

# Predicting Acute Liver Failure using Clinical and Histopathological Data

## A PROJECT REPORT

Submitted as part of BBD451 BTech Major Project

Submitted by:

**Yukti Makhija**

**2019BB10067**

Guided by:

**Prof. Ishaan Gupta**



**DEPARTMENT OF BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY  
INDIAN INSTITUTE OF TECHNOLOGY DELHI**

**2022-2023**

## **DECLARATION**

I certify that

- a) the work contained in this report is original and has been done by me under the guidance of my supervisor(s).
- b) I have followed the guidelines provided by the Department in preparing the report.
- c) I have conformed to the norms and guidelines given in the Honor Code of Conduct of the Institute.
- d) whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the report and giving their details in the references. Further, I have taken permission from the copyright owners of the sources whenever necessary.

**Signed by:** Yukti Makhija (2019BB10067)

## **CERTIFICATE**

It is certified that the work contained in this report titled “Predicting acute liver failure using clinical and histopathological data” is the original work done by Ms Yukti Makhija and has been carried out under my supervision.

**Signed by:** Prof. Ishaan Gupta, IITD

**Date:** 02/01/23

## **TABLE OF CONTENTS**

<b>DECLARATION</b>	<b>1</b>
<b>CERTIFICATE</b>	<b>2</b>
<b>TABLE OF CONTENTS</b>	<b>3</b>
<b>List of Tables</b>	<b>4</b>
<b>List of Figures</b>	<b>5</b>
<b>List of Abbreviations</b>	<b>6</b>
<b>Chapter 1: Introduction</b>	<b>7</b>
Stages of Chronic Liver Disease	7
Upper Gastrointestinal Bleeding	8
Histological Analysis	9
Objectives	9
<b>Chapter 2: Literature Review</b>	<b>10</b>
Longitudinal Studies	10
Histological Studies (Liver Biopsies)	11
<b>Chapter 3: Methods</b>	<b>13</b>
Longitudinal Study	13
Data Description	13
Types of CLD	13
Preprocessing the Dataset	14
Parser to Extract Information from EMRs	15
Histological Study	17
Dataset Description	17
NAS Scoring System	18
Architecture	19
<b>Chapter 4: Results</b>	<b>21</b>
Longitudinal Study	21
Projection and Clustering	21
Machine Learning	23
Histological Study	29
<b>Chapter 5: Conclusion</b>	<b>31</b>
<b>References</b>	<b>33</b>

## **List of Tables**

<b>Table</b>	<b>Title</b>	<b>Page</b>
1	Dataset diagnosis distribution	13
2	Dataset after preprocessing, descriptions of parameters	14
3	NAS Score Descriptions	19
4	CLD-Ethanol Prediction: Performance metrics	23
5	CLD-NASH/Ethanol/Infection Prediction: Performance metrics	25
6	Bleeder Prediction: Performance metrics	27
7	NAS Score Prediction Results a) Activity and Steatosis Score b) Fibrosis Score	30

## List of Figures

<b>Figure</b>	<b>Title</b>	<b>Page</b>
1	Different stages of NASH	8
2	(Yu et al.) Univariate COX regression analysis	10
3	(González et al.) Statistic significance of mortality predictors	11
4	(Roy et al.) DELINEATE Architecture	12
5	(Arjmand et al.) Histopathological Features	12
6	Symptom Detection	15
7	Information Extraction from Patient History	16
8	The different classes used to train the CNN models	17
9	The architecture used	19
10	CLD Ethanol Classification-UMAP projection	21
11	CLD NASH/Ethanol/Infectious -UMAP projection	22
12	Bleeder Prediction- UMAP projection coloured.	22
13	CLD-Ethanol Prediction: Logistic Regression Plots	23
14	CLD-Ethanol Prediction: Decision Tree	24
15	CLD-Ethanol: SVM (Linear Kernel) plots	24
16	CLD-Ethanol Prediction: SVM (RBF Kernel plots)	25
17	CLD-NASH/Ethanol/Infectious: Logistic Regression	26
18	CLD-NASH/Ethanol/Infectious: Decision Tree (Splits Visualised)	26
19	Bleeder Prediction: Decision Tree (Splits Visualised)	27
20	Bleeder Prediction: SHAP Feature Importance	28
21	Bleeder Prediction: Logistic Regression Feature Weights	28

### **List of Abbreviations**

<b>Abbreviation</b>	<b>Description</b>
ACLF	Acute-on-Chronic Liver Failure
CLD	Chronic Liver Disease
EMR	Electronic Medical Record
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
ILBS	Institute of Liver and Biliary Sciences
MELD	Model of End-stage Liver Disease
NAFLD	Non-alcoholic Fatty Liver Disease
NASH	Non-alcoholic steatohepatitis
UGIB	Upper Gastrointestinal Bleed

# Chapter 1: Introduction

---

Chronic Liver Disease (CLD) is the continuous worsening of liver functions which lasts more than 6-7 months [1]. In this process, there is a generation of harmful proteins and clotting factors, inflammation of liver parenchyma that results in cirrhosis and fibrosis. In some cases, bile is excreted, and detoxification of dangerous products of the metabolism is also observed. Cirrhosis is considered the last phase of the prognosis because it causes changes to the liver structure and triggers neoangiogenesis and nodule formation.

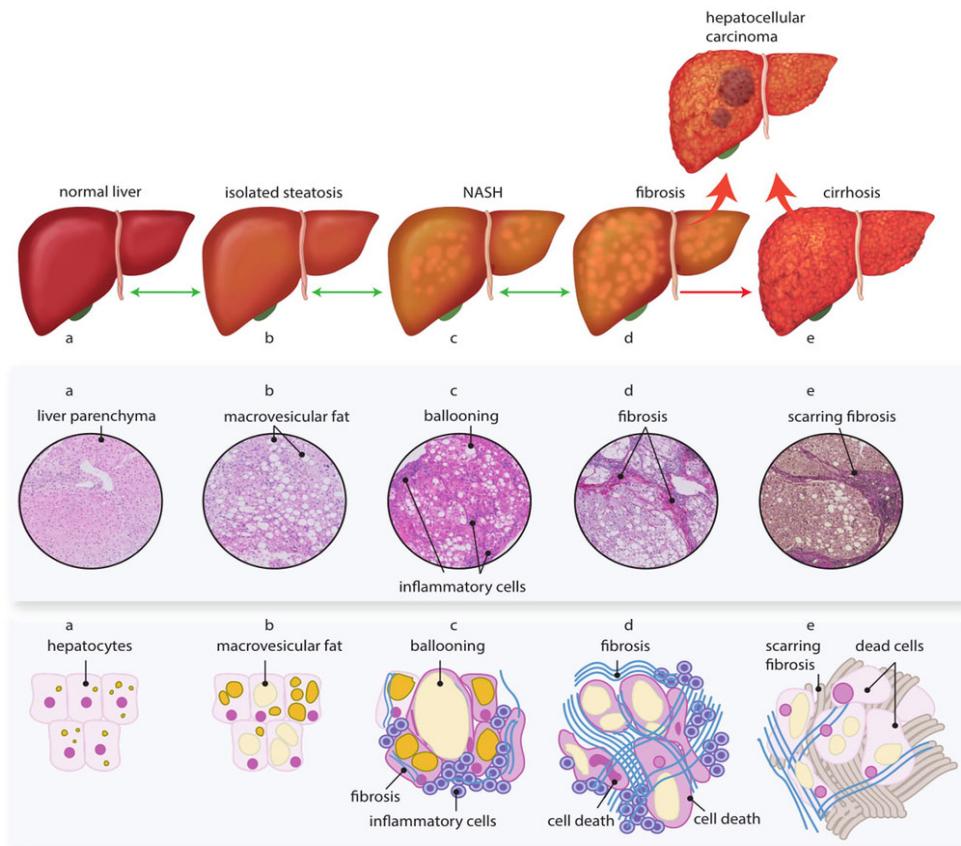
Around 50% of patients with Chronic Liver Failure develop life-threatening conditions like multiple organ failure as the illness progresses. Due to the high mortality rate observed worldwide, there are many scoring systems which use clinical parameters at the time of diagnosis like MELD (Model of End stage Liver Disease) [2], CTP (Child Turcotte Pugh) [3], and the European Association's CLIF-C ACLF [4]. MELD Score is widely used to rank the patients for liver transplant and is calculated using [5]:

- **Serum Sodium:** Measures the fluid-balance regulation in the body.
- **Bilirubin:** Bile is passed through the liver and finally excreted. Measuring the level of bilirubin in blood will tell us how effectively it is being cleared by the liver.
- **Creatinine:** Quantifies kidney function.
- **Internal Normalized Score (INS):** Indicative of liver functioning and measures the formation of proteins that leads to blood clotting.

In most cases, it has been observed that we get a confirmed prognosis between the third and seventh day after hospitalisation and can plan for liver transplant based on the clinical scores (like MELD).

## Stages of Chronic Liver Disease

The four main stages of CLD are inflammation, fibrosis, cirrhosis, and liver failure. Figure 1 illustrates the different stages of Non-Alcoholic SteatoHepatitis (NASH) and the structural changes in the liver as observed in tissue biopsies. During the inflammation stage, the liver is swollen due to the high level of toxins in the blood. Inflammatory cells and ballooning is observed in the tissue in this phase. Patients only experience mild symptoms; hence, the disease often goes undetected and progresses to the fibrosis stage. Tissue scarring and restricted flow of blood in the liver are indicators of this stage. Beyond this stage, the disease is untreatable. Once it reaches the cirrhosis stage, the liver tissue is irreversibly scarred. The fibrosis stage lasts for multiple years, and the transition to cirrhosis is slow, so it can be detected with regular tests. After this, the two possible outcomes are hepatocellular carcinoma or liver failure. In both cases, a liver transplant is the only hope.



**Figure 1: Different stages of NASH**

## Upper Gastrointestinal Bleeding

Upper gastrointestinal bleeding is the most common emergency observed in patients with advanced chronic liver disease. A population study conducted in the UK in 2011 reported an in-patient mortality rate of 10% [6]. The situation hasn't improved much in the past thirty years despite advancements in medicine and diagnosis. The bleeding can be divided into two classes, the first being gastro-oesophageal variceal bleeding, which results from portal hypertension, and the second is non-variceal, which is caused by lesions, ulcers, erosions due to gastritis, and tumours. It is common for patients with advanced CLD to develop varices in their oesophagus, but not all patients suffer from bleeding. The biggest risk factor is an increase in portal hypertension, which is linked to the pressure in the portal vein caused by resistance to out-flow. The development of varices can be seen as the mechanism to bring the flow back to the systemic circulation by decompressing the hypertensive vein. Other factors include alcohol consumption even after diagnosis, the presence of red patches, and the size of the varices. Over 60% of the bleeding situations in cirrhosis patients are variceal, and the patient's survival depends on how severe the CLD is [7]. The remaining cases fall under the category of non-variceal bleeding and are less studied than variceal. Non-variceal upper gastrointestinal bleeding is more common as it can occur in patients with renal, neurological, and cardiac comorbidities. It occurs when the

sub-mucosa is exposed to pepsin and gastric acid after the mucosa barrier has been broken by peptic ulcers. There are very few non-invasive tests to differentiate between the two groups of bleeders before performing an endoscopy.

