Cirrhosis Prediction

By: Sheryl, Anjali, Qiao Shi
Lab Group: W132

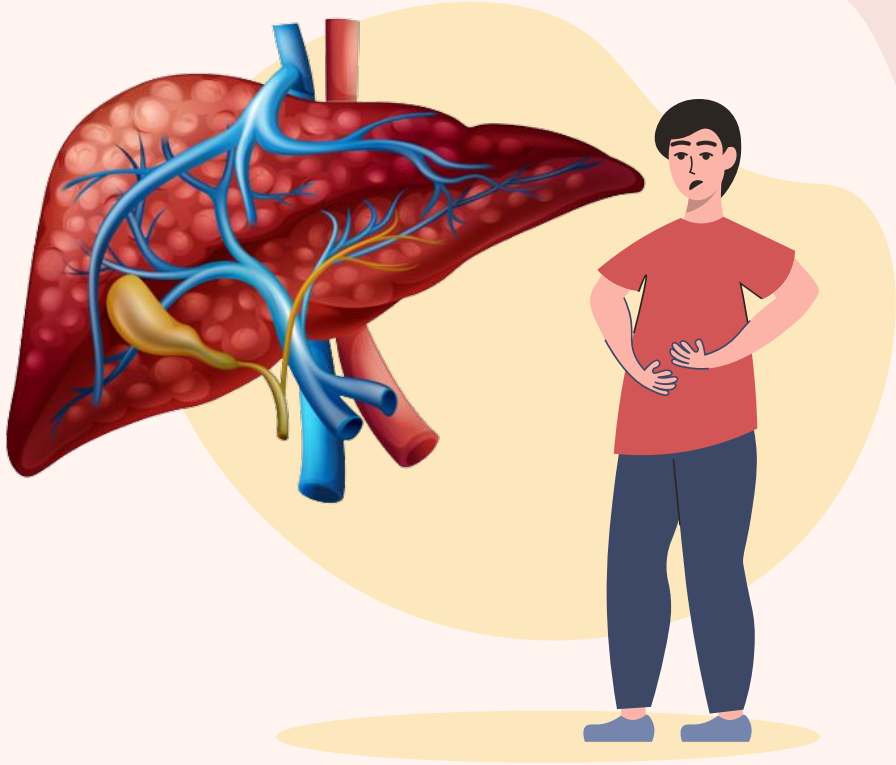
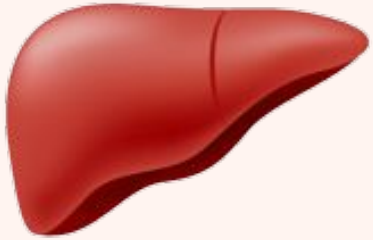


TABLE OF CONTENTS

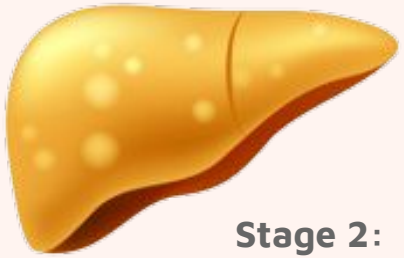
What is Cirrhosis?

Cirrhosis is a **chronic liver disease** where the liver is scarred and permanently damaged

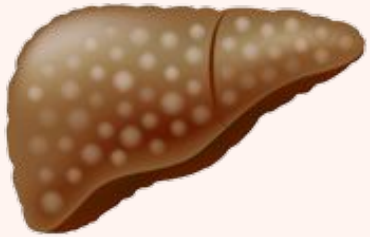
Stages of disease



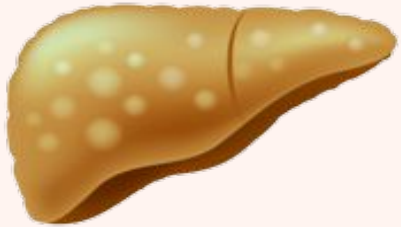
Stage 1:
Healthy Liver



Stage 2:
Fatty Liver

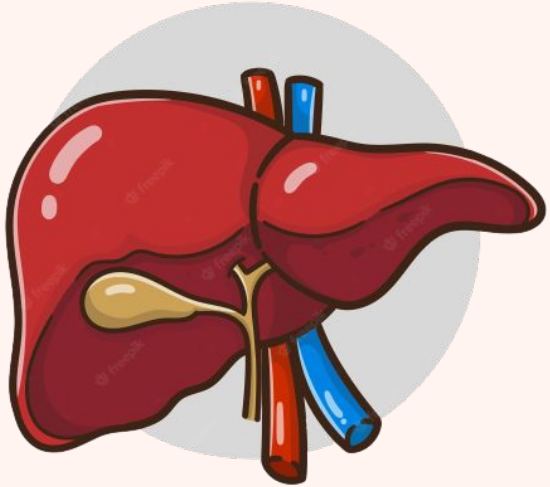


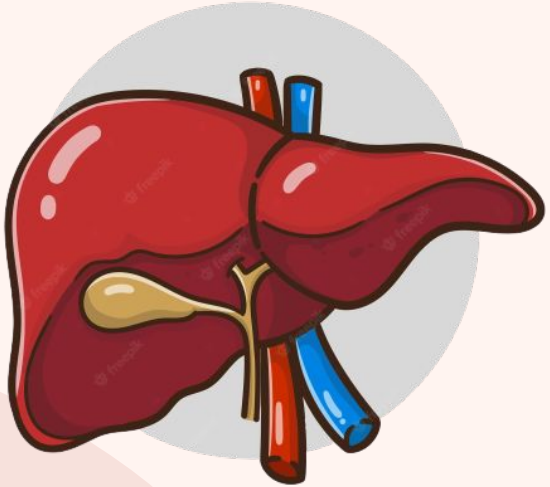
Stage 4:
Cirrhosis



Stage 3:
Liver Fibrosis

Can we predict if a patient is in early or late stage of liver cirrhosis?





Which factors will be most useful in assisting our prediction?

Mayo Clinic Trial

418

Patient Records

20

Information attributes

Dealing with NULL Values

```
In [68]: liverData.isnull().sum()

Out[68]: ID                0
         N_Days            0
         Status            0
         Drug            106
         Age               0
         Sex               0
         Ascites         106
         Hepatomegaly    106
         Spiders         106
         Edema             0
         Bilirubin         0
         Cholesterol     134
         Albumin           0
         Copper          108
         Alk_Phos        106
         SGOT            106
         Tryglicerides   136
         Platelets        11
         Prothrombin       2
         Stage             6
         dtype: int64
```

1. Remove rows with NULL value for stage

```
In [68]: liverData.isnull().sum()

Out[68]: ID                0
         N_Days            0
         Status            0
         Drug            106
         Age               0
         Sex               0
         Ascites         106
         Hepatomegaly    106
         Spiders         106
         Edema             0
         Bilirubin         0
         Cholesterol     134
         Albumin           0
         Copper          108
         Alk_Phos        106
         SGOT            106
         Tryglicerides   136
         Platelets        11
         Prothrombin       2
         Stage             6
         dtype: int64
```

