



Funded by  
the European Union

# Untargeted analyses by LC-HRMS of contaminants with the use of AI

Marco Blokland - Thursday 14 July 2026

## Description of task 2.3

Taken from the project proposal:

Deep Learning for feature detection in LC-HRMS. Data tools for anomaly screening and marker identification will be developed.



# Deep Learning for feature detection in **LC-HRMS**. Data tools for anomaly screening and marker identification will be developed.

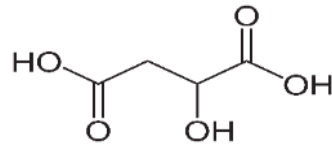


Maize ~ 20000

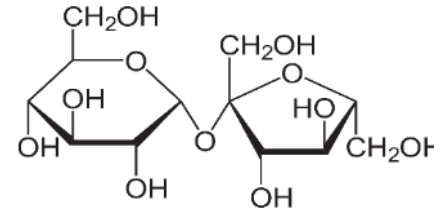


Meat ~ 40000

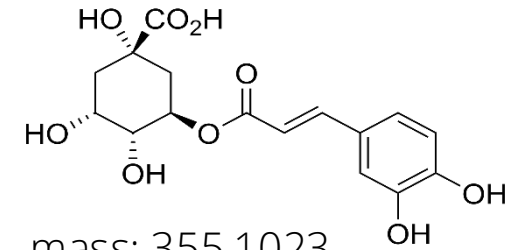
**LC**



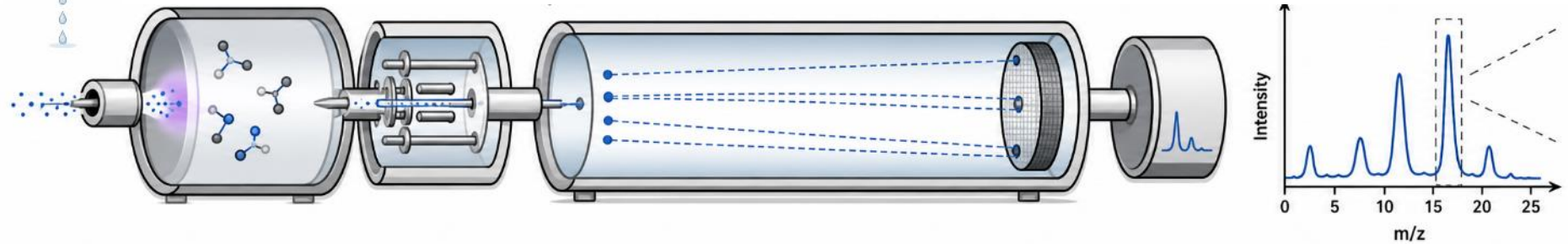
mass: 135.0288



mass: 343.1234



mass: 355.1023

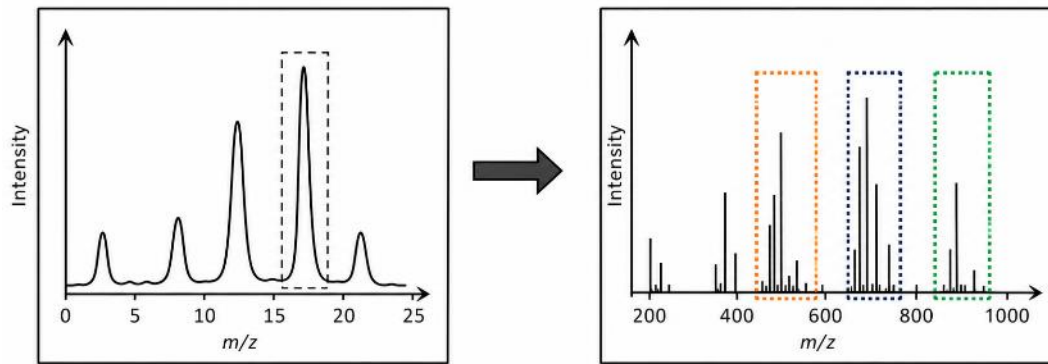


Deep Learning for **feature detection** in LC-HRMS. Data tools for anomaly screening and marker identification will be developed.

## Untargeted HRMS Data

- Complex data files
- Many features in a sample but limited number of samples
- Thousands of chemical signals
- No ground truth

