Exploring Machine Learning Models To Improve The Classification Of Displaced Hadronic Jets In The ATLAS Calorimeter

March 31st, 2020

Rodrigue de Schaetzen

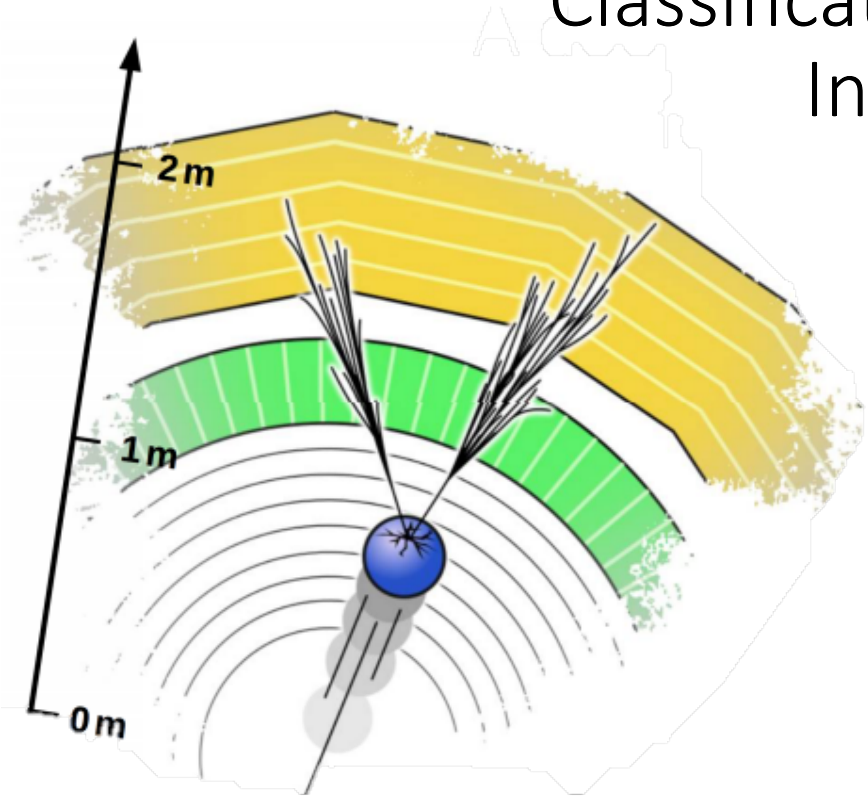


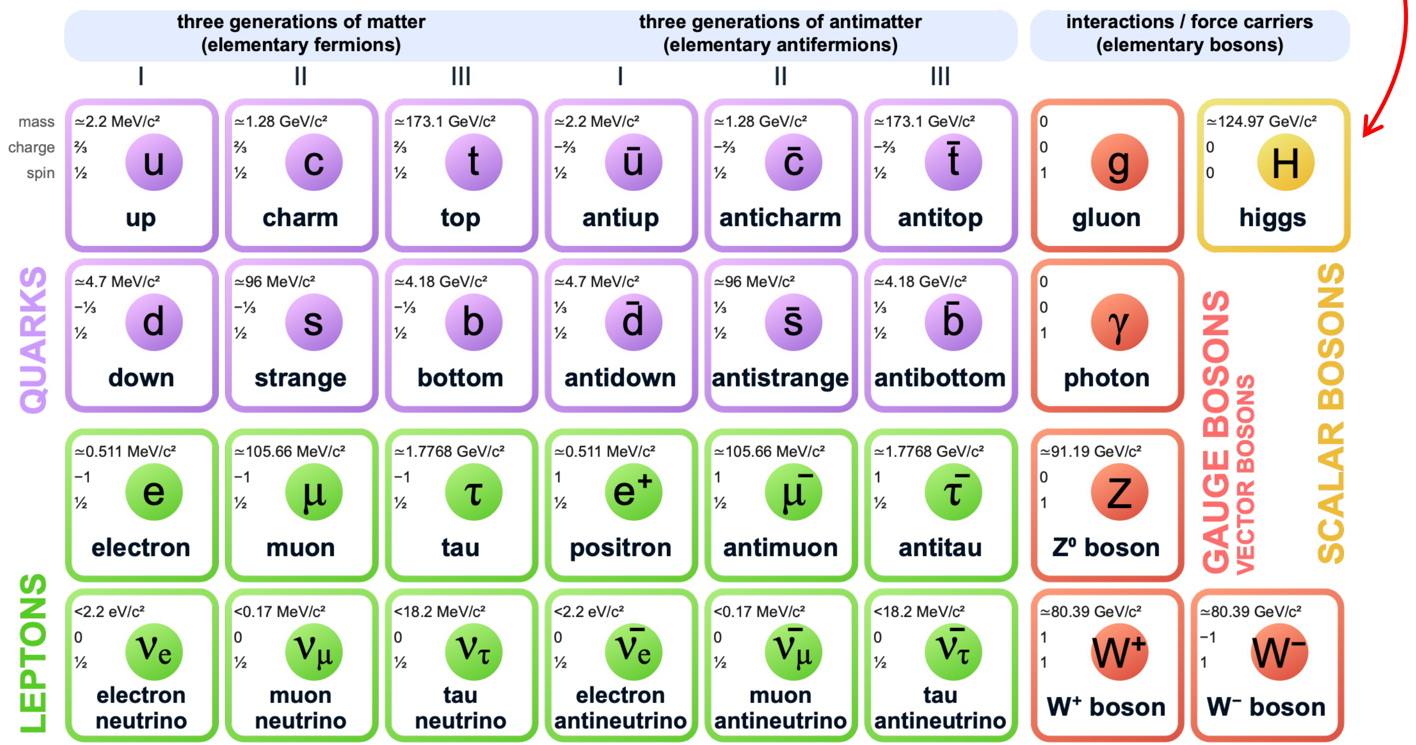
Image from CMS Collaboration

Agenda

1. Introduction and Motivation
2. Machine Learning
3. Study #1: Does p_T ordering matter
4. Study #2: Modifying network architecture
5. Study #3: Optimizing hyperparameters
6. Conclusions and Discussions

Introduction and Motivation

Confirmed in 2012!

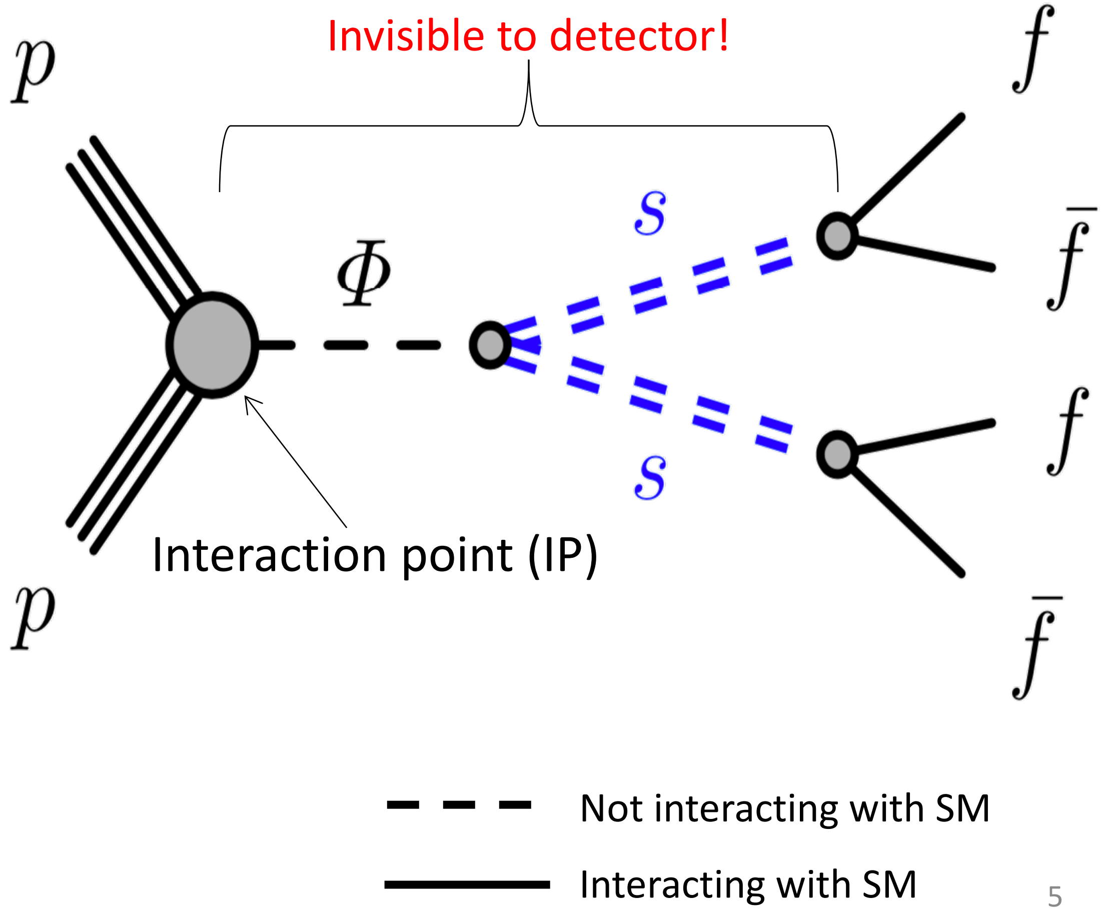


Standard Model (SM)

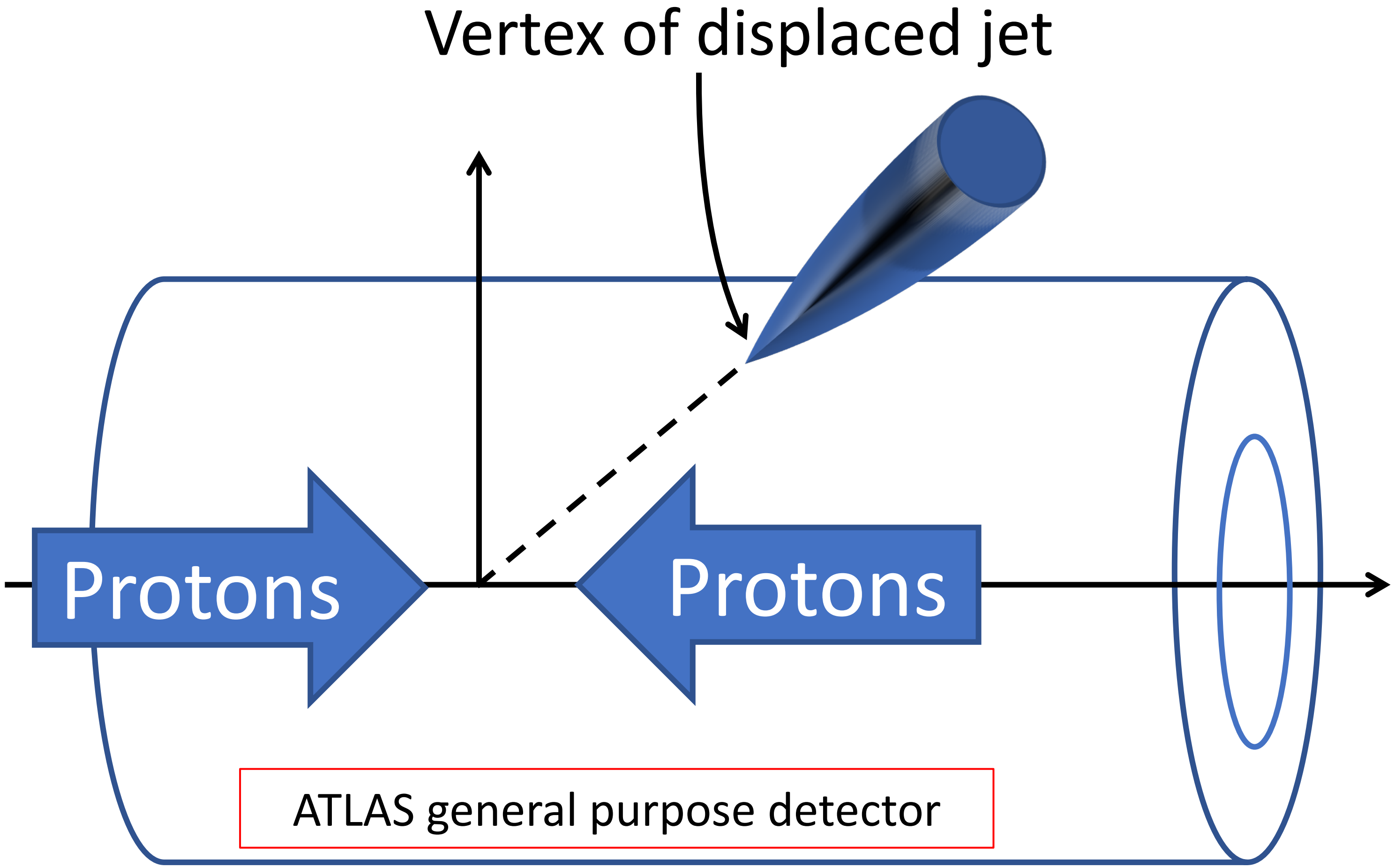
- Current understanding of the fundamental particles and their interactions
- Cannot answer many large questions e.g. what is Dark Matter made of?
- Big search for extensions to the SM

Long-lived particle search

- Long-lived particles (LLPs) occur in many extensions to the SM e.g. the Hidden Sector
- **Theorized model:**
    - Two protons (*p*) collide to form a heavy boson (*Φ*)
    - Boson decays to two long-lived particles (*s*) which decay to two fermion-antifermion pairs (*f*)
    - Model serves as a benchmark for searching paired LLPs



