

Analysis of historical bioprocessing data gives clues to microbial culture performance

Duygu Dikicioglu and Nishanthi Gangagharan from the University of Cambridge used BioProNET business interaction voucher funding to analyse historical bioprocessing data from Johnson Matthey. The predictive model they developed using this data could be used make current bioprocess operations more efficient, as it was able to identify successful versus poor performing microbial cultures at an earlier stage.

The challenge: Bioprocessing development generates large amounts of data, such as such as process parameters, culture properties and phenotypic characteristics of microorganisms. But this data is often not used as fully as it could be. This can be because the data highly heterogeneous; it is often collected over long periods of time, from different projects, different product types and different host cell types, and is also impacted by changes in instrumentation, inconsistent sampling and missing data.

Project aim: The aim of this project was to pre-process historical bioprocessing data generated from several enzyme-producing microbial systems at Johnson Matthey. Data pre-processing converts unstructured data into a structured format that is then suitable for processing and modelling.

Key methods: A dataset from 87 microbial cultures from two geographical locations collected over a 5-year period (2014-2018) was collated into a structured format. First, missing data was handled using a hybrid machine learning method. This was then followed by clustering of parameters to see which parameters behaved similarly; parameters were then ranked in the order of their importance. Following these pre-processing steps, the collaborators were able to carry out additional work: modelling was performed on the pre-processed data to make predictions of the culture output and performance.

“Early indicators of fermentation run quality were identified that allow us to reduce both failed runs and increase the efficiency of runs” Johnson Matthey

Key outcomes: The collaborators now have an optimised pipeline and relevant models that can predict the performance of a fermentation (both in terms of product quality and quantity) at an earlier onset – as early as halfway through culture. Furthermore, the pipeline can identify the most important parameters that act as indicators of culture performance from an earlier stage.

Further testing on different datasets will help determine if this pipeline has the potential to become a universal tool for bioprocess data analysis.

Next steps:

- The tools developed by the academic partner will be adopted by the industrial partner as part of its bioprocessing pipeline
- Johnson Matthey will sponsor the research of a new PhD student at the University of Cambridge
- The academic partner is investigating grant applications to continue the project and to validate further the modelling approach