2. Replace NULL values with the mode of each column

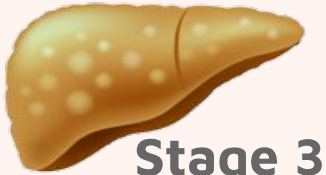
```
In [68]: liverData.isnull().sum()

Out[68]: ID                0
         N_Days            0
         Status            0
         Drug            106
         Age               0
         Sex               0
         Ascites         106
         Hepatomegaly    106
         Spiders         106
         Edema             0
         Bilirubin         0
         Cholesterol     134
         Albumin           0
         Copper          108
         Alk_Phos        106
         SGOT            106
         Tryglicerides   136
         Platelets        11
         Prothrombin       2
         Stage             6
         dtype: int64
```

Early Stage

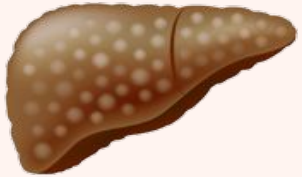
Stage 1:
Healthy Liver

Stage 2:
Fatty Liver

Late Stage



Stage 3:
Liver
Fibrosis

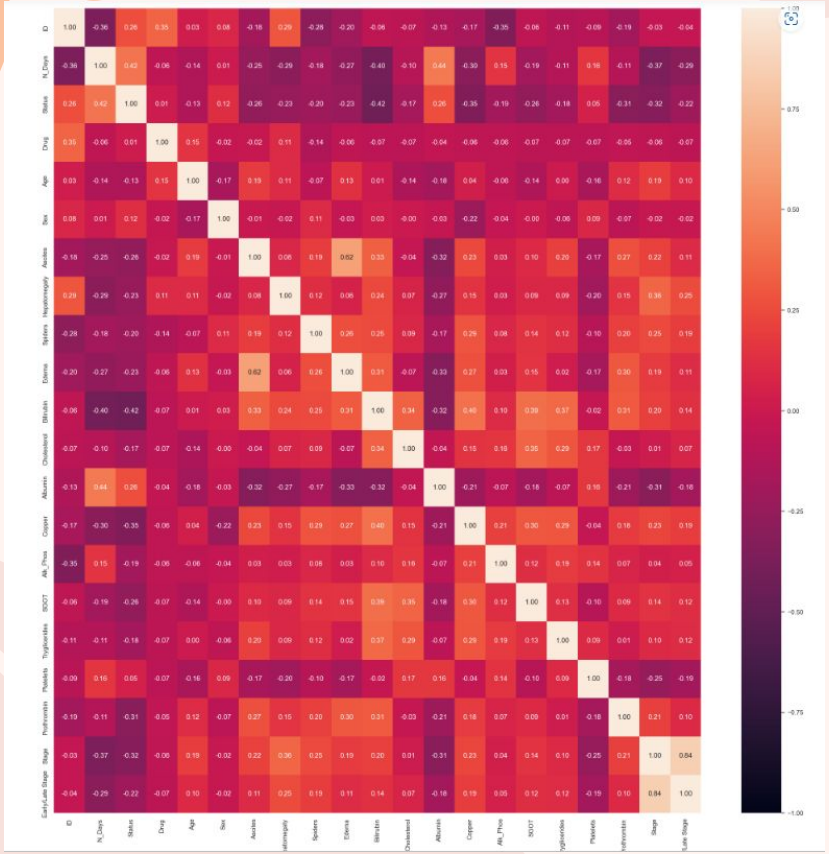


Stage 4:
Cirrhosis

Creating a new column

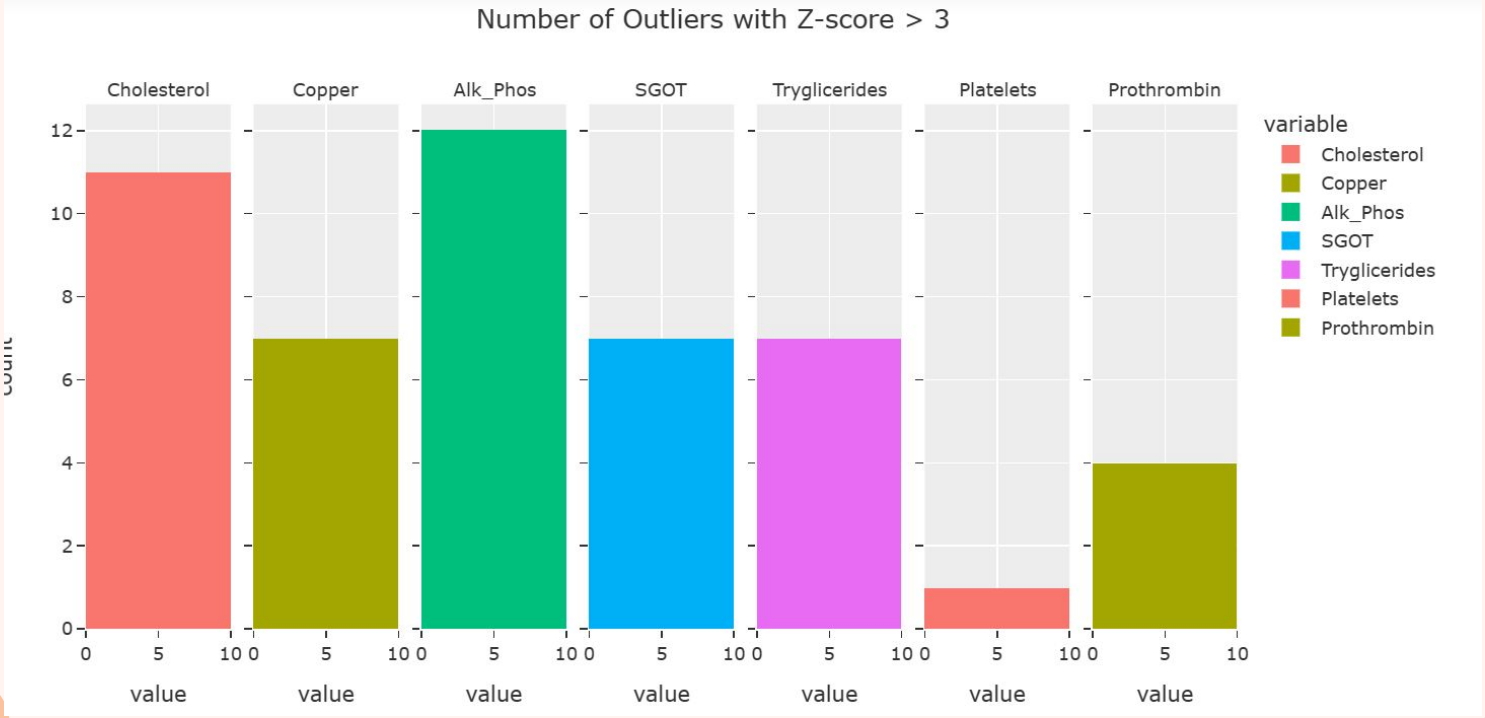
Create a new column on dataset to see whether the patient is in early or late stage of disease.