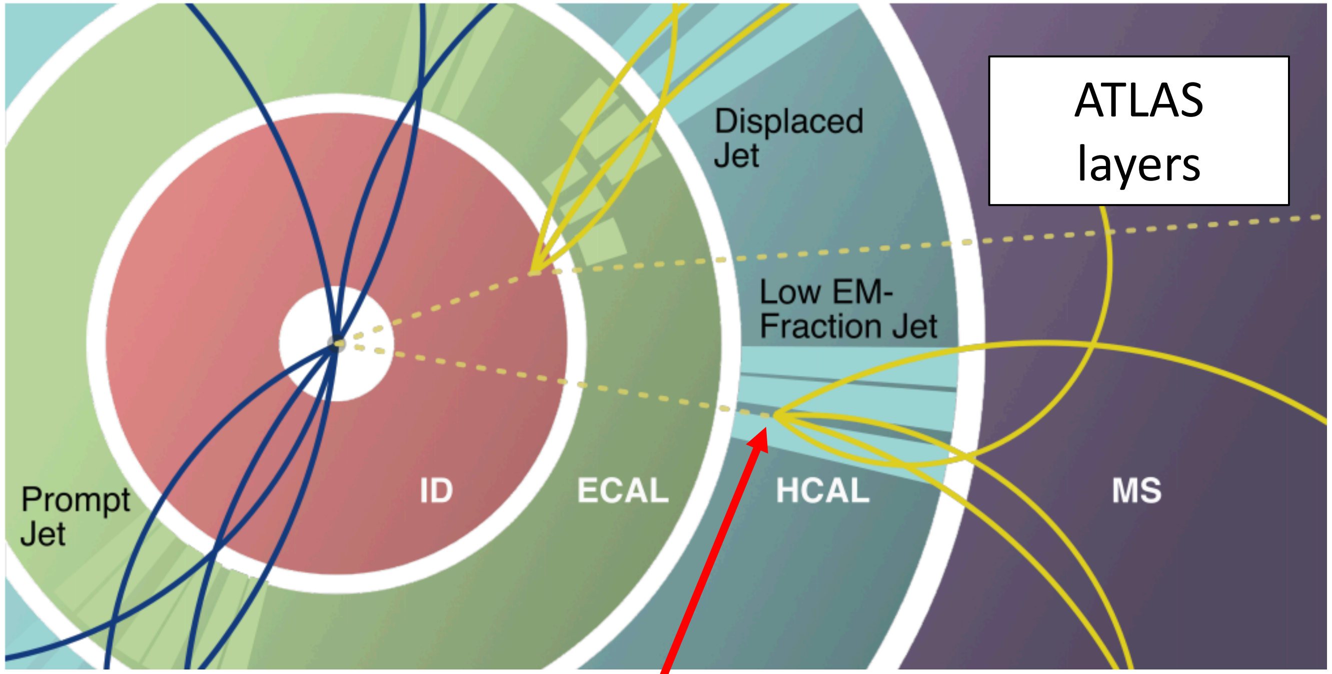
Long-lived particle search



- This search uses data from ATLAS detector at the LHC
- Jets are bunches of particles that can be clustered in a single object
- Each LLP decaying to a fermion-antifermion pair will produce a displaced jet

Signal jet

- Signature generated by LLPs is the signal
- Characteristics of a signal jet:
  - No tracks
  - Displaced jet
  - High ratio of energy deposited in the HCal to energy deposited in the ECal

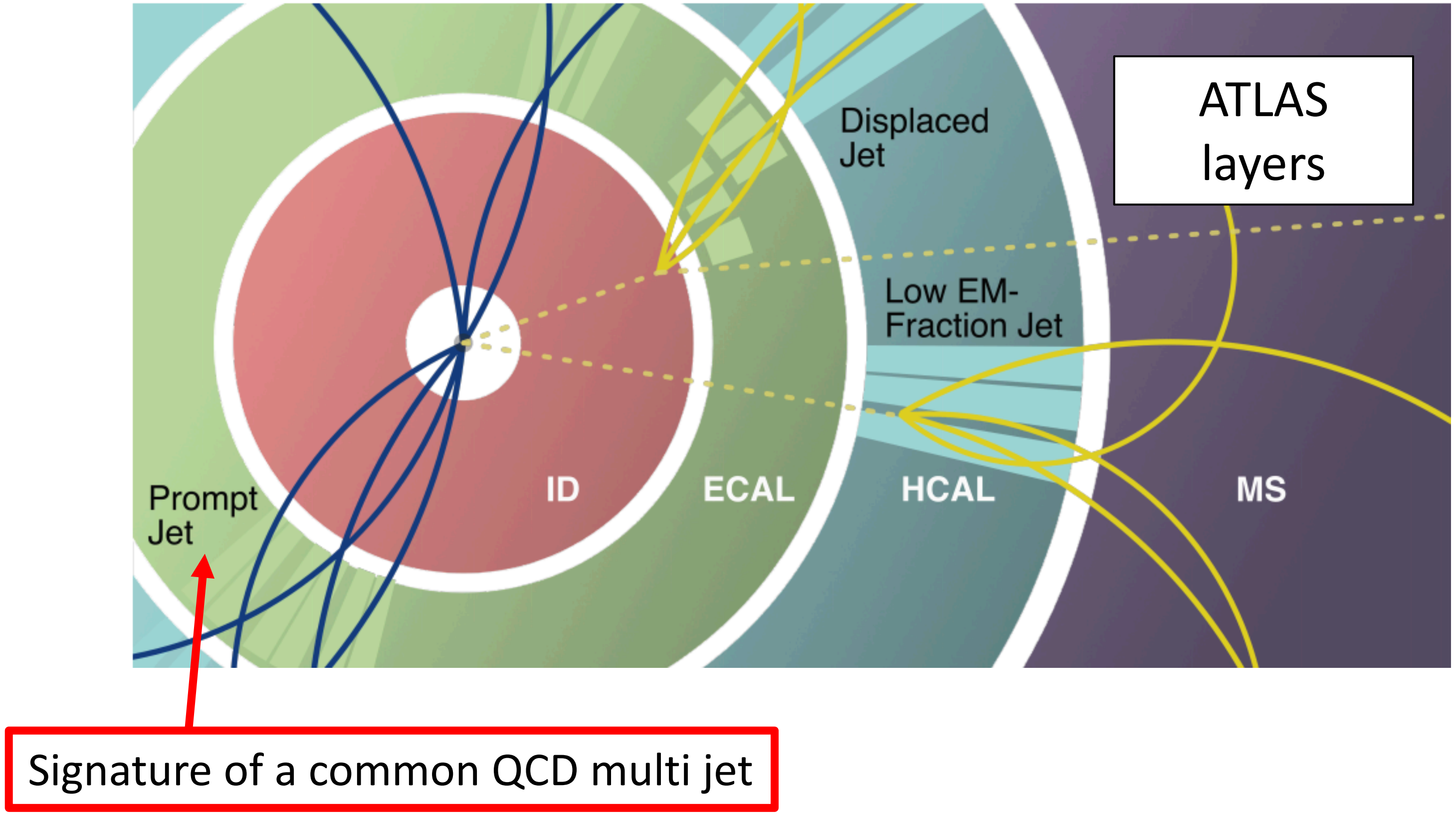


Signature of a signal jet

Background jets

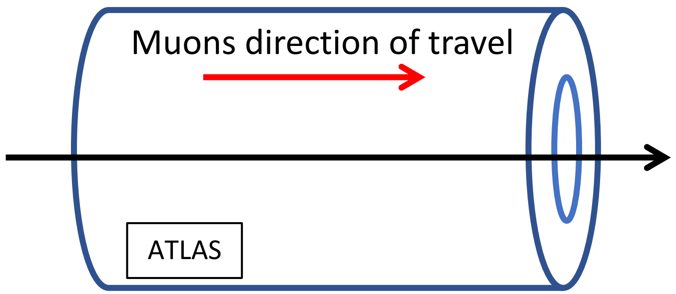
Which particle signatures mimic signal?

- **Background jet #1:** QCD multi jets are decays to the SM from proton-proton collisions
- Least probable to resemble signal but most abundant
- Cluster of neutrons decaying to other particles in the HCAL mimics signal

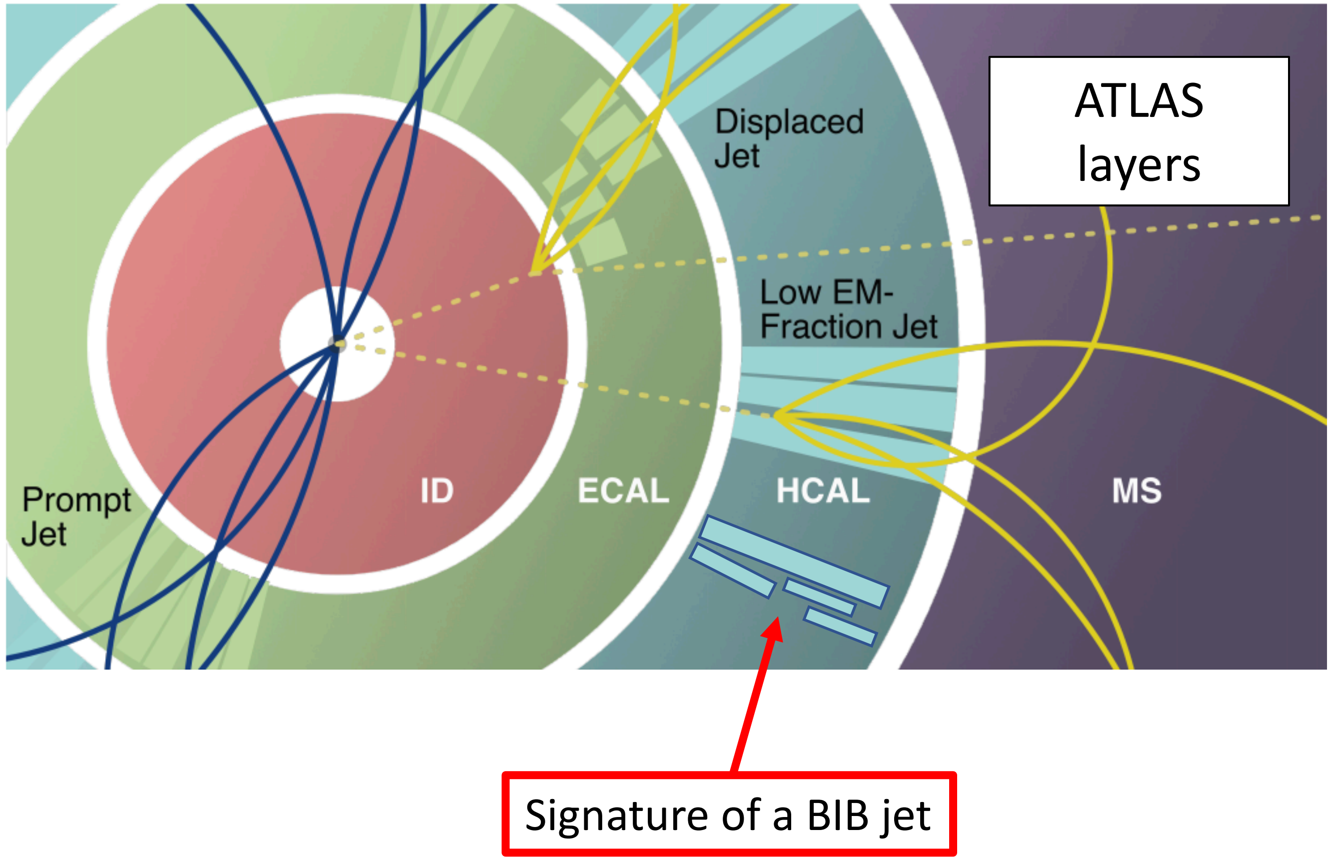


Background jets

Which particle signatures mimic signal?



- **Background jet #2:** Beam-induced background (BIB) is from muons generated from proton interactions before the proton-proton collision in ATLAS
- Muons travelling parallel to beam pipe could deposit energy in calorimeters



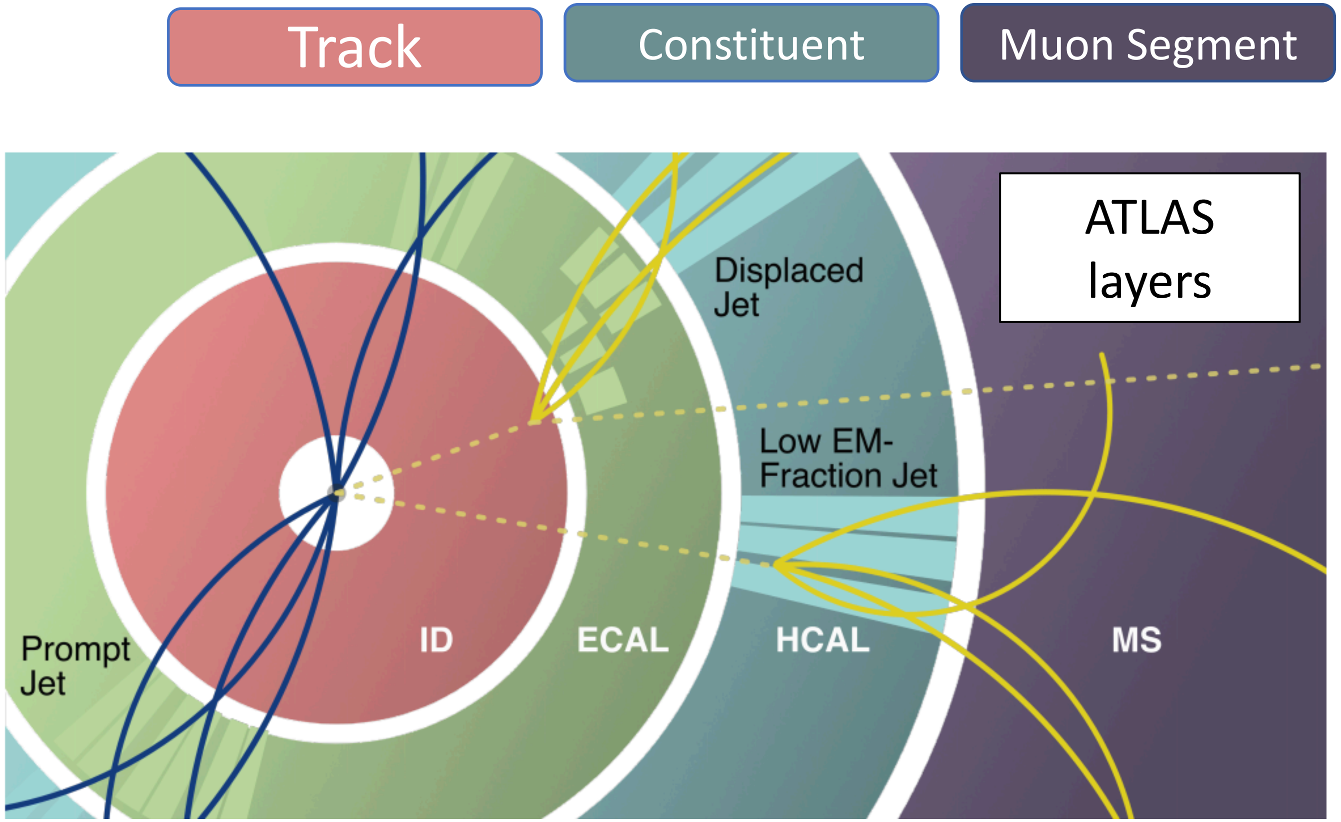
Differentiating signal from background

- Need a multi-class classification model to classify jet as either signal, QCD, or BIB
- Previous published analysis on the search for LLPs used Boosted Decision Trees (BDT)
- Ongoing analysis uses Recurrent Neural Networks (RNN)
- Main goal of this project: improve classification model

