Cirrhosis Mortality Prediction as Classification

Chenxi Ge: Department of Biomedical Engineering, Northwestern University, Evanston, IL, USA

Diego Klabjan: Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, USA

Daniela Ladner, Parker Stride: Transplant Outcomes Research Collaborative (NUTORC), Comprehensive Transplant Center, Northwestern University, Chicago, IL, USA

1. Background

Cirrhosis is a condition in which the liver slowly deteriorates and is unable to function normally due to a chronic injury, as defined by The National Institute of Diabetes and Digestive and Kidney Diseases.[1] Currently, liver transplantation serves as an effective treatment of liver cirrhosis, but the number of livers requiring transplantation largely exceeds the number of organ donors. To prioritize recipients of liver transplantation, the mortality of each cirrhotic patient is predicted by the Model of End-stage Liver Disease, or MELD, index which uses the serum bilirubin level, serum creatinine level and international normalized ratio as components.[2]

The MELD model is proven clinically useful,[3-5] but new researches point out that serum levels of other indicators could help the mortality prediction,[6] and for this reason we postulate that using a much more comprehensive feature domain could result in a higher prediction accuracy of mortality.

Novel data analysis methods like machine learning facilitate the exploration of such a feature domain. Machine learning models learn patterns from past data and then make predictions on unseen data. They are widely used in both academia and industry, and show great potential in medical research.

In this study, we collect 2,322 cirrhotic patients from the Northwestern Enterprise Data Warehouse and apply different machine learning models to predict if cirrhotic patients die within one year. The most significant features in the model are their last MELD score, their T0 age, and if they have Current Procedural Terminology (CPT) code 75894. We take one further novel step to analyze temporal trends of patient's physiological status. Conventionally, only snapshots (e.g., last

record, first record) or statistical values (e.g., mean, median, maximum, minimum) of historical records were used for such analyses. We use Frequent Subgraph Mining (FSM) to capture the trends in physiological measurements.[7] The detected trends are then used as features to train another model, and the model is ensembled with the model without trends. The FSM-based model and the underlying ensemble model are only used on a subset of the cohort with sufficient physiological measurements.

All our models statistically significantly outperform the traditionally used MELD-score model. Among all of the models, the ensemble model outperforms the model based only on statistical features and this behavior is statistically significant.

Our main contributions are that we much more accurately predict one-year mortality of cirrhotic patients awaiting liver transplant, and evaluate the performance of different machine learning algorithms on this particular task. As the very first study to use machine learning methods to predict one-year mortality of cirrhotic patients, our research shows that machine learning greatly improves the accuracy of prediction and largely outperforms the MELD model. The improvement on average is 10%, with an increase of AUC from 0.822 to 0.904. Another contribution is the novel idea of creating ‘patient slices,’ i.e., from a single patient we create several ‘auxiliary’ patients with the survived label. This enlarges the training data set and improves the quality of the models. The last contribution is a combination of techniques based on FSM and ensemble that yields the best performing model.

2. Prior Work

2.1 Conventional cirrhosis mortality prediction

Several models are used to predict the mortality of cirrhotic patients. Historically, the Child-Pugh score was wildly used for prioritizing the patients awaiting liver transplantation.[8] The Child-Pugh score uses ascites, Hepatic Encephalopathy (HE), serum bilirubin, serum albumin and INR as predictors. Later in 2000, Malinchoc et al. created a model to predict the survival of patients undergoing transjugular intrahepatic portosystemic shunts (TIPS), which takes the serum bilirubin, serum creatinine and INR as predictors.[5] The model is known as Model of End-stage Liver Disease, or MELD, and is considered as a reliable measure of mortality risk in patients with end-stage liver disease. Then in 2004, Ruf et al. proposed that the addition of serum sodium to MELD results in a more accurate way of mortality prediction than MELD alone.[9]

Effective as it is, the MELD score has problems with the prediction of mortality rate in patients with low MELD score (defined as $MELD < 20$), as patients with persistent ascites and a low serum sodium level have a higher than expected mortality rate despite having low MELD scores.[6] Recent studies show that serum sodium or ascites are better predictors of mortality for low MELD patients.[10] No studies have shown whether other clinical predictors or phenotypic patterns can better predict mortality in low MELD cirrhotic patients.

Some of the other factors related to the mortality of cirrhotic patients are: Esophageal Varices,[11] HE,[12] Infection,[12] Hepatorenal Syndrome.[13] Each of these symptoms also have their own factors: glutamine can be used to predict the development of HE,[14] platelet count is accurate in predicting Esophageal Varices when combined with albumin and histologic levels,[15] C-reactive protein can be used to identify Infection.[16] These factors are indirectly related to the mortality of cirrhotic patients.

Knowing that there is a large collection of unexplored factors, we decide to use a comprehensive feature space in our research. The features we consider fall into four main categories: demographic features, comorbidities (co-occurrence of other diseases), clinical procedures and laboratory records. All features used by the MELD model fall in the ‘laboratory records’ category.

2.2 Machine learning in liver disease

As a powerful analytical method, machine learning has been extensively used in liver disease research. One of the main foci is disease diagnosis. Various machine learning approaches have been proven useful in previous studies. Singal et al. 2013 used a random forest model to predict the development of Hepatocellular Carcinomas (HCC) that outperformed the conventional regression models.[17] Sakiyama et al. 2008 compared various machine learning models to predict the human liver microsomal stability and Sartakhti et al. 2012 proposed a machine learning approach to assist in the diagnosis of hepatitis.[18-19] Machine learning was also used to identify the importance of features. In 2003, Ye et al. used supervised machine learning algorithms to investigate the gene expression change associated with HCC, and identified important genes that are relevant to patient survival.[20] While all these studies use machine learning for liver related diseases, none of them focuses on cirrhotic. In addition, we use novel features based on patterns and ensembles of models.

2.3 Frequent Subgraph Mining in pattern recognition

To extract information from historical records, snapshots or statistical measurements are often used. But these approaches fail to recognize the temporal trends of test results. In our study, we first represent the historical laboratory tests for each patient as graphs, then use a subgraph mining method to analyze the change of patient's physiological status. After normalizing of measurements, we use a FSM method to find patterns of physiological change. FSM is an effective pattern recognition method in identifying common structures in graphs, and is used in tracking patient's status with frequently recorded data.[7] By using FSM, we are able to identify patterns like 'serum bilirubin level stable for 6 consecutive months' as features used in our machine learning models.

3. Materials and Methods

3.1 Cohort Definition

For cohort identification and data collection, we access the Northwestern University Electronic Data Warehouse which has patient information from Northwestern hospitals since year 2000. We conclude that data prior to year 2009 is inaccurate, thus we only consider the data from year 2009 to 2014.

Initially, we extract 27,804 patients who are either diagnosed or close to developing cirrhosis. Cirrhotic patients are defined as those that ever had a cirrhosis related International Classification of Diseases code (ICD-9 code) of 571.2/571.5/571.6. Patients that are close to developing cirrhosis are defined as those not having these ICD-9 codes, but with Fibrosis-4 (Fib-4) score higher than 3.25. The Fib-4 score reflects the scar tissue level of the liver, and a threshold value of greater than 3.25 has the specificity of 98% in confirming cirrhosis. We hypothesize that these patients are cirrhotic, but their ICD-9 codes are not correctly recorded. We then apply keyword search in notes and reports among these patients to assert if they are cirrhotic, which is explained later.

The cohort we used is selected from the initial cohort (n = 27,804) using the following criteria: (1) ever had at least one hepatologist visit between year 2009 and 2014, (2) ever had one of the complications defined in Table 1 between year 2009 and 2014, (3) did not have a liver transplantation earlier than year 2010, to make sure at least one year of laboratory records between year 2009 and the transplantation date, (4) did not have liver cancer (identified by having ICD-9 codes of comorbidity 'Solid tumor without metastasis,' see Appendix 1), (5) if the patient is deceased, the cause of death has to be liver disease related.

