ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Hyperparameter Optimization for Deep Learning Model Using High Performance Computing

Speaker : Muhammad Numan Anwar

PhD Student at POLIBA and INFN, Bari

On behalf of the FCC Collaborations at ICSC and Spoke2, 10-12 dec 2024, Physics Dept and INFN, Catania

INFN
Istituto Nazionale di Fisica Nucleare

de' remi facemmo ali 1990 POLITECNICO DI BARI
Politecnico di Bari

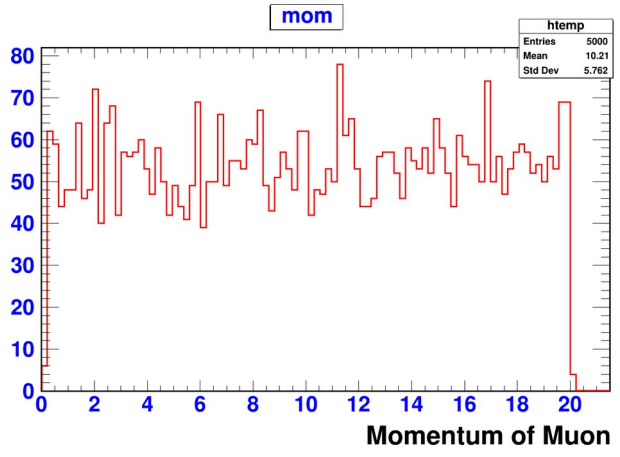
Outline

- Simulation based on Garfield ++ for 2024 real test beam data
- Training of Long Short Term Memory (LSTM) Model Using HPC Resources for Peak Finding Algorithm
- Training of Convolutional Neural Network(CNN) Model Using HPC Resources for Clusterization Algorithm
- Preliminary Results related to GPU's and Future Planning

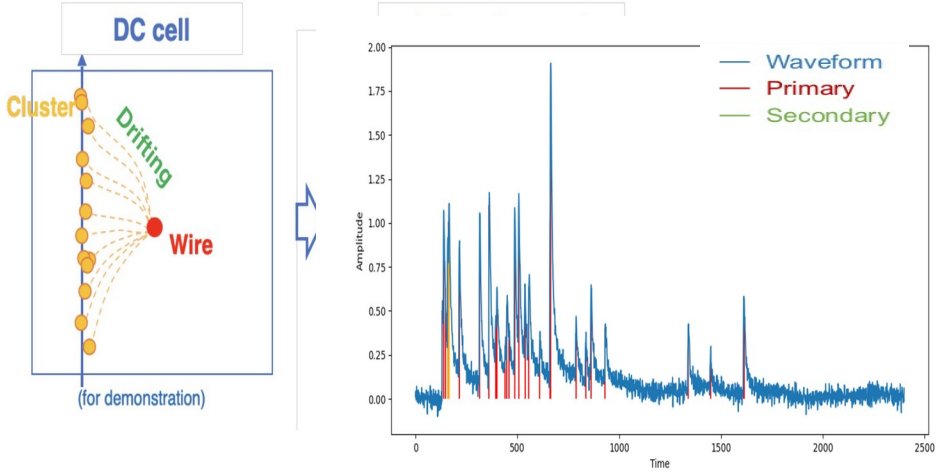
Main Goal of the Talk

- The main goal of the talk is to train neural network models, such as the Long Short-Term Memory (LSTM) Model and Convolutional Neural Network (CNN) Model on the mometa ranges from 0.2 to 20 GeV/C and then we applied this trained model on the sample of 2, 4, 6, 8, and 10 GeV/c momenta as testing to check the performance of the models. These models are trained for a two-step reconstruction algorithm, which involves peak finding and clusterization
- So, I designed a task involving the simultaneous submission of several jobs using local HPC Resources. The purpose of this task is to train Long Short-Term Memory (LSTM) models to classify signals from background, a process known as a classification task. To achieve this task, I utilized various hyperparameters, including activation functions, optimizer, Epochs, batch size, patience, and dropout rates etc . Additionally, I managed different HPC resources such as memory requests, Job duration, and CPU Usage etc
- For the peak finding algorithm, I selected best trained LSTM model based on the higest area under the curve value among all configurations which is further used to discriminate signals (primary and secondary peaks) from the noise in the waveform, addressing a classification problem. SGD as an optimizer, relu and sigmoid as an activation functions, 32 neurons in LSTM layer, 32 neurons in the dense layer and 1 neuron in the output layer, 200 epochs and 250 batch size were selected for the best LSTM peak finding model
- Concurrently, I used the above logics again to select best CNN model based on the lowest mean square error value among all configurations. Convolutional Neural Network model is utilized to determine the number of primary clusters based on the detected peaks, dealing with a regression problem. rmsprop as an optimizer, selu and selu as an activation functions, 32 and 64 filters in two convolutional layers, 32 neurons in dense layer and 1 neuron in output layer, 50 epochs and 150 batch size were selected for the best CNN regression model
- It should be noted that the best trained models (LSTM and CNN) are applied to simulations based on Garfield++ for 10, 8, 6 4, and 2 GeV/c momenta and the all those other parameters which were used for the 2024 test beam data.
- Follow the below link for the details