## Histological Analysis

Histological liver biopsy images are known to carry useful information which can assist in confirming diagnosis, managing patients, predicting prognosis, and devising treatment plans. NASH, a type of CLD, is caused by the buildup of lipids in the liver cells (liver steatosis). The accurate measurement of fat (steatosis) accumulation is necessary to estimate the stage of CLD and make treatment decisions. This information is needed by the liver transplant registry to find a suitable donor. Patients with higher steatosis are often removed from the transplant lists as the chances of graft/organ rejection and kidney failure are very high. Pathologists examine the H&E stain tissue biopsies to grade the steatosis, but the results aren't always reproducible and variability in estimates with the observer is a common problem. Quantification of steatosis is a challenging problem because fat droplets don't have clear boundaries and can be of irregular shapes and sizes. Many hospitals are setting up automated image processing pipelines to avoid such discrepancies. Deep Learning models are commonly used in computer vision for tasks ranging from image classification to segmentation as they can easily learn characteristic features from a set of images.

With the rise in machine learning applications in healthcare, we can filter out the best clinical parameters to predict prognosis. Trained models can be deployed to get predictions at different stages of treatment and automate the decision-making in a clinical setting. Currently, few models exist for liver failure, and even fewer have been tested on Indian patients. This project aims to understand the existing models and develop one using data obtained from Indian hospitals. Effective disease management strategies and early diagnosis can significantly lower the mortality rate.

## Objectives

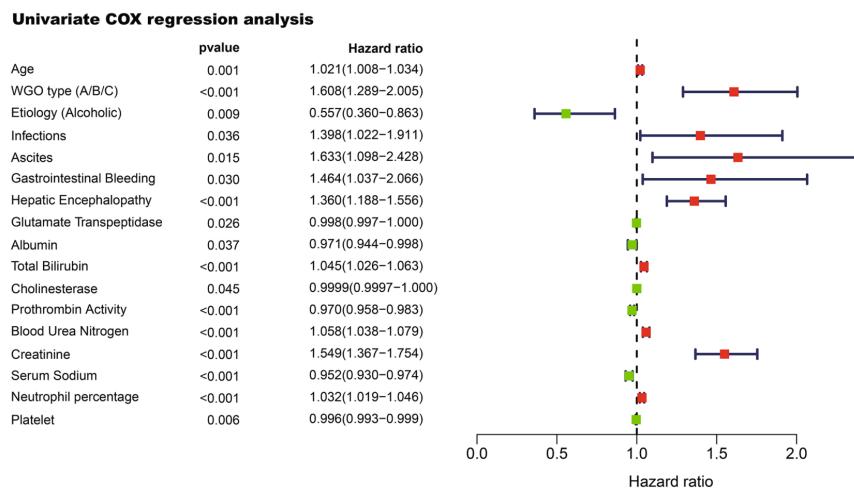
- Using longitudinal clinical data on liver disease patients with some of them having Acute liver failure to predict the cause of CLD.
- Making a parser using pre-trained NLP models for entity recognition on biomedical texts that extracts useful information from EMRs and stores it in a tabular format.
- Train models to distinguish between nonvariceal bleeders and non-bleeders using the clinical features extracted from EMRs and blood clotting factors.
- Study the progression of CLD by training neural networks to classify liver biopsy images on the degree of inflammation, fibrosis, steatosis and ballooning.

# Chapter 2: Literature Review

---

## Longitudinal Studies

The dynamic prediction model for the prognosis of ACLF given by Yu et al. [8] is one of the most recent approaches to predicting liver failure. These worked with time-series data to evaluate whether changes occurring in the features over time were connected to ACLF prognosis. These temporal evolutions were analysed using Univariate and Multivariate Cox Regression. Patient data used by them was collected over four weeks. Their analysis showed that variations in bacterial infections, age, WGO, gastrointestinal bleeding, etc. were related to ACLF. Their key results are summarised in the following diagram.



**Figure 2:** (Yu et al.) Univariate COX regression analysis

There are several other papers where the different factors causing chronic liver diseases have been studied.

Another paper which is pertinent to our work assessed the changes in the prevalence of the causes of CLD from 1988 to 2008 [9]. According to them, there has been an increase in the number of CLD cases in these twenty years. Over time, the CLD cases due to Non-alcoholic Fatty Liver Disease (NAFLD) were drastically increasing with the increase in obesity. This major shift led to a decrease in the percentage of cases caused by HCV infections. NASH (Non-Alcoholic SteatoHepatitis) is the most severe type of NAFLD and often evolves into cirrhosis. NASH is very common in diabetic and obese patients and is the reason behind the high mortality rate.

A short retrospective study from 2013 [10] analysed clinical features that can distinguish between variceal and nonvariceal bleeding in 340 patients, with nearly 80% in the nonvariceal

group. The analysis performed was quite simple and involved comparing the value of clinical parameters for both groups. The main parameters included symptoms, treatment (medication) history, endoscopy results, and previous haemorrhage incidents (rebleeds). The study showed a connection between nonvariceal bleeding and the regular usage of NSAIDs (high dosage of aspirin). Patients who had previously undergone anti-coagulation, steroid, and anti-platelet treatments also experienced nonvariceal. They found no concrete relation between gender and another haemorrhage in 3 days and the 1-month mortality rate. On the other hand, higher Alanine transaminase (ALT) and lower bilirubin and albumin were associated with variceal bleeding. Other key factors were low initial BP (blood pressure) and a diagnosis of cirrhosis.

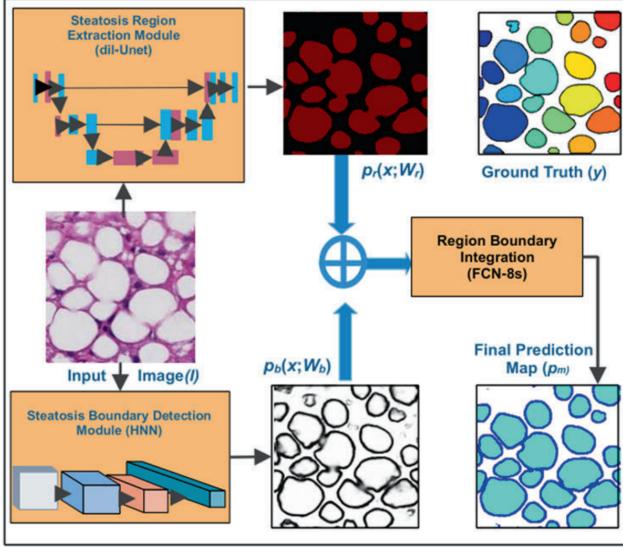
Another extensive study from 2011 [11] focussed on identifying differences in predictors of nonvariceal bleeding in cirrhotic and non-cirrhotic patients. The 2217 patients enrolled in this study had a wide set of comorbidities including renal failure, chronic lung diseases, and malignancies. Predictors were identified by performing multivariate Cox regression. They showed that the mechanism of development of peptic ulcers varies in cirrhotic and non-cirrhotic patients. Portal vein pressure and alcohol usage only affect the outcome in patients with cirrhosis. As expected, fewer cases of multiple haemorrhage and rebleeding incidents occurred in non-cirrhotics. In case of non-cirrhotics, old age, severe haemorrhage, and type of comorbidity, played a role in outcome prediction. On the other hand, the prime predictors for cirrhotics were ulcer severity, albumin levels, metabolic disorders, and the cryptogenic nature of CLD.

Predictors	Odds ratio	95% CI	p
Cryptogenic etiology	3.1	1.06 - 9.01	<b>0.038</b>
BUN	2.6	0.8 - 7.2	0.152
Serum albumin	0.312	0.13 - 0.71	<b>0.006</b>
Active bleeding at ulcer base	9.6	1.8 - 50.2	<b>0.007</b>
Endoscopic treatment	2.0	0.5 - 5.8	0.38

**Figure 3:** (González et al.) Statistic significance of mortality predictors

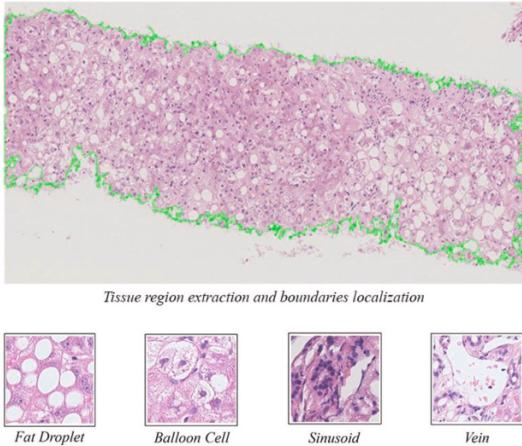
## Histological Studies (Liver Biopsies)

A recent paper [12] published in Nature proposed the first end-to-end model, DELINEATE, which can segment and quantify steatosis droplets in Whole Slide Images (WSI). The two main steps involved are identifying the fat accumulation region by locating its boundary, followed by segmentation of overlapping droplets. Figure 3 illustrates the main stages in the architecture. They trained their model using WSI obtained from 36 patients which had been manually annotated by experts. The model was able to segment steatosis with 98% precision and 0.94 F1-Score.



**Figure 4:** (Roy et al.) DELINEATE Architecture

CNN-based models have also been used to identify histopathological features and use them for classification of liver biopsies. The four key features present in these images (Figure 4a) are steatosis, fibrosis, hepatocyte ballooning, and inflammation.



Feature Category	Description
1 Shape-based	Eccentricity, Extent, Major/Minor axis length
2 Texture-based	Mean intensity, Mean/StD pixel value, Solidity
3 Position	Centroid (x,y coordinates)
4 Magnitude	Area, Diameter, Perimeter

**Figure 5:** (Arjmand et al.) Histopathological Features

Arjmand et al. [13] devised an architecture that first extracts the image features which capture information about the shape, texture and position of the object. Then using image annotations it identifies the four types of objects (fat cells, balloon cells, sinusoid, veins). Lastly, the fat ratio is calculated for WSIs. The training was performed on 13 WSIs (7305 training samples) at 40x magnification but 28 images at 20x were used for evaluation. The mean absolute error at the testing stage was less than 2%.

# Chapter 3: Methods

---

## Longitudinal Study

### Data Description

The data used in this study was collected at the Institute of Liver and Biliary Sciences (ILBS) in New Delhi. All the patients had presented with long-term liver problems, but all the patients did not have CLD. We received the data in two forms, the first was an excel sheet containing 3747 patients and 41 features, but complete information about the diagnosis was given only for 1104 patients. The features given for each patient include physical parameters like sex, gender and age, and multiple blood clotting parameters obtained after performing a thromboelastometry test. In some cases, multiple entries are also present for a patient. We decided to use them as augmentations to increase the size of our dataset.

