Medical Mining Topic – Summer Term 2019

Multifactorial Diseases & Disorders

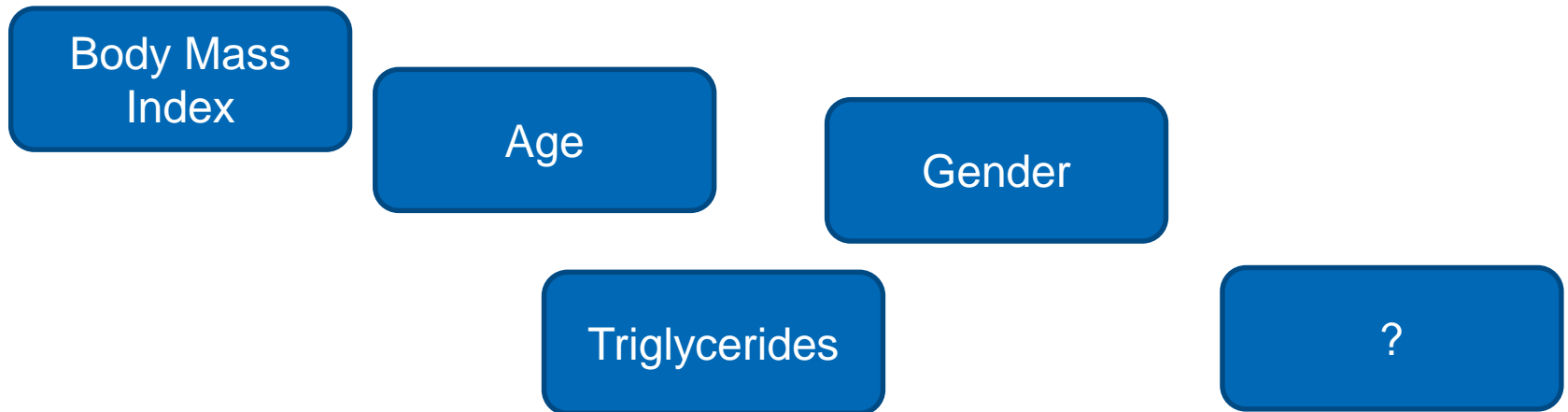**Which risk factors and characteristics are associated with a disease or disorder?**

- Lifestyle factors like daily alcohol intake

- Genetic factors about individual predispositions

- Environmental factors like water pollution

Longitudinal epidemiological studies

- Multiple study waves

- Same-participant followups

- Repeated medical examination programmes

OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

INF

FAKULTÄT FÜR
INFORMATIK

Hepatic Steatosis

- "Fatty liver"

- Example of a multifactorial disorder

- Multiple independent risk factors and important features

- Potential predictors according to [BBM+06]:

Body Mass Index

Age

Gender

Triglycerides

?

OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

INF

FAKULTÄT FÜR
INFORMATIK

How do we find relevant factors? – One example:

- Given a study participant dataset with feature set F and target disease C:

    (1) Enumerate a number of subspaces of F

    (2) Use each subspace to predict the C

    (3) Compare the performance of each space and analyze the best ones

- (1), (2) and (3) is easy when a lot of ground truth is available

- However, how do we proceed when only **limited** ground truth w.r.t. the disease under study is available?

- One way is to enumerate and evaluate subspaces according to their best match between internal structure of the data and the limited knowledge w.r.t. the medical outcome under study

OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

INF

FAKULTÄT FÜR
INFORMATIK

SP : Integrating Subspace Clustering with Constraints-based Feature Selection

How to utilize limited knowledge about the target disease/disorder?

- Limited knowledge stating that two participants have or have not the disease / disorder under study (Must-Link / Not-Link constraints)

- Assessing relevance of subspaces

TODO

- Literature has to be reviewed and compared: Find and review different subspace clustering methods.

- Integrate the subspace clustering method into a given feature selection framework for constraints exploitation.

- Evaluate your method on real-world data.

The End

Thank you very much!