

Data pertaining to prognostics of T2D postpartum in women previously diagnosed with gestational diabetes mellitus

Raw data normalized:

- removed variables where more than 5% of values were missing
- remaining missing values were estimated as half of the lowest positive values
- log transformation carried out to get a normal/semi-normal data distribution

Final data being used is composed of ~100 analytes

PCA analysis conducted to determine influence of confounding factors in the final dataset

Random forest used to determine top 10 analytes with greatest influence on T2D prognostics

Various supervised machine learning based classification techniques used to identify prognostics

Applying deep learning classifier to identify the prognostics

Applying clustering algorithms (K-means and Hierarchical clustering) to identify disease heterogeneity

Based on outcome, determine which machine learning analysis can be used to determine prognostics of T2D postpartum in women previously diagnosed with gestational diabetes.