Optimization of Event Selection for $H^{\pm\pm}$ Search at ATLAS using Machine Learning Techniques

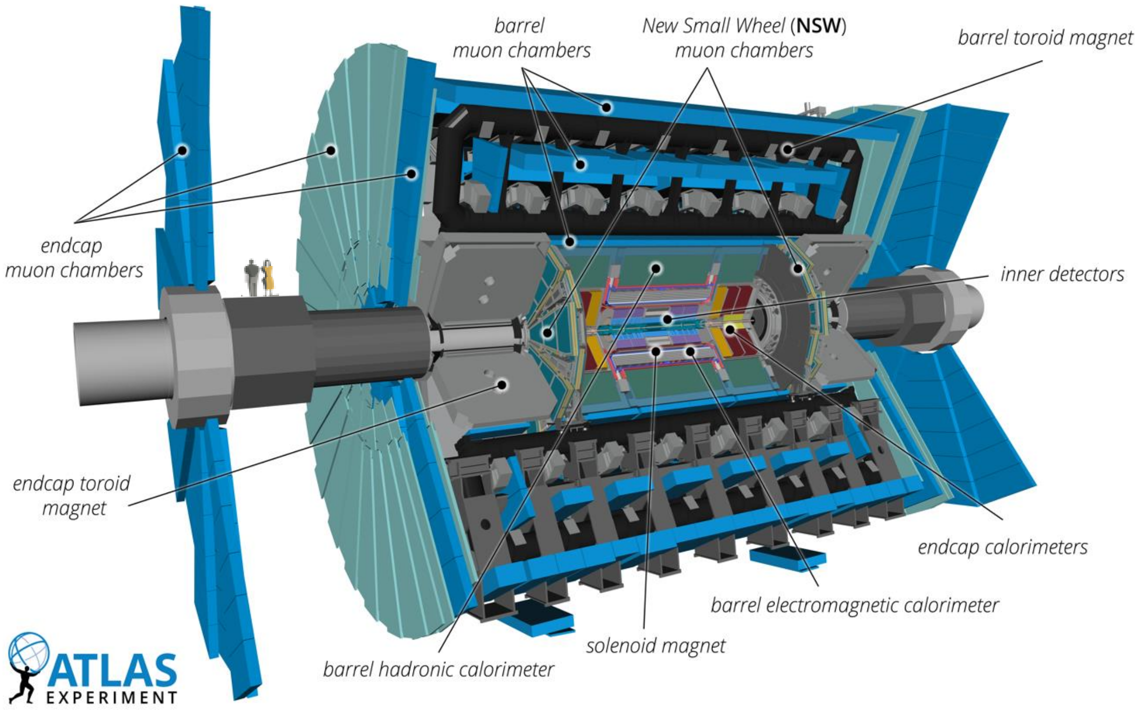
Adrienne Scott
University of Victoria

February 15th, 2025

Winter Nuclear Particle Physics Conference 2025

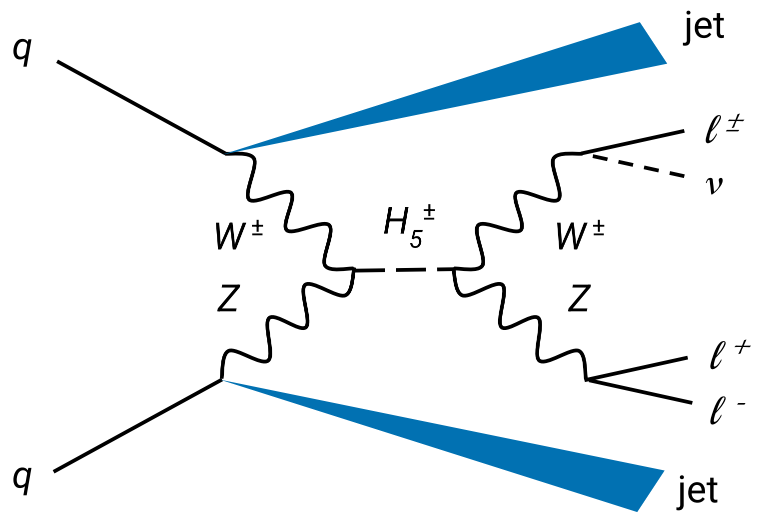
The ATLAS Experiment

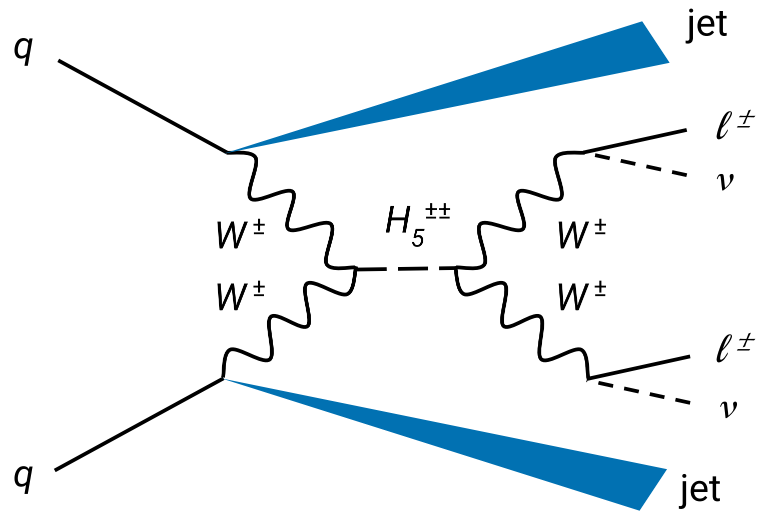
- ATLAS data from p-p collisions at the LHC
  - Run 2 (2015-2018): $\sqrt{s}$ = 13 TeV, 140 fb^{-1}
  - Run 3 (2022-2026): $\sqrt{s}$ = 13.6 TeV, 183 fb^{-1} so far
- **In this talk, only using Run 2 simulated events**
- Events are reconstructed using
  - Charged particle tracks in the inner detector
  - Energy deposits in the calorimeters
  - Hits in the muon spectrometer



Charged Higgs Search

- The Georgi-Machacek (GM) model extends the Higgs sector, allowing for a quintuplet (H_5^0, $H_5^\pm$, $H_5^{\pm\pm}$) which are degenerate in mass (m_{H_5})[1]