Ascites, ICD-9 code = {789.5, 789.51, 789.59, 568.82}
Hepatic encephalopathy, ICD-9 code = {572.2, 348.31, 348.30, 348.39, 349.82}
Varices, ICD-9 code = {456.1, 456.2, 456.21, 456, 456.8}
Gastrointestinal bleeding, ICD-9 code = {456.0, 456.20, 578, 578.9, 578.1}
Creatinine > 1.3
Platelets < 150
INR > 1.2
Albumin < 3.5

Table 1. Definition of Complication

	survived 1 year	deceased within 1 year	Overall cohort	p-value
	n = 2,003 (86%)	n = 319 (14%)	n = 2,322 (100%)	
age	59 (53-66)	62 (55-70)	59 (53-66)	<0.01
T0 age	57 (51-64)	61 (54-69)	58 (52-65)	<0.01
Female	857 (43%)	120 (38%)	977 (42%)	0.08
Race				
American Indian or Alaskan Native	5 (1%)	1 (1%)	6 (1%)	0.83
Asian	60 (3%)	8 (3%)	68 (3%)	0.63
Black or African American	196 (10%)	22 (7%)	218 (9%)	0.10
Hispanic	23 (1%)	0	23 (1%)	0.05
Native Hawaiian or Other Pacific Islander	2 (1%)	0	2 (1%)	0.57
Unknown	587 (29%)	85 (27%)	672 (29%)	0.33
White	1130 (56%)	203 (64%)	1333 (57%)	0.02
Ethnic Group				
Hispanic or Latino	258 (13%)	26 (8%)	284 (12%)	0.02
Not Hispanic or Latino	1445 (72%)	202 (63%)	1647 (71%)	<0.01
Unknown	300 (15%)	91 (29%)	391 (17%)	<0.01
Alcohol Use				
Yes	231 (12%)	46 (14%)	277 (12%)	0.14
No	1046 (52%)	83 (26%)	1129 (49%)	<0.01
Unknown	726 (36%)	190 (60%)	916 (39%)	<0.01
Drug Use				
Yes	60 (3%)	9 (3%)	69 (3%)	0.86
No	1078 (54%)	97 (30%)	1175 (51%)	<0.01
Unknown	865 (43%)	213 (67%)	1078 (46%)	<0.01
Smoking Status				
Passive Smoker	11 (1%)	3 (1%)	14 (1%)	0.40
Former Smoker	685 (34%)	98 (31%)	783 (34%)	0.22
Heavy Smoker	217 (11%)	28 (9%)	245 (11%)	0.27
Light Smoker	67 (3%)	3 (1%)	70 (3%)	0.02
Never Smoker	768 (38%)	69 (22%)	837 (36%)	<0.01
Smoker	0	1 (1%)	1 (1%)	0.01
Unknown	255 (13%)	117 (37%)	372 (16%)	<0.01

Table 2. Demographics characteristics of the cohort by different outcome groups. Discrete variables are presented as counts (percentages); continuous variables are presented as mean (25th

– 75th percentile). We define the T0 date as the date that any of the complications in Table 1 first occurred, and T0 age is calculated by subtracting the T0 date with birth date.

A total of 2,322 qualified patients are collected. Whether they deceased within one year after their last laboratory record or not is defined as the outcome. Demographics characteristics of the cohort by different outcome groups are listed in Table 2. Statistical test shows that the alive group and deceased group for many features do not have the same demographics characteristics.

This study has been approved by the Institutional Review Board at Northwestern University (study number: STU00098092).

3.2 Patient identification from textual medical records

As previously mentioned, the initial cohort includes patients with cirrhosis related ICD-9 codes (571.2, 571.5, 571.6), and patients with only Fib-4 score > 3.25. The Fib-4 score helps clinicians estimates the amount of scarring in the liver. For patients with high Fib-4 scores, we hypothesize that some of them were diagnosed with cirrhosis, but their ICD-9 codes were not correctly collected. We believe by analyzing textual clinical notes including radiology and biopsy reports we could find previously unnoticed cirrhotic patients, thus enlarging our cohort size. The textual information we analyze includes CT, MRI and biopsy reports.

We use keywords search. Liver transplant clinicians from Northwestern medicine provided a dictionary with words that are highly related to cirrhosis, and commonly seen misspelled variants of these words (e.g. ‘cirrhosis’ as ‘cirhosis’ or ‘cirrosis,’ and ‘hypertension’ as ‘hypertention)’ to search for.

Finding keywords alone is not enough, as these words can be mentioned with different key phrases, such as ‘not having cirrhosis’ or ‘no evidence of cirrhosis.’ We collect major key phrases and manually decide if the evidence is positive (e.g., ‘show clear evidence of’), negative (e.g., ‘not having’) or ambiguous. We only search for keywords that are paired with positive key phrases (see Appendix 2). We further narrow the search field of key phrase-keyword pairs to certain sections of a report: ‘History,’ ‘Indication’ and ‘Impression.’

With keyword search, 1,095 patients (8.5% of the Fib-4 only group) are identified as cirrhotic. Sensitivity of keyword search is 79.6% according to a test subset that has been manually inspected with precision 100% due to the choice of the data.

3.3 Feature Engineering

Recall that we are predicting if cirrhotic patients die within one year of their last data recorded prior to death.

The features used fall into 4 categories, which are demographic information, comorbidities, clinical procedures and laboratory records.

Demographic features are extracted directly from clinical records of patients. The clinical records include regular demographic information together with select behavioral attributes of patients (e.g., if the patient consumes alcohol).

Comorbidities reflects co-occurrence of other severe diseases that a patient has. Each comorbidity is indicated by a group of ICD-9 codes that a patient has, as shown in Appendix 1. We use binary features to indicate if the patient has a certain comorbidity or not with a total of 45 different comorbidities included.

Clinical procedures may influence a patient's chances of survive, thus they are used as features in our study. All clinical procedures are defined by specific codes, also known as Current Procedural Terminology (CPT) codes. We use binary features to indicate if a patient has had a certain procedure. A total of 24 different procedures are included, as shown in Appendix 3.

As for laboratory items, we take the first value, last value, mean and standard deviation of each item as features. We call these 4 types of features as statistical features.

All of the considered features are listed in Table 3. Note that MELD is one of the features despite being a derived value from other features. It is well known that combined features can improve model performance. We also use a subgraph mining method to analyze temporal trends of laboratory items which yields additional features.

Feature Domains	Individual Variables

Demographics	Age, T0 age, Gender, Race, Ethnic Group, Alcohol Use, Drug Use, Smoking Status
Comorbidity	Alcohol Abuse, Alcohol-related Liver Disease, Ascites, Cardiac Arrhythmias, Cholestasis, Chronic Pulmonary Disease, Deficiency Anemia, Depression, Diabetes Complicated, Diabetes Uncomplicated, Esophageal Varices, Fluid and Electrolyte Disorders, HCV, HE, Hepatic Hydrothorax, Hep B, Hepatopulmonary Disease, HRS, Hypertension Complicated, Hypertension Uncomplicated, Hypothyroidism, Jaundice, Lymphoma, MACE, Malnutrition, Metastatic Cancer, NASH, Obesity, Other Neurological Disorders, Paralysis, Peptic Ulcer Disease Excluding Bleeding, PVD, Portal Hypertension, Psychoses, Pulmonary Circulation Disorders, Renal Failure, RA/CVD, SBP, Valvular Disease
Clinical Procedures	32554, 32555, 32557, 37182, 37204, 37243, 43205, 43227, 43235, 43236, 43243, 43244, 43255, 47120, 47122, 47125, 47130, 49082, 49083, 75894, 77778, 79445
Laboratory Items	Albumin, ALP, AFP Tumor marker, ALT, AST, Bilirubin total, Creatinine, GFR, Hemoglobin, INR, MELD score, Platelet Count, PT, Serum Sodium, White Cell Count

