

The backstage of the third Health Data Challenge: Introducing an Organizational Template and Custom, Temporary and Local Codabench Instances.

Nicolas HOMBERG¹, Magali RICHARD¹ and Florent CHUFFART²

¹ Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France

² Univ. Grenoble Alpes, INSERM U1209, CNRS UMR5309, IAB, 38000 Grenoble, France

Corresponding Author: nicolas.homberg@univ-grenoble-alpes.fr

Keywords

Data challenge, Benchmark, Scientific competition, Reproducibility,

Abstract

Codabench [1] is an open-source web platform that aims toward reproducible research and benchmarking. Codabench, established in 2023, represents the upgraded iteration of the competition platform CodaLab, which originated in 2013. Alongside EvalAI, Alcrowd, and Kaggle, CodaLab stands out as one of the leading platforms for academic competitions today [2], renowned for its user-friendly interface and open-source framework. Codabench is a platform of choice for creating data-centric Artificial Intelligence (AI) challenges which engage several participants with substantial datasets, often requiring heavy computation. Organizers of challenges can use the public instance of Codabench at no cost to host their own competitions, provided they adhere to the platform's terms and conditions. Nevertheless, the tasks of setting up a data challenge and managing computational resources remain time-consuming and inherently challenging. This poster serves three main purposes: i) to introduce the Codabench platform; ii) to guide users through challenges generation, emphasizing a suggested

organizational template and iii) to demonstrate the installation process of a local instance of Codabench.

Our organizational template relies on bipartite git repositories devoid of binary files. The first public git repository ensures fairness for all participants by providing source files that ingest and evaluate submission along with the docker recipe and a basic working submission. Additionally, this git repository also includes a script to generate the bundle ready to be submitted to Codabench, featuring either simulated test data or the challenge target data. The latter data depends on a private git repository, accessible only to the administrators of the competition. This private repository contains the recipe to create the challenge target data that will be used during the competition. In parallel, we are developing a novel procedure to automate the launch of custom-sized, temporary, and local Codabench instances via OpenStack based virtual machines (VMs) hosted by the GRICAD mesocentre of the University of Grenoble Alpes (UGA). This organizational template and the automated Codabench instance will be deployed from December 2nd to 6th, 2024, for the third Health Data Challenge (HADACA3). This challenge will aim to explore and potentially discover new methods to resolve cell-type deconvolution problems.

References

1. Xu Z, Escalera S, Pavão A, Richard M, Tu WW, Yao Q, Zhao H, Guyon I. Codabench: Flexible, easy-to-use, and reproducible meta-benchmark platform. *Patterns* (N Y). 2022 Jun 24;3(7):100543. doi: 10.1016/j.patter.2022.100543. PMID: 35845844; PMCID: PMC9278500
2. Carlens, H, "State of Competitive Machine Learning in 2023", ML Contests Research, 2024.