Original Paper
Wei-Huan Hu¹, MS; Heng-Chih Pan², MD; Ming-Hung Tsai³, MD; Yuh-Jyh Hu⁴, PhD; Shang-Hwa Hsu⁵, PhD; Yung-Chang Chen²,⁶, MD

¹ Institute of Computer Science and Engineering of National Chiao Tung University, Hsinchu, Taiwan
² Division of Nephrology, Department of Internal Medicine, Chang Gung Memorial Hospital, Keelung, Taiwan
³ Division of Gastroenterology, Chang Gung Memorial Hospital, Linkou, Taiwan
⁴ Institute of Biomedical Engineering and the Department of Computer Science of National Chiao Tung University, Hsinchu, Taiwan.
⁵ Department of Industrial Engineering and Management of National Chiao Tung University, Hsinchu, Taiwan
⁶ Chang Gung University College of Medicine, Taoyuan, Taiwan

Authorship:
Guarantor of the article: Wei-Huan Hu.
Author contributions: YCC and MHT contributed to the conception, design and interpretation of data. HCP provided patient information, participated in the design and coordination, and helped to draft the manuscript. YJH and SHH provided intellectual content to the study and were involved in editing and revising the manuscript. All of the authors discussed, contributed to and approved the final version of the manuscript.

Corresponding Author:
Yung-Chang Chen
Division of Nephrology,
Department of Internal Medicine,
Chang Gung Memorial Hospital,
222 Mai-Chih Road, Keelung 20401, Taiwan
Tel.: 886-2-24313131 ext. 3170
Fax: 886-2-24335342
E-mail: cyc2356@gmail.com
Development of 365-day Life Expectancy Models: Application of Machine Learning Methods to a Prospective Study of Critically Ill Cirrhotic Patients

Abstract

**Background:** The mortality rate of cirrhotic patients in intensive care units (ICUs) is usually high. Analyzing high dimensional data and incorporating accurate life expectancy indices using machine learning methods may improve the accuracy of the prediction of the long-term prognosis of critically ill cirrhotic patients.

**Objective:** To develop highly accurate 365-day candidate life expectancy models based on machine learning methods, and compare their accuracy with CLIF-SOFA scores in critically ill cirrhotic patients.

**Methods:** This study analyzed 141 prospective predictor variables on day 1 of 294 cirrhotic patients admitted to a 10-bed specialized hepato-gastroenterology ICU in a 2000-bed tertiary care referral hospital from September 2010 to August 2013. The least absolute shrinkage and selection operator (LASSO) method was used to select subsets of predictor variables. Model validation was applied to compare multiple life-expectancy models with CLIF-SOFA scores.

**Results:** The overall in-hospital and 365-day mortality rates were 55.4% (163/294) and 82.3% (242/294), respectively. CLIF-SOFA score, hepatic encephalopathy grade and serum creatinine were the most critical risk predictors. The final models were developed by incorporating the most important 16 predictors based on Support Vector Machine (SVM), Random Sample Consensus (RANSAC), Ensemble, Random Forest and Decision Tree methods. The SVM model showed excellent discrimination (0.901±0.021) and outperformed CLIF-SOFA score (0.841±0.024) as well as the others. The RANSAC model had the best Youden index, and the Ensemble model had the highest overall correctness of prediction.

**Conclusions:** The 365-day life expectancy models based on SVM/Random Forest/RANSAC/Ensemble classifiers were excellent prognostic evaluation tools for critically ill cirrhotic patients.

**Keywords:** Machine learning; Long-term prognosis; Support Vector Machine; Random Sample Consensus; Ensemble; Random Forest; CLIF-SOFA
Introduction
The prognosis of patients with cirrhosis in intensive care units (ICUs) is poor. The high mortality rate is attributed to multiple organ dysfunction [1-3]. Disturbed systemic circulation in end-stage liver disease is caused by arterial vasodilation in the splanchnic circulation, reduced total peripheral vascular resistance and arterial pressure causing a secondary increase in cardiac output [4-6]. These abnormalities can result in severe crucial cirrhotic complications such as hepatocardiac syndrome, hepatorenal syndrome, hepatic encephalopathy, hepatopulmonary syndrome, coagulopathy and jaundice [7-10].