https://agenda.infn.it/event/43914/contributions/246942/attachments/127268/188223/Cluster%20counting.pdf

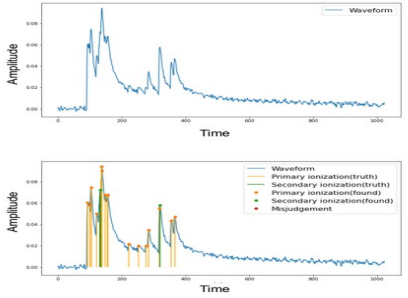


Cluster Counting in Drift Chambers and Simulation Parameters based on Garfield ++



- **A muon particle is passed through mixture of gases (90% He and 10% C_4H_{10}) generate electron-ion pairs causing a read out signal (induce current). The simulation package creates analog induced current waveforms from ionizations (HEED). The digitization package incorporates electronics responses taken from experimental measurements and generates realistic digital waveforms**

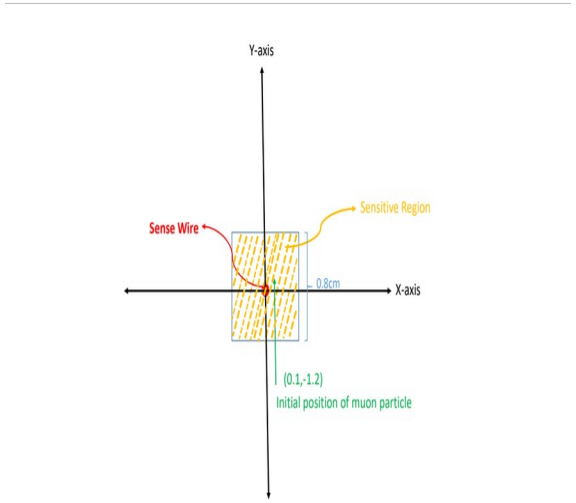
Two Step Reconstruction Algorithm:

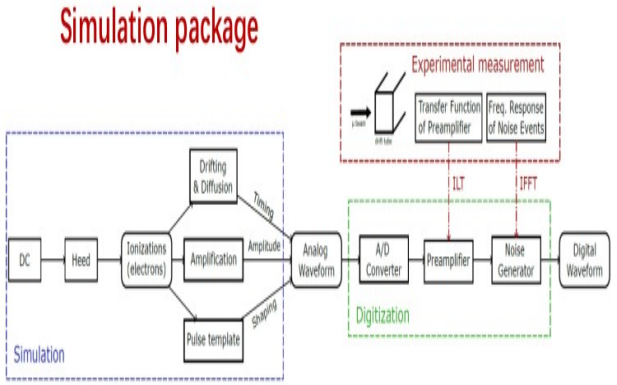


Step1. Peak Finding
Discriminate peaks (both primary and secondary) from the noises (classification problem)

Step2. Clusterization:
Determine the number of clusters (N_{cls}) from the detected peaks (regression problem)

