Improving Signal Classification for HNL with τ using Transfer ML

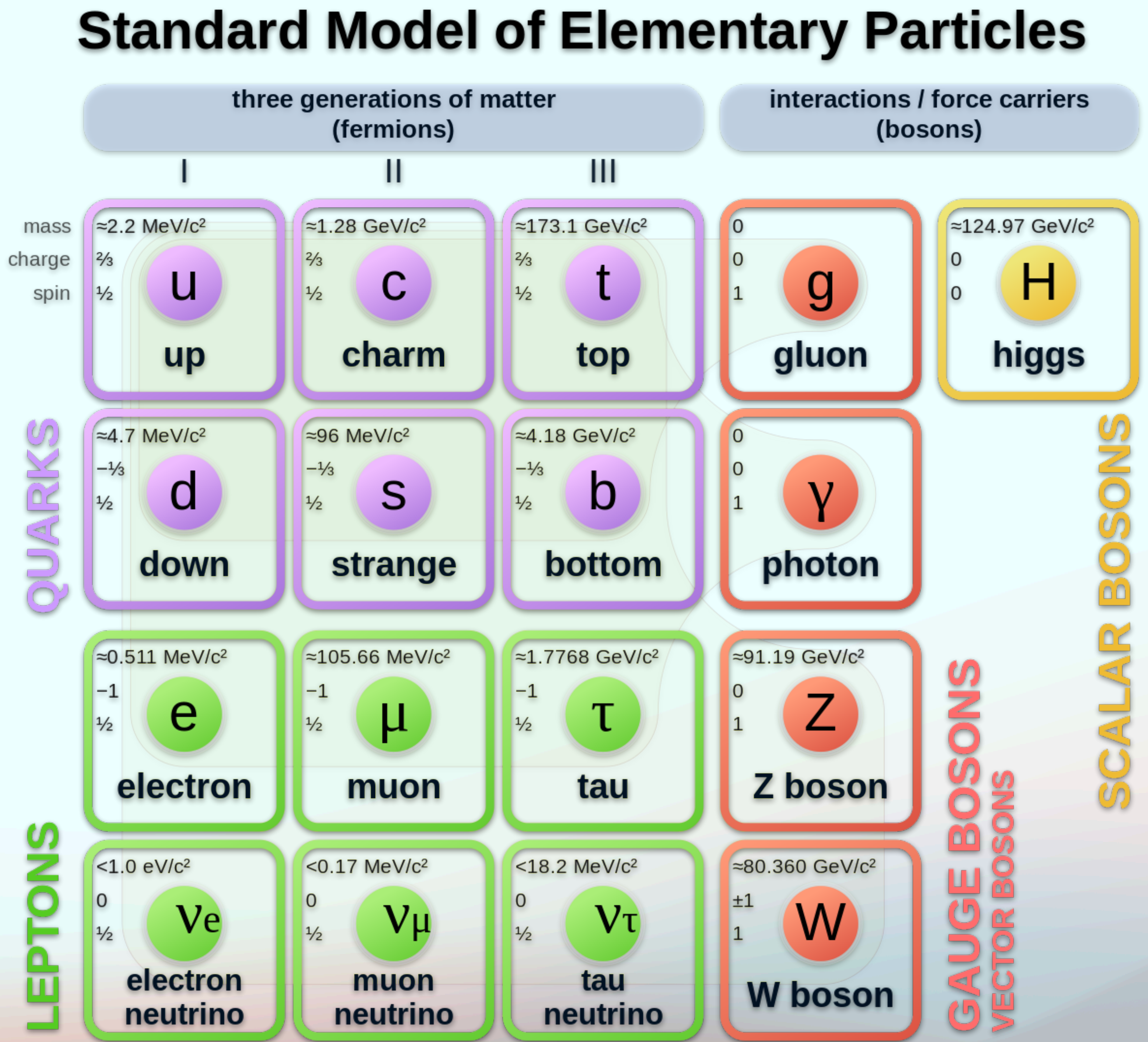
September 4, 2023

Dmitri Demler

Supervisor: Konstantin Androsov

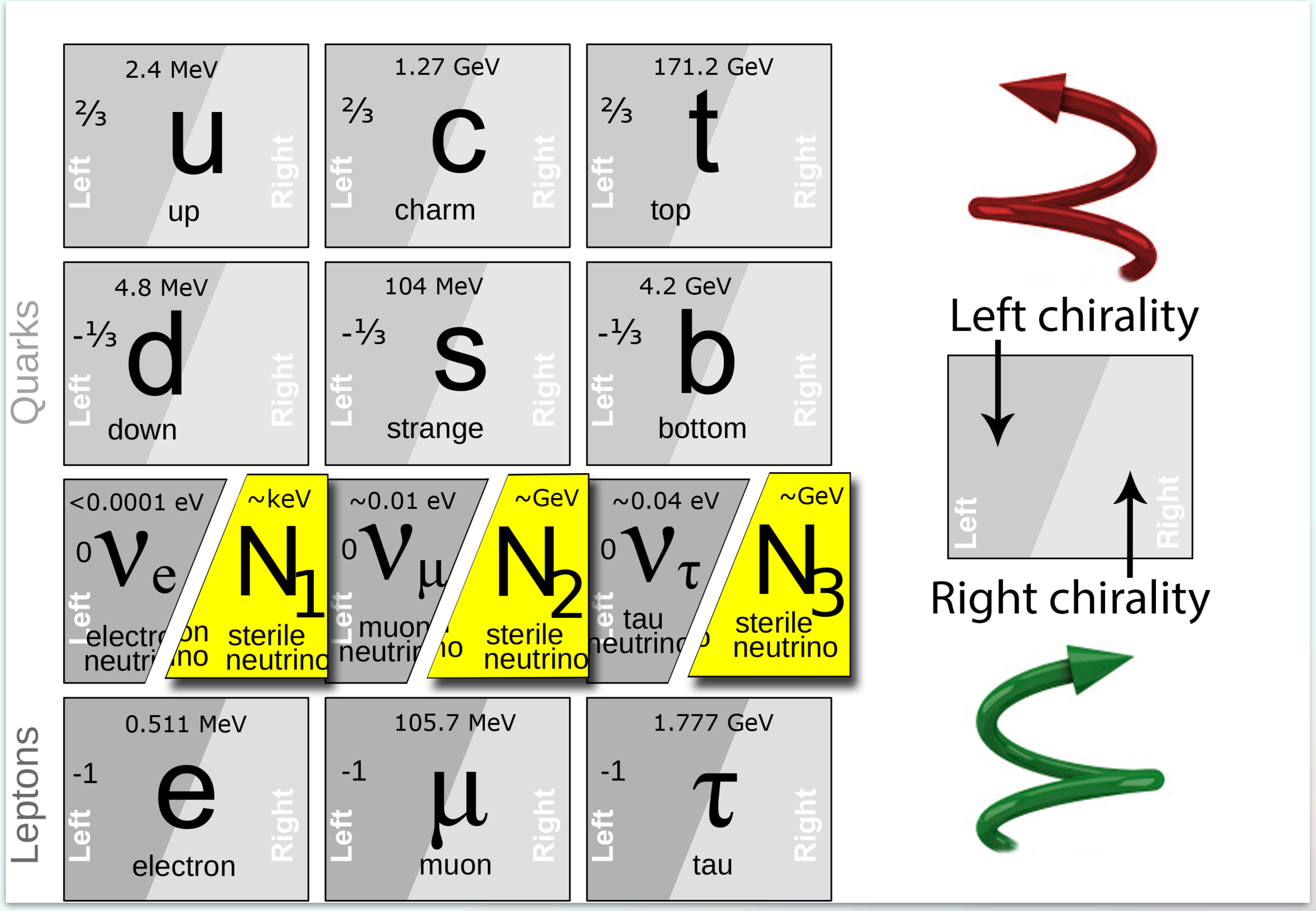
Standard Model

- Successful at predicting interaction between particles
- Struggles to explain dark matter to the mass of neutrinos



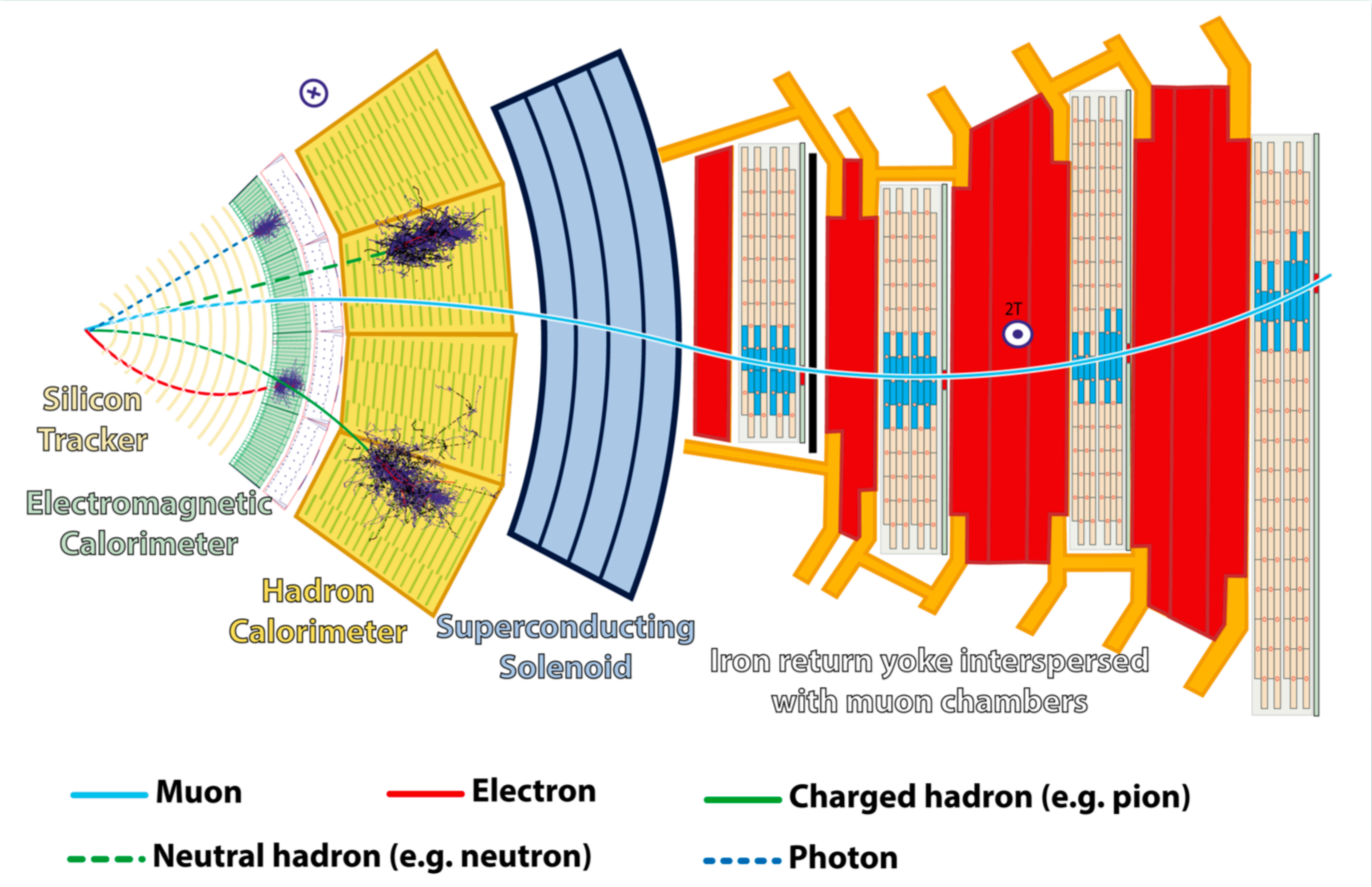
νMSM

- Neutrino minimal standard model
- Heavy Neutral Leptons
  - Introduces three right-handed, colorless neutrinos
  - Don’t interact electromagnetically or via strong force



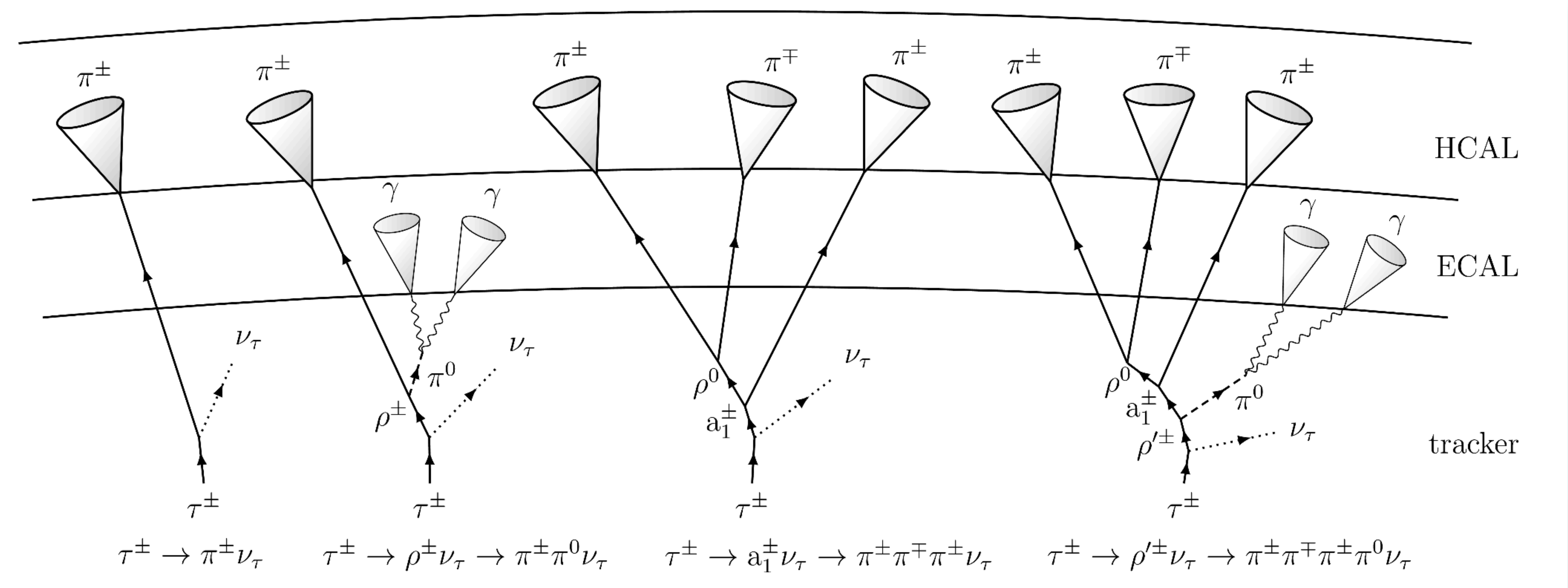
Our task

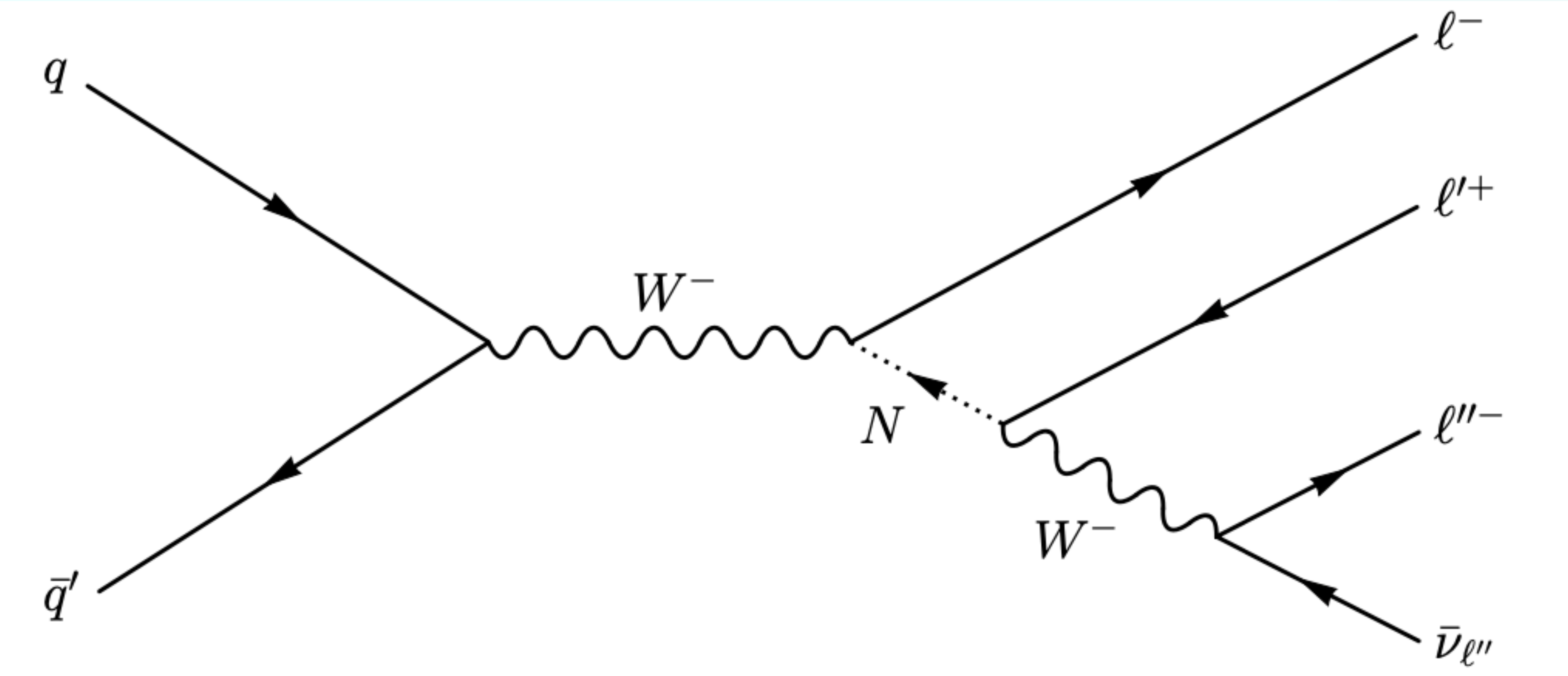
- Study prompt decays of HNL
  - Kinematic signatures
- Distinguish HNLs and background SM processes with similar signatures in the detector.
- Use processed 2018 CMS data



