



UMEÅ UNIVERSITET

Chemometric Strategies for Supervised Multi-Model Analysis

Edvin Forsgren

Akademisk avhandling

som med vederbörligt tillstånd av Rektor vid Umeå universitet för avläggande av filosofie doktorexamen framläggs till offentligt försvar i Stora Hörsalen (KBE303), KBC, onsdagen den 16 april, kl. 09:00.

Avhandlingen kommer att försvaras på engelska.

Fakultetsopponent: Professor, Federico Marini,
University of Rome, Rom, Italien.

Kemiska institutionen

Organization
Umeå University
Department of Chemistry

Document type
Doctoral thesis

Date of publication
26 March 2025

Author
Edvin Forsgren

Title
Chemometric Strategies for Supervised Multi-Model Analysis

Abstract

Understanding biological processes is inherently complex. The cellular machinery and biochemical pathways present significant challenges in scientific research. Advances in data collection, such as high-content imaging and omics technologies, have enabled deeper insights, but extracting meaningful conclusions from these complicated datasets remains a challenge. In this thesis, the focus has been on developing chemometric strategies and supervised modelling approaches to improve data interpretation, aiming to aid scientists in drawing conclusions from their data.

In **Paper I**, we show that cell imaging data, combined with chemometric tools, can effectively characterize treatment effects, leading to the development of a metric called Equivalence (Eq.) scores. This work raised two main questions: Are fluorescent labels necessary for meaningful characterization? Can living cells, imaged over time, provide deeper insights? In **Paper III**, we address these questions by investigating an approach based on label-free live-cell imaging data where we extended the Eq. scores to time series data. We demonstrate that time-dependent analysis reveals both early and late cellular responses and improves the prediction of drug mechanisms.

In **Paper II**, we address challenges arising when Orthogonal Projections to Latent Structures-Discriminant Analysis (OPLS-DA) models are used to analyse several classes, such as subtypes of diseases or different treatments. We introduce OPLS-Hierarchical Discriminant Analysis (OPLS-HDA), a method that integrates hierarchical clustering analysis (HCA) with two-class OPLS-DA models to create an OPLS-based decision tree. We demonstrated that OPLS-HDA is a strong classifier compared to eight other established methods while maintaining interpretability. Additionally, we provide Python scripts that are integrated with SIMCA®, offering a user-friendly interface for broader accessibility.

Extracting reliable insights from complex data requires intentional and structured approaches. This work highlights the benefits of modular and interpretable modelling solutions, ensuring that results are both understandable and trustworthy. By breaking down complex analytical challenges and building tools that enhance interpretability, this work contributes to the broader goal of accelerating data-driven discoveries in life sciences.

Keywords

Label-free live-cell imaging, Morphological profiling, Multi-class classification

Language	ISBN	Number of pages
English	print: 978-91-8070-642-1 PDF: 978-91-8070-643-8	58 + 3 papers