs.1800600817
3. Kamath PS, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. Hepatology 2001;33:464–470. doi:10.1053/jhep.2001.22172
4. Kamath PS, Kim WR; Advanced Liver Disease Study Group. The model for end-stage liver disease (MELD). Hepatology 2007;45:797–805. doi:10.1002/hep.21563
5. Cholongitas E, Senzolo M, Triantos C, Samonakis D, Patch D, Burroughs AK. MELD is not enough--enough of MELD? J Hepatol 2005;42:475–479. doi:10.1016/j.jhep.2005.02.002
6. Cholongitas E, Papatheodoridis GV, Vangeli M, Terreni N, Patch D, Burroughs AK. Systematic review: The model for end-stage liver disease--should it replace Child-Pugh's classification for assessing prognosis in cirrhosis? Aliment Pharmacol Ther 2005;22:1079–1089. doi:10.1111/j.1365-2036.2005.02691.x
7. Bernardi M, Gitto S, Biselli M. The MELD score in patients awaiting liver transplant: strengths and weaknesses. J Hepatol 2011;54:1297–1306. doi:10.1016/j.jhep.2010.11.008
8. Johnson PJ, Berhane S, Kagebayashi C, et al. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach-the ALBI grade. J Clin Oncol 2015;33:550–558. doi:10.1200/JCO.2014.57.9151
9. Toyoda H, Lai PBS, O'Beirne J, et al. Long-term impact of liver function on curative therapy for hepatocellular carcinoma: application of the ALBI grade. Br J Cancer 2016;114:744–750. doi:10.1038/bjc.2016.33
10. Hiraoka A, Kumada T, Michitaka K, et al. Usefulness of albumin-bilirubin grade for evaluation of prognosis of 2584 Japanese patients with hepatocellular carcinoma. J Gastroenterol Hepatol 2016;31:1031–1036. doi:10.1111/jgh.13250
11. Pinato DJ, Sharma R, Allara E, et al. The ALBI grade provides objective hepatic reserve estimation across each BCLC stage of hepatocellular carcinoma. J Hepatol 2017;66:338–346. doi:10.1016/j.jhep.2016.09.008
12. Peng Y, Qi X, Tang S, et al. Child-Pugh, MELD, and ALBI scores for predicting the in-hospital mortality in cirrhotic patients with acute-on-chronic liver failure. Expert Rev Gastroenterol Hepatol 2016;10:971–980. doi:10.1080/17474124.2016.1177788
13. Zou D, Qi X, Zhu C, et al. Albumin-bilirubin score for predicting the in-hospital mortality of acute upper gastrointestinal bleeding in liver cirrhosis: A retrospective study. Turk J Gastroenterol 2016;27:180–186. doi:10.5152/tjg.2016.15502
14. Chen RC, Cai YJ, Wu JM, et al. Usefulness of albumin-bilirubin grade for evaluation of long-term prognosis for hepatitis B-related cirrhosis. J Viral Hepat 2017;24:238–245. doi:10.1111/jvh.12638
15. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. Hepatology. 2006 Jun;43(6):1317–25.
16. Guha IN, Harris R, Berhane S, et al. Validation of a Model for Identification of Patients With Compensated Cirrhosis at High Risk of Decompensation. Clin Gastroenterol Hepatol 2019;17:2330-2338.e1. doi:10.1016/j.cgh.2019.01.042
17. Schneider ARP, Schneider CV, Schneider KM, et al. Early prediction of decompensation (EPOD) score: Non-invasive determination of cirrhosis decompensation risk. Liver Int 2022;42:640–650. doi:10.1111/liv.15161
18. Appenrodt B, Grünhage F, Gentemann MG, Thyssen L, Sauerbruch T, Lammert F. Nucleotide-binding oligomerization domain containing 2 (*NOD2*) variants are genetic risk factors for death and spontaneous bacterial peritonitis in liver cirrhosis. Hepatology 2010;51:1327–1333. doi:10.1002/hep.23440
19. Bruns T, Peter J, Reuken PA, et al. NOD2 gene variants are a risk factor for culture-positive spontaneous bacterial peritonitis and monomicrobial bacterascites in cirrhosis. Liver Int 2012;32:223–230. doi:10.1111/j.1478-3231.2011.02561.x
20. Huang H, Fang M, Jostins L, et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. Nature 2017;547:173–178. doi:10.1038/nature22969
21. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001;411:599–603. doi:10.1038/35079107
22. Hampe J, Cuthbert A, Croucher PJ, et al. Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. Lancet 2001;357:1925–1928. doi:10.1016/S0140-6736(00)05063-7
23. Yasaka K, Akai H, Kunimatsu A, Abe O, Kiryu S. Deep learning for staging liver fibrosis on CT: a pilot study. Eur Radiol 2018;28):4578–4585. doi:10.1007/s00330-018-5499-7
24. Choi KJ, Jang JK, Lee SS, et al. Development and Validation of a Deep Learning System for Staging Liver Fibrosis by Using Contrast Agent-enhanced CT Images in the Liver. Radiology 2018;289:688–697. doi:10.1148/radiol.2018180763
25. Decharatanachart P, Chaiteerakij R, Tiyarattanachai T, Treeprasertsuk S. Application of artificial intelligence in chronic liver diseases: a systematic review and meta-analysis. BMC Gastroenterol 2021;21:10. doi:10.1186/s12876-020-01585-5
