Final Project Liver Disease Prediction

Group 2: Selina Meng, Tina Zhang, Tianchuzi Qin, Yuyang Chen

In [1]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import scipy.stats as stats
```

In [2]:
```
df = pd.read_csv('Liver Patient Dataset (LPD)_train.csv', encoding='ISO-8859-1
```

EDA

In [3]:
```
df
```

Out[3]:

	Age of the patient	Gender of the patient	Total Bilirubin	Direct Bilirubin	Alkphos Alkaline Phosphotase	Sgpt Alamine Aminotransferase	Sgot Aspartate Aminotransferase
0	65.0	Female	0.7	0.1	187.0	16.0	18.0
1	62.0	Male	10.9	5.5	699.0	64.0	100.0
2	62.0	Male	7.3	4.1	490.0	60.0	68.0
3	58.0	Male	1.0	0.4	182.0	14.0	20.0
4	72.0	Male	3.9	2.0	195.0	27.0	59.0
...	...	...	...	...	...	...	...
30686	50.0	Male	2.2	1.0	610.0	17.0	28.0
30687	55.0	Male	2.9	1.3	482.0	22.0	34.0
30688	54.0	Male	6.8	3.0	542.0	116.0	66.0
30689	48.0	Female	1.9	1.0	231.0	16.0	55.0
30690	30.0	Male	3.1	1.6	253.0	80.0	406.0

30691 rows × 11 columns

In [4]:
```
print(df.columns)
```

```
Index(['Age of the patient', 'Gender of the patient', 'Total Bilirubin',
       'Direct Bilirubin', ' Alkphos Alkaline Phosphotase',
       ' Sgpt Alamine Aminotransferase', 'Sgot Aspartate Aminotransferase',
       'Total Protiens', ' ALB Albumin',
       'A/G Ratio Albumin and Globulin Ratio', 'Result'],
      dtype='object')
```

In [5]:
```
df = df.rename({'Age of the patient': 'Age',
                'Gender of the patient': 'Gender'}, axis='columns')
```

```
df.columns = df.columns.str.strip()
```

In [6]:

```
 # statistics summary
df.describe()
```

Out[6]:

	Age	Total Bilirubin	Direct Bilirubin	Alkphos Alkaline Phosphotase	Sgpt Alamine Aminotransferase	Sgot A Aminotra
count	30689.000000	30043.000000	30130.000000	29895.000000	30153.000000	30229
mean	44.107205	3.370319	1.528042	289.075364	81.488641	111
std	15.981043	6.255522	2.869592	238.537589	182.158850	28
min	4.000000	0.400000	0.100000	63.000000	10.000000	10
25%	32.000000	0.800000	0.200000	175.000000	23.000000	26
50%	45.000000	1.000000	0.300000	209.000000	35.000000	42
75%	55.000000	2.700000	1.300000	298.000000	62.000000	88
max	90.000000	75.000000	19.700000	2110.000000	2000.000000	4929

In [7]:

```
# check null
df.isnull().sum()
```

Out[7]:

```
Age                                     2
Gender                                902
Total Bilirubin                       648
Direct Bilirubin                      561
Alkphos Alkaline Phosphotase          796
Sgpt Alamine Aminotransferase         538
Sgot Aspartate Aminotransferase       462
Total Protiens                        463
ALB Albumin                           494
A/G Ratio Albumin and Globulin Ratio  559
Result                                  0
dtype: int64
```

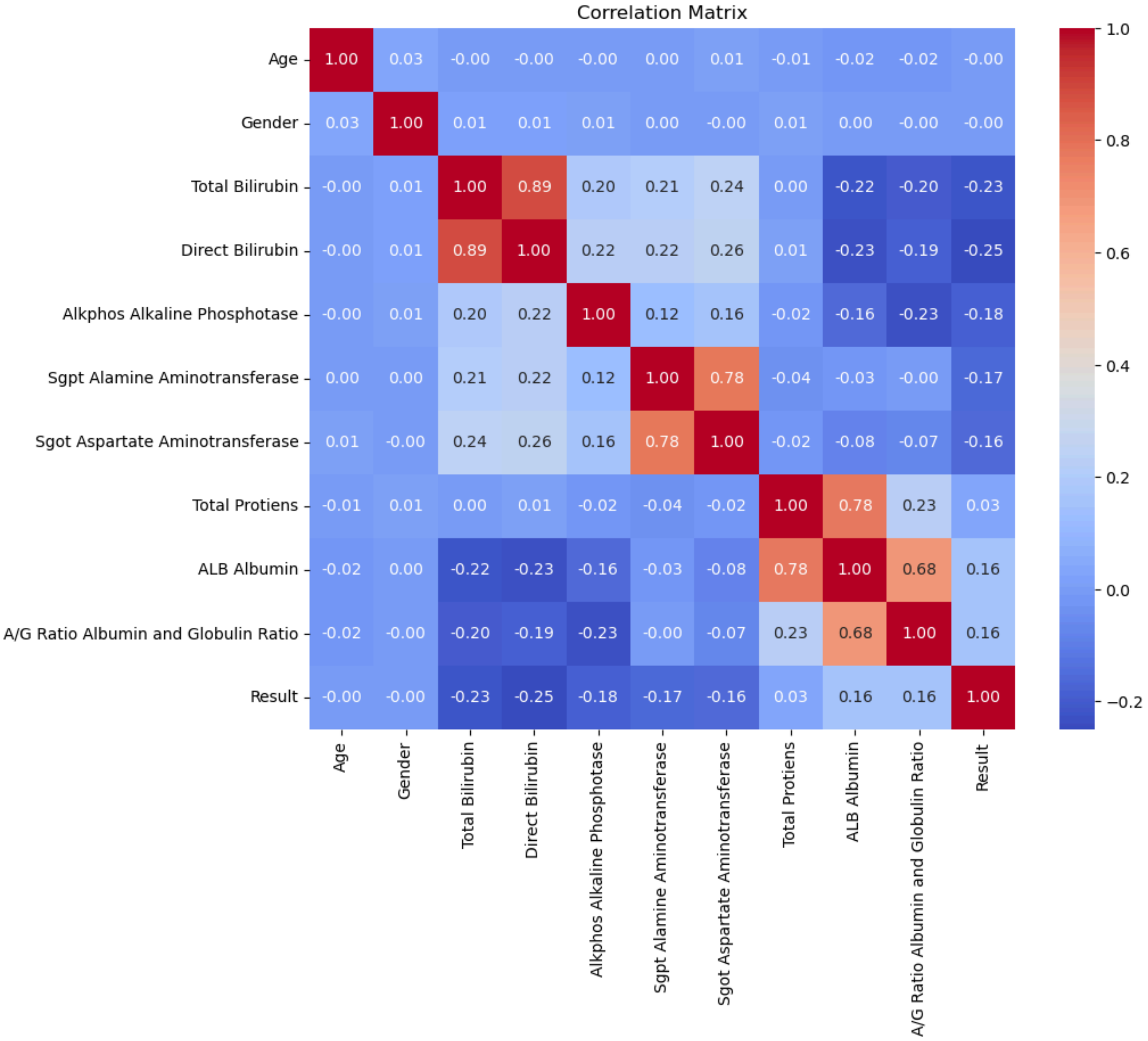
In [8]:

```
# check data types
print(df.dtypes)

# convert categorical x
df['Gender'] = df['Gender'].map({'Male': 1, 'Female': 0})
```

```
Age                                     float64
Gender                                   object
Total Bilirubin                         float64
Direct Bilirubin                        float64
Alkphos Alkaline Phosphotase            float64
Sgpt Alamine Aminotransferase           float64
Sgot Aspartate Aminotransferase         float64
Total Protiens                          float64
ALB Albumin                             float64
A/G Ratio Albumin and Globulin Ratio    float64
Result                                    int64
dtype: object
```

In [9]:

```
# correlation matrix
plt.figure(figsize=(10, 8))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Matrix')
plt.show()
```

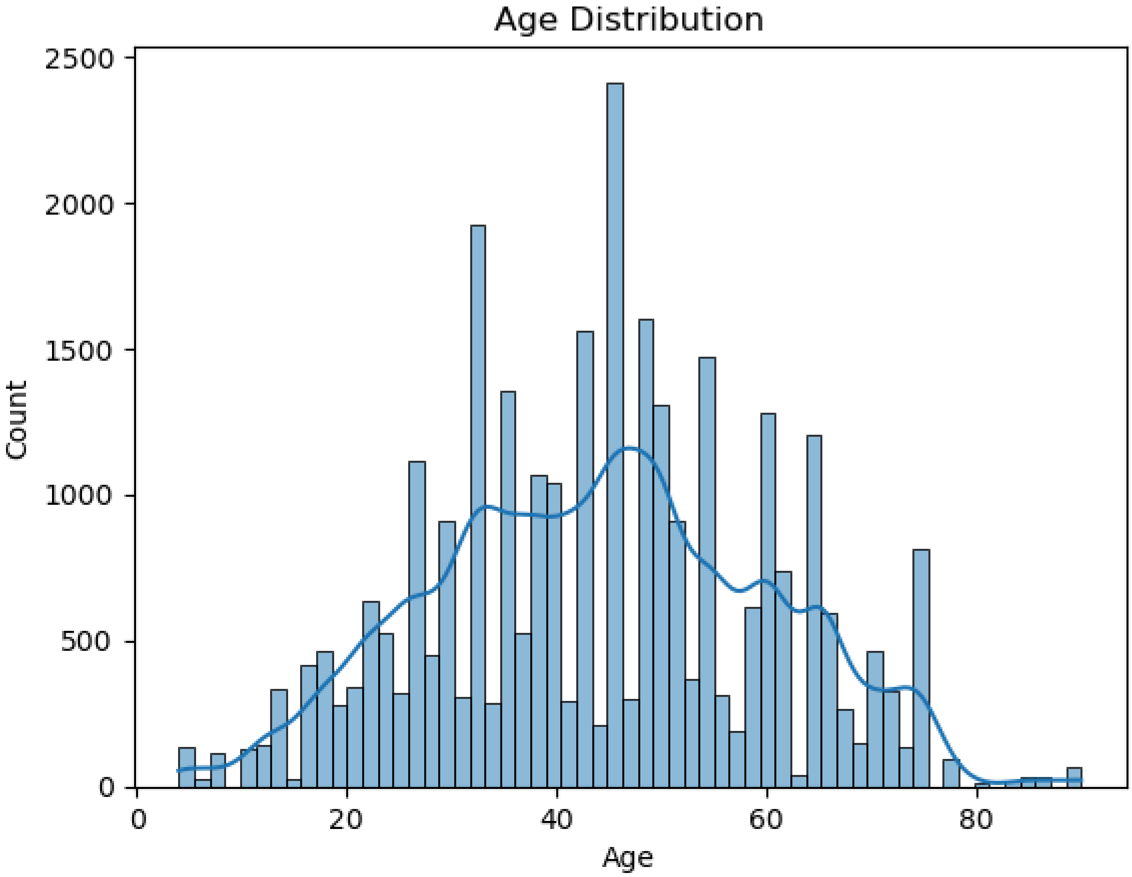


In [10]:

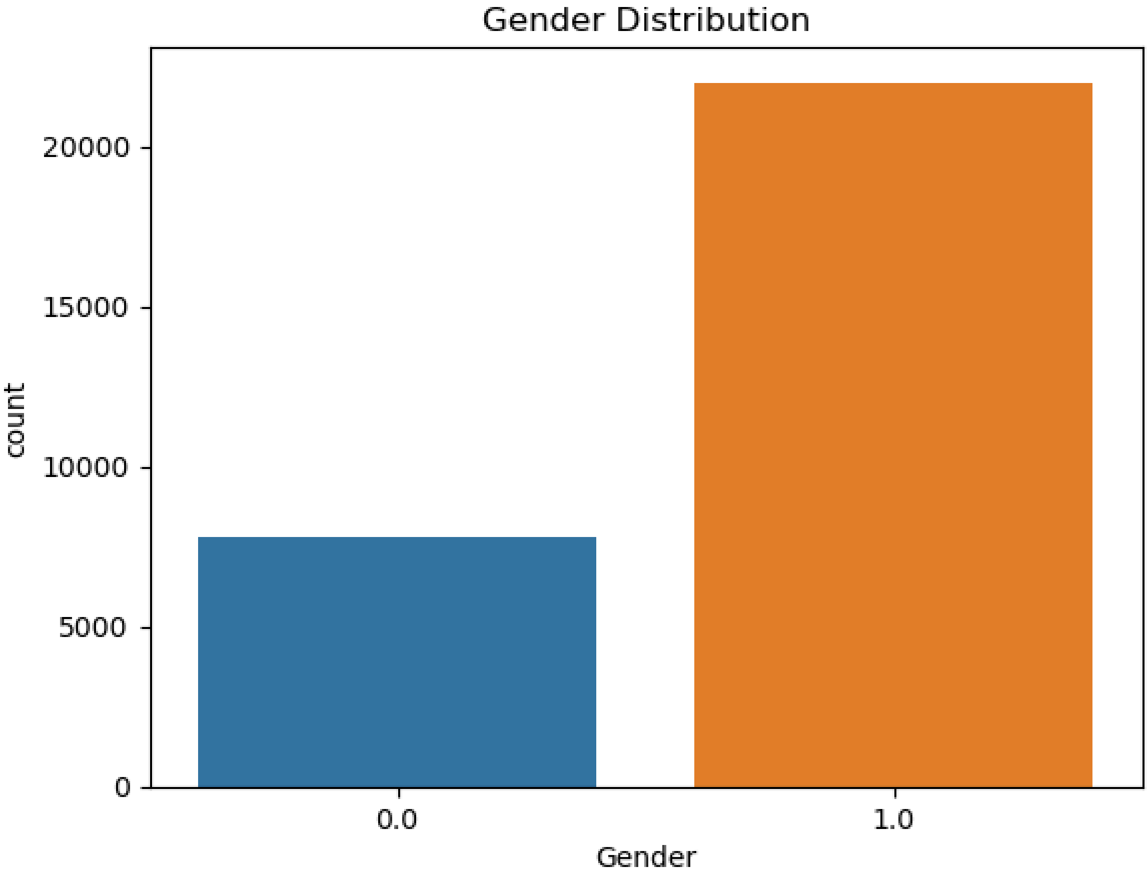
```
# visualization
sns.histplot(df['Age'], kde=True)
plt.title('Age Distribution')
plt.show()

sns.countplot(x='Gender', data=df)
plt.title('Gender Distribution')
plt.show()