1647 health records were provided in the form of word documents which would be used to extract additional information about the patients.

790 out of 1104 patients had a confirmed diagnosis of CLD, and even the type of CLD was mentioned. These subcategories have been formed based on the cause of CLD. The distribution of patients is as follows:

Type of CLD	Number of Patients
CLD Ethanol	459
CLD NASH	159
CLD HBV	41
CLD HCV	37
CLD Cryptogenic	41

**Table 1:** Dataset diagnosis distribution

### Types of CLD

- 1) **CLD Ethanol:** Excessive drinking and alcohol consumption for several years leads to this condition. It is the most common cause of CLD.
- 2) **CLD NASH:** This is the most severe type of NAFLD. Patients' lifestyle choices are responsible to a large extent as this affects the obese. Controlling calorie and fat intake and leading an active

life can reduce the chances of developing NASH. The symptoms of NASH appear once the liver has deteriorated beyond repair [14].

- 3) **CLD HBV:** CLD is caused by Hepatitis B viral infection.
- 4) **CLD HCV:** When CLD is caused by hepatitis C viral infection (hepatotropic RNA virus). Nearly 50% of cases of Hepatitis C infections lead to long-term liver problems.
- 5) **CLD Cryptogenic:** CLD causes irreparable changes to the liver structure and presents in the form of cirrhosis. The reason behind the CLD is unknown.

## Preprocessing the Dataset

While cleaning the data, we removed columns where more than 15% of the values were present. This was followed by imputing missing values for the remaining columns with the mean. After this, we dropped the rows of patients with unknown diagnoses.

The final dataset contained 1104 patients and 26 features. Description of the features present in the preprocessed dataset [15][16]:

Clinical Parameters	Description
Clotting time (CT)	Time after which clot formation starts after adding the start reagent to the blood.
Clot Formation time (CFT)	Time taken for the clot firmness to reach 20mm after CT.
A5, A10, A15, A20, A25, A30 values	Clot amplitude (firmness) after 5,10,15,20,25,30 mins.
Maximum Clot Firmness (MCF)	Largest observed amplitude.
Alpha-angle	Tangential angle 0 and the curve when clot firmness has reached 20mm.
Maximum Lysis (ML)	Percent of clot stability lost wrt MCF at the end of the test.
Lysis Index after 30 mins (LI 30)	Clot stability wrt MCF (%) is measured thirty minutes after clotting time.
MaxV	Maximum velocity of clot formation

**Table 2:** Dataset after preprocessing, descriptions of parameters

## Parser to Extract Information from EMRs

The medical text provided has patient history, treatment details, and test results in paragraph format. This needs to be made tabular before analysis can be performed on it.

I also worked on a parser to extract useful information from these EMRs. It uses the famous biomedical NLP library, SciSpacy [17], to process medical texts. SciSpacy has multiple pre-trained models for entity detection and identifying dependency relations in EMRs. I applied the **en\_ner\_bc5cdr\_md** model, which can segment diseases, symptoms and chemicals appearing in the text. This model has been trained on the BC5CDR corpus. It can be used to create the final diagnosis column of the patient.

Some examples of the results obtained after applying these models are given below.

- Symptom detection:

### Chief Complaint

Fever since past since past 15 days

Non productive cough since past 15 days

Altered sensorium since past 2 days

```
[{"text": "Fever since past since past 15 days", "bc5cdr": {"Fever": "DISEASE"}, "craft": {}, "nlp": {"past 15 days": "DATE"}}, {"text": "Non productive cough since past 15 days", "bc5cdr": {"cough": "DISEASE"}, "craft": {}, "nlp": {"past 15 days": "DATE"}}, {"text": "Altered sensorium since past 2 days", "bc5cdr": {}, "craft": {}, "nlp": {"past 2 days": "DATE"}}, {"text": "The End", "bc5cdr": {}, "craft": {}, "nlp": {}}]
```

**Figure 6: Symptom Detection (a):** Example of some symptoms written in the Docx files.

**(b):** The same passed through multiple NLP models to detect those symptoms.

- Outcome Extraction

We searched for the following terms in the documents to find the bleeders: ['bleed','mucosal','cutaneous','hemorrhage','coagulopathy','melena'] and the remaining patients were treated as non-bleeders. Variceal bleeders were treated as non-bleeders. Out of 330 patients, only 41 were bleeders.

- Extraction of Information from Patient History

### History

Mrs XXX XXXX 50 yrs old female who is a known hypertensive , and had hypothyroidism , non diabetic had an index issue in the form of fever since past 15 days which was high grade , intermittent associated with gernalised body aches (Backaches +) . patient also complaint of non productive cough since past 15 days . later patient also developed bleeding PV since past 2-4 days , fever settled . following that patient again developed cough and also he had episode of altered sensorium since past 2 days. patient was evaluated outside and was found to have dengue NS1 positive and typhoid IgM positive . Patient was admitted and was being managed conservatively with RDPC, iv fluids and other supportive measures . patient came to ILBS with above mentioned complaints for further evaluation and management . There is no , vomiting, abdominal pain, altered bowel habits, hematemesis, and malena, burning micturition or decreased urine output. There is no h/o any intoxications, indigenous medications, major surgeries, blood transfusions or IV drug abuse prior to onset of the disease. There is no h/o CAD/TB/COPD.

```

{'bc5cdr': {'hypertensive': 'DISEASE', 'hypothyroidism': 'DISEASE', 'diabetic': 'DISEASE', 'fever': 'DISEASE', 'aches': 'DISEASE', 'cough': 'DISEASE', 'bleeding': 'DISEASE', 'dengue': 'DISEASE', 'typhoid': 'DISEASE', 'vomiting': 'DISEASE', 'abdominal pain': 'DISEASE', 'hematemesis': 'DISEASE', 'malena': 'DISEASE', 'drug abuse': 'DISEASE'}, 'craft': {'PV': 'GGP', 'drug': 'CHEBI'}, 'nlp': {'50 yrs old': 'DATE', 'past 15 days ': 'DATE', 'past 15 days ': 'DATE', 'PV': 'ORG', '2-4 days': 'DATE', 'past 2 days': 'DATE', 'RDPC': 'ORG', 'malena': 'GPE', 'CAD/TB/COPD': 'ORG'}}}

```

**Figure 7: Information Extraction from Patient History (a):** Example of patient history written in the Docx files. **(b):** The same passed through multiple NLP models to extract information about the history like previous diseases and basic information about the patient.

List of symptoms and comorbidities extracted by pre-trained models:

['pneumonia', 'clot', 'necrosis', 'sepsis', 'infection', 'sirs', 'sah', 'portal hypertension', 'diabetes']

They were later stored as categorical features.

These models were not able to extract continuous valued blood parameters and so I had to use regex filters to locate certain blood parameters and scores in the document.

**Scores:** MELD, CTP, CHILD

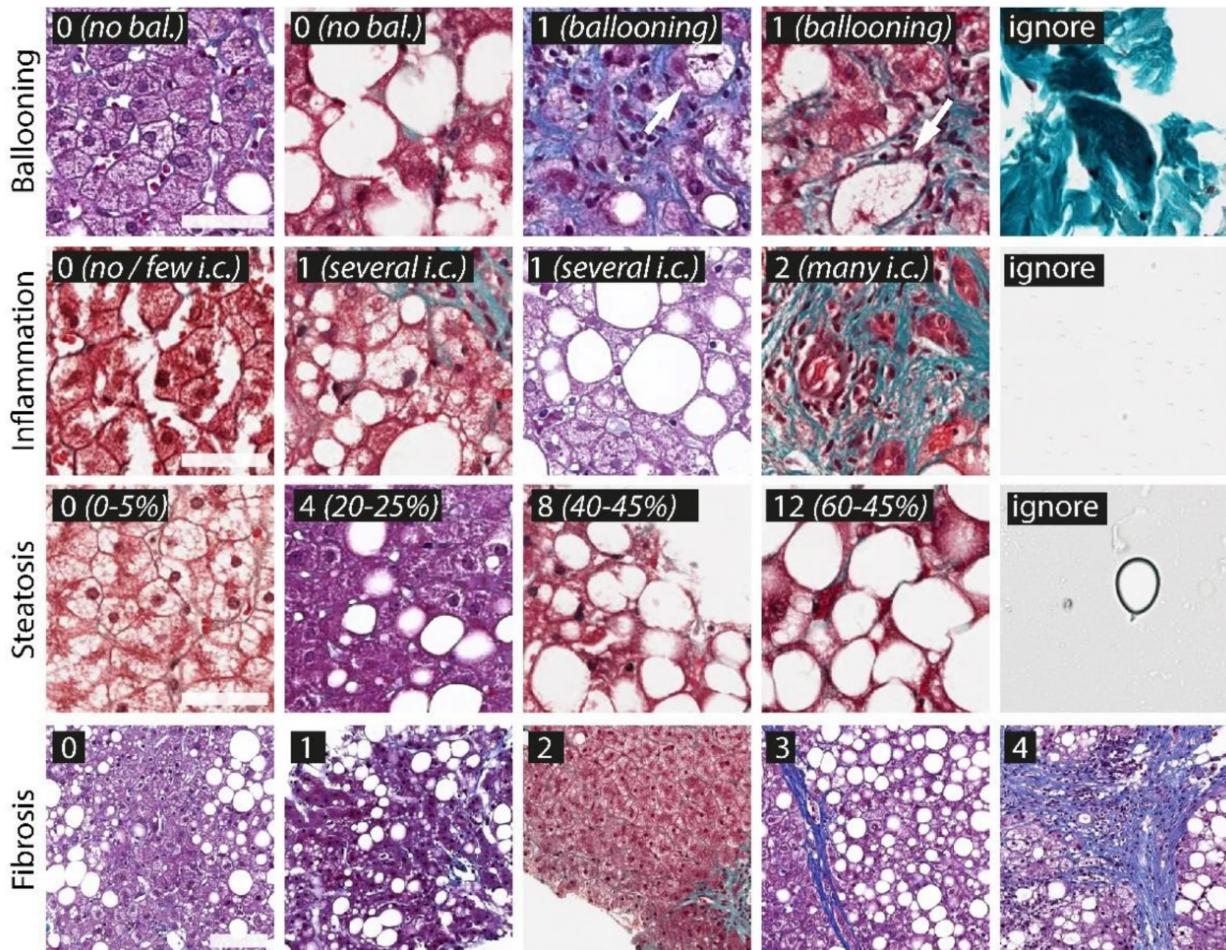
**Blood Tests:** Creatinine, Bilirubin, AST, Gamma-glutamyl transferase (GGTP), Sodium, Potassium, Albumin, etc.

## Histological Study

The objective for this part of my project is to train a CNN-based model to identify four main histopathological features in liver tissue biopsies of NASH patients: steatosis, fibrosis, inflammation, and ballooning, and quantify them using the Kleiner and Brunt scoring system.

### Dataset Description

We used the public dataset containing 338 samples collected at Duke University and 72 samples from Medizinische Hochschule Hannover [18]. The pathologists at these institutes had annotated the WSIs and the ground truth Kleiner NAS score is provided in the form of a CSV.