Focus on τ Leptons and HNLs

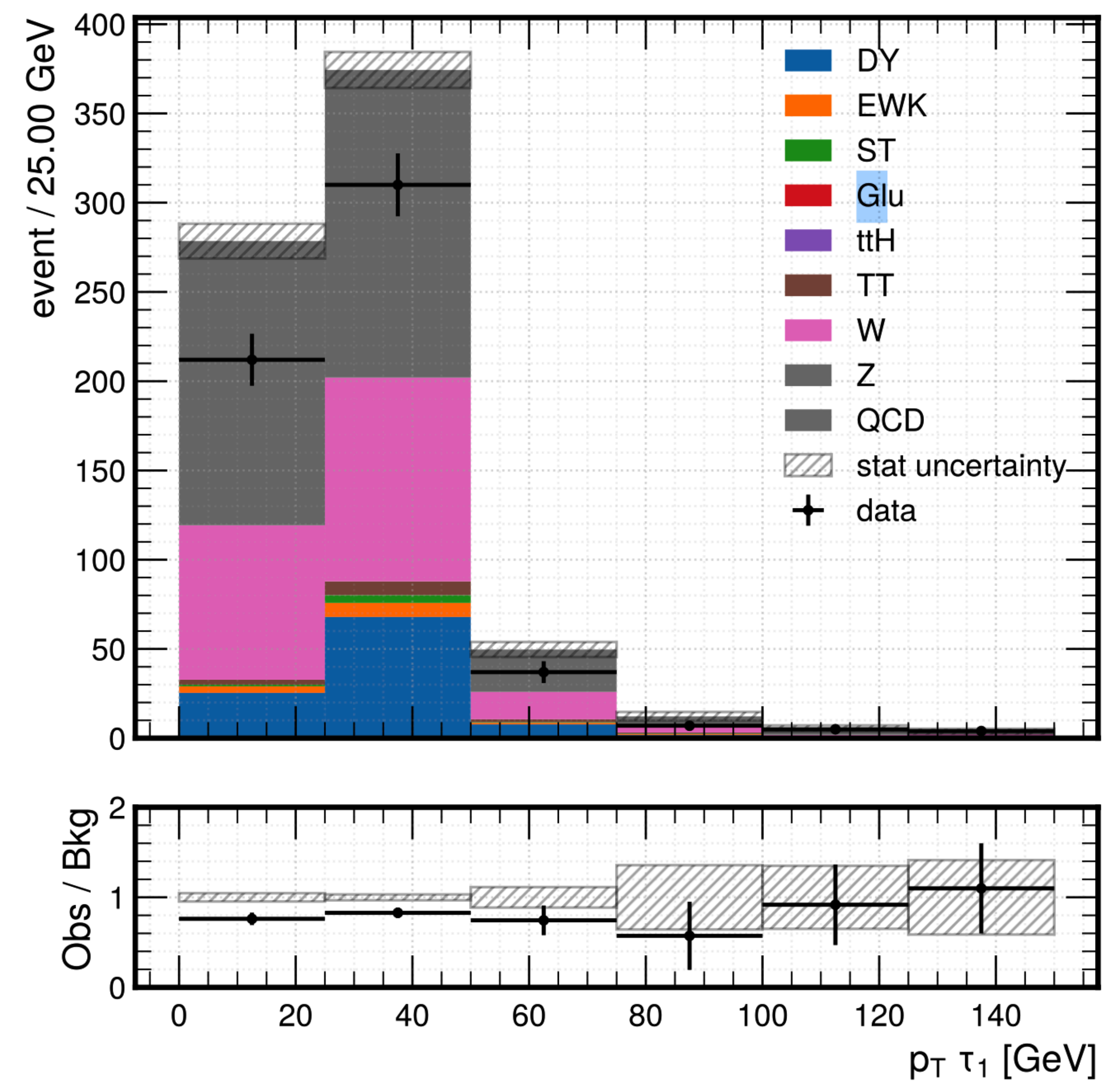
- Focus
  - HNLs to 3 lepton decay
  - Targeting |VτN|
- Importance of Hadronic τ's
  - 65% decay into hadrons
- CMS sensors
  - Tracker for charged particles
  - ECAL for e & γ
  - Hadronic calorimeter for hadrons

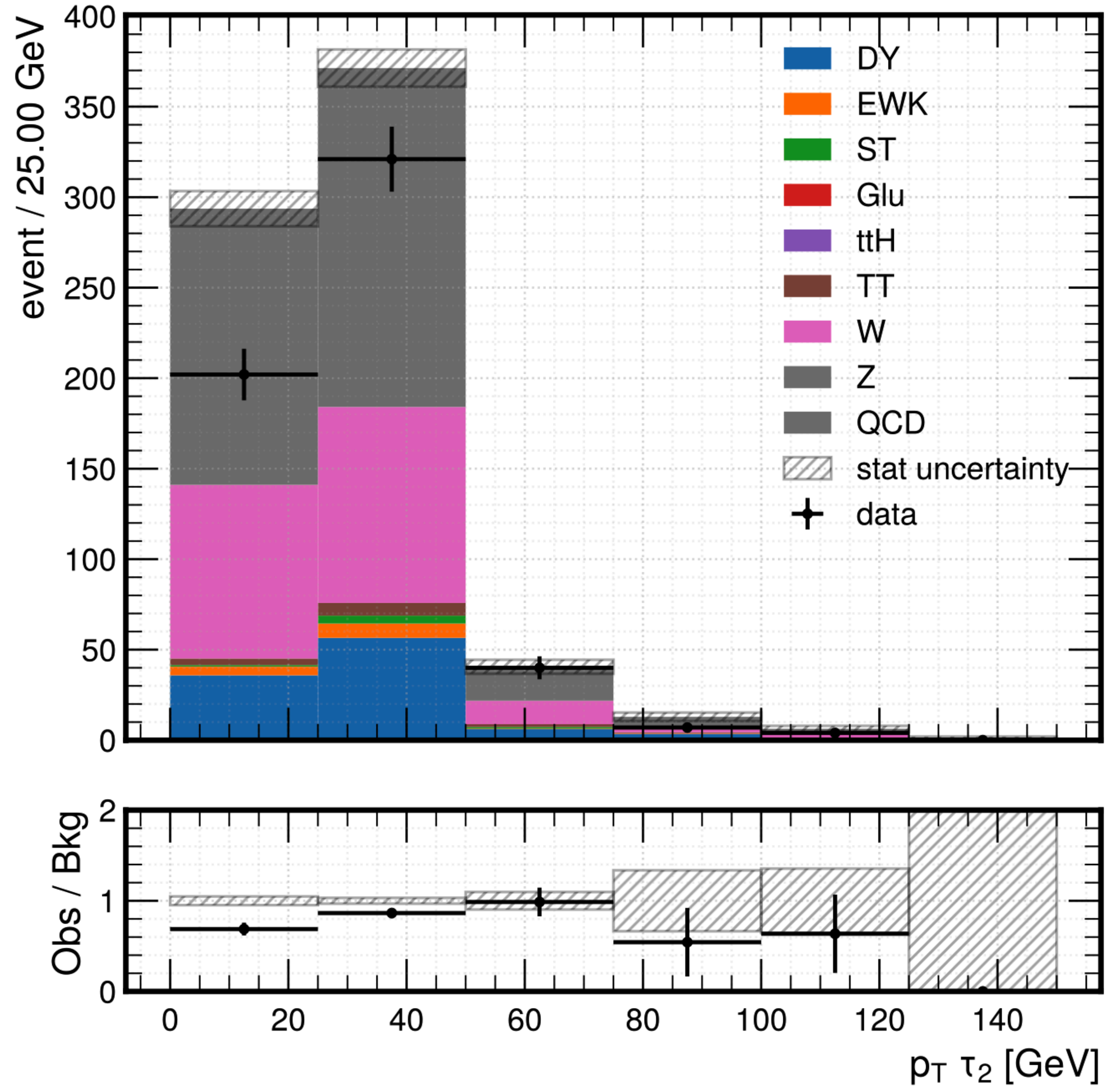


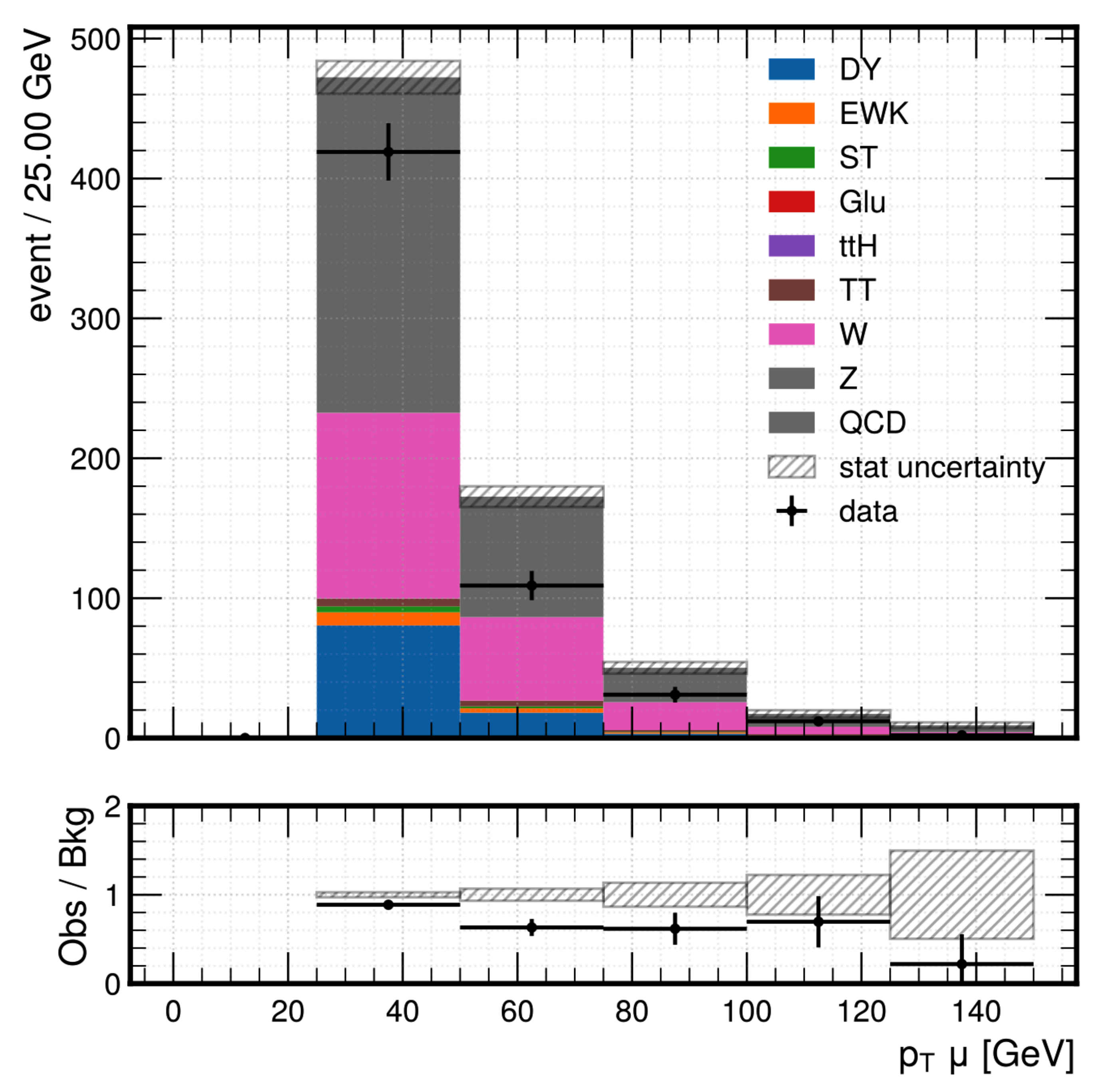


Event Preselection & Channels

- Event reconstruction
  - 3 well reconstructed isolated leptons
  - $\Delta R \leq 0.5$
- Channels and triggers
  - 5 channels: $ll'l'' \in \{\tau\tau\mu, \tau\tau e, \tau\mu\mu, \tau\mu e, \tau ee\}$
- Data preprocessing
  - Deep Tau discriminator score
- Lepton requirements
  - $p_T^{e,\mu} > 10$ GeV
  - $p_T^{\tau} > 20$ GeV