Accurate life expectancy prediction is critical when making clinical decisions, as it helps physicians to assess the benefits and risks of alternative care strategies and determine the best choice for each patient. Failure to consider life expectancy leads to poor quality of care and waste of medical resources. The Sequential Organ Failure Assessment (SOFA) score is an excellent tool for assessing six vital organ systems, and it has been demonstrated to have good discriminatory power for predicting mortality in critically ill cirrhotic patients [11-13]. The European Association for the Study of Liver/Chronic Liver Failure Consortium further modified the parameters of SOFA according to specific factors associated with end-stage liver disease and proposed the Chronic Liver Failure Sequential Organ Failure Assessment (CLIF-SOFA) score to replace the SOFA score and define acute-on-chronic-liver-failure (ACLF) [14]. Our previous study validated that the CLIF-SOFA score was independently associated with in-hospital mortality and had better discriminatory power than the SOFA score for predicting 6-month mortality in critically ill cirrhotic patients [4]. However, the accuracy of the long-term prognosis of critically ill cirrhotic patients is still insufficient.

Recent innovations in prognostic healthcare have shown that machine learning can generate actionable insights based on massive and high dimensional prospective data, and it has been explored as a means to improve the treatment of infections and cancer, identify adverse drug events, measure the quality of asthma care, and predict cancer outcomes. For example, International Business Machines Watson Health represents a new technique to analyze medical data [16]. To improve the accuracy of the prediction of the long-term prognosis of critically ill cirrhotic patients, we conducted this investigation incorporating life expectancy indices using machine learning. Our goals were to: (1) present a set of approaches for machine learning and analysis of high dimensional prospective data, (2) develop highly accurate 365-day candidate life expectancy models, and (3) externally validate the efficacy of the candidate models and compare their accuracy with the CLIF-SOFA score in predicting the prognosis of critically ill cirrhotic patients.

Materials and Methods

Ethics Statement
This clinical study was conducted in full compliance with the ethical principles of the Declaration of Helsinki and was consistent with the Good Clinical Practice guidelines.
and applicable local regulatory requirements. The local institutional review board of Chang Gung Memorial Hospital approved the study protocol (approval No. 98-3658A3). Patients who met the inclusion criteria were invited to participate in this study on their first day of ICU admission. Trained physicians evaluated their mental status during screening and conducted informed consent procedures. Written informed consent was obtained from all of the mentally competent patients or the next of kin of compromised patients before their participation.

**Patient Information and Data Collection**

This study was conducted from September 2010 to August 2013 in a 10-bed specialized ICU (Hepato-Gastroenterology ICU) at a 2000-bed tertiary care referral hospital in Taiwan. A total of 294 cirrhotic patients admitted to this ICU were enrolled. Prospective data were collected including 141 predictor variables (physiological variables: 23, laboratory results: 35, demographic data: 78, and medications: 5, see Suppl. Table 1) and treatment outcomes. The primary study outcome was 365-day mortality rate. Follow-up after hospital discharge was performed via telephone interview or by analyzing the chart records.

**Definitions**

Cirrhosis was diagnosed on the basis of the results of liver histology or a combination of physical signs and symptoms and findings from biochemical analysis and ultrasonography. The severity of liver disease on admission to the ICU was determined using the Child-Pugh and MELD scoring systems. The severity of illness was also assessed using Acute Physiology And Chronic Health Evaluation (APACHE) II, APACHE III, SOFA and CLIF-SOFA scores. The CLIF-SOFA score was based on a six organ system [4]. The worst physiological and biochemical values determined on the first day of ICU admission were recorded.

**Ascertainment of 365-day Survival**

The primary outcome was death within 12 months of admission to the hepatogastroenterology ICU.

**Predictor Variables**

We extracted 141 distinct predictor variables derived from the prospective data of the 294 enrolled patients. These variables included 23 physiological variables, 35 biochemical variables, 78 demographic variables, and five interventions determined on the first day and during the period of ICU admission.

**Demographic Data**

We extracted 78 demographic variables from the prospective data, including age, gender, past history, source of admission, reason for admission to the ICU, and cause of cirrhosis (see Suppl. Table 1).

**Physiological Data**

We extracted 23 physiological variables from the prospective data, including body temperature, respiratory rate, heart rate, blood pressure, urine output, body weight, and fraction of inspired oxygen (see Suppl. Table 1).
Biochemical Data
We extracted 35 laboratory results from the prospective data, including serum bilirubin, albumin, alanine aminotransferase, aspartate aminotransferase, creatinine, blood urea nitrogen, estimated glomerular filtration rate, electrolyte, international normalized ratio, platelet count, white blood cell, hemoglobin, hematocrit, prothrombin time, glucose, and blood gas analysis (see Suppl. Table 1).

Intervention Data
We extracted five interventions from the prospective data, including terlipressin, renal replacement therapy, anesthetic drugs, emergency surgery, and ventilator support.