**Figure 8: The different classes used to train CNN models**

## NAS Scoring System

a) **Fibrosis**

<b>Fibrosis Score</b>	<b>Characteristic</b>
0	No fibrosis
1	Centrilobular
2	Periportal + Centrilobular
3	Bridging
4	Cirrhosis

b) **Ballooning:** Characterised by the number of swollen cells present in the tissue sample.

<b>Ballooning Score</b>	<b>Number of ballooning cells</b>
0	0
1	Few
2	Many

c) **Inflammation:** Defined after the assessment of neutrophils and mononuclear cells.

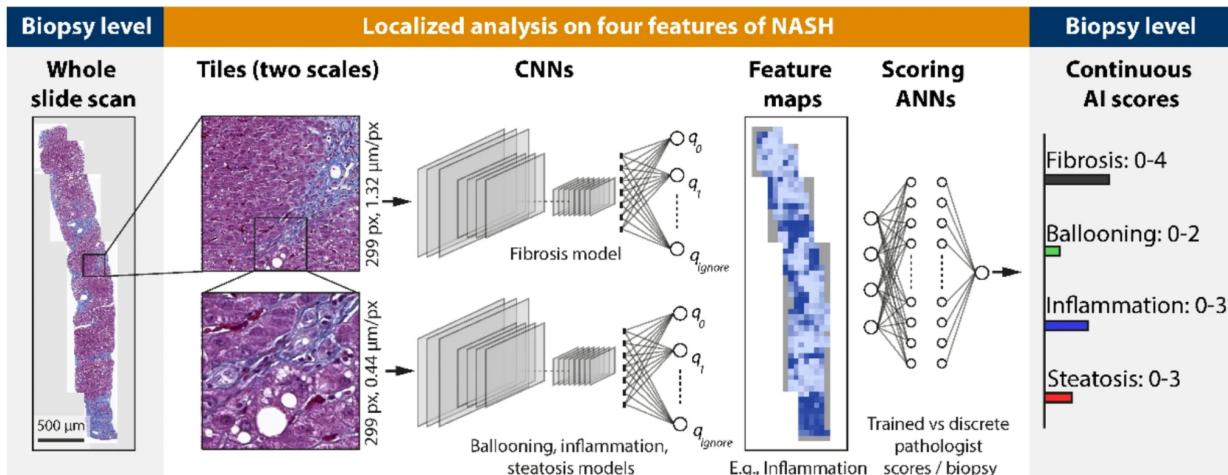
<b>Inflammation Score</b>	<b>Number of inflammatory foci</b>
0	0
1	0-2
2	3-5
3	>5

d) **Steatosis:** The total area where fat accumulation (steatosis) is observed.

Steatosis Score	Percentage Coverage (Area)
0	0-5%
1	5-33%
2	34-66%
3	>67%

**Table 3: NAS Score Descriptions** a) fibrosis b)ballooning c) inflammation d) steatosis

## Architecture



**Figure 9: The architecture used in [18]**

Our architecture is similar to the above, but in our case, we want to combine inflammation score and ballooning score into an activity score.

The architecture consists of a Convolutional Neural Network (CNN) (InceptionV3) followed by a multilayer perceptron (MLP) scoring head, consisting of 2 layers with 100 neurons each and an output layer with 1 neuron. The architecture is the same for all 3 scores (Fibrosis, Steatosis, Activity) but the weights of the CNN and scoring head are different for each.

There is batch normalisation before the images pass through the InceptionV3 model. They are then passed through a global average pooling layer and a dropout layer with  $p=0.5$  for robustness. Finally, the 2D output is flattened and a linear layer is used to reduce dimension to the number of classes for each score and a softmax is used to get the probability of each class.

Between the first two layers of the scoring head, there is additionally a dropout layer with dropping probability = 0.2.

ELU activation is used after the first layer and ReLU after the second layer to provide non-linearity to the model.

# Chapter 4: Results

---

## Longitudinal Study

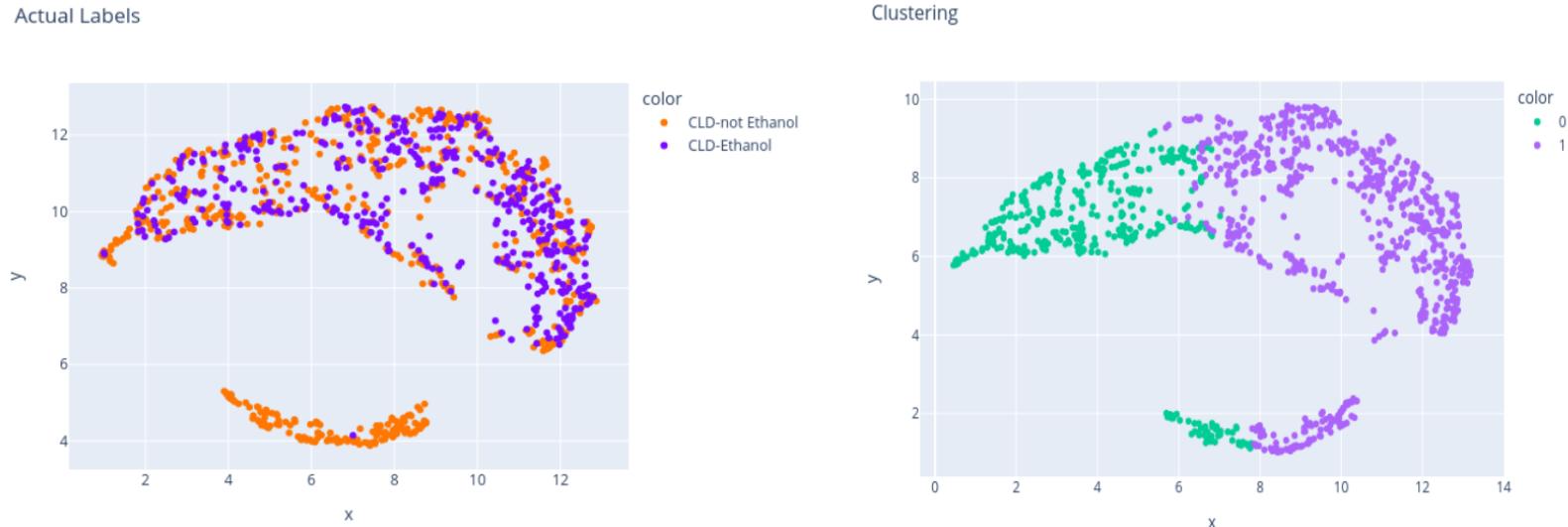
### Projection and Clustering

I have used UMAP to project the high-dimensional data (with 26 features) onto a 2D plot to see whether there is any rough similarity between different groups. UMAP has recently been shown to be consistently better and more stable than other projection methods like t-SNE.

#### CLD-Ethanol Classification Problem (Binary)

However, from the projection plot with the actual labels, we can see that there is little separation between the two classes (CLD-Ethanol and CLD-non Ethanol). So, we can expect the supervised results not to do too well either, as the data is non-clusterable.

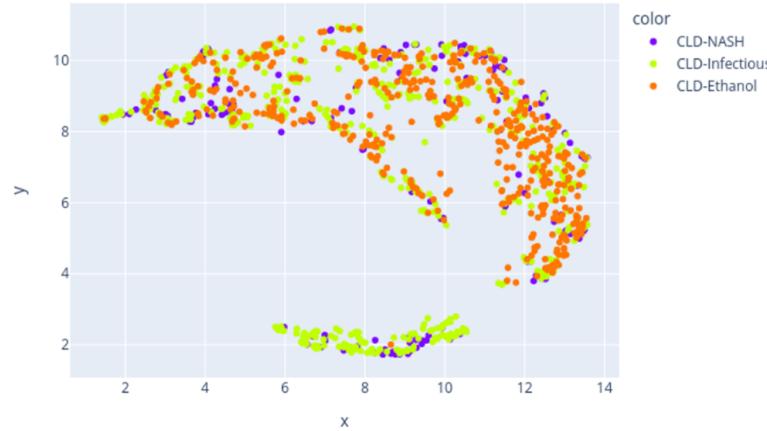
I also do an unsupervised K-Means clustering of the data and show the results on the same projection.



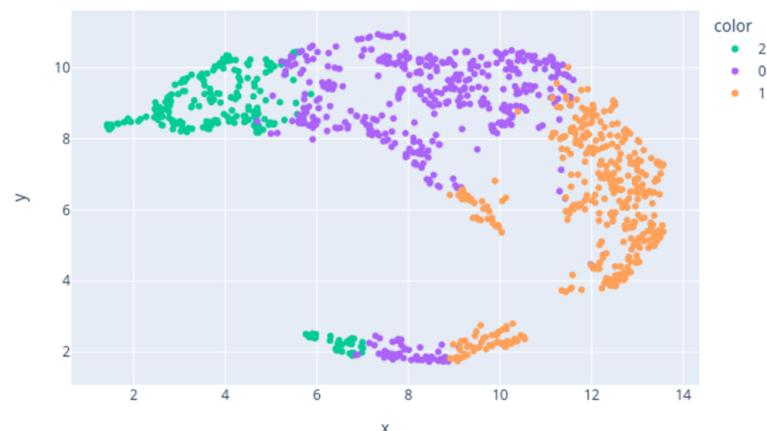
**Figure 10: CLD Ethanol vs Not Ethanol Classification-UMAP projection coloured with (a): actual labels (b): K-Means clustering labels.**

## CLD NASH/Ethanol/Infectious Classification Problem (Multiclass)

Actual Labels



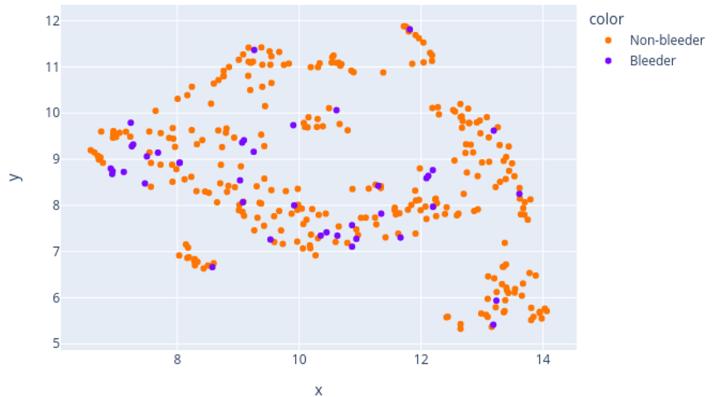
Clustering



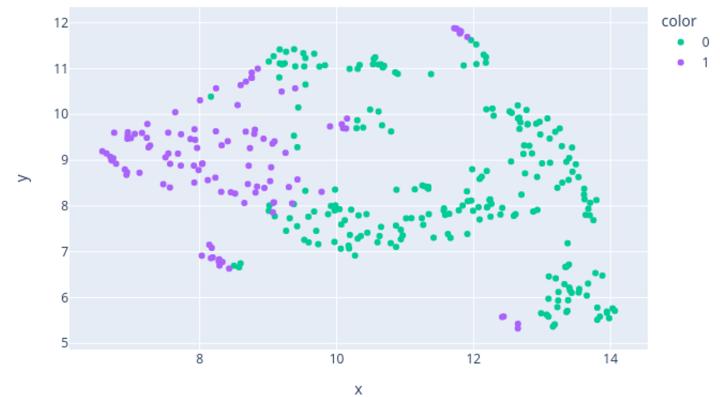
**Figure 11:** *CLD NASH/Ethanol/Infectious -UMAP projection coloured with (a): actual labels (b): K-Means clustering labels.*