### Feature

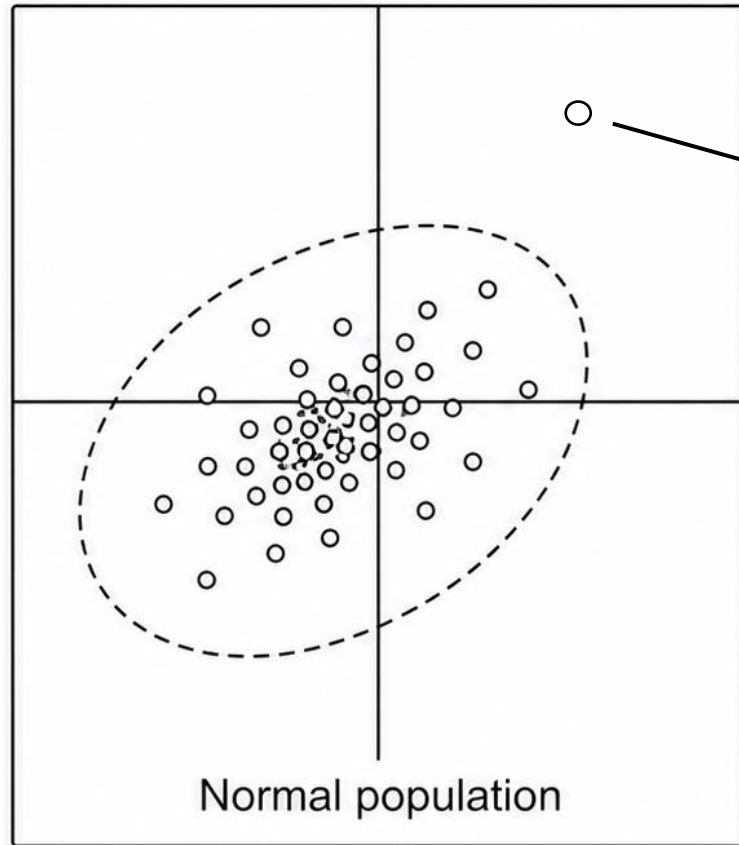


### Feature tables

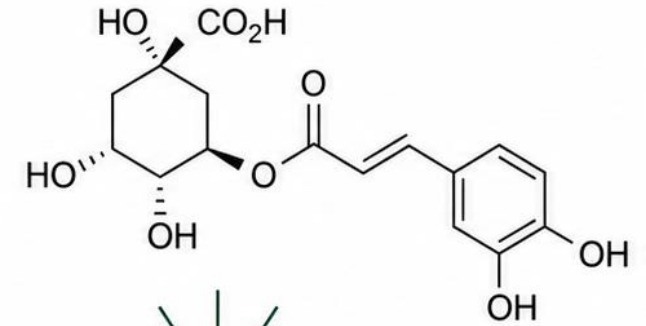
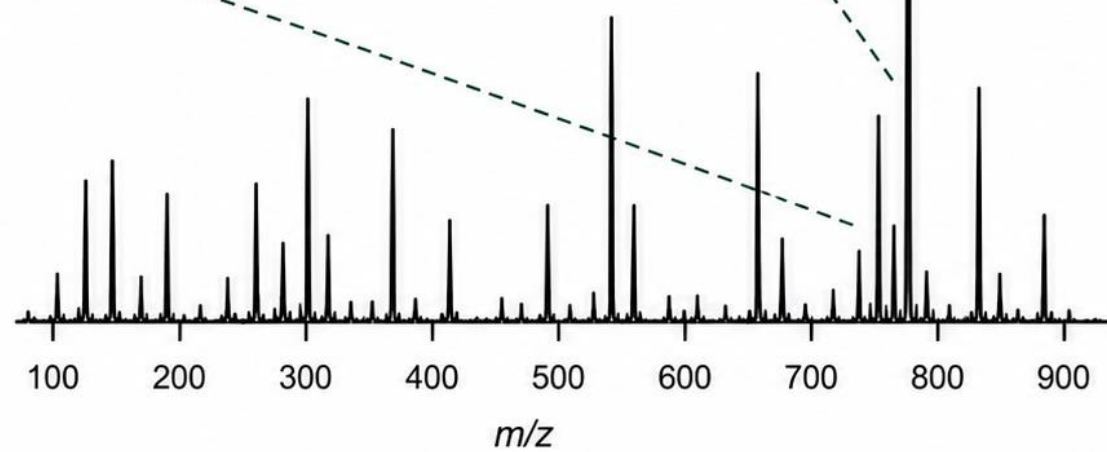
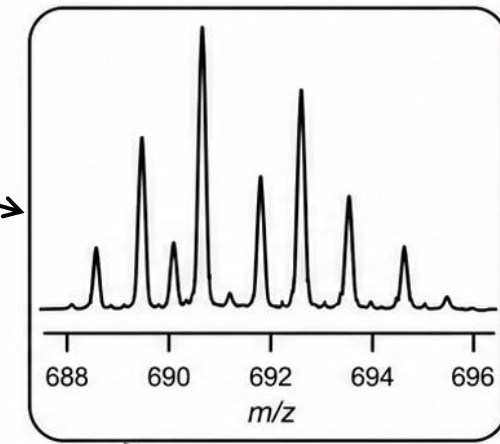
Feature ID	Mass	RT (min)	Intensity Sample 1	Intensity Sample 2	...
F001	301.2167	5.43	1.20E+06	9.80E+05	...
F002	523.2741	7.81	3.50E+05	4.10E+05	...
F003	150.0583	1.92	8.90E+04	7.20E+04	...

Deep Learning for feature detection in LC-HRMS. Data tools for **anomaly screening** and **marker identification** will be developed.