In [9]:
```
# If 0, patient is in early stage.
# If 1, patient is in late stage.
# We first expressed it in numeric form for comparison.

def Early_Late_Stage(liverData):
    if (liverData["Stage"] == 1.0) or (liverData["Stage"] == 2.0):
        return 0
    else:
        return 1

liverData["Early/Late Stage"] = liverData.apply(lambda liverData: Early_Late_Stage(liverData), axis=1)
```

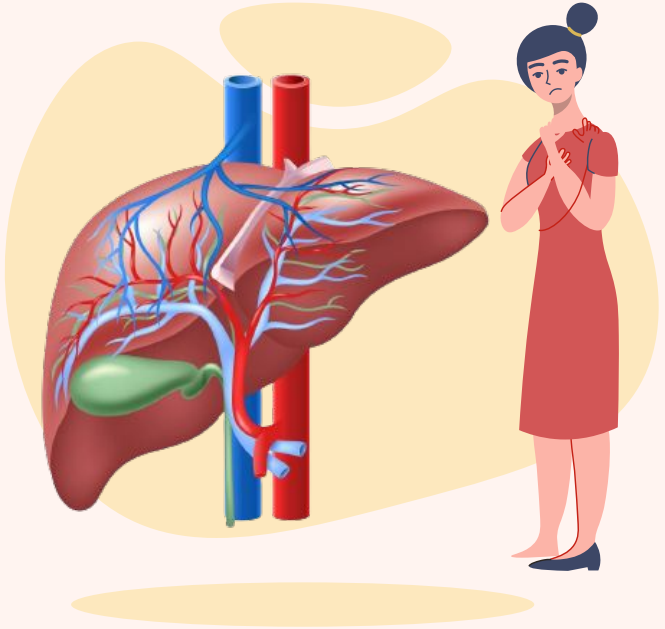


Remove the ID column

Removing outliers



02 Exploratory Analysis



Exploratory Data Analysis

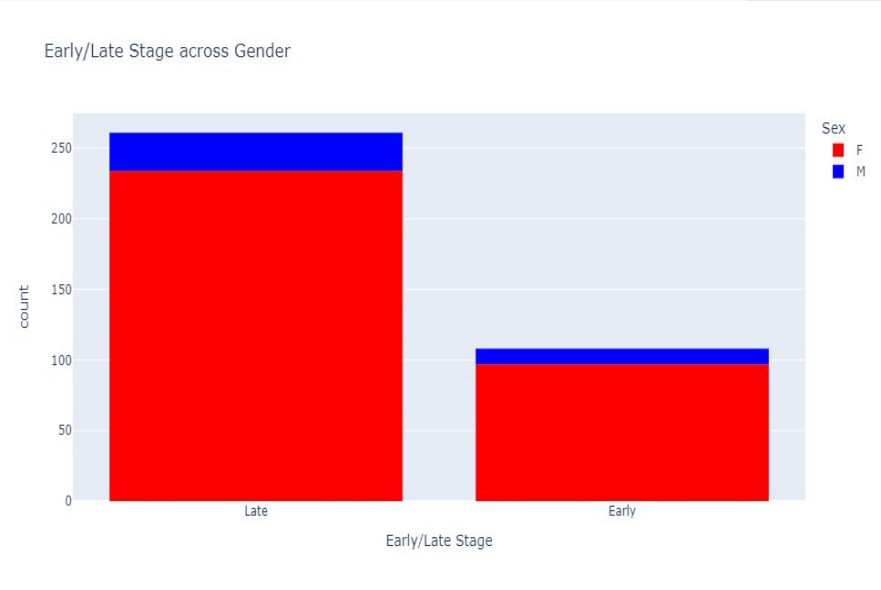
Contribution of Factors Across Stages

Correlation of Factors with Stage

Exploring Discrete and Continuous data

Contribution of Factors Across Stages

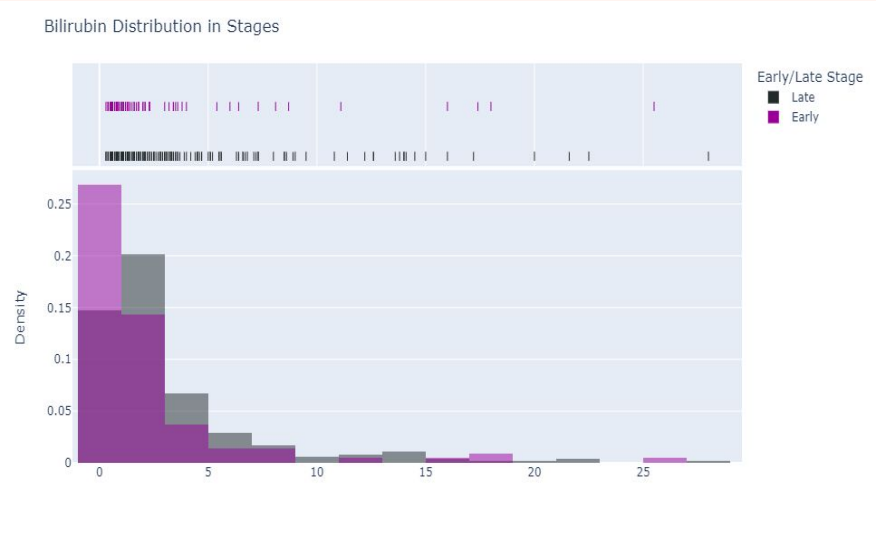
- Females have a higher tendency of developing liver disease than males



