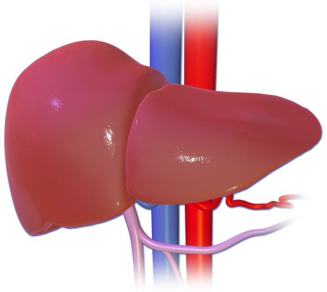
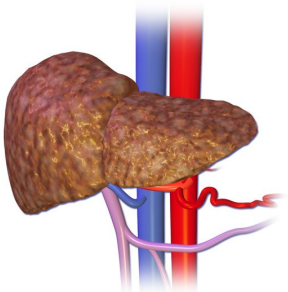
Machine Learning Analysis for Improve Cirrhosis Prediction



Normal Liver



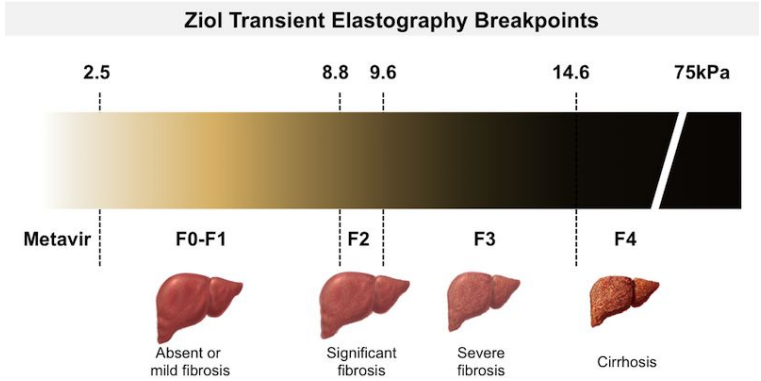
Liver Cirrhosis

By: Sophia Giustino, Emma Kim, Gabriel Miranda Contreras, Sydney Pham, Yunxiang Zhang

Introduction

Cirrhosis

- **12th** leading cause of Death in the US [1]
- **Liver disease** related to damage of the tissue
- **Disrupts** the normal function of the liver
- Causes **problems** in a systematic level



Stages of Cirrhosis:

- **Stage 1:** Steatosis
- **Stage 2:** Liver Scarring (fibrosis)
- **Stage 3:** Cirrhosis
- **Stage 4:** Liver Failure

Compensated | **Decompensated**

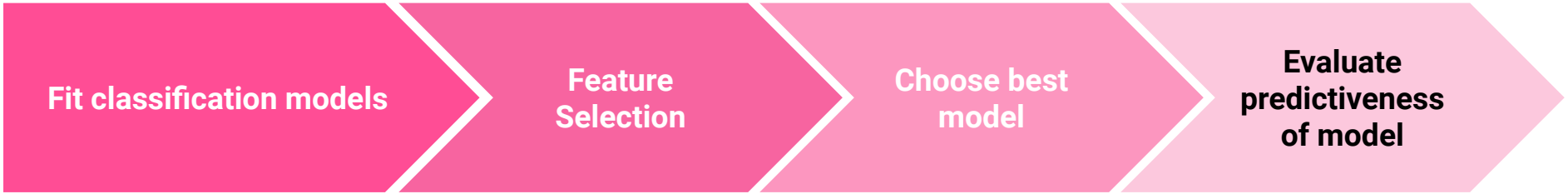
Reversible Damage | **Permanent Damage**

[1] Smith, A., Baumgartner, K., & Bositis, C. (2019). *Cirrhosis:Diagnosis and Management. 100*(12).
[2] D'Amico, G., Bernardi, M., & Angeli, P. (2022). Towards a new definition of decompensated cirrhosis. *Journal of Hepatology*, *76*(1), 202–207. https://doi.org/10.1016/j.jhep.2021.06.018
[3] Karthik Kumar, M. (2022, April 7). 4 stages of cirrhosis of the liver: 18 symptoms, causes & treatment. MedicineNet. https://www.medicinenet.com/what_are_the_4_stages_of_cirrhosis_of_the_liver/article.htm

Introduction

Problem – Current symptoms of cirrhosis occur in mild or mild-advance stages of the disease