Feature Selection and Machine Learning
Feature selection can be used to reduce the number of variables while retaining the predictive power of the original variables [25]. When analyzing all of the prospective data, we used a least absolute shrinkage and selection operator (LASSO) approach to identify a subset of variables [26]. An advantage of this approach is that some of the effects of the variables in these models are estimated to be exactly zero. This represents variables that have no discriminatory power between the two classes, while those with nonzero coefficients represent variables that can dichotomize 365-day mortality successfully. Thus, a byproduct of the approach is a subset of variables. We used machine learning methods including Decision Tree, Ensemble learning, Random Sample Consensus (RANSAC), Random Forest, and Support Vector Machines (SVMs), respectively, as the basis for classifiers to predict 365-day mortality. Ensemble learning is performed using multiple machine-learning methods to approximate better predictive performance than any one method could achieve independently. Through multiple machine-learning algorithms, ensemble learning uses ensemble methods to obtain predictions more effectively than any one alone. In this study, we used Decision Tree, Random Forest and SVMs to obtain predictions. The Decision Tree classifier is a non-parametric supervised learning method used for classification and regression [20]. In this study, we used an optimized version of the Classification and Regression Three algorithm. RANSAC is an iterative method used to estimate parameters of a mathematical model from a set of observed data, in which the maximum residual for a data sample is classified as being an inlier (Alive) if it is lower than the residual threshold, otherwise the sample is classified as being an outlier (Death) [21]. The Random Forest classifier is a powerful machine learning classifier, which has the key advantages of a having non-parametric nature, high classification accuracy, and the capability to determine variable importance [22]. SVMs are a popular machine learning methods for classification (support vector classification, SVC), regression (support vector regression, SVR), and other learning tasks. The SVC classifier is an implementation of SVM for a linear kernel which is based on LIBSVM [23]. Cross-validation (CV) is a popular strategy to select an algorithm. The main idea behind CV is to split data, once or several times, to estimate the risk of each algorithm. Part of the data (the training sample) is used to train each algorithm, and the remaining part (the validation sample) is used to estimate the risk of the algorithm. CV is then used to
select the algorithm with the smallest estimated risk. Compared to resubstitution errors, CV avoids overfitting because the training sample is independent from the validation sample (at least when the data are independent and identically distributed). In this study, we used Leave-One-Out to evaluate the models [24].

**Statistical Analysis**

Continuous variables from prospective data are summarized as means and standard deviations unless otherwise stated. The Student's t-test was used to compare the means of continuous variables and normally distributed data; otherwise, the Mann-Whitney U-test was used. Categorical data were tested using the chi-square test. Correlations of paired-group results of the life expectancy models were assessed using linear regression and Spearman rank correlation test. Calibration was assessed using the Hosmer-Lemeshow goodness-of-fit test to compare the number of observed and predicted deaths in the risk groups for the entire range of death probabilities.

Discriminatory predictive power was assessed using the area under the receiver operating characteristic curve (AUROC). Pairwise comparisons of the AUROC were performed using the Hanley and McNeil method. AUROC analysis was also performed to calculate the sensitivity, specificity, and percentage of correctness of the prognosis models and CLIF-SOFA scores. Finally, cut-off points were calculated by obtaining the best Youden index (sensitivity + specificity – 1) [27]. Correlations of paired-group variables were assessed using linear regression and Spearman rank correlation test. Cumulative survival curves as a function of time were plotted using the Kaplan-Meier approach and were compared using the log-rank test. All of the statistical tests were two-tailed, and a value of \( P < 0.05 \) was considered to be statistically significant. Data were analyzed with the Statistical Package for the Social Sciences software, version 19.0 (SPSS Inc., Chicago, IL, USA), Python 3.6 with a machine learning package (scikit-learn version 0.18), and Matrix Laboratory (Matlab version R2016.a) for Windows.

**Results**

**Patient Characteristics**

We enrolled 294 cirrhotic patients who were admitted to the specialized hepatogastroenterology ICU at our institution from September 2010 to August 2013. Overall, the in-hospital and 365-day mortality rates were 55.4% (163/294) and 82.3% (242/294), respectively. Table 1 compares the demographic data and clinical characteristics of the 365-day survivors and non-survivors.