DC Layout



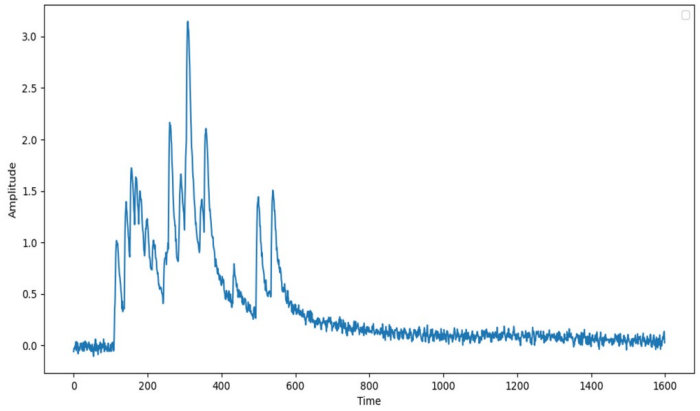


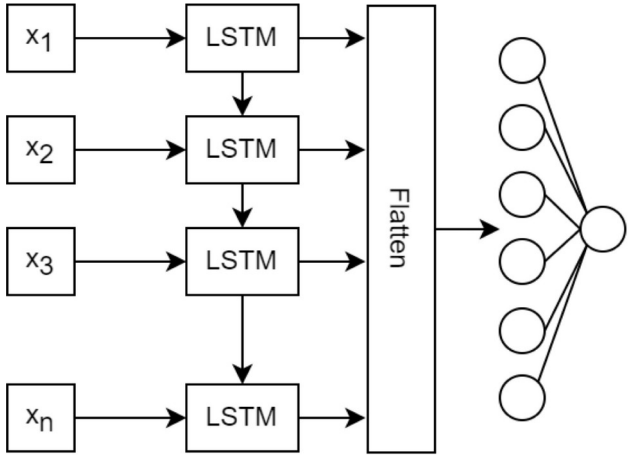
Sampling Rate	2 GHz
Gas Mixture	He (90%) & C_4H_{10} (10%)
Cell Size	0.8 cm
Momentum (GeV/C)	10, 8, 6, 4, 2
Angle between the z axis of drift tube chamber and track of the muon particle	45 degree
Particle	Muon

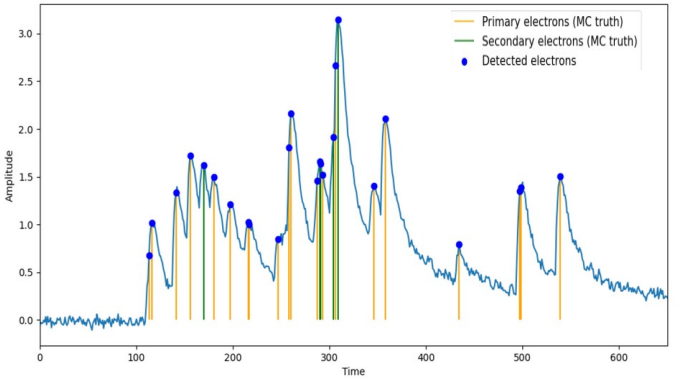
- **All the simulation parameters are shown in the table above**

Two-Step Reconstruction Algorithm

Step1: Peak Finding Algorithm



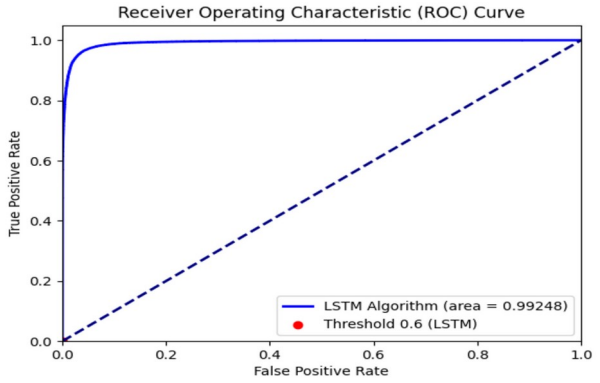




- The task of peak finding can be framed as a classification problem in machine learning
- The waveforms are divided into segments, each comprising 15 bins. Each segment can represent either a signal or a nois
- The list of the amplitudes of a segment, subtracted by their mean and normalized by their standard deviation, is served as the input feature for the neural network
- The data of waveform is time sequence data, which suitable for especially Long Short Term Memory Model

- We applied a Long Short-Term Memory (LSTM) model to the waveform to classify signals (primary and secondary electrons) from the Noise using a peak-finding algorithm known as classification
- Detected peaks from both primary and secondary electrons are shown by blue dots

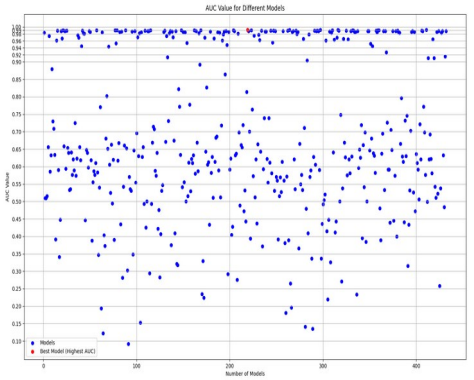
Performance of the Best LSTM Model and HPC Resources