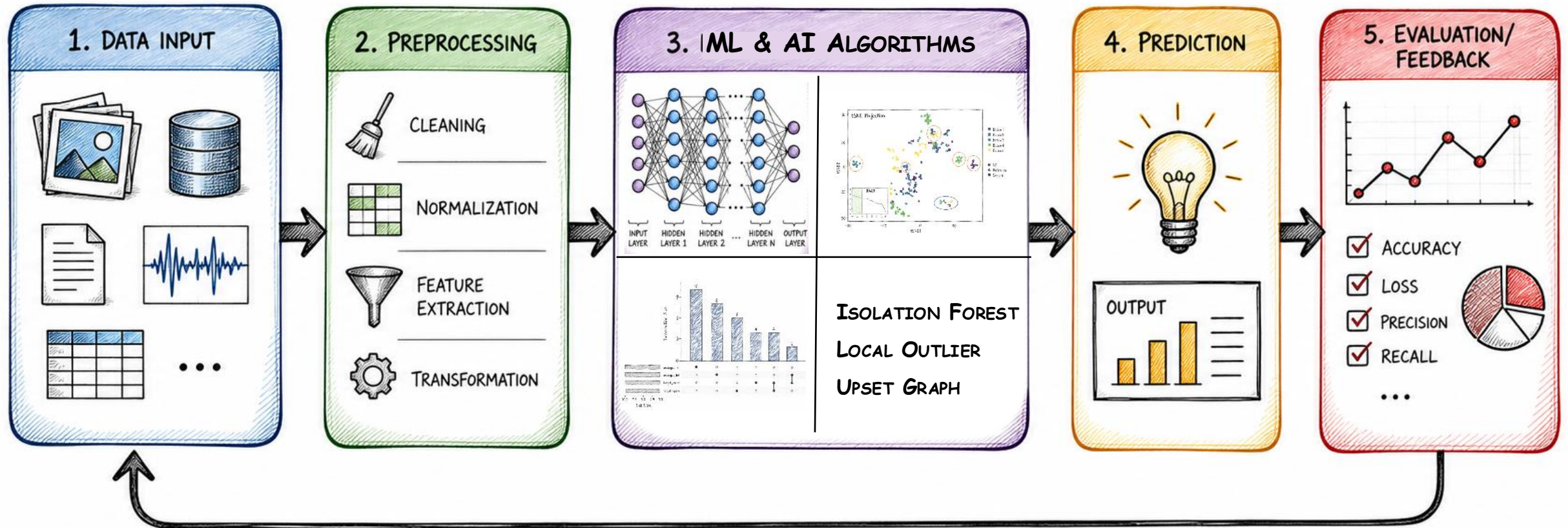
### Anomaly screening



### Marker identification

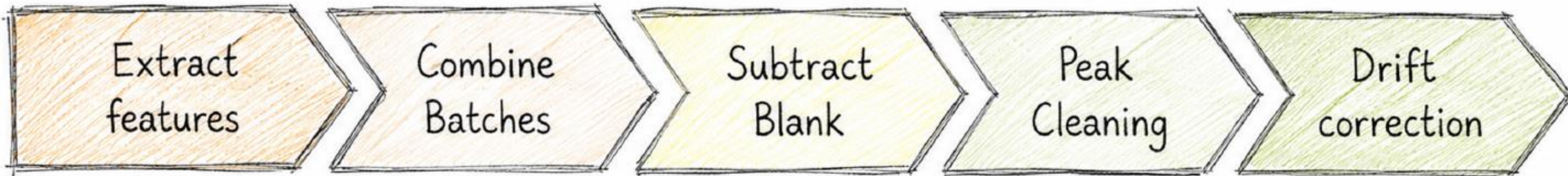


**Deep Learning** for feature detection in LC-HRMS. Data tools for anomaly screening and marker identification will be developed.