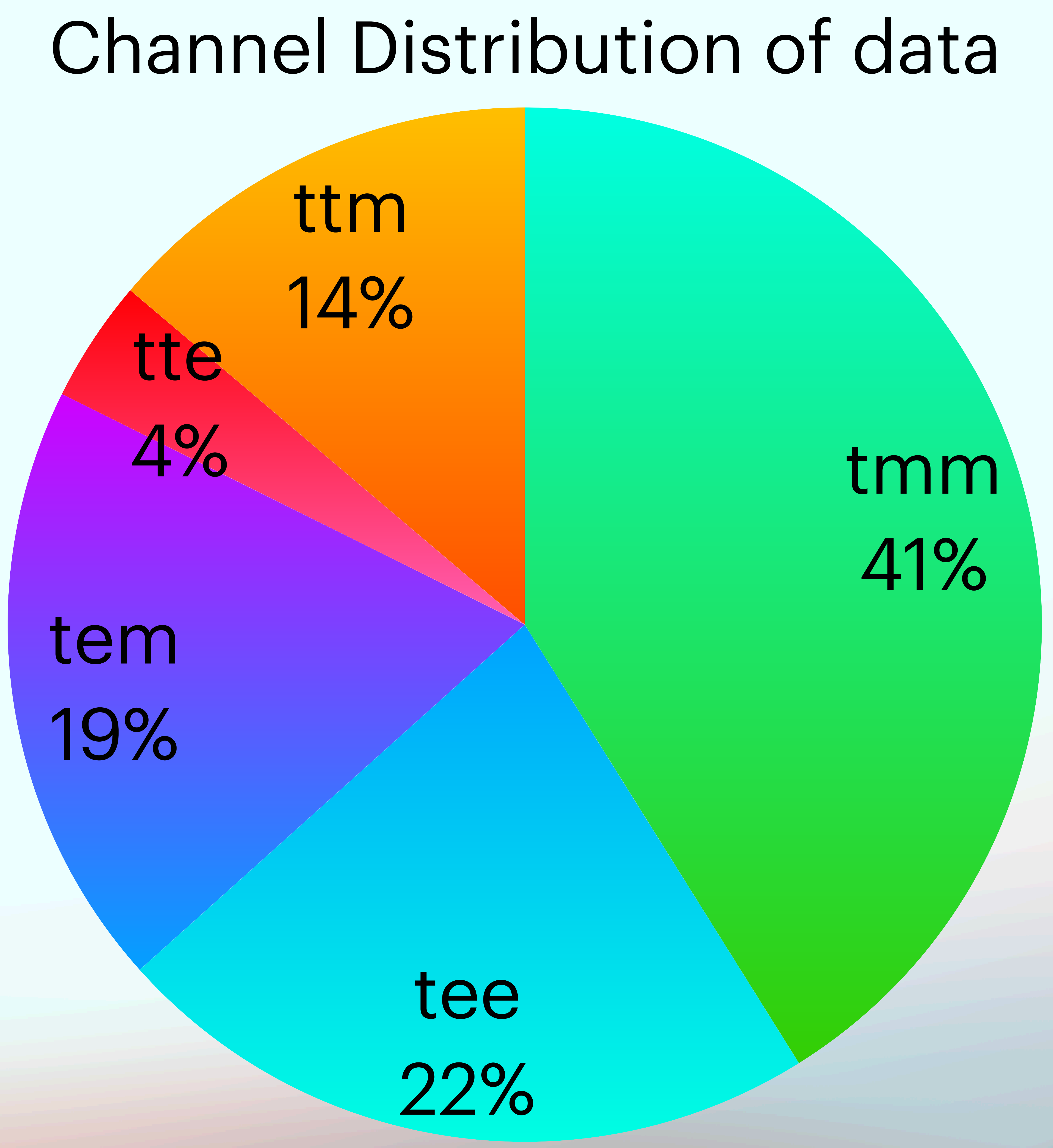




Distribution p_T for $\tau\tau\mu$ channel
From Luca Hartman report

Dataset

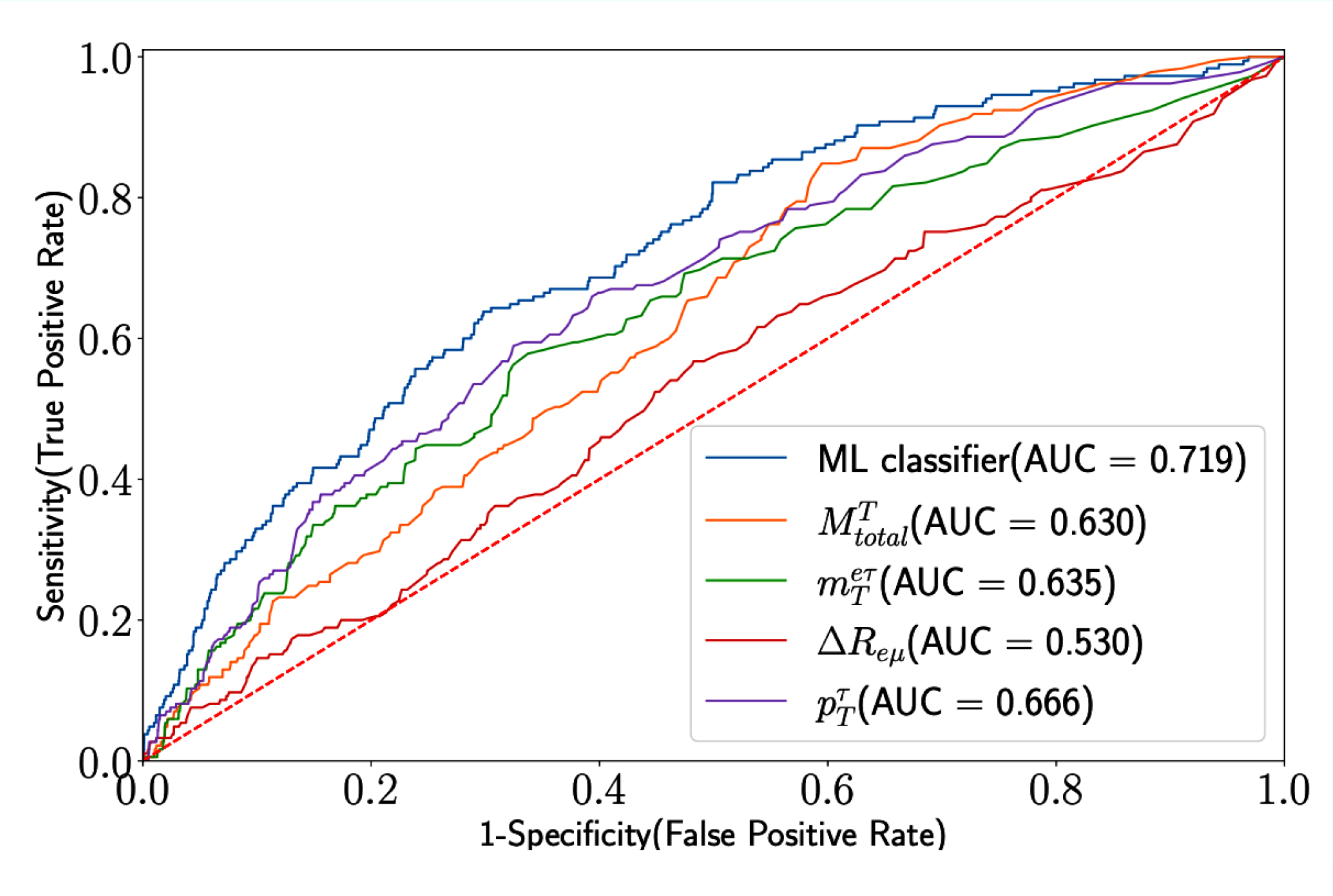
- HNL Mass Hypothesis
  - $m_{\text{HNL}}^{\text{hyp}} \in [85, 1000]$ GeV
- Data Generation
  - MadGraph, Pythia, Geant4
- Data Size and Cuts
  - 216k signal events, 1.4M background
- Potential Challenges
  - Risk of overfitting
- Variable and Feature
  - m, p_T, ϕ, η
  - Weight, channel, $m_{\text{HNL}}^{\text{hyp}}$



Previous Work

Lucas Mollier

- Machine Learning algorithm
  - XGBoost
- Training Approach
  - Classifiers for each mass and channel
- Inputs:
  - 40 classical observables



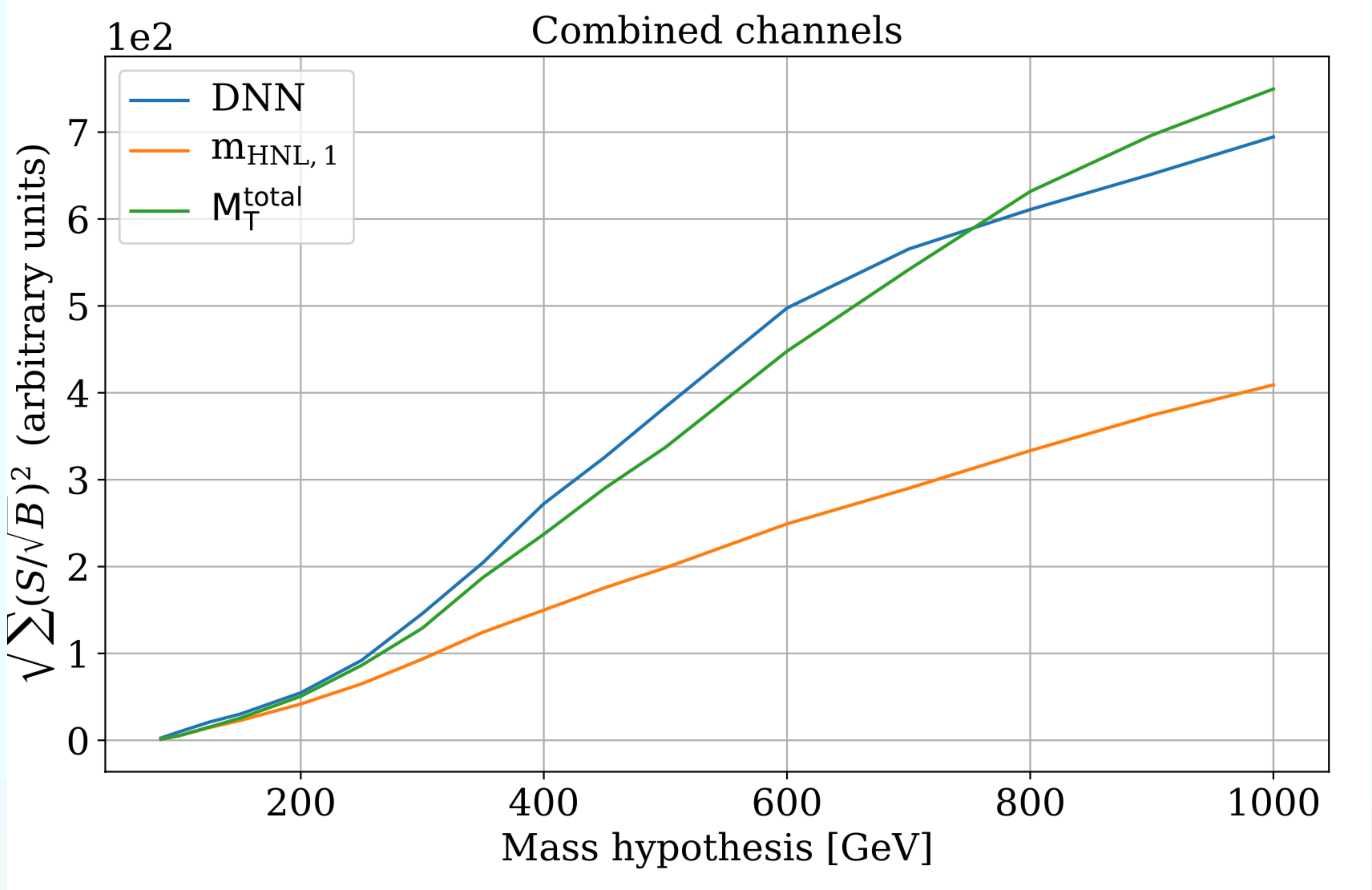
ROC curves and AUC values for HNL mass 250 GeV

Classical observables: calculated + raw kinematic variables ($m, \Delta R, M_T^{tot}$, etc...)

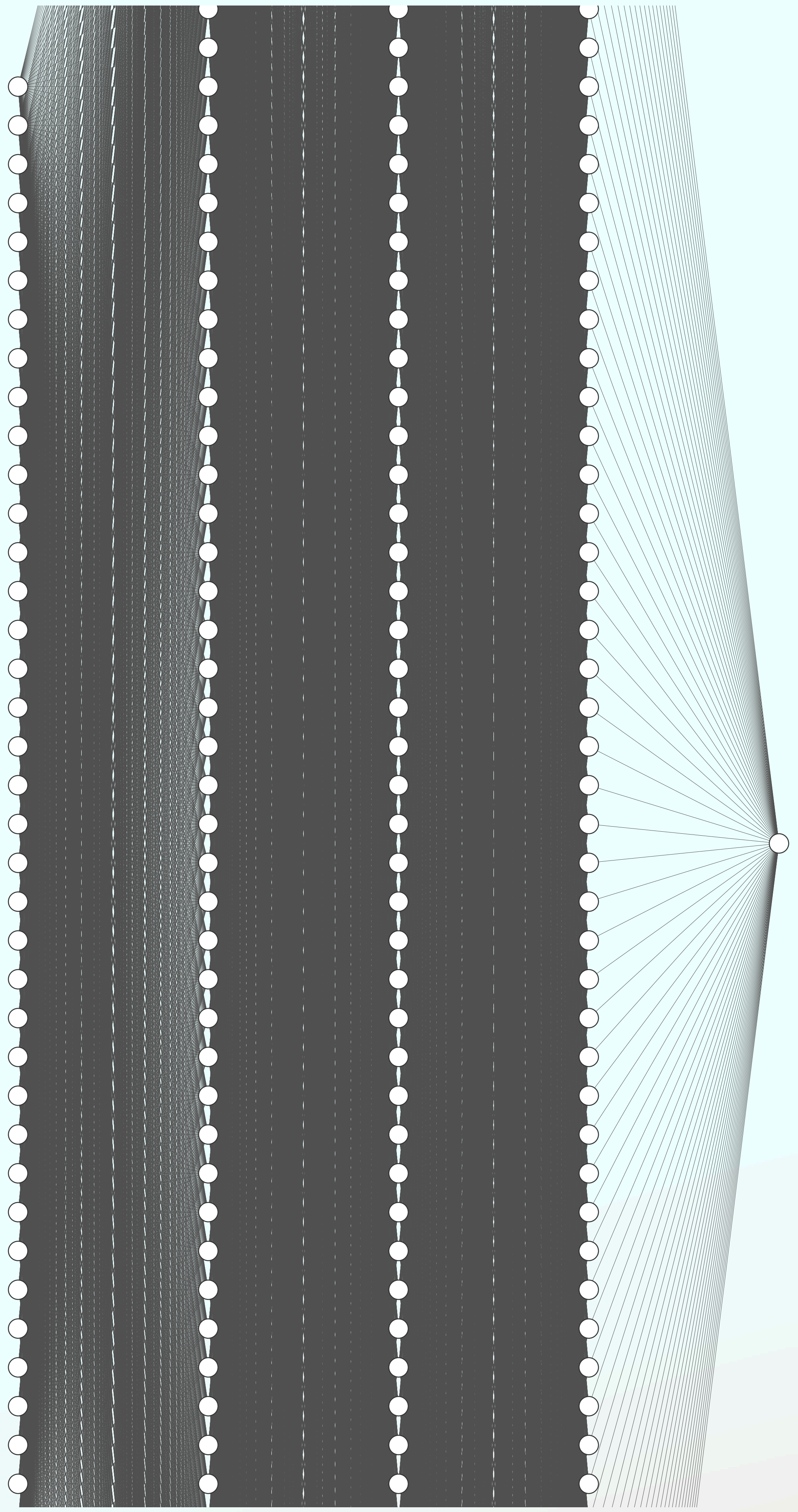
Previous Work

Nelson Glardon

- Machine Learning algorithm
  - Deep Neural Network
- Input:
  - 85 input features
- Training Approach
  - One classifier for all channels and m_{hyp}



Significance Estimator for DNN score



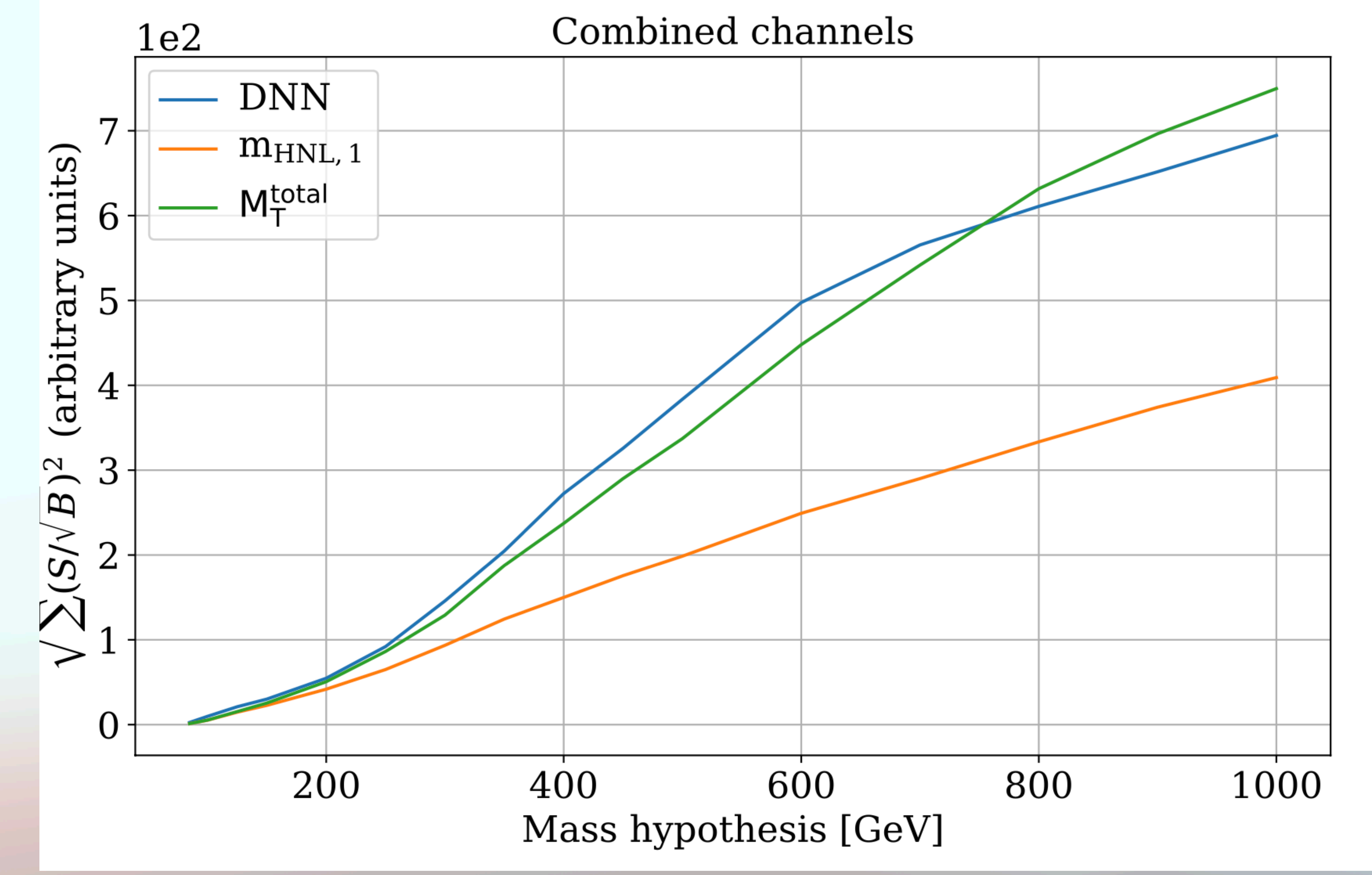
Visualization of the DNN model

Previous Work

Nelson Glardon

- Machine Learning algorithm
  - Deep Neural Network
- Input:
  - 85 input features
- Training Approach
  - One classifier for all channels and m_{hyp}