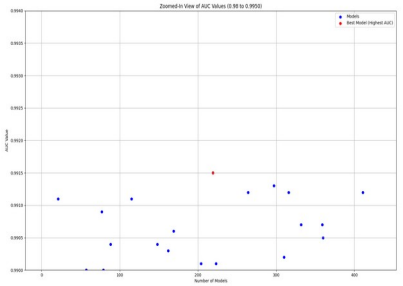
- The above plot show ROC curve for the best LSTM model with Area under the curve value 0.992 which show a best classification to discrimate sugnal from background

		Prediction	
		Sig	Noise
Truth	Sig	TP	FN
	Noise	FP	TN

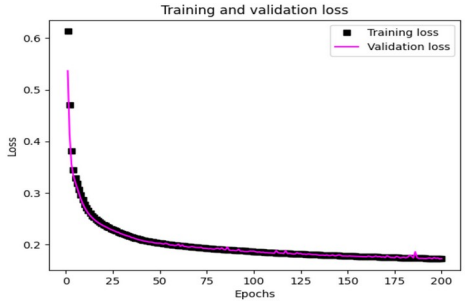
- The above table tell us about the concept of classification (TP, TN) and misclassification (FP, FN)



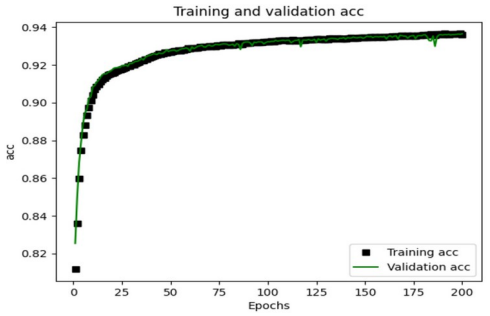
- **The above plot shows us different configuration models with Area under the curve value.The red dot shows us the best model among all**



- **Zoom region to specify the best model among all configurations clearly**



- **The above plot loss VS epoch show us that the training and validation loss decreases over the epochs and then it become approximately constant which shows a best trained model**

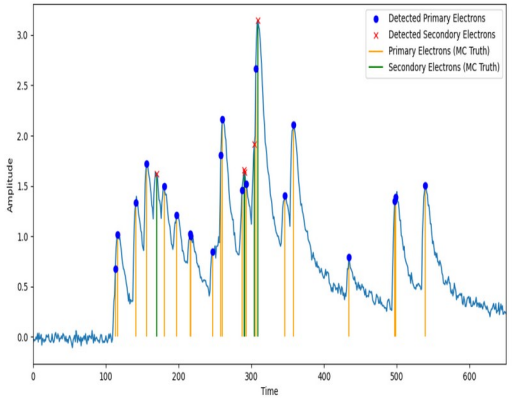


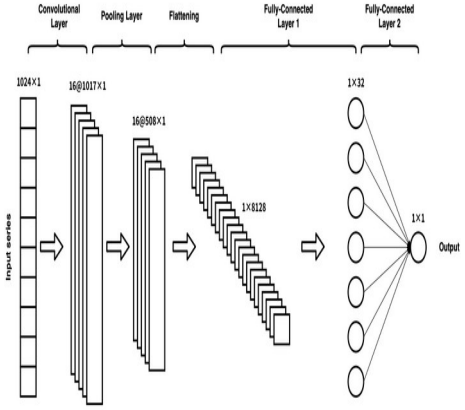
- **The above plot Accuracy VS Epoch show us that the training and validation accuracy increases over the epochs and then it become approximately constant which shows a best trained model**

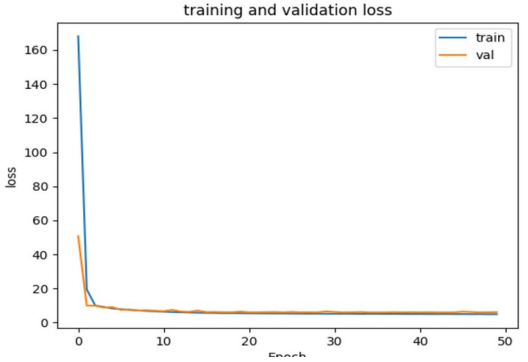
Partionable Resources	Usage
CPUS	4
Memory (MB)	108.5
Run Remote Usage	3847.28 sec