Table 3. Feature Domains and Individual Variables. Age is calculated by subtracting the date of their Last Laboratory record date with birth date; AFP- Alpha Fetoprotein; ALP - Alkaline Phosphatase; ALT - Alanine Aminotransferase; AST - Aspartate Aminotransferase; GFR - Glomerular Filtration Rate; INR - International Normalized Ratio; HCV - Hepatitis C; HE - Hepatic Encephalopathy; Hep B - Hepatitis B; HRS - Hepatorenal Syndrome; NASH - Nonalcoholic Steatohepatitis; SBP - Spontaneous Bacterial Peritonitis; PT - Prothrombin Time; PVD - Peripheral Vascular Disorders; RA/CVD - Rheumatoid Arthritis Or Collagen Vascular Diseases

The subgraph mining algorithm is a pattern mining method to find frequently occurring structures in graphs, thus the first step is to convert historical laboratory records into a graph representation, as shown in Figure 1. We average the records of each laboratory item every 2 months (defined as a node) to get a graph with fixed intervals corresponding to nodes and the averaged values to be the weights of the nodes.

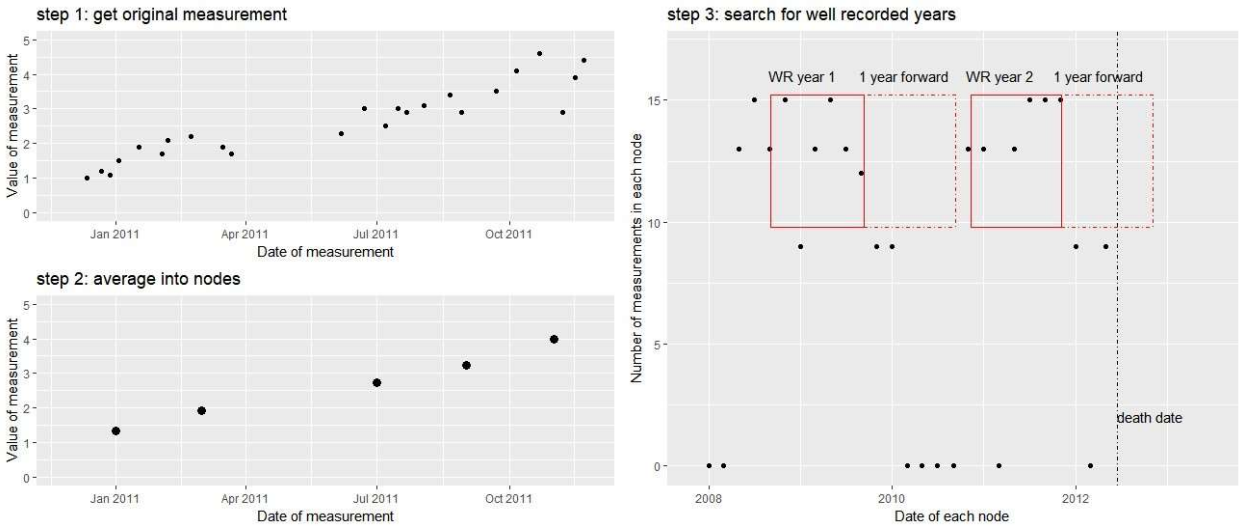

Figure 1. Identification of well recorded years from historical records

However, since laboratory values for a patient are often sparsely recorded and not evenly distributed in time, we decide to only take frequently recorded years of the data. After generating the initial graph, we use a search window with width of 6 nodes (corresponding exactly to one year) to search for well recorded years. We define a well recorded year as having at least 3 nodes in that year with 10 out of 15 laboratory items. (We have 15 laboratory items in total.)

Some patients may have multiple well recorded years. To further utilize the data, we create patient record slices according to well recorded years. For each well recorded year, we duplicate the patient to create an artificial patient or a patient record slice. For such an artificial patient, the features are only considered up to the point in time at the end of the well recorded year. Since our focus is finding one-year mortality, we search forward one year from the end of the well recorded year to decide the outcome of the artificial patient.

By this definition, 1,728 patient record slices out of 1,170 patients with at least one well recorded year are created, and are later used for subgraph mining methods.

3.4 Models

3.4.1 Statistical feature based model

Our statistical feature based model takes demographic records, comorbidity records, clinical procedures and laboratory tests as features. It only uses statistical features to describe the

laboratory tests, hence the origin of the name of the model. The missing values of a laboratory test are imputed by Multivariate Imputation by Chained Equations, or MICE.[22]

The statistical feature based model (Stat model for short) can be trained and tested on all patients, or on patient record slices, mainly for model comparison and ensemble. When referring to this model, we always specify the underlying data set.

3.4.2 Subgraph Enhanced model

For patient record slices which are created by well recorded years, in addition to the features used by the Stat model, we also use an FSM method to find patterns in the change of the laboratory records, and use the patterns as features. FSM is a method for graph pattern recognition. Intuitively, patients with similar physiological conditions share similar trajectories of laboratory records. Thus, the occurrence of certain patterns could be used to identify patients with certain physiological conditions. We use the Subgraph Enhanced model (SE model for short) only on patient record slices (since for others there is not enough laboratory data to perform FSM).

We next provide further details specifically on the subgraph mining process.

As previously mentioned, the historical laboratory records are first converted to graphs. However, the definition of a well recorded year does not require a year to be fully recorded, so we first use MICE imputation to impute the missing values in the nodes, to make sure each graph has 6 nodes, and we have all 15 laboratory values.

Next, we discretize the node values to get the graph representation. To do this, we use a customized z-score, where all values within the normal range are considered as 0. For values larger than the upper bound of the normal range, we use H1 and H2 to represent the 33% and 66% percentile. A value larger than H2 is considered as 3, between H1 and H2 is considered as 2, and between the upper bound of the normal range and H1 is 1. For values smaller than the lower bound of the normal range, we apply the same approach except we use -3, -2, -1 with -3 representing the lowest 33% of the values.

Subgraph miner MoSS is then used to identify frequent subgraphs among all graphs.[23] A total of 2,907 subgraphs that occur with an empirically chosen frequency are found. Note that when the miner captures a frequent subgraph, all its subgraphs are also identified. Thus, graphs with smaller structures outnumber those with larger structures. We use the strategy that if a subgraph occurs in a patient, all subgraphs of it are not considered for this patient. After mining, for each patient we have a set of

frequent subgraphs. In other words, we obtain a matrix with rows corresponding to patients and columns to subgraphs. The value in the matrix is the count. This matrix has a large number of subgraphs, i.e., columns.

We take one further step by applying Non-Negative Matrix Factorization (NMF) to this matrix to construct latent groups of subgraphs. NMF is a clustering method that is efficient in grouping subgraphs by different patient groups.[7,24] An occurrence of each group is used as a feature in the SE model. We empirically choose to group the subgraphs into 20 groups, and the value of the corresponding feature is obtained from the NMF factorization.

3.4.3 Ensemble Model

For patient record slices, the predictions can be made by either the Stat or SE model. We ensemble the two models to further improve the prediction accuracy. Ensemble methodology is to build a predictive model by integrating multiple models and is well known for improving prediction performance. We ensemble our models as follows. For each patient record slice, we use the Stat and SE models to separately calculate the probability of death. We then consider a weighted average of the results. The weights between the two models are decided by 10-fold cross-validation to get the highest prediction accuracy (on validation sets). We expect the ensemble model to have a better performance than either model since the weights can always be selected towards the best model of the two.

4. Results

4.1 Model training and testing

For evaluation, we follow the standard process of model training and testing by using 10-fold cross validation. For the ensemble model we further set aside the validation subset from the training data.