- Our analysis is performing a search for $H_5^\pm$ and $H_5^{\pm\pm}$ produced by vector boson fusion (VBF) and decaying leptonically via $W^\pm Z$ or $W^\pm W^\pm$
  - **In this talk, focusing on $H_5^{\pm\pm}$**

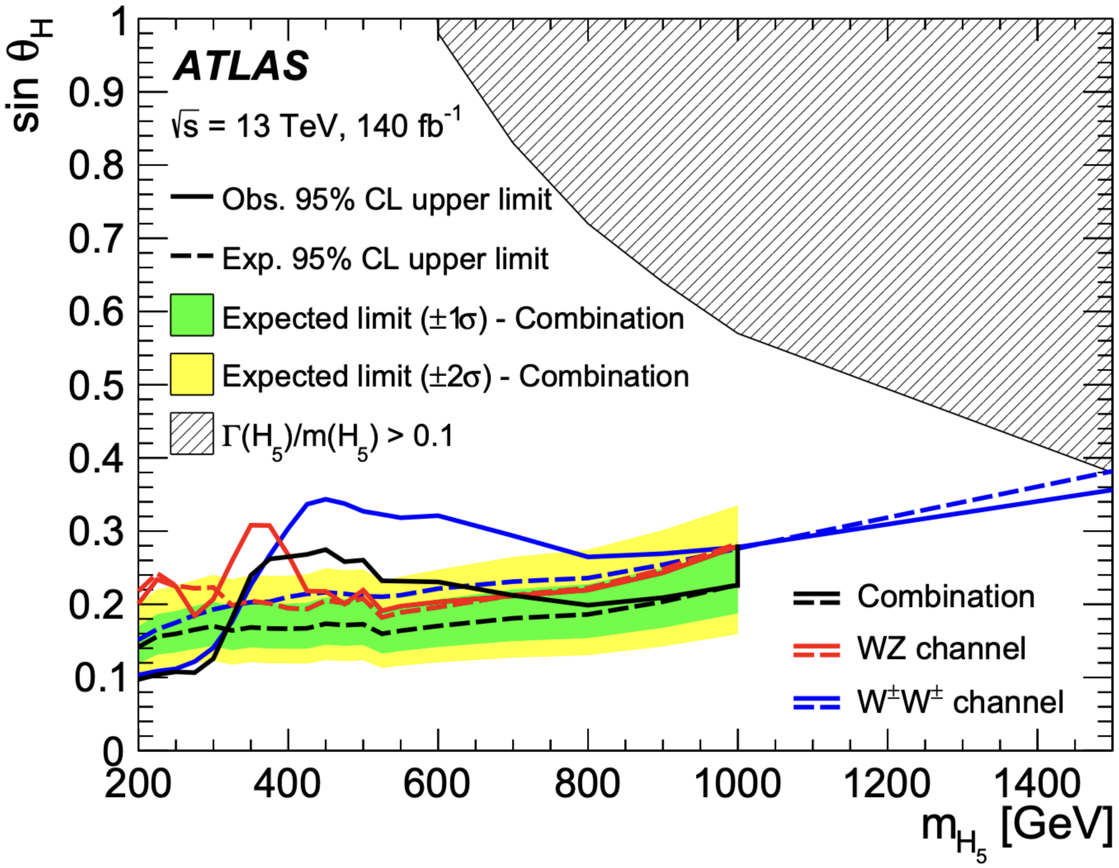




1 Nucl. Phys. B 262 (1985) 463

Charged Higgs Search

- The Georgi-Machacek (GM) model extends the Higgs sector, allowing for a quintuplet (H_5^0, $H_5^\pm$, $H_5^{\pm\pm}$) which are degenerate in mass (m_{H_5})[1]
- Our analysis is performing a search for $H_5^\pm$ and $H_5^{\pm\pm}$ produced by vector boson fusion (VBF) and decaying leptonically via $W^\pm Z$ or $W^\pm W^\pm$
  - **In this talk, focusing on $H_5^{\pm\pm}$**
- Motivated by excess observed in Run 2 in $W^\pm Z$ and $W^\pm W^\pm$ channels
- Improve limits by optimizing search for GM model, using Run 2 and partial Run 3 (2022-2023) data