Machine Learning

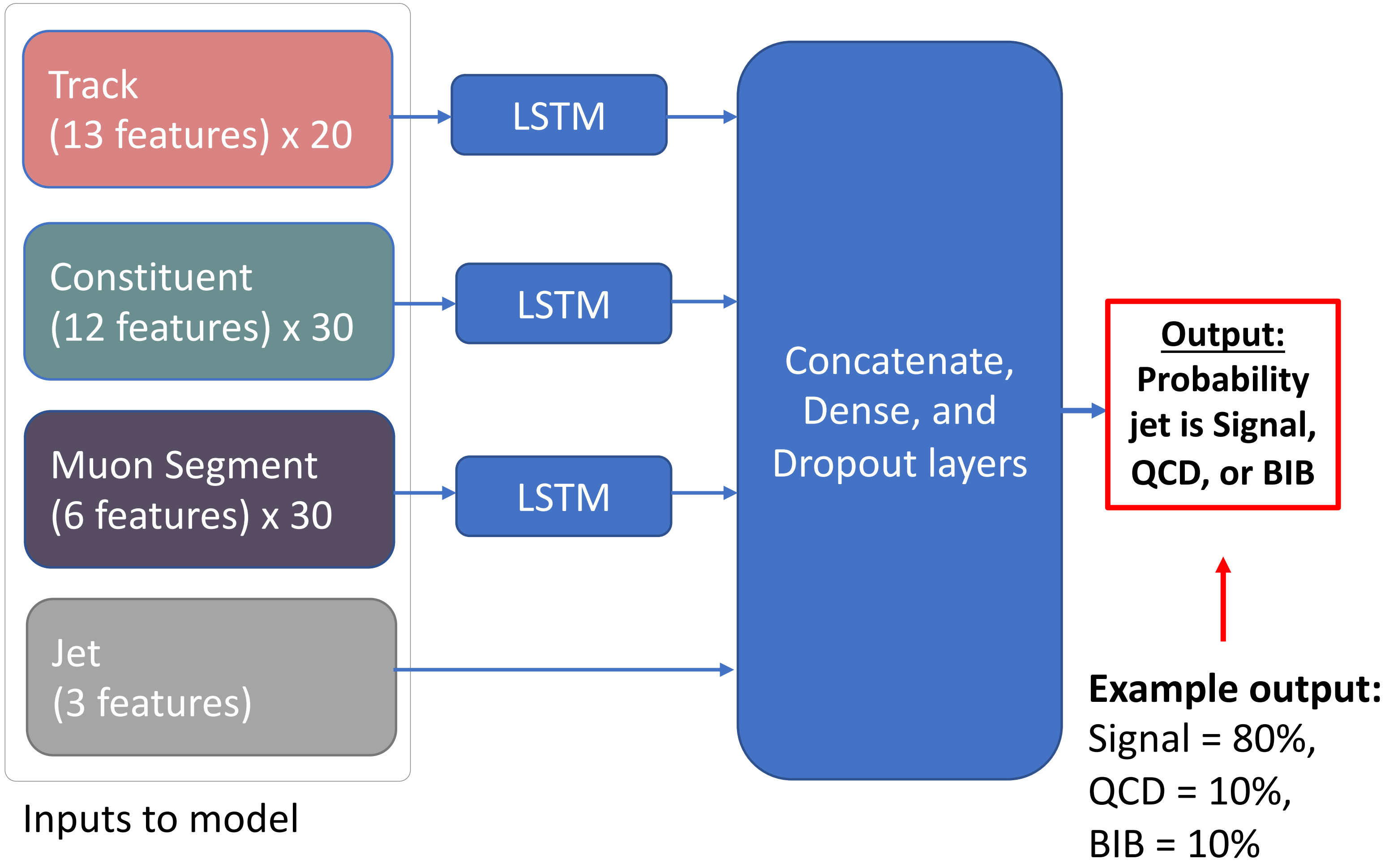
Machine Learning terminology

- **Features** are the variables that describe an input e.g. track, constituent, muon segment
- **Labels** are the outputs the model is trying to predict e.g. Signal, QCD, BIB
- Predicted labels are compared to true labels during the **training phase**



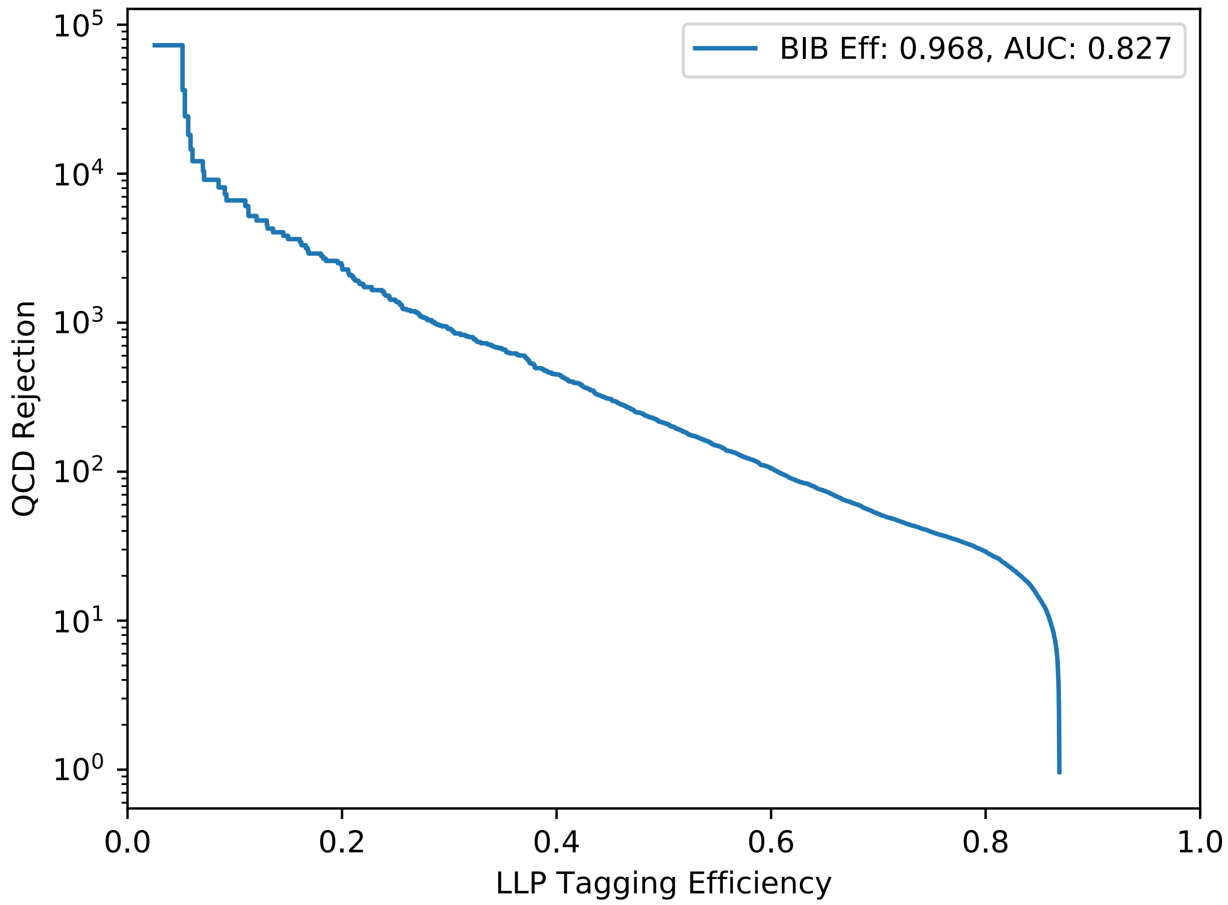
Current model architecture

- Input to model is a jet with up to 20 tracks, 30 constituents, and 30 muon segments
- Leverages Long Short-Term Memory networks (LSTM), capable of learning long-term dependencies
- Good at making predictions based on ordered data e.g. time series



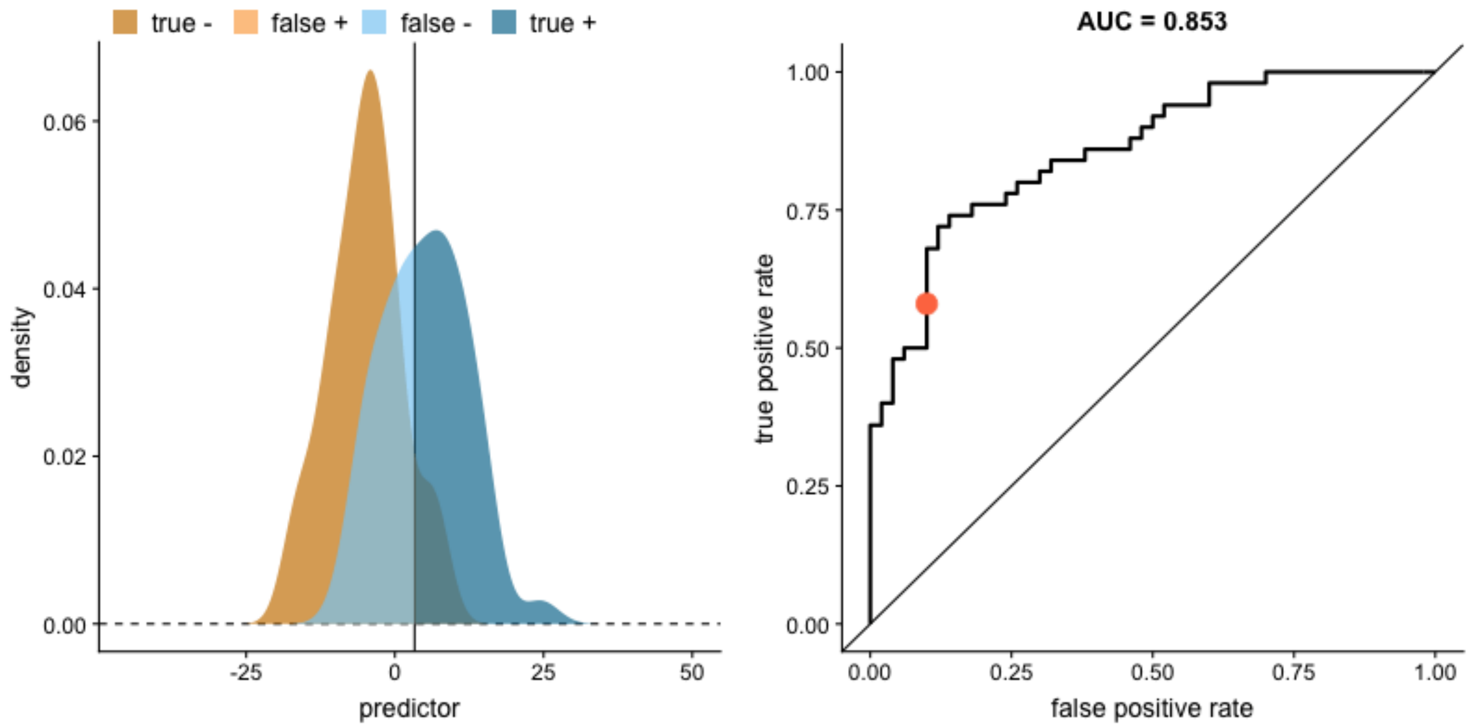
Metrics for model performance

- Accuracy
  - Percentage of correctly labeled jets
- ROC Area under curve (AUC)
  - X-axis: Proportion of times correctly tagged jet as signal
  - Y-axis: 1 / FPR where false positive rate (FPR) is the proportion of jets tagged as signal but were actually QCD
  - BIB efficiency is the proportion of true BIB jets classified as BIB
  - Higher AUC = better network



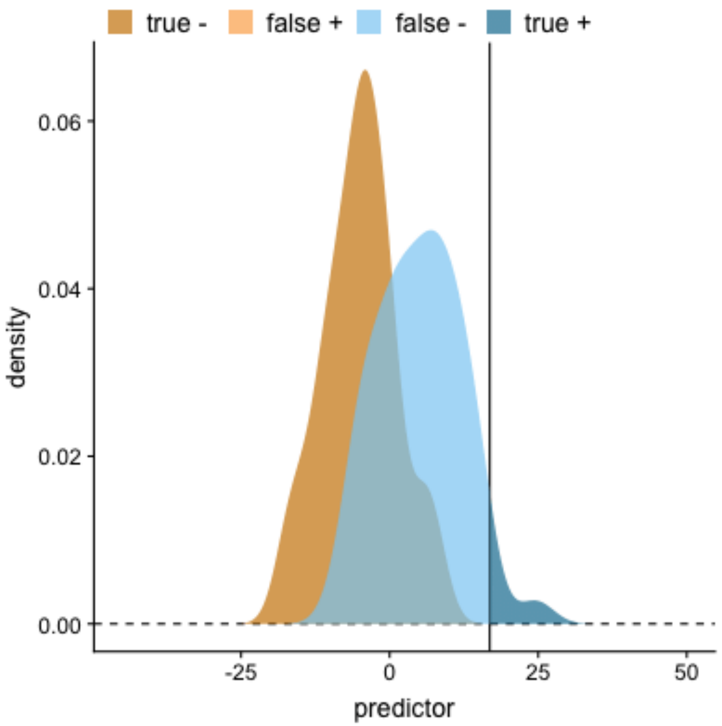
Visualizing ROC AUC

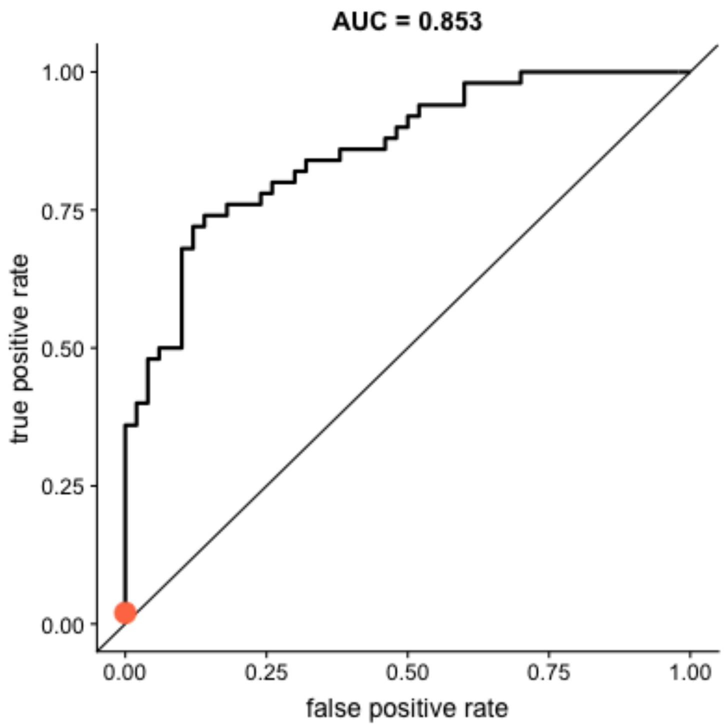
- Extend idea below to 3 cuts i.e. signal, QCD, and BIB
- Axes are slightly different than previous figure, but idea is the same