For the Stat model, we first standardize the features (subtract the mean and divided by standard deviation over the training data). We then apply feature selection method mRMR to optimize the model performance while also preserving the features for model interpretability (instead of using principal component analysis).[25] Several supervised machine learning algorithms are trained to predict the outcome, including Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting Classifier (GBC), and Artificial Neural Network (ANN).

The same procedure is followed during the training and testing of the SE model, except before training, we deal with the imbalance of the two classes. The patient record slice has a one-year mortality of 6.4%. We use the oversampling method SMOTE to synthesize patients who died within a year after their well recorded year.[26] We empirically choose the final balance ratio to be 10%.

4.2 Performance evaluation

The performance evaluation is based on how accurate the model predicts the outcome of a patient given specified features. The Area Under the receiver operating characteristics Curve, or AUC, is used as the criterion, as it is a commonly used measurement of the accuracy of discrimination performance. For each fold, we separately calculate the AUC on test, and then use a t-test to compare the results of the different models.

For the pure MELD model, we use MELD as the only feature within the logistics regression model, and then calculate AUC. Note that this is equivalent to using a threshold based on the MEDL score except that the use of the single feature logistic regression model enables the use of out-of-the-box AUC.

4.3 Model comparison

4.3.1 Stat Model on original cohort

The overall AUC performance of different algorithms we use in the Stat model and the baseline MELD model are shown in Figure 2. The model is applied on the entire cohort.

The t-test shows all models outperform the MELD model which has the AUC of 0.822. Based on the t-test among the algorithms, LR is statistically significantly better than others, followed by ANN and SVM. The performance of RF and GBC algorithms are statistically significantly indifferent, and are the worst of the five.

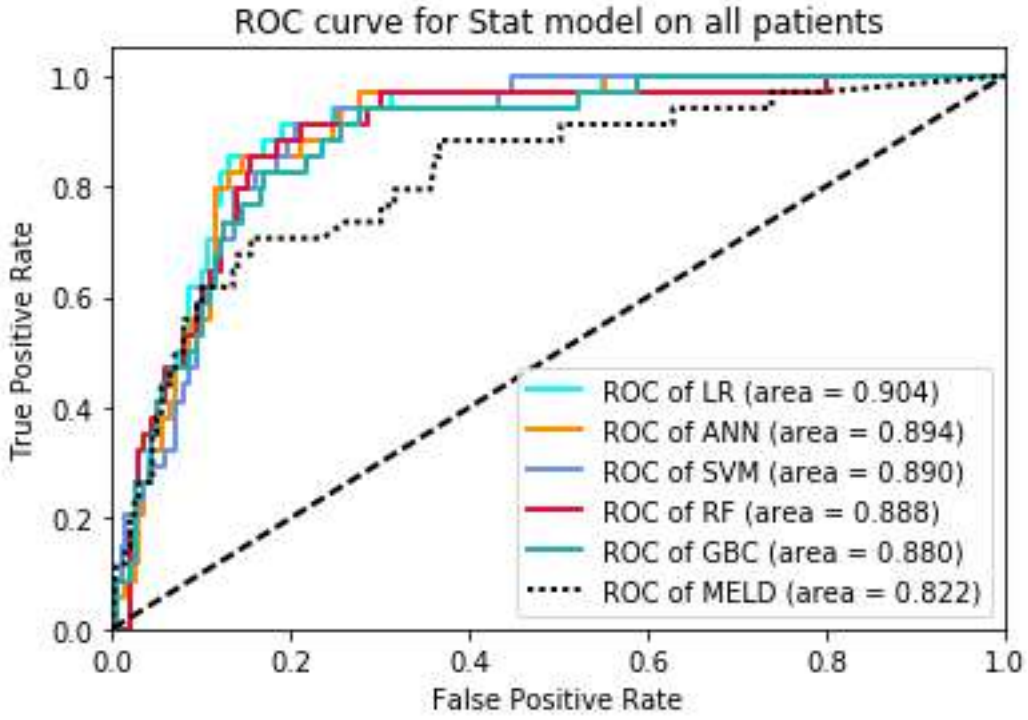

Figure 2. ROC curve for Stat model on all patients

4.3.2 SE Model on patient record slices

The AUC performance of the different algorithms we use for the SE model are shown in Figure 3. Note that only a subset of the patients is considered here (those having at least one well recorded year).

	SE model	Stat model	p-value
LR	**0.901**	0.898	0.173
ANN	**0.865**	0.864	0.686

RF	**0.855**	0.843	0.274
SVM	**0.840**	0.837	0.263
GBC	**0.862**	0.838	0.200

Table 4. AUC values of all algorithms and different models.

Although the AUC values of the SE models are always higher than those of the Stat models across all algorithms as shown in Table 4, the t-test shows that the subgraph mining algorithm does not yield a statistically significant improvement over the Stat model.

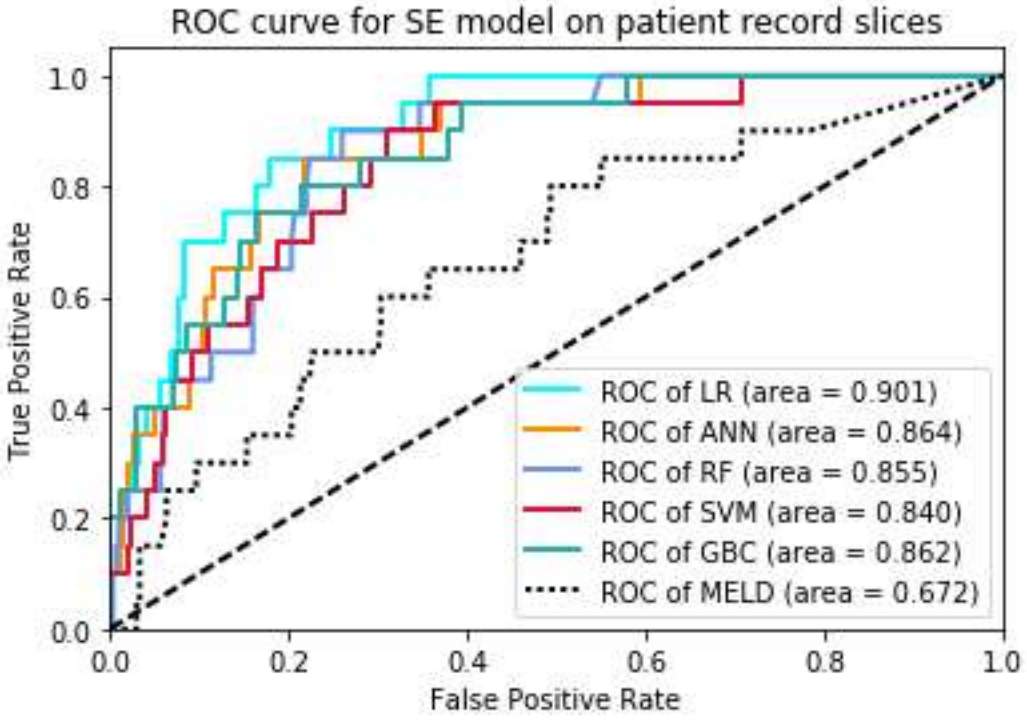

Figure 3. ROC curve for SE model on patient record slices

4.3.3 Ensemble Model

We take the best performed Stat model and the best performed SE model for model ensemble. Both models turn out to be logistic regression. The AUC performance of the ensemble model and the two logistic regression models are shown in Figure 4.

The ensemble model has a higher performance than either model. Furthermore, the t-test shows that ensemble is statistically better than the Stat model (with the p-value of 0.0027).