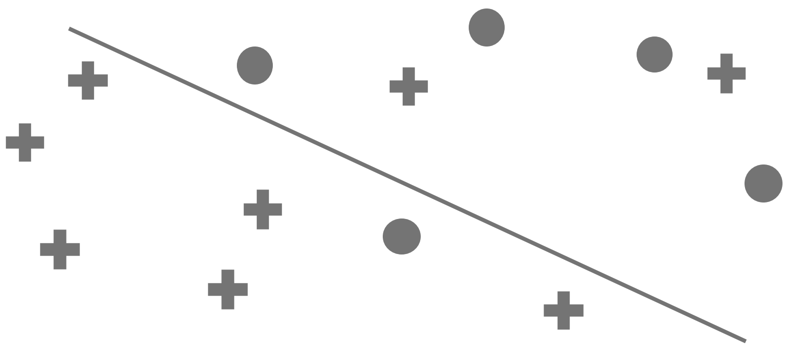


Phys. Lett. B 860 (2025) 139137.

1 Nucl. Phys. B 262 (1985) 463

Signal Region Optimization for $H^{\pm\pm}$

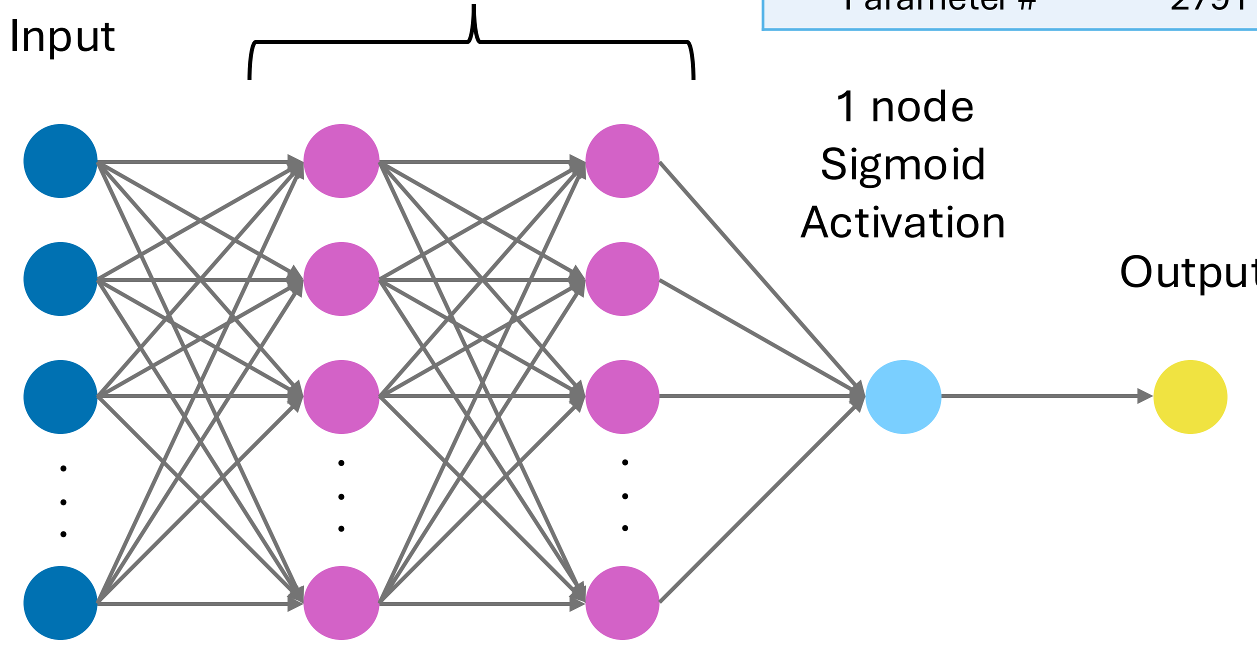
- Improving the signal region(SR) selection will improve our sensitivity to $H^{\pm\pm}$
- High-dimensional problem with large quantity of labelled simulation events is ideal for machine learning (ML)
- *Can ML do better than existing cut-based SR selection?*



