



DEEPHEALTH

Deep-Learning and HPC to Boost Biomedical Applications for Health

HPC, Big Data, IoT and AI future industry-driven collaborative strategic topics virtual workshop – HPC / HPDA spectrum

Monica Caballero (Project Manager) - monica.caballero.galeote@everis.com

Jon Ander Gómez (Technical Manager) – jon@upv.es

Eduardo Quiñones (HPC expert) – eduardo.quinones@bsc.es

Marco Aldinucci (HPC expert) - aldinuc@di.unito.it

July 3th 2020



About DeepHealth

Aim & Goals

- Put **HPC computing power at the service of biomedical applications** with DL needs and apply DL techniques on large and complex image biomedical datasets to support **new and more efficient ways of diagnosis, monitoring and treatment of diseases.**
- Facilitate the daily work and **increase the productivity of medical personnel and IT professionals** in terms of image processing and the **use and training of predictive models** without the need of combining numerous tools.
- Offer a **unified framework** adapted to exploit underlying heterogeneous **HPC and Cloud architectures** supporting state-of-the-art and next-generation **Deep Learning (AI) and Computer Vision algorithms** to enhance European-based medical software platforms.

Key facts



Duration: 36 months
Starting date: Jan 2019



Budget 14.642.366 €
EU funding 12.774.824 €



22 partners from **9 countries:**
Research centers, Health organizations,
large industries and SMEs

Research Organisations



UNIVERSITAT POLITÈCNICA DE VALÈNCIA



BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación



list lab



UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA



Karolinska Institutet



CRS4



UNIVERSITÀ DEGLI STUDI DI TORINO



OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

Health Organisations



Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana



CHUV



SPITALUL CLINIC PROF. DR. THEODOR BURGHELE BUCUREȘTI



Stockholms läns landsting

Large Industries



everis



PHILIPS

an NTT DATA Company



THALES



SIMAVI
Software Imagination & Vision

SMEs



WINGS
ICT SOLUTIONS



TREE
TECHNOLOGY



prodesign
electronic growth



Stelar
research

Developments & Expected Results

- **The DeepHealth toolkit**

- Free and open-source software: 2 libraries + front-end.
 - **EDDLL**: *The European Distributed Deep Learning Library*
 - **ECVL**: *the European Computer Vision Library*
- Ready to run algorithms on Hybrid HPC + Cloud architectures with heterogeneous hardware (Distributed versions of the training algorithms)
- Ready to be integrated into end-user software platforms or applications



- **HPC infrastructure** for an efficient execution of the training algorithms which are computationally intensive by making use of heterogeneous hardware in a transparent way
- Seven enhanced **biomedical and AI software platforms** provided by EVERIS, PHILIPS, THALES, UNITO, WINGS, CRS4 and CEA that integrate the DeepHealth libraries to improve their potential
- Proposal for a structure for anonymised and pseudonymised data lakes
- **Validation** in 14 use cases (*Neurological diseases, Tumor detection and early cancer prediction, Digital pathology and automated image annotation*).



DEEPHEALTH

DeepHealth perspective Guiding questions



The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.



Incorporation of HPC in use cases & Field impact

Data / Workflows / HPC-Cloud infrastructure / AI-ML training



DEEPHEALTH

- DeepHealth **incorporates HPC parallelizing the training operations of AI/ML use-cases models** on top of HPC infrastructures using *COMPSS distributed framework (BSC)* and *StreamFlow (UNITO)*
 - Abstract the parallel execution from the underlying infrastructure.
 - Promotes a "**codified approach**" to HPC
- **DATA:** High impact in two dimensions:
 - **HPC/Cloud:** Issue on allowing health data out of health institutions (ethical, privacy and internal and national policies). Anonymized and **pseudonymized data, public data** (and **specific training techniques**) needed to allow exploiting HPC/cloud infra outside health organizations.
 - **AI:** Without **quality and shareable-interoperable** data between partners it is difficult to develop pilot test cases.
- **WORKFLOWS:** important effort in defining efficient pipelines (a.k.a. data-flows) by simply providing in a description file: (1) the URLs of the data sources of each sample or subset of samples, and (2) the computing infrastructure elements; with a twofold reason:
 - to easily manage the data
 - to describe the parallelism exposed by the training operations, with the overall objective of increasing the productivity of computer/data scientist working in any sector and efficiently exploit the underlying HPC infrastructure
 - To promote portability and lock-in avoidance
- **AI/ML and training:** a core objective of the DeepHealth project is the development of a European Deep Learning library able to perform distributed/federated learning on HPC/Cloud infrastructures.