## Bleeder Prediction Problem (Binary)

Actual Labels



Clustering



**Figure 12:** *Bleeder Prediction- UMAP projection coloured with (a): actual labels (b): K-Means clustering labels.*

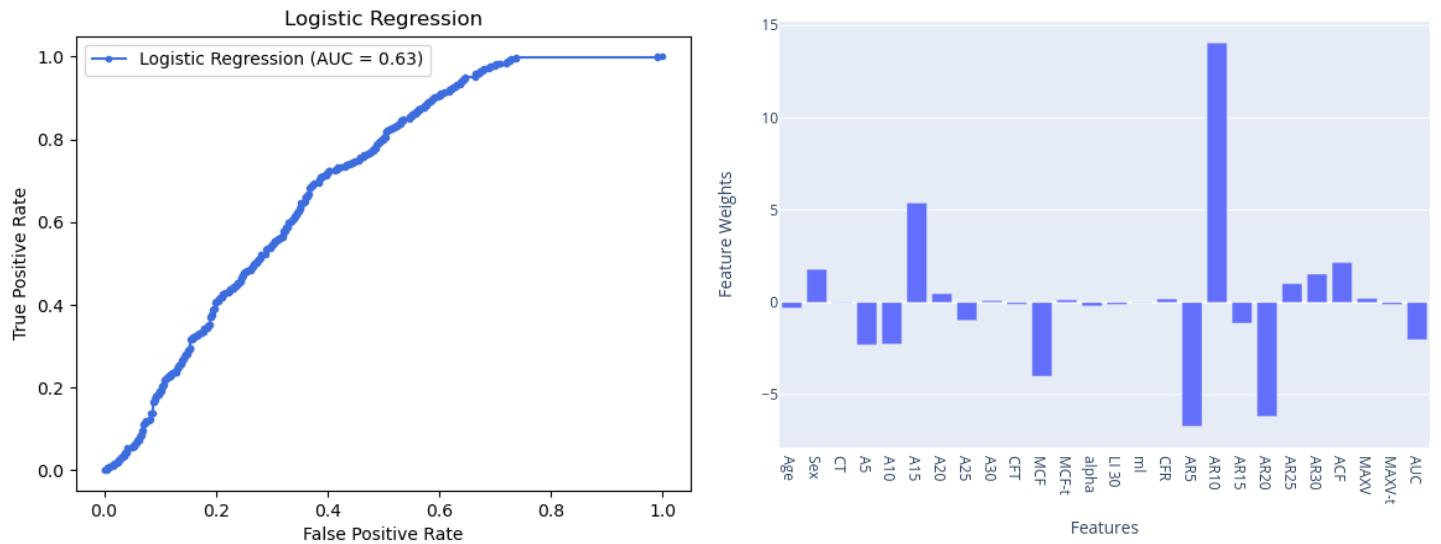
## Machine Learning

We trained different types of models on the preprocessed dataset and evaluated their performance using multiple metrics. The results for logistic regression, decision trees, support vector machines (SVM), and XGBoost can be found below. 10-fold stratified cross-validation was performed to reduce bias that might occur from train-test random splitting.

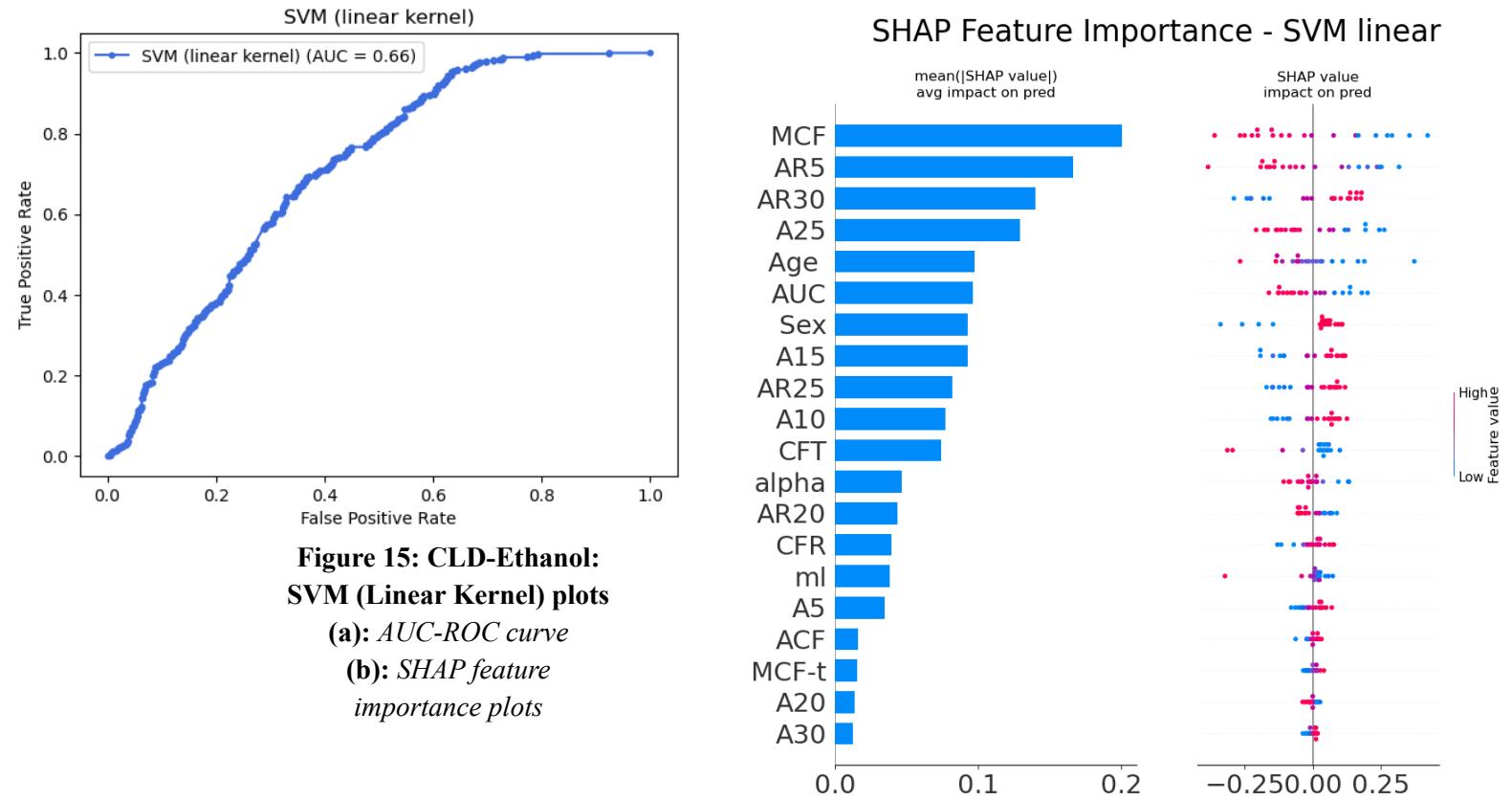
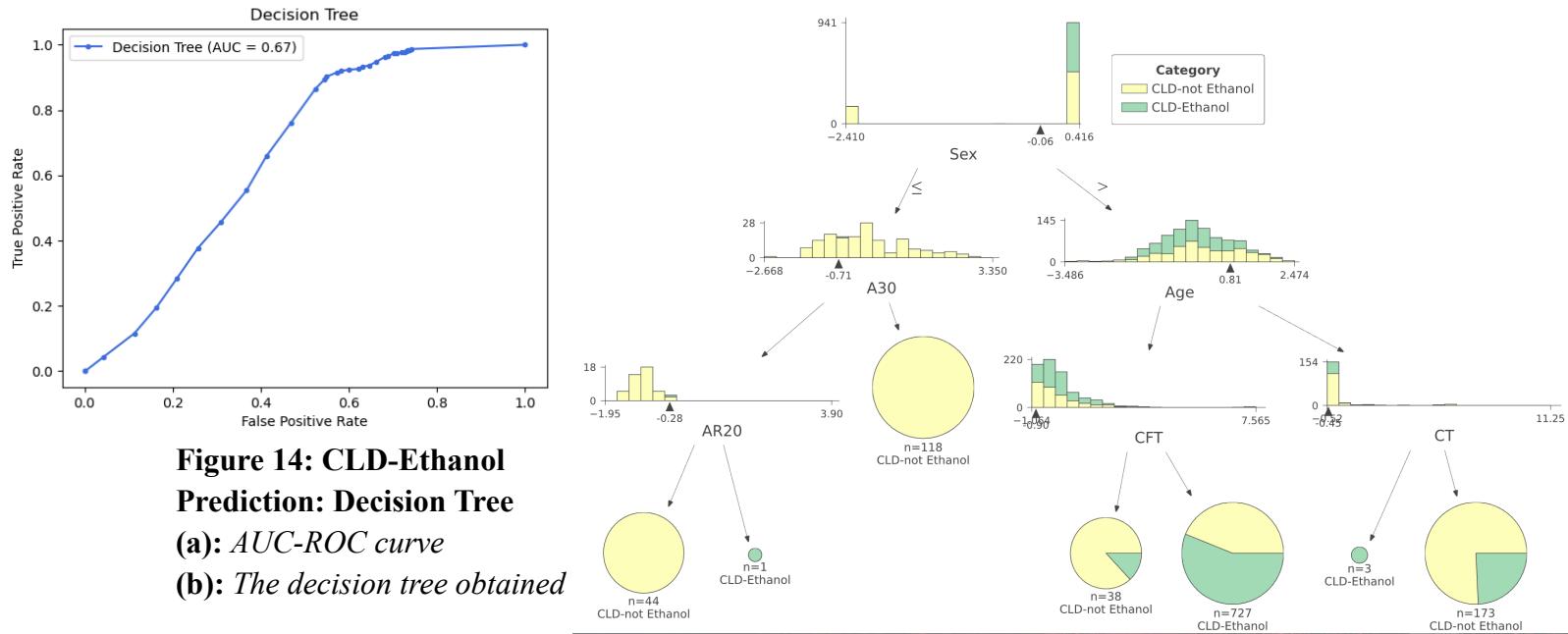
### CLD-Ethanol Prediction Problem