Improve and iterate

# LC-HRMS Preprocessing Workflow



*XCMS*  
Features  
extracted per  
batch

*XCMS Combiner*  
Merges all data  
in one  
experiment

Remove features  
removed from  
process blanks

Removes low-  
quality peaks  
from feature  
table

*batchCorr*

- Alignment of features across batches
- Correct intensity drift

## HeRMeS Processing Platform

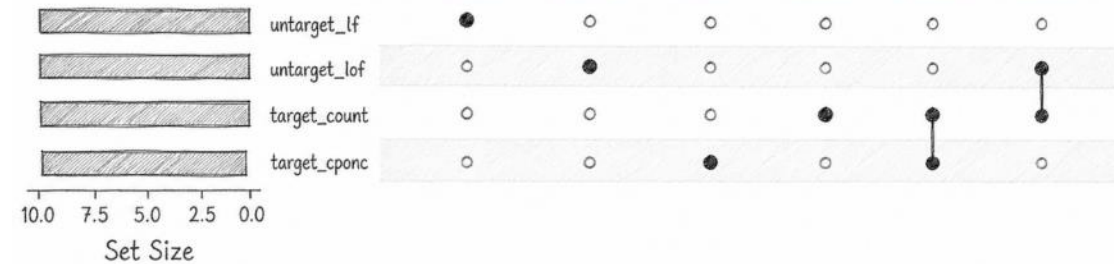
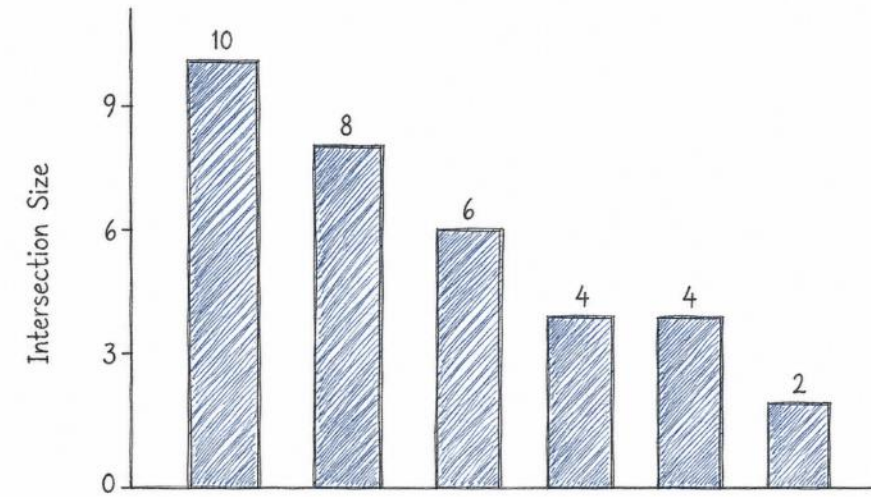
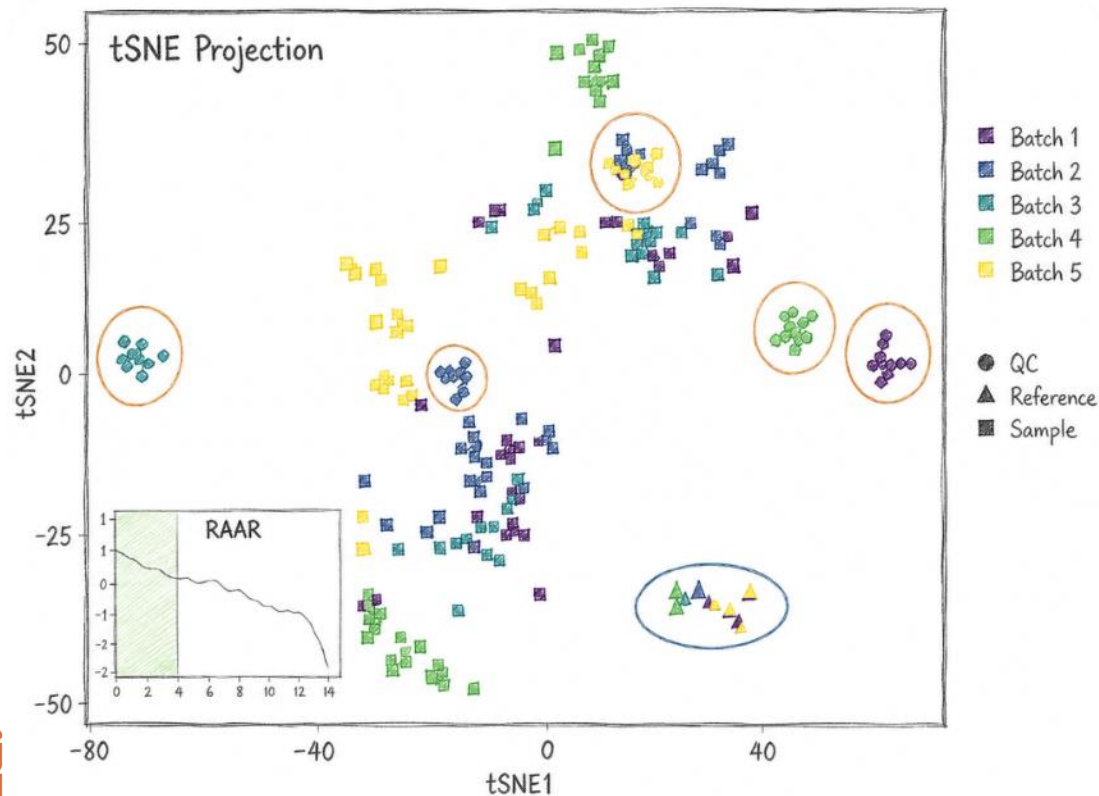
- User friendly interface to support feature extraction
- Integration of long-term reference samples for multi-batch analysis

# Characterization of the data



## Characterize data, maximize group differences

- Projection of feature table into low-dimensional space
- Reveals group structure and batch effects
- Identifies global and local patterns