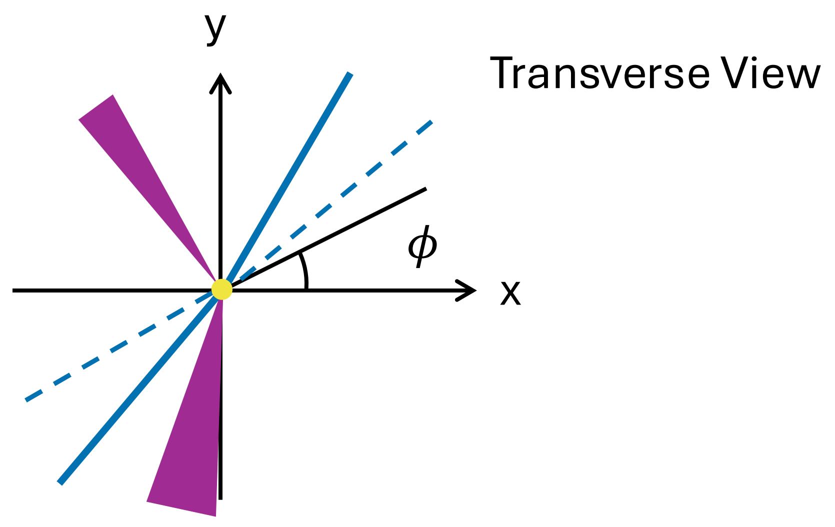
Machine Learning Approach

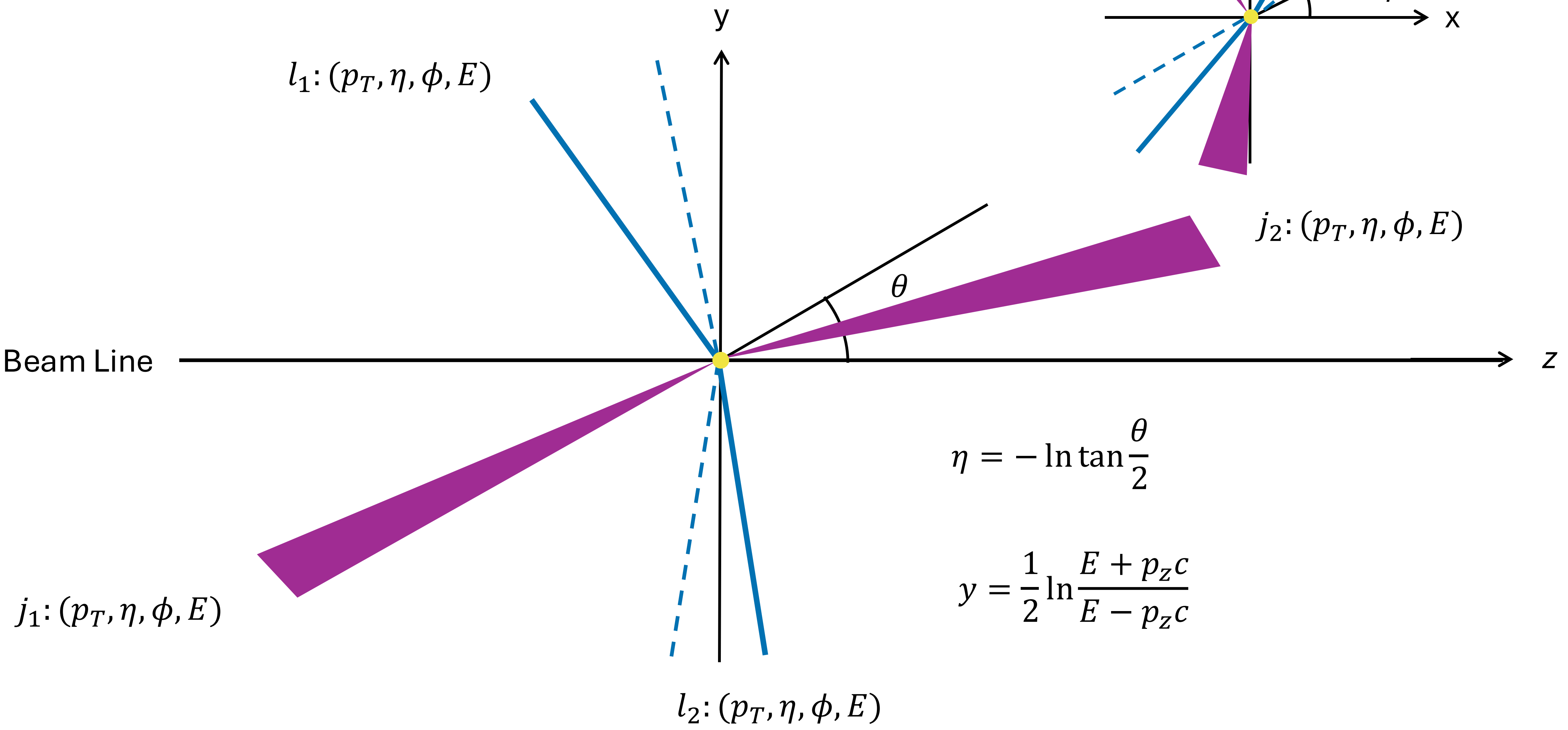
Optimizer	Adam
Learning Rate	0.001
Loss Function	BCE
Epoch #	100
Batch Size	1000
Dropout	20%
Parameter #	2791

- Classifier neural network (NN) implemented in Keras Tensorflow
- Simulated event samples from Run 2
- Apply a ‘basic’ SR cut prior to training
- Signal (1) – 277,393 events
  - 200 GeV – 3 TeV $H^{\pm\pm}$ produced via VBF
  - Omitting every other mass point for interpolation studies
- Backgrounds (0) – 879,119 events
  - EW & QCD (& interference) $W^{\pm}W^{\pm}$ production with 2 jets in the final state
  - EW & QCD (& interference) $W^{\pm}Z$ production with 2 jets in the final state
- Balance signal & background using class weights



