Deep Learning Framework Accelerates Computational Enzyme Characterization

Background/Objective

- Estimation of enzymatic activities relies heavily on experimental assays, which can be cost and time-intensive. Predictive models that enable quantitative annotation of enzyme kinetics could be enabling for enzyme characterization in the same manner AlphaFold has become for structure prediction.

Approach

- This work presents CatPred, a deep probabilistic learning framework for predicting *in vitro* enzyme kinetic parameters, including turnover numbers (k_{cat}), Michaelis constants (K_m), and inhibition constants (K_i).
- CatPred tackles major challenges in the field, including the absence of standardized datasets, the need for reliable performance on enzyme sequences dissimilar to the training data, and the importance of quantifying model uncertainty.

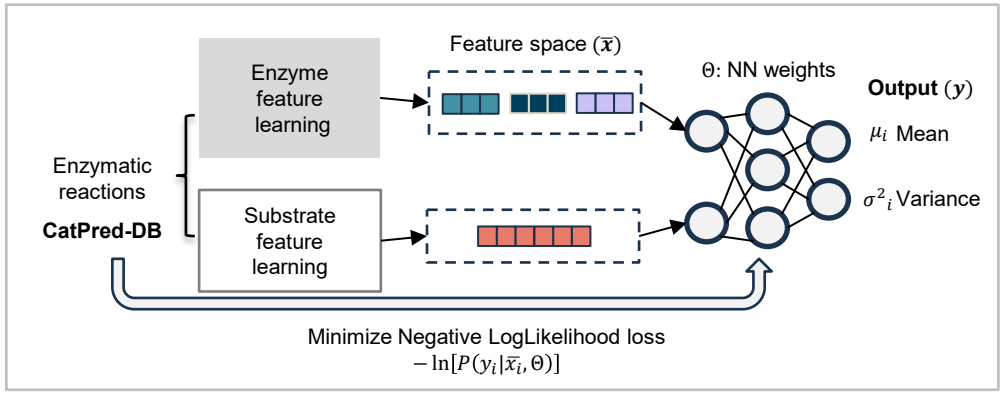
Results

- CatPred predicts k_{cat} with up to 40% more explained variance than state-of-the-art models on highly dissimilar sequences.
- CatPred delivers accurate predictions with query-specific uncertainty, where lower variance indicates higher accuracy.
- CatPred-DB offers benchmark datasets with broad coverage: ~23 k, 41 k, and 12 k data points for k_{cat}, K_m, and K_i respectively.

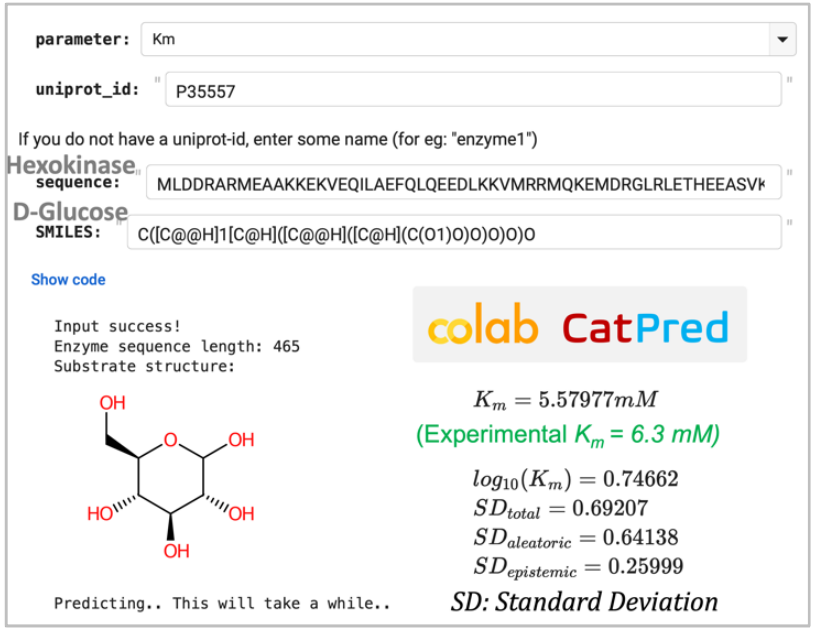
Significance/Impacts

- CatPred lays the groundwork for accelerating novel enzyme discovery, *in silico* directed evolution, and kinetic metabolic modeling. It can be readily adapted to tackle a wide range of enzyme design challenges, including those aimed at improving biofuel production.

Boorla et al., Nat Commun 2025. doi: 10.1038/s41467-025-57215-9



CatPred models are trained on CatPred-DB datasets utilizing both substrate & enzyme feature learning modules using probabilistic regression. The enzyme & substrate features are fed as inputs, and the kinetic parameters are predicted as outputs.



CatPred production models are made available through the Google-Colab interface for ease of access by the scientific community.