

An HPC-Driven Data Science Platform to Speed-up Time Series Data Analysis of Patients with the Acute Respiratory Distress Syndrome

C. Barakat, S. Fritsch, M. Riedel, S. Brynjólfsson

2021-09-29



UNIVERSITY OF ICELAND
FACULTY OF INDUSTRIAL ENGINEERING,
MECHANICAL ENGINEERING AND COMPUTER SCIENCE



Status Quo



- Advances in computing and storage technologies:
 - Large scale databases of medical data¹.
 - Availability of data for analysis and information mining^{2,3}.
 - Room for new applications of well-established Machine Learning (ML) and Deep Learning (DL) techniques^{4,5}.
- These advances have highlighted some major shortcomings:
 - Increased need for efficient data storage, versioning, and communication.
 - Increased need for computing power in medical big data analytics.

Available Solutions



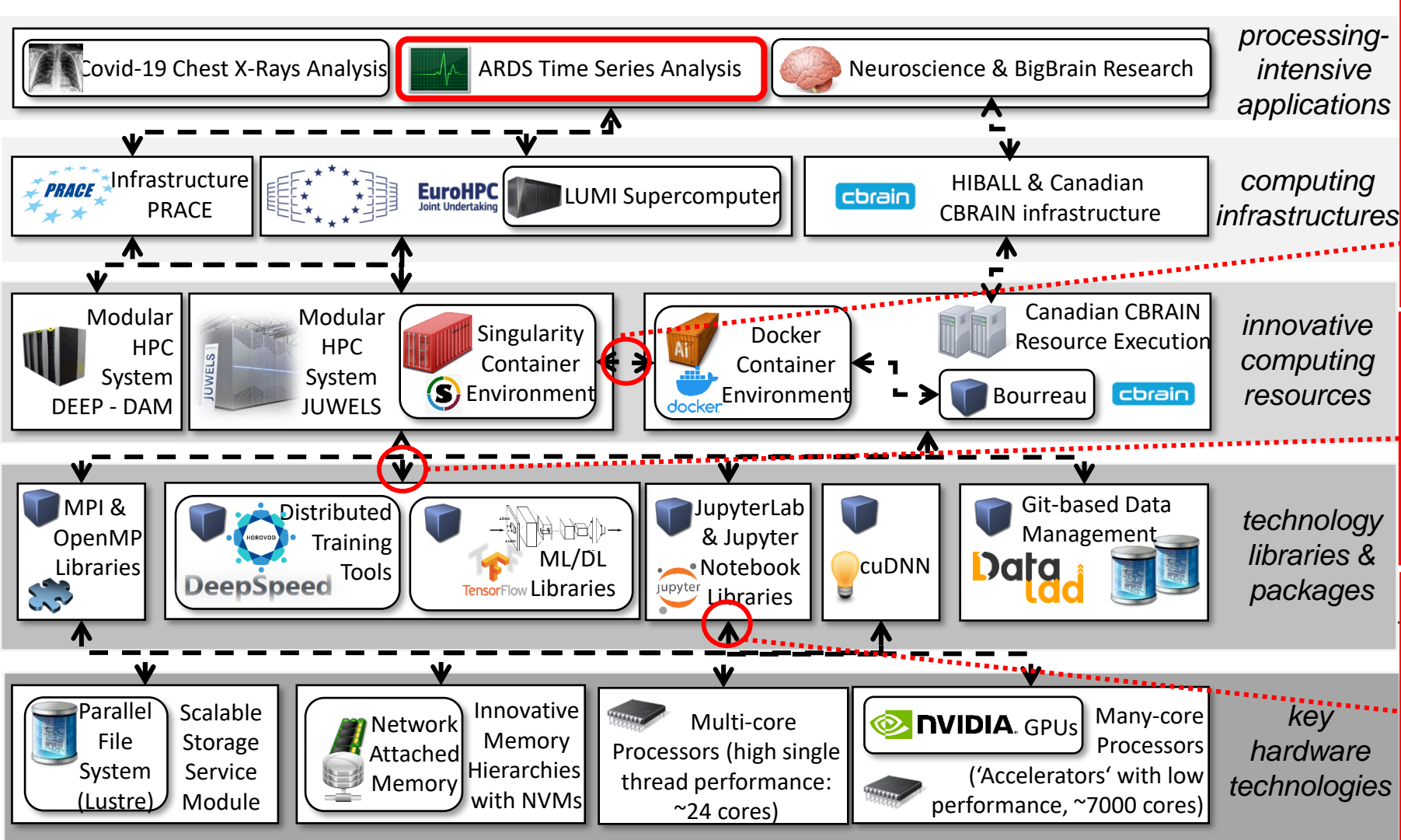
- Making the available technologies more accessible:
 - Medical Institutions gaining access to or setting up their own High Performance Computing (HPC) resources.
 - Cloud computing resources available through private vendors.
 - Medical data made available in open-source repositories for research purposes^{1,2,3,6}.
- However:
 - Medical staff can not be expected to become acquainted with HPC.
 - Large applications cripple the hardware of medical institutions.