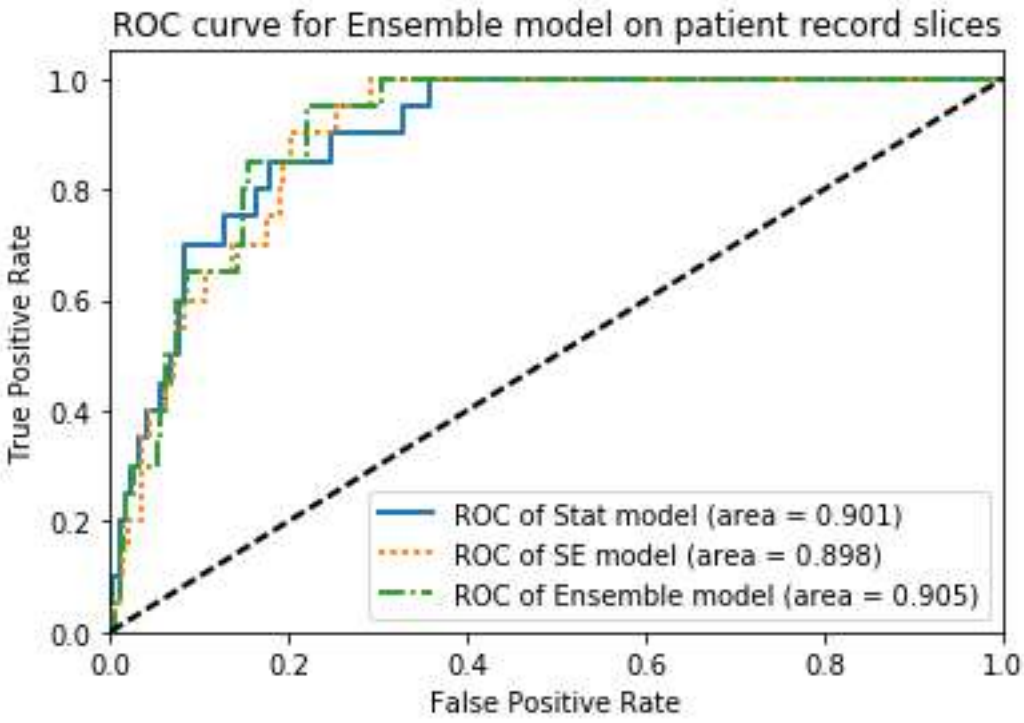

Figure 4. ROC curve for Ensemble model on patient record slices

4.3.4 Feature importance

Using the coefficients of trained logistic regression models, we gain more information about important features and how they influence the probability of death.

Appendix 4 shows the 10 most influential features of mortality prediction, together with the actual characteristics of each feature in alive group and deceased group. "Coefficient" indicates the change of the logit of probability of death, which is $\log(\frac{probability\ of\ death}{1-probability\ of\ death})$. For continuous features like 'last measurement of MELD,' whenever the value increases by 1 unit (8.118), the logit of the patient's probability of death would be increased by 1.745. For discrete feature like 'no alcohol use,' if the feature changes from 0 to 1 (clinically it means that the patient stops drinking alcohol), then the logit of the probability of death would be increased by 0.442.

All these features are known to be correlated with cirrhosis, and have a statistically significant difference between the alive and deceased groups. Among them, only 'Bilirubin Total' is a component of the MELD score.

5. Concluding discussions

As the first to apply modern machine learning algorithms and data analysis methods on predicting one-year mortality of cirrhosis patients, we have built a model that predicts whether a patient would die within a year fairly accurately. The model can serve as a second opinion when clinicians decide whether the patient should get a liver transplantation.

5.1 The Models

Previous studies showed that MELD could serve as a mortality predictor, and we confirm that by only using the MELD score the AUC reaches 0.82. By using additional features and machine learning models however, all our models outperform the MELD-only model.

One important reason of the improvement we believe is that the feature space we use is larger than the one used in previous models.

We expect the SE model to outperform the Stat model since the SE model uses all features included in the Stat model and extra subgroup features. This is the case on average, however, the t-test does not show a statistically significant improvement of the model performance. The ensemble model outperforms both the SE and Stat models and the t-test shows that this is statistically significant.

We believe that patterns of laboratory records do not significantly help in our study due to the sparsity of the records. On average, 36.8% of the data in a well recorded year is missing.

5.2 Important features

The 10 most influential features exhibited in Appendix 4 have been confirmed by other research that they are factors of cirrhosis. However, our study is the first to quantitively show the magnitude of feature importance. Our model provides an understanding of how a change of one feature would influence mortality. For example, ‘No alcohol use,’ although other researches have point out a strong relationship between alcohol intake and cirrhosis, we also quantitively show that if a patient stops drinking alcohol, how much would the probability of death change.

6. Acknowledgment

We are obliged to Ye Xue, a PhD candidate in computer science at Northwestern University, for his assistance in text mining. In addition, we acknowledge Professor Yuan Luo for contributing his subgraph mining implementation. Partial financial support has been provided by the Data Science Institute at Northwestern University which is gratefully acknowledged.

7. Reference list

1 National Institute of Diabetes and Digestive and Kidney Diseases. *Cirrhosis*. https://www.niddk.nih.gov/health-information/liver-disease/cirrhosis (accessed Nov 2017).

2 Malinchoc M, Kamath PS, Gordon FD, et al. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology*. 2000 Apr;31(4):864-71.

3 Suman A, Barnes DS, Zein NN, et al. Predicting outcome after cardiac surgery in patients with cirrhosis: a comparison of Child-Pugh and MELD scores. *Clin Gastroenterol Hepatol*. 2004 Aug;2(8):719-23.

4 Botta F, Giannini E, Romagnoli P, et al. MELD scoring system is useful for predicting prognosis in patients with liver cirrhosis and is correlated with residual liver function: a European study. *Gut*. 2003 Jan;52(1):134-9.

5 Kamath PS, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. *Hepatology*. 2001 Feb;33(2):464-70.

6 Heuman DM, Abou-Assi SG, Habib A, et al. Persistent ascites and low serum sodium identify patients with cirrhosis and low MELD scores who are at high risk for early death. *Hepatology*. 2004 Oct;40(4):802-10.

7 Luo Y, Xin Y, Joshi R, et al. Predicting ICU Mortality Risk by Grouping Temporal Trends from a Multivariate Panel of Physiologic Measurements. In: *Proc Conf AAAI Artif Intell 2016*. 2016 Feb 12-17; Phoenix, Az.

8 Pugh RN, Murray-Lyon IM, Dawson JL, et al. Transection of the oesophagus for bleeding oesophageal varices. *Br J Surg*. 1973 Aug;60(8):646-9.

9 Ruf AE, Kremers WK, Chavez LL, et al. Addition of serum sodium into the MELD score predicts waiting list mortality better than MELD alone. *Liver Transpl*. 2005 Mar;11(3):336-43.

10 Prohic D, Mesihovic R, Vanis N, Prognostic Significance of Ascites and Serum Sodium in Patients with Low Meld Scores. *Med Arch*. 2016 Feb;70(1):48–52.

11 Goldberg E, Chopra S. Cirrhosis in adults: Etiologies, clinical manifestations, and diagnosis. *UpToDate*. https://www.uptodate.com/contents/cirrhosis-in-adults-etiologies-clinical-manifestations-and-diagnosis (accessed Nov 2017).

12 Charif I, Saada K, Mellouki I., et al. Predictors of Intra-Hospital Mortality in Patients with Cirrhosis. *Open Journal of Gastroenterology*. 2014 Mar;4(3):141-48.

13 Rodrigues-Pinto E, Freitas-Silva M. Hepatorenal syndrome, septic shock and renal failure as mortality predictors in patients with spontaneous bacterial peritonitis. *GE Jornal Português de Gastrenterologia*. 2012 Nov-Dec;19(6):278-83.

14 Romero-Gómez M, Boza F, García-Valdecasas MS, et al. Subclinical hepatic encephalopathy predicts the development of overt hepatic encephalopathy. *Am J Gastroenterol*. 2001 Sep;96(9):2718-23.