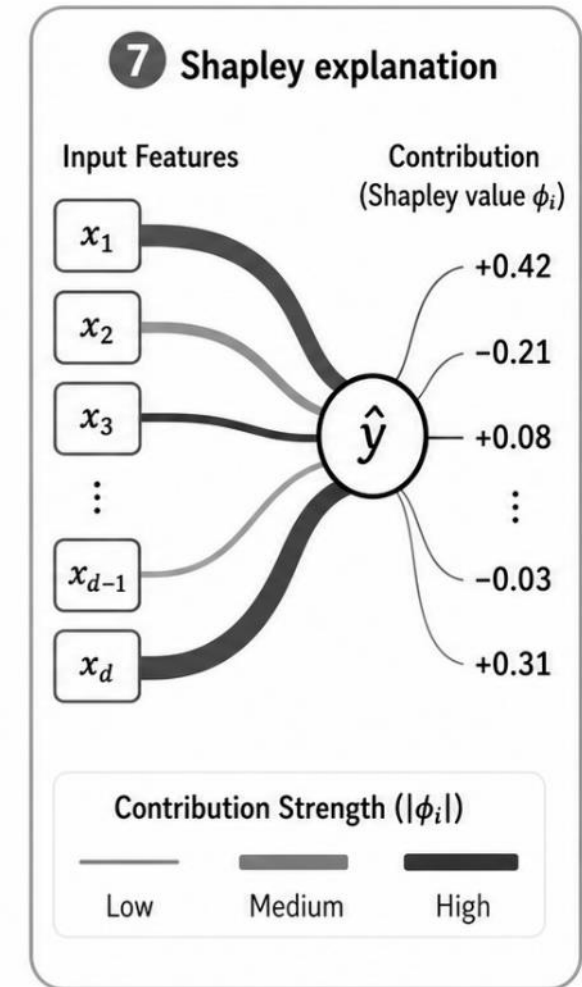
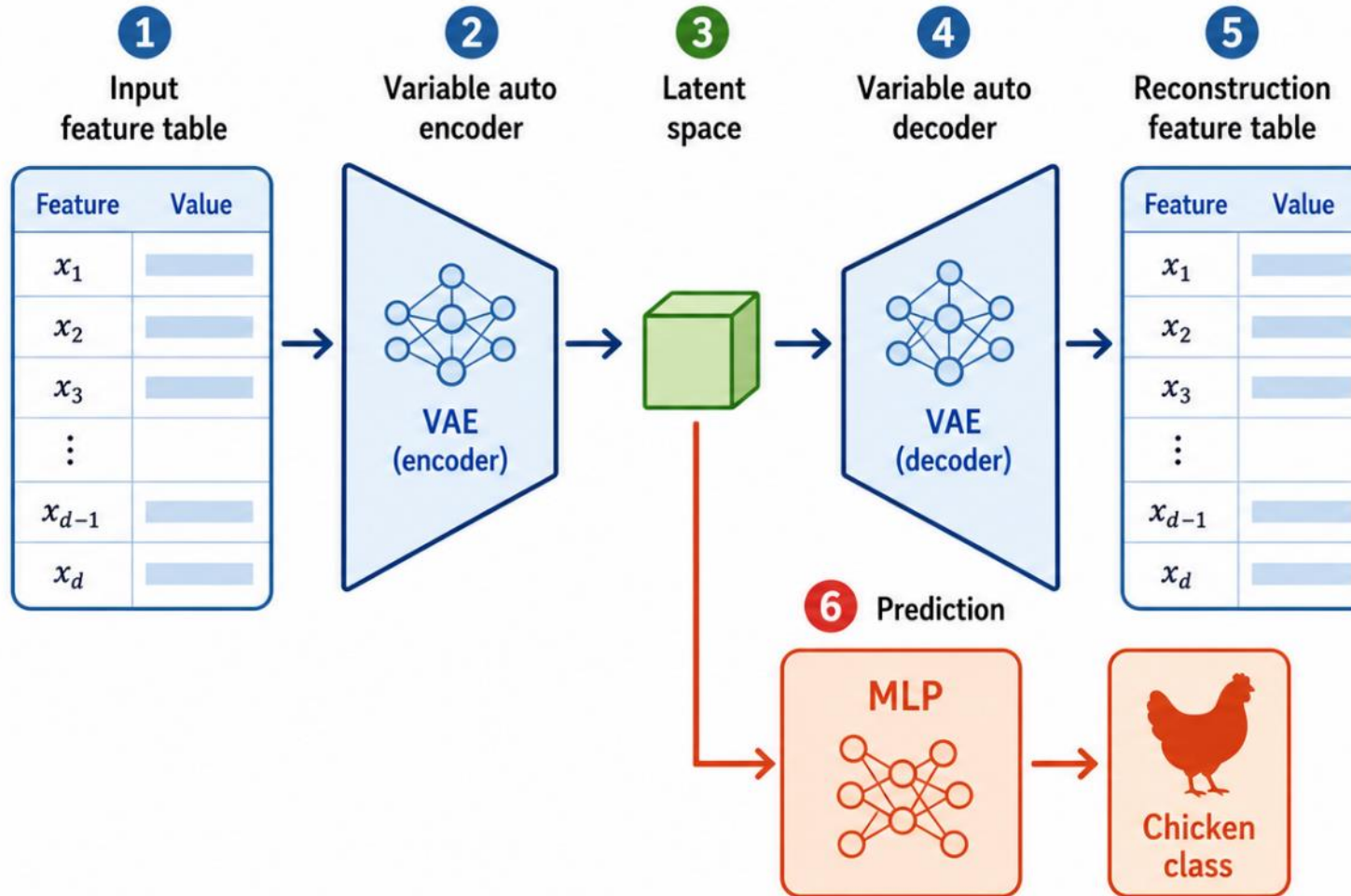
## Multimodal anomaly detector

- Detect outliers using multiple models (VAE, isolation forest, LOF)
- Anomaly ranking
- Combine results for robust detection
- Visualization via UpSet plots

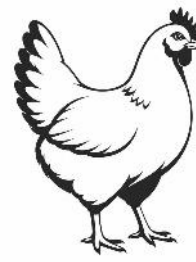
# Building a predictive AI-Model

## The model

Branched Beta Variational Autoencoder



# AI-Based classification of

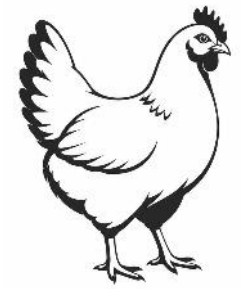


# LC-HRMS data

- ▶ Dataset
  - ▶ Blank (0) vs spiked (1) samples
  - ▶ Levels: 0.1× and 1× MRL
  - ▶ Multiple compounds
- ▶ Model setup
  - ▶ Train/validation split
  - ▶ Independent test set
- ▶ Key findings
  - ▶ Most samples correctly classified
  - ▶ Day effect impacts performance
  - ▶ More data needed → pretrain model

uncontaminated	19	5 FP
contaminated	2 FN	24
	uncontaminated	contaminated

# Pretrained model fails without fine-tuning for



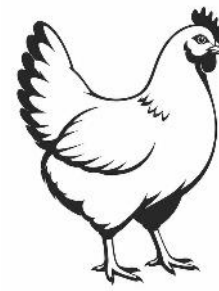
- Setup
  - Pre-training dataset LC-HRMS dataset cow urine
  - ~1200 samples, balanced data-set
  - Transfer of model trained on cow data only
  - Same chicken dataset as baseline