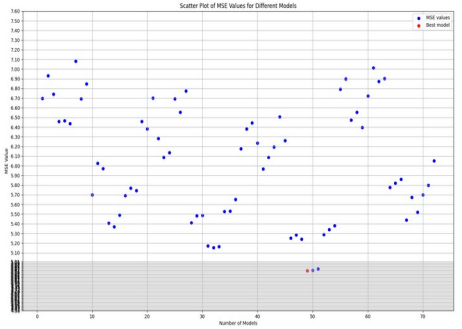
HPC Resources are shown in the table above

Step2: Clusterization and Performance of Best CNN Model









- **A regression problem to predict Number of primary clusters based on the primary detected peaks by using Convolutional Neural Network (CNN) model**
- **The peaks found by peak finding algorithm would be training sample of this algorithm**

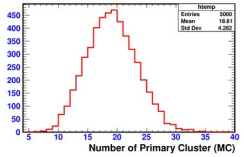
- **Labels: Number of clusters from MC truth**
- **Features: Time list of the detected times in the previous step encoding in an (1024, 1) array.**
- **A regression problem**

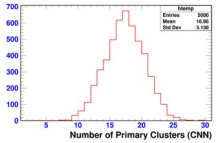
The above plots show us that theresult of lowest mean square error value (4.918) and also showed by the red dot among all the configurations

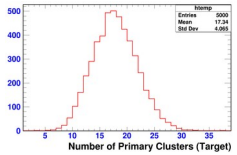
Partionable Resources	Usage
CPUS	4
Memory (MB)	93
Run Remote Usage	1561 sec

HPC Resources are shown in the table above

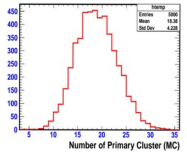
Final Results of the Reconstruction for Different Momenta_

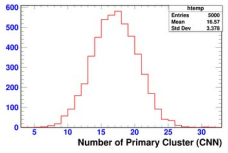


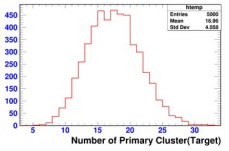




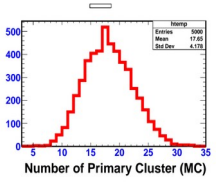
- The above distrubitions shows us the number of primary cluster (MC), number of primary clusters detected by CNN and Target(LSTM) for 10 GeV Momentum

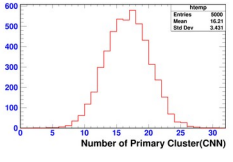


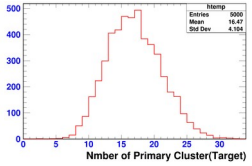




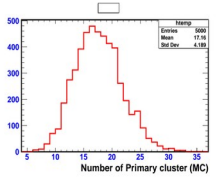
- The above distrubitions shows us the number of primary cluster (MC), number of primary clusters detected by CNN and Target(LSTM) for 8 GeV Momentum

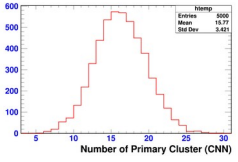


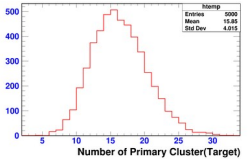




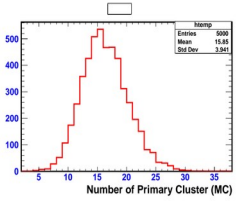
- The above distrubitions shows us the number of primary cluster (MC), number of primary clusters detected by CNN and Target(LSTM) for 6 GeV Momentum

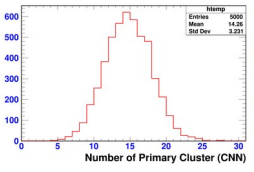


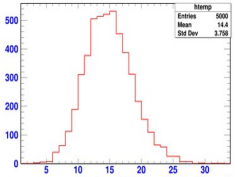




- The above distrubitions shows us the number of primary cluster (MC), number of primary clusters detected by CNN and Target(LSTM) for 4 GeV Momentum







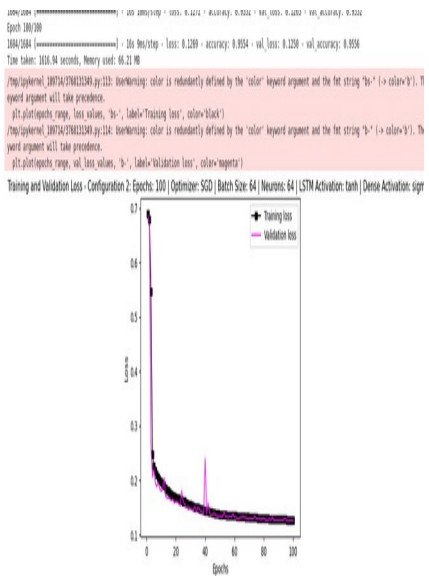
- The above distrubitions shows us the number of primary cluster (MC), number of primary clusters detected by CNN and Target(LSTM) for 2 GeV Momentum

