University of Amsterdam
Data Science Centre

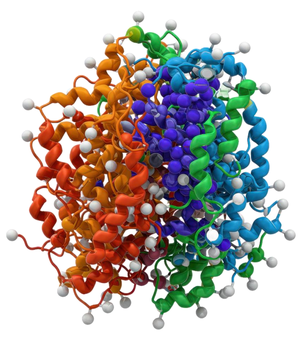
AI and Data Science at UvA

# Protein discovery for diabetes & periodontitis using oral rinse samples

Elena Stamatelou, PhD Candidate @ ACTA, Data Scientist

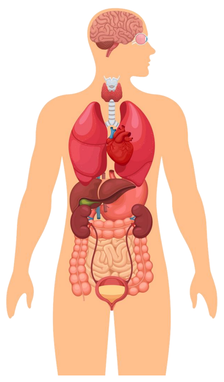
## Introduction

**Oral rinse samples**



**Proteomic Analysis**

Olink



**Periodontitis**

**Diabetes type 2**

## Datasets

DE MAASTRICHT STUDIE

**176** participants

**92** proteins
Target 96
Cardiometabolic panel

**20** diabetes (high HbA1c )*
**156** no diabetes (low HbA1c )*
**glycosylated hemoglobin (HbA1c) cutoff 48mmol/mol*

ACTA

**54** participants

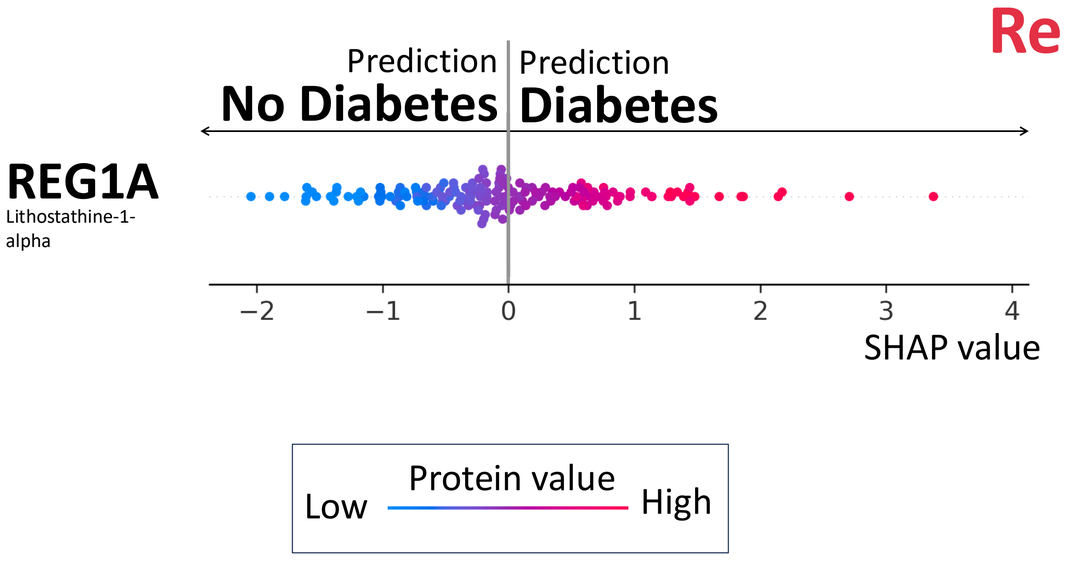
**184** proteins
Target 96 Inflammatory & Immuno-oncology panels

**38** periodontitis
**16** no periodontitis

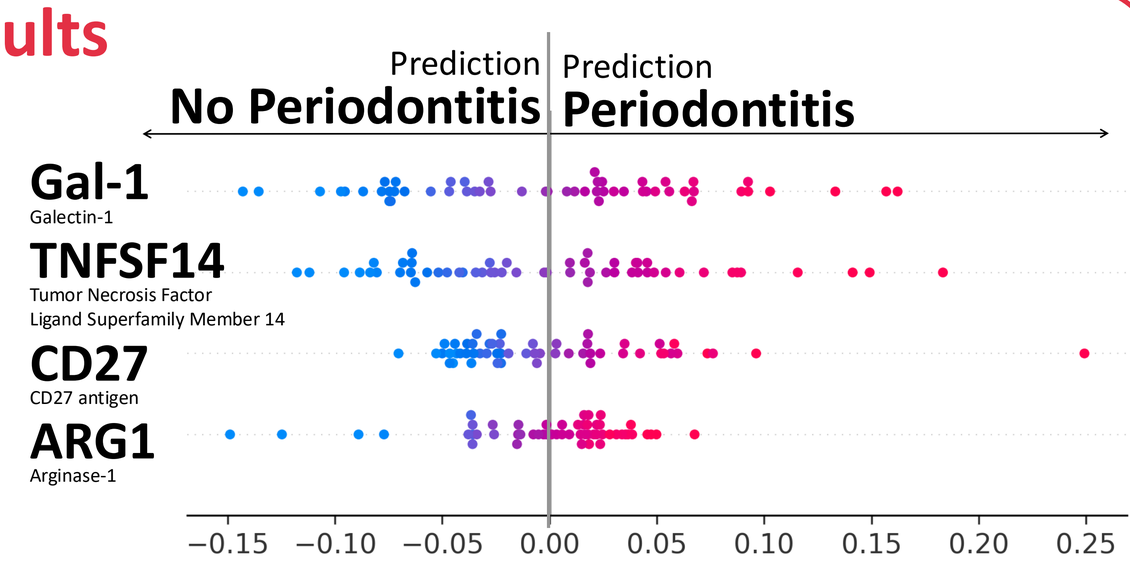
## Methodology

Data → **Repeated nested cross validation**
Protein Selection
*Boruta*
Hyperparameter tuning & Training & Validation
*Logistic Regression, Random Forest, XGBoost*
→ Consensus model
*model with max(ROC AUC) most frequently selected proteins*
→ Interpretation
*SHAP*

## Results




**Logistic Regression**
ROC AUC: **0.79**±0.09;
Sensitivity: **0.68**±0.22; Specificity: **0.76**±0.08




**Logistic Regression**
ROC AUC: **0.86**±0.13;
Sensitivity:**0.76**±0.16; Specificity:**0.80**±0.23

## Conclusions

**REG1A** emerged as the top **diabetes** protein, while **Gal-1**, **TNFSF14**, **CD27**, and **ARG1** were top for **periodontitis**. Oral rinse samples are suitable for protein discovery for oral and systemic diseases.