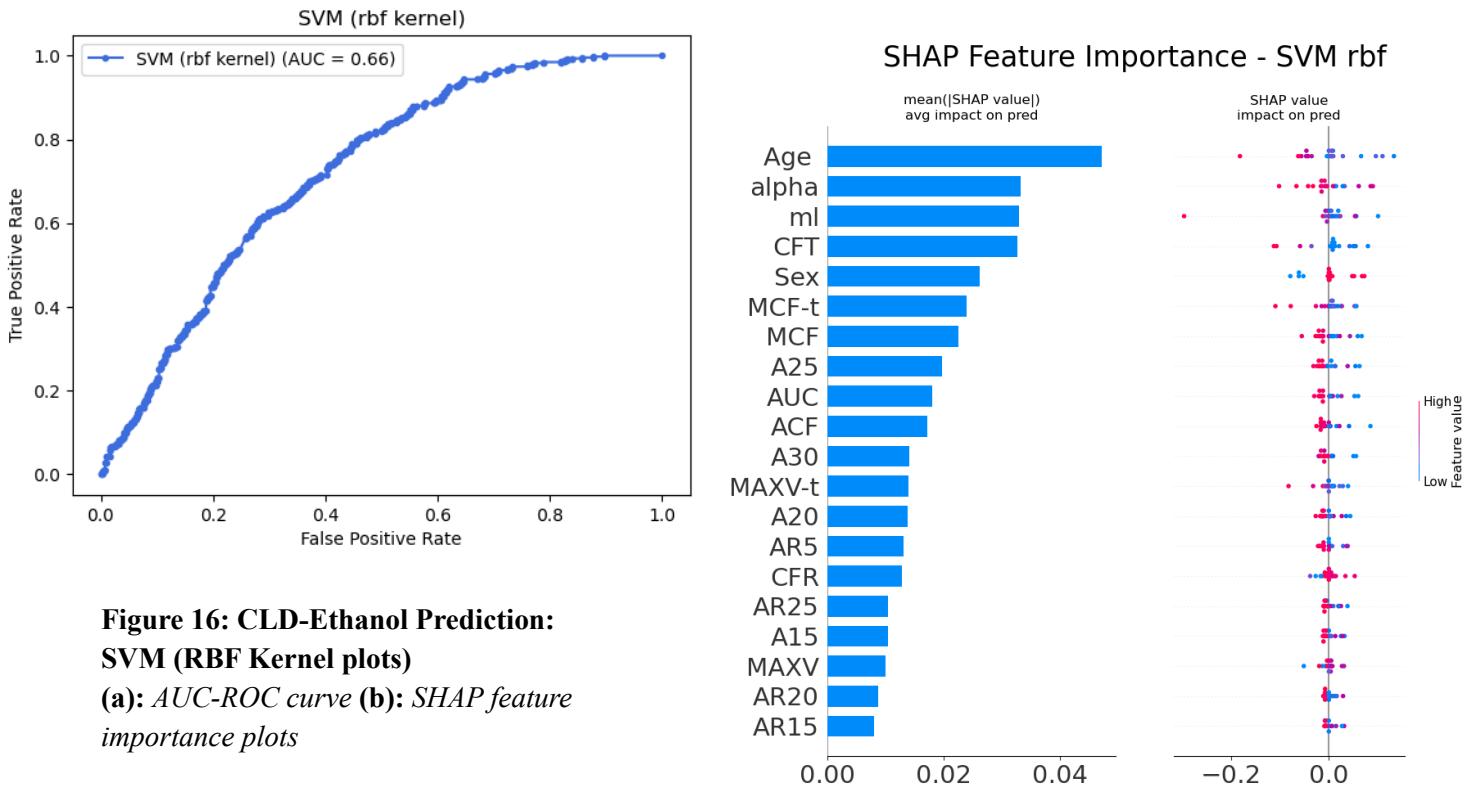
Metrics	Logistic Regression	Decision Tree	SVM (linear kernel)	SVM (RBF Kernel)	XGBoost
Training Accuracy	67.381	66.576	66.938	70.853	<b>78.14</b>
Testing Accuracy	63.578	63.758	65.659	66.567	<b>66.96</b>
Precision	0.559	0.540	0.573	0.593	<b>0.617</b>
Recall	0.588	0.865	0.682	0.623	<b>0.543</b>
F-score	0.573	0.665	0.623	0.608	<b>0.578</b>
AUC-ROC	0.629	0.670	0.660	0.660	<b>0.651</b>

**Table 4: CLD-Ethanol Prediction: Performance metrics**



**Figure 13: CLD-Ethanol Prediction: Logistic Regression Plots (a): AUC-ROC (b): Feature Weights**



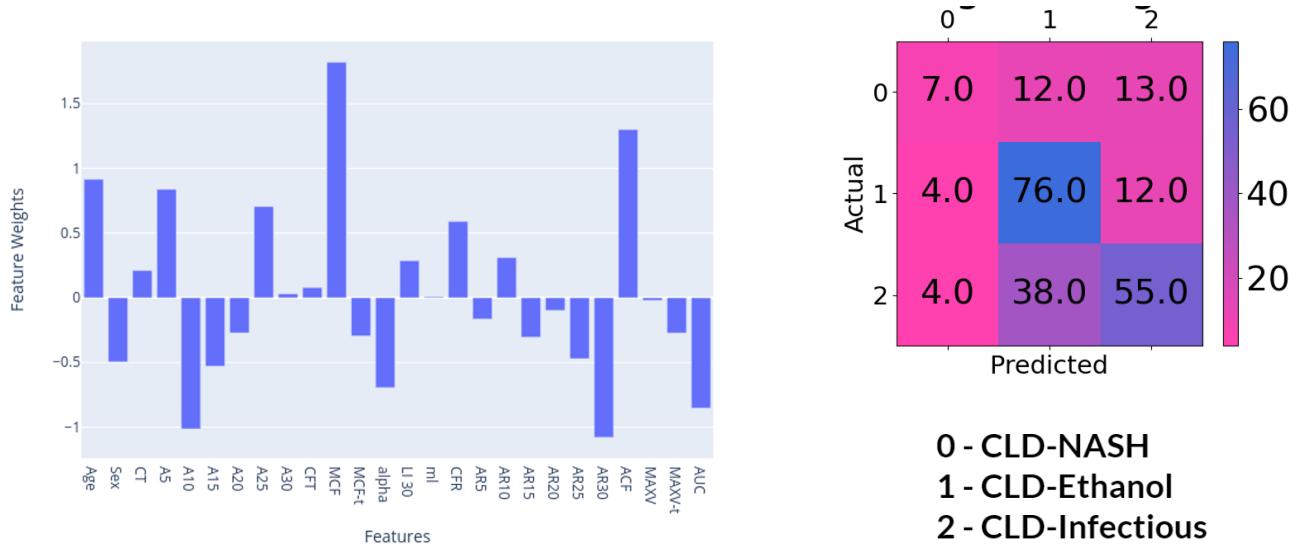


**Figure 16: CLD-Ethanol Prediction:**  
**SVM (RBF Kernel plots)**  
**(a): AUC-ROC curve (b): SHAP feature importance plots**

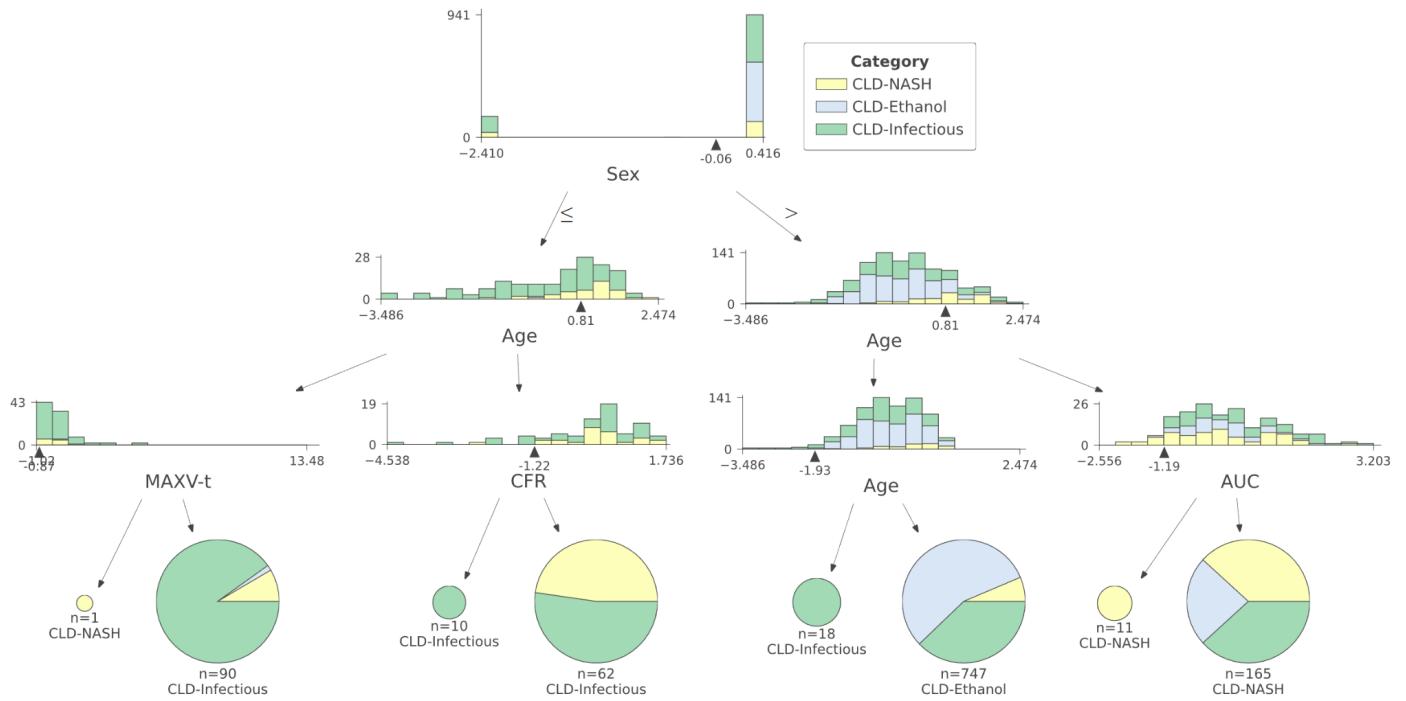
### CLD-NASH/Ethanol/Infectious Prediction (Multiclass)

Metrics	Logistic Regression	Decision Tree	SVM (linear kernel)	SVM (RBF Kernel)	XGBoost
Training Accuracy	<b>58.75</b>	57.41	59.10	60.02	65.74
Testing Accuracy	<b>62.44</b>	52.48	61.54	59.28	59.11
Precision	<b>0.585</b>	0.492	0.544	0.404	0.472
Recall	<b>0.537</b>	0.456	0.516	0.466	0.467
AUC-ROC	<b>0.663</b>	0.601	0.651	0.615	0.613

