INTELLIGENT HEALTHCARE DATA ANALYTICS FOR HEPATIC STEATOSIS PREDICTION

Background and Objective

- Metabolic dysfunction-associated fatty liver disease (MAFLD) is important because it is a common and significant risk factor for cardiovascular diseases, diabetes, and liver-related morbidity and mortality, impacting global health.
- Our group has been studying MAFLD for years. Panigrahi and Deo, 2022 proposed a set of ML algorithms for the detection of Hepatic Steatosis (HS) i.e., simply fatty liver in the MAFLD paradigm, using the NHANES III survey data.
- The inputs of the developed models were age, sex, BMI, HDL, glucose (GLU), AST, ALT, and ASP. The output was the prediction of HS. The HS was assessed using ultrasound images categorized for medical personnel.
- They found good performance on these algorithms. However, they did not employ the weights adjustment that the Center for Disease Control and Prevention (CDC) suggested to do for analysis.
- Therefore, we want to evaluate the use of sample weights for an HS screening model to be used as a clinical decision support tool using physiological and liver biochemistry parameters.

INTELLIGENT HEALTHCARE DATA ANALYTICS FOR HEPATIC STEATOSIS PREDICTION

Methodology

- We follow the methodology described by Deo, 2022:

  Raw data is coming from the NHANES III survey (31311 samples).

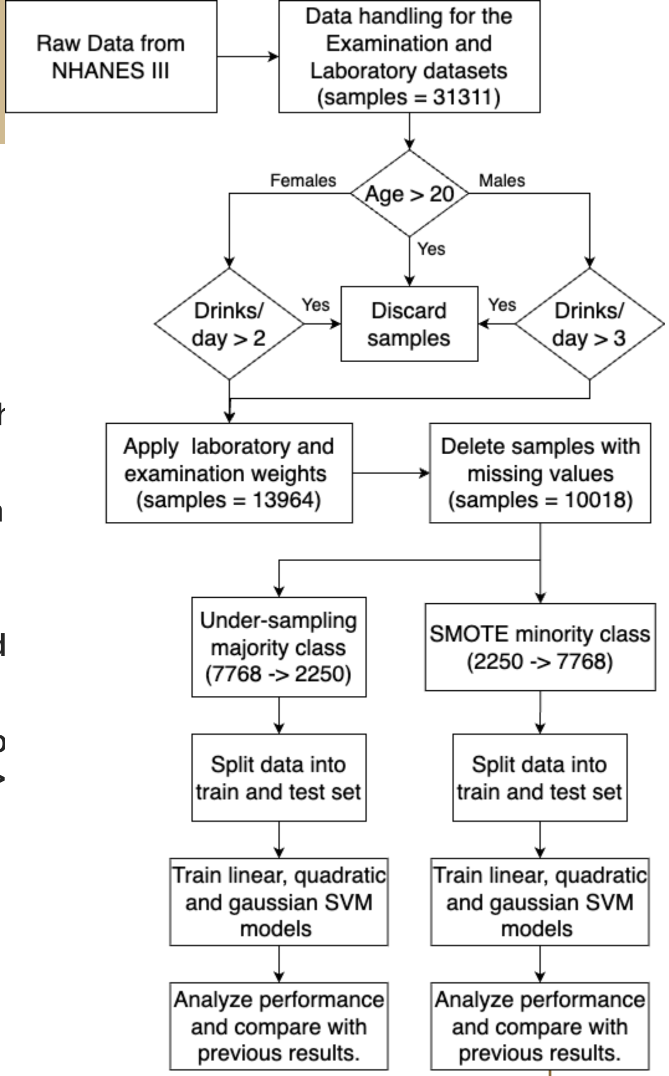
  We kept the predictors Age, Sex, BMI, HDL, GLU, AST, ALT, and ASP. Th objective is HS (Hepatic Steatosis) from ultrasound assessments.

  We discarded samples from people over 20, women, and men with m than 2 and 3 drinks per day (13964 samples left). Also, we deleted samples with missing values (10018 samples left).

  We applied the respective weights for examination (Age and BMI) and laboratory datasets (HDL, GLU, AST, ALT, and ASP).

  Two sampling methods were tested: 1) randomly under-sample the no HS class (7768 -> 2250) and 2) apply SMOTE to the HS class (2250 -> 7768) to equalize the classes in both cases.

  SVM with linear, quadratic, and Gaussian kernels were trained.



INTELLIGENT HEALTHCARE DATA ANALYTICS FOR HEPATIC STEATOSIS PREDICTION

Results and Future work

- The quadratic kernel SVM with weights obtained a comparable perfomance than the models without weights.
- The sample weights introduced an improvement in the sensitivity metric, but they performed worse in specificity.
- In general, the sample weights did not introduce an improvement in the perfomance of the SVM models.
- Future work include to try more algorithms that Deo, 2022 tested looking for an improvement in the performance of HS prediction.

SVM		Deo, 2022			This work		
Kernel	Sampling	Acc	Sen	Spe	Acc	Sen	Spe
Linear	Under.	.697	.670	.724	0.60	0.87	0.33
	Smote	.74	.739	.742	0.60	0.89	0.30
Quad.	Under.	.699	.661	.737	0.69	0.70	0.68
	Smote	.731	.742	.719	0.68	0.76	0.61
Gaussian	Under.	.698	.667	.729	0.62	0.79	0.44
	Smote	.722	.728	.716	0.70	0.75	0.65