# total Bilirubin levels
sns.histplot(df['Total Bilirubin'], kde=True)
plt.title('Total Bilirubin Distribution')
plt.show()
sns.boxplot(x='Result', y='Total Bilirubin', data=df)
plt.title('Total Bilirubin by Disease Status (Result)')
plt.show()

# highly correlated variables
```

```
sns.boxplot(x='Result', y='Sgpt Alamine Aminotransferase', data=df)
plt.title('Sgpt Alamine Aminotransferase by Disease Status (Result)')
plt.show()

sns.boxplot(x='Gender', y='Alkphos Alkaline Phosphotase', data=df)
plt.title('Alkaline Phosphotase by Gender')
plt.show()
```

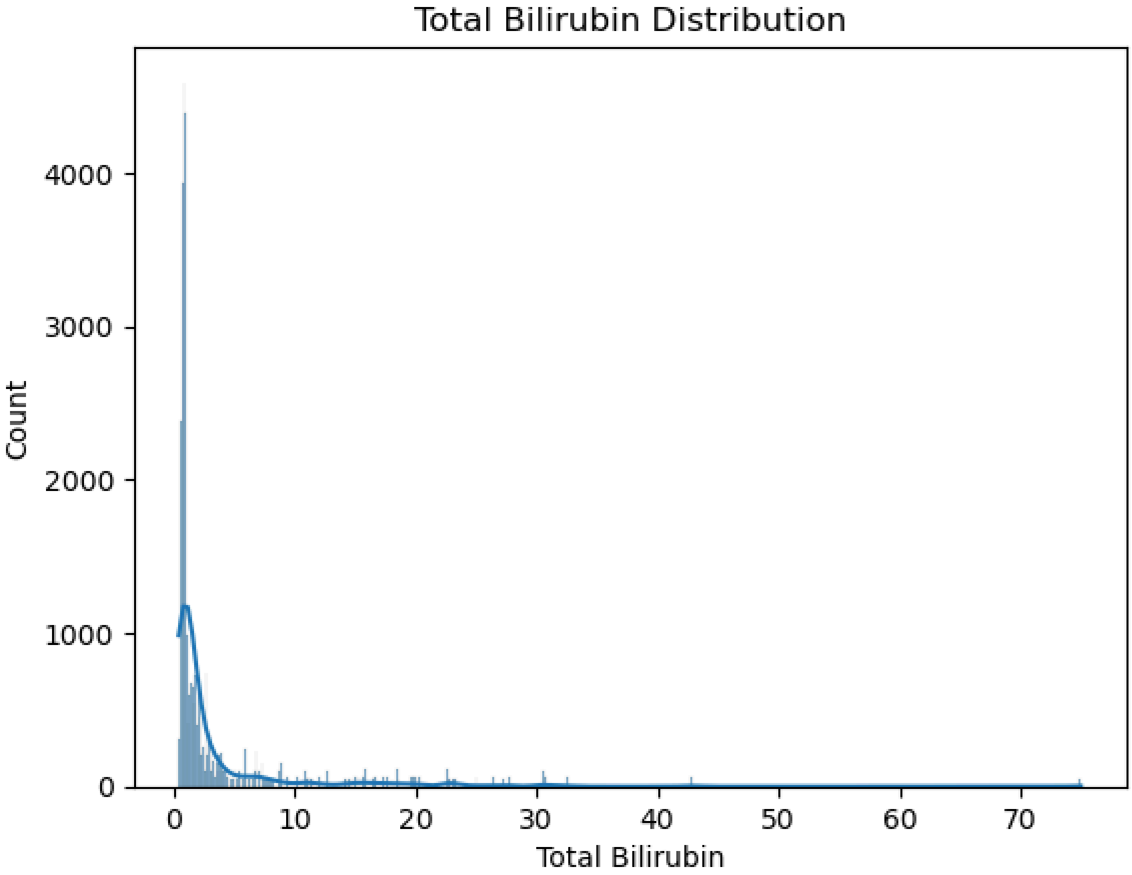
```
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1119: F
utureWarning: use_inf_as_na option is deprecated and will be removed in a futu
re version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```



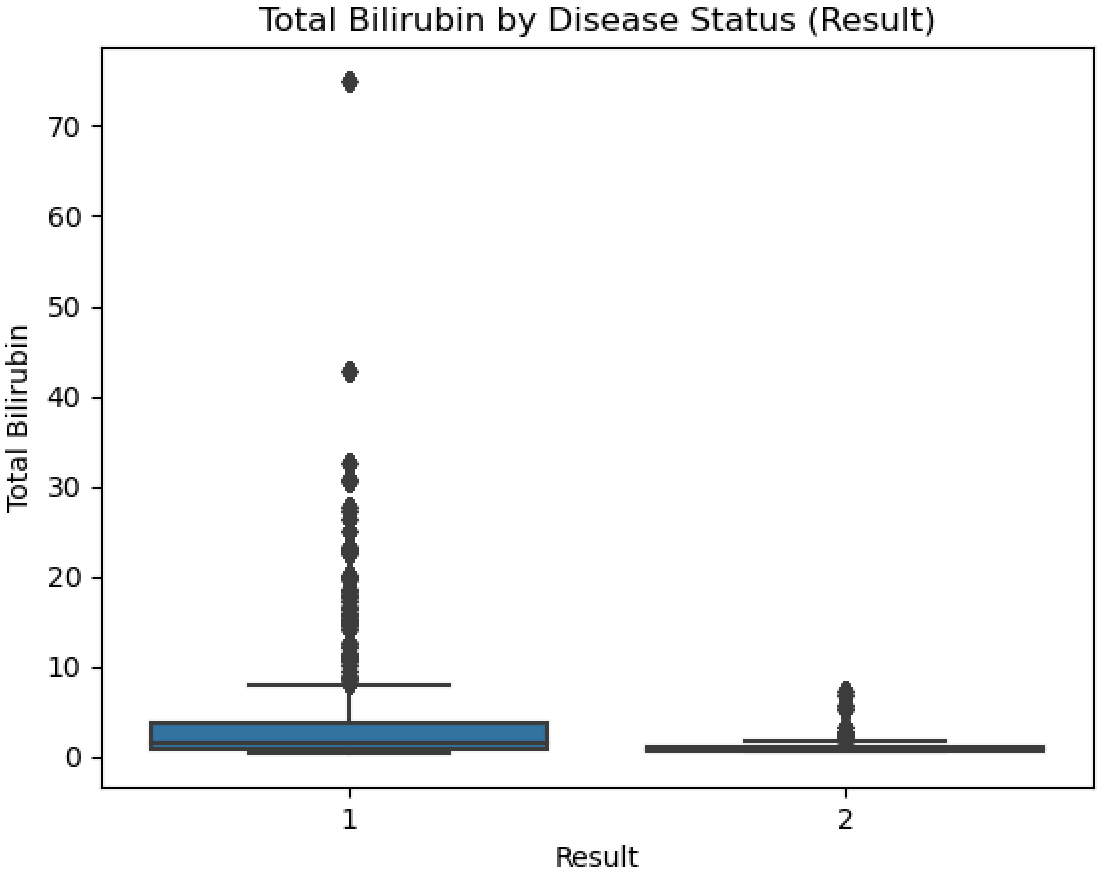
```
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
```