Visualizing ROC AUC

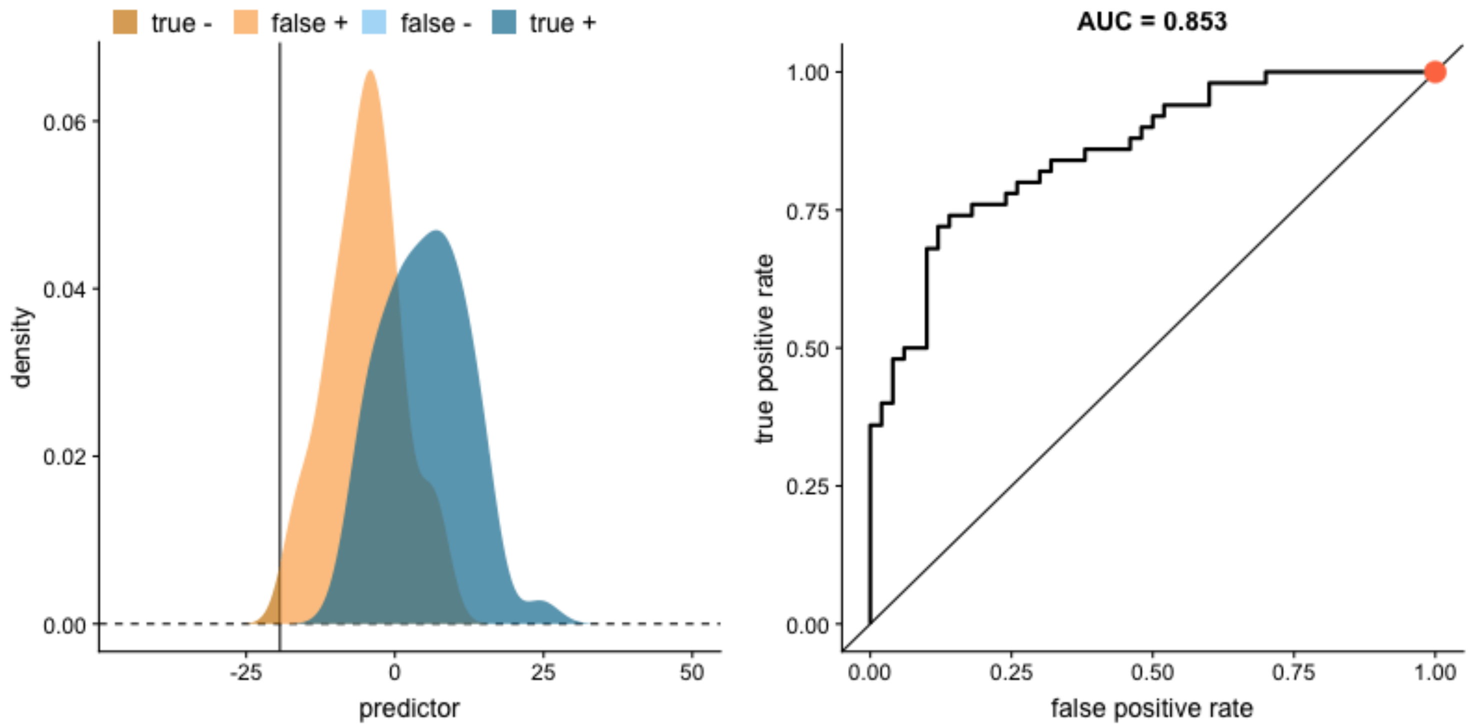
- Extend idea below to 3 cuts i.e. signal, QCD, and BIB
- Axes are slightly different than previous figure, but idea is the same





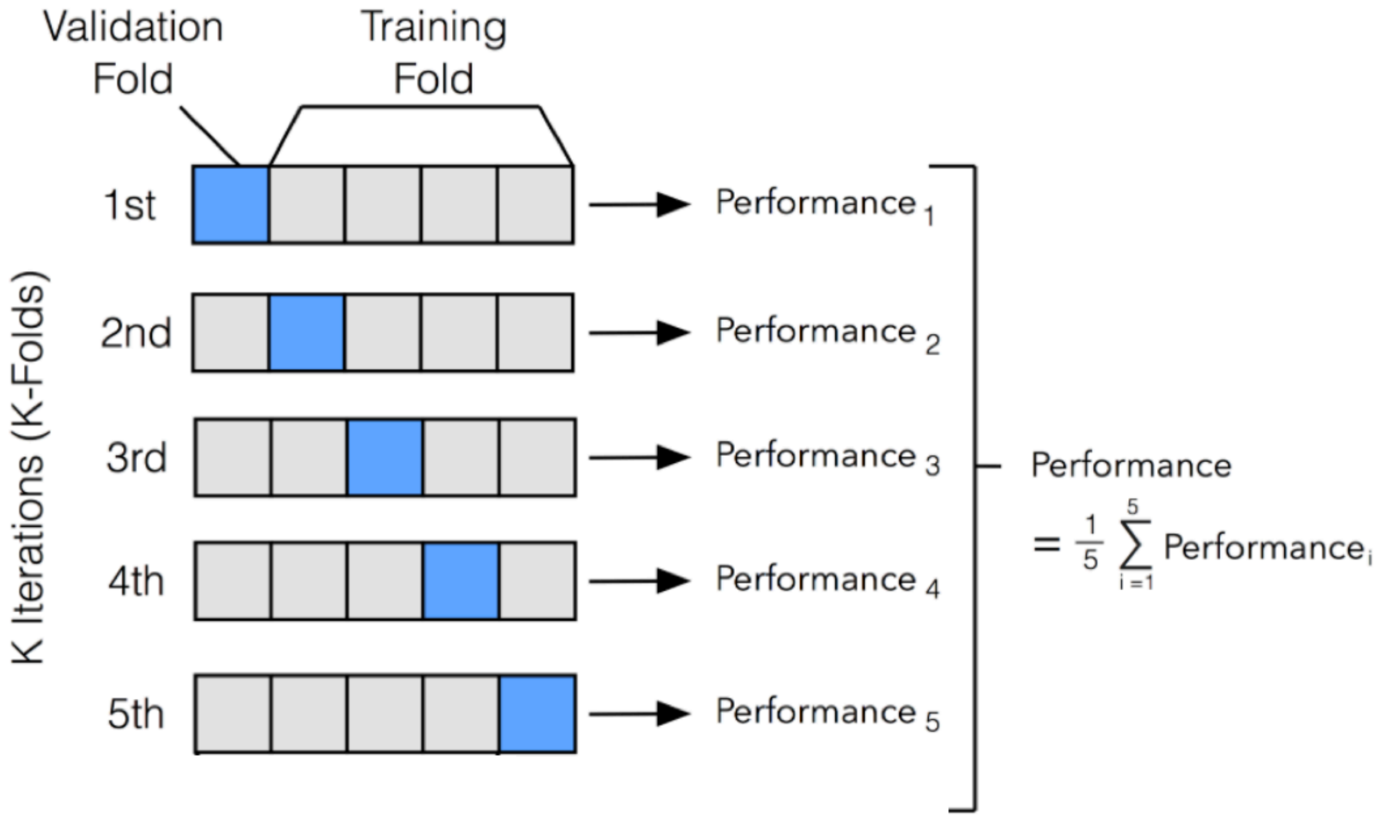
Visualizing ROC AUC

- Extend idea below to 3 cuts i.e. signal, QCD, and BIB
- Axes are slightly different than previous figure, but idea is the same



Comparing model performances

- **K-Fold Cross Validation**: Statistical technique used to evaluate performance of model
- Results in a less biased or less optimistic estimate of the model skill than other methods

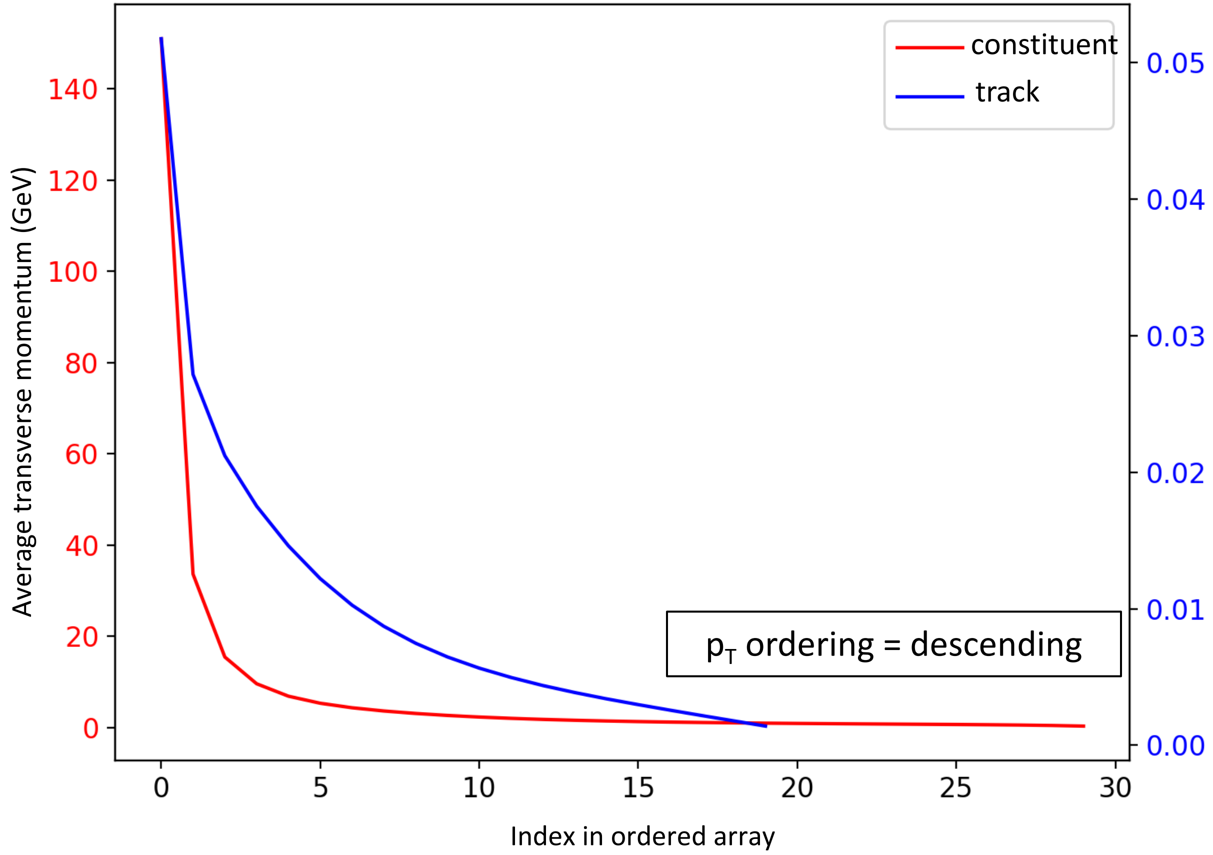


Study #1

Does the ordering of transverse momentum matter?

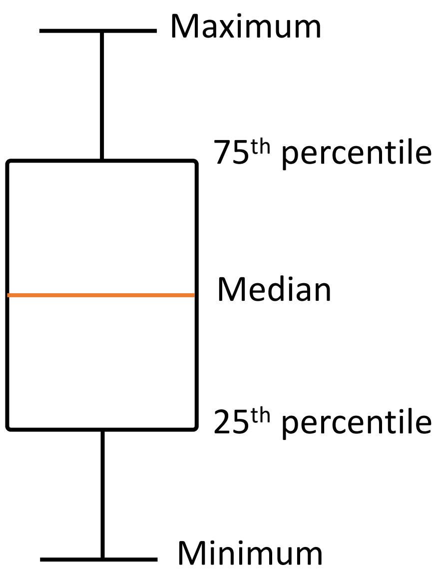
Changing the order of transverse momentum

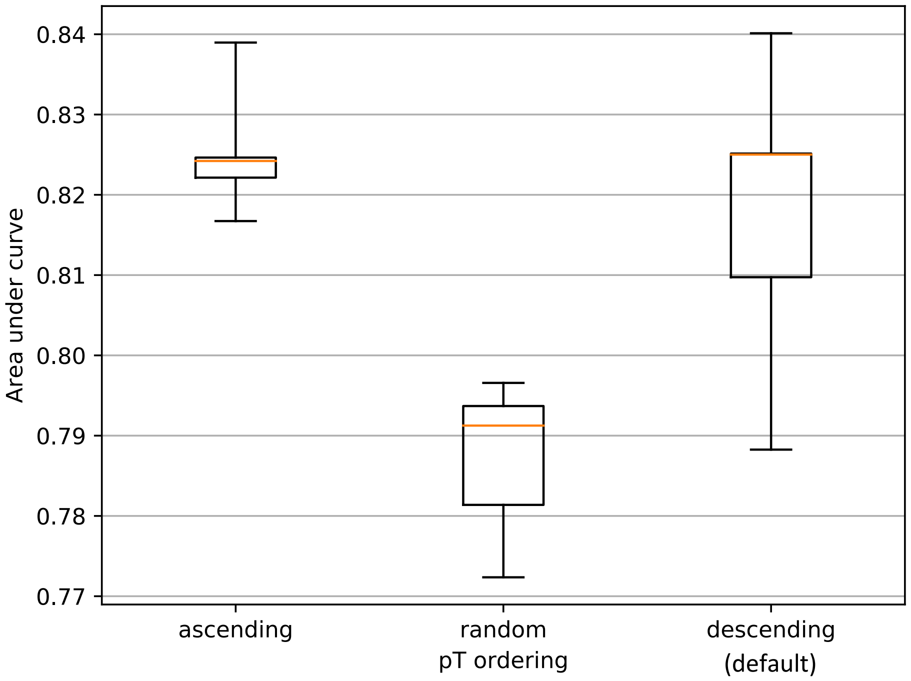
- Transverse momentum (p_T) is the momentum of a particle in the transverse plane
- Current ordering of tracks, and constituents is by descending p_T
- This study will compare performance of inputs with different ordering of p_T

