



HAL
open science

codabench, a web-plateform to organize scientific competitions

Florent Chuffart, Anne-Catherine Letournel, Isabelle Guyon, Magali Richard

► **To cite this version:**

Florent Chuffart, Anne-Catherine Letournel, Isabelle Guyon, Magali Richard. codabench, a web-plateform to organize scientific competitions. JOBIM2024, Jun 2024, Toulouse, France. hal-04664118

HAL Id: hal-04664118

<https://hal.science/hal-04664118v1>

Submitted on 29 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

codabench, a web-plateform to organize scientific competitions

Florent CHUFFART¹, Anne-Catherine LETOURNEL², Isabelle GUYON³ and Magali RICHARD⁴

¹ Université Grenoble-Alpes, Inserm, CNRS, Institute for Advanced Biosciences, Grenoble, France

² CNRS, Université Paris-Saclay, LISN Laboratoire Interdisciplinaire des sciences du Numériques, Gif-sur-Yvette, France

³ Université Paris-Saclay, France and ChaLearn, USA

⁴ Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France

Corresponding author: magali.richard@univ-grenoble-alpes.fr

Abstract

Background

Codabench [1] is an open-source web-based portal developed to address the growing difficulties of reproducibility and comparability in the data science communities. Codabench is a new and improved version of "Codalab competition" [2], the most widely used platform in the world for organizing machine learning challenges [3]. While maintaining full backward compatibility with Codalab, Codabench features reworked code to efficiently manage competitions and benchmarks. Despite their solid mathematical foundations, data analysis methods often require empirical evaluation to confirm their effectiveness and reliability. This trend is intensifying with the growing complexity of the methods, particularly with the emergence of deep neural networks, which are difficult to explain and interpret. Empirical evaluation is becoming indispensable, because of the complexity of the algorithms and the unpredictable nature of the data. In addition, discrepancies in hardware/software environments, code and data used, make computational approaches difficult to compare. These challenges underline the need to develop tools and methodologies to guarantee the reproducibility and comparability of computational research results. Codabench aims to solve these problems by providing a flexible and reproducible platform for the comparative analysis of algorithms and computational methods. Codabench enable the organization of scientific competitions (also known as "challenges"). These scientific competitions enables large-scale benchmarking of methods as well as the collaborative development of new algorithms to solve complex methodological problems.

Results

Codabench focuses on tasks (i.e. solving a data analysis problem), giving organizers the flexibility to implement various mathematical challenges, to use different data formats and APIs, and even to integrate data-generating models, making it easier to organize challenges in fields such as reinforcement learning. The platform can be used to create both benchmarks and competitions. A distinctive feature of Codabench is its support for the submission of results and code, enabling rigorous and reproducible evaluation of methods, ensuring execution in a controlled environment on the server and thus preventing cheating. Codabench offers "compute workers" to process submissions, but it also enables organizers to create customized queues and attach their own compute resources, guaranteeing great flexibility and avoiding cost overruns. The use of Docker environments for secure execution of participants' code ensures consistency across different environments and enhanced reproducibility of results.

Codabench offers a user-friendly interface which guides organizers step-by-step in the preparation of data and code, facilitating their management. Organizers can download and update “competition bundles”, ZIP files containing all the elements required to create a competition, making the creation process simple and flexible. Codabench also offers flexible competition management by allowing competitions to be structured into several phases, each with its own distinct parameters, dates, data and scoring programs, making it easy to manage and segment the various stages of the competition. The highly customizable dashboard can handle multiple scoring functions simultaneously, enabling a more in-depth evaluation of participants’ results.

Finally, Codabench’s well-organized documentation, categorized according to the different players involved, ensures efficient use of the platform by participants, organizers, administrators and contributors. Compared with its main competitors (see Table 1), Codabench offers much-appreciated flexibility in research and teaching, with unique features that meet the needs of researchers and educators alike.

Tab. 1. Comparison of various reproducible science platforms, table originally published in Xu et al [1]. ‘Bundle’ means whether a wrap up is provided for a benchmark such that we could reuse or share. ‘Result/Code/Dataset’ submit means whether different submissions are supported to enable flexible tasks. ‘Compute queue’ means, where public or private computation resources could be provided or linked for convenient deployment.