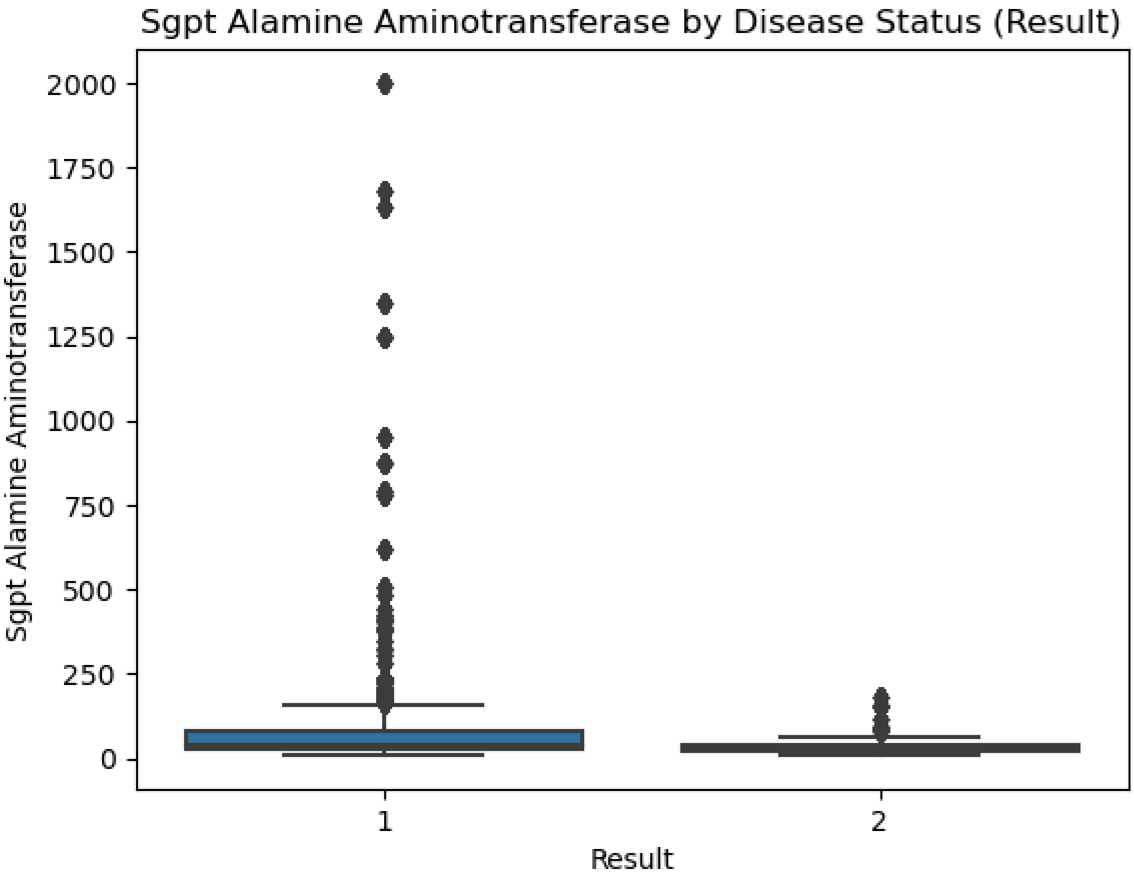
```
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1119: F
utureWarning: use_inf_as_na option is deprecated and will be removed in a futu
re version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```



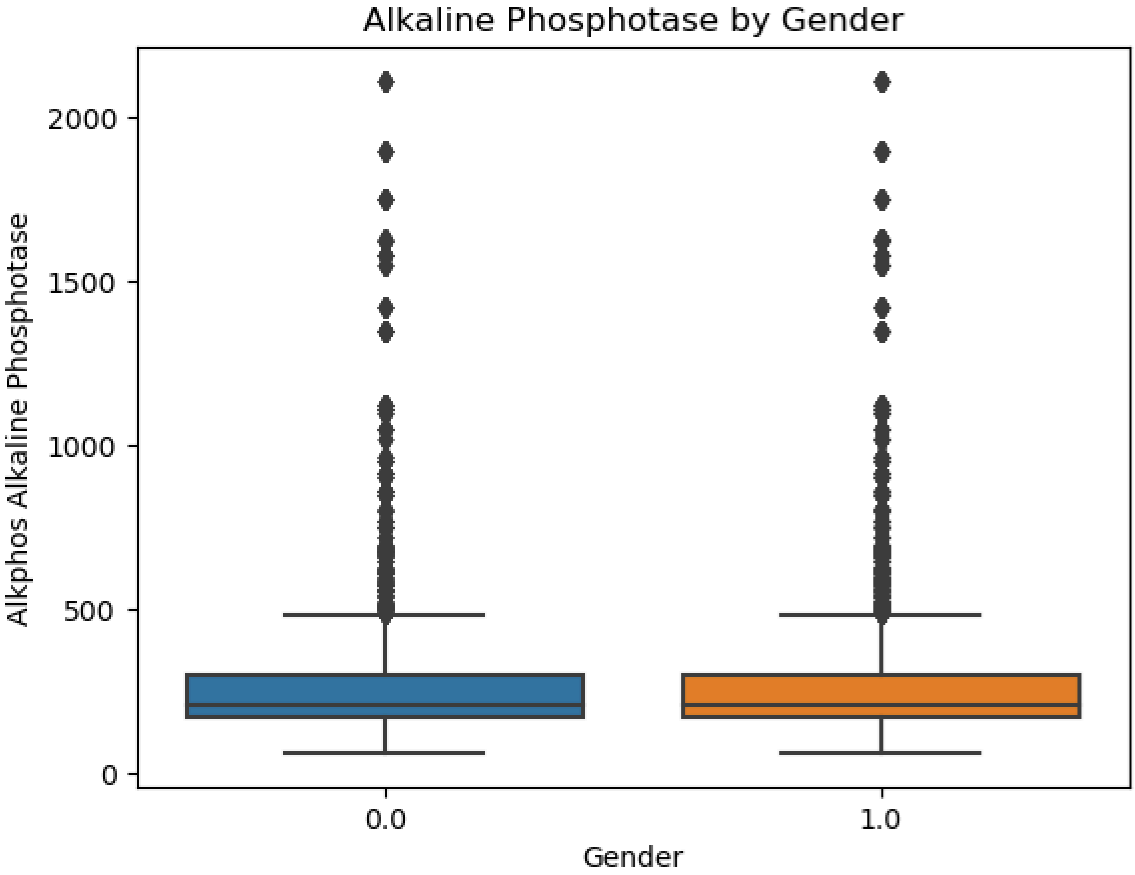
```
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
```