**Table 5: CLD-NASH/Ethanol/Infection Prediction: Performance metrics**



**Figure 17: CLD-NASH/Ethanol/Infectious: Logistic Regression** a) Feature Weights b) Confusion Matrix



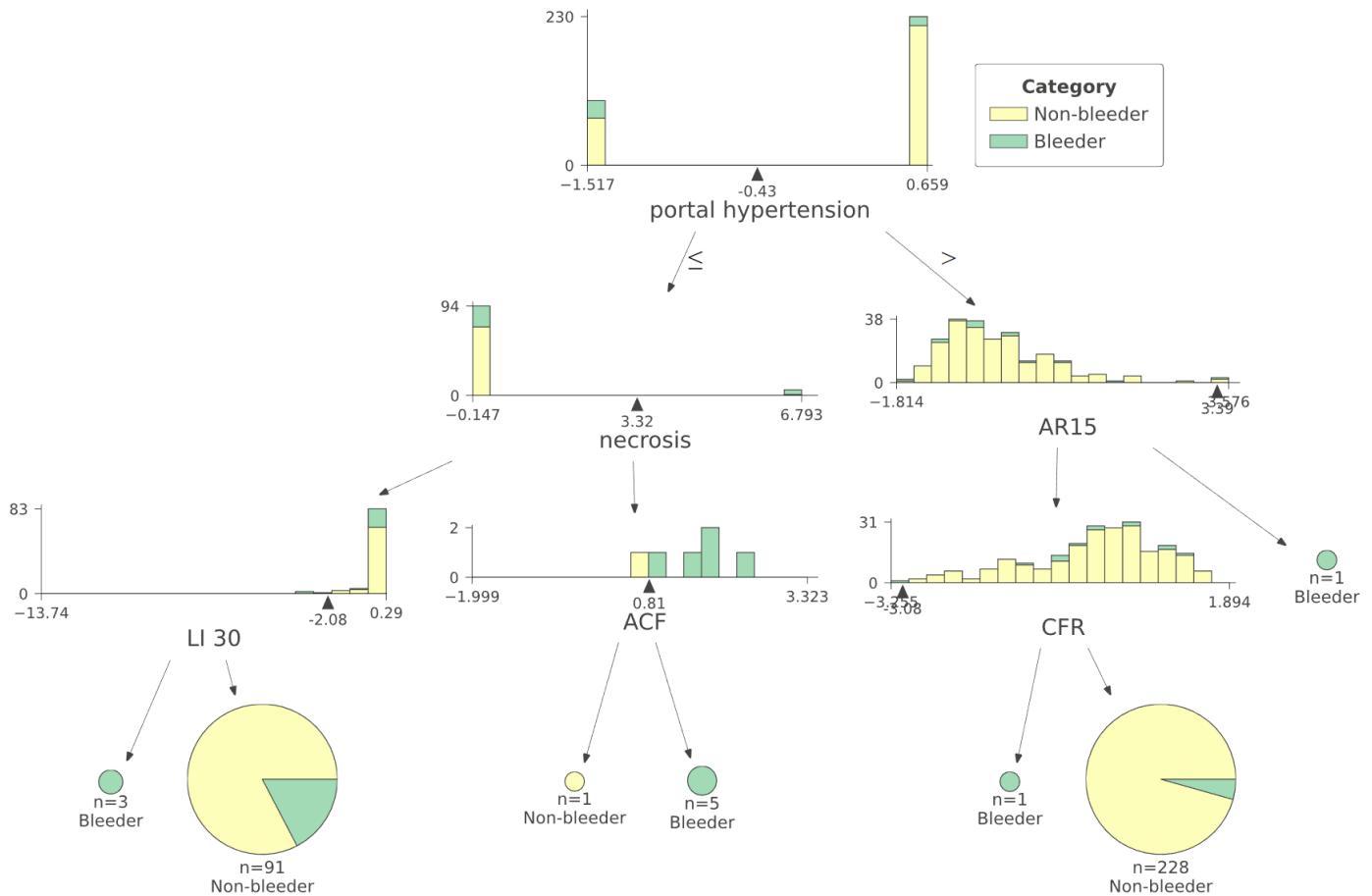
**Figure 18: CLD-NASH/Ethanol/Infectious: Decision Tree (Splits Visualised)**

## Bleeder Prediction

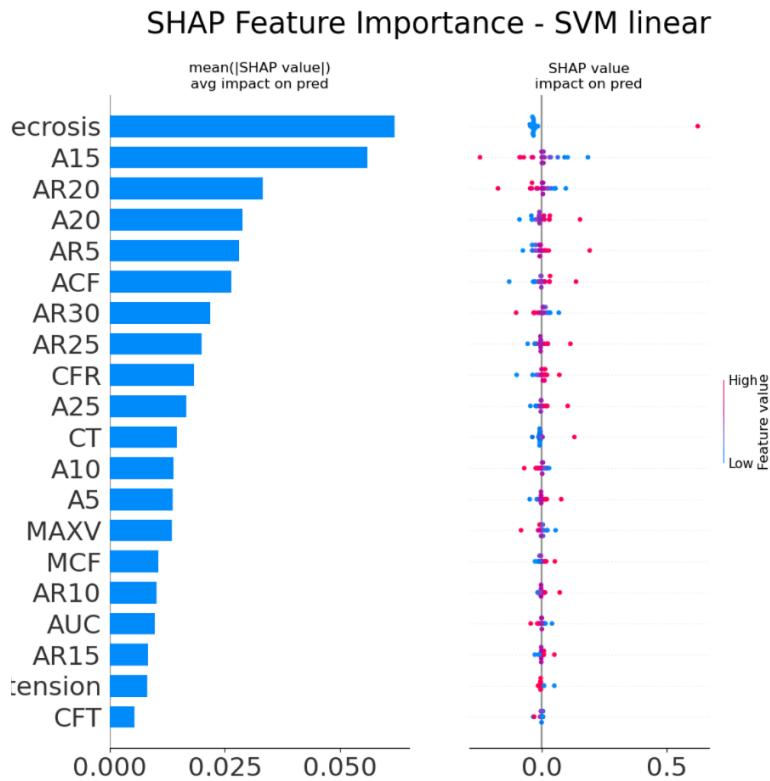
Since less than 15% patients were bleeders, we only used a subset of non-bleeders to balance the ratio of the two groups.

Metrics	Logistic Regression	Decision Tree	SVM (linear kernel)
Training Accuracy	91.191	<b>86.99</b>	87.895
Testing Accuracy	80.45	<b>84.567</b>	80.53
Precision	0.387	<b>1.0</b>	0.294
Recall	0.293	<b>0.136</b>	0.1724
AUC-ROC	0.6	<b>0.5366</b>	0.5317

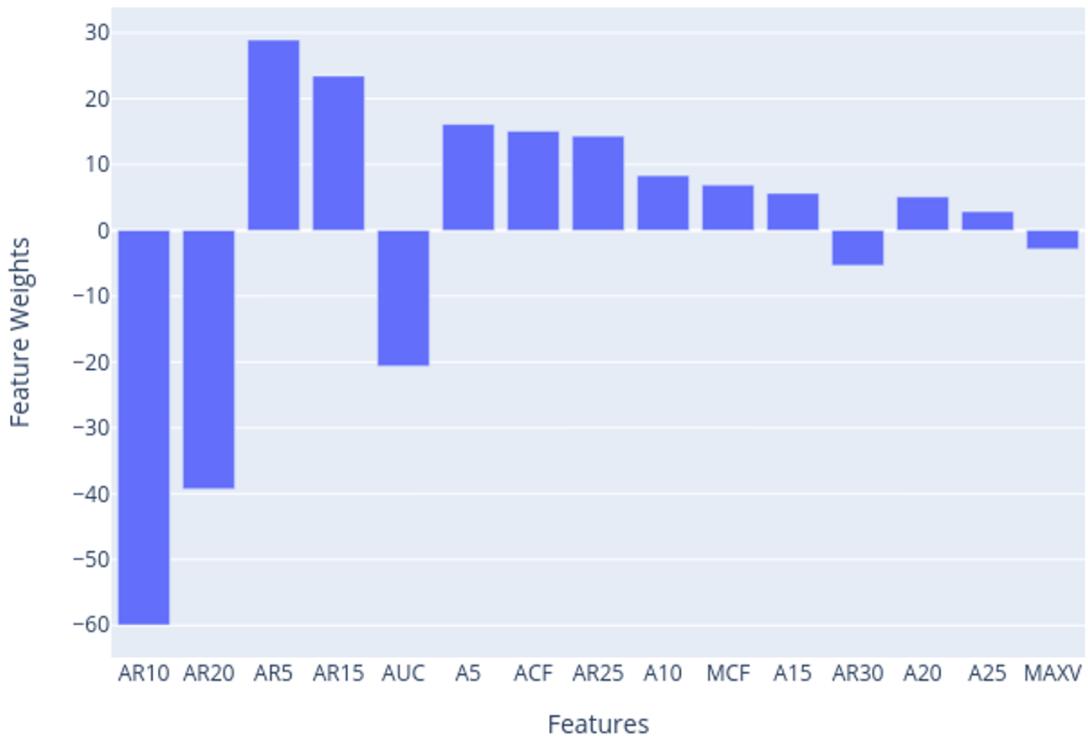
**Table 6: Bleeder Prediction: Performance metrics**



**Figure 19: Bleeder Prediction: Decision Tree (Splits Visualised)**

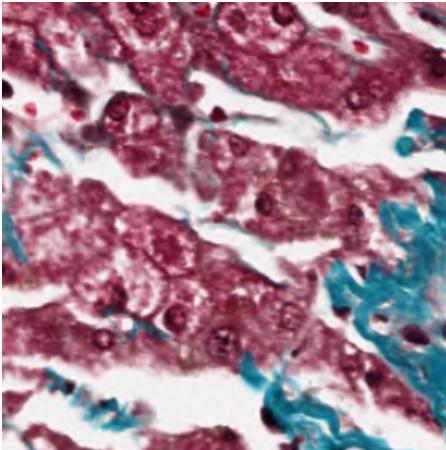
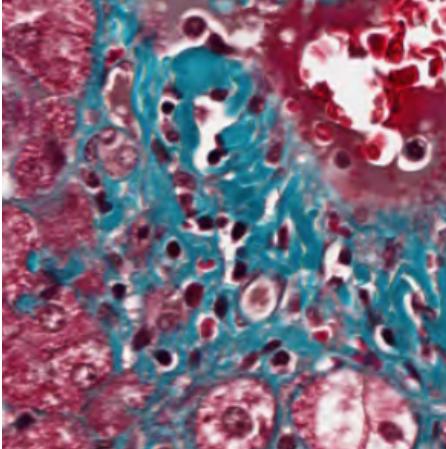
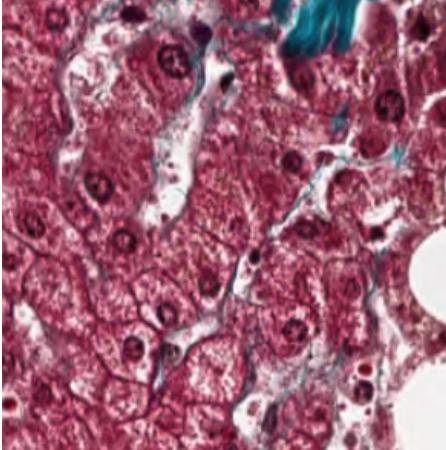