Using Plotly

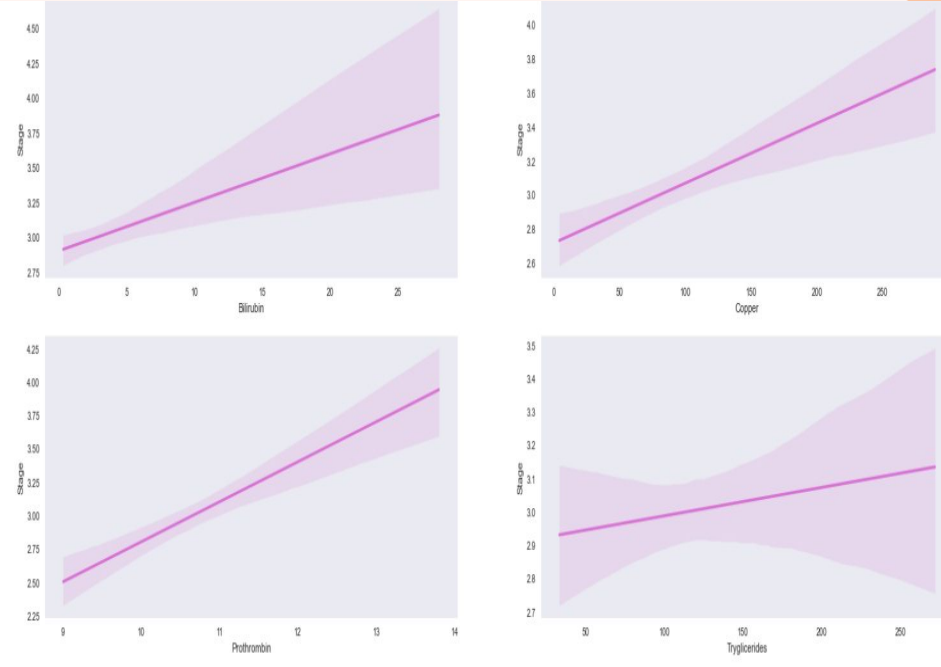
Contribution of Factors Across Stages

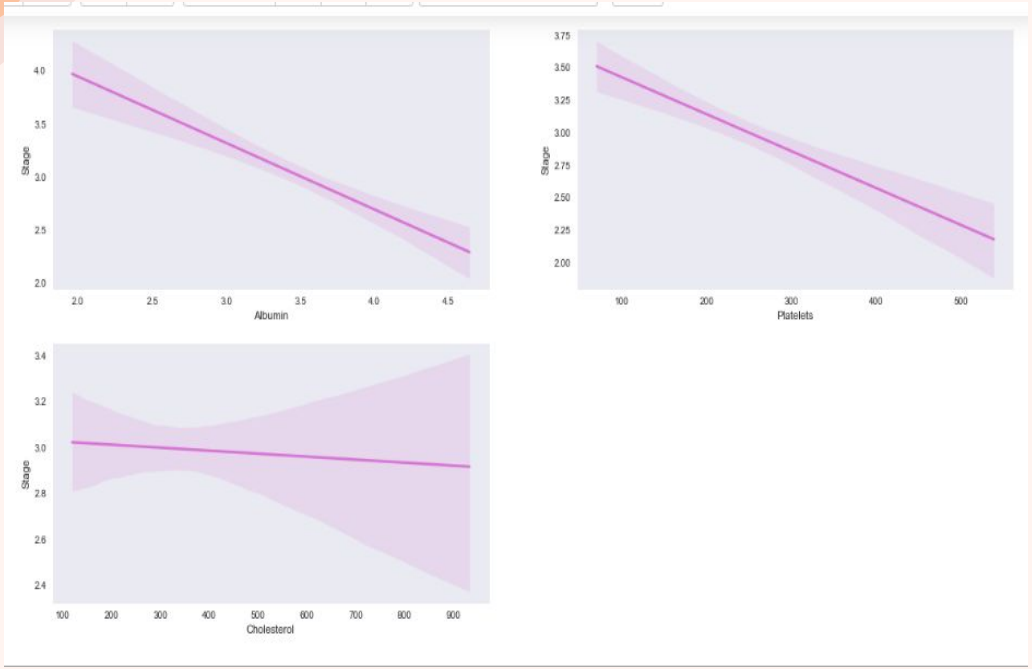
```
fig = px.histogram(liverData, x='Bilirubin', color='Early/Late Stage', nbins=20, marginal='rug', barmode='overlay',
                   histnorm='probability density', color_discrete_sequence=['#222A2A', '#990099', '#ecb4d4'])
fig.update_layout(title='Bilirubin Distribution in Stages',
                  xaxis_title='', yaxis_title='Density', xaxis_ticks='', yaxis_ticks='')
```



Positive Correlation

- ★ Bilirubin
- ★ Triglyceride
- ★ Copper
- ★ Prothrombin

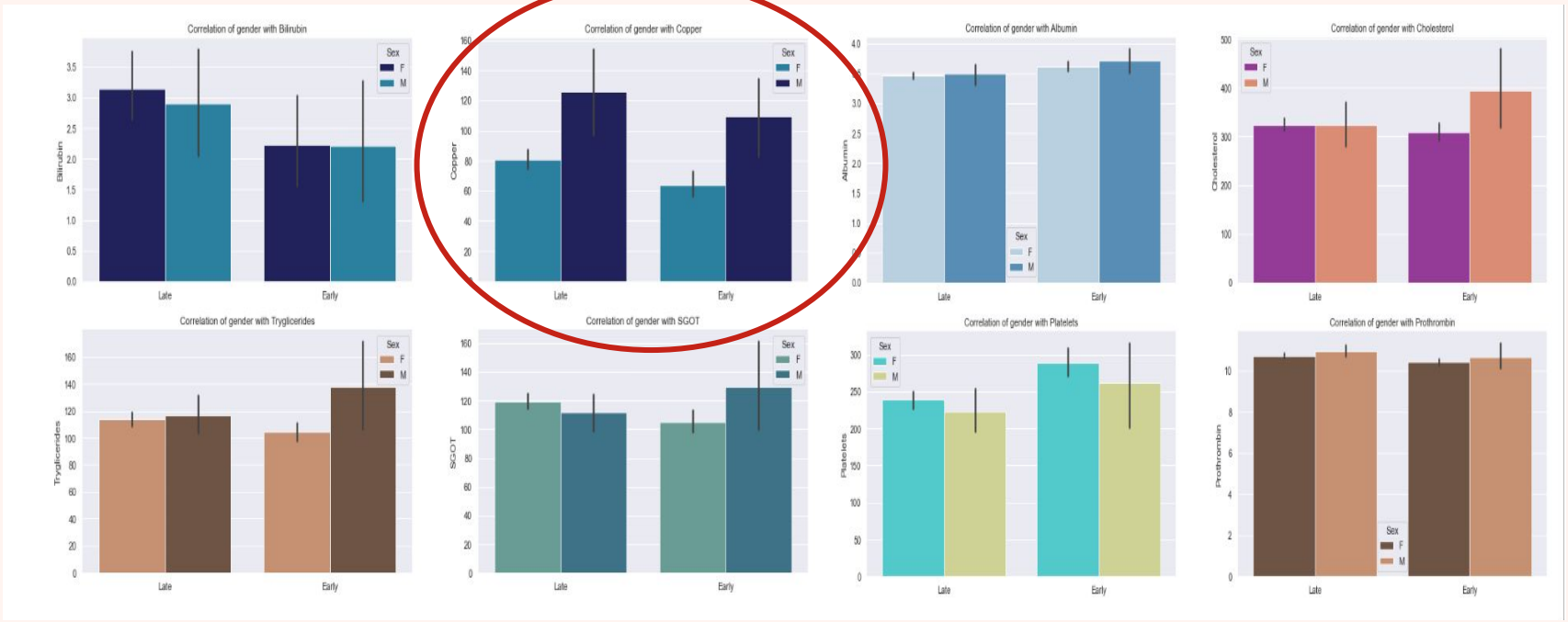




Negative Correlation

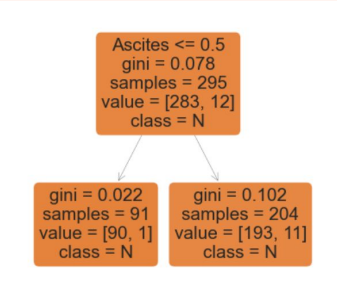
- ★ Albumin
- ★ Platelets
- ★ Cholesterol