Some patient characteristics and some prognostic predictors such as Age, Length of ICU stay (days), Length of Hospital stay (days), and CLIF-SOFA, etc. were compared between the study groups using Mann-Whitney test; Gender (M/F), Use of Terlipressin (Yes/No), Oliguria (Yes/No), and Acute renal failure (Yes/No), etc. were compared between the study groups using Chi-square test; Coagulopathy (Yes/No), Hepatorenal syndrome (Yes/No), Admission due to respiratory failure (Yes/No), and
History of alcoholic plus hepatitis C (Yes/No) were compared between the study groups using Fisher’s exact test. The mean age of the patients was 58 years; 219 patients were men (74.5%) and 75 were women (25.5%). The mean length of ICU stay was 9 days. There was no significant difference in the age or gender between the survivors and non-survivors.

Table 1 | Patients’ demographic data and prognostic predictors according to 365-day mortality

| Feature Selection (LASSO) identified the most relevant 15 features according to 365-day mortality |

<table>
<thead>
<tr>
<th></th>
<th>All patients (n = 294)</th>
<th>Survivors (n = 52)</th>
<th>Non-survivors (n = 242)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>58 ± 14</td>
<td>57 ± 13</td>
<td>59 ± 14</td>
<td>NS (.341)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Gender (M/F)</td>
<td>219/75</td>
<td>41/11</td>
<td>178/64</td>
<td>NS (.428)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Length of ICU stay (days)</td>
<td>9 ± 9</td>
<td>6 ± 6</td>
<td>9 ± 10</td>
<td>.0069&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Length of hospital stay (days)</td>
<td>25 ± 24</td>
<td>22 ± 20</td>
<td>26 ± 24</td>
<td>NS (.232)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>CLIF-SOFA (mean ± s.d.)</td>
<td>12.0 ± 5.3</td>
<td>7.1 ± 3.3</td>
<td>13.1 ± 5.1</td>
<td>&lt; .0001&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Feature selection (LASSO) identified the most relevant 15 features according to 365-day mortality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Terlipressin (Yes/No)</td>
<td>43/251</td>
<td>20/32</td>
<td>23/219</td>
<td>&lt; .0001&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Serum albumin (g/dL)</td>
<td>2.4 ± .6</td>
<td>2.7 ± .5</td>
<td>2.4 ± .6</td>
<td>0.002&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Serum creatinine (mg/dL)</td>
<td>2.7 ± 2.2</td>
<td>1.2 ± 1.0</td>
<td>3.0 ± 2.3</td>
<td>&lt; .0001&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Oliguria (Yes/No)</td>
<td>101/193</td>
<td>19/33</td>
<td>82/160</td>
<td>NS (.715)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Coagulopathy (Yes/No)</td>
<td>24/270</td>
<td>0/52</td>
<td>24/218</td>
<td>0.0111&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Hepatorenal syndrome (Yes/No)</td>
<td>27/267</td>
<td>0/52</td>
<td>27/215</td>
<td>0.0066&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Acute renal failure (Yes/No)</td>
<td>191/103</td>
<td>16/36</td>
<td>175/67</td>
<td>&lt; .0001&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Ascites (No/Grade I/Grade II/ Grade III)</td>
<td>119/162/3/10</td>
<td>31/20/0/1</td>
<td>88/142/3/9</td>
<td>.0191&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Hepatic encephalopathy (0/1/2/3/4)</td>
<td>50/70/36/36/102</td>
<td>25/16/7/4/0</td>
<td>25/54/29/32/102</td>
<td>&lt; .0001&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Admission due to respiratory failure (Yes/No)</td>
<td>21/273</td>
<td>1/51</td>
<td>20/222</td>
<td>NS (.141)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Admission due to upper gastrointestinal bleeding (Yes/No)</td>
<td>199/95</td>
<td>33/19</td>
<td>62/180</td>
<td>&lt; .0001&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Source of admission (0=general ward 1=transferred from other hospital)</td>
<td>178/116</td>
<td>20/32</td>
<td>158/84</td>
<td>.0003&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>History of hepatitis B (Yes/No)</td>
<td>119/175</td>
<td>11/41</td>
<td>108/134</td>
<td>.0018&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>History of hepatocellular carcinoma (Yes/No)</td>
<td>98/196</td>
<td>9/43</td>
<td>89/153</td>
<td>.0070&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>History of alcoholic plus hepatitis C (Yes/No)</td>
<td>7/287</td>
<td>3/49</td>
<td>4/238</td>
<td>NS (.108)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

* a: Mann-Whitney test; b: Chi-square test; c: Fisher’s exact test

**Risk Predictors for 365-day Mortality**

The LASSO method identified that 15 of the 141 prospective predictor variables (Table 1) were good prognostic predictors, including the source of admission, two reasons for ICU admission (respiratory failure, upper gastrointestinal bleeding), a medical history of three conditions (hepatitis B, alcoholism plus hepatitis C, hepatocellular carcinoma), five comorbidities (ascites, coagulopathy, acute renal...
failure, hepatorenal syndrome, hepatic encephalopathy), two laboratory results (serum albumin, serum creatinine), one medication (terlipressin), and one physiological variable (oliguria). Calibration of the 15 variables and CLIF-SOFA scores are listed in Table 2.