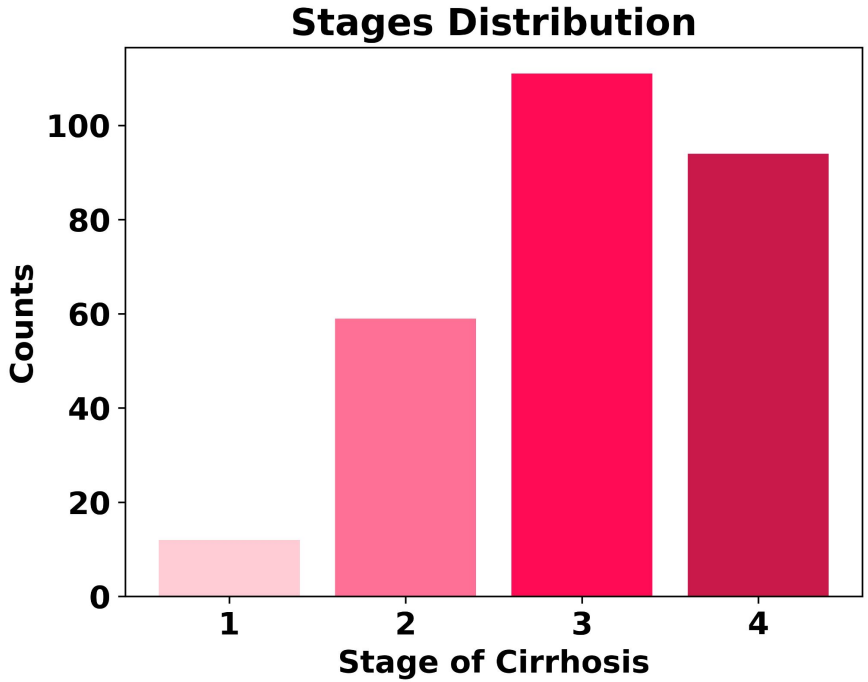
Proposal – Build models to predict the stage of Cirrhosis from physiological data



Data Set

Cirrhosis Patient Survival Prediction Dataset

- Data including **17 clinical features** from **patients** of the Mayo Clinic
- Content
  - 4 conditions of liver physiognomy
  - 9 samples of different compounds in blood
  - Stage of liver damage in 4 stages
  - Patient characteristics
  - Drug treatment
  - Classifier for the study



Data set obtained from Kaggle

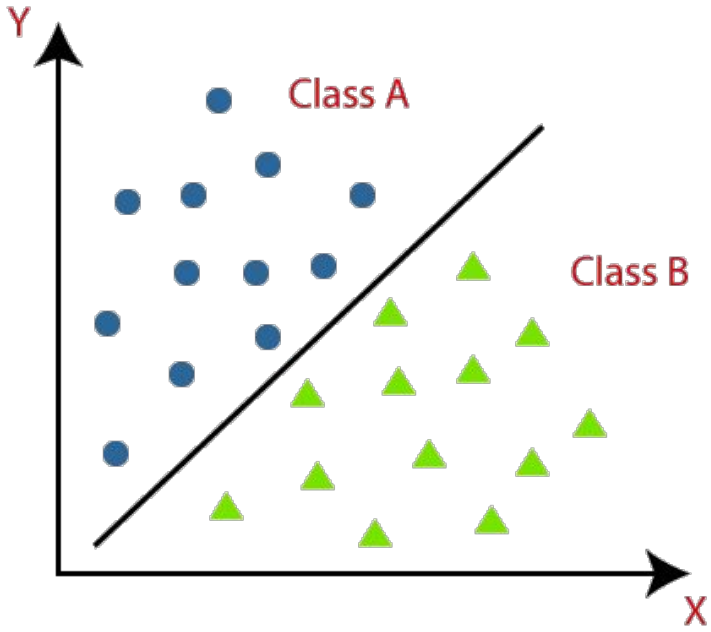
Methods

Classification Methods:

- Decision Tree Classification
- Logistic Regression Classification (L2 regularization)
- K-Nearest Neighbors Classification
- Random Forest
- Gradient Boosting
- Support Vector Machine Classification