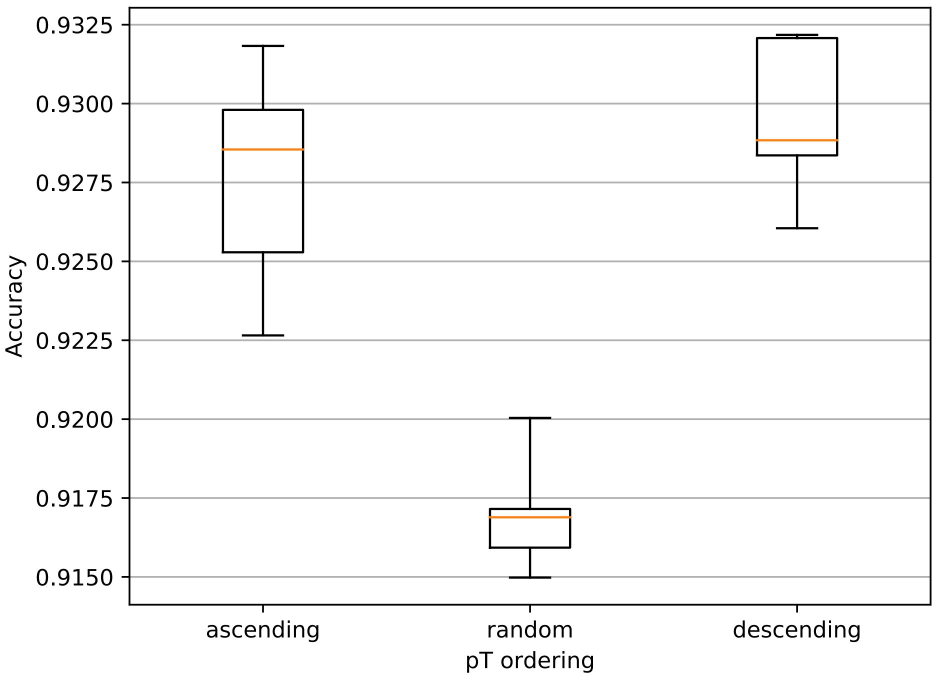


K-Fold results

- Ascending and descending order perform better than random
- Can conclude p_T is a good ordering for jet constituents and jet tracks





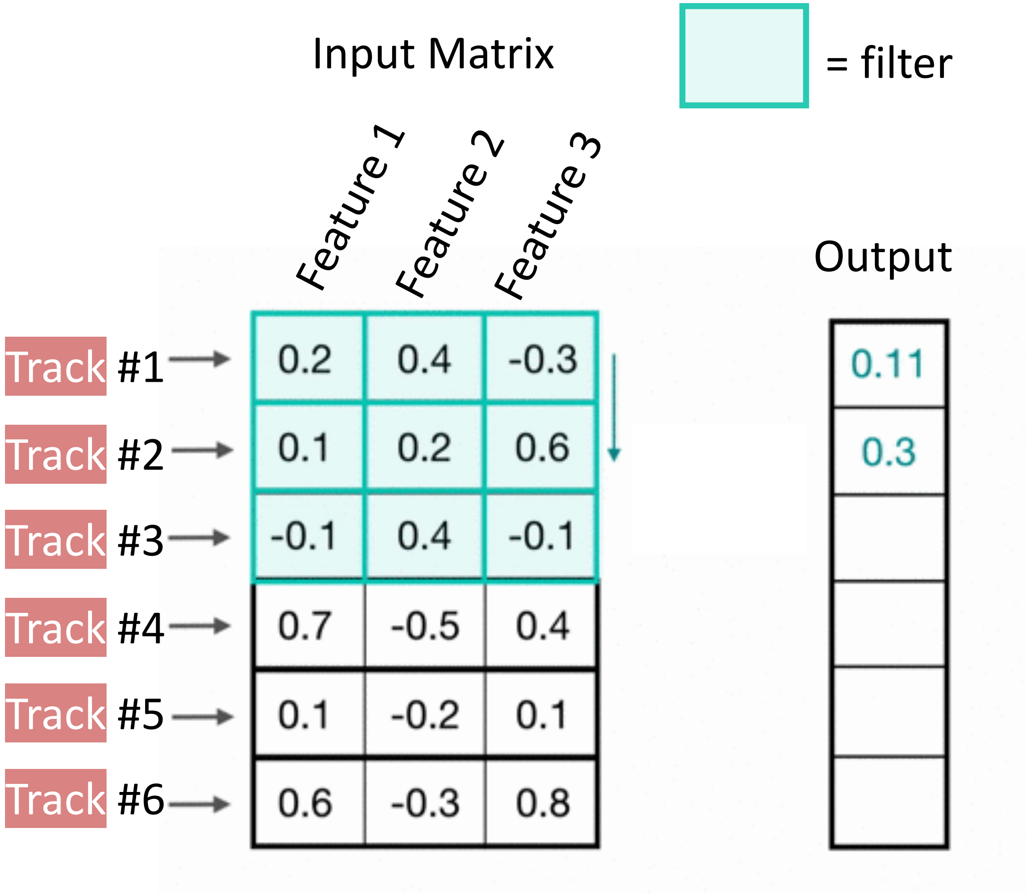


Study #2

Modifying architecture to improve model performance

1D Convolutions

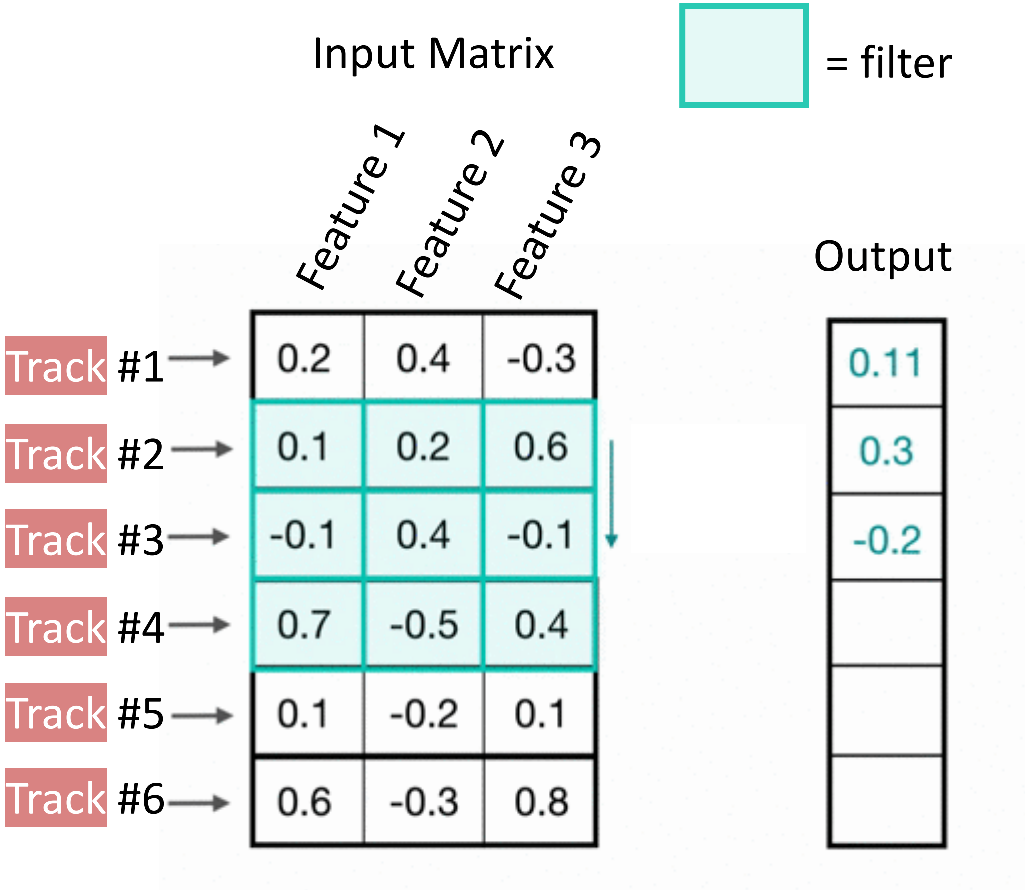
- Add 1D convolutional layers between inputs track, constituent, muon segment and LSTM layers
- Additional 1D convolutional layers act as feature extraction and compression



1D filter sliding down matrix

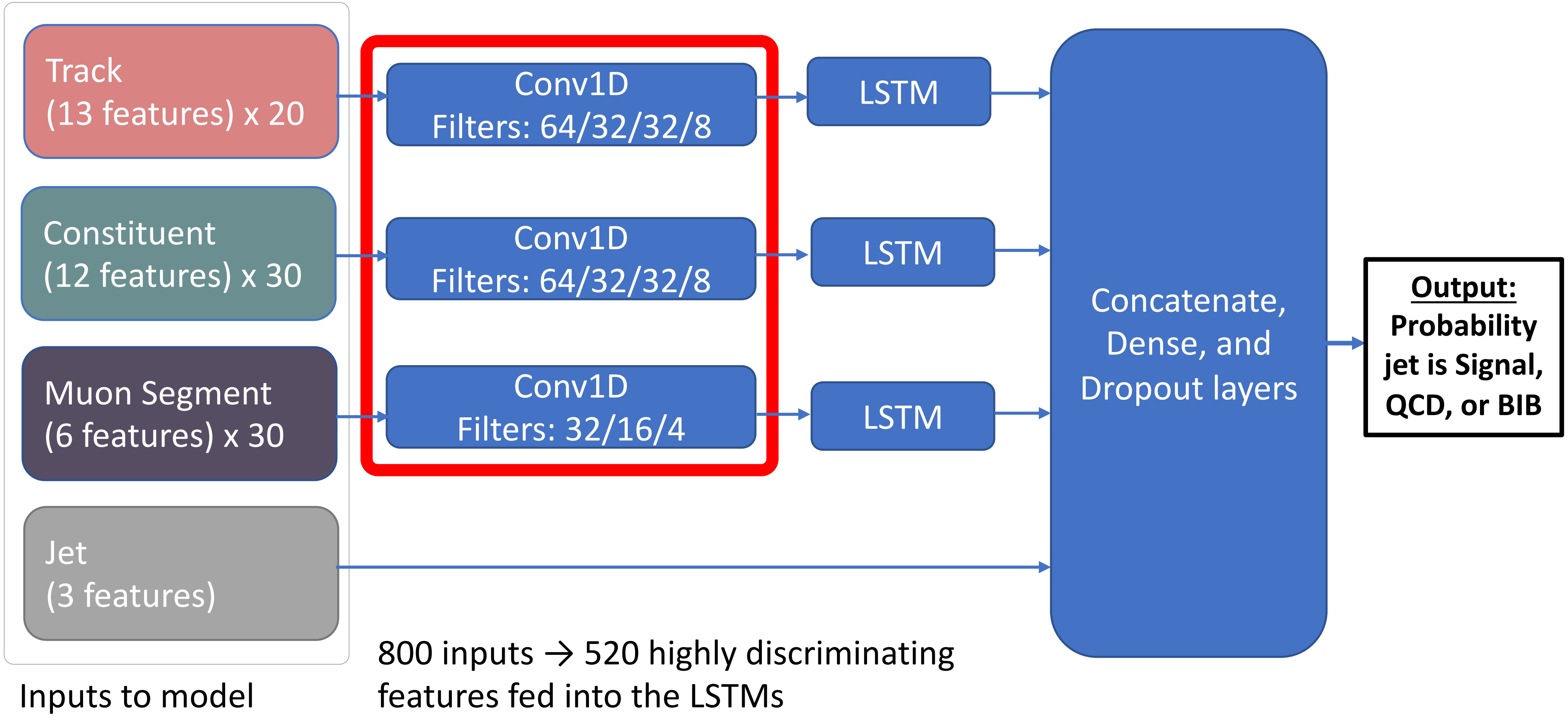
1D Convolutions

- Add 1D convolutional layers between inputs track, constituent, muon segment and LSTM layers
- Additional 1D convolutional layers act as feature extraction and compression