- Best Model Specifications:
  - Input: 29 features
  - Depth: 3
  - Width: 58
  - Optimizer: Adam
  - Dropout: 0.2

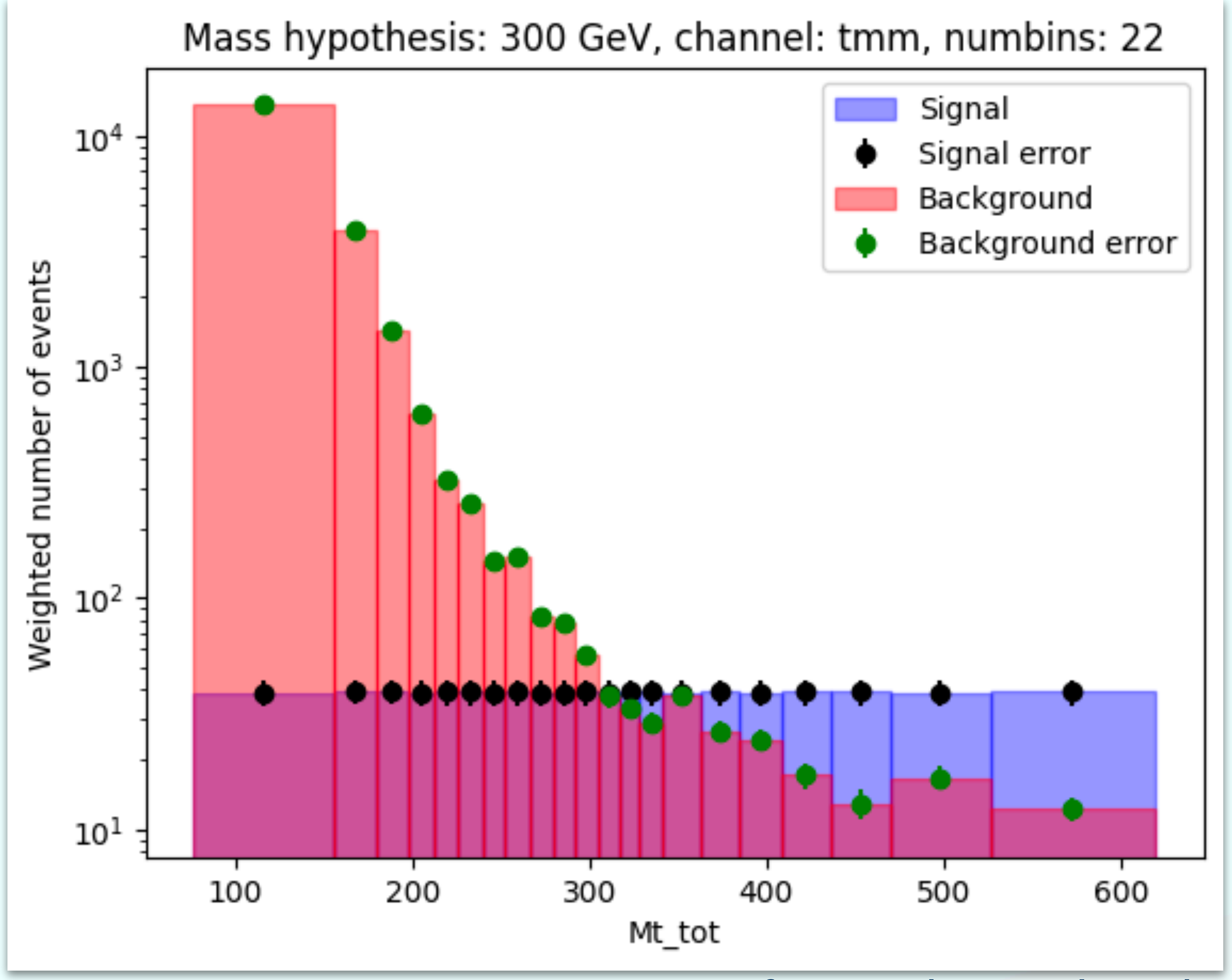


Significance Estimator for DNN score

Model comparison

Histograms

- Objective
  - Have a clear metric to compare various models and features at different mass hypotheses
- Histograms
  - Model scores (or features) vs Event count at a specific m_{hyp}
- Statistical Certainty
  - Relative weighted uncertainty for each bin
  - $\sqrt{\frac{\Sigma w^2}{\Sigma w}} < 0.15$

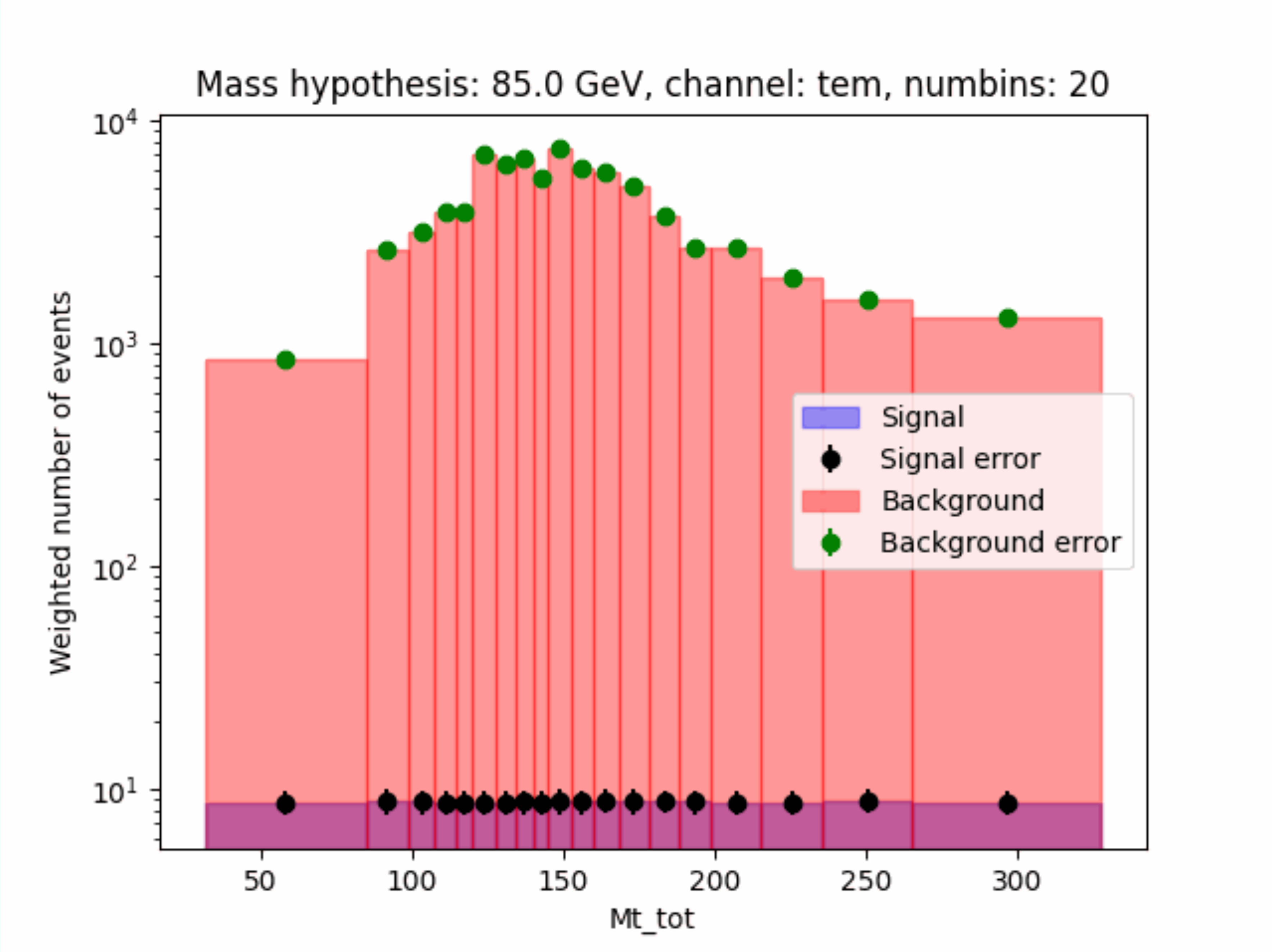


Constant-signal histogram of $\tau\mu\mu$ channel and m_{hyp} 300 GeV

Model comparison

Histogram Binning

- Objective
  - Have a clear metric to compare various models and features
- Histograms
  - Model scores (or features) vs Event count
- Statistical Certainty
  - Relative weighted uncertainty for each bin
  - $\sqrt{\frac{\Sigma w^2}{\Sigma w}} < 0.15$



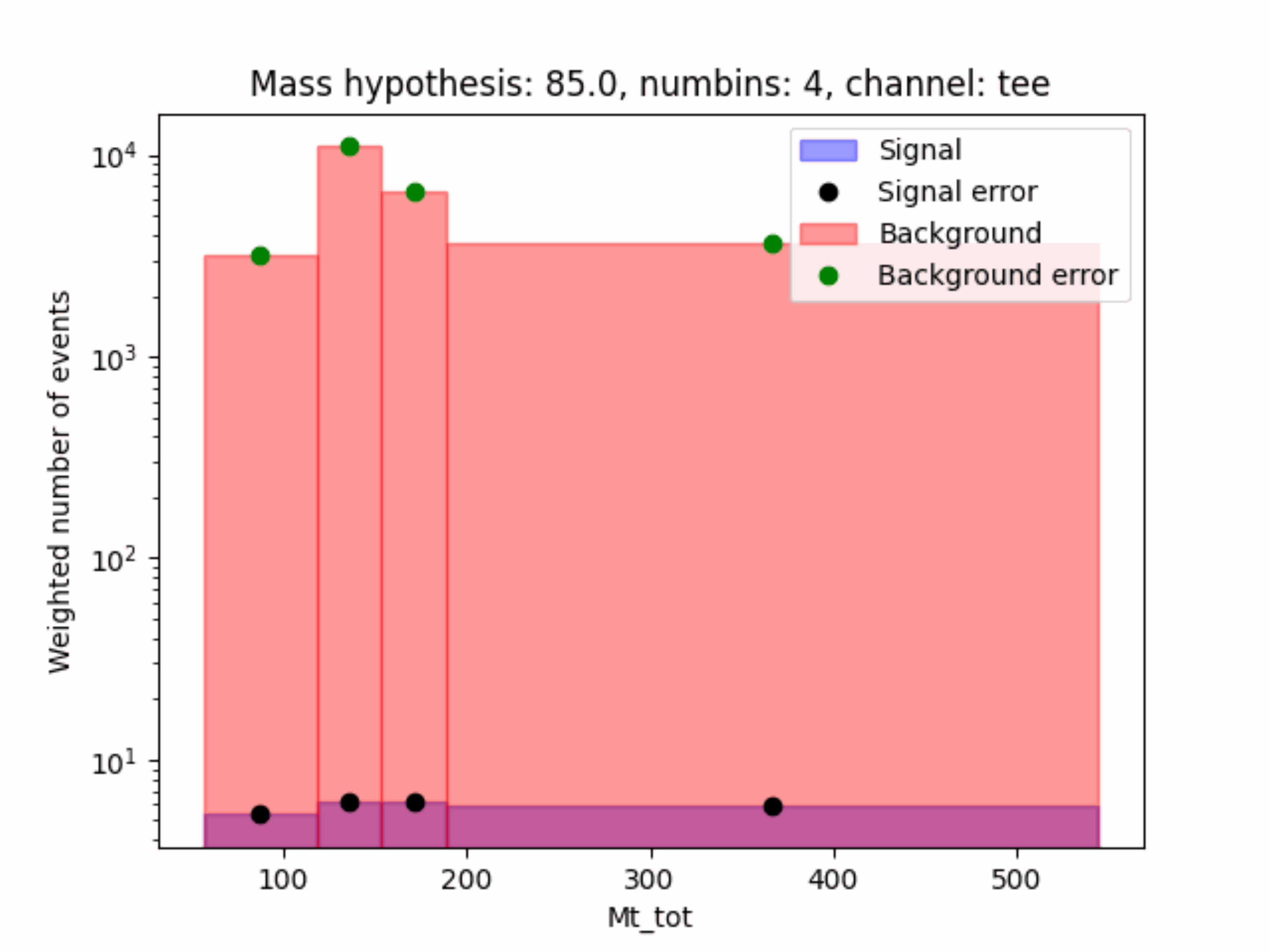
Constant-signal histogram of $\tau e \mu$ channel of M_t^{tot} for different m_{hyp}

- Constant-signal histogram
  - Left to right
  - Signal heigh stays the same
  - Easy to compare at wide range of x values

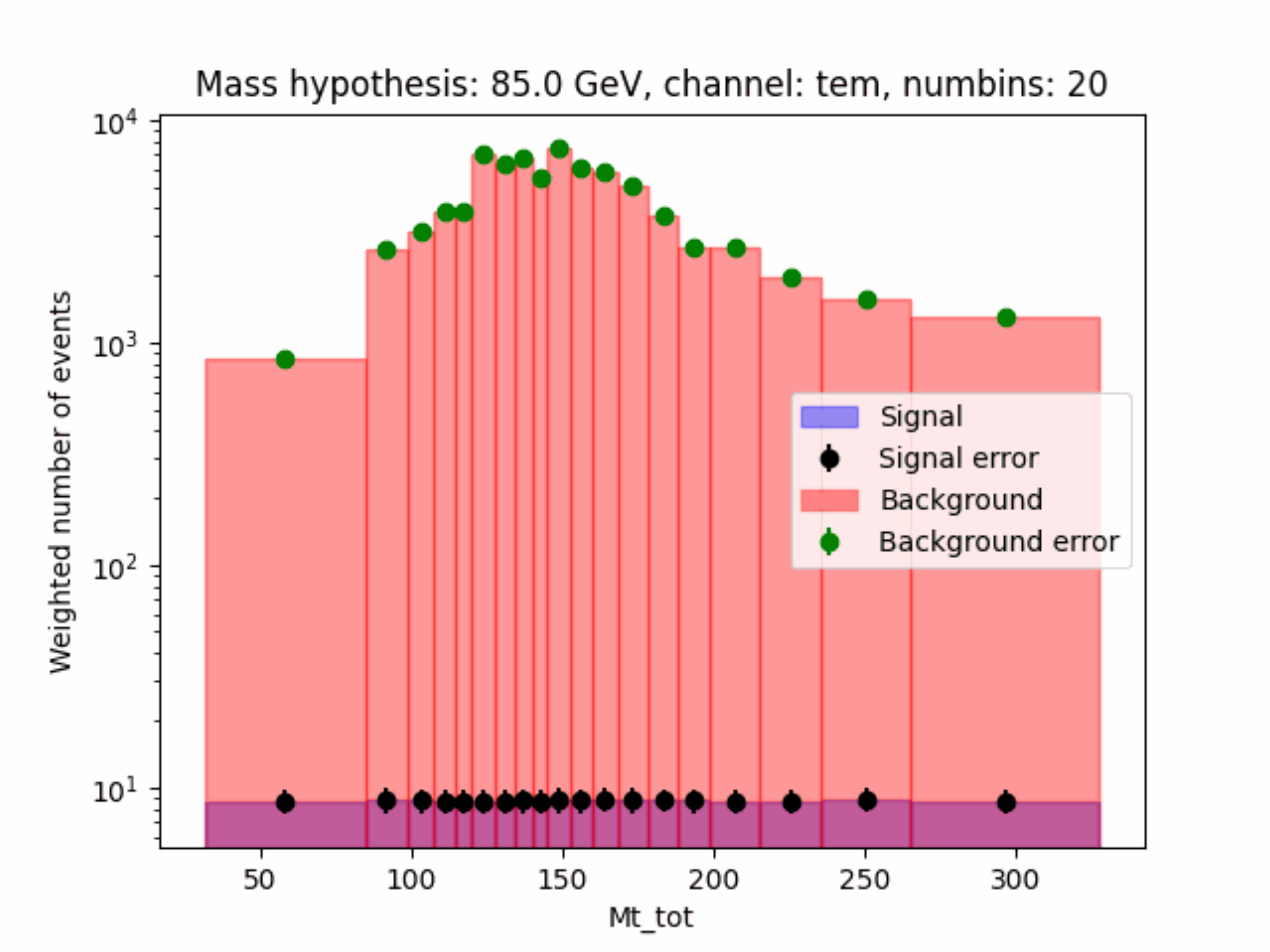
Model comparison

Histogram Binning

- increasing-signal histogram
  - Right to left
  - Signal height increases
  - Better comparison at high x-values



Increasing-signal histogram of $\tau e \mu$ channel of M_t^{tot} for different m_{hyp}