Input Features

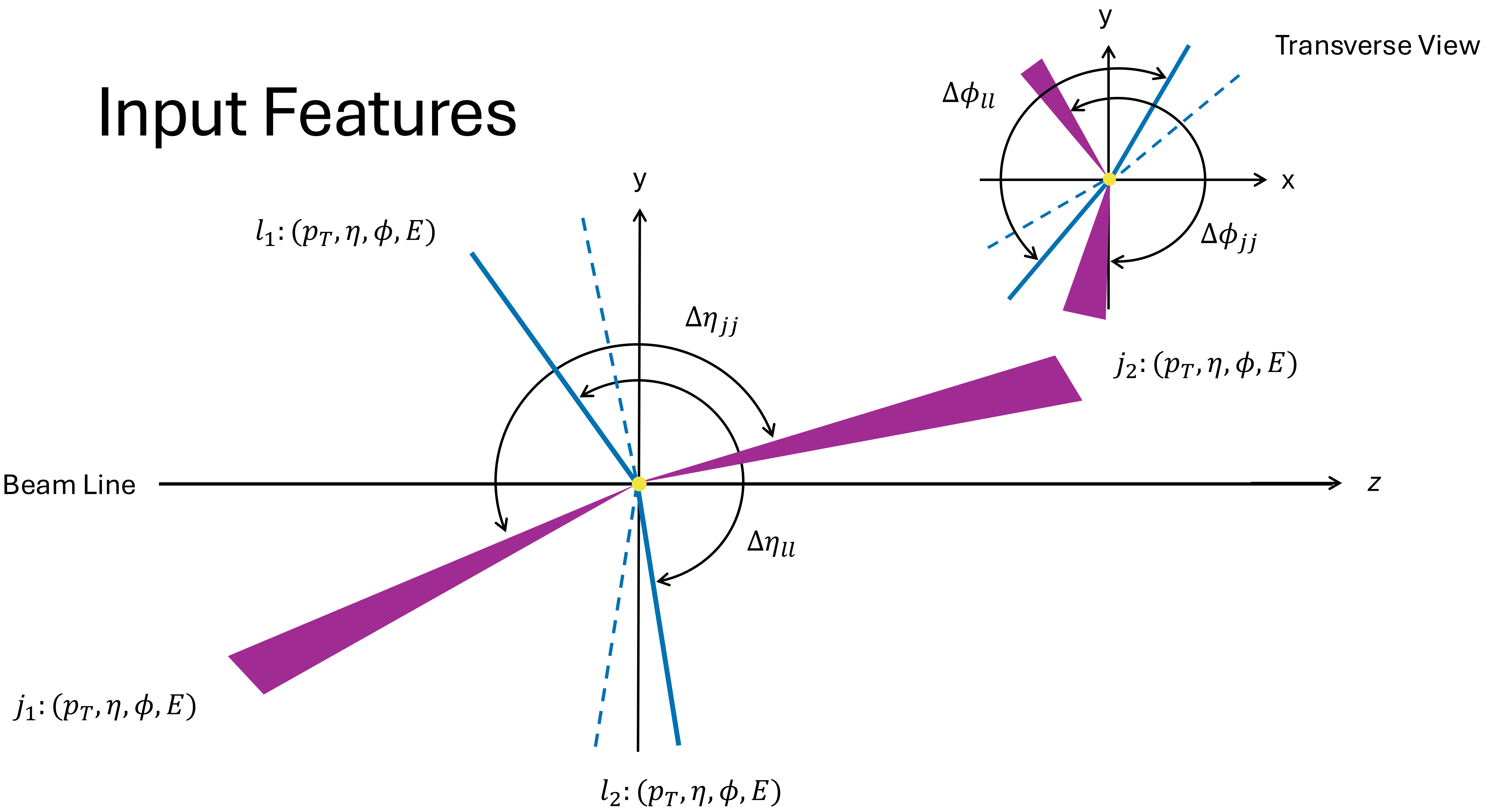




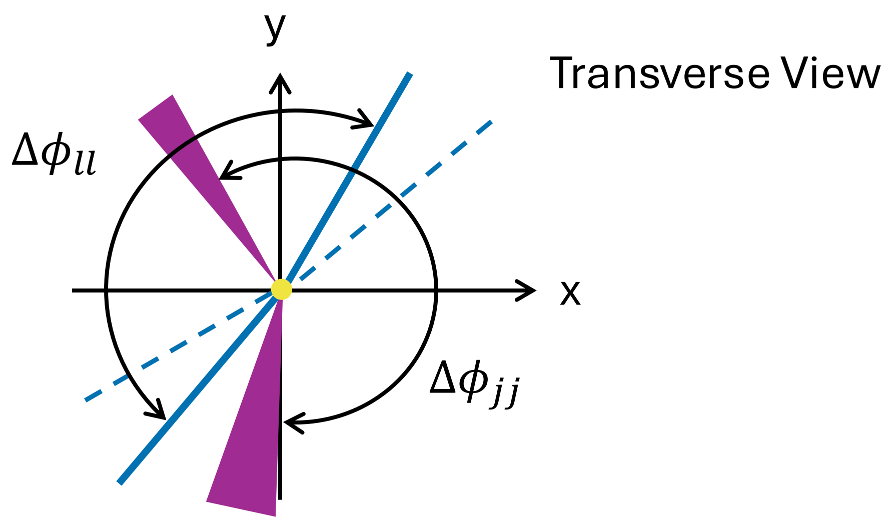
$$\eta = -\ln \tan \frac{\theta}{2}$$

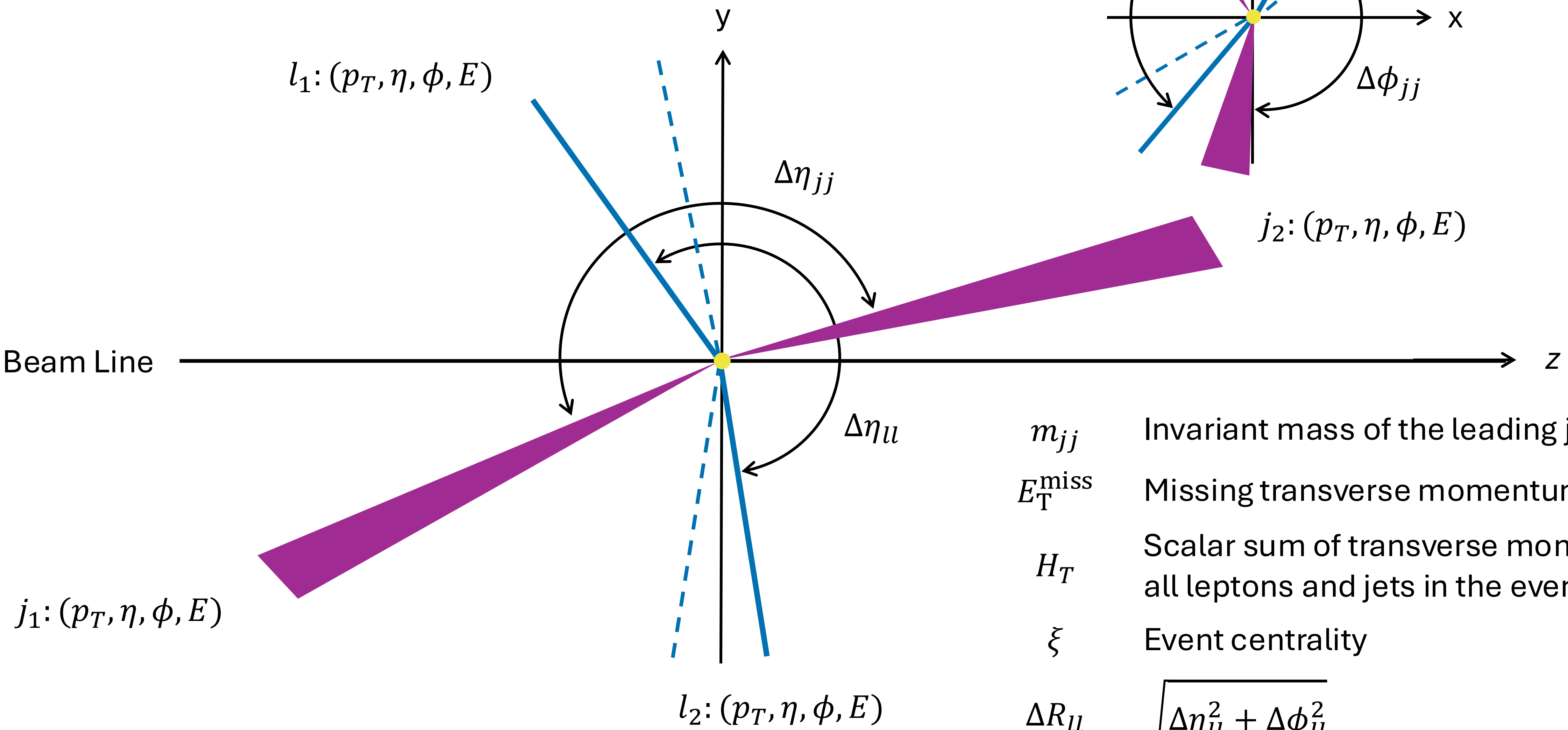
$$y = \frac{1}{2} \ln \frac{E + p_z c}{E - p_z c}$$

Input Features



Input Features





m_{jj}	Invariant mass of the leading jet pair
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse momentum ($\lvert \boldsymbol{p}_{\mathrm{T}}^{\mathrm{miss}} \rvert$)
H_T	Scalar sum of transverse momentum of all leptons and jets in the event
ξ	Event centrality
ΔR_{ll}	$\sqrt{\Delta\eta_{ll}^2 + \Delta\phi_{ll}^2}$