Hyperparameter tuning

- Randomized Search for Hyperparameter tuning with 5-Fold cross validation



Software used: DecisionTreeClassifier; LogisticRegression; KNeighborsClassifier;RandomizedSearchCV;SVC;GradientBoostingClassifier;RandomForestClassifier

Methods

Feature Selection: SelectKBest

- Score function: chi2

Software used: SelectKBest

Results

Classification Models: 4-stage Classification Performance over Test set

Method	Accuracy	ROC AUC	Precision Score	Recall Score
KNN	0.375	0.6126	0.3939	0.375
Decision Tree	0.4464	0.6786	**0.4419**	0.4464
Random Forest	**0.5357**	**0.7199**	0.4403	**0.5357**
Gradient Forest	0.4642	0.6677	0.3904	0.4642
Logistic Regression	0.4643	0.6508	0.3908	0.4643
Support Vector Machine	0.4643	0.6351	0.3083	0.4643

Software used: DecisionTreeClassifier; LogisticRegression; KNeighborsClassifier;RandomForestClassifier;SupportVectorMachine;GradientBoostingClassifier

Results

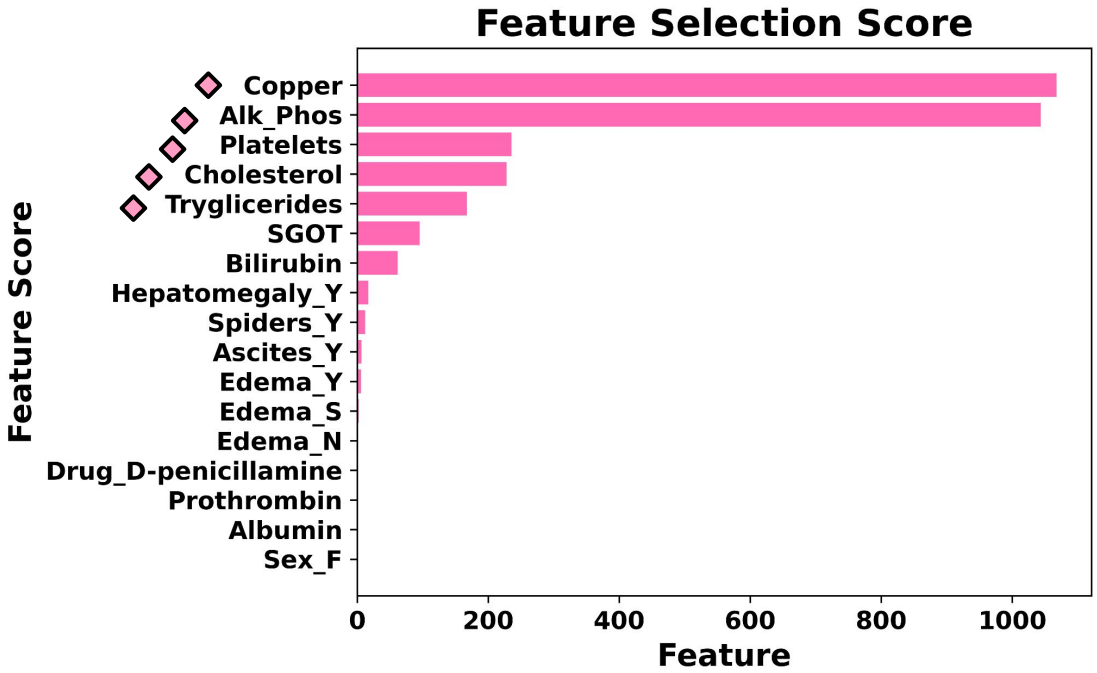
Classification Models: Reversible/Irreversible Stage Classification Performance over Test set

Method	Accuracy	ROC AUC	Precision Score	Recall Score
KNN	0.7143	0.5757	0.7142	0.7142
Decision Tree	0.6964	**0.5946**	0.7229	0.6964
Random Forest	0.7321	0.5871	**0.7239**	0.7321
Gradient Forest	0.6964	0.5037	0.6661	0.6964
Logistic Regression	0.7143	0.5151	0.6755	0.7142
Support Vector Machine	**0.75**	0.4773	0.6111	**0.75**