Table 2. Calibration of the 16 predictor variables: predicted and observed 365-days mortality by risk groups

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>N ≥ 55 years old (Alive)</th>
<th>Observed (Dead)</th>
<th>Predicted (Dead)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>29</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>29</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>29</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>29</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>29</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>28</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>26</td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>21</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>17</td>
<td>9</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>5</td>
<td>11</td>
<td>5</td>
</tr>
</tbody>
</table>

Chi-squared 1.9755
Degree of freedom 8
Significance level $P = .9818$

AUROC analysis was used to assess the discriminatory power of the 15 variables and CLIF-SOFA scores. CLIF-SOFA score (c-statistic: 0.841), hepatic encephalopathy grade (c-statistic: 0.818) and serum creatinine (c-statistic: 0.788) had the best predictive performance (Table 3).

Table 3. Comparison of AUROC to predict 365-day mortality using predictors evaluated on the first day of ICU admission

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>AUROC</th>
<th>Confidence Interval</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLIF-SOFA</td>
<td>.841</td>
<td>.794-.880</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Hepatic encephalopathy grade</td>
<td>.818</td>
<td>.769-.860</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Serum creatinine</td>
<td>.788</td>
<td>.737-.833</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Acute renal failure</td>
<td>.708</td>
<td>.652-.759</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Admission due to upper gastrointestinal bleeding</td>
<td>.689</td>
<td>.633-.742</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Use of Terlipressin</td>
<td>.645</td>
<td>.587-.699</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Serum albumin</td>
<td>.636</td>
<td>.578-.691</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Source of admission</td>
<td>.634</td>
<td>.576-.689</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>History of hepatitis B</td>
<td>.617</td>
<td>.559-.673</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Ascites</td>
<td>.602</td>
<td>.562-.676</td>
<td>.0015</td>
</tr>
<tr>
<td>Hepatocellular carcinoma</td>
<td>.597</td>
<td>.539-.654</td>
<td>.0015</td>
</tr>
<tr>
<td>Hepatorenal syndrome</td>
<td>.556</td>
<td>.497-.613</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Coagulopathy</td>
<td>.55</td>
<td>.491-.607</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Admission due to respiratory failure</td>
<td>.532</td>
<td>.473-.590</td>
<td>.0154</td>
</tr>
<tr>
<td>History of alcoholic plus hepatitis C</td>
<td>.521</td>
<td>.462-.579</td>
<td>.2214</td>
</tr>
<tr>
<td>Oliguria</td>
<td>.513</td>
<td>.455-.572</td>
<td>.7198</td>
</tr>
</tbody>
</table>
Comparisons of the Various Life Expectancy Models and CLIF-SOFA Scores

The 15 predictors and CLIF-SOFA scores were further incorporated to develop 365-day life expectancy models through machine learning methods. The predictive accuracies of the models based on SVM, Random Forest, RANSAC, Ensemble, and Decision Tree methods are listed in Table 4. Comparisons between the discriminatory values of the five models and CLIF-SOFA scores are also listed in Table 4. The SVM method was an excellent life expectancy model for predicting patient outcome with higher discrimination (c statistic: 0.901, 95% CI: 0.861-0.933) than the Ensemble method (c statistic: 0.885, 95% CI: 0.843-0.919), RANSAC method (c statistic: 0.897, 95% CI: 0.857-0.929), Random Forest method (c statistic: 0.892, 95% CI: 0.850-0.925) and CLIF-SOFA score (c statistic: 0.841, 95% CI: 0.794-0.880). Based on analysis of the AUROC curves, the discriminatory power of the SVM model was the best and significantly better than those of CLIF-SOFA score and Decision Tree method.