Proposed Solution



- Establish an HPC-enabled Platform:
 - Bridges the gap between research institutions and medical centres.
 - Makes use of high efficiency data storage, analysis, and transfer architecture.
 - Performs necessary computationally-intensive tasks on specialized hardware.
 - Keeps technical complexity at a low degree.
- We highlight an initial use case concerning missing value prediction in medical data, to assist in the detection of onset of Acute Respiratory Distress Syndrome (ARDS).

Platform Blueprint



processing-intensive applications

computing infrastructures

innovative computing resources

technology libraries & packages

key hardware technologies

Neuroscience Image Analysis

```

Some preparation
$ mkdir winterschool winterschool_cache winterschool_tmp
$ chmod +w winterschool_cache
$ export SINGULARITY_CACHEDIR=$(mktemp -d -p "$(pwd)/winterschool_cache")
$ export SINGULARITY_TMPDIR=$(mktemp -d -p "$(pwd)/winterschool_tmp")

Pull the docker image:
$ cd winterschool
$ singularity pull hws.sif docker://glatard/hws

Step into the container
$ singularity shell ./hws.sif
(the prompt changes to `>Singularity`)

download a dataset:
$ git config --global user.name "Your name"
$ git config --global user.email "peturhelgi@gmail.com"

Singularity> datalad install https://github.com/CONP-PCNO/conp-dataset.git
    
```

DataLad

cbrain

ARDS Time Series Analysis

Jupyter

```

GRU_Layer_1_input: InputLayer
input: [(?, 1000, 6)]
output: [(?, 1000, 6)]

GRU_Layer_1: GRU
input: (?, 1000, 6)
output: (?, 1000, 32)

GRU_Layer_2: GRU
input: (?, 1000, 32)
output: (?, 32)

Output_Layer: Dense
input: (?, 32)
output: (?, 1)
    
```

Covid-19 Chest X-Ray Analysis

Covid-Net

Jupyter

Covid-X Dataset

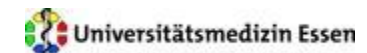
```

#!/bin/bash
# Load required modules
module purge
module use $OTHERSTAGES
module load Stages/2020
module load GCCcore/.9.3.0
module load Python/3.8.5
module load TensorFlow/2.3.1-Python-3.8.5
module load OpenCV/4.5.0-Python-3.8.5
# Activate Python virtual environment
source /p/project/training2104/ingolfsson1/jupyter/kernels/ingolfsson1_kernel/bin/activate
# Ensure python packages installed in the virtual environment are always preferred
export PYTHONPATH=/p/project/training2104/ingolfsson1/jupyter/kernels/ingolfsson1_kernel/lib
exec python -m ipykernel $@
    
```

SMITH and ARDS



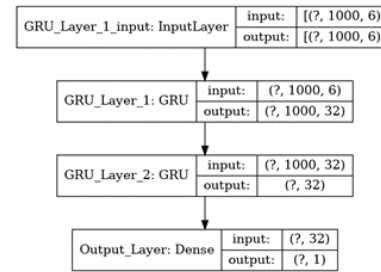
- Acute Respiratory Distress Syndrome (ARDS) is a rare condition that affects ICU patients, but has a high mortality rate^{7,8}.
- There is consensus on how to diagnose the condition, but not how to treat it⁹.
- This is one of the use cases of the Smart Medical Information Technology for Healthcare (SMITH) consortium grouping major research institution in Germany¹⁰.
- The aim is to develop algorithms that can efficiently and accurately diagnose the onset of ARDS, and potentially provide suggestions for treatment.