15 Zein CO, Lindor KD, Angulo P. Prevalence and predictors of esophageal varices in patients with primary sclerosing cholangitis. *Hepatology*. 2004 Jan;39(1):204-10.

16 Papp M, Vitalis Z, Altorjay I, et al. Acute phase proteins in the diagnosis and prediction of cirrhosis associated bacterial infections. *Liver Int*. 2012 Apr;32(4):603-11. doi: 10.1111/j.1478-3231.2011.02689.x. Epub 2011 Dec 6.

17 Singal AG, Mukherjee A, Elmunzer BJ, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. *Am J Gastroenterol*. 2013 Nov;108(11):1723-30. doi: 10.1038/ajg.2013.332. Epub 2013 Oct 29.

18 Sakiyama Y, Yuki H, Moriya T, et al. Predicting human liver microsomal stability with machine learning techniques. *J Mol Graph Model*. 2008 Feb;26(6):907-15. Epub 2007 Jun 27.

19 Sartakhti JS, Zangooei MH, Mozafari K. Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVM-SA). *Comput Methods Programs Biomed*. 2012 Nov;108(2):570-9. doi:10.1016/j.cmpb.2011.08.003. Epub 2011 Oct 2.

20 Ye QH, Qin LX, Forgues M, et al. Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nat Med*. 2003 Apr;9(4):416-23.

21 Joshi R, Szolovits P. Prognostic Physiology: Modeling Patient Severity in Intensive Care Units Using Radial Domain Folding. In: *AMIA Annu Symp Proc*. 2012;2012:1276-83. Epub 2012 Nov 3.

22 Van Buuren S, Groothuis-Oudshoorn K. Mice: Multivariate Imputation by Chained Equations. *J Stat Softw*. 2011 Dec;45(3):1–67

23 Borgelt C, Berthold MR. Mining molecular fragments: finding relevant substructures of molecules. In *Proc IEEE Int Conf Data Min 2002*. 2002:51-58.

24 Hofree M, Shen JP, Carter H, et al. Network-based stratification of tumor mutations. *Nat Methods*. 2013 Nov;10(11):1108–1115.

25 Peng HC, Long FH, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell*. 2005 Jun;27(8):1226-1238

26 Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Intell Res*. 2002 Jan;16(1):321-357

8. Appendices

Appendix 1. Definition of Complications of Cirrhosis

Complication	Corresponding ICD-9 code
HCV	070.41, 070.44, 070.51, 070.54, 070.70, 070.71, V02.62
Hepatitis B	070.20, 070.21, 070.22, 070.23, 070.30, 070.31, 070.32, 070.33, 070.42, V02.61
Alcohol	571.0, 571.1, 571.2, 571.3
NASH	571.8, 571.9
HCC	155.0
Cholestasis	571.6, 576.1
Portal hypertension	572.3
Ascites	789.5, 789.51, 789.59, 568.82
HE	572.2, 348.31, 348.30, 348.39, 349.82
Jaundice	782.4, 277.4
Esophageal Varices	456.1, 456.21
Variceal bleeding	456.0, 456.20
SBP	567.23, 567.0, 567.21, 567.29, 567.89, 567.9
Hepatorenal	572.4
Hepatopulmonary	573.5
Hepatic hydrothorax	511.8, 511.9, 511.89
Malnutrition	263.9, 728.2, 263.0, 263.1, 263.2, 263.8, 799.4, 260, 261, 262, 783.2, 783.21, 783.22, 783.3
Congestive heart failure	398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 425.4, 425.5, 425.7, 425.8, 425.9, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9
Cardiac Arrhythmias	426.0, 426.13, 426.7, 426.9, 426.10, 426.12, 427.0, 427.1, 427.2, 427.31, 427.32, 427.41, 427.42, 427.60, 427.61, 427.69, 427.81, 427.89, 427.9, 785.0, 996.01, 996.04, V45.00, V45.01, V53.31, V53.32, V53.39

Valvular Disease	093.20, 093.21, 093.22, 093.23, 093.24, 394.0, 394.1, 394.2, 394.9, 395.0, 395.1, 395.2, 395.9, 396.0, 396.1, 396.2, 396.3, 396.8, 396.9, 397.0, 397.1, 397.9, 424.0, 424.1, 424.2, 424.3, 424.90, 424.91, 424.99, 746.3, 746.4, 746.5, 746.6, V42.2, V43.3
Pulmonary Circulation Disorders	415.0, 415.11, 415.12, 415.13, 415.19, 416.0, 416.1, 416.2, 416.8, 416.9, 417.0, 417.8, 417.9
Peripheral Vascular Disorders	093.0, 437.3, 440.0, 440.1, 440.20, 440.21, 440.22, 440.23, 440.24, 440.29, 440.30, 440.31, 440.32, 440.4, 440.8, 440.9, 441.00, 441.01, 441.02, 441.03, 441.1, 441.2, 441.3, 441.4, 441.5, 441.6, 441.7, 441.9, 443.1, 443.21, 443.22, 443.23, 443.24, 443.29, 443.81, 443.82, 443.89, 443.9, 447.1, 557.1, 557.9, V43.4
Hypertension Uncomplicated	401.0, 401.1, 401.9
Hypertension Complicated	402.00, 402.01, 402.10, 402.11, 402.90, 402.91, 403.00, 403.01, 403.10, 403.11, 403.90, 403.91, 404.00, 404.01, 404.02, 404.03, 404.10, 404.11, 404.12, 404.13, 404.90, 404.91, 404.92, 404.93, 405.01, 405.09, 405.11, 405.19, 405.91, 405.99
Paralysis	334.1, 342.00, 342.01, 342.02, 342.10, 342.11, 342.12, 342.80, 342.81, 342.82, 342.90, 342.91, 342.92, 343.0, 343.1, 343.2, 343.3, 343.4, 343.8, 343.9, 344.00, 344.01, 344.02, 344.03, 344.04, 344.09, 344.1, 344.2, 344.30, 344.31, 344.32, 344.40, 344.41, 344.42, 344.5, 344.60, 344.61, 344.9
Other Neurological disorders	331.9, 332.0, 332.1, 333.4, 333.5, 333.92, 334.0, 334.1, 334.2, 334.3, 334.4, 334.8, 334.9, 335.0, 335.10, 335.11, 335.19, 335.20, 335.21, 335.22, 335.23, 335.24, 335.29, 335.8, 335.9, 336.2, 340, 341.0, 341.1, 341.20, 341.21, 341.22, 341.8, 341.9, 345.00, 345.01, 345.10, 345.11, 345.2, 345.3, 345.40, 345.41, 345.50, 345.51, 345.60, 345.61, 345.70, 345.71, 345.80, 345.81, 345.90, 345.91, 348.1, 780.31, 780.32, 780.33, 780.39, 784.3
Chronic Pulmonary Disease	416.8, 416.9, 490, 491.0, 491.1, 491.20, 491.21, 491.22, 491.8, 491.9, 492.0, 492.8, 493.00, 493.01, 493.02, 493.10, 493.11, 493.12, 493.20,