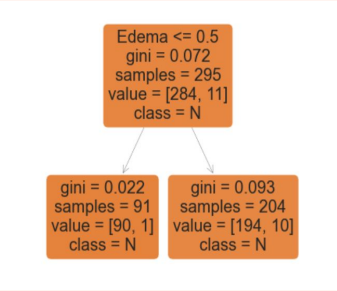
Correlation of Gender With Factors

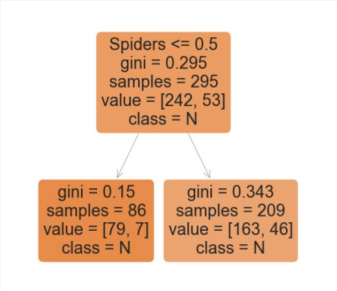


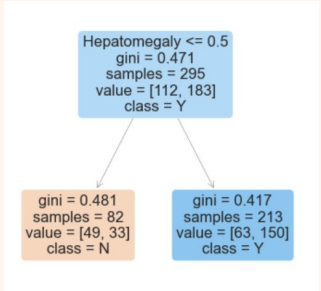
Exploring Discrete Data

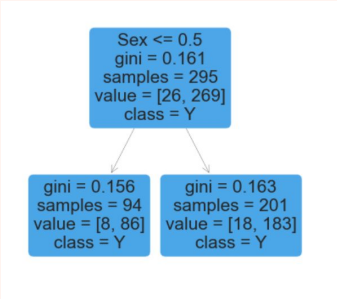
Decision Tree Classifier









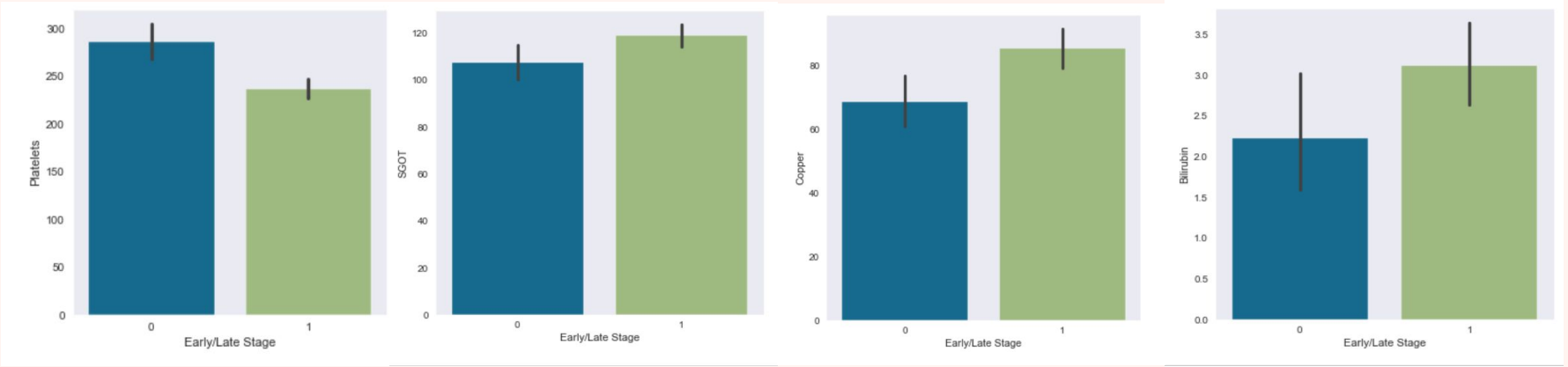


Low Gini Index

- Ascites
- Edema
- Spiders

Exploring Continuous Data

Bar Plot

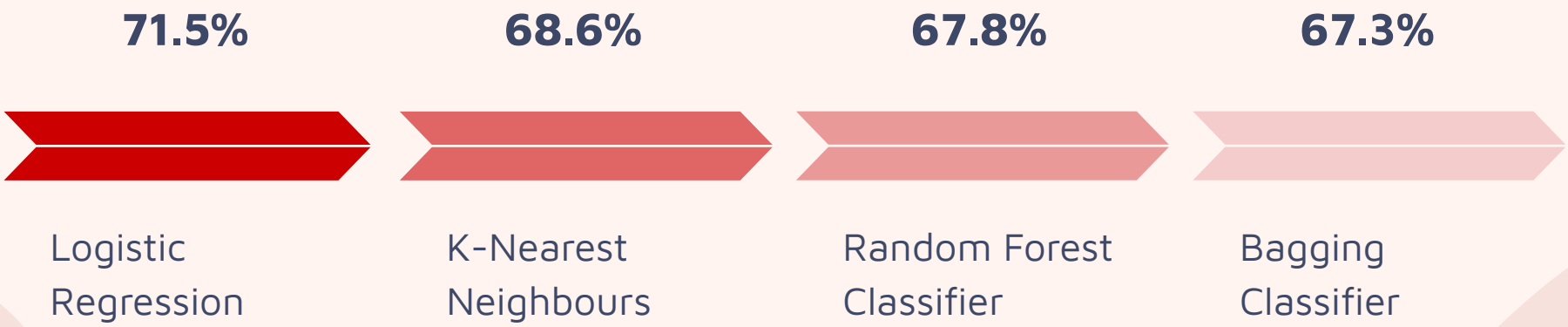


03 Machine Learning

Machine Learning Models Used

1. Random Forest Classifier
2. Logistic Regression
3. K-Nearest Neighbour (KNN) Classifier
4. Bagging Classifier

Accuracy of Models



Receiver Operating Characteristic Curve

Receiver Operating Characteristic Curve

1. Area Under Curve (AUC)

The higher the AUC, the better the overall performance

2. Difference between training and validation accuracy

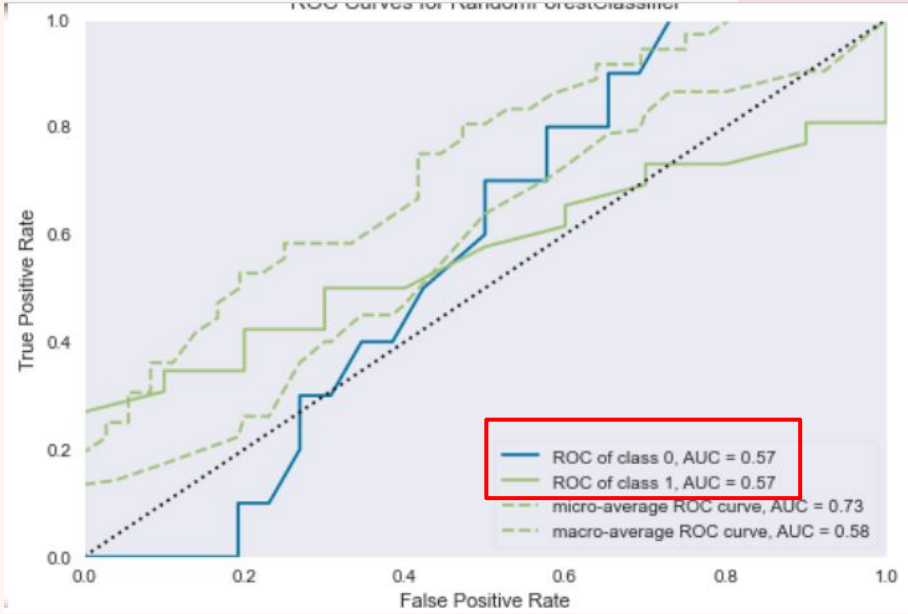
The smaller the difference, the lower the tendency to overfit, thus it works better for new, unseen data