Prioritization of the four fields

In terms of complexity and importance for R&I calls in Europe



DEEPHEALTH

1) Data

2+3) HPC + Workflows

4) AI/ML

- The availability of FAIR **data** is still a big challenge:
 - Difficult to make data providers from the same sector (e.g. Health sector) to collect data following standard protocols (still to be defined in most of cases) to make datasets corresponding to the same disease collected from different hospitals to be interoperable to use them together train AI/ML models.
 - Difficult to access data outside health organizations (limiting the exploitation of available data)
- **HPC & (AI+HPC) workflows** needed to be boosted to increase the productivity of expert-users (data-scientists)
 - Facilitating the definition of AI workflows capable of exploiting the underlying parallel capabilities of the HPC and hybrid cloud-HPC infrastructures.
- **AI/ML**: it is a mature enough research area. But still a long way to go regarding the improvement of model accuracy.





Plans & Specific contributions of DeepHealth partners

Data / Workflows / HPC-Cloud infrastructure / AI-ML training



DEEPHEALTH

- **DATA:** definition of a data-lake structure and organization. Additionally, anonymization procedures are being defined and will be tested in terms of robustness.
 - The exploration of federated/split learning techniques to **avoid the need of moving/centralize the data** and preserve **privacy**
- **AI/ML training:** DeepHealth toolkit including ECV and EDDL libraries ready to run on HPC/Cloud infrastructures in a transparent way for computer/data scientist working in the Health sector, or any other sector
- **HPC+cloud:** Supporting DeepHealth libraries HPC heterogenous & hybrid cloud-HPC computing infrastructure
 - **Heterogenous HPC** computing infrastructure featuring GPUs, FPGAs, and other HW accelerators
- **WORKFLOWS:**
 - **Portability:** Definition of AI+HPC workflows for training & inference operations relying on task-based programming models (COMPSs) and hybrid cloud-HPC cross-application workflows (StreamFlow) capable of efficiently expressing the existing parallelism of AI/ML workflows at different granularities levels
 - **Usability:** Design and development of a toolkit to make it easy the daily work of computer/data scientist working in the Health sector with no deep knowledge of ML and HPC management



Industry in shaping future HPC strategy

Unique HPC needs of industrial partners (IT partners serving the Health industry)

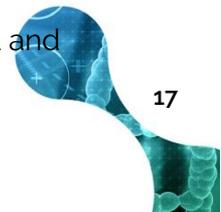
- **Time-to-solution: reducing processing times** for **incorporating AI/ML predictive models to their applications and platforms** to solve health use-cases to support the diagnosis, treatment and monitoring of diseases
- **Easy to use:** if properly engineered (e.g. cloudified), HPC is highly desired to allow easy update of AI/ML models to adapt to new use-cases and improve models fast with new available data

How do you think that industry is engaged to the above-mentioned areas?

- **Expectancy on how they can benefit from HPC** technologies in their AI strategy, applications and services and demanding data, workflows and AI/ML tools
- Most industrial partners have only temporary needs of high processing power (generate the model, update it), thus HPC solutions provided **as a service (e.g. cloudified HPC)**, or low-power (e.g. FPGA-based) inference for embedded systems could be of interest for them

What are your ideas about a commercialization of the product results?

- The **DeepHealth toolkit** is conceived as free and open-source software available on a public repository, with a sustainability plan based on *services and advice to any company or academic institution interested in using any of the software components.*
- **HPC+cloud results**, commercialization exploitation for different results by industrial partners developing FPGA and hybrid cloud solutions, and for non-profit organizations for COMPS and resources managers.





DEEPHEALTH Questions?

Project Coordinator: Mónica Caballero monica.caballero.galeote@everis.com

Technical Manager: Jon Ander Gómez jon@upv.es

HPC Expert: Eduardo Quiñones eduardo.quinones@bsc.es

HPC Expert and Dissemination Manager: Marco Aldinucci aldinuc@di.unit.it

<https://deephealth-project.eu>



@DeepHealthEU