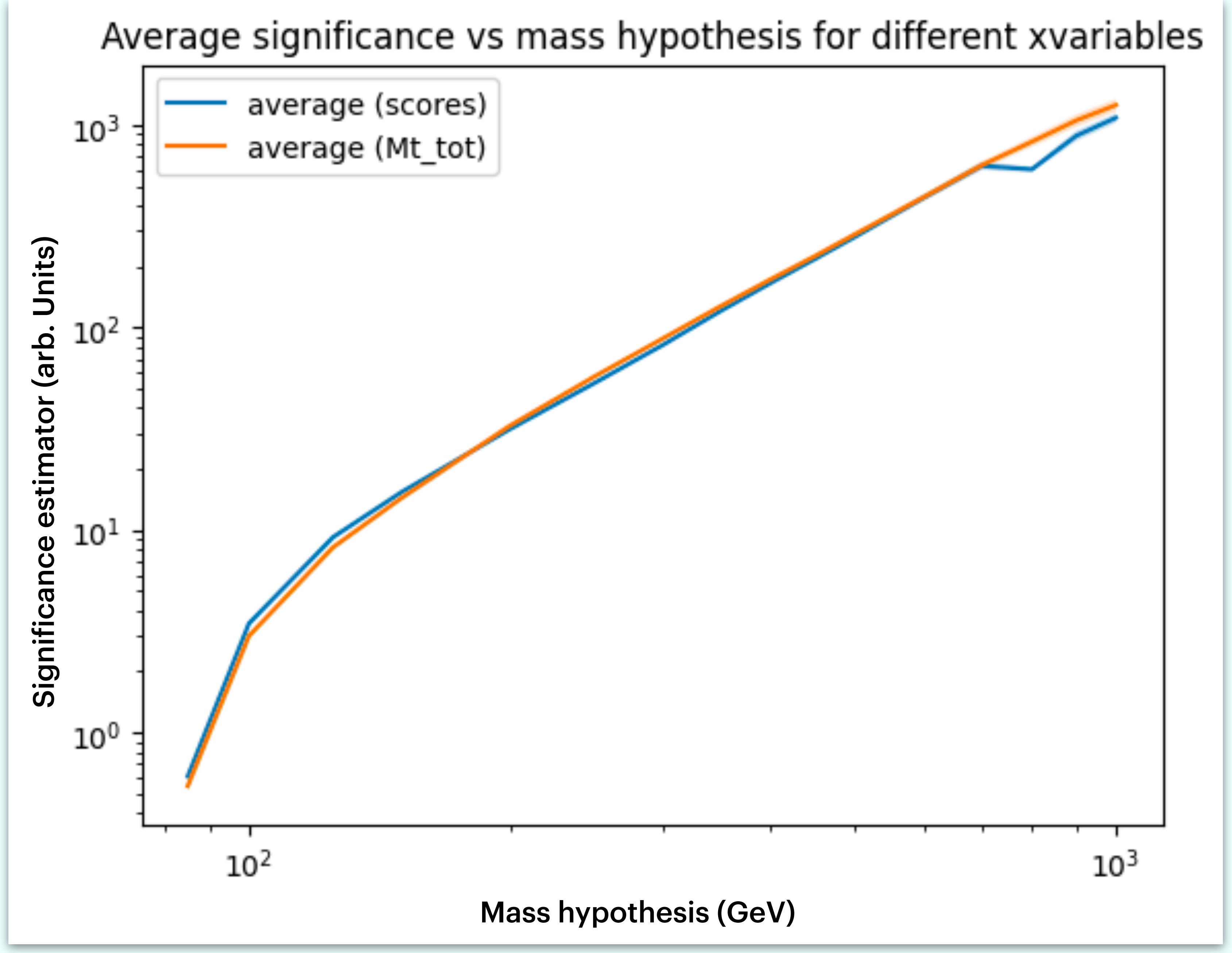
Constant-signal histogram of $\tau e \mu$ channel of M_t^{tot} for different m_{hyp}

- Constant-signal histogram
  - Left to right
  - Signal heigh stays the same
  - Easy to compare at wide range of x values

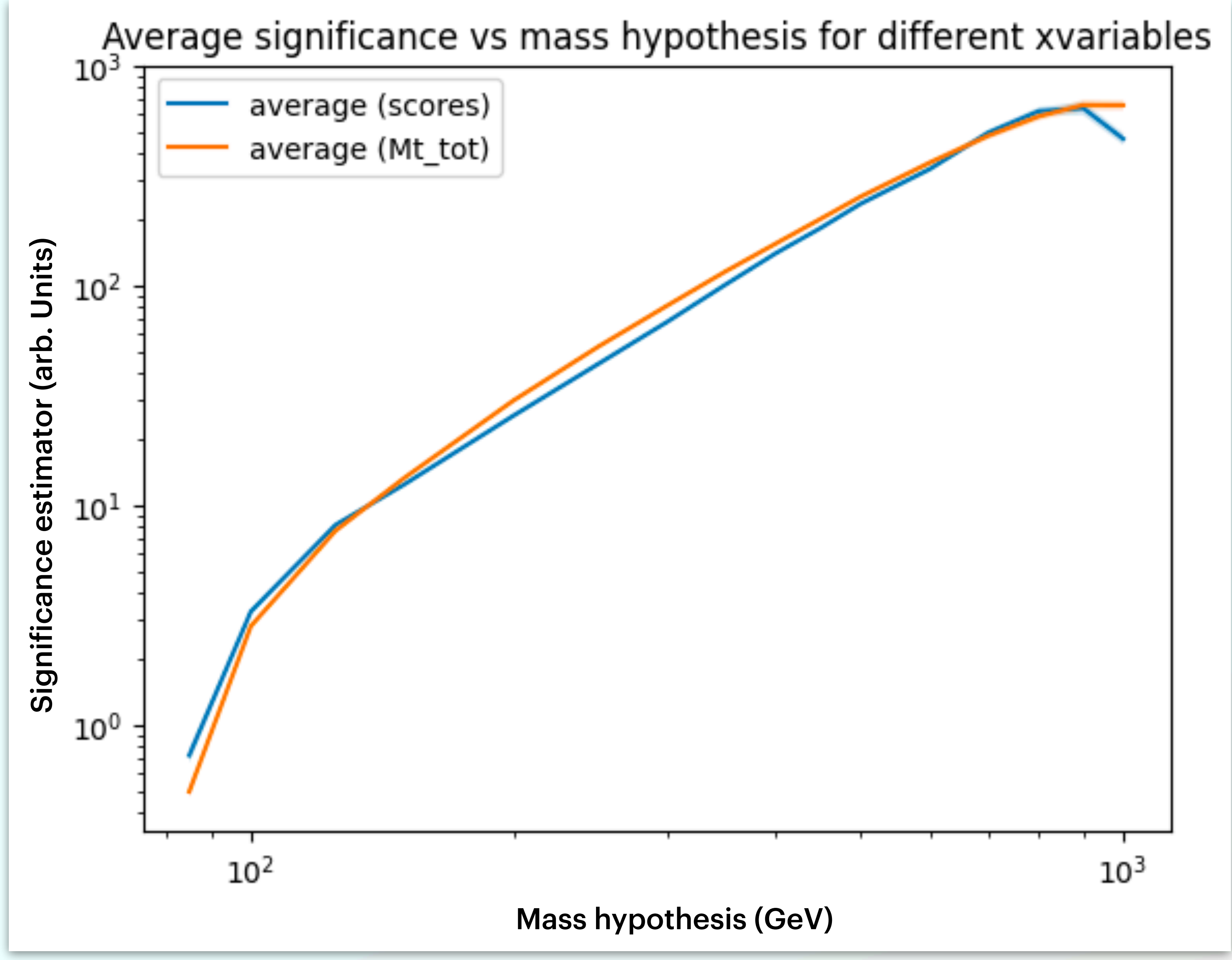
Model comparison

Significance plotting

- Standard Formula for Significance
  - $\text{Sig} = \dfrac{S}{\sqrt{B}}$ with custom bins
  - Sig. has arbitrary units
- Significance plot structure
  - X-axis: Mass Hypothesis
  - Y-axis: Average of significance scores



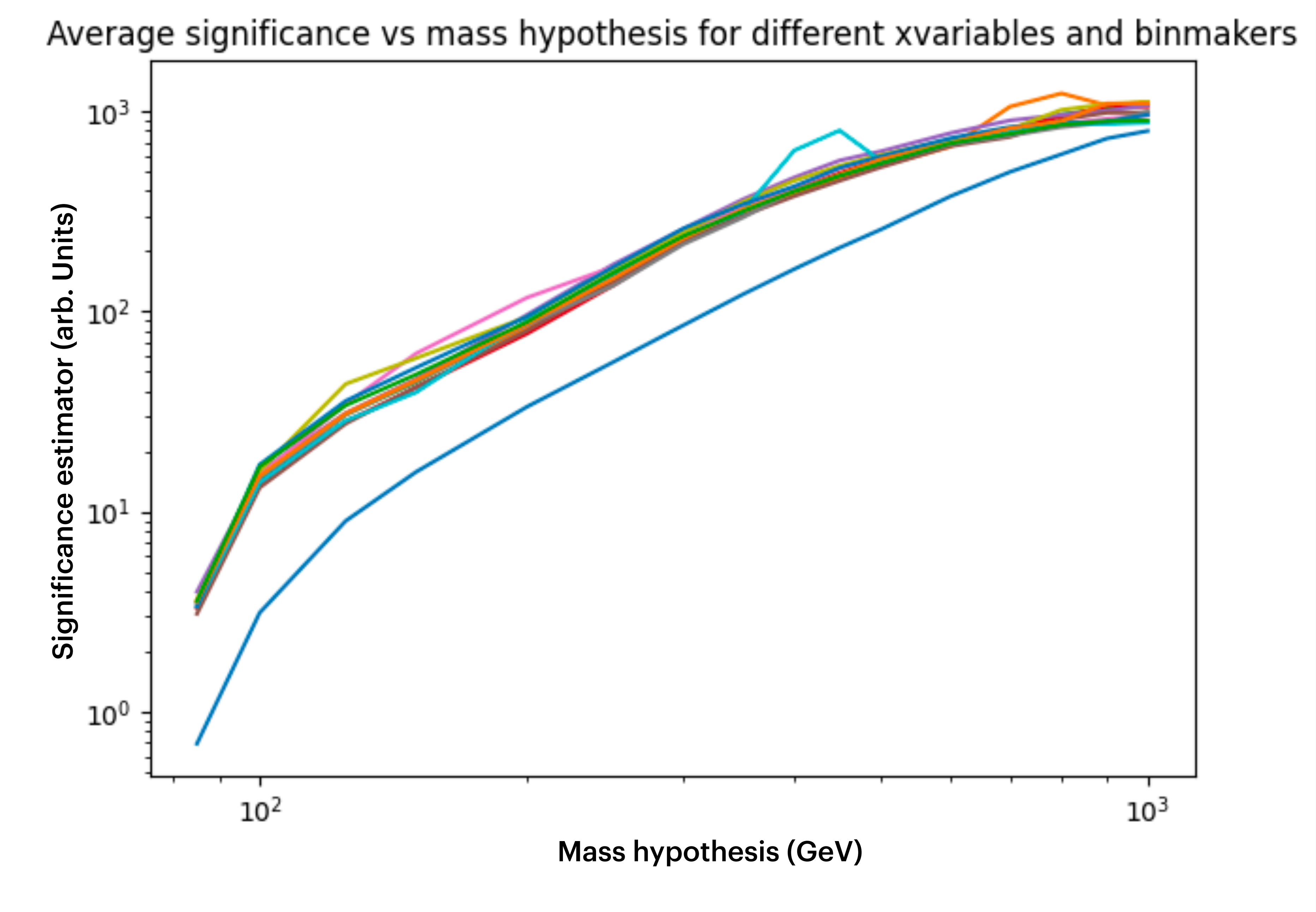
Increasing-Signal binning



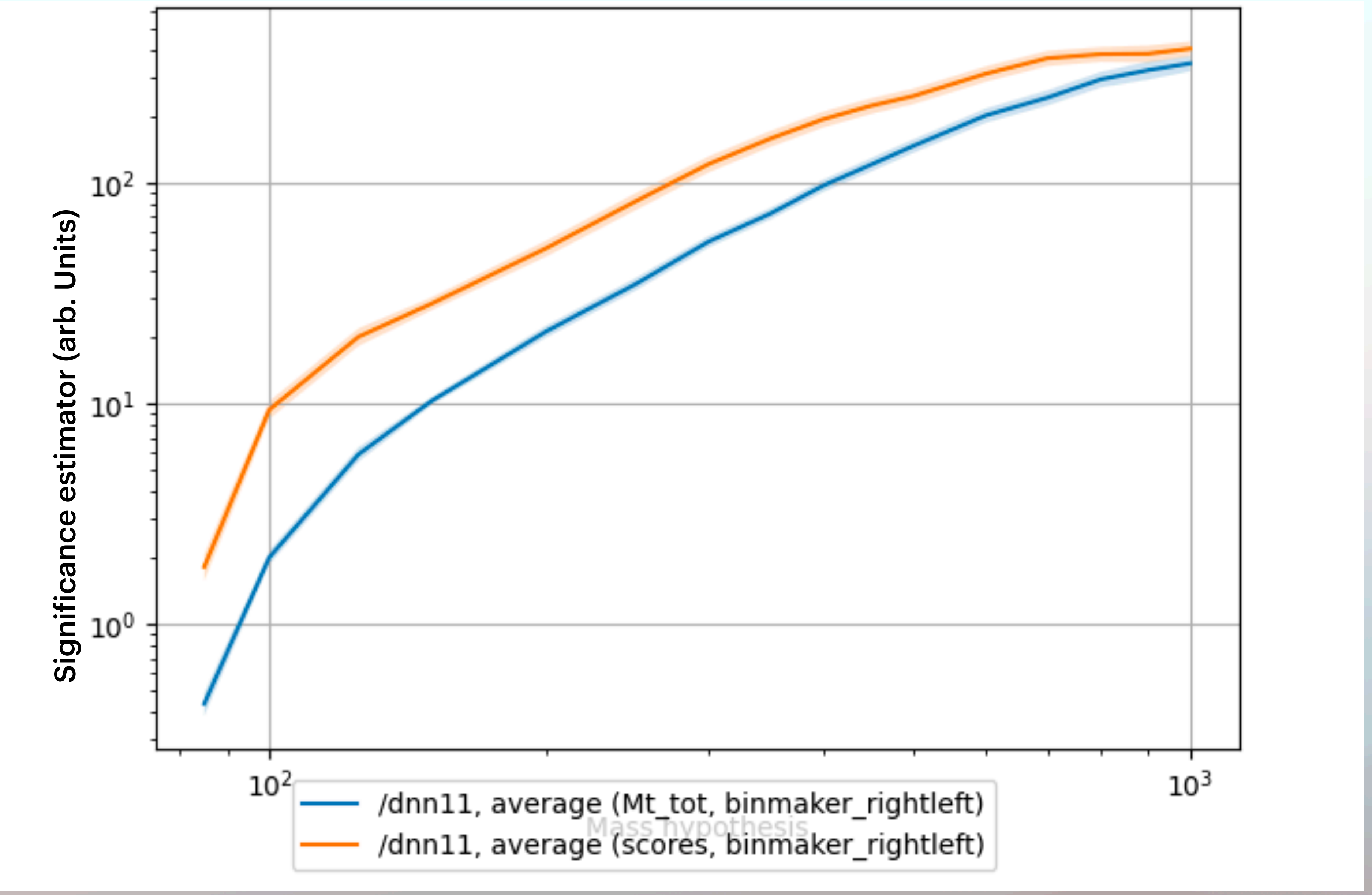
Constant Signal binning

DNN Training

* Initial goal
  * Beat m_T^{tot} across all mass hypotheses
* Methodology
  * Normalize inputs
  * Try different depth and width combinations
* Findings
  * Best model: 85 features, 2 layers [83, 30]



Comparing different models



Best Model

Transfer Learning

Introduction

DNN Kinematic feature

Give a ~~man~~ a ~~fish~~, and you feed him for a day. Teach a ~~man~~ to ~~fish~~, and you feed him for a lifetime.