- Observation
  - Predicts nearly all samples same class

- Issue
  - No exposure to chicken data
  - Fine-tuning is required

uncontaminated	23	1 FP
contaminated	26 FN	0
	uncontaminated	contaminated

# Fine-tuned model significantly improves performance for

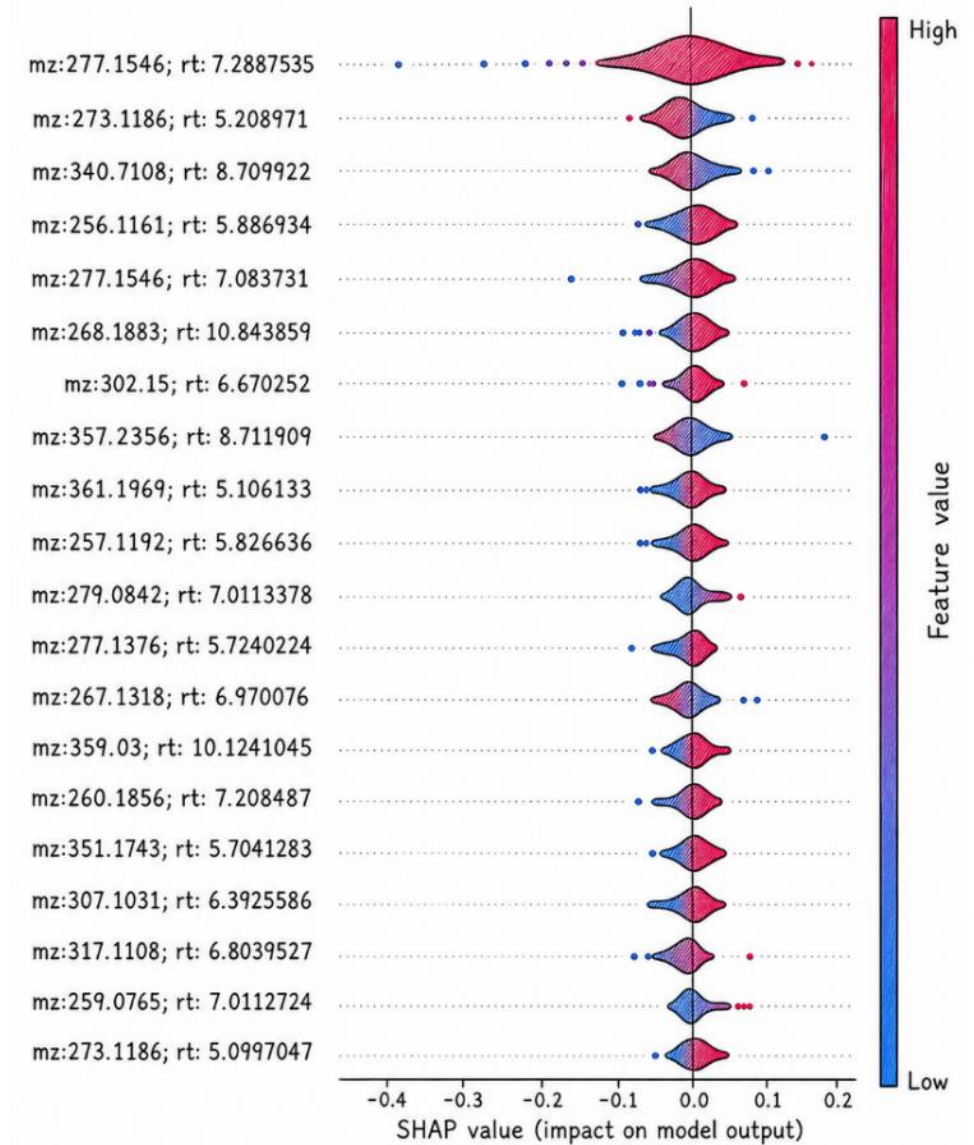


- ▶ Data
  - ▶ LC-HRMS dataset (train/validation + independent test)
- ▶ Preprocessing
  - ▶ Feature selection (WFSR library + blank comparison)
  - ▶ Normalization: PQN, log transform, z-scoring
- ▶ Training
  - ▶ Pre-trained model fine-tuned on chicken data
  - ▶ No information leakage (train-only preprocessing)
- ▶ Successful applied transfer learning

uncontaminated	23	1 FP
contaminated	0 FN	26
	uncontaminated	contaminated

# Which Features Drive the Model? (SHAP Analysis)

- ▶ What does SHAP show?
  - ▶ Explains how each feature influences predictions
  - ▶ Positive values → push toward class 1
  - ▶ Negative values → push toward class 0
- ▶ Analysis levels
  - ▶ Global: important features for the full model
  - ▶ Local: explanation for a single sample



