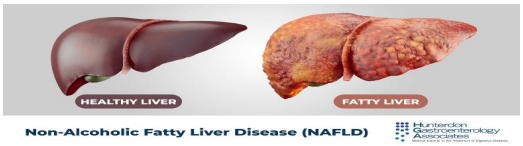
Non-alcoholic Fatty Liver Predictive Analytics

Orit Goldman, Ofir Ben-Assuli, Arie Jacobi, Ono Academic College
Shira Zelber-Sagi, University of Haifa,
Shani Shenhar-Tsarfaty, Tel-Aviv Sourasky Medical Center

Background

Non-Alcoholic Fatty Liver Disease (NAFLD) is the most common liver disease worldwide (20%-30% prevalence in the general population)



https://hunterdongastro.com/non-alcoholic-fatty-liver-disease-nafld/

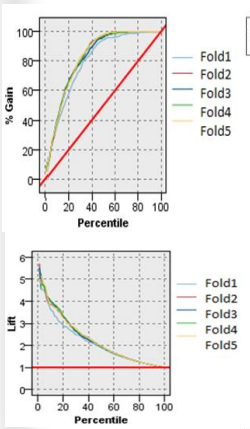
Research Setting

Prospective cohort from the **T**el **A**viv **M**edical **C**enter **I**nflammation **S**urvey (**TAMCIS**)
>more than 600 variables per visit
>more than 14,500 individuals
>up to 10 annual follow-up visits for each person

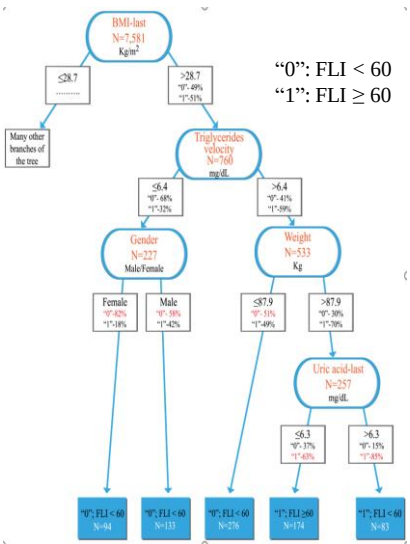
Results (I)

-Our technique outperformed the predictive power of existing methods
-New features were identified as powerful factors in the predictive process

AUC 0.8486
Accuracy 0.8573



FLI - Performance Evaluation



"0": FLI < 60
"1": FLI ≥ 60

Decision tree for FLI-CHAID model

Research Objective

Stratify the risks for NAFLD by using:
>Fatty Liver Index (FLI)
>NAFLD fibrosis 4 calculator (FIB-4)
to predict changes over time and suggest preventive medical decisions

Research Plan

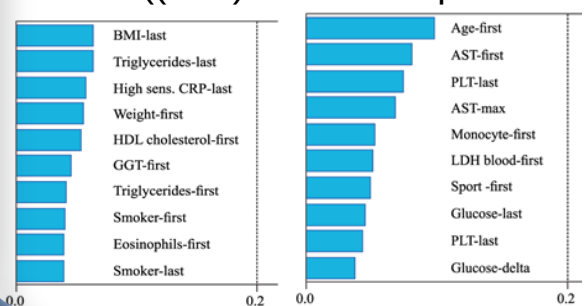
>Extensive data preparation
Computation of time series variables
Build 2 panels in person-level
Define 2 dependent variables
FLI ≥ 60 diagnosed as 'YES'
FIB4 ≥ 1.45 diagnosed as 'YES'
>Machine learning classification models to predict individual risk

Research Methods

>Decision trees to develop classification models for the two dependent variables (FLI and FIB4)
>Fivefold cross validation to avoid over-fitting
>Best results : CHAID decision tree
>evaluation: Lift, Gains, AUC and Accuracy (0.5 threshold)

Results (II)

Unexpected influential features:
Smoking as a FLI predictor
Exercise ((lack) as a FIB-4 predictor



This work was funded by the Israel National Institute For Health Policy Research, Grant# 2018/52/א.