

EARLY STAGE BIOTECH BUSINESS GAINS VALUABLE SUPPORT IN MACHINE LEARNING FOR EFFECTIVE ENZYME IDENTIFICATION AND RESEARCH

THE BRIEF

Entropix Ltd is an early-stage life science company based at Sci-Tech Daresbury in Cheshire, UK. The company is developing enzymes with improved performance using synthetic biology and computational analysis techniques. Entropix supports multiple industry sectors including diagnostics, pharmaceuticals, personal care, food, flavours, fragrances, specialty chemicals and environmental remediation/recycling.

To improve the performance of enzymes, Entropix uses a process known as directed evolution in which changes are made to the gene sequence associated with a given enzyme. These changes are then screened for an improvement in enzyme function. Entropix Ltd approached the CW4.0 team at the Virtual Engineering Centre (VEC) to support an investigation into the potential use of using Machine Learning (ML) techniques to relate gene sequence to enzyme function.

THE APPROACH

A proof-of-concept study was created based on a well-studied enzyme with known functional changes related to specific gene sequence modifications. The Industrial Digitisation Team at VEC used their unique Machine Learning and Artificial Intelligence (AI) expertise to investigate how mainstream Machine Learning tools could be applied in supporting the existing data sets and sequence analysis tools used by Entropix.

The VEC team was able to demonstrate that the Machine Learning approach can effectively identify and categorise patterns in the gene sequence that give rise to different functional variants of the enzyme. The tools used by the VEC team were successfully integrated into the Entropix portfolio of computational techniques, ensuring that they can be used and managed easily beyond the support project.

THE BENEFITS

Overall, this project will enable Entropix to scale up their already impressive operations, analysing even more data to reach predictions, results and recommendations more effectively. The ability to predict potential improvements in the functional properties of enzymes (and other proteins) based on gene sequence information offers enormous potential to accelerate developments in the Medtech sector and elsewhere.

The knowledge exchange workshops have outlined and demonstrated the type of open-source libraries used in the work packages and the methods of data analysis and modelling alongside the types of Machine Learning required for gene sequence analysis. This ensures that the Entropix team is well equipped to capitalise on the benefits of this approach in future projects.

The teams also hope to extend the findings of the proof-of-concept project to study variants of the Covid-19 virus, using the wealth of genetic information now available worldwide. By comparing the gene sequences and infectivity of different variants, it may be possible to predict future mutations of the spike protein that could potentially become "variants of concern".

"The support we have received from the CW4.0 team at VEC has been great for Entropix. The teams have learnt a lot in such a short space of time, which is crucial for a young business like Entropix, where time is everything. We not only have a proof-of-concept, but we can now reach results even quicker than before using the new tools and techniques developed by the team at VEC."

- Rob Rule, CEO and co-founder of Entropix Ltd