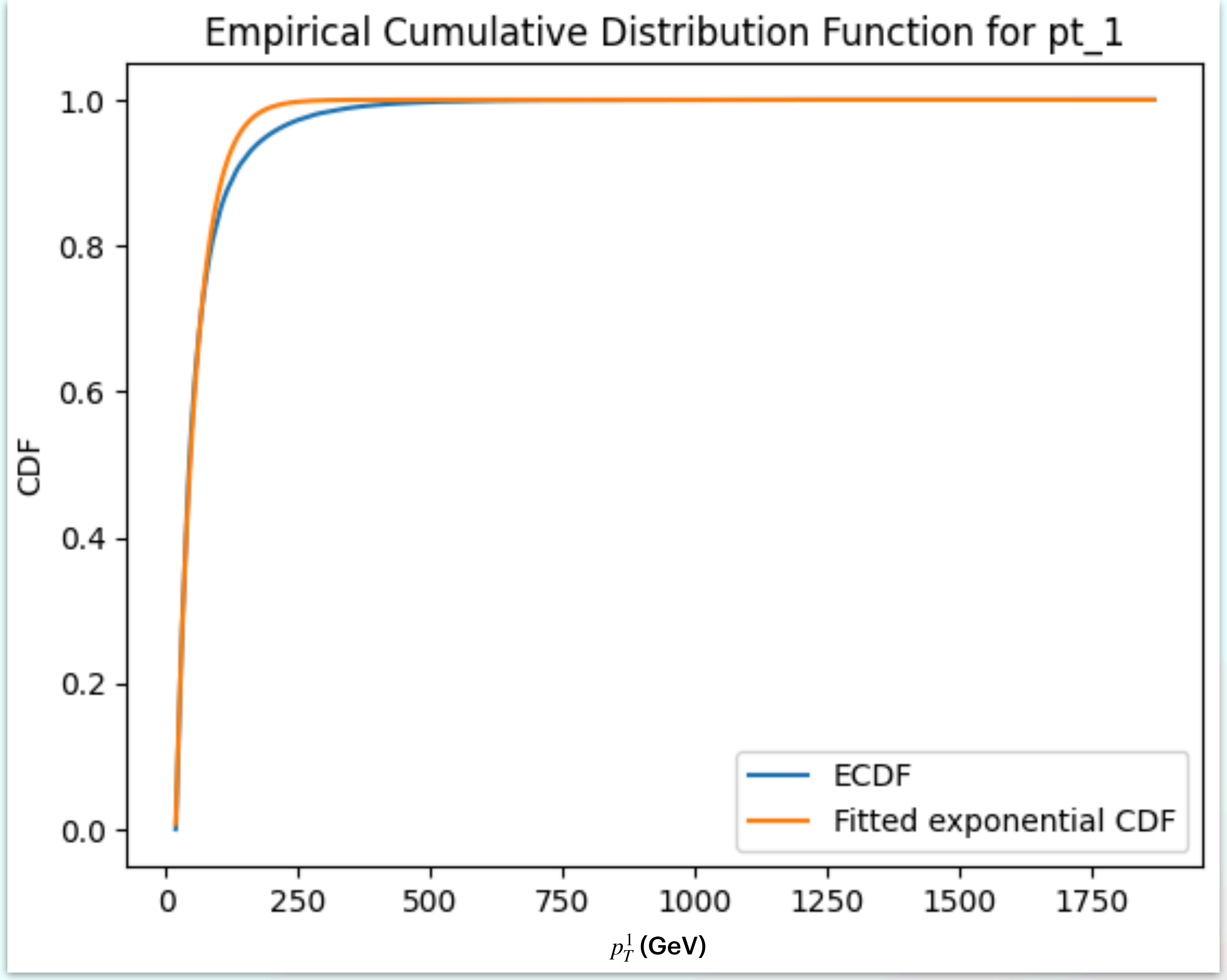
DNN Kinematic feature

- Data Size Issue
- New Approach
  - **Transfer Learning**
- Regression DNN: predict calculated kinematic features
- Classification DNN: use regression model as input
- Advantages
  - Infinite synthetic event generation

Transfer Learning

Data Generation

- Additional output features
  - Mother particle kinematic values & E_{tot}
- Data cuts
  - Cut 0.03rd and 99.7th percentiles of real data
  - $\approx 27\,\%$ data removed
- Logical Limits
  - $\eta \in [-2.5, 2.5]$
  - $\phi \in [-\pi, \pi]$
  - p_T : exponential CDF



Fitted Exponential CDF

Transfer Learning

Regression Training

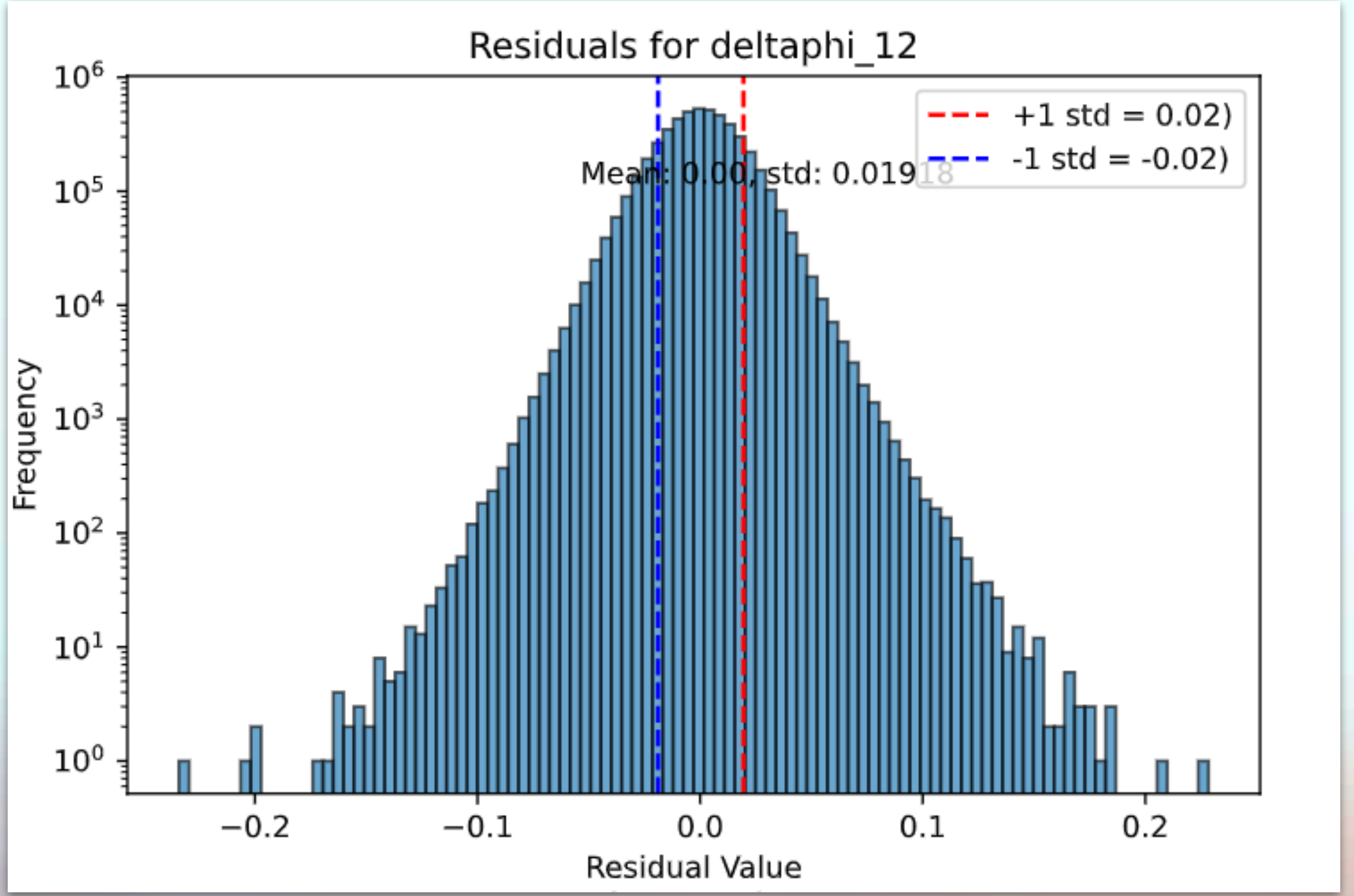
- Network
  - 1024 nodes
  - 25 layers
  - ~25M parameters
- Normalization
  - GeV vars divided by E_{tot}
- Training
  - New data every epoch
  - Loss function: MSE & Relative MSE
  - Optimizer: Adam + Decaying Learning Rate

Best model validation loss

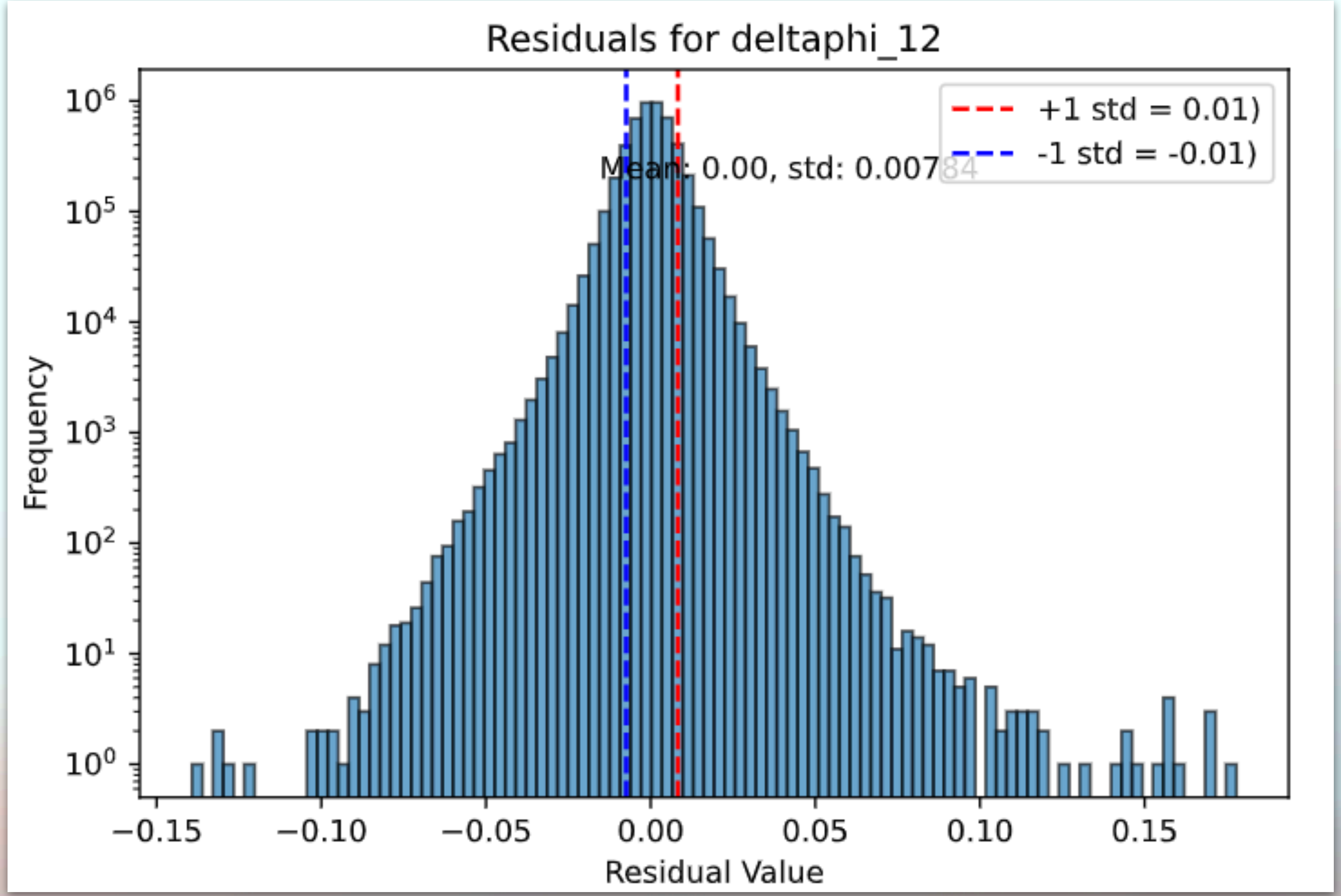
Transfer Learning

Drop-in technique

- Challenge
  - Losing input feature values
- Solution
  - “Drop-In”: Reintroduce inputs every 3 layers