Feature Selection Procedure

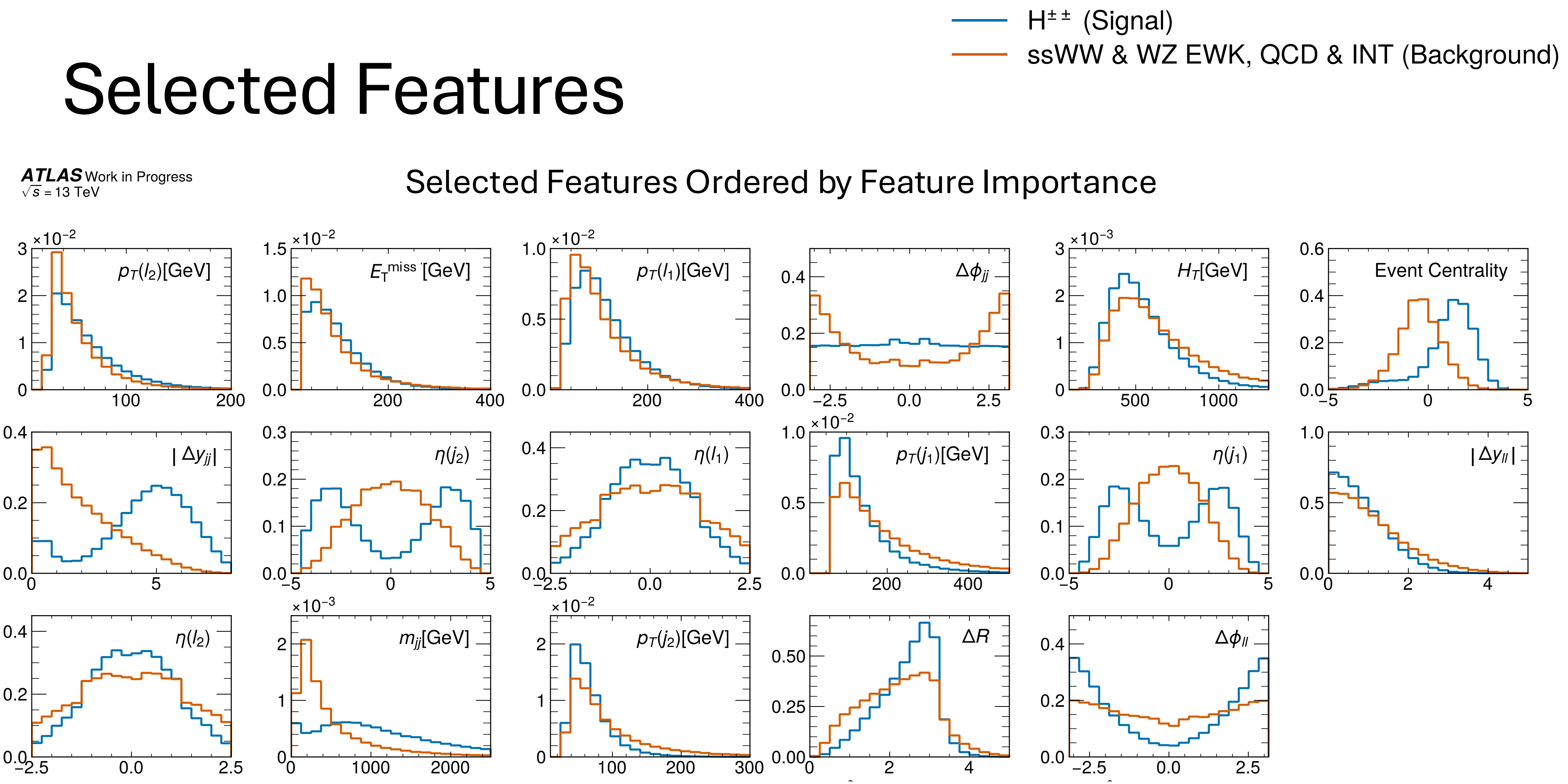
1. Train NN with all features
2. Train NN with each feature scrambled one at a time
3. Rank features by feature importance

$$\textbf{Feature Importance (FI)} = \frac{\textbf{Area Under ROC Curve with Feature}}{\textbf{Area Under ROC Curve without feature}}$$

FI > 1: The network performs worse without the feature
FI < 1: The network performs better without the feature