```
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
```



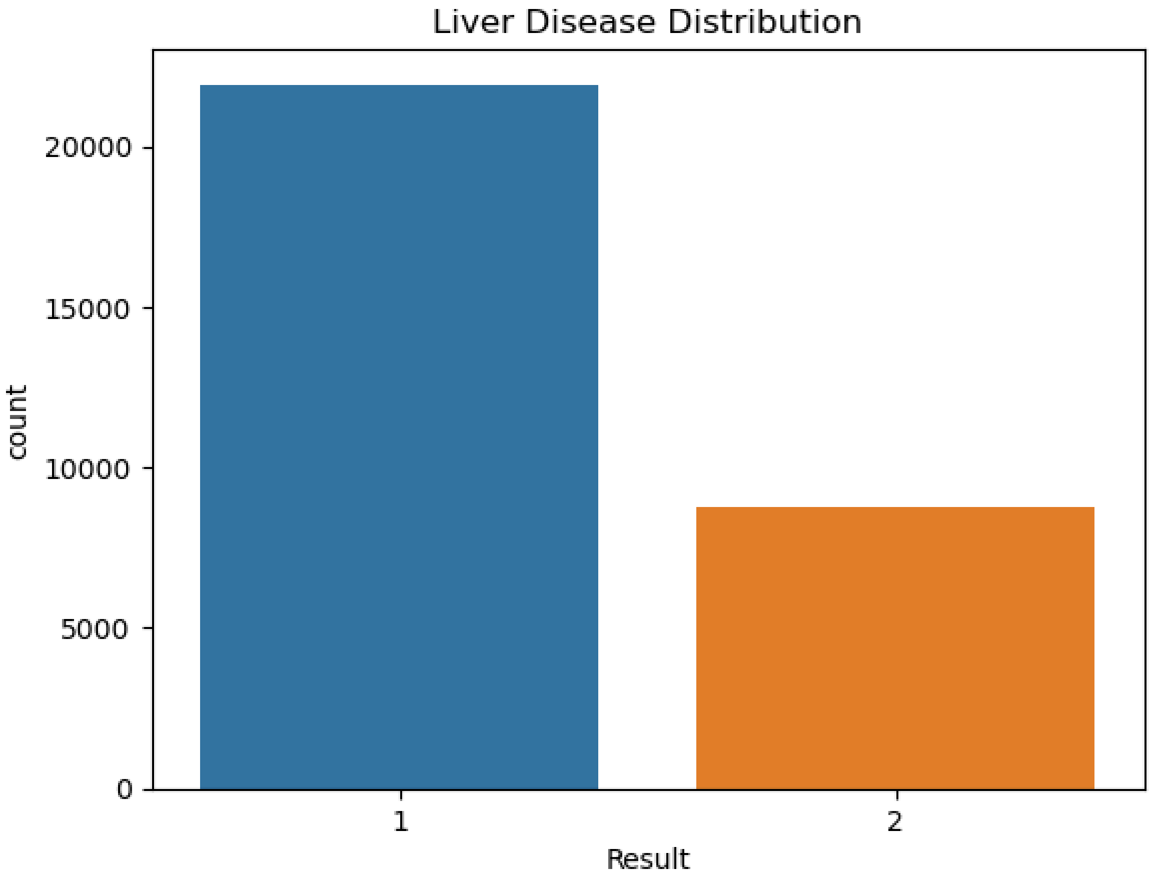
```
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
```



In [11]:

```
# y
sns.countplot(x='Result', data=df)
plt.title('Liver Disease Distribution')
plt.show()
```

```
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
```



In [12]:

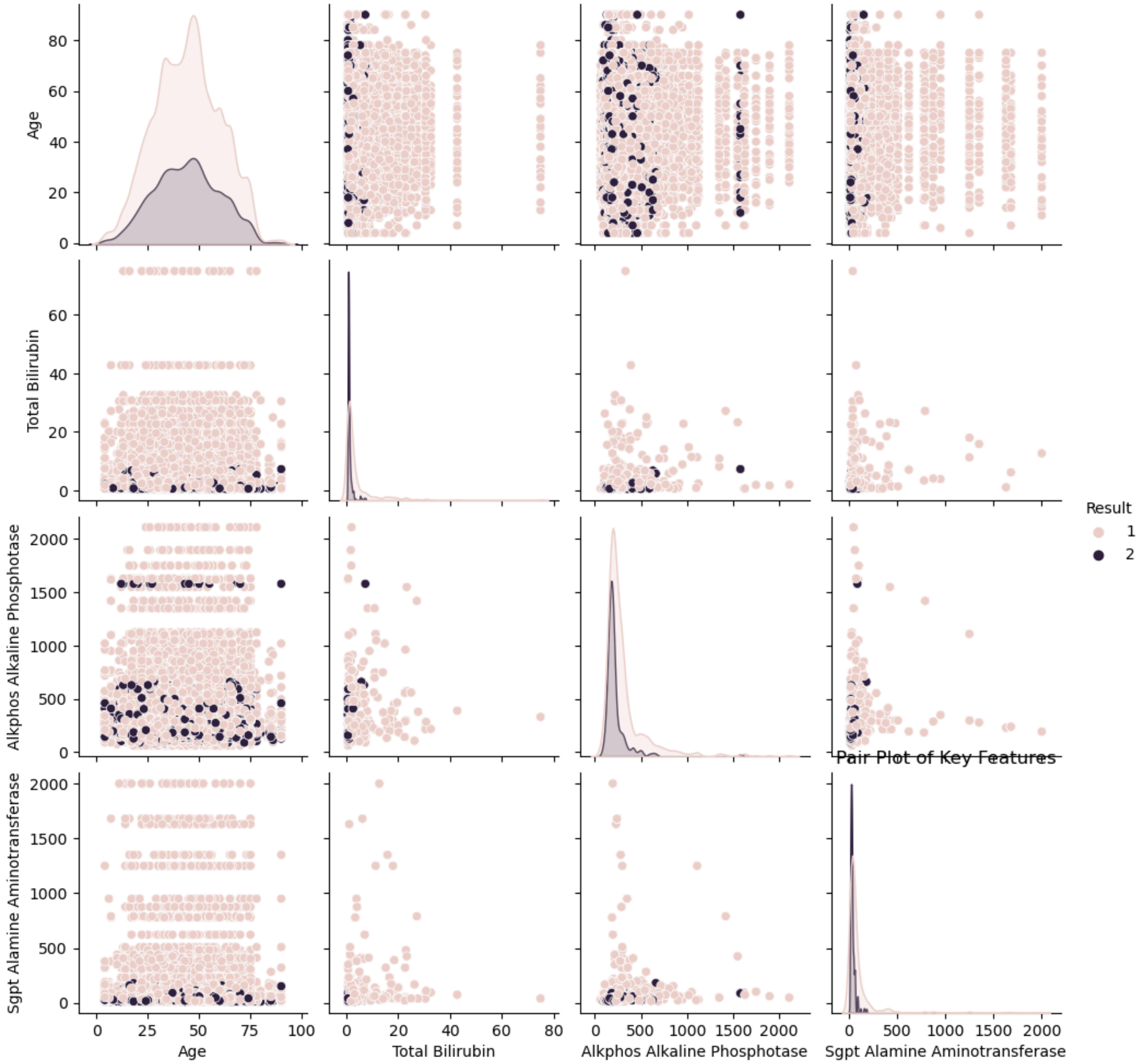
```
# key features
sns.pairplot(df[['Age', 'Total Bilirubin', 'Alkphos Alkaline Phosphotase', 'Sg
plt.title('Pair Plot of Key Features')
plt.show()
```

```
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1119: F
utureWarning: use_inf_as_na option is deprecated and will be removed in a futu
re version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1119: F
utureWarning: use_inf_as_na option is deprecated and will be removed in a futu
re version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1119: F
utureWarning: use_inf_as_na option is deprecated and will be removed in a futu
re version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

```
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1119: F
utureWarning: use_inf_as_na option is deprecated and will be removed in a futu
re version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
```

```
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
```

```
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/axisgrid.py:118: Us
erWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```



Data Preparation/Feature Engineering

In [13]:

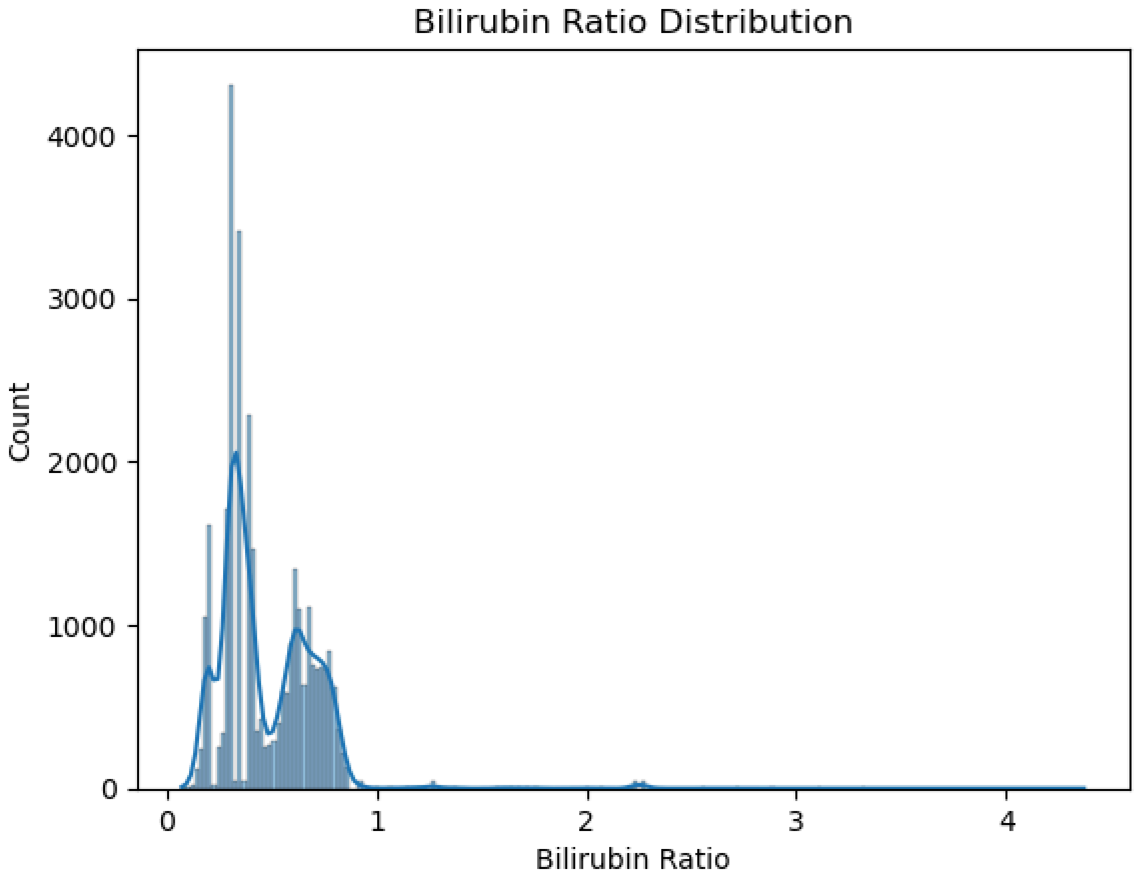
```
import numpy as np

# Handle missing values
df['Gender'] = df['Gender'].fillna(df['Gender'].mode()[0])

num_cols = ['Total Bilirubin', 'Direct Bilirubin', 'Alkphos Alkaline Phosphota
            'Sgpt Alamine Aminotransferase', 'Sgot Aspartate Aminotransferase'
            'Total Protiens', 'ALB Albumin', 'A/G Ratio Albumin and Globulin Ra

for col in num_cols:
    df[col] = df[col].fillna(df[col].median())

# Log transformation (avoid skewness)
log_transform_cols = ['Total Bilirubin', 'Direct Bilirubin', 'Sgpt Alamine Amin
for col in log_transform_cols:
    df[col] = np.log1p(df[col])  # np.log1p(x) = log(1 + x) to avoid log(0) er

# Create new features
df['Bilirubin Ratio'] = df['Direct Bilirubin'] / df['Total Bilirubin']
```

```
df['SGOT/SGPT Ratio'] = df['Sgot Aspartate Aminotransferase'] / df['Sgpt Alami
df['Protien Ratio'] = df['ALB Albumin'] / df['Total Protiens']

# Group by age
bins = [0, 18, 40, 60, 100]
labels = ['Child', 'Young', 'Middle-aged', 'Old']
df['Age Group'] = pd.cut(df['Age'], bins=bins, labels=labels)

# One-hot encode categorical variable
df = pd.get_dummies(df, columns=['Age Group'], drop_first=True)

# Remove highly correlated features to avoid multicollinearity
df = df.drop(columns=['Direct Bilirubin'])  # Highly correlated with Total Bil
```

In [14]:

```
sns.histplot(df['Bilirubin Ratio'], kde=True)
plt.title('Bilirubin Ratio Distribution')
plt.show()
```

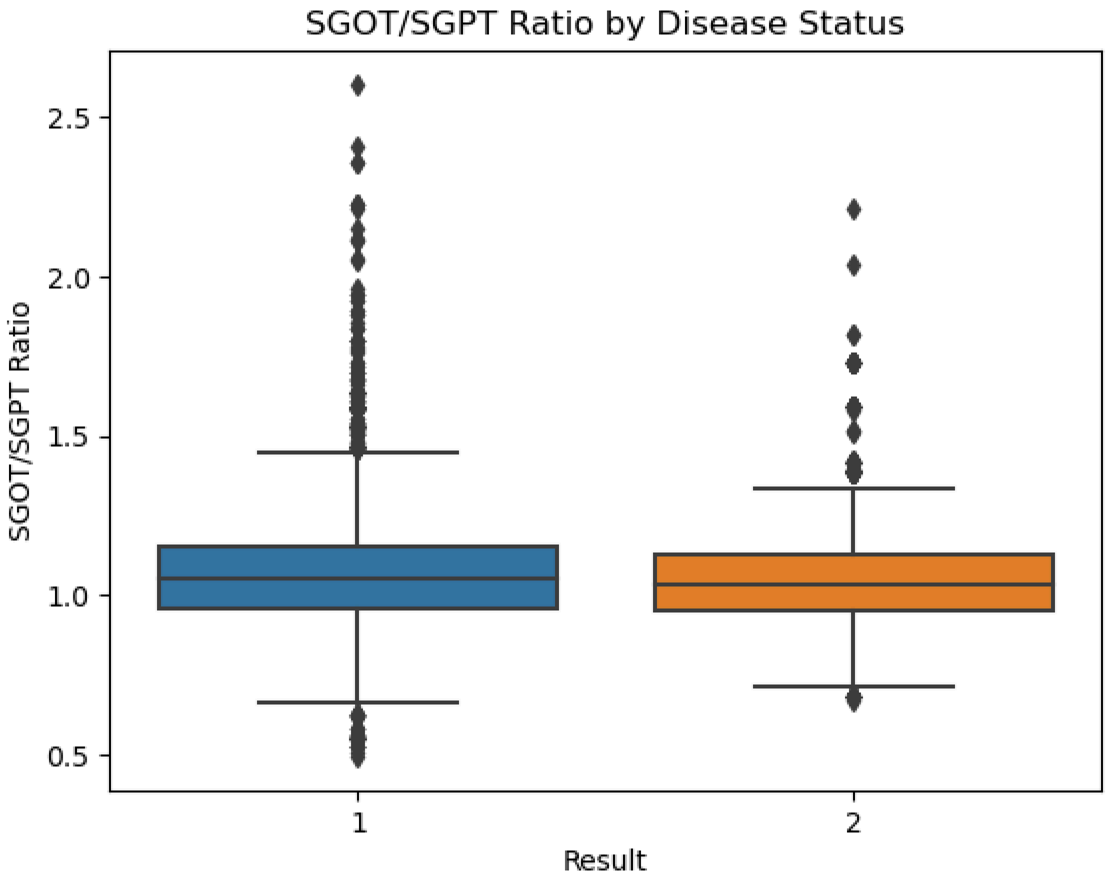
```
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1119: F
utureWarning: use_inf_as_na option is deprecated and will be removed in a futu
re version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```



In [15]:

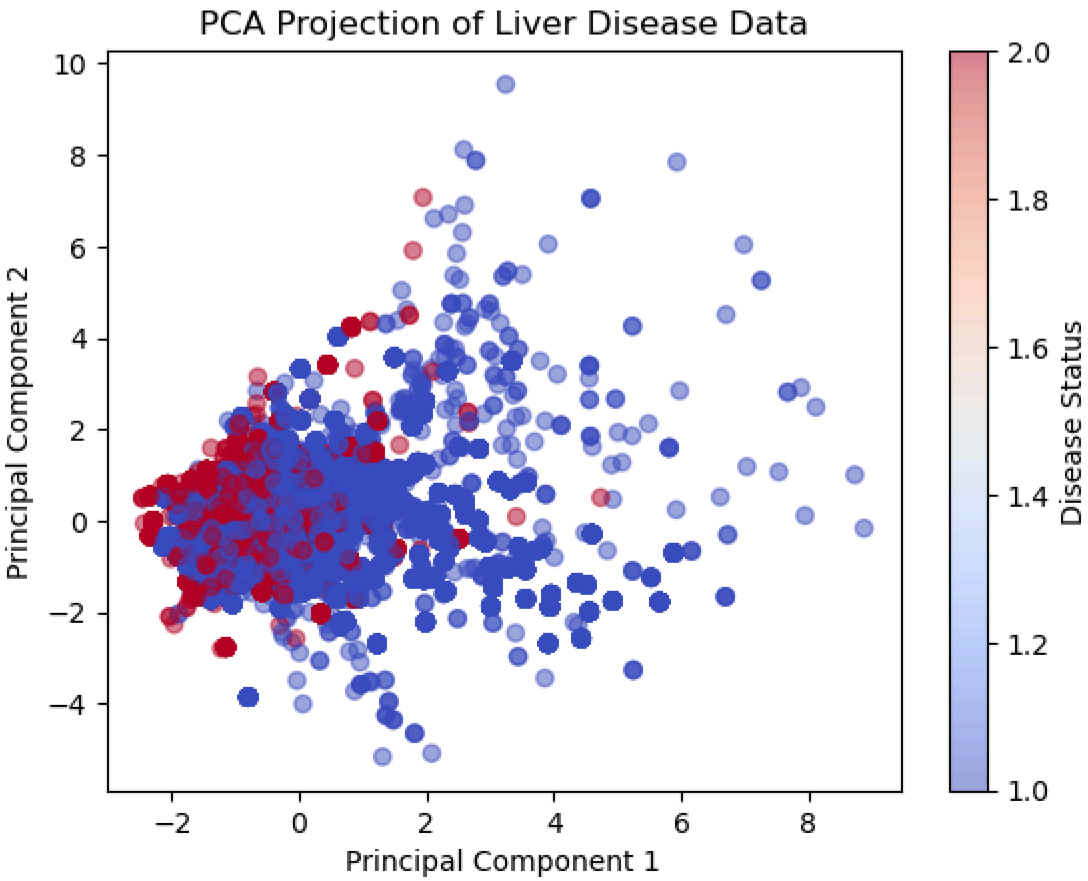
```
sns.boxplot(x='Result', y='SGOT/SGPT Ratio', data=df)
plt.title('SGOT/SGPT Ratio by Disease Status')
plt.show()
```

```
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
/Users/tina/anaconda3/lib/python3.11/site-packages/seaborn/_oldcore.py:1498: F
utureWarning: is_categorical_dtype is deprecated and will be removed in a futu
re version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
```



In [16]:

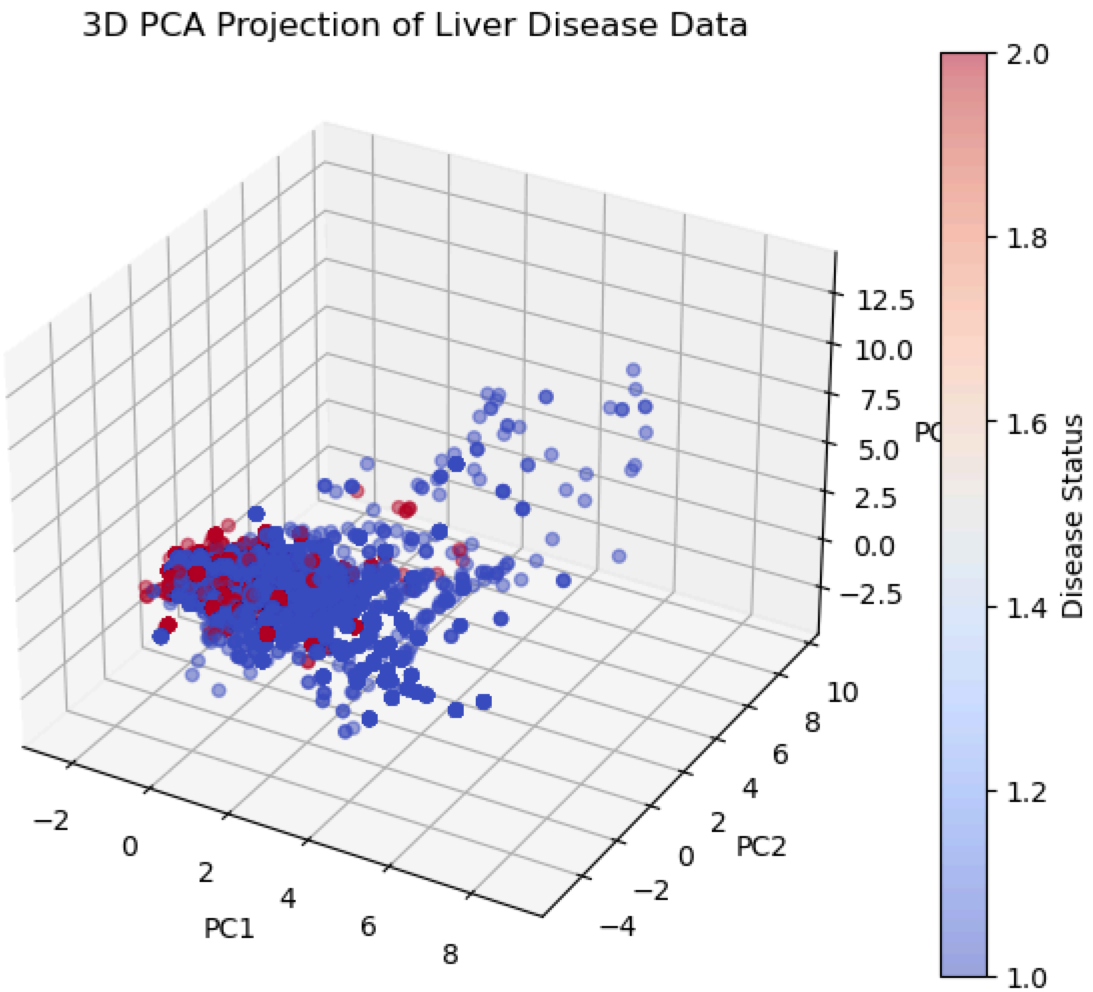
```
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

# choose features for PCA
features = ['SGOT/SGPT Ratio', 'Bilirubin Ratio', 'Sgpt Alamine Aminotransfera
X = df[features]

# standard
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# PCA
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_scaled)

# visualization for PCA
plt.scatter(X_pca[:, 0], X_pca[:, 1], c=df['Result'], cmap='coolwarm', alpha=0
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.title('PCA Projection of Liver Disease Data')
plt.colorbar(label='Disease Status')
plt.show()

# variance ratio of PCA principal components
print(f"Explained variance ratio: {pca.explained_variance_ratio_}")
```



```
Explained variance ratio: [0.52513792 0.29101084]
```

In [17]:

```
from mpl_toolkits.mplot3d import Axes3D

pca = PCA(n_components=3)
X_pca = pca.fit_transform(X_scaled)

fig = plt.figure(figsize=(8,6))
ax = fig.add_subplot(111, projection='3d')
scatter = ax.scatter(X_pca[:,0], X_pca[:,1], X_pca[:,2], c=df['Result'], cmap=
ax.set_xlabel('PC1')
ax.set_ylabel('PC2')
ax.set_zlabel('PC3')
plt.title('3D PCA Projection of Liver Disease Data')
plt.colorbar(scatter, label='Disease Status')
plt.show()
```



In [18]:

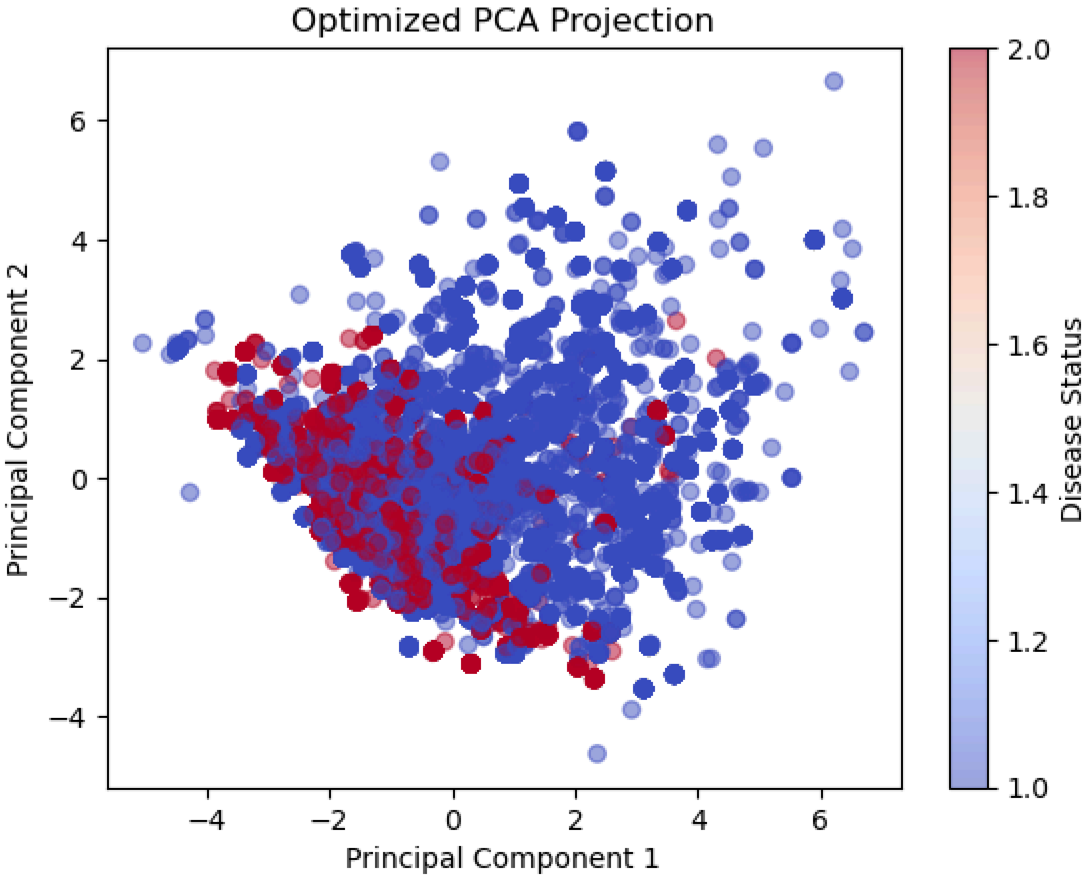
```
from sklearn.preprocessing import StandardScaler, MinMaxScaler

scaler = StandardScaler()
df_scaled = df.copy()
df_scaled[df.columns] = scaler.fit_transform(df)
```

In [19]:

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.feature_selection import RFE

X = df.drop(columns=['Result'])
y = df['Result']
X = X.fillna(X.median())
model = RandomForestClassifier(n_estimators=100, random_state=42)
selector = RFE(model, n_features_to_select=10)  # 10 features

X_selected = selector.fit_transform(X, y)

selected_features = X.columns[selector.support_]
print("Selected Features:", selected_features)
```

```
Selected Features: Index(['Total Bilirubin', 'Alkphos Alkaline Phosphotase',
       'Sgpt Alamine Aminotransferase', 'Sgot Aspartate Aminotransferase',
       'Total Protiens', 'ALB Albumin', 'A/G Ratio Albumin and Globulin Rati
o',
       'Bilirubin Ratio', 'SGOT/SGPT Ratio', 'Protien Ratio'],
      dtype='object')
```

In [20]:
```
from sklearn.decomposition import PCA

pca = PCA(n_components=2)
X_pca = pca.fit_transform(df_scaled[selected_features])

plt.scatter(X_pca[:, 0], X_pca[:, 1], c=y, cmap='coolwarm', alpha=0.5)
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.title('Optimized PCA Projection')
plt.colorbar(label='Disease Status')
plt.show()
```


Optimized PCA Projection
Principal Component 2
Principal Component 1
Disease Status