DeepHealth contribution to Chapter 1 – General

Monica Caballero https://orcid.org/0000-0002-5723-249X
Marco Aldinucci https://orcid.org/0000-0001-8788-0829
Jon Ander Gómez https://orcid.org/0000-0002-4174-3762

1.1. The DeepHealth approach to HPC and AI convergence

The DeepHealth - Deep-Learning and HPC to Boost Biomedical Applications for Health - project (https://deephealth-project.eu/) is one of the innovation actions supported by the EU to boost AI and HPC leadership and promote large scale pilots. DeepHealth is a 3-year project, kicked-off in January 2019. DeepHealth aims to foster the use of technology in the Healthcare sector by reducing the current gap between mature enough AI-medical imaging solutions and their deployment in real scenarios. Its main goal is to put HPC power at the service of biomedical applications that require the analysis of large and complex biomedical datasets and apply DL and Computer Vision (CV) techniques to support new and more efficient ways of diagnosis, monitoring and treatment of diseases.

DeepHealth motivation originates from two different main observations: the decoupled AI and HPC worlds and the benefits these two technologies together can bring to the healthcare domain.

The DeepHealth perspective on the current relation between HPC and AI is as follows: the ability of AI-related techniques to analyze data accurately is growing at a breakneck pace. Among these techniques, Deep Learning (DL) has benefited from crucial results in machine learning theory and the large availability of data that is intimately linked to the ability to generalize this data and transform it into useful knowledge. The accuracy of the process is crucially related to the quality and quantity of data and the computing power needed to digest the data. For this High-Performance Computing (HPC) is an enabling platform for AI. On the other hand, supercomputers are shifting to GPUs because of their better energy efficiency and need for more and more GPU-enabled workloads, such as DL.

Despite this potential, supercomputers are rarely used for AI. They are not yet equipped to effectively support specific AI software tools nor to acquire large amounts of data securely as required from medical field applications. Additionally, AI researchers are not used to the batch execution model used in supercomputers. Notwithstanding, the two communities need each other and are fated to meet and both technologies are at the cross-road of the European digital sovereignty challenge, which is one of the key items in the EU commission agenda.

With regards to the healthcare domain, this is one of the key sectors of the global economy, making any improvement in healthcare systems to have a high impact on the welfare society. European public health systems are generating large datasets of biomedical data, in particular images that constitute a large unexploited knowledge database, since most of its value comes from the interpretations of the experts. Nowadays, this process is still performed manually in most cases. In the context of automating and accelerating the analysis of the health data and processes, moving towards the so-called "fourth paradigm of science", unifying the traditionally separated environments of HPC, Big Data analytics and AI can overcome current issues and foster innovative solutions, in a clear path to more efficient healthcare, benefitting to people and public budgets.

In this context, following its main goal, the DeepHealth project proposes a unified European framework that offers DL and CV capabilities, wholly adapted to exploit underlying heterogeneous HPC and cloud architectures by taking advantage of Big Data software tools to distribute computing workload.

DeepHealth advocates system software tools and developer libraries as the cornerstones of the European value chain on digital technologies and of the AI-HPC convergence and its applications in several verticals, such as the medical domain.

The framework is composed, on the one hand, by the DeepHealth toolkit, a free open-source

software (https://github.com/deephealthproject/) that includes two libraries, the European Distributed Deep Learning Library (EDDLL) and the European Computer Vision Library (ECVL), both libraries are ready to be integrated into any software platform to facilitate the development and deployment of new applications for specific problems. On the other hand, DeepHealth also provides HPC and cloud infrastructure support, with a focus on usability and portability, so the procedure for training predictive models could be efficiently distributed on Hybrid and Heterogeneous HPC, Big Data and Cloud architectures in a transparent manner. For that, it relies in the StreamFlow Workflow Manager System [1] and the COMPs [1] parallel programming framework.

The DeepHealth framework allows data scientists and IT experts to train models over hybrid HPC, cloud and Big Data architectures without a profound knowledge of Deep Learning, HPC, Big Data, distributed computing or Cloud computing, and increase their productivity reducing the required time to do it. Additionally, DeepHealth widens the use of and facilitates the access to advanced HPC, Big Data and cloud infrastructures to any company or institution. Furthermore, the usefulness of the DeepHealth proposed framework goes beyond the Health sector, being it applicable to any application domain or industry.

Bibliography

[1] I. Colonnelli, B. Cantalupo, I. Merelli and M. Aldinucci, "StreamFlow: cross-breeding cloud with HPC," *IEEE Transactions on Emerging Topcis in Computing,* 2020.

[2] F. Lordan, E. Tejedor, J. Ejarque, R. Rafanell, J. Álvarez, F. Marozzo, D. Lezzi, R. Sirvent, D. Talia and R. Baida, "ServiceSs: an interoperable programming framework for the Cloud," *Journal of Grid Computing,* vol. 12, no. 1, pp. 67-91, 2014.