26. Spann A, Yasodhara A, Kang J, et al. Applying Machine Learning in Liver Disease and Transplantation: A Comprehensive Review. Hepatology 2020;71:1093–1105. doi:10.1002/hep.31103
27. Bosch J, Chung C, Carrasco-Zevallos OM, et al. A Machine Learning Approach to Liver Histological Evaluation Predicts Clinically Significant Portal Hypertension in NASH Cirrhosis. Hepatology 2021;74:3146–3160. doi:10.1002/hep.32087
28. Banerjee R, Das A, Ghoshal UC, Sinha M. Predicting mortality in patients with cirrhosis of liver with application of neural network technology. J Gastroenterol Hepatol 2003;18:1054–1060. doi:10.1046/j.1440-1746.2003.03123.x
29. Cucchetti A, Vivarelli M, Heaton ND, et al. Artificial neural network is superior to MELD in predicting mortality of patients with end-stage liver disease. Gut 2007;56:253–258. doi:10.1136/gut.2005.084434
30. Guo A, Mazumder NR, Ladner DP, Foraker RE. Predicting mortality among patients with liver cirrhosis in electronic health records with machine learning. PLoS One 2021;16:e0256428. doi:10.1371/journal.pone.0256428

31. Reiniš J, Petrenko O, Simbrunner B, et al. Assessment of portal hypertension severity using machine learning models in patients with compensated cirrhosis. J Hepatol 2023;78:390–400. doi:10.1016/j.jhep.2022.09.012
32. Casper M, Mengel M, Fuhrmann C, et al. The INCA trial (Impact of NOD2 genotype-guided antibiotic prevention on survival in patients with liver Cirrhosis and Ascites): study protocol for a randomized controlled trial. Trials 2015;16:83. doi:10.1186/s13063-015-0594-4
33. Reichert MC, Ripoll C, Casper M, et al. Common NOD2 risk variants as major susceptibility factors for bacterial infections in compensated cirrhosis. Clin Transl Gastroenterol 2019;10:e00002. doi:10.14309/ctg.0000000000000002
34. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. J Mach Lear Res 2011;12:2825–2830.
35. Breiman L. Random Forests. Machine Learning 2001;45:5–32. doi:10.1023/A:1010933404324
36. Ward JH. Hierarchical Grouping to Optimize an Objective Function. J Am Stat Assoc 1963;58:236–244. doi:10.1080/01621459.1963.10500845
37. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res 2012;13281-305.
38. Konerman MA, Beste LA, Van T, et al. Machine learning models to predict disease progression among veterans with hepatitis C virus. PLoS One 2019;14:e0208141. doi:10.1371/journal.pone.0208141
39. Wang D, Wang Q, Shan F, Liu B, Lu C. Identification of the risk for liver fibrosis on CHB patients using an artificial neural network based on routine and serum markers. BMC Infect Dis. 2010 Aug 24;10:251- doi:10.1371/journal.pone.0208141
40. Wei R, Wang J, Wang X, et al. Clinical prediction of HBV and HCV related hepatic fibrosis using machine learning. EBioMedicine 2018;35:124–132. doi:10.1016/j.ebiom.2018.07.041
41. Angulo P, Bugianesi E, Bjornsson ES, et al. Simple noninvasive systems predict long-term outcomes of patients with nonalcoholic fatty liver disease. Gastroenterology 2013;145:782-789.e4. doi:10.1053/j.gastro.2013.06.057
42. Vergniol J, Foucher J, Terrebonne E, et al. Noninvasive tests for fibrosis and liver stiffness predict 5-year outcomes of patients with chronic hepatitis C. Gastroenterology 2011;140:1970–1979.e1-e3. doi:10.1053/j.gastro.2011.02.058
43. de Lédinghen V, Vergniol J, Barthe C, et al. Non-invasive tests for fibrosis and liver stiffness predict 5-year survival of patients chronically infected with hepatitis B virus. Aliment Pharmacol Ther 2013;37:979–988. doi:10.1111/apt.12307
44. Kim WR, Biggins SW, Kremers WK, et al. Hyponatremia and Mortality among Patients on the Liver-Transplant Waiting List. N Engl J Med 2008;359:1018–1026. doi:10.1056/NEJMoa0801209
45. Biggins SW, Kim WR, Terrault NA, et al. Evidence-Based Incorporation of Serum Sodium Concentration Into MELD. Gastroenterology 2006;130:1652–1660. doi:10.1053/j.gastro.2006.02.010
46. Ripoll C, Groszmann R, Garcia-Tsao G, et al. Hepatic venous pressure gradient predicts clinical decompensation in patients with compensated cirrhosis. Gastroenterology 2007;133:481–488. doi:10.1053/j.gastro.2007.05.024
47. de Franchis R, Bosch J, Garcia-Tsao G, Reiberger T, Ripoll C; Baveno VII Faculty. Baveno VII – Renewing consensus in portal hypertension. J Hepatol 2022;76:959–974. doi: 10.1016/j.jhep.2021.12.022
48. Bothou C, Rüschenbaum S, Kubesch A, et al. Anemia and Systemic Inflammation Rather than Arterial Circulatory Dysfunction Predict Decompensation of Liver Cirrhosis. J Clin Med 2020;9:1263. doi:10.3390/jcm9051263
49. Kubesch A, Quenstedt L, Saleh M, et al. Vitamin D deficiency is associated with hepatic decompensation and inflammation in patients with liver cirrhosis: A prospective cohort study. PLoS One 2018;13:e0207162. doi:10.1371/journal.pone.0207162