# Transfer learning enables generalisation to

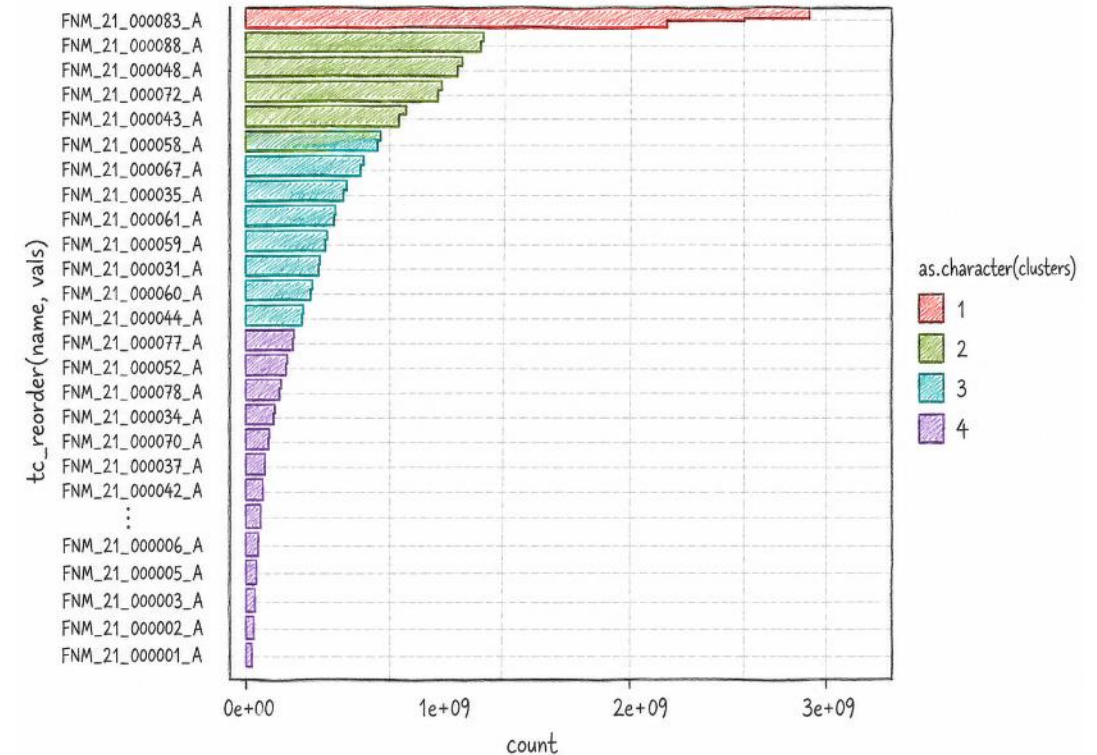
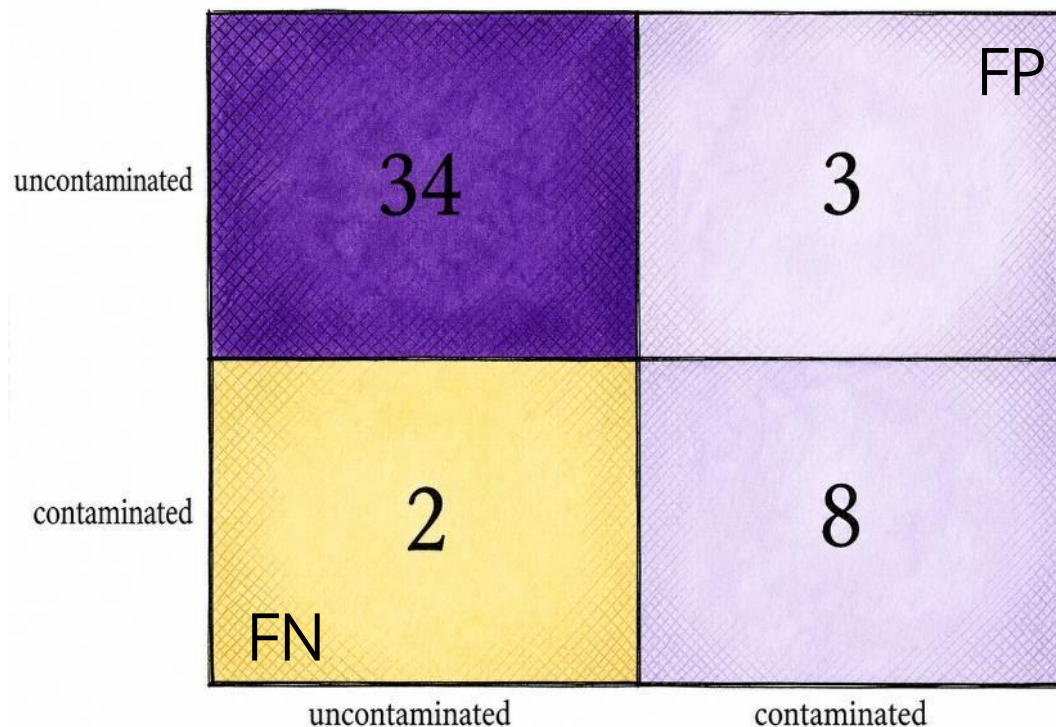


