**SUPPORTING INFORMATION**

***Data Preprocessing***

Before the AI models could be built and used, the readmission and death dataset needed to be preprocessed for analysis. This preprocessing consisted of two primary steps: cleaning the dataset and then mining for new features. Cleaning the dataset involved handling incorrect, redundant, or missing values in the data, while feature mining involved identifying new features from the data.

Data cleaning was important to validate the quality of the data prior to its use in AI analysis. Mis-recorded information was carefully adjusted under the supervision of the clinical investigators to adjusted units for laboratory values. Finally, only labeled data was kept because supervised learning was performed with the AI models. 1-month readmission data was available for 2010 out of 2170 samples. All 2170 samples were labeled for 3-month death, but only 2077 were labeled for 3-month readmission. Thus, 110 samples were dropped for the 1-month readmission analysis and 93 samples were dropped for 3-month readmission analysis. Once the data cleaning was complete, some feature mining was performed. For the development of AI models, dummy variables were created for the nominal variables of race and cirrhosis etiology so that they could be fed into the algorithms. For a nominal variable with k categories, only k-1 dummy variables were created to avoid high multicollinearity.

***Model Development***

The three AI techniques we chose were Logistic Regression (LR), Support Vector Machine (SVM) and Random Forest classifier (RFC), which all use different and common AI algorithms.

LR falls into the category of generalized linear models. It fits a linear combination of the original features onto a sigmoid curve. The sigmoid function has the value range [0,1], which means that the output on the curve can be interpreted directly as a probability. To encourage better generalizability to unseen data, we added an L2 regularization term (with the inverse of regularization strength C=1.0) to the LR formulation to penalize large coefficients for its linear transformation.

An SVM model estimates the optimal hyperplane to separate data from different outcome classes with the maximum margin. The maximum-margin principle contributes to SVM’s generalizability for unseen data. When the outcome classes are not linearly separable, kernel tricks are usually applied to implicitly map the original features to a higher dimensional space for greater separability. We used the popular radial basis function (RBF) kernel, with the kernel coefficient set to be the inverse of the number of predictors. An L2 regularization term (C=1.0) was also added.

The RFC model consists of multiple decision trees that all independently make binary predictions using random subsets of the features and training samples. Ideally, all training samples are independent and all features are independent. If each individual decision tree is able to make a prediction better than random chance, then the ensemble learning nature of the RFC will make its overall performance grow as the number of trees grows. To combat the fact that the independence assumptions among samples and features are not fully realized, we adjusted the number of trees to be 20 and the depth of each tree to be 5.

***Model Evaluation***

We used 10-fold cross-validation to evaluate the performance of our AI models. In each of the 10 folds, the dataset was randomly shuffled and split to 90% training samples and 10% testing samples. All features were standardized based on mean and standard deviation computed from the training data. The statistical tests were performed on the training samples, and only significant variables (p < 0.05) were chosen as predictors. Patients with incomplete records of the chosen predictors from both training and testing were dropped at this point. We used area under the receiver operating characteristic curve (AUC) as our evaluation metric for the testing data. For each of the testing samples with each of the algorithms we used (LR, SVM, and RFC), their classification labels were decided by thresholding their class probability estimates. AUC is the area under the curve obtained by varying this threshold and plotting the true positive rate against the false positive rate. A classifier performing as well as random chance will have an AUC of around 0.5, while a perfect classifier will have an AUC of 1.