1D filter sliding down matrix

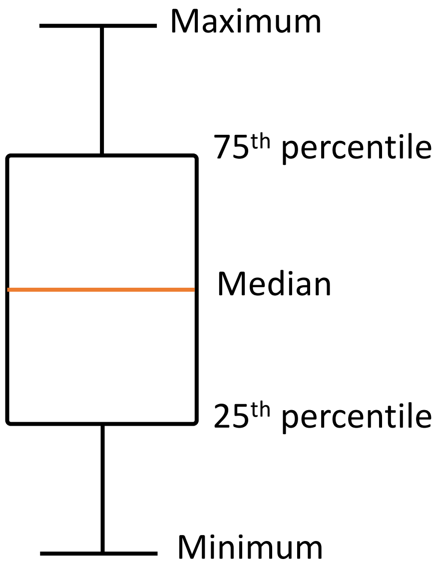
Modifying the current architecture

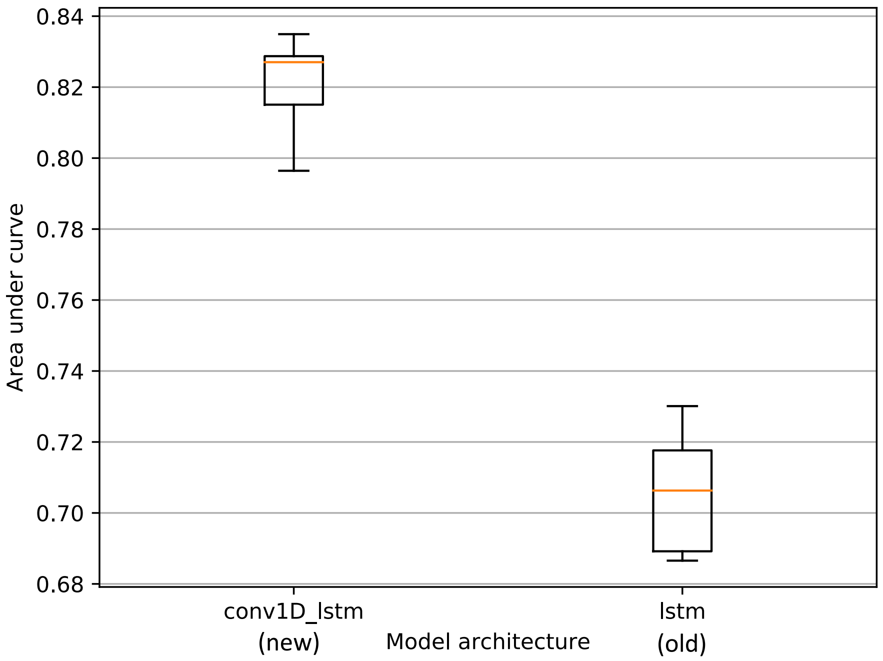


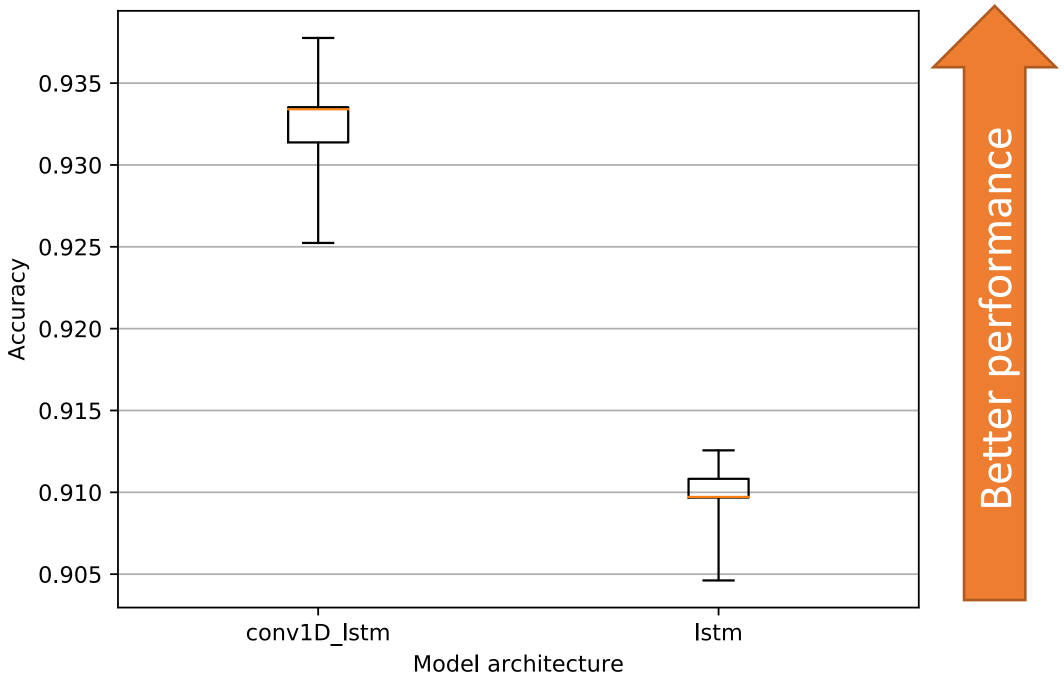
800 inputs → 520 highly discriminating features fed into the LSTMs

K-Fold results

- New architecture (*conv1D_lstm*) performed better in both metrics





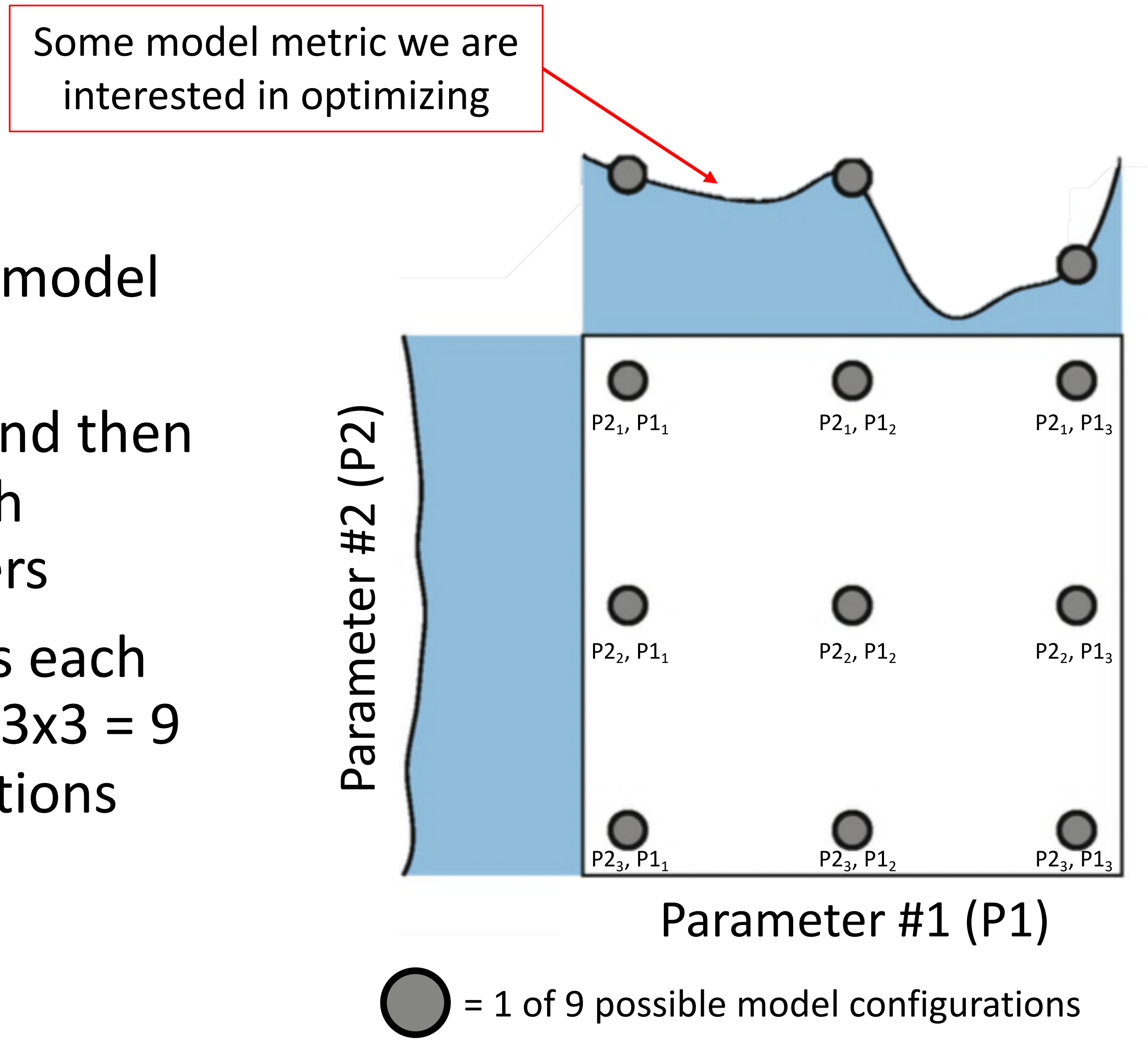


Study #3

Grid search to optimize hyperparameters on new model architecture

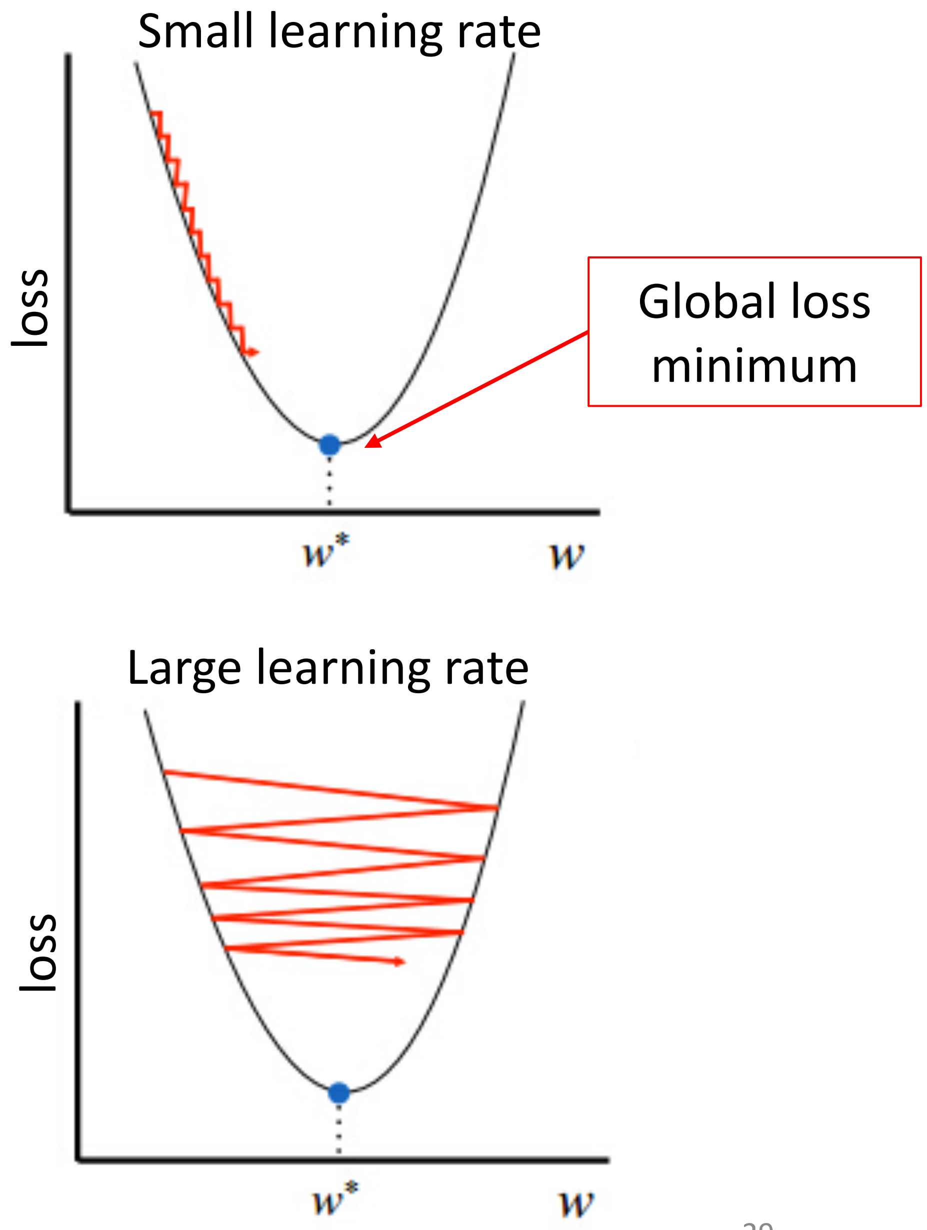
Grid Search

- Technique for optimizing model hyperparameters
- Decide on search space and then generate a model for each combination of parameters
- Example: Two parameters each with 3 possible values so 3x3 = 9 possible model configurations

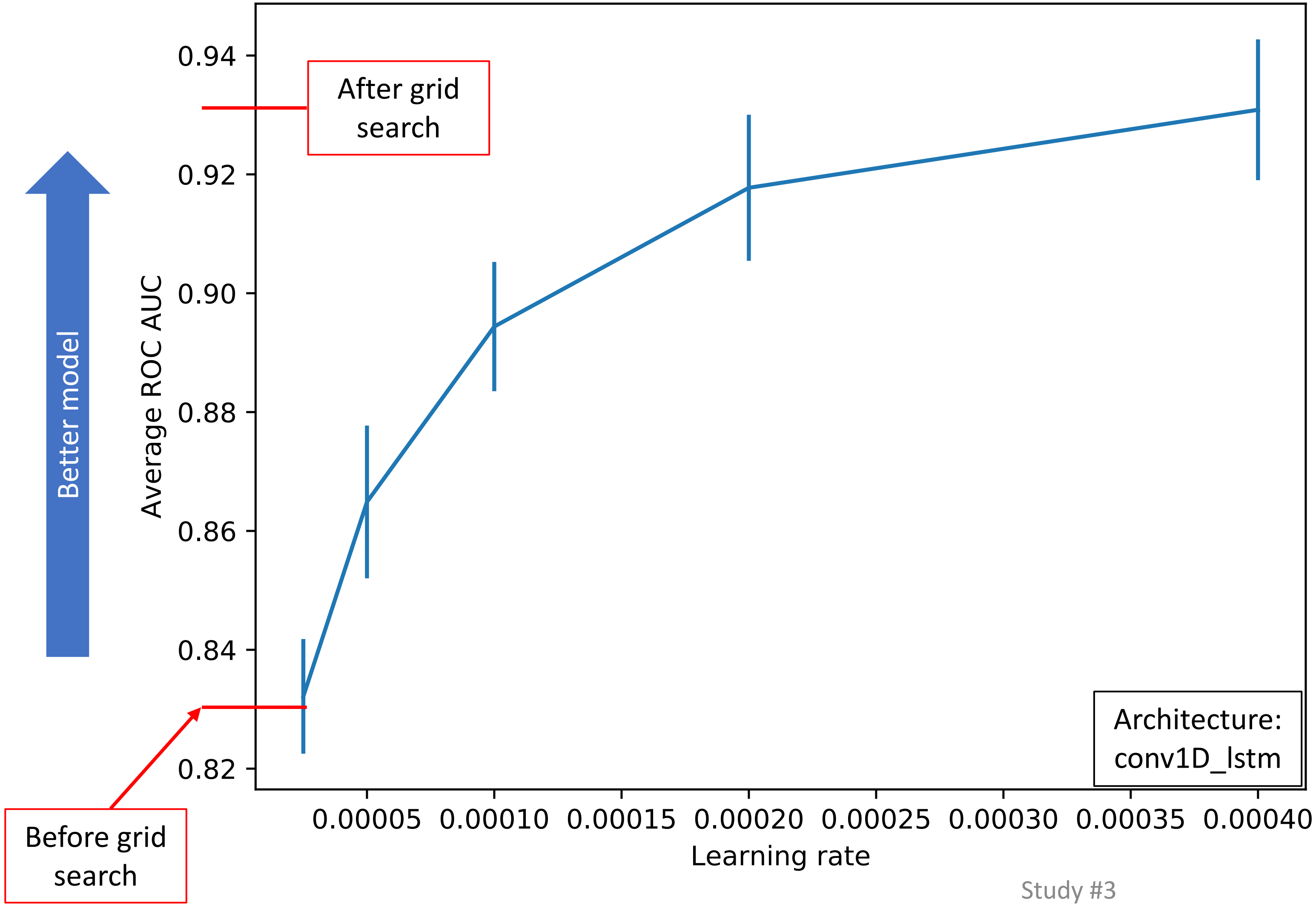


Grid Search

- Search space: 5 values of learning rate, 4 values of regularization, and 3 values for number of nodes in final Conv1D layer
- 5x4x3 = 60 unique models



