Predicting Metabolic Syndrome

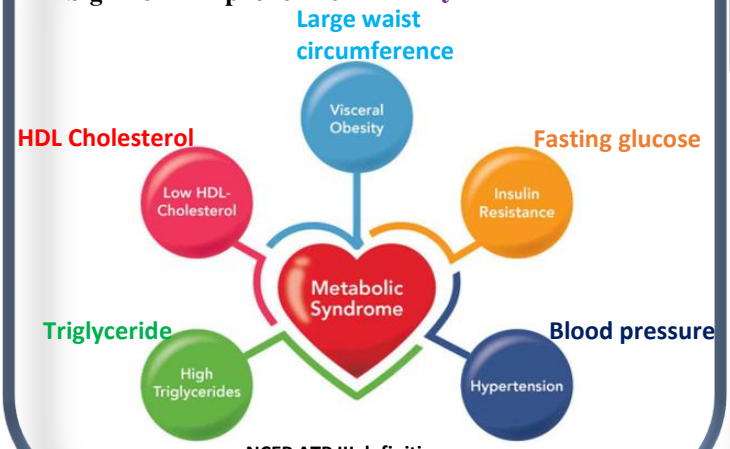
Machine Learning Techniques for Improved Preventive Medicine

Orit Goldman, Ofir Ben-Assuli, Shimon Ababa, Ono Academic College
Ori Rogowski, Shlomo Berliner Tel-Aviv Sourasky Medical Center

Background

Metabolic Syndrome (MetS)

Significant impact on health Any 3 of the 5 criteria



NCEP ATP III definition

National Cholesterol Education Program Adult Treatment Panel III

Research Objective

Predict the risks associated with **MetS** to enable medical personnel to make more optimal **preventive medical decisions**

Research Plan

>Extensive data preparation

Pre-processing/Refinement data

Computation of time series variables

Build panel at person-level

Define a dependent variable

MetS='YES': All visitS ='NO' ,last visit ='YES'

MetS='NO': All visitS ='NO'

>Machine learning classification models to predict individual risk

Research Setting

A prospective cohort from the **T**el **A**viv **M**edical **C**enter **I**nflammation **S**urvey (**TAMCIS**)

September 2002 to July 2023

> More than 600 variables per visit

> More than 14,500 individuals

> Up to 10 annual follow-up visits for each person

Research Methods

>**ML algorithms** in different methods to develop classification models to predict the presence of **MetS**

>Fivefold cross-validation to avoid over-fitting

>Best results: **Gradient Boosting**

>Evaluation: AUC, Recall, Precision, Accuracy (0.5 threshold), Lift and Gains

Results (I)

The results in our best model (Gradient Boosting): AUC = 0.947, Accuracy = 0.947

Outperformed existing methods

Model	AUC	Accuracy	Recall	Precision
Neural Network	0.920	0.934	0.492	0.644
Logistic Regression	0.943	0.944	0.550	0.728
Gradient Boosting	**0.947**	**0.947**	0.482	**0.818**

Lifestyle features, among others, were identified as powerful factors in the predictive process:

Smoking No. of **cigarettes** per day

Exercise (lack) Weekly **sport** exercise average

Results (II)

Gradient Boosting

Performance Evaluation

Lift

Top decile: 8.48% x 6.6 = 56%

Gains

Top decile: 68%