	493.21, 493.22, 493.81, 493.82, 493.90, 493.91, 493.92, 494.0, 494.1, 495.0, 495.1, 495.2, 495.3, 495.4, 495.5,494.6, 495.7, 495.8, 495.9, 496, 500, 501, 502, 503, 504, 505, 506.4, 508.1, 508.8
Diabetes, Uncomplicated	250.00, 250.01, 250.02, 250.03, 250.10, 250.11, 250.12, 250.13, 250.20, 250.21, 250.22, 250.23, 250.30, 250.31, 250.32, 250.33
Diabetes Complicated	250.40, 250.41, 250.42, 250.43, 250.50, 250.51, 250.52, 250.53, 250.60, 250.61, 250.62, 250.63, 250.70, 250.71, 250.72, 250.73, 250.80, 250.81, 250.82, 250.83, 250.90, 250.91, 250.92, 250.93
Hypothyroidism	240.9, 243, 244.0, 244.1, 244.2, 244.3, 244.8, 244.9, 246.1, 246.8
Renal failure	403.01, 403.11, 403.91, 404.02, 404.03, 404.12, 404.13, 404.92, 404.93, 585.1, 585.2, 585.3, 585.4, 585.5, 585.6, 585.9, 586, 588.0, V42.0, V45.1, V56.0, V56.1, V56.2, V56.31, V56.32, V56.8
Peptic Ulcer disease excluding bleeding	531.70, 531.71, 531.90, 531.91, 532.70, 532.71, 532.90, 532.91, 533.70, 533.71, 533.90, 533.91, 534.70, 534.71, 534.90, 534.91
AIDS/HIV	042
Metastatic Cancer	196.0, 196.1, 196.2, 196.3, 196.5, 196.6, 196.8, 196.9, 197.0, 197.1, 197.2, 197.3, 197.4, 197.5, 197.6, 197.7, 197.8, 198.0, 198.1, 198.2, 198.3, 198.4, 198.5, 198.6, 198.7, 198.81, 198.82, 198.89, 199.0, 199.1, 199.2
Lymphoma	200.00, 200.01, 200.02, 200.03, 200.04, 200.05, 200.06, 200.07, 200.08, 200.10, 200.11, 200.12, 200.13, 200.14, 200.15, 200.16, 200.17, 200.18, 200.20, 200.21, 200.22, 200.23, 200.24, 200.25, 200.26, 200.27, 200.28, 200.30, 200.31, 200.32, 200.33, 200.34, 200.35, 200.36, 200.37, 200.38, 200.40, 200.41, 200.42, 200.43, 200.44, 200.45, 200.46, 200.47, 200.48, 200.50, 200.51, 200.52, 200.53, 200.54, 200.55, 200.56, 200.57, 200.58, 200.60, 200.61, 200.62, 200.63, 200.64, 200.65, 200.66, 200.67, 200.68, 200.70, 200.71, 200.72, 200.73, 200.74, 200.75, 200.76, 200.77, 200.78, 200.80, 200.81, 200.82, 200.83, 200.84, 200.85, 200.86, 200.87, 200.88, 201.00, 201.01, 201.02, 201.03, 201.04, 201.05, 201.06, 201.07, 201.08, 201.10, 201.11, 201.12, 201.13, 201.14, 201.15, 201.16, 201.17, 201.18, 201.20, 201.21, 201.22, 201.23, 201.24, 201.25, 201.26, 201.27, 201.28, 201.40, 201.41, 201.42, 201.43, 201.44, 201.45, 201.46, 201.47, 201.48,

	201.50, 201.51, 201.52, 201.53, 201.54, 201.55, 201.56, 201.57, 201.58, 201.60, 201.61, 201.62, 201.63, 201.64, 201.65, 201.66, 201.67, 201.68, 201.70, 201.71, 201.72, 201.73, 201.74, 201.75, 201.76, 201.77, 201.78, 201.90, 201.91, 201.92, 201.93, 201.94, 201.95, 201.96, 201.97, 201.98, 202.00, 202.01, 202.02, 202.03, 202.04, 202.05, 202.06, 202.07, 202.08, 202.10, 202.11, 202.12, 202.13, 202.14, 202.15, 202.16, 202.17, 202.18, 202.20, 202.21, 202.22, 202.23, 202.24, 202.25, 202.26, 202.27, 202.28, 202.30, 202.31, 202.32, 202.33, 202.34, 202.35, 202.36, 202.37, 202.38, 202.40, 202.41, 202.42, 202.43, 202.44, 202.45, 202.46, 202.47, 202.48, 202.50, 202.51, 202.52, 202.53, 202.54, 202.55, 202.56, 202.57, 202.58, 202.60, 202.61, 202.62, 202.63, 202.64, 202.65, 202.66, 202.67, 202.68, 202.70, 202.71, 202.72, 202.73, 202.74, 202.75, 202.76, 202.77, 202.78, 202.80, 202.81, 202.82, 202.83, 202.84, 202.85, 202.86, 202.87, 202.88, 202.90, 202.91, 202.92, 202.93, 202.94, 202.95, 202.96, 202.97, 202.98, 203.00, 203.01, 203.02, 203.10, 203.11, 203.12, 203.80, 203.81, 203.82, 238.6
Solid tumor without metastasis	140.0, 140.1, 140.3, 140.4, 140.5, 140.6, 140.8, 140.9, 141.0, 141.1, 141.2, 141.3, 141.4, 141.5, 141.6, 141.8, 141.9, 142.0, 142.1, 142.2, 142.8, 142.9, 143.0, 143.1, 143.8, 143.9, 144.0, 144.1, 144.8, 144.9, 145.0, 145.1, 145.2, 145.3, 145.4, 145.5, 145.6, 145.8, 145.9, 146.0, 146.1, 146.2, 146.3, 146.4, 146.5, 146.6, 146.7, 146.8, 146.9, 147.0, 147.1, 147.2, 147.3, 147.8, 147.9, 148.0, 148.1, 148.2, 148.3, 148.8, 148.9, 149.0, 149.1, 149.8, 149.9, 150.0, 150.1, 150.2, 150.3, 150.4, 150.5, 150.8, 150.9, 151.0, 151.1, 151.2, 151.3, 151.4, 151.5, 151.6, 151.8, 151.9, 152.0, 152.1, 152.2, 152.3, 152.8, 152.9, 153.0, 153.1, 153.2, 153.3, 153.4, 153.5, 153.6, 153.7, 153.8, 153.9, 154.0, 154.1, 154.2, 154.3, 154.8, 155.1, 155.2, 156.0, 156.1, 156.2, 156.8, 156.9, 157.0, 157.1, 157.2, 157.3, 157.4, 157.8, 157.9, 158.0, 158.8, 158.9, 159.0, 159.1, 159.8, 159.9, 160.0, 160.1, 160.2, 160.3, 160.4, 160.5, 160.8, 160.9, 161.0, 161.1, 161.2, 161.3, 161.8, 161.9, 162.0, 162.2, 162.3, 162.4, 162.5, 162.8, 162.9, 163.0, 163.1, 163.8, 163.9, 164.0, 164.1, 164.2, 164.3, 164.8, 164.9, 165.0, 165.8, 165.9, 170.0, 170.1, 170.2, 170.3, 170.4, 170.5, 170.6,

