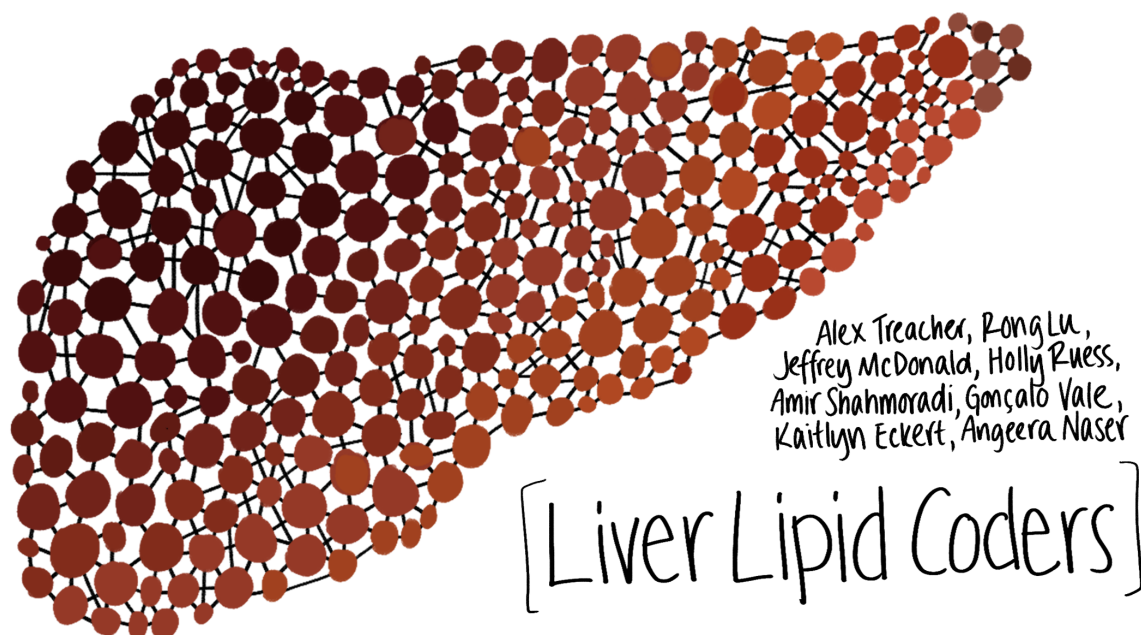


LiverLipidCoders

The focus of this project is to develop a machine learning algorithm using the plasma lipid data provided to provide staging for a patient's liver disease diagnosis.



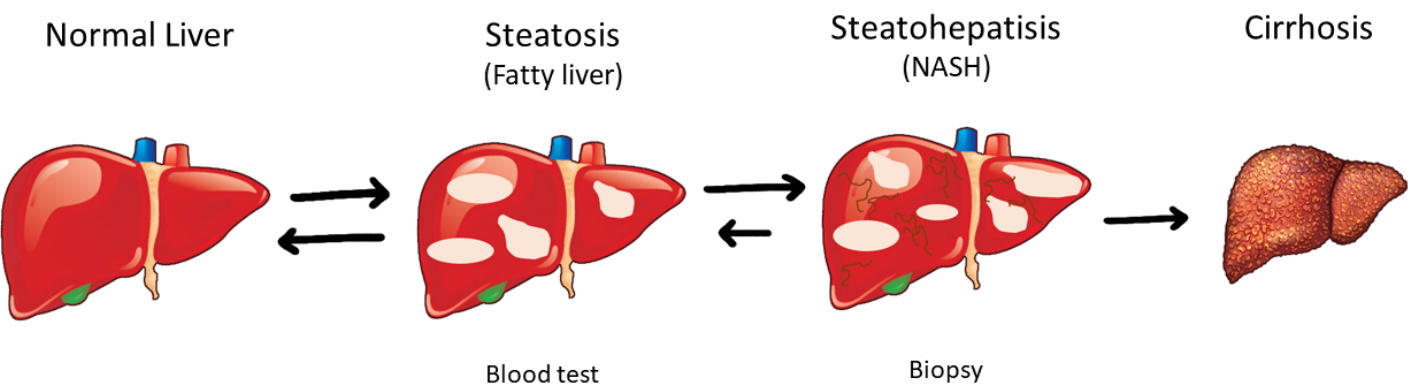
Logo

Introduction

There are four main stages of liver disease: 1. Normal: Healthy condition. 2. Steatosis: Liver is fatty. Reversible. Can be diagnosed via a blood test. 3. Steatohepatitis: Liver is fatty and inflamed. Reversible. Can only be diagnosed via a biopsy. 4. Cirrhosis: Liver is dying. Non-reversible.