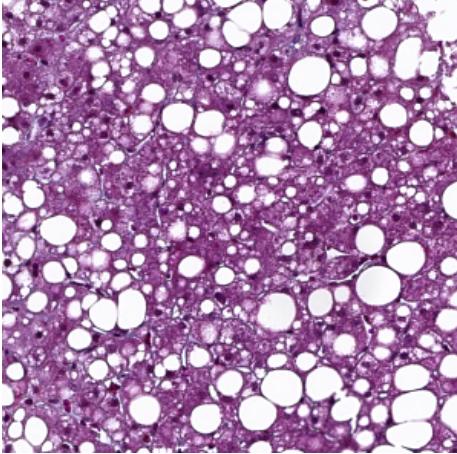
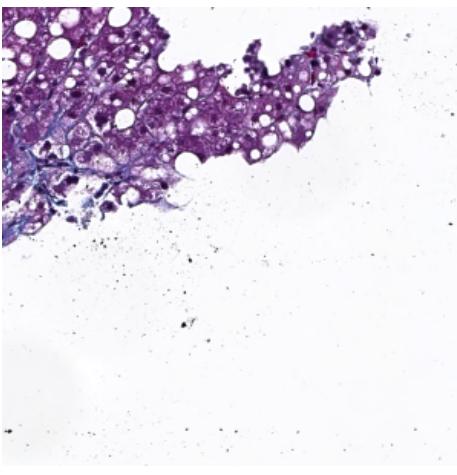
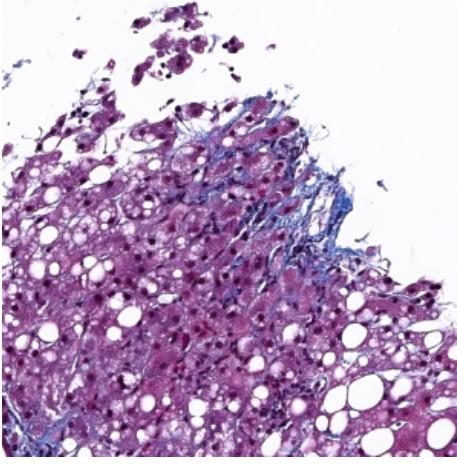
**Figure 20: Bleeder Prediction: SHAP Feature Importance**



**Figure 21: Bleeder Prediction: Logistic Regression Feature Weights**

## Histological Study

Image	Score	Predicted (% confidence)	Actual
	Steatosis	0-5% (84%)	0.51%
	Activity	0 (71%)	1
	Steatosis	0-5% (95%)	0.51%
	Activity	3 (92%)	3
	Steatosis	5-10% (49%)	0.51%
	Activity	1 (47%)	1

	Fibrosis	0 (97%)	0
	Fibrosis	1 (72%)	1
	Fibrosis	2 (69%)	1

**Table 7: NAS Score Prediction Results** a) Activity and Steatosis Score b) Fibrosis Score  
 This model is able to achieve 68.2% accuracy by using pretrained weights and 57.5% by training from scratch. We were able to reproduce the results presented in the paper, but combining the inflammation and ballooning into activity score resulted in a decrease in the performance.

## Chapter 5: Conclusion

---

The analysis we have performed so far has given us the most important of the parameters that predict whether a patient has CLD-Ethanol or some other type of CLD. These include factors like age, sex, AR5-30 values (clot amplitude), ML (maximum lysis), alpha (angle), CFT (clot formation time (20mm)), CT (clotting time) etc. from the SHAP feature importance plots and the splitting nodes of the decision tree. Several features including MCF, age and AR values were top features in both the binary and multiclass CLD-cause classification problems.

We can also observe that the AUC of all these models is not very high across all the classification problems. This can be explained to a great degree by the largely overlapping nature of the data (as seen in the UMAP projections), so it would not be possible to separate those points with an explainable/simple hyperplane.

This CLD-type classification can be very useful in determining the medication for patients who do not want to share/lie about their drinking habits and alcoholism, as most of the time, the medication for treating Hepatitis infections would cause harm to the said alcoholic patient. The cause of CLD plays an important role in ranking patients for transplant and clinical decision making. More extensive studies are needed where data collection takes place on a larger scale.

With the help of pretrained trained Large Language Models and regex filters, I was able to make a parser that can locate and tabulate relevant information. I also got a chance to write a parser to extract biopsy images and NAS activity scores from pathology reports at ILBS.

The bleeder prediction problem is of use to gastroenterologists as variceal/esophageal bleeds are easily preventable. They are often caused by excess blood transfusion in hospitalised CLD patients with esophageal varices or triggered by alcohol consumption. If we can identify the right set of non-invasive predictors to distinguish between nonvariceal and variceal bleeders before performing an endoscopy then it can save more lives. Variceal bleeds are the most common UGIB in cirrhosis patients with a high mortality rate. While performing literature review, I didn't find any major papers where Machine Learning has been employed to perform this classification, and is still an open research problem in the biomedical domain. The biggest drawback with the current analysis is that we had complete information for only 41 bleeders (nonvariceal). The most performant model trained was the Decision Tree Classifier and there was small difference between training and testing that shows that the model doesn't overfit and is generalisable. The first split was made about Portal Hypertension which is related to variceal bleeding. The other important features identified by this model were AR15 and necrosis. These features were seen in the SHAP importance plot of the SVM and were identified by logistic regression also. More complex models like XGBoost were also tried but didn't perform well.

The paper presents a decent automated way to predict various liver scores to quantify NASH using deep learning. We have modified the architecture to fit our needs as we need to predict activity score instead, which combines inflammation and ballooning scores. For example, we don't need to train four copies of the same CNN and a completely different MLP head for each score. Instead, we could use one shared CNN (as most of the image features would be common) and then either use multiple separate heads or just have a shared MLP with the output layer changing as per what scores we want to learn. I had visited ILBS during the break to work on a parser to extract images from pathology reports. However, there are multiple challenges and they can't be directly used to train such models. This is because they have cropped and saved only a few segments of the WSI. The public dataset I worked with has annotations for all tiles of a WSI which are missing in the ILBS reports. We have ground truth for the WSI but only access to cropped images. It will be difficult to estimate the surface area of steatosis without the whole slide image and tile annotations. These are some interesting problems which need attention.

Code is available at [https://github.com/yuktimakhija/cld\\_ml](https://github.com/yuktimakhija/cld_ml).

## References

---

1. Website. Available: Sharma A, Nagalli S. Chronic Liver Disease. [Updated 2022 Jul 4]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK554597/>
2. Kamath PS, Kim WR. The model for end-stage liver disease (MELD). *Hepatology*. 2007;45. doi:10.1002/hep.21563
3. Tsoris A, Marlal CA. Use Of The Child Pugh Score In Liver Disease. StatPearls [Internet]. StatPearls Publishing; 2022.
4. Engelmann C, Thomsen KL, Zakeri N, Sheikh M, Agarwal B, Jalan R, et al. Validation of CLIF-C ACLF score to define a threshold for futility of intensive care support for patients with acute-on-chronic liver failure. *Crit Care*. 2018;22: 1–8.
5. Understanding MELD Score for Liver Transplant. In: UPMC | Life Changing Medicine [Internet]. [cited 23 Sep 2022]. Available: <https://www.upmc.com/services/transplant/liver/process/waiting-list/meld-score>
6. Hearnshaw SA, Logan RFA, Lowe D, Travis SPL, Murphy MF, Palmer KR. Acute upper gastrointestinal bleeding in the UK: patient characteristics, diagnoses and outcomes in the 2007 UK audit. *Gut*. 2011. pp. 1327–1335. doi:10.1136/gut.2010.228437
7. Garcia-Tsao G, Sanyal AJ, Grace ND, Carey W, Practice Guidelines Committee of the American Association for the Study of Liver Diseases, the Practice Parameters Committee of the American College of Gastroenterology. Prevention and management of gastroesophageal varices and variceal hemorrhage in cirrhosis. *Hepatology*. 2007. pp. 922–938. doi:10.1002/hep.21907
8. Yu Z, Zhang Y, Cao Y, Xu M, You S, Chen Y, et al. A dynamic prediction model for prognosis of acute-on-chronic liver failure based on the trend of clinical indicators. *Sci Rep*. 2021;11: 1–13.
9. Younossi ZM, Stepanova M, Afendy M, Fang Y, Younossi Y, Mir H, et al. Changes in the prevalence of the most common causes of chronic liver diseases in the United States from 1988 to 2008. *Clin Gastroenterol Hepatol*. 2011;9. doi:10.1016/j.cgh.2011.03.020
10. Kim BJ, Kim J, Kim S-J, Choi S, Kang K. How do we differentiate variceal bleeding from non-variceal bleeding before emergency endoscopy in patients with acute upper gastrointestinal bleeding? *Eur J Intern Med*. 2013;24: e85.
11. González-González JA, García-Compean D, Vázquez-Elizondo G, Garza-Galindo A, Jáquez-Quintana JO, Maldonado-Garza H. Nonvariceal upper gastrointestinal bleeding in patients with liver cirrhosis. Clinical features, outcomes and predictors of in-hospital mortality. A prospective study. *Ann Hepatol*. 2011;10: 287–295.
12. Roy M, Wang F, Vo H, Teng D, Teodoro G, Farris AB, et al. Deep-learning-based accurate hepatic steatosis quantification for histological assessment of liver biopsies. *Lab Invest*. 2020;100: 1367–1383.

13. An evolutionary algorithm-based optimization method for the classification and quantification of steatosis prevalence in liver biopsy images. 2021;11: 100078.
14. Nonalcoholic fatty liver disease. In: Mayo Clinic [Internet]. 22 Sep 2021 [cited 23 Sep 2022]. Available:  
<https://www.mayoclinic.org/diseases-conditions/nonalcoholic-fatty-liver-disease/symptoms-causes/syc-20354567>
15. Crochemore T, de Toledo Piza FM, dos Reis Rodrigues R, de Campos Guerra JC, Ferraz LJR, Corrêa TD. A new era of thromboelastometry. Einstein . 2017;15: 380.
16. Thromboelastometry. Wikimedia Foundation, Inc.; 15 May 2009 [cited 23 Sep 2022]. Available:  
<https://en.wikipedia.org/wiki/Thromboelastometry>
17. Neumann M, King D, Beltagy I, Ammar W. Scispacy: Fast and Robust Models for Biomedical Natural Language Processing. 2019 [cited 23 Sep 2022]. Available:  
<https://www.aclweb.org/anthology/W19-5034.pdf>
18. Heinemann F. Deep learning-based quantification of NAFLD/NASH progression in human liver biopsies. 2022 [cited 2 Jan 2023]. Available: <https://osf.io/8e7hd/>
19. Kormilitzin A, Vaci N, Liu Q, Nevado-Holgado A. Med7: A transferable clinical natural language processing model for electronic health records. Artif Intell Med. 2021;118.  
doi:10.1016/j.artmed.2021.102086