Bardia Mojra – 1000766739

02/28/2022 – Assignment 02 – Application 01

Supercomputer #4 – JUWELS Booster Module Supercomputer

An HPC-Driven Data Science Platform to Speed-up Time Series Data Analysis of Patients

with the Acute Respiratory Distress Syndrome

In recent years, the use of large and readily available datasets and cutting-edge machine learning (ML) methods have become popular among medical researchers as it has shown to result in increased accuracy and overall treatment of the patient. More importantly, recent advents in deep learning (DL), biomedical engineering, and High-Performance Computing (HPC) have allowed researchers to train DL models capable of learning the dynamics of large multivariate time series medical datasets. The deployment of Recurrent Neural Networks has been shown to predict patient conditions with relatively high accuracy. This is particularly helpful in the analysis, diagnosis, and treatment of patients with diseases related to the respiratory system e.g. Acute Respiratory Distress Syndrome (ARDS).

Multiple solutions have been proposed; but, as the authors point out, using such tools and datasets is difficult for medical professionals for multiple reasons. First, although there is an abundance of publicly available medical data, these datasets are scattered across the web and are often stored in different Electronic Health Records (EHR) formats. Second, deploying ML and DL analysis methods and prediction tools on large datasets requires expert knowledge of the field and programming skills that medical professionals are not widely trained on. Third, the growing amount of data available through EHRs requires an HPC-enabled application that is both scalable over time and seamlessly integrates into the ICU environment and provides patient-specific recommendations.

With a 40% mortality rate, ARDS is a life-threatening medical condition that on average involves 12% of mechanically ventilated (MV) ICU patients. Per the Berlin definition, the onset of ARDS is defined as a prolonged ratio of arterial oxygen potential to fraction of inspired oxygen (P/F ratio) of 300 mmHg or less. Multiple papers have pointed out the correlation between early detection of the onset of ARDS and the patient’s rate of survival. Moreover, several papers have demonstrated the effectiveness of MV protocols, e.g. “low tidal volume” and “high Peak End-Expiratory Pressure (PEEP)”, for ICU patients suffering from ARDS. However, practical use of such techniques requires a well-integrated system within the hospital ICU that can analyze patient data and make inferences on the onset of ARDS. To address this, the authors propose an HPC-based application that is tailored to support clinical researchers in understanding ARDS, and in the future, it could potentially be deployed to real ICUs and clinical research i.e. German Smart Medical Information Technology for Healthcare (SMITH) project.

This application is designed to trains its DL model on two HPC systems, the DEEP prototype, or the JUWELS system for increased computational loads. The JUWELS Booster Module (subject of this assignment) is the world’s 8th highest-performing supercomputer, per the TOP500 list. The system is built in Germany by Atos and utilizes an AMD EPYC processor and NVIDIA A100 for accelerated GPU operations. The system is the most powerful in Europe with 44.1 Pflop/s.

It is important to note that the trained model could be downloaded with a much smaller size and stored locally to make inferences on the onset of ARDS in ICU patients. This is an important factor as it keeps the patient training data in a secure location by the host HPC system which avoids unnecessary distribution. Moreover, the model could be used for inference and diagnosis at local hospitals and clinics without reliance on the cloud which could not be reached if communication deteriorates. Thus, they define two main activities or applications *ARDS Time Series Analysis and Model Training* and *ARDS Time Series Analysis and Model Inference*.

This cloud-based ARDS early diagnosis application uses DataLad, a git-based data management system that features transparent and traceable access to patients' datasets on the premise of the hospital. For development purposes, the authors used the Medical Information Mart for Intensive Care - III (MIMIC-III) database for training their prediction model and stored in the Scalable Storage Service Module (SSSM) and could be used by researchers to train new models without the need to download the database. Their platform supports an array of local and cloud-based deep learning training tools, e.g. DeepSpeed and Jupiter Notebook, that could be used by researchers to develop more innovative models or run GUI-based analysis on a local machine.

The authors trained their model on datasets collected from nearly 20,000 ICU patients who received MV during their stay. First, they make predictions for some missing values, and in some cases with many data points missing they completely drop the feature. They discovered the following six medical metrics are well preserved, Respiratory Rate (RR), Heart Rate (HR), Systolic Arterial Pressure (SAP), Diastolic Arterial Pressure (DAP), Mean Arterial Pressure (MAP), and Blood Oxygen Saturation (SpO2). As mentioned earlier, the diagnosis of a patient depends on the ratio of arterial oxygen potential (PaO2) to the Fraction of Inspired Oxygen (FiO2). The FiO2 is known since it is set on the ventilator and it is recorded automatically and the goal is to predict the value of SpO2 which is directly related to PaO2.

|  |
| --- |
| Figure 1 |

The author deployed three learning and prediction approaches to study the performance of the proposed HPC-based ARDS prediction system. First, they used the Gated Recurrent Unit (GRU) model which consists of two GPU layers with 32 units each, followed by a dense layer. The hyperparameters that define the structure of the model were chosen empirically and the authors point out the crucial role HPC resources played. The system is able to train a total of 10,209 parameters with Mean Absolute Error (MAE) converging to 0.7432 after about 15 epochs, all under 405 seconds. The second method deploys a One-Dimensional Convolutional model that consists of three convolution layers with 128 filters each, a stride of 9, and a 1D max-pooling layer after each layer except the last which has a 1D global max-pooling layer. This model has a total of 302,337 parameters and the MAE stabilizes at 0.725 after only 40 seconds of training. The third approach uses a mixed model with two 1-D convolution layers with 64 filters each, followed by a GRU layer with 32 units, followed by a dense output layer. This model has a total of 49,889 trainable parameters with stabilizing around similar MAE losses in 45 seconds. Figure 1 shows the performance for all three models. The authors highlight the benefits of their mixed model which combines quick training time and high accuracy shown by the other two models.

The authors point out the quick training time provided by the DEEP and JUWELS HPC systems and the role they played in finding the optimal model structure in all three cases. Furthermore, they have made their HPC-based platform available to the public and researchers can train new models for ARDS early detection applications in seconds using only a few lines of code. Moreover, they show the promising results 1D convolution models produce and outline how such a system could be one day integrated into the ICU environment seamlessly.

Corresponding Authors:

[c.barakat@fz-juelich.de](mailto:c.barakat@fz-juelich.de) (C. Barakat), [sfritsch@ukaachen.de](mailto:sfritsch@ukaachen.de) (S. Fritsch), [morris@hi.is](mailto:morris@hi.is) (M. Riedel), [sb@hi.is](mailto:sb@hi.is) (S. Brynjólfsson)

References:

1. Baraka, Fritsch, Riedel, and Brynjolfsson, “An HPC-Driven Data Science Platform to Speed-up Time Series Data Analysis of Patients with the Acute Respiratory Distress Syndrome,” 44th International Convention on Information, Communication and Electronic Technology (MIPRO), 2021, PP. 311–316, IEEE.