Table 4. Pairwise comparison of AUROC for prediction of 365-day mortality using machine learning methods and scoring method measured or calculated on the day 1 of ICU

<table>
<thead>
<tr>
<th>Basis Method</th>
<th>AUROC</th>
<th>Confidence Interval</th>
<th>P</th>
<th>AUROC Difference</th>
<th>Standard Error</th>
<th>Confidence Interval</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLIF-SOFAa</td>
<td>.841</td>
<td>.794-.880</td>
<td>&lt; .001</td>
<td>.0608</td>
<td>.0250</td>
<td>.0188-.1100</td>
<td>.0150</td>
</tr>
<tr>
<td>Random-</td>
<td>.892</td>
<td>.850-.925</td>
<td>&lt; .001</td>
<td>.0097</td>
<td>.0190</td>
<td>-.0275-.0470</td>
<td>.6084</td>
</tr>
<tr>
<td>Forest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RANSACb</td>
<td>.897</td>
<td>.857-.929</td>
<td>&lt; .001</td>
<td>.0042</td>
<td>.0242</td>
<td>-.0433-.0517</td>
<td>.8621</td>
</tr>
<tr>
<td>Ensemblec</td>
<td>.885</td>
<td>.843-.919</td>
<td>&lt; .001</td>
<td>.0160</td>
<td>.0167</td>
<td>-.0168-.0487</td>
<td>.3394</td>
</tr>
<tr>
<td>Decision</td>
<td>.808</td>
<td>.758-.851</td>
<td>&lt; .001</td>
<td>.0933</td>
<td>.0358</td>
<td>.0231-.164</td>
<td>.0092</td>
</tr>
<tr>
<td>Tree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVMd</td>
<td>.901</td>
<td>.861-.933</td>
<td>&lt; .001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AUROC, area under the receiver operator characteristic curve

a CLIF-SOFA, Chronic Liver Failure – Sequential Organ Failure Assessment

b RANSAC, Random sample consensus
c Ensemble, Combination of the predictions of SVM, Random Forest and Decision Tree built with a method
d SVM, Support vector machine
e All six prognostic models were significant in predicting 365-day mortality. The accuracy of the SVM method was the best and significantly better than CLIF-SOFA scores and Decision Tree method.

To assess the validity of these methods, we tested the sensitivity, specificity and overall correctness of prediction at cut-off points that provided the best Youden index (Table 5). The model based on RANSAC had the best Youden index, and the Ensemble method had the highest overall correctness of prediction.

Table 5. Prediction of subsequent 365-day mortality on the first day of ICU admission

<table>
<thead>
<tr>
<th>Method</th>
<th>Cut-off Valued</th>
<th>Youden Indexe</th>
<th>Sensitivity(%)</th>
<th>Specificity(%)</th>
<th>Overall Correctness(%)b</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLIF-SOFAa</td>
<td>≤ 11</td>
<td>.6132</td>
<td>92.31</td>
<td>69.01</td>
<td>73.13</td>
</tr>
<tr>
<td>Random Forest</td>
<td>≤ .6848</td>
<td>.6462</td>
<td>92.31</td>
<td>72.31</td>
<td>75.85</td>
</tr>
<tr>
<td>RANSACb</td>
<td>≤ .4059</td>
<td>.6958</td>
<td>92.31</td>
<td>77.27</td>
<td>79.93</td>
</tr>
<tr>
<td>Method</td>
<td>≤</td>
<td>.6734</td>
<td>.6410</td>
<td>82.69</td>
<td>81.40</td>
</tr>
<tr>
<td>----------</td>
<td>-----</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Ensemble</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision Tree</td>
<td>≤</td>
<td>.4973</td>
<td>.5529</td>
<td>78.85</td>
<td>76.45</td>
</tr>
<tr>
<td>SVM</td>
<td>≤</td>
<td>.7089</td>
<td>.6848</td>
<td>90.38</td>
<td>78.10</td>
</tr>
</tbody>
</table>

\(^a\) CLIF-SOFA, Chronic Liver Failure – Sequential Organ Failure Assessment
\(^b\) RANSAC, Random sample consensus
\(^c\) Ensemble, Combination of the predictions of SVM, Random Forest and Decision Tree built with a method
\(^d\) SVM, Support vector machine
\(^e\) All six prognostic models were significant in predicting 365-day mortality. The accuracy of the SVM method was the best and significantly better than CLIF-SOFA scores and the Decision Tree method.
\(^f\) Optimal cut-off points for predicting for 365-day mortality were derived from receiver operator characteristic analysis. \(^g\) The RANSAC method had the best Youden index.
\(^h\) The Ensemble method had the highest overall correctness of prediction

Figures 1 to 3 show that the model based on SVM had strong and positive correlations in 365-day mortality with the Ensemble method \((r^2 = 0.796, P < .001)\), RANSAC method \((r^2 = 0.547, P < .001)\) and Random Forest method \((r^2 = 0.678, P < .001)\).
Figure 1. Correlations of SVM and Ensemble methods for 365-day mortality\textsuperscript{a,b}.

