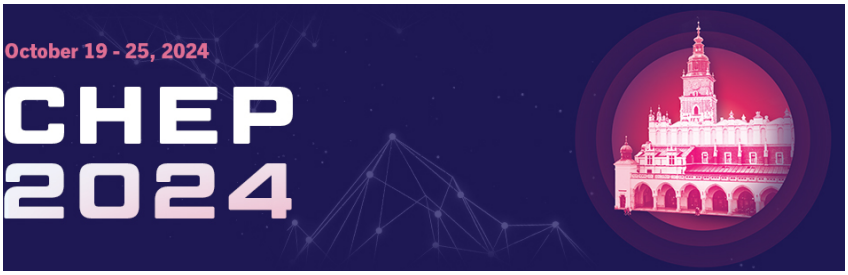

Contribution ID: **268** Contribution code: **TUE 12** Type: **Poster**

Decode the Workload: Training Deep Learning Models for Efficient Compute Cluster Representation

Tuesday 22 October 2024 16:00 (15 minutes)

Monitoring the status of a high throughput computing cluster running computationally intensive production jobs is a crucial yet challenging system administration task due to the complexity of such systems. To this end, we train autoencoders using the Linux kernel CPU metrics of the cluster. Additionally, we explore assisting these models with graph neural networks to share information across threads within a compute node. The models are compared in terms of their ability to: 1) Produce a compressed latent representation that captures the salient features of the input, 2) Detect anomalous activity, and 3) Make distinction between different kinds of jobs run at Jefferson Lab. The goal is to have a robust encoder whose compressed embeddings are used for several downstream tasks. We extend this study further by deploying these models in a human-in-the-loop production-based setting for the anomaly detection task and discuss the associated implementation aspects such as continual learning and the criterion to generate alarms. This study represents a first step in the endeavor towards building self-supervised large-scale foundation models for computing centers.

Primary author: Dr MOHAMMED, Ahmed (Thomas Jefferson National Accelerator Facility)

Co-authors: Mr HESS, Bryan (Thomas Jefferson National Accelerator Facility); Mrs MCSPADDEN, Diana (Thomas Jefferson National Accelerator Facility); RAJPUT, Kishansingh (Thomas Jefferson National Accelerator Facility); Ms HILD, Laura (Thomas Jefferson National Accelerator Facility); Dr SCHRAM, Malachi (Thomas Jefferson National Accelerator Facility); Mr JONES, Mark (Thomas Jefferson National Accelerator Facility); Mr MOORE, Wesley (Thomas Jefferson National Accelerator Facility); Dr DAI, Zhenyu (Thomas Jefferson National Accelerator Facility)

Presenter: JESKE, Torri

Session Classification: Poster session

Track Classification: Track 7 - Computing Infrastructure