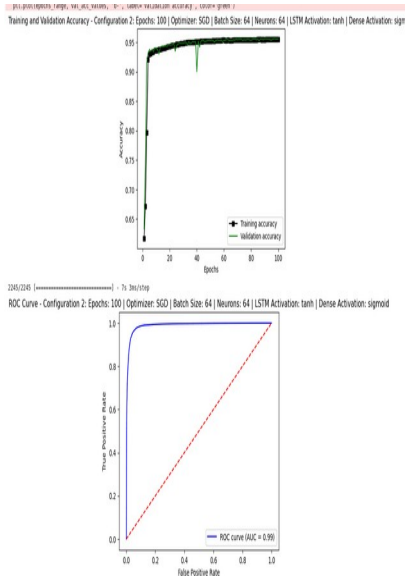
Momentum of Muon	Primary Cluster(MC)	Standard Deviation (MC)	Cluster Size (Full Range)	Primary Cluster(LSTM	Standard Deviation (LSTM)	Primary Cluster (CNN)	Standard Deviation (CNN)
2 Gev/c	15.85	3.9	1.55	14.4	3.75	14.26	3.2
4 GeV/c	17.16	4.189	1.54	15.85	4.015	15.77	3.42
6 GeV/c	17.65	4.178	1.605	16.47	4.104	16.21	3.43
8 GeV/c	18.38	4.228	1.54	16.96	4.05	16.57	3.37
10 Gev/c	18.61	4.282	1.54	17.34	4.065	16.86	3.13

- **The above table show us different number of primary clusters (MC), cluster size, number of primay cluster (Target) and primary cluster detected by CNN with standard Deviations for different momenta are shown in the table**

Preliminary Results related to GPU's

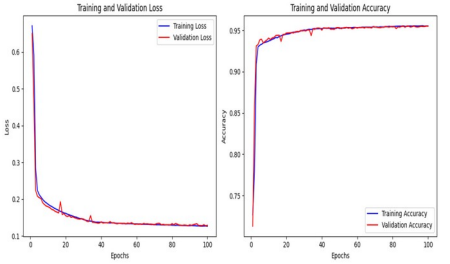
Best configuration among all to train LSTM Model by Using Two GPU's

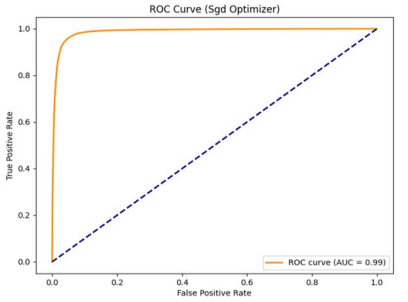




Partionable Resources	Usage
GPUS	2.0
Memory (MB)	66.21
Run Remote Usage	1616.94 sec

- **Best plots (Accuracy & Loss VS epochs and ROC curve) among configurations are shown by using 2 GPU's, training time taken for this job, and memory usage are shown in the screenshots**
- **It is faster than CPU's because I used exactly the same hyperparameters for LSTM model in both cases**
- **The above table shows us the HPC recas resources related to GPU's**





- **Plots (Accuracy & Loss VS epochs and ROC curve) are shown by using 4 CPU's, training time taken for this job, and memory usage are shown in the screenshots**
- **It is slower than GPU's because I used exactly the same hyperparameters for LSTM model in both cases**

Partionable Resources	Usage
CPUS	4.0
Memory (MB)	98.50
Run Remote Usage	2847.28

- **The above table shows us the HPC recas resources related to CPU's**

Future Planing

 Then, I would apply best Long Short Term Memory (LSTM) model to the waveform to classify the signal from the background, a process known as classification. This algorithm is referred to as the peak finding algorithm.

 I would repeat the same logic to select the best Convolutional Neural Networks (CNN) model based on the lowest mean square error value among the all configurations by Using HTCondor to submit GPU jobs

 Similarly, I will apply the best Convolutional Neural Networks (CNN) model to the detected peaks (primary and secondary) of 2, 4, 6, 8, and 10 GeV testing samples to identify the number of primary clusters and evaluate the model's performance

Thank you