\textsuperscript{a} SVM was significantly correlated ($r^2 = 0.796$, $P < .001$) with the Ensemble method for the whole group and the non-survivor groups (Total group, $n = 294$; Death group, $n = 242$, Alive group, $n = 52$).

\textsuperscript{b} SVM, support vector machine; Ensemble, combination of the predictions of SVM, Random Forest and Decision Tree built with a classifier

Figure 2. Correlations of SVM and RANSAC methods for 365-day mortality\textsuperscript{a,b}.

\textsuperscript{a} SVM was significantly correlated ($r^2 = 0.547$, $P < .001$) with RANSAC for the whole group and non-survivor groups (Total group, $n = 294$; Death group, $n = 242$, Alive group, $n = 52$).

\textsuperscript{b} SVM, support vector machine; RANSAC, Random sample consensus

Figure 3. Correlations of SVM and Random Forest methods for 365-day mortality\textsuperscript{a,b}.
SVM was significantly correlated ($r^2 = 0.678$, $P < .001$) with Random Forest for the whole group and non-survivor groups (Total group, $n = 294$; Death group, $n = 242$, Alive group, $n = 52$).

Figure 4 illustrates that the cumulative survival rates differed significantly for the patients classified by the cut-off value of the SVM model on the first day of ICU admission.

Figure 4. Survival functions for SVM.

\[ \text{Survival Function} \]

\[ \begin{array}{c}
\text{Days} \\
0 & 30 & 60 & 90 & 120 & 150 & 180 & 210 & 240 & 270 & 300 & 330 & 360 \\
\text{Percent survival} & 1 & 0.8 & 0.6 & 0.4 & 0.2 & 0.0 \end{array} \]

\[ \text{SVM score is less than or equal to 0.7089} \]

\[ \text{SVM score is more than 0.7089} \]

\[^a\] Kaplan-Meier survival analysis of 294 critically ill cirrhotic patients according to SVM score on the first day of ICU admission. Cumulative survival rates differed significantly for the patients with a SVM score $\leq 0.7089$ ($n = 100$) and those with a SVM score $> 0.7089$ ($n = 194$) on the first day of ICU admission.

**Discussion**

**Principal Results**

The overall 365-day mortality rate was 82.3% (242/294). CLIF-SOFA score, hepatic encephalopathy grade and serum creatinine were the strongest predictive variables for 365-day mortality in this investigation (Table 3). We developed a series of life expectancy models including CLIF-SOFA score and the 15 most relevant predictors (hepatic encephalopathy grade, serum creatinine, serum albumin, admission due to upper gastrointestinal bleeding, admission due to respiratory failure, use of terlipressin, source of admission, history of hepatitis B, ascites, history of hepatocellular carcinoma, history of alcoholism plus hepatitis C, presence of acute renal failure, presence of hepatorenal syndrome, presence of coagulopathy, presence of oliguria) that successfully predicted 365-day mortality of the critically ill cirrhotic patients (Table 1). The predictive accuracy of the models based on SVM had the highest discriminative power (Table 4). Moreover, the RANSAC method had the best Youden index, and the Ensemble method had the best overall correctness of prediction (Table 5).

Life expectancy models can be inaccurate if imperfect prospective data are used. More accurate indices often include additional critical information, however this can lead to difficulties if there is too much raw data from a prospective study or if the data are high dimensional. Using high dimensional prospective data to predict life
expectancy in critically ill cirrhotic patients could address the limitations of existing scoring systems. However, analyzing a large amount or high dimensional clinical data is challenging. Therefore, we used machine learning methods to generate meaningful predictors from prospective data and developed highly accurate and clinically actionable 365-day life expectancy model. There were positive strong correlations between the models based on SVM, Ensemble, RANSAC and Random Forest (Figures 1-3). The cumulative survival rates differed significantly when classified by the SVM model (Figure 4). The results also showed that the predictive abilities of the SVM, RANSAC, Random Forest and Ensemble models were stronger than CLIF-SOFA scores (Table 5), which we previously found was the best model available [4].