Without Drop-Ins

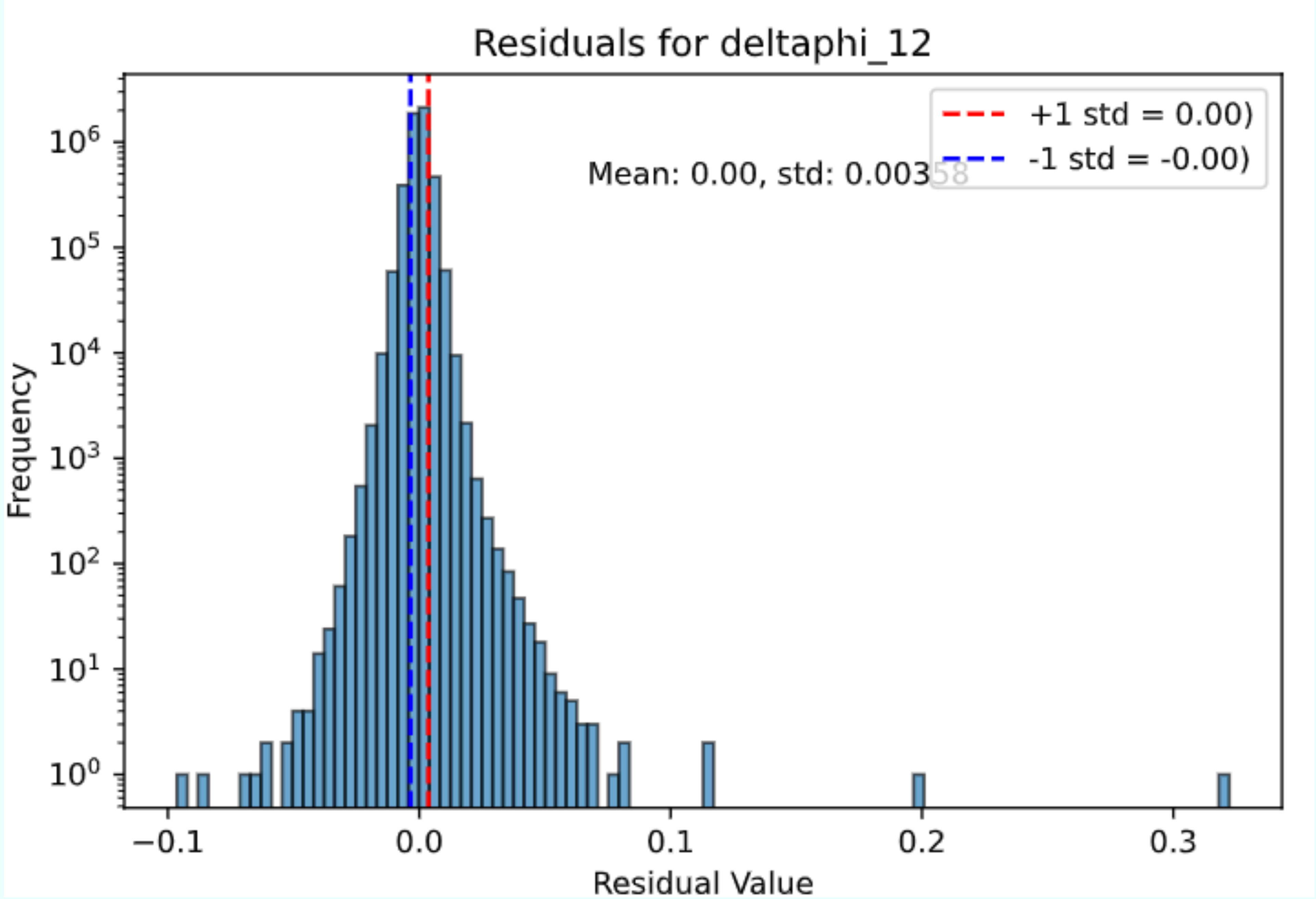


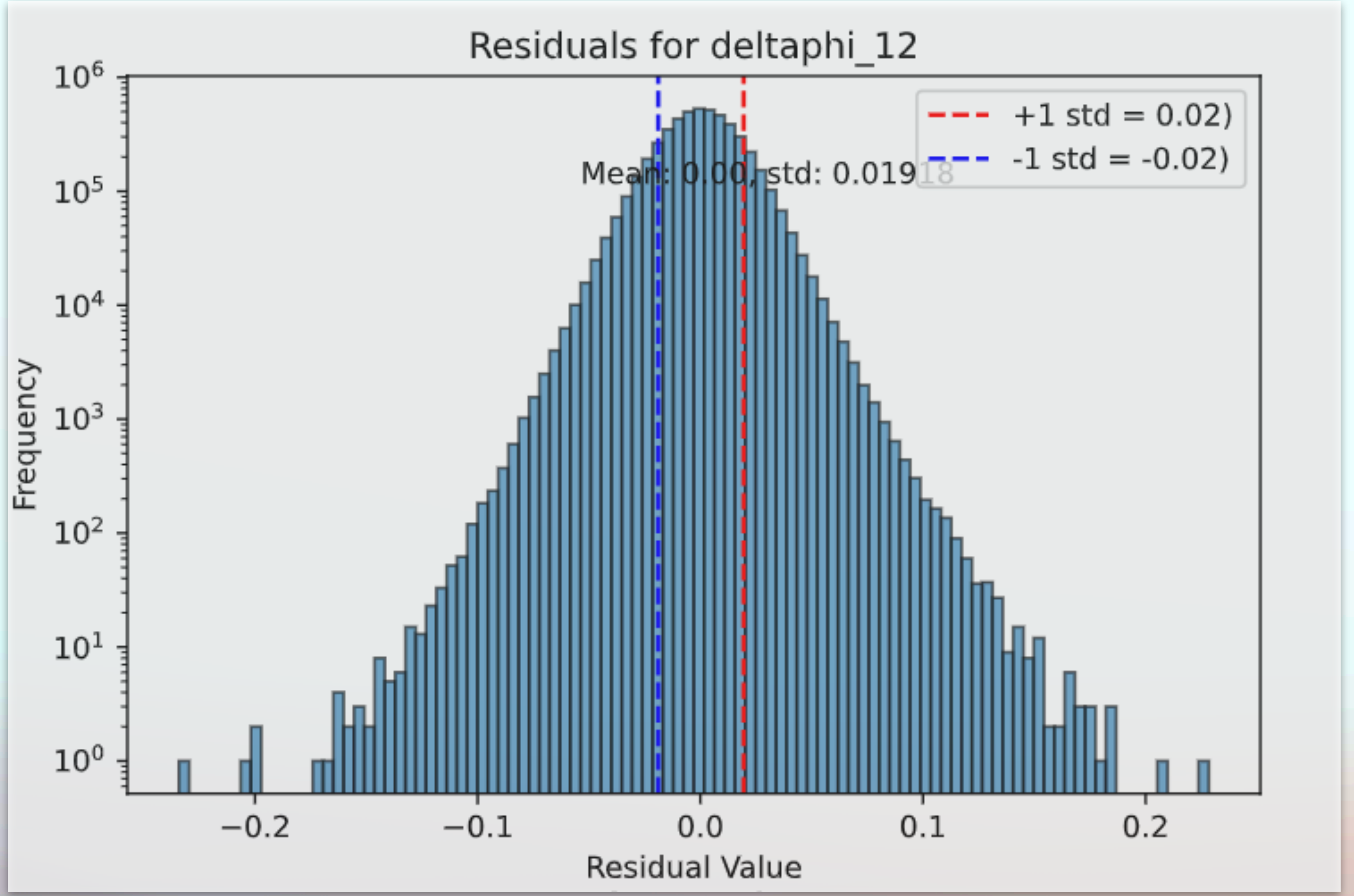
With Drop-Ins

Transfer Learning

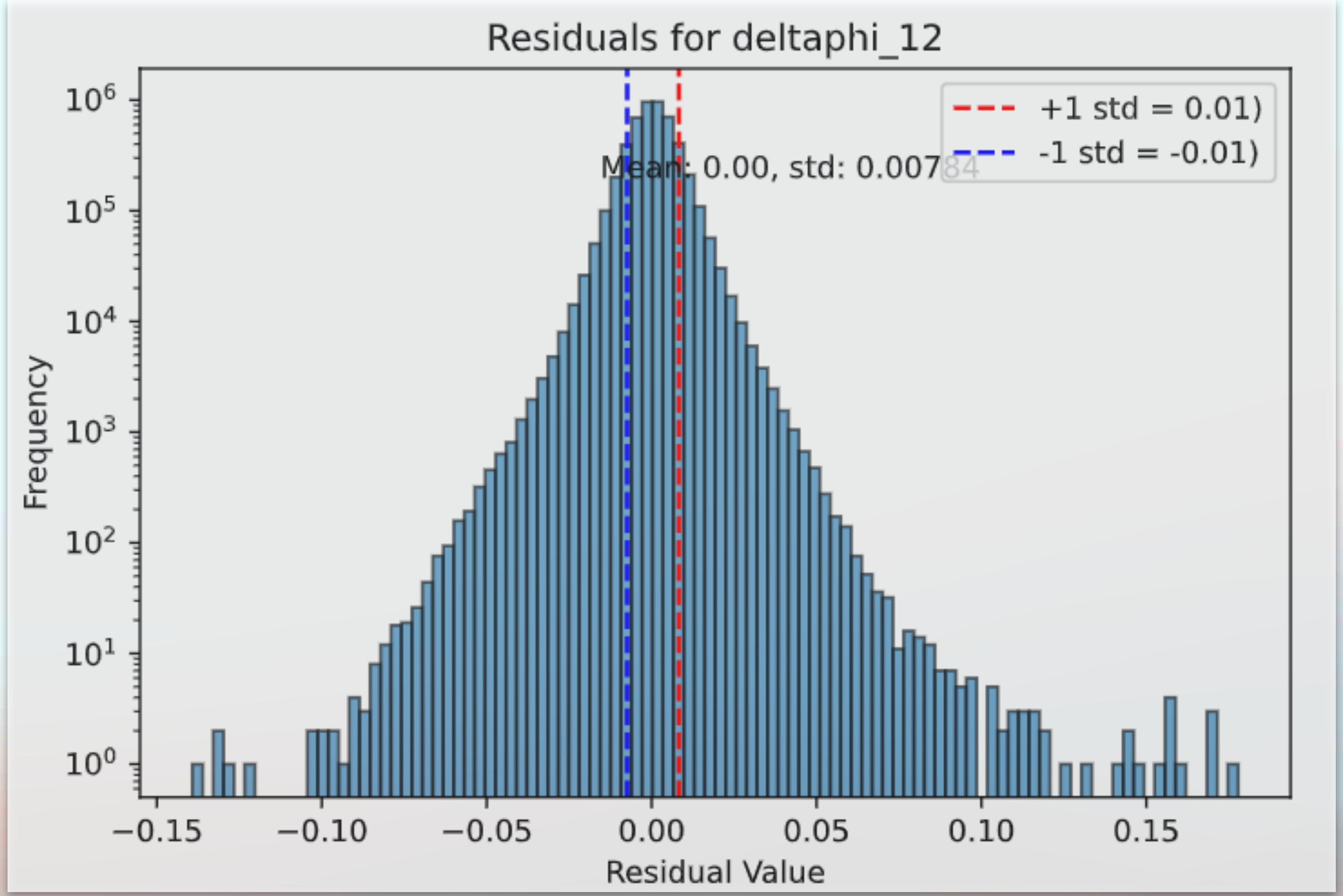
Drop-in technique

Best model:





Without Drop-Ins

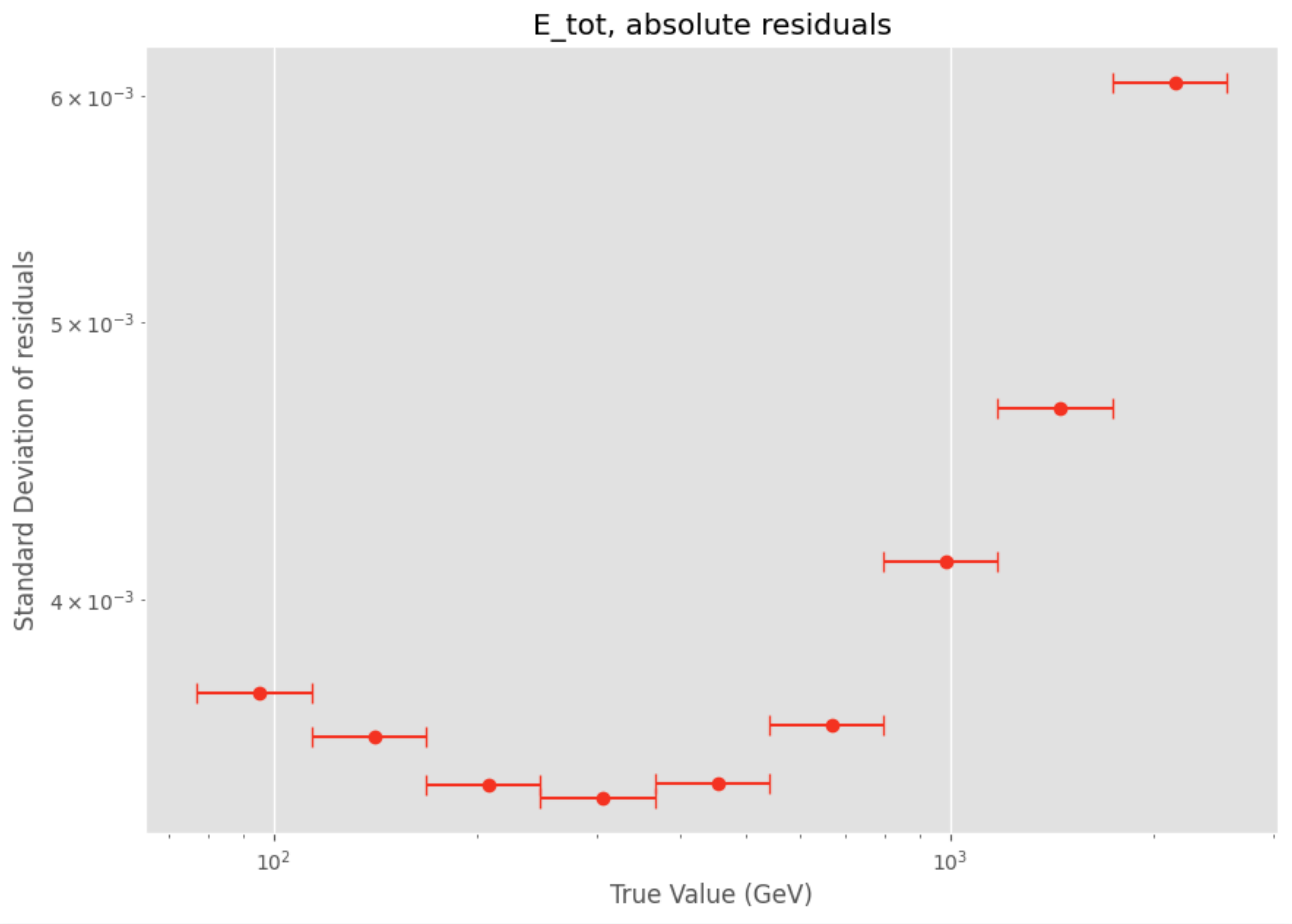


With Drop-Ins

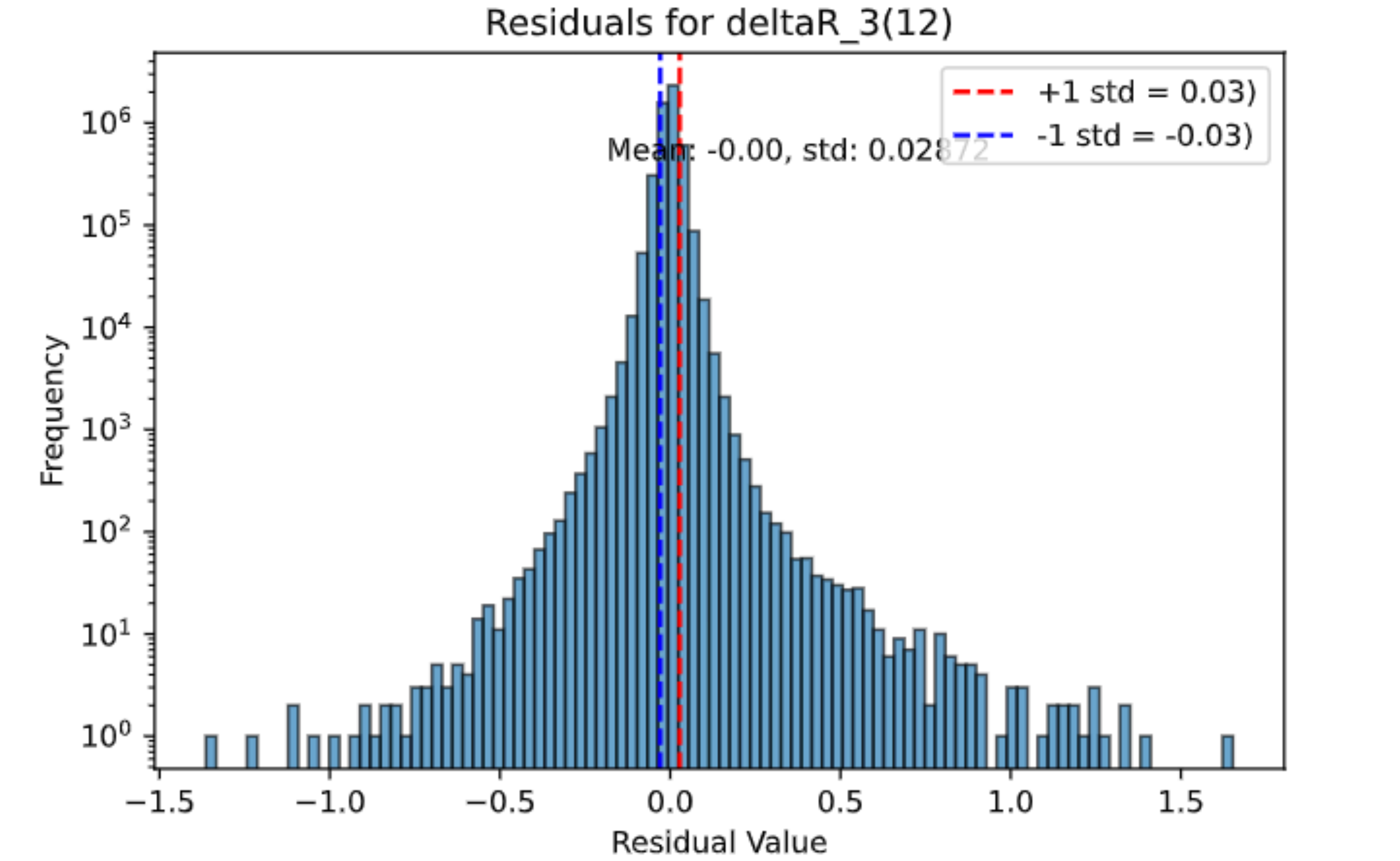
Transfer Learning

Regression Results

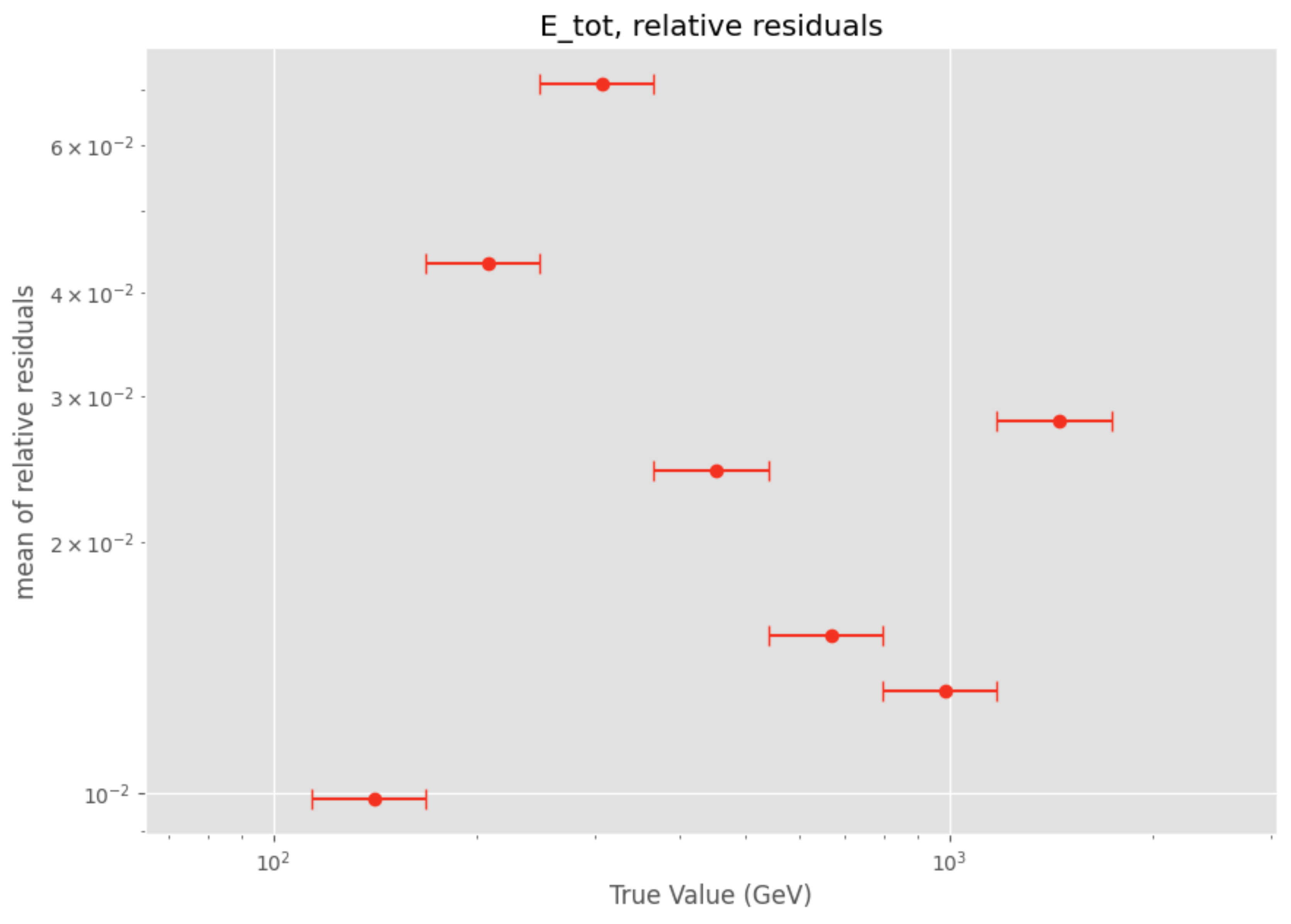
- Safety check
  - Make sure E_tot is being predicted well
- Challenges



