BIOTRANS
LA ROCHELLE 2023

N°1288 / PC
TOPIC(s) : Enzyme discovery and engineering / Artifical intelligence / computational methods

Highly Accurate Enzymatic Activity Predictions Based on a Machine Learning Algorithm Trained on a Transaminase Screening Data Set

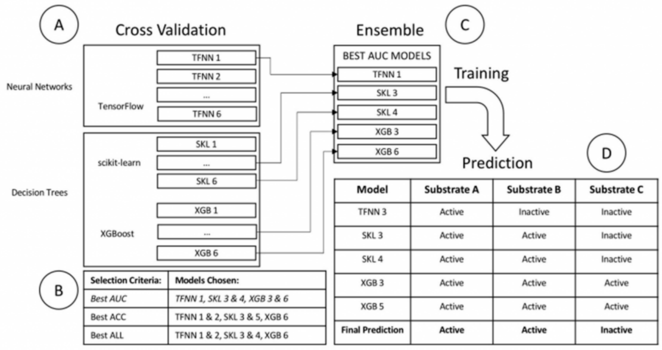
AUTHORS

Oscar ALVIZO / CODEXIS, 200 PENOBSCOT DRIVE, REDWOOD CITY

PURPOSE OF THE ABSTRACT

Directed evolution is a well-established method for improving the performance of an enzyme catalyst. One of the challenges in the process is identifying a sequence that can serve as the starting sequence for the evolution campaign. Ideally, this sequence must be active on the target substrate and produce the desired product with the intended stereochemistry. The duration of an evolutionary campaign is highly dependent on the initial activity of the starting enzyme under the desired process conditions. The more active the starting sequence, the less rounds of evolution required to identify a highly performing variant that meets the required commercial criteria. Here we discuss a machine learning approach used to identify enzymes that can act on the target substrate. The ML model was trained on the activity of transaminases from multiple directed evolutionary lineages and dozens of substrates. The best results were observed when an ensemble of models was used for the predictions. The final optimized pipeline was shown to be 80-90% accurate.

FIGURES



Selection Criteria:	Models Chosen:
Best AUC	*TFNN 1, SKL 3 & 4, XGB 3 & 6*
Best ACC	TFNN 1 & 2, SKL 3 & 5, XGB 6
Best ALL	TFNN 1 & 2, SKL 3 & 4, XGB 6

Model	Substrate A	Substrate B	Substrate C
TFNN 3	Active	Inactive	Inactive
SKL 3	Active	Active	Inactive
SKL 4	Active	Active	Inactive
XGB 3	Active	Active	Active
XGB 5	Active	Active	Active
Final Prediction	**Active**	**Active**	**Inactive**

FIGURE 1

Architectural Flow Diagram

A diagram depicting the steps used to generate the machine learning model for the prediction of active enzymes.

FIGURE 2

KEYWORDS

Machine Learning | Enzyme Engineering | Transaminases

BIBLIOGRAPHY

[1] Yang, K.K., Z. Wu, and F.H. Arnold, Nat Methods, 2019. 16(8): p. 687-694.
[2] Singh, N., et al., Emerg Top Life Sci, 2021. 5(1): p. 113-125.
[3] Wu, Z., et al., Proc Natl Acad Sci U S A, 2019. 116(18): p. 8852-8858.