In addition to CLIF-SOFA score, hepatic encephalopathy grade, and serum creatinine were the other two critical risk factors affecting outcomes in this study. Several studies have reported that the occurrence of hepatic encephalopathy in ACLF indicates a severe systemic inflammatory response and increased risk of cerebral edema and mortality [30]. Acute renal failure is a common complication and has been correlated with an extremely poor prognosis in critically ill cirrhotic patients [28-29]. Interestingly, we found that the predictive performance of serum creatinine (c-static: 0.788) was better than that of acute renal failure (c-static: 0.708), which was defined as Kidney Disease Improving Global Outcomes (KDIGO) stage 3 according to serum creatinine and urine output criteria [28]. In addition, the predictive performance of oliguria (c-static: 0.513) was significantly lower than that of serum creatinine (c-static: 0.788) in our study. This result is consistent with the consensus of the International Club of Ascites, in that the amount of urine is not a good indicator to assess renal function in cirrhotic patients. With regards to the nature of the disease, patients with cirrhosis and ascites may have a reduced amount of urine with avid sodium retention while maintaining a relatively normal glomerular filtration rate [29]. Taking both serum creatinine and urine amount into account might, at least partially, explain why the predictive performance of acute renal failure was relatively lower than that of using serum creatinine alone in this study.

In cirrhotic patients, complications requiring emergency care and ICU admission often occur leading to a vicious cycle. The cascade of events may even be life threatening, even if the patients receive the best possible technological support [17]. Even though incorporating accurate life expectancy predictions into clinical decision making could improve the quality of care and decrease costs, few clinicians actually do this – perhaps because existing life expectancy models are inaccurate and/or difficult to perform. In recent years, machine learning techniques have successfully linked the assessment of prospective data to diagnoses and appropriate clinical decision making [31-34]. The prognostic predictors identified through machine learning can provide highly discriminative information without being burdensome, and therefore clinicians should feel comfortable using this highly discriminative, well calibrated, non-burdensome critical prognostic method to assist with clinical decision making.
To date, prospective studies of critically ill cirrhotic patients have usually included a study period of 180 days but have lacked data on the long-term (365-day) prognosis and prognostic predictors that can be used for clinically useful decision making for personalized medicine. Research has shown that different properties of the underlying distributions of variables define whether they will be good predictive or significant variables [35]. If prediction is the goal, significance should not be the only selection standard. In this study, we offer an alternative approach that was not focused on significance.

The main strength of this study is the use of models based on machine learning validated by leave-one-out cross-validation using prospective data, which demonstrated the clinical utility of the SVM/Ensemble/RANSAC/Random Forest methods in predicting survival [36-39]. We believe that our results may have positive clinical implications, because the information provided is objective and can be used in family meetings to help set expectations for reasonable clinical care. Our life expectancy models could thus be used to optimize ICU practice. Over half of our patients ≥ 55 years old had a predicted 365-day mortality of < 50% and were likely to benefit from intensive care practices, despite their advanced age.

Limitations

In spite of the encouraging results observed in our study, several potential limitations should also be considered. First, the study was conducted on patients from just one academic tertiary care medical center, which limits the generalizability of our findings. Our results may be unsuitable for direct extrapolation to other hospitals with different patient populations. Second, we hepatitis B viral infection was the leading cause of liver cirrhosis in our patients. This means that our results cannot be applied to patients with liver disorders in North America and European countries because liver diseases in these regions are largely attributed to hepatitis C viral infection or alcoholism.

Conclusions

In summary, our 365-day life expectancy models based on SVM/Ensemble/RANSAC/Random Forest methods outperformed both the CLIF-SOFA scoring system and life expectancy model based on decision trees. In addition, our life expectancy models were more discriminative and less burdensome than similar scoring systems reported in critically ill cirrhotic patients.
Acknowledgements

Authorship
Guarantor of the article: Wei-Huan Hu.
Author contributions: YCC and MHT contributed to the conception, design and interpretation of data. HCP provided patient information, participated in the design and coordination, and helped to draft the manuscript. YJH and SHH provided intellectual content to the study and were involved in editing and revising the manuscript. All of the authors discussed, contributed to and approved the final version of the manuscript.

Declaration of Personal Interests
None.

Declaration of Funding Interests
This study was supported by the Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. MOST 106-2321-B-182-002.

Conflicts of Interest
None declared.

Abbreviations
ACLF: acute-on-chronic-liver-failure
APACHE: acute physiology and chronic health evaluation
AUROC: area under the receiver operating characteristic curve
CLIF-SOFA: chronic liver failure-sequential organ failure assessment
CV: cross-validation
ICU: intensive care unit
KDIGO: kidney disease improving global outcomes
LASSO: least absolute shrinkage and selection operator
MELD: The model for end-stage liver disease
RANSAC: random sample consensus
SOFA: sequential organ failure assessment
SVC: support vector machine for classification
SVM: support vector machine
SVR: support vector machine for regression

References


