

Paper Validation Report

for **dnaZyme**



Correspondence: Yes

Percentages: 92.0%

Conclusion:

The codebase provides clear implementations of the key methods, models, and algorithms described in the paper's experiments section. It covers the extraction and normalization of biochemical and sequence features (k-mer, autoencoder, descriptors), construction of feature sets, and machine learning pipelines for predicting DNAzyme activity (LightGBM, Random Forest, etc.). The code also implements the same dataset splits, feature selection, descriptor engineering, and cross-validation procedures as the paper. Evaluation metrics such as R2 and RMSE are present. However, the exact reproduction of figures/tables and access to the raw or augmented dataset are not directly verifiable from the code snippets provided. Also, end-to-end scripts for generating main result plots/outputs seem to be absent or fragmented, introducing minor barriers to full reproducibility. Nevertheless, the repository aligns closely with the scientific claims, justifying a high score.