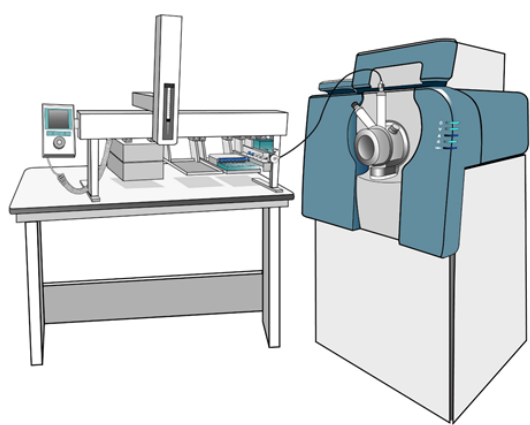
Through machine-learning-based staging of disease and identification of relevant biomarkers in plasma, patients can avoid the process of an intrusive biopsy (for stage 3) and still receive an accurate diagnosis of the disease. Our team conducted statistical analyses on a small sample of patient data to determine whether further investigation of similar techniques is beneficial.

Non-alcoholic fatty liver disease



Are there any lipid biomarker(s) in blood that can be used for NASH diagnosis?

Stages of Liver Disease



Mass spectrometry based workflow

- Global lipids screening analysis.
- Detailed information about lipid species

Cohort: 87 patients

G1 (Normal)	30
G2 (Fatty Liver)	17
G3 (NASH)	20
G4 (Cirrhosis)	20

Named Lipids	400
Unknown Molecules	1400

The Cohort

Lipid Markers for Stage 3

We focused on feature selection and dimension reduction so we could identify lipids especially relevant to staging. We created a list of known and unknown lipid signals that contribute the most to the automated classification of steatohepatitis (stage 3):

Lipid Markers of Stage 3 Classifier

Top known lipid signal:

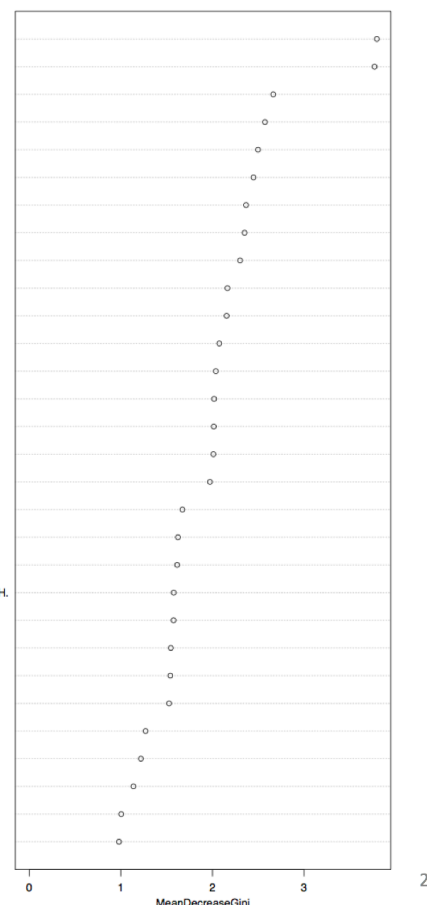
- TAG(48:0) – NL FA(18:0)
- TAG(48:0) – NL FA(16:0)
- DAG(18:1_18:1)
- TAG(52:4) – NL FA(20:4)
- TAG(54:4) – NL FA(22:4)
- DAG(18:2_18:1)
- PC(33:1) or PC(P-34:0) or PC(0-34:1)
- Cer(15:0/18:1{d8})
- TAG(50:4) – NL FA(16:1)

Top unknown lipid signal (mass pairs):