	170.7, 170.8, 170.9, 171.0, 171.2, 171.3, 171.4, 171.5, 171.6, 171.7, 171.8, 171.9, 172.0, 172.1, 172.2, 172.3, 172.4, 172.5, 172.6, 172.7, 172.8, 172.9, 174.0, 174.1, 174.2, 174.3, 174.4, 174.5, 174.6, 174.8, 174.9, 175.0, 175.9, 176.0, 176.1, 176.2, 176.3, 176.4, 176.5, 176.8, 176.9, 179, 180.0, 180.1, 180.8, 180.9, 181, 182.0, 182.1, 182.8, 183.0, 183.2, 183.3, 183.4, 183.5, 183.8, 183.9, 184.0, 184.1, 184.2, 184.3, 184.4, 184.8, 184.9, 185, 186.0, 186.9, 187.1, 187.2, 187.3, 187.4, 187.5, 187.6, 187.7, 187.8, 187.9, 188.0, 188.1, 188.2, 188.3, 188.4, 188.5, 188.6, 188.7, 188.8, 188.9, 189.0, 189.1, 189.2, 189.3, 189.4, 189.8, 189.9, 190.0, 190.1, 190.2, 190.3, 190.4, 190.5, 190.6, 190.7, 190.8, 190.9, 191.0, 191.1, 191.2, 191.3, 191.4, 191.5, 191.6, 191.7, 191.8, 191.9, 192.0, 192.1, 192.2, 192.3, 192.8, 192.9, 193, 194.0, 194.1, 194.3, 194.4, 194.5, 194.6, 194.8, 194.9, 195.0, 195.1, 195.2, 195.3, 195.4, 195.5, 195.8
Rheumatoid arthritis	446.0, 446.1, 446.20, 446.21, 446.29, 446.3, 446.4, 446.5, 446.6, 446.7, 701.0, 710.0, 710.1, 710.2, 710.3, 710.4, 710.8, 710.9, 711.2, 714.0, 714.1, 714.2, 714.30, 714.31, 714.32, 714.33, 714.4, 714.81, 714.89, 714.9, 719.3, 720.0, 720.1, 720.2, 720.81, 720.89, 720.9, 725, 728.5, 728.89, 729.30
Obesity	278.00, 278.01, 278.02, 278.03
Fluid and Electrolyte disorders	253.6, 276.0, 276.1, 276.2, 276.3, 276.4, 276.50, 276.51, 276.52, 276.61, 276.69, 276.7, 276.8, 276.9
Blood Loss Anemia	280.0
Deficiency Anemia	280.1, 280.8, 280.9, 281.0, 281.1, 281.2, 281.3, 281.4, 281.8, 281.9
Alcohol Abuse	265.2, 291.0, 291.1, 291.2, 291.3, 291.4, 291.5, 291.81, 291.82, 291.89, 291.9, 303.0, 303.9, 305.0, 357.5, 425.5, 535.5, 980.0, 980.1, 980.2, 980.3, 980.8, 980.9, V11.3
Drug Abuse	292.0, 292.11, 292.12, 292.2, 292.81, 292.82, 292.83, 292.84, 292.85, 292.89, 292.9, 304.00, 304.01, 304.02, 304.03, 304.10, 304.11, 304.12, 304.13, 304.20, 304.21, 304.22, 304.23, 304.30, 304.31, 304.32, 304.33, 304.40, 304.41, 304.42, 304.43, 304.50, 304.51, 304.52, 304.53, 304.60, 304.61, 304.62, 304.63, 304.70, 304.71, 304.72, 304.73, 304.80, 304.81, 304.82, 304.83, 304.90, 304.91, 304.92, 304.93, 305.20, 305.21, 305.22,

	305.23, 305.30, 305.31, 305.32, 305.33, 305.40, 305.41, 305.42, 305.43, 305.50, 305.51, 305.52, 305.53, 305.60, 305.61, 305.62, 305.63, 305.70, 305.71, 305.72, 305.73, 305.80, 305.81, 305.82, 305.83, 305.90, 305.91, 305.92, 305.93, V65.42
Psychoses	293.8, 295.00, 295.01, 295.02, 295.03, 295.04, 295.05, 295.10, 295.11, 295.12, 295.13, 295.14, 295.15, 295.20, 295.21, 295.22, 295.23, 295.24, 295.25, 295.30, 295.31, 295.32, 295.33, 295.34, 295.35, 295.40, 295.41, 295.42, 295.43, 295.44, 295.45, 295.50, 295.51, 295.52, 295.53, 295.54, 295.55, 295.60, 295.61, 295.62, 295.63, 295.64, 295.65, 295.70, 295.71, 295.72, 295.73, 295.74, 295.75, 295.80, 295.81, 295.82, 295.83, 295.84, 295.85, 295.90, 295.91, 295.92, 295.93, 295.94, 295.95, 296.04, 296.14, 296.44, 296.54, 297.0, 297.1, 297.2, 297.3, 297.8, 297.9, 298.0, 298.1, 298.2, 298.3, 298.4, 298.8, 298.9
Depression	296.2, 296.3, 296.5, 300.4, 309.0, 309.1, 309.21, 309.22, 309.23, 309.24, 309.28, 309.29, 309.3, 309.4, 309.81, 309.82, 309.83, 309.89, 309.9, 311

Appendix 2. Positive key phrases used in NLP process

Key Phrases
consistent with (history of)
compatible with (history of)
suggest(ive|ing)
there is
the liver is (a) | there is (a)
evidence of
stable findings of
the liver is again noted to be
indicate of
presumed
reflect
re-identified is
morphologic changes of

Appendix 3. Definition of Clinical procedures

CPT code	Clinical Procedure
32554	Thoracentesis, needle or catheter, aspiration of the pleural space; without imaging guidance
32555	Thoracentesis, needle or catheter, aspiration of the pleural space; with imaging guidance
32556	Pleural drainage, percutaneous, with insertion of indwelling catheter; without imaging guidance
32557	Pleural drainage, percutaneous, with insertion of indwelling catheter; with imaging guidance
37182	Insertion of transvenous intrahepatic portosystemic shunt(s) (TIPS)
37204	Embolization code (deleted)
37243	Vascular embolization or occlusion, inclusive of all radiological supervision and interpretation, intraprocedural road mapping, and imaging guidance necessary to complete the intervention; for tumors, organ ischemia, or infarction
43204	Esophagoscopy, flexible, transoral; with injection sclerosis of esophageal varices
43205	Esophagoscopy, flexible, transoral; with band ligation of esophageal varices
43227	Esophagoscopy, flexible, transoral; with control of bleeding, any method
43235	Esophagogastroduodenoscopy, flexible, transoral; diagnostic, including collection of specimen(s) by brushing or washing, when performed (separate procedure)
43236	Esophagogastroduodenoscopy, flexible, transoral; with directed submucosal injection(s), any substance
43243	Esophagogastroduodenoscopy, flexible, transoral; with injection sclerosis of esophageal/gastric varices
43244	Esophagogastroduodenoscopy, flexible, transoral; with band ligation of esophageal/gastric varices
43255	Esophagogastroduodenoscopy, flexible, transoral; with control of bleeding, any method

47120	Hepatectomy, resection of liver; partial lobectomy
47122	Hepatectomy, resection of liver; trisegmentectomy
47125	Hepatectomy, resection of liver; total left lobectomy
47130	Hepatectomy, resection of liver; total right lobectomy
49082	Abdominal paracentesis (diagnostic or therapeutic); without imaging guidance
49083	Abdominal paracentesis (diagnostic or therapeutic); with imaging guidance
75894	Transcatheter therapy, embolization, any method, radiological supervision and interpretation
77778	Interstitial radiation source application, complex, includes supervision, handling, loading of radiation source, when performed
79445	Radiopharmaceutical therapy, by intra-arterial particulate administration

Appendix 4. Top 10 most influential features and characteristics among different patient groups.

	coefficient	unit	alive (n = 2,003)	deceased (n = 319)
last MELD	1.745	8.118	12.38 (6.70-15.25)	22.65 (15.00-30.49)
cpt_75894	1.615	1	9.44%	27.90%
t0 age	1.397	11.28	57.02 (51.00-64.00)	60.95 (54.00-69.00)
last Alpha Fetoprotein Tumor	1.243	299.77	34.23 (2.70-8.50)	210.50 (2.70-20.80)
mean White Cell Count	1.192	2.854	5.80 (4.09-6.91)	7.29 (4.50-8.72)
standard deviation of Sodium	1.172	1.348	2.31 (1.52-2.80)	3.26 (2.14-4.17)
last Bilirubin Total	1.114	5.879	2.31 (0.80-2.20)	8.05 (1.60-8.95)
last AST	1.1	109.56	60.60 (28.00-68.00)	133.92 (44.00-124.00)
No alcohol use	0.442	1	52.22%	26.02%
last Albumin	0.498	0.777	3.51 (3.00-4.10)	2.69 (2.20-3.20)

cpt_75894: Under Transcatheter Diagnostic Radiology (Diagnostic Imaging) Procedures; discrete variables are presented as percentages; continuous variables are presented as mean (25th – 75th percentile).