Software used: DecisionTreeClassifier; LogisticRegression; KNeighborsClassifier;RandomForestClassifier;SupportVectorMachine;GradientBoostingClassifier

Results

Feature Selection: Score Function Chi2



Software used: SelectKBest, chi2

Results

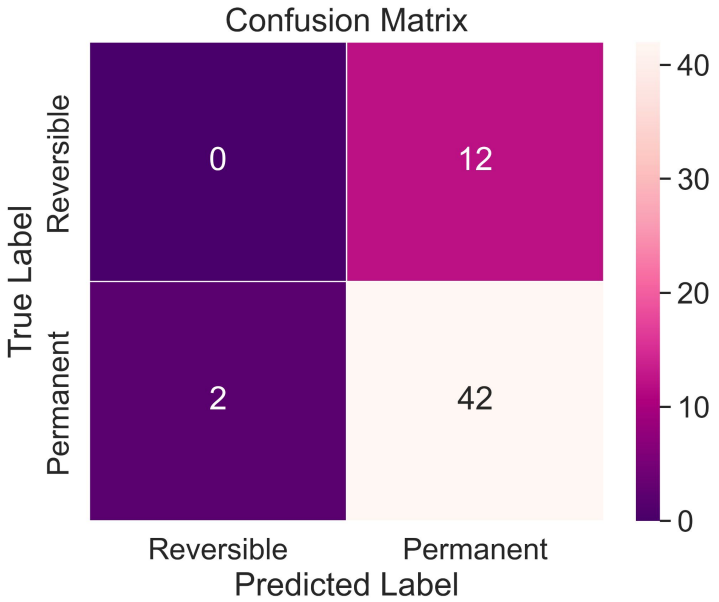
Feature Selection of Best Model

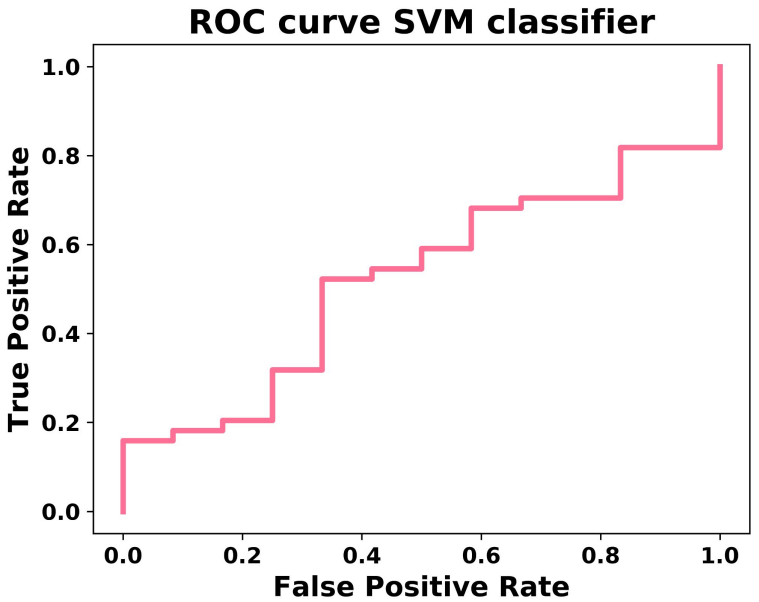
Method	Accuracy	ROC AUC	Precision Score	Recall Score
Random Forest	0.7321	0.4962	0.659	0.7321
Support Vector Machine ◆	**0.75**	0.4773	0.6111	**0.75**

Software used: RandomizedSearchCV; RandomForestClassifier; SVC

Results

Support Vector Machine Metrics





Software used: roc_curve, confusion_matrix

Conclusion

- **Classification:** Reversible Cirrhosis or Irreversible Cirrhosis
- **Model:** Support Vector Machine
- **Predictive Features:**
  - Copper
  - Alk_Phos
  - Platelets
  - Cholesterol
  - Triglycerides
- **Accuracy:** 0.75
- Errors are mostly false positives (irreversible when truly reversible)