Standard deviation bar plot of Absolute residual



Residual distribution ΔR_3 in frame 1,2

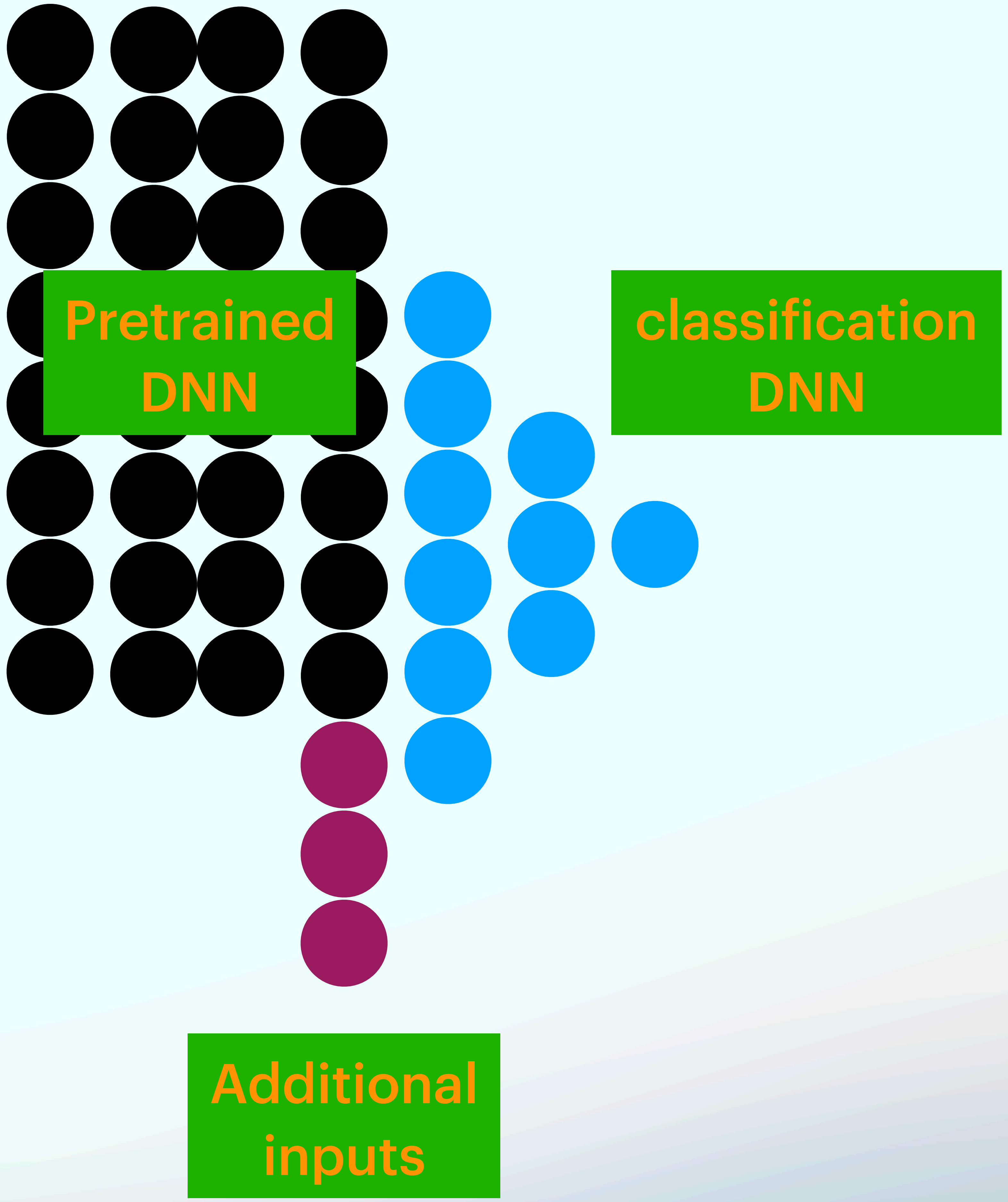


Mean relative residual bar plot

Transfer Learning

Classification!

- Start point
  - Best multivariate regression model (pretrained)
- Data Prep
  - Remove last output of pretrained DNN
- Additional Inputs
  - Channel, Mass Hypothesis, particle charges
- Classification DNN
  - Depth 3

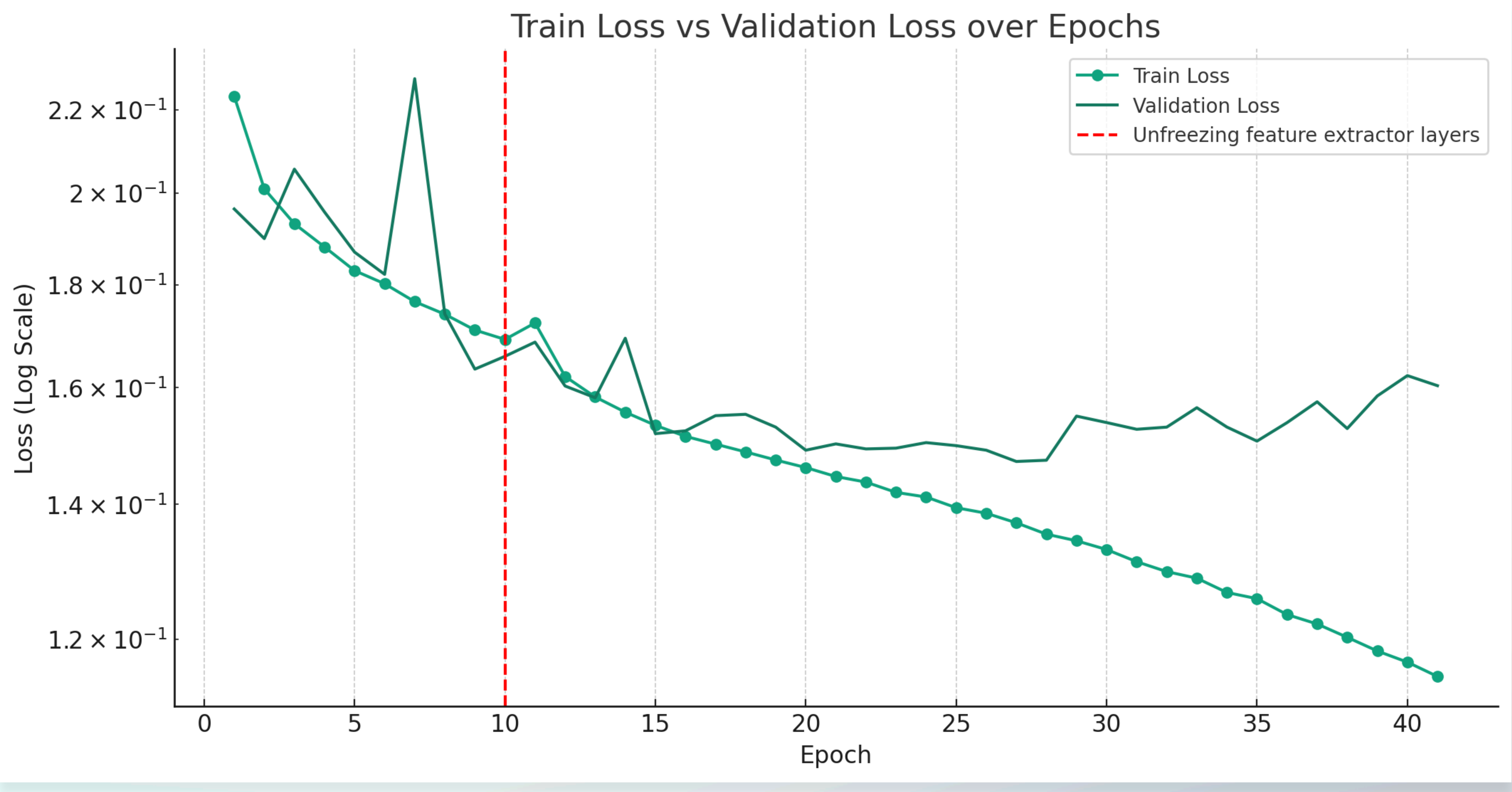


Visualization of Transfer Learning model

Transfer Learning

Transfer Learning Strategies & Overfitting

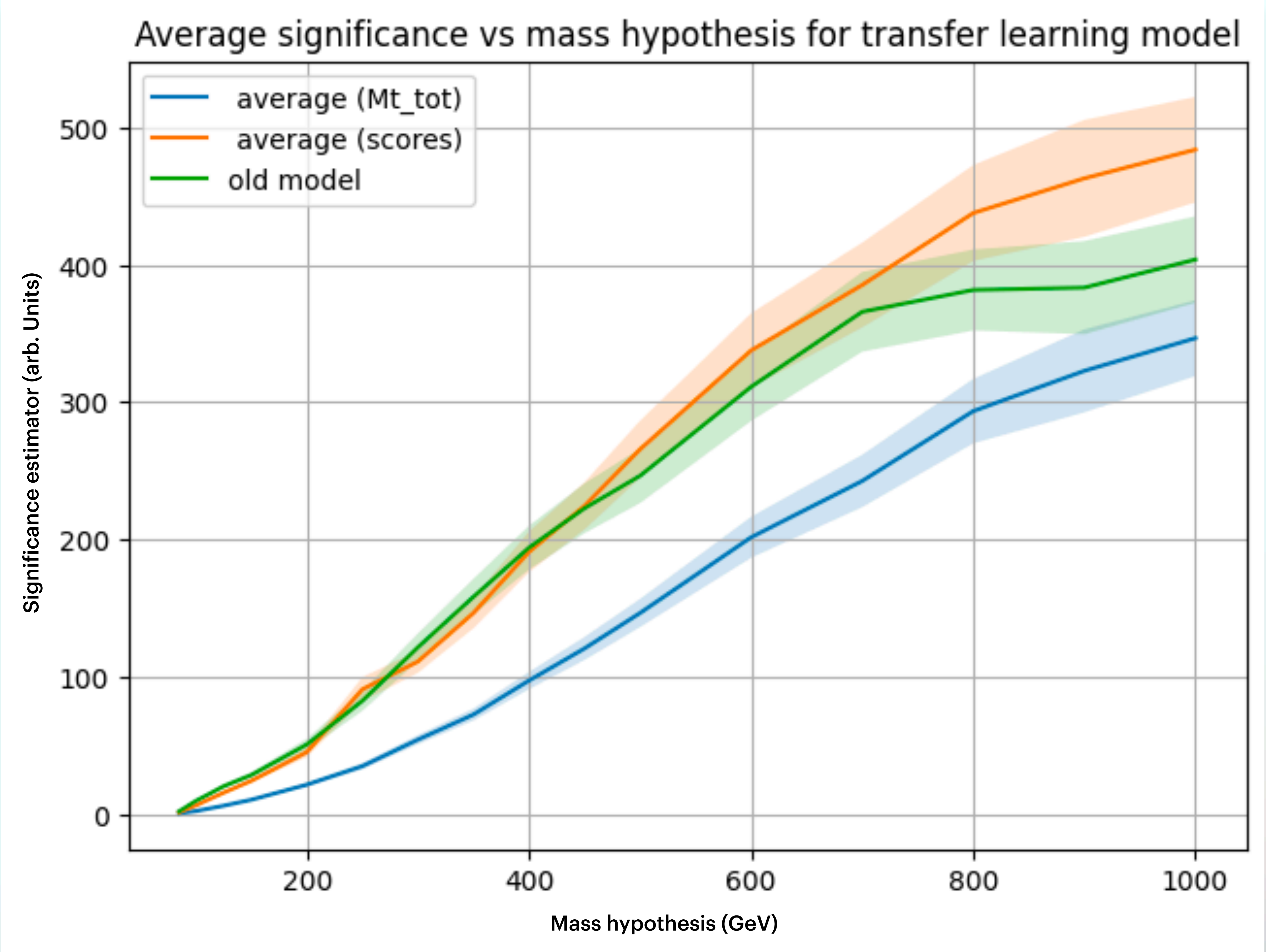
- Options for transfer learning
  - Fixed weights
  - Unfrozen weights
  - Partially unfrozen weights
- Overfitting Risks
  - Escalates when weights are unfrozen
- Mitigation techniques
  - Adamw with L2 regularization
  - Dropout layers
  - Pruning



Transfer Learning

Model

- ✴Pretrained model
  - 1024 width best model
  - Added dropout layers (50% chance)
- ✴Whole model:
  - Unfrozen weights at epoch 5
  - Binary Cross Entropy loss
  - One-hot encoded channel input
  - Hidden layers: [128,128]

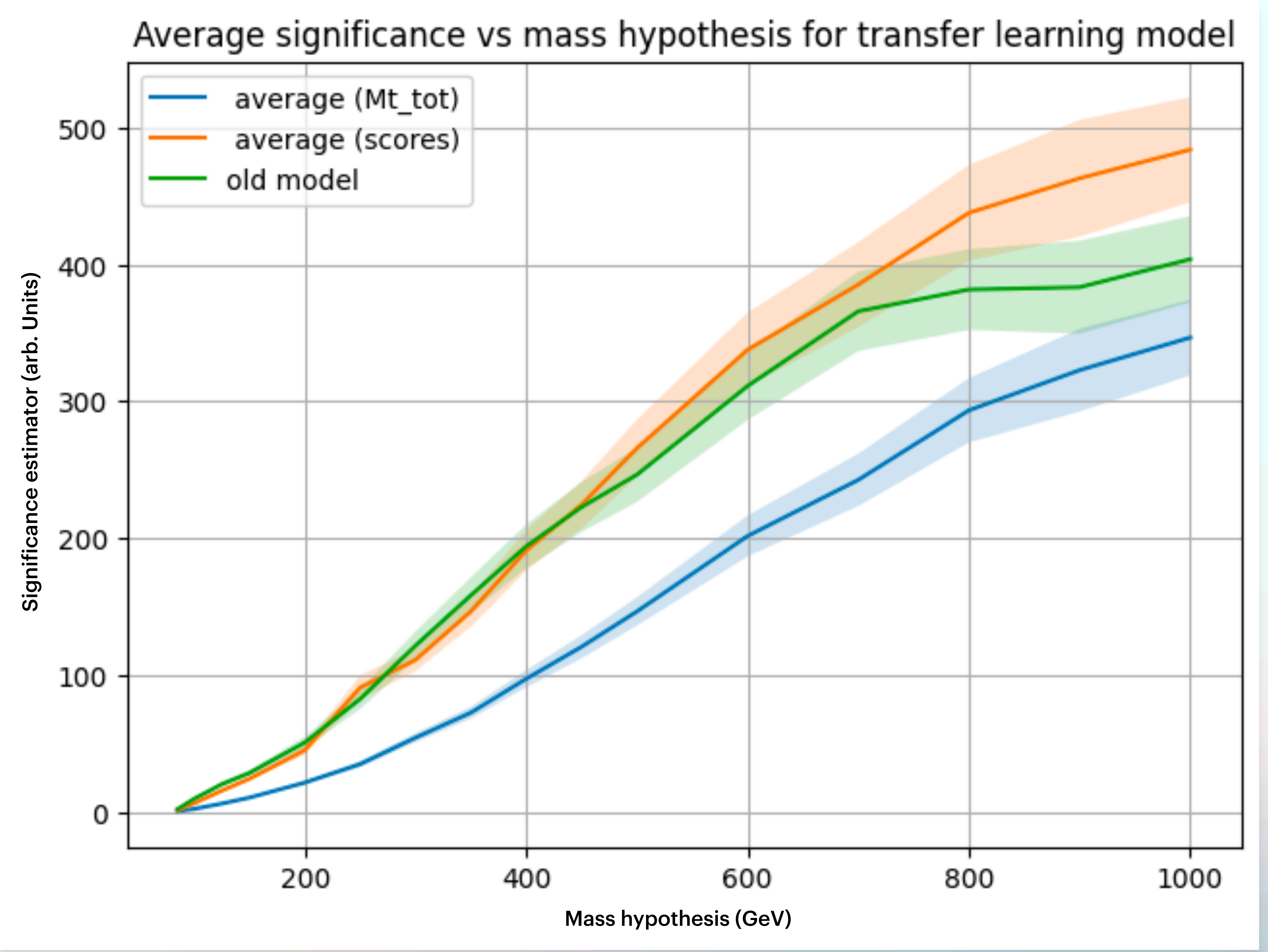


Significance estimator for Transfer learning model

Transfer Learning

Analysis

- Analysis
  - Model performs better than simple DNN at higher mass hypotheses
  - Similar to simple DNN at smaller values
- Possible Improvements
  - Try out other overfitting techniques
  - Stronger Regularization
  - Smaller regression network



Significance estimator for Transfer learning model

Conclusion