Transfer model to maize data

Same model applied to maize LC-HRMS data

Shared feature table

No retraining on maize

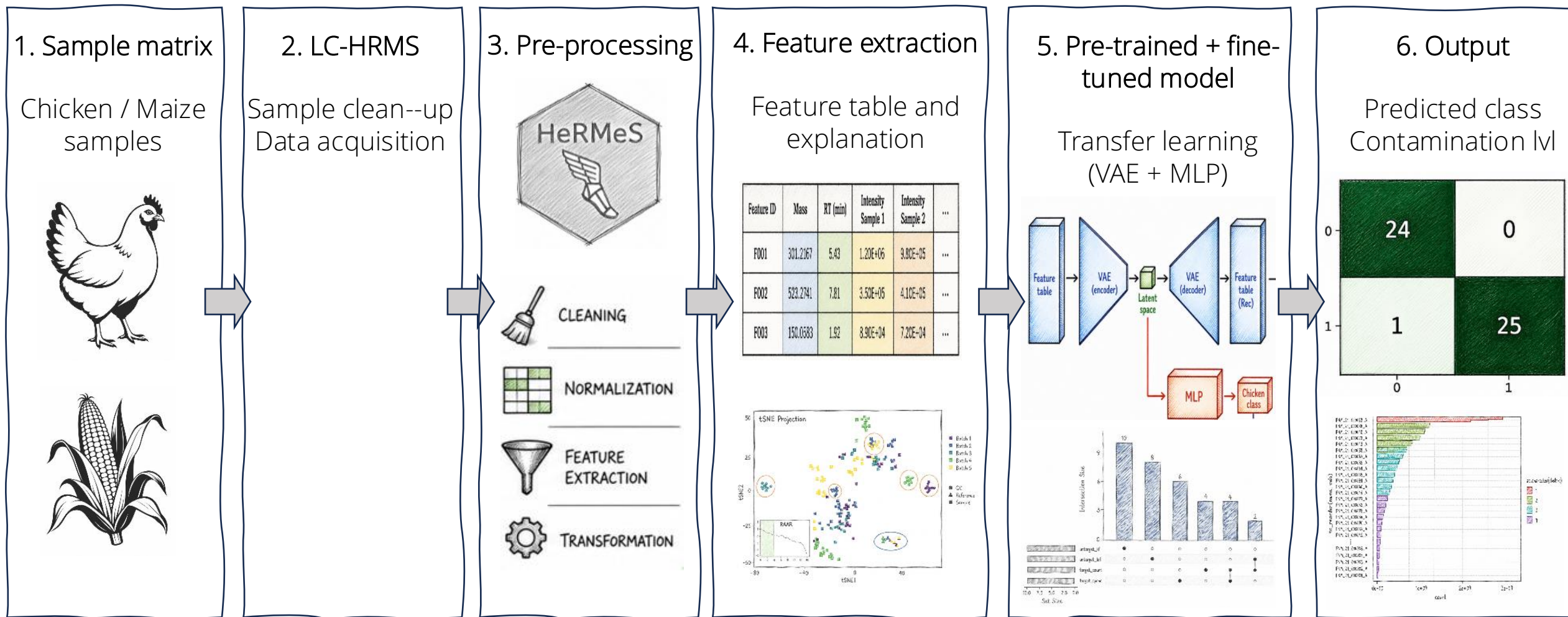


Contamination structure

Continuous contamination levels observed

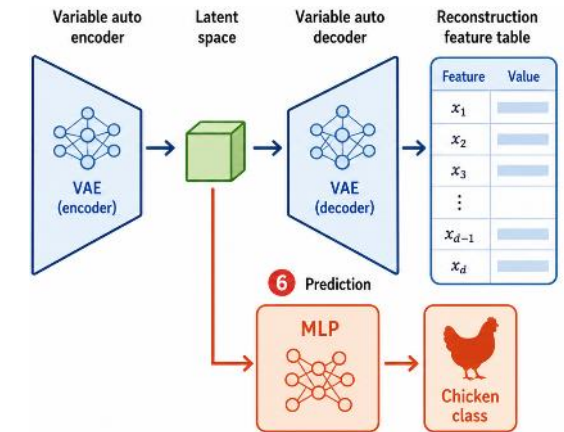
Hierarchical clustering identifies distinct groups

# Developed AI Pipeline for LC-HRMS classification



# Conclusions and takeaways

- Several (AI)-tools were developed for LC-HRMS data evaluation
- Transfer learning enables AI modelling with limited data
- Pre-training on cow data and fine-tuning gives accurate classification
- Model decisions are interpretable and biologically meaningful
- Validate model further in real-life routine analysis
- Developed tools capable of detecting anomalies related to contamination in chicken and maize



# People involved

- Ane Arrizabalaga Larranaga (WFSR)
- Melissa Broeren (WFSR)
- Andrea Junior Carnoli (WFSR) →Poster presentation
- Jakob Kuehn (BfR)
- Xenia Pascari (BfR) →Poster presentation
- Stefan Weigel (BfR)

# Thank you

# Questions?

This project has received funding from the European Union's Horizon Europe research and innovation program under grant agreement No 101059813 and was supported by the Dutch Ministry of Agriculture, Fisheries, Food Security and Nature through the Food Safety Knowledge development program (project KB-56-002-004)