Platform	Flexibility			Easy-to-use				Reproducibility
	Bundle	Result/Code submit	Dataset submit	Easy creation	Open-source/free access	API access	Compute queue	
Kaggle	✗	✓	✗	✓	✗	✓	✓	✓
Tianchi	✗	✓	✗	✓	✗	✗	✓	✓
CodaLab	✓	✓	✗	✓	✓	✗	✓	✓
UCI	✗	✗	✓	✗	✓	✗	✗	✓
OpenML	✗	✓	✓	✓	✓	✓	✗	✓
PapersWithCode	✗	✓	✗	✓	✓	✗	✗	✓
DAWNBench	✗	✓	✗	✗	✓	✗	✗	✓
Codabench	✓	✓	✓	✓	✓	✓	✓	✓

To demonstrate the usage of Codabench, we will present two case studies illustrating how Codabench can effectively contribute to both teaching and scientific research within the bioinformatics community.

Algorithms evaluation and collaborative science. As part of the Health data challenge program ([HADACA](#), mainly funded by the European EIT Health program), we organized several data challenges [4,5]. The objective of this program was to provide : firstly, to establish a framework for evaluating newly developed algorithms tailored for the analysis of health data; and secondly, to introduce innovative methodologies aimed at training students, researchers in the collaborative development of computational methods for biological data analysis. All these scientific events were dedicated to the quantification of intra-tumor heterogeneity using appropriate statistical methods on transcriptomic and

methylation data. Here we will provide a comprehensive account of our experiences and insights gained from these competitions.

Data challenge as a pedagogical approach. Codabench is playing an increasingly important role in education. Every year, students in the Master's program in Artificial Intelligence at the Université Paris-Saclay organize challenges using Codalab, which are then solved by other students. Some of these challenges even serve as the basis for international research competitions. Here we will present and discuss the challenge [EPISMOKE2.0](#), a challenge specially designed for students of the Master in Data Science at the University of Grenoble. Students were asked to predict the smoking status of patients based on the methylation profile of their blood DNA. This challenge offers a valuable opportunity for students to apply their theoretical knowledge to a concrete problem, while developing their skills in data analysis and machine learning.

Usefull resources. [Codabench platform](#) and [documentation](#).

Conclusions

In conclusion, Codabench represents a significant advance in the field of data science, offering a flexible and reproducible platform for the comparative analysis of computational methods. Its user-friendliness, flexibility and ability to guarantee reproducibility make it a valuable tool for researchers working in bioinformatics and computational biology. By facilitating comparability of results and fostering reproducibility, Codabench helps to promote high-quality research and accelerate progress in data science.

References

- [1] Xu Z, Escalera S, Pavão A, Richard M, Tu WW, Yao Q, et al. Codabench: Flexible, easy-to-use, and reproducible meta-benchmark platform. *Patterns*. 2022 Jun;0(0). Publisher: Elsevier. Available from: [https://www.cell.com/patterns/abstract/S2666-3899\(22\)00146-5](https://www.cell.com/patterns/abstract/S2666-3899(22)00146-5).
- [2] Pavao A, Guyon I, Letournel AC, Tran DT, Baro X, Escalante HJ, et al. CodaLab Competitions: An Open Source Platform to Organize Scientific Challenges. *Journal of Machine Learning Research*. 2023;24(198):1-6. Available from: <http://jmlr.org/papers/v24/21-1436.html>.
- [3] Carlens H. The State of Competitive Machine Learning;. Available from: <https://mlcontests.com/state-of-competitive-machine-learning-2023/>.
- [4] HADACA consortium, Decamps C, Privé F, Bacher R, Jost D, Waguët A, et al. Guidelines for cell-type heterogeneity quantification based on a comparative analysis of reference-free DNA methylation deconvolution software. *BMC Bioinformatics*. 2020 Dec;21(1):16. Available from: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-3307-2>.
- [5] Decamps C, Arnaud A, Petitprez F, Ayadi M, Baurès A, Armenoult L, et al. DECONbench: a benchmarking platform dedicated to deconvolution methods for tumor heterogeneity quantification. *BMC bioinformatics*. 2021 Oct;22(1):473.