Random Forest Classifier

Random Forest

```
[[ 0 10]
 [ 3 23]]

Training Acc.  : 100.0%
Validation Acc.: 63.89%
```

	precision	recall	f1-score	support
0	0.00	0.00	0.00	10
1	0.70	0.88	0.78	26
accuracy			0.64	36
macro avg	0.35	0.44	0.39	36
weighted avg	0.50	0.64	0.56	36



Tendency to Overfit: 36.11 percentage points
Area Under Curve (AUC): 0.57

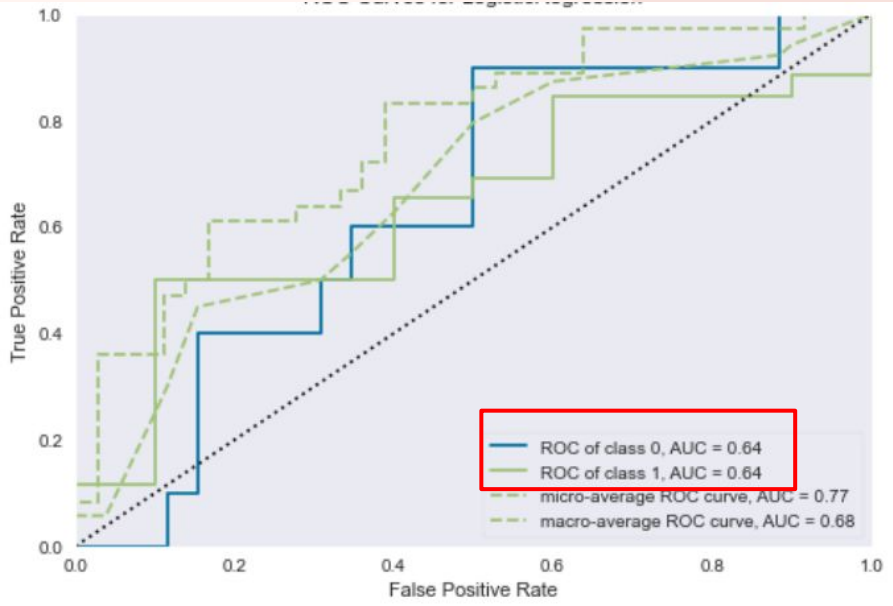
Logistic Regression Model

Logistic Regression

```
[[ 1  9]
 [ 3 23]]
```

Training Acc. : 73.87%
Validation Acc.: 66.67%

	precision	recall	f1-score	support
0	0.25	0.10	0.14	10
1	0.72	0.88	0.79	26
accuracy			0.67	36
macro avg	0.48	0.49	0.47	36
weighted avg	0.59	0.67	0.61	36

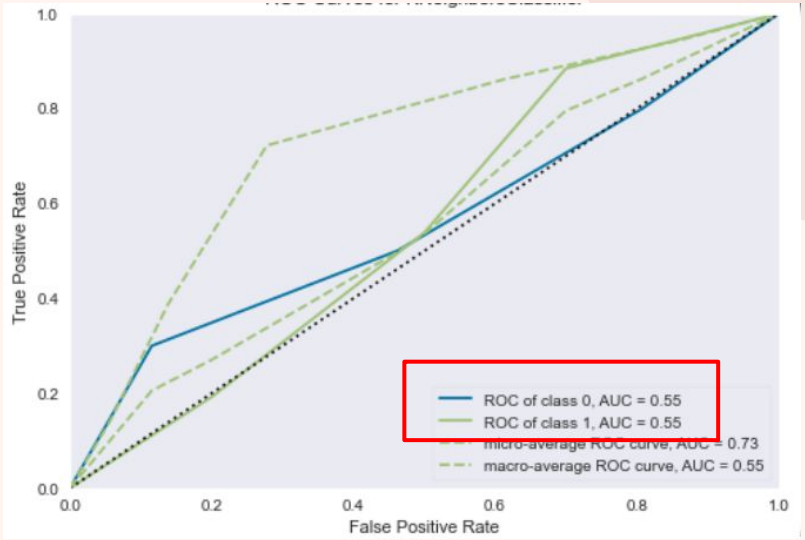


Tendency to Overfit: 7.2 percentage points
Area Under Curve (AUC): 0.64

K-Nearest Neighbours

KNN

```
[[ 3  7]
 [ 3 23]]

Training Acc.  : 74.77%
Validation Acc.: 72.22%
              precision    recall  f1-score   support

           0       0.50      0.30      0.37        10
           1       0.77      0.88      0.82        26

    accuracy                           0.72        36
   macro avg       0.63      0.59      0.60        36
weighted avg       0.69      0.72      0.70        36
```



Tendency to Overfit: 2.55 percentage points
Area Under Curve (AUC): 0.55

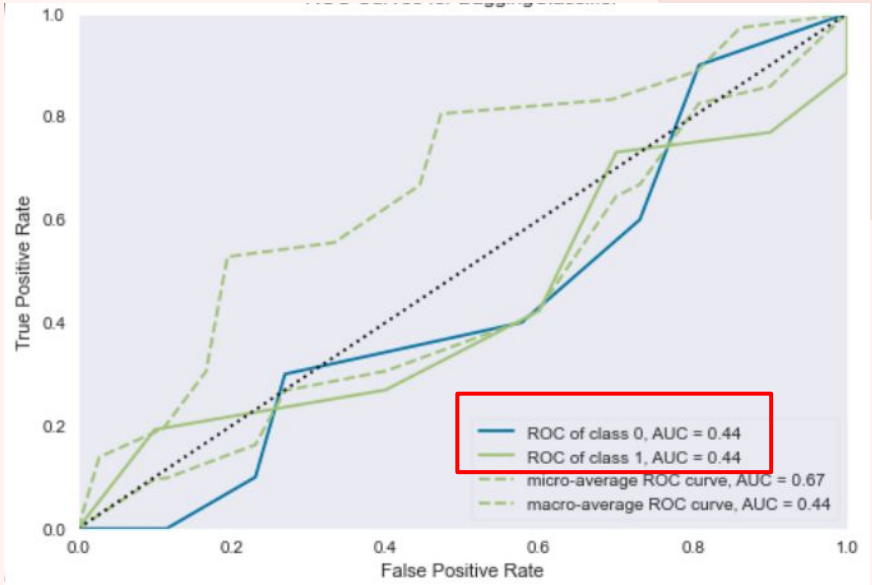
Bagging Classifier

Bagging Classifier

```
[[ 2  8]
 [ 5 21]]
```

Training Acc. : 98.5%
Validation Acc.: 63.89%

	precision	recall	f1-score	support
0	0.29	0.20	0.24	10
1	0.72	0.81	0.76	26
accuracy			0.64	36
macro avg	0.50	0.50	0.50	36
weighted avg	0.60	0.64	0.62	36