Grid Search results



- Positive correlation found between learning rate and model performance
- 12 different models were trained for each learning rate
- Reached optimum value since curve is plateauing

Conclusions and Discussions

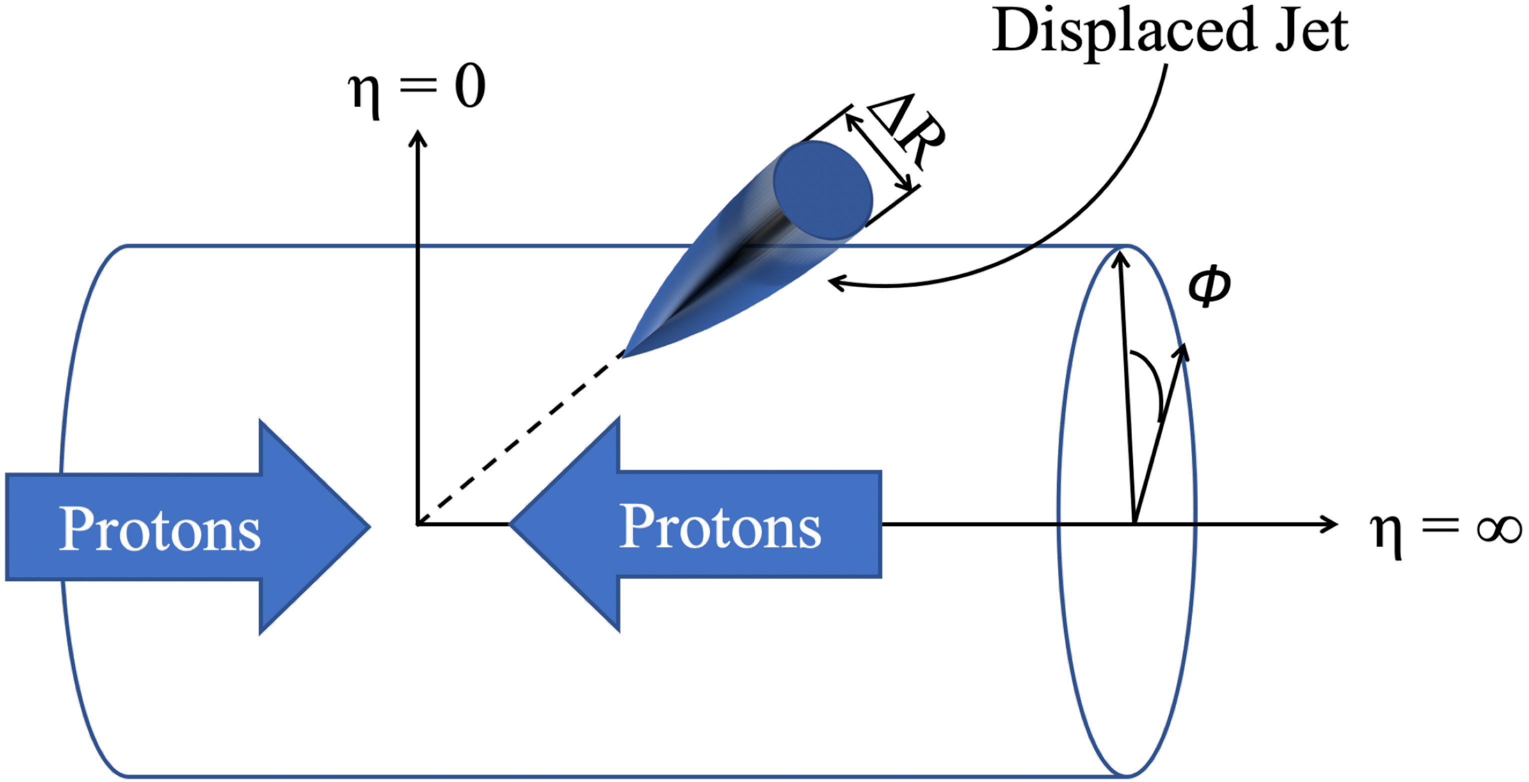
- Conclusion of results:
  - p_T ordering is important
  - Conv1D + LSTM outperforms old architecture
  - Bigger learning rate significantly improved model performance
  - Hopeful results will be used in the overall ATLAS analysis!
- To explore further:
  - Are there other metrics than accuracy or AUC to better quantify model performance?
    - Signal efficiency at specific QCD rejection
    - **S**/sqrt(**Background**), **S** = # of signal events and **Background** = # of background events

Acknowledgements

Thank you to Alison Lister and Felix Cormier for their continued support

Backup slides

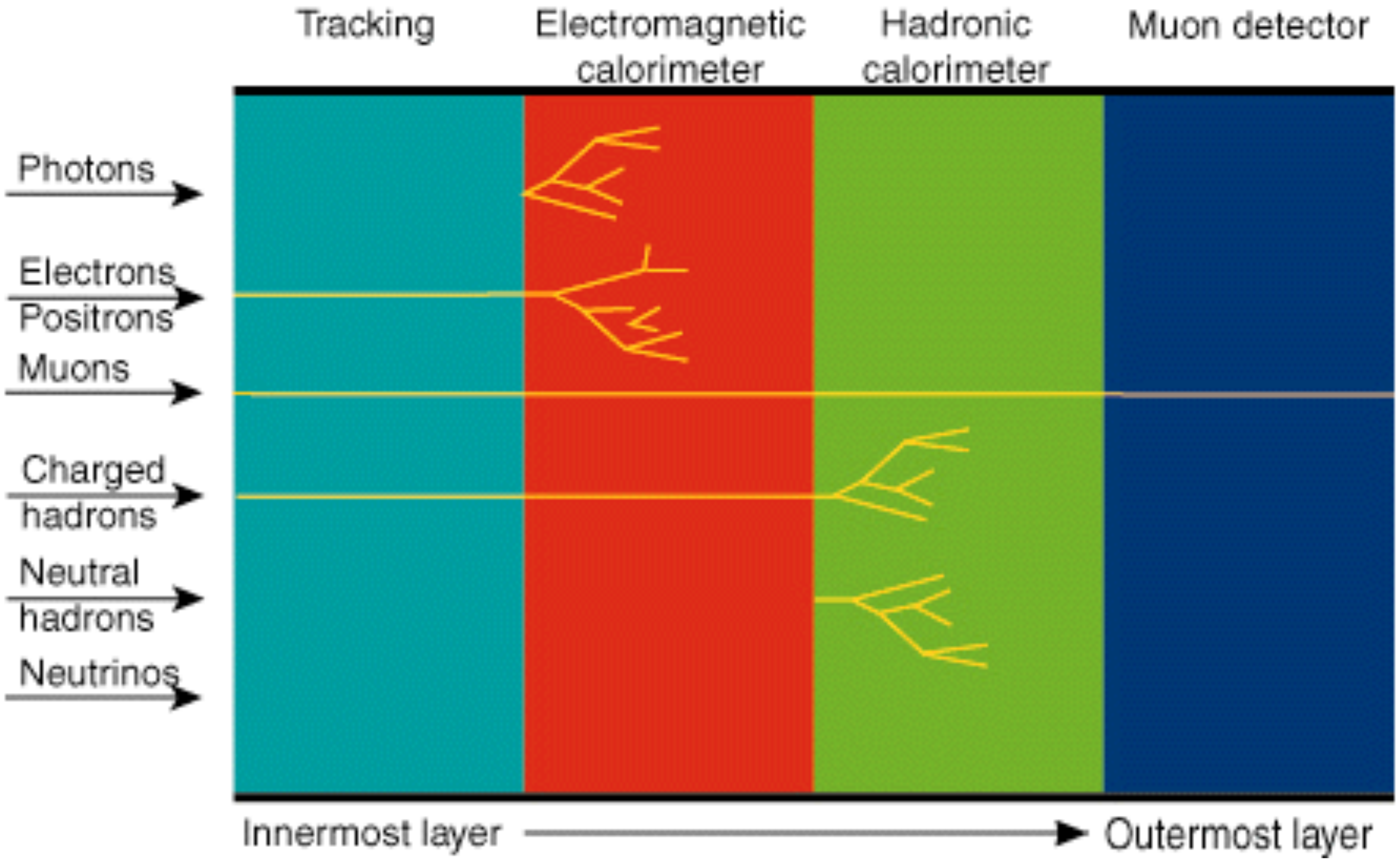
Important Variables



- **Pseudorapidity (η)** - angle of the particle in relation to the axis of the detector cylinder
- **Angle (Φ)** - Angle of a particle in the transverse plane
- **Transverse momentum (p_T)** - Momentum of a particle in the transverse plane
- **ΔR** - Width of jet

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$$

Particle signatures



Features

Constituent:

- pt
- eta
- phi
- l1hcal
- l1ecal
- l2hcal
- l2ecal
- l3hcal
- l3ecal
- l4hcal
- l3ecal
- time

Track:

- pt
- eta
- phi
- Vertex_nParticles
- D0
- Z0
- chiSquared
- PixelShared
- SCTShared
- SCTHoles
- SCTHoles
- PixelHits
- SCTHits

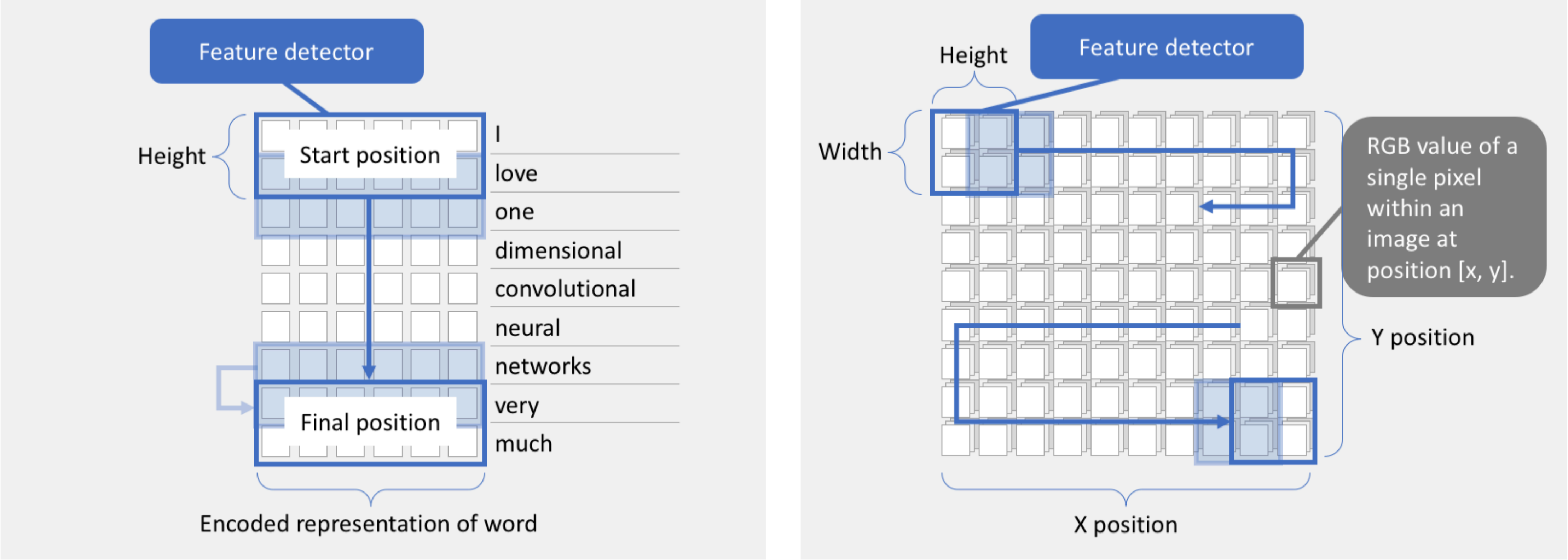
Muon Segment:

- etaPos
- phiPos
- etaDir
- phiDir
- chiSquared
- t0

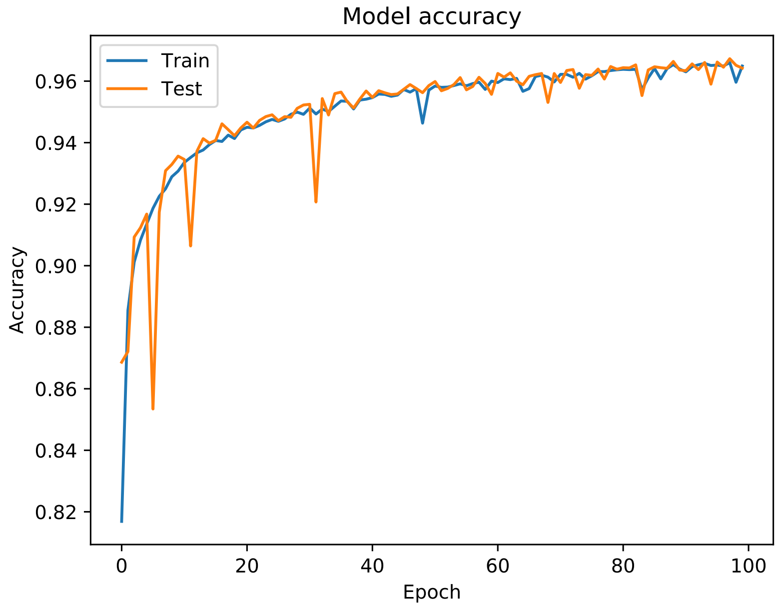
Jet:

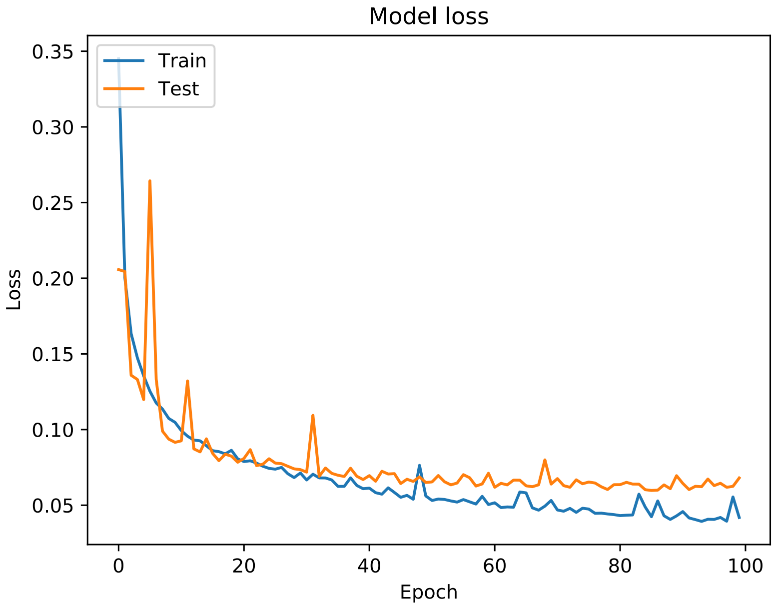
- pt
- eta
- phi

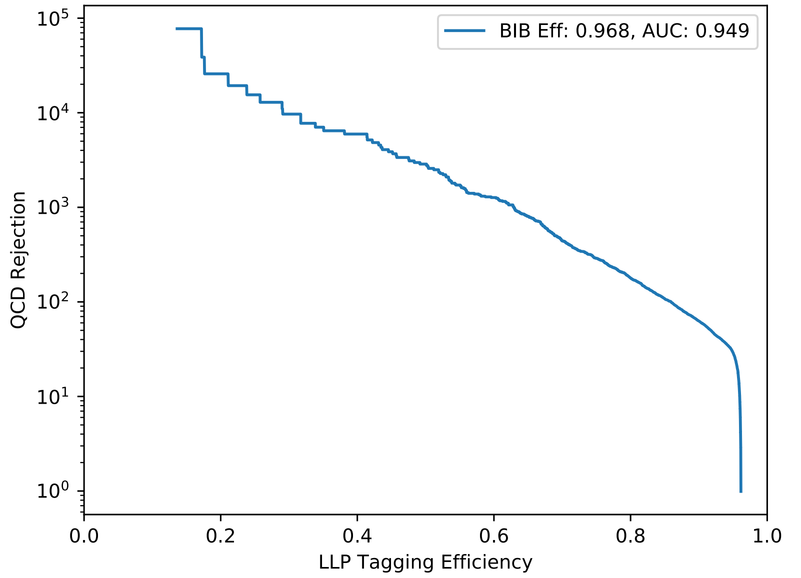
1D vs. 2D Convolution



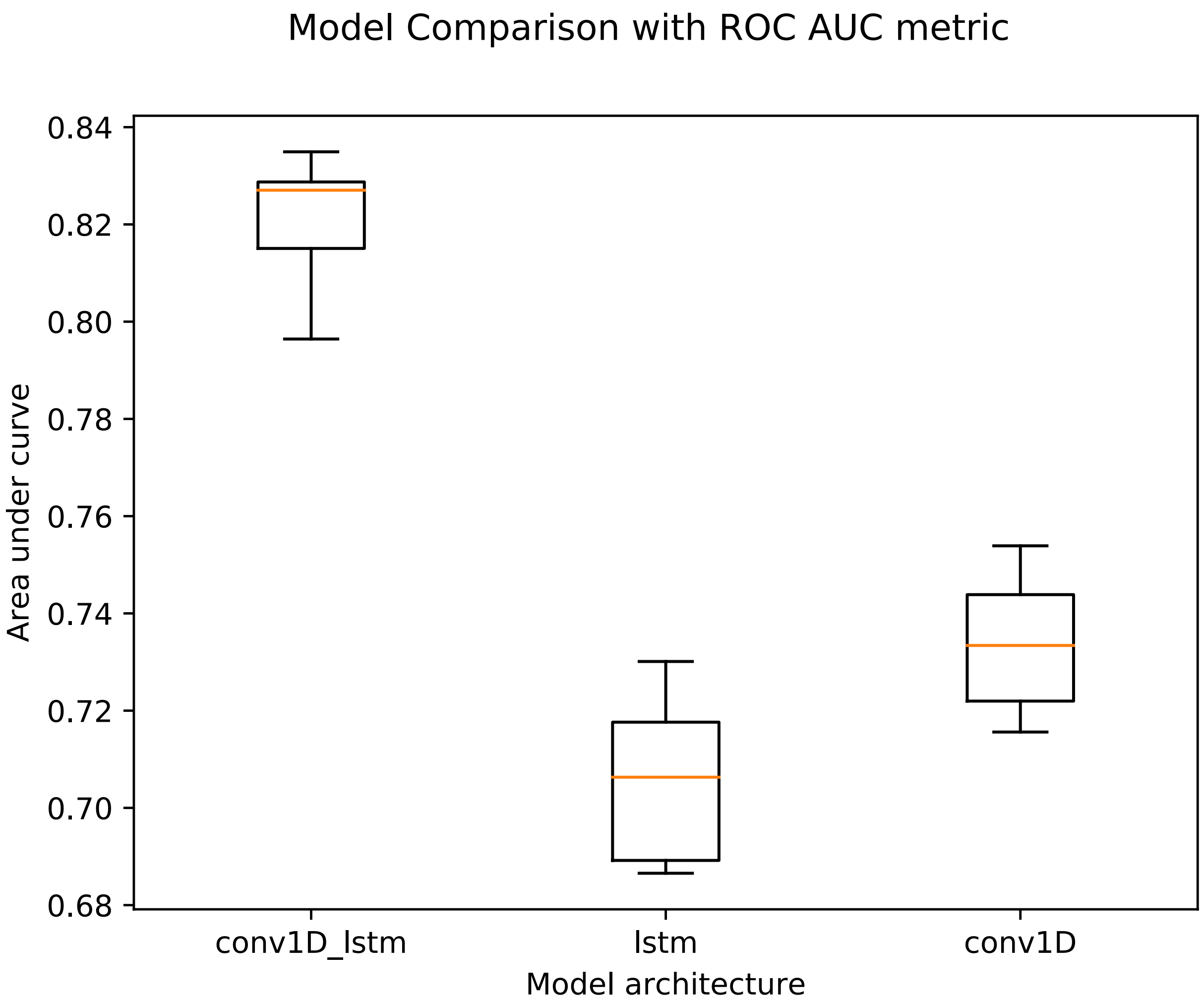
Plots for best model found





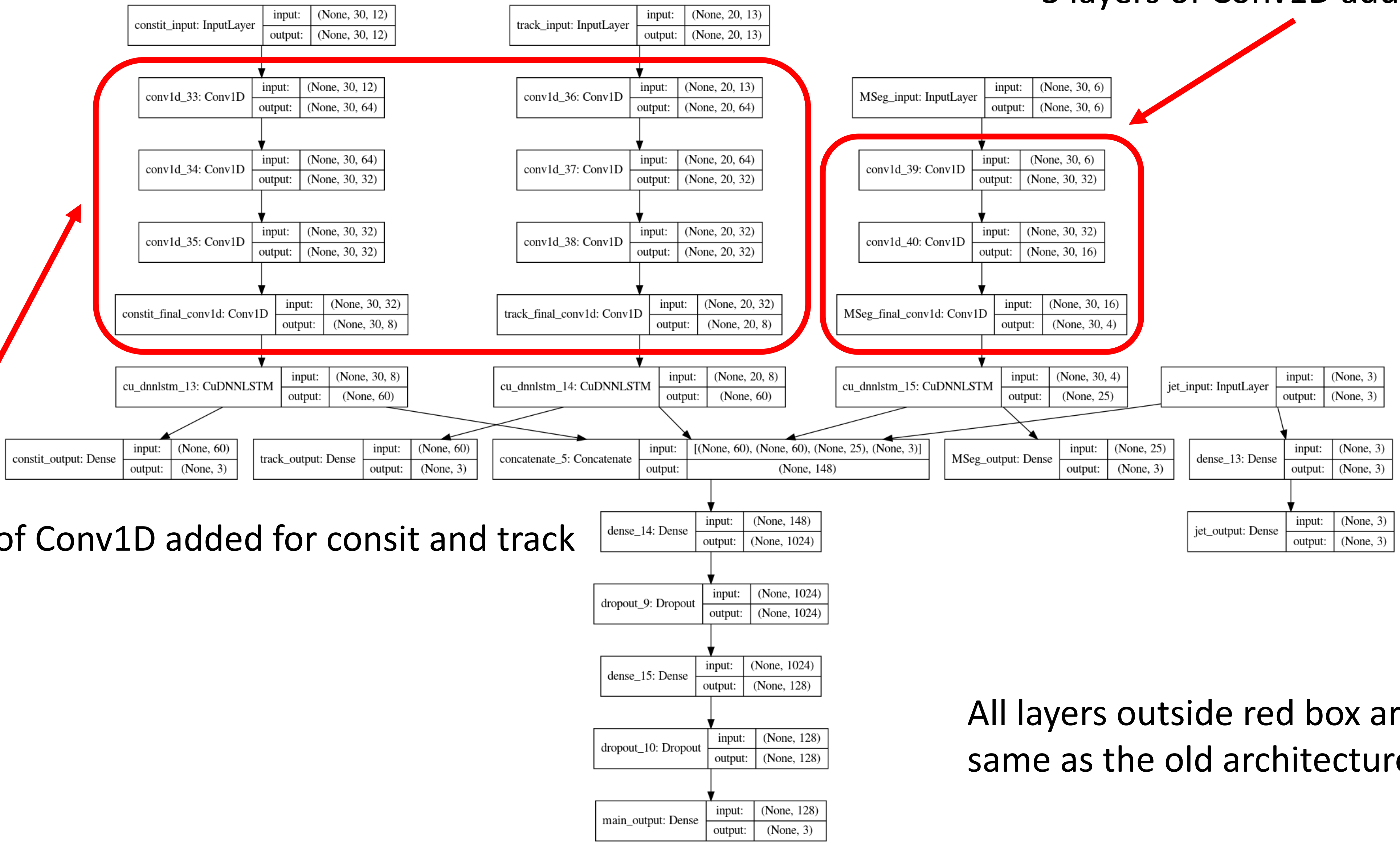


K-Fold results



New architecture

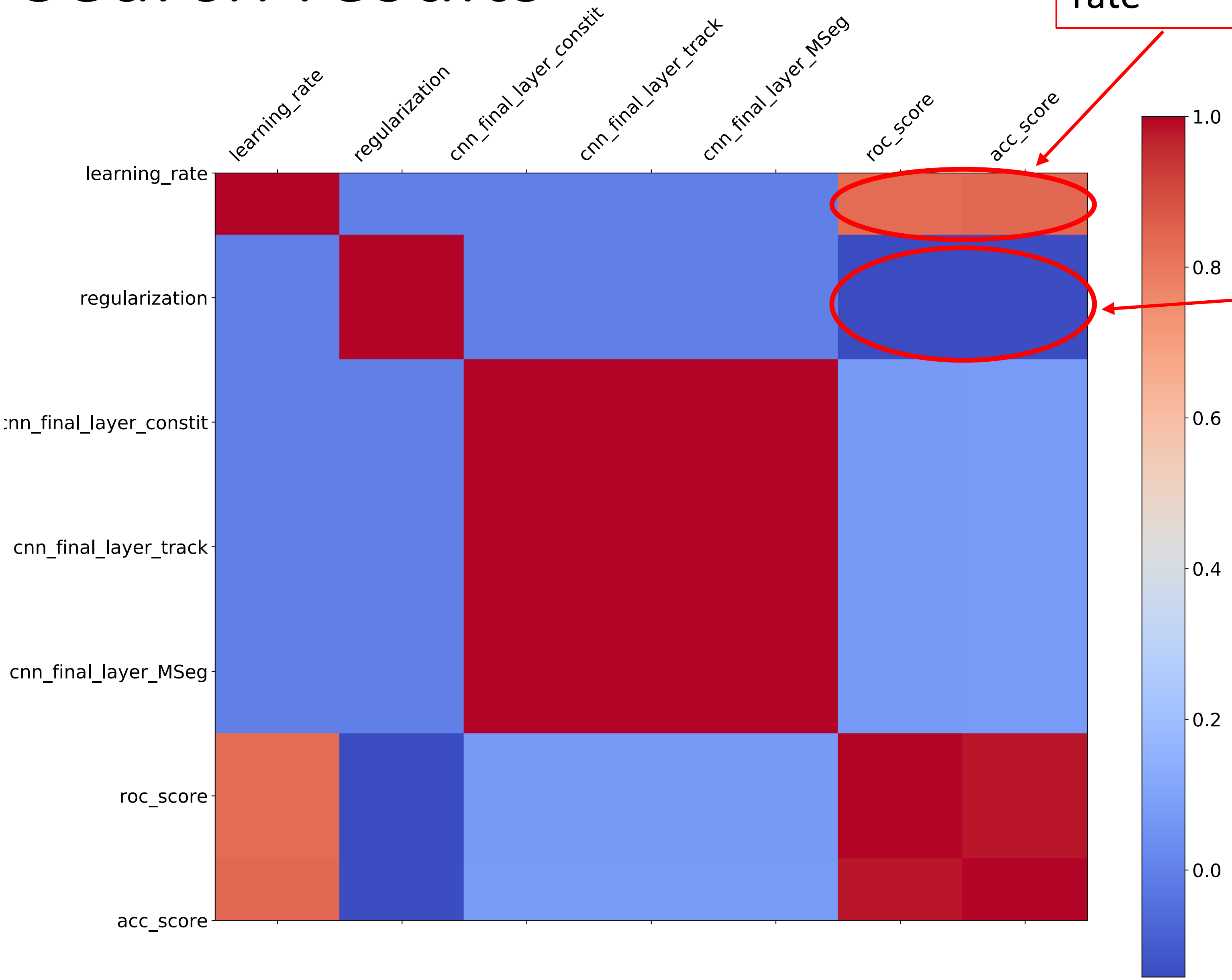
3 layers of Conv1D added for MSeg



4 layers of Conv1D added for consit and track

All layers outside red box are the same as the old architecture

Grid Search results



Positive correlation (0.8) between model performance and learning rate

Negative correlation (-0.14) but minimal so will ignore

Grid Search results