4. Remove low ranking features and repeat until AUC is maximized

Selected Features



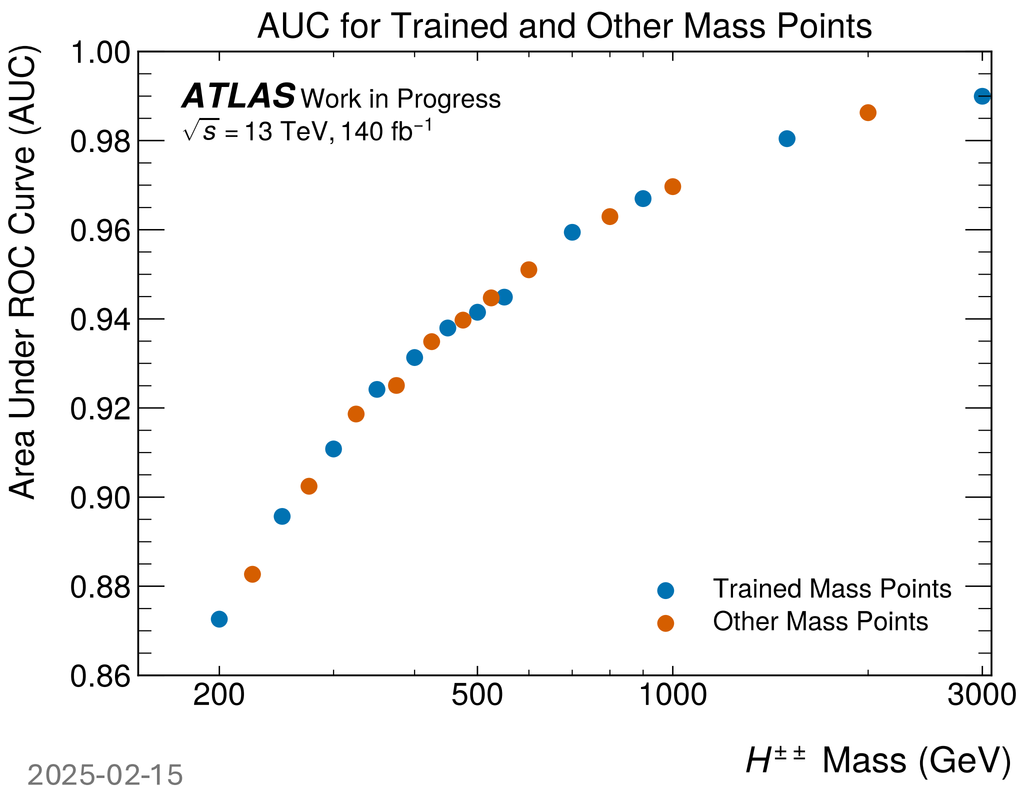
Charged Higgs Mass Coverage

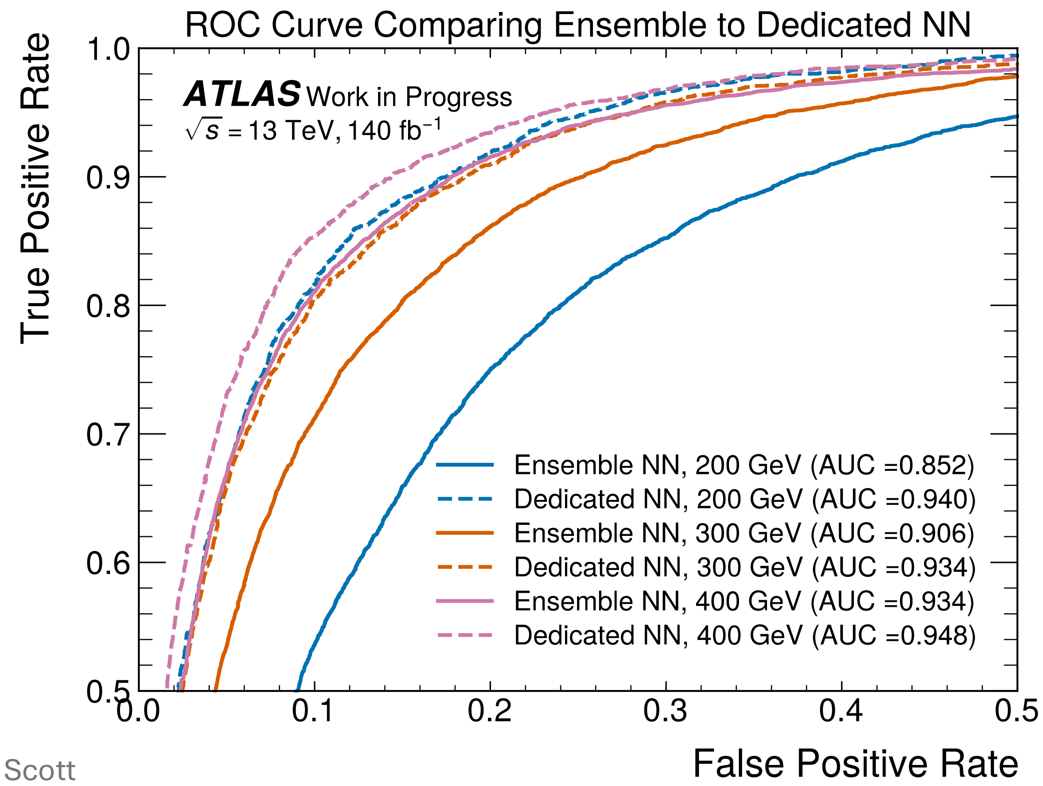
- Performing search for $H^{\pm\pm}$ masses ranging from 200 GeV to 3 TeV
- Simplest approach is to use a single NN for entire mass range
- So far: using an **ensemble** of signals across the mass range

1. How well is the ensemble NN able to interpolate between simulated mass points used in training?
2. How does the ensemble NN performance depend on $H^{\pm\pm}$ mass?
3. At a given mass point, how does the ensemble NN performance compare to a NN dedicated for this mass point?

Charged Higgs Mass Coverage

1. Ensemble demonstrates excellent interpolation power
2. Performance decreases significantly at low mass
3. Gap between dedicated NN and ensemble NN increases at low mass





Event Weighting