Tendency to Overfit: 34.61 percentage points
Area Under Curve (AUC): 0.44

Summary of Models

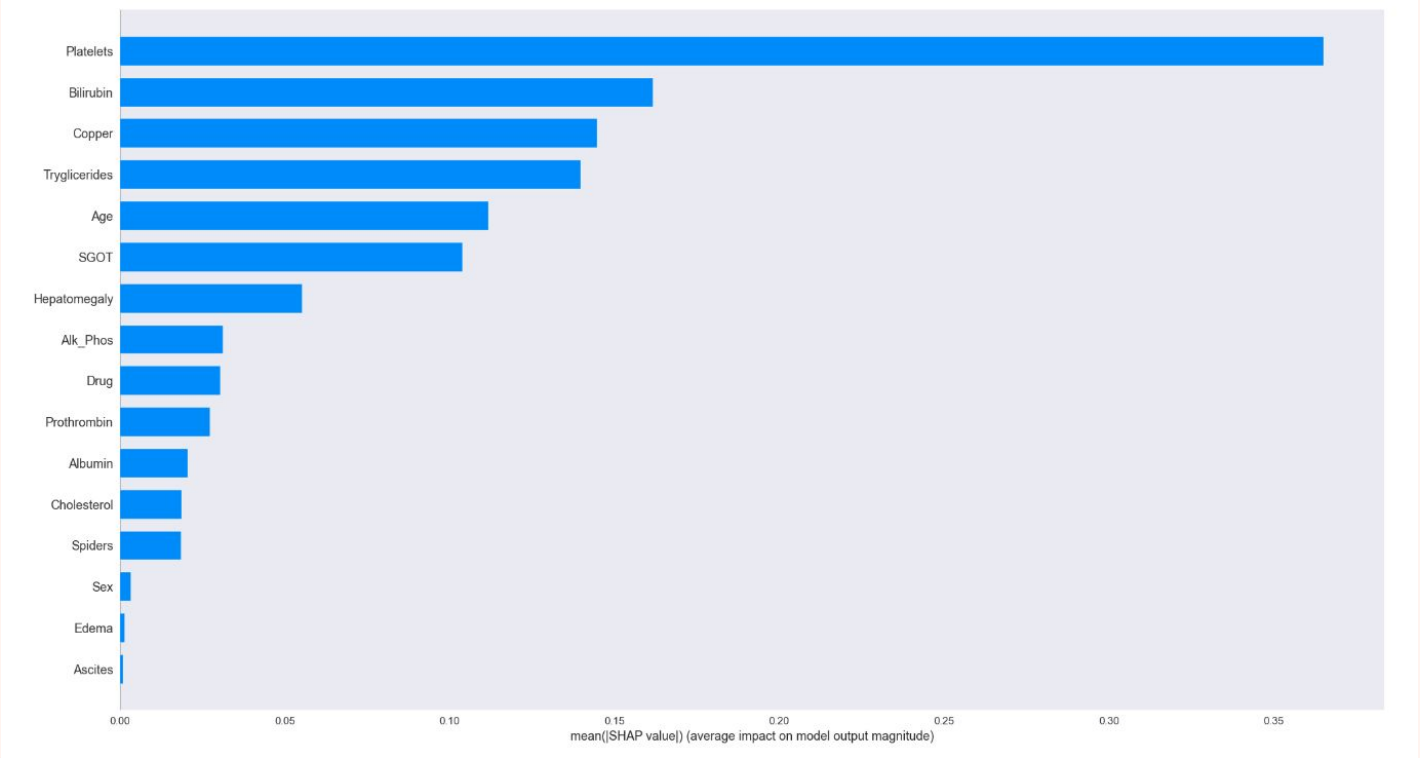
Model	Random Forest	Logistic Regression	K-Nearest Neighbours	Bagging Classifier
Accuracy	67.8%	71.5%	68.6%	67.3%
Tendency to Overfit	Highest	Low	Lowest	High
Area Under Curve (AUC)	0.57	0.64	0.55	0.44

Summary of Models

Model	R[illegible] Forest	Logistic Regression	K-Nearest Neighbours	Ba[illegible] Classifie
Accuracy	7[illegible]%	71.5%	68.6%	7[illegible]%
Tendency to Overfit	High[illegible]st	Low	Lowest	High
Area Under Curve (AUC)	0.57	0.64	0.55	0.44

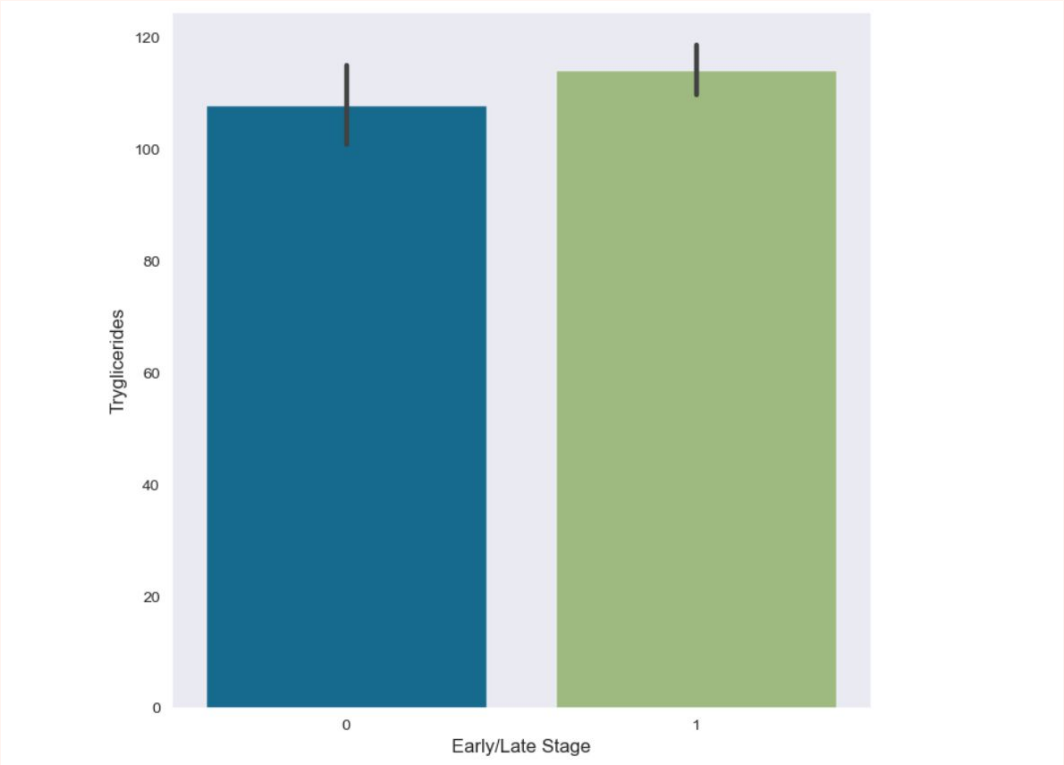
04 Conclusion

Identifying key factors

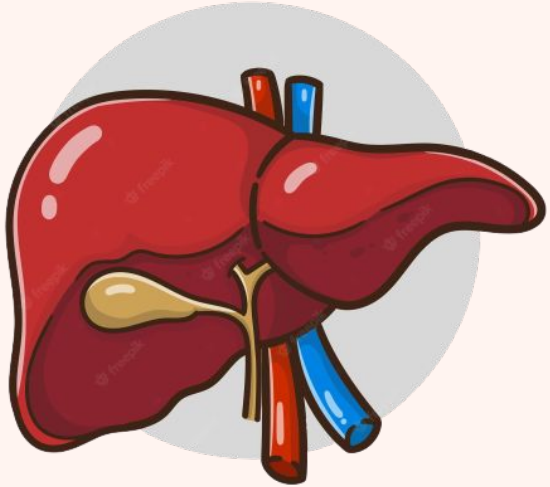


Identifying key factors

Triglicerides?