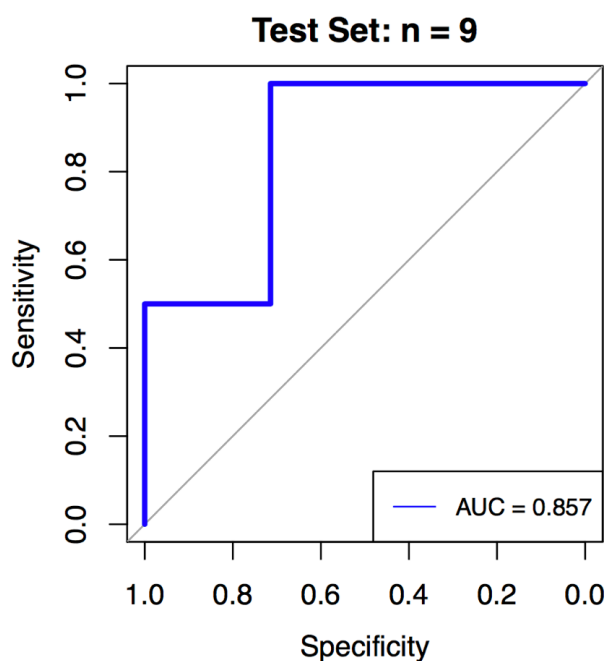
- U1 (729.58/570.49)
- U2 (877.73/604.48)
- U3 (877.73/578.46)
- U4 (667.52/164.13)
- U5 (844.70/261.19)

11/10/18

TAG.48.0..NL.18.0....NH4.
TAG.48.0..NL.16.0....NH4.
u729.58_570.4902
u877.73_604.4830
u877.73_578.4698
u667.52_164.1375
u844.70_261.1963
u798.65_577.4597
u668.52_261.2273
DAG.36.2..NL.18.1.18.1..NH4.
u638.49_603.4743
u877.73_240.2178
u888.74_577.4612
u877.73_239.2149
u891.74_570.4918
u1028.88_170.1385
u222.08_192.1198
TAG.52.4..NL.20.4....NH4.
TAG.54.4..NL.22.4....NH4.
DAG.36.3..NL.18.2.18.1..NH4.
PC.33.1....H..or.PC.34.0..P....H..or.PC.34.1..O....H.
Cer.41.1..d18.1.23.0..H.
TAG.50.4..NL.16.1....NH4.
u872.72_189.1449
u925.78_575.4482
u836.69_563.4485
u824.68_553.4568
u876.73_876.7106
u824.68_239.2157
u895.75_598.4339



Stage 3 Random Forest Classifier Performance in Testing Set:



11/10/18

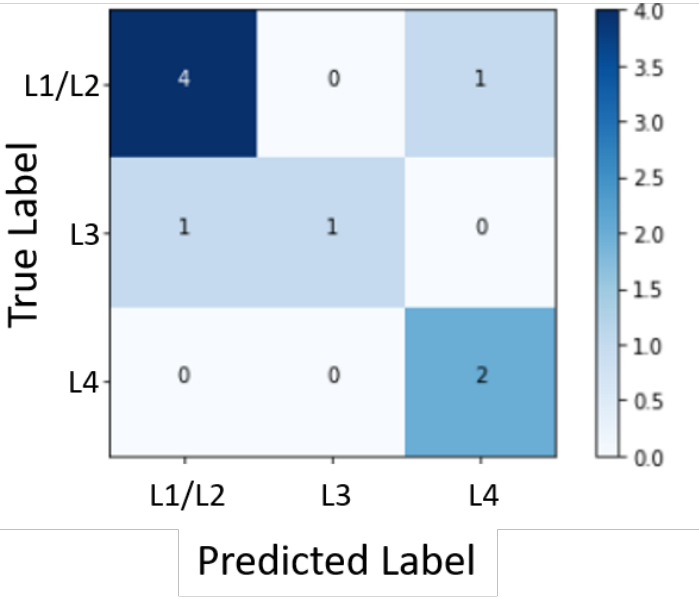
Common Marker of different classifiers:

- TAG(48:0) – NL FA(18:0)
- TAG(56:8) – NL FA (22:6)
- Unknown (680.53/551.27)
- Unknown (828.68/523.42)
- Unknown (902.75/339.25)

4

Dense Neural Network

Using all the named lipids, we normalized the data and trained 281 different architectures and used the best to classify between F1/F2, F3, and F4.



Metric	Value
Accuracy	.78
F1 macro score	.76