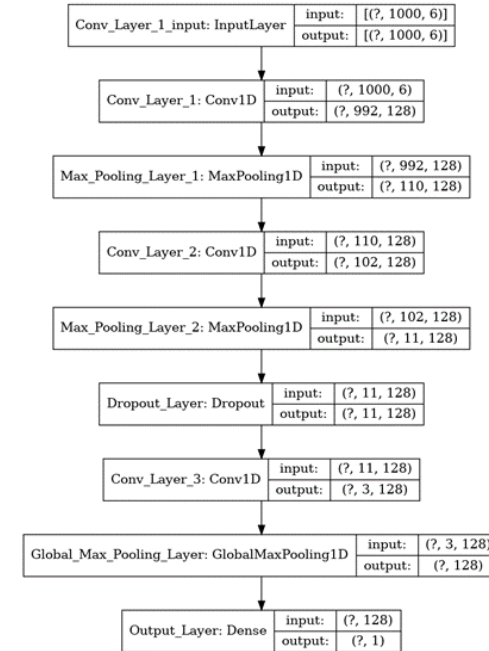
Applying the Platform



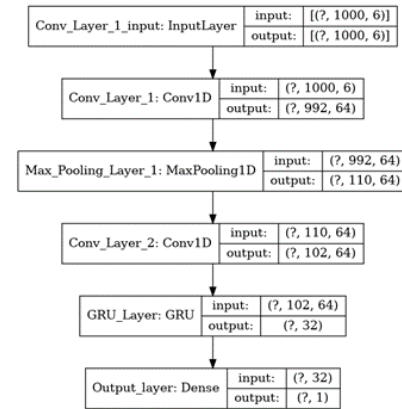
- Perform data cleaning and preparation using the established platform.
- Set up sequence analysis models using modules available through Keras and Tensorflow.
- Train a sequence prediction model on the cleaned time series data.
- Training results show promise in terms of speed-up and reduced prediction errors (compared to other methods)



a) GRU Model Structure



b) 1D-Conv Model Structure



c) Mixed Model Structure

Conclusions



- ARDS time series prediction is one case study from the platform.
- The goal is to introduce HPC into healthcare applications, without the complexities that usually come with it.
- Currently setting up the frameworks for two further medical applications
- Also improving the use cases currently in place.

Bibliography



1. A. E. Johnson et al., "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, no. 160035, May 2016.
2. <https://aimi.stanford.edu/research/medical-imagenet>
3. V. Huddar, B. K. Desiraju, V. Rajan, S. Bhattacharya, S. Roy and C. K. Reddy, "Predicting Complications in Critical Care Using Heterogeneous Clinical Data," *IEEE Access*, vol. 4, pp. 7988-8001, October 2016.
4. H. Sun, Z. Liu, G. Wang, W. Lian, and J. Ma, "Intelligent analysis of medical big data based on deep learning," *IEEE Access*, vol. 7, pp. 142022–142037, 2019.
5. W. Wang et al., "Can computer simulators accurately represent the pathophysiology of individual COPD patients?," *Intensive Care Medicine Experimental*, vol. 2, no. 23, 2014.
6. N. S. Punn and S. Agarwal, "Automated diagnosis of covid-19 with limited posteroanterior chest x-ray images using fine-tuned deep neural networks," *Applied Intelligence*, 2020.
7. D. G. Ashbaugh, D. B. Bigelow, T. sL. Petty, and B. E. Levine, "Acute respiratory distress in adults," *The Lancet*, vol. 290, no. 7511, pp. 319 – 323, 1967, Originally published as Volume 2, Issue 7511.
8. J. Villar et al., "The ALIEN study: incidence and outcome of acute respiratory distress syndrome in the era of lung protective ventilation," *Intensive Care Medicine*, vol. 37, no. 12, pp. 1932–1941, December 2011.
9. The ARDS Definition Task Force, "Acute Respiratory Distress Syndrome: The Berlin Definition of ARDS," *JAMA*, vol. 307, no. 23, pp. 2526–2533, June 2012.
10. A. Winter et al., "Smart medical information technology for healthcare (SMITH)," *Methods Inf Med*, vol. 27, no. 1, 2018.