Can we predict if a patient is in early or late stage of liver cirrhosis?



Yes

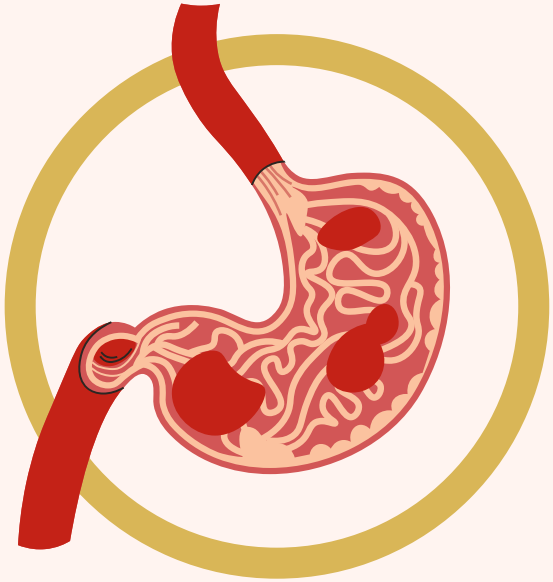
Logistic Regression Model

Key factors for prediction

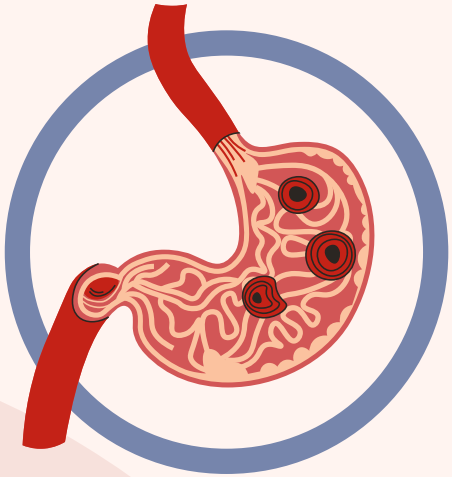
1 – Platelets

2 – Bilirubin

3 – Copper



With the derived results, what are some recommendations?



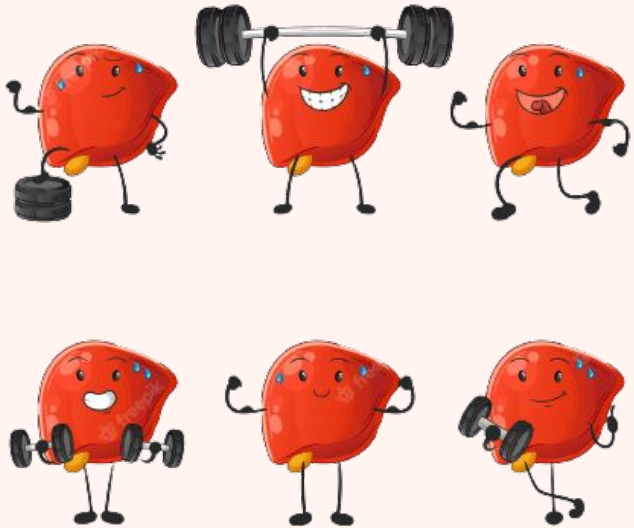
Recommendation (model):
Larger training data

Recommendations: Healthcare providers

To improve time efficiency, nurses can just collect patients' data regarding these 3 factors

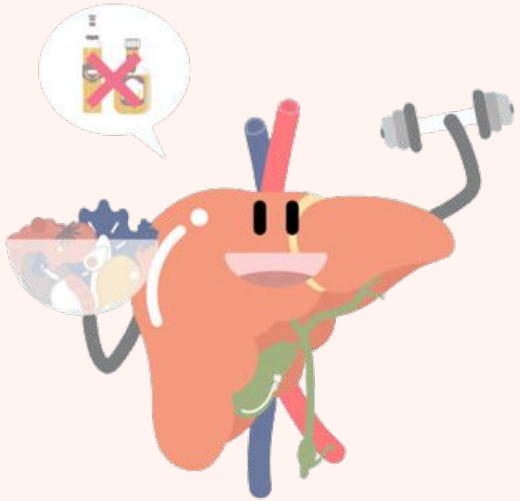
Late stage Cirrhosis patients should :

1. Exercise regularly to help blood flow and increase platelet count



It is imperative for late stage Cirrhosis patients to :

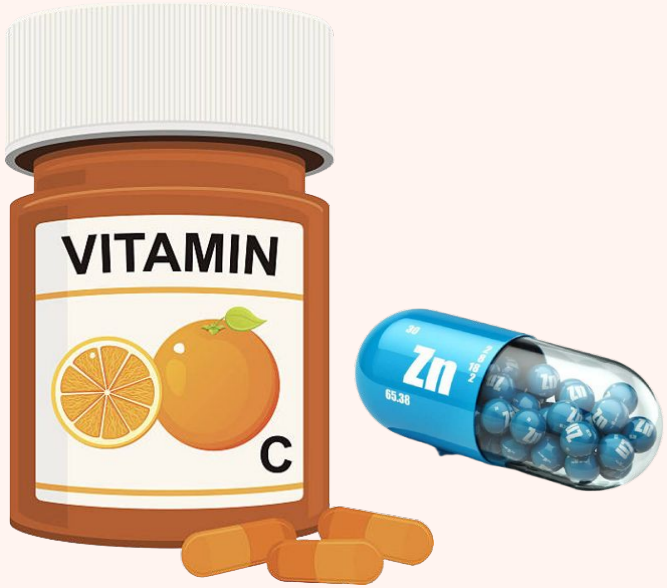
2. Have a healthier diet
   - Drink more water
   - Cut back on alcohol consumption
   - Eat more fruits and vegetables
   - Eat fewer processed foods



Late stage Cirrhosis patients should :

3. Consume food low in copper

4. Under doctor's advice
   - Take more vitamin Cs
   - Take zinc supplements



Thank you!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics and images by **Freepik**

References

https://images.app.goo.gl/ffuJVPmjK38prsZz5
https://images.app.goo.gl/dhBFCCSS9w3UqMZe9
Premium Vector | Liver exercise set character cartoon mascot vector (freepik.com)
Healthy Liver Internal Organs Anatomy Body Part Nervous System Infographic Health Care Concept Stock Illustration - Download Image Now - iStock (istockphoto.com)
Vitamin C Bottle Stock Illustration - Download Image Now - Vitamin C, Nutritional Supplement, Bottle - iStock (istockphoto.com)
Pills with Zinc Zn Element Dietary Supplements. Vitamin Capsules Stock Illustration - Illustration of dietary, diet: 62351501 (dreamstime.com)