Revolution Liver care: Predicting Liver Cirrhosis using Advanced Machine Learning techniques

Introduction:

Liver cirrhosis is a progressive and irreversible liver disease often resulting from chronic liver conditions such as hepatitis, alcohol abuse, and fatty liver disease. Early detection is crucial for effective treatment and patient survival. Traditional diagnostic methods are often invasive, expensive, or too late in detecting the onset of cirrhosis.

In this project, we harness the power of **machine learning** to build a predictive model capable of identifying potential liver cirrhosis cases using clinical and biochemical data. This approach aims to aid medical professionals in early diagnosis, minimize the need for invasive tests, and ultimately improve patient outcomes.

Description:

Objective: To develop a machine learning model that can accurately predict liver cirrhosis from patient medical data.

Dataset Used:
The Indian Liver Patient Dataset (ILPD) from the UCI Machine Learning Repository.
Attributes include:

- Age
- Gender
- Total Bilirubin
- Direct Bilirubin
- Alkaline Phosphotase
- Alamine Aminotransferase (ALT)
- Aspartate Aminotransferase (AST)
- Total Proteins
- Albumin

- Albumin and Globulin Ratio
- Class (1 for patients with liver disease, 2 for no disease)

 Steps Involved:

1. Data Preprocessing & Cleaning
2. Feature Engineering
3. Splitting dataset
4. Model training using algorithms like Random Forest, XGBoost, or SVM
5. Model Evaluation (Accuracy, Precision, Recall, F1-Score, AUC)
6. Visualization of results
7. Deployment (optional)

Code:

```
# Upload the CSV file to Colab from local machine
from google.colab import files

uploaded = files.upload()
# Data manipulation
import pandas as pd
import numpy as np

# Data visualization
import matplotlib.pyplot as plt
import seaborn as sns

# For better table display
from IPython.display import display

# Set Seaborn style
sns.set(style="whitegrid")
```

```
# Make sure plots display inline
%matplotlib inline

# Read the uploaded dataset
df = pd.read_excel("HealthCareData.xlsx")

# Display first 5 rows
display(df.head())

# Print info about the dataset
print("\nDataset Info:\n")
df.info()

# Print shape
print("\nDataset Shape:", df.shape)
```

	S.NO	Age	Gender	Place(location where the patient lives)	Duration of alcohol consumption(years)	Quantity of alcohol consumption (quarters/day)	Type of alcohol consumed	Hepatitis B infection	Hepatitis C infection	Diabetes Result	...	Indirect (mg/dl)	Total Protein (g/dl)	Albumin (g/dl)	G
0	1	55	male	rural	12	2	branded liquor	negative	negative	YES	...	3.0	6.0	3.0	
1	2	55	male	rural	12	2	branded liquor	negative	negative	YES	...	3.0	6.0	3.0	
2	3	55	male	rural	12	2	branded liquor	negative	negative	YES	...	3.0	6.0	3.0	
3	4	55	male	rural	12	2	branded liquor	negative	negative	NO	...	3.0	6.0	3.0	
4	5	55	female	rural	12	2	branded liquor	negative	negative	YES	...	3.0	6.0	3.0	

5 rows × 42 columns

```
Dataset Info:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 950 entries, 0 to 949
Data columns (total 42 columns):
 #   Column                                  Non-Null Count  Dtype
```

```
# Check the shape of the dataset
print("Dataset Shape:", df.shape)
```

```
# Check data types and non-null counts
print("\nDataset Info:\n")
df.info()

# Check for missing values
print("\nMissing Values per Column:\n")
print(df.isnull().sum())
# List numerical columns (excluding the target if needed)
numerical_cols = df.select_dtypes(include=['int64', 'float64']).columns.tolist()
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(15, 10))
for i, col in enumerate(numerical_cols):
    plt.subplot((len(numerical_cols)+2)//3, 3, i+1)
    sns.boxplot(x=df[col], color='lightblue')
    plt.title(col)
    plt.tight_layout()
plt.show()
from google.colab import drive
drive.mount('/content/drive')
```

```
print("\nMissing Values per Column:\n")
print(df.isnull().sum())
```

```
Dataset Shape: (950, 42)

Dataset Info:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 950 entries, 0 to 949
Data columns (total 42 columns):
 #   Column                                                Non-Null Count  Dtype
---  ------                                                --------------  -----
 0   S.NO                                                  950 non-null    int64
 1   Age                                                   950 non-null    int64
 2   Gender                                                950 non-null    object
 3   Place(location where the patient lives)               816 non-null    object
 4   Duration of alcohol consumption(years)                950 non-null    int64
 5   Quantity of alcohol consumption (quarters/day)        950 non-null    int64
 6   Type of alcohol consumed                              950 non-null    object
 7   Hepatitis B infection                                 950 non-null    object
 8   Hepatitis C infection                                 950 non-null    object
 9   Diabetes Result                                       950 non-null    object
 10  Blood pressure (mmhg)                                 950 non-null    object
 11  Obesity                                               950 non-null    object
 12  Family history of cirrhosis/ hereditary               950 non-null    object
 13  TCH                                                   591 non-null    float64
 14  TG                                                    591 non-null    object
 15  LDL                                                   591 non-null    object
 16  HDL                                                   582 non-null    float64
 17  Hemoglobin  (g/dl)                                    950 non-null    float64
 18  PCV  (%)                                              920 non-null    float64
 19  RBC  (million cells/microliter)                       398 non-null    float64
```

Conclusion:

This project demonstrates the potential of machine learning in revolutionizing liver care. By leveraging clinical data, healthcare systems can detect liver cirrhosis early, improve decision-making, and enhance patient outcomes. This initiative shows how data science can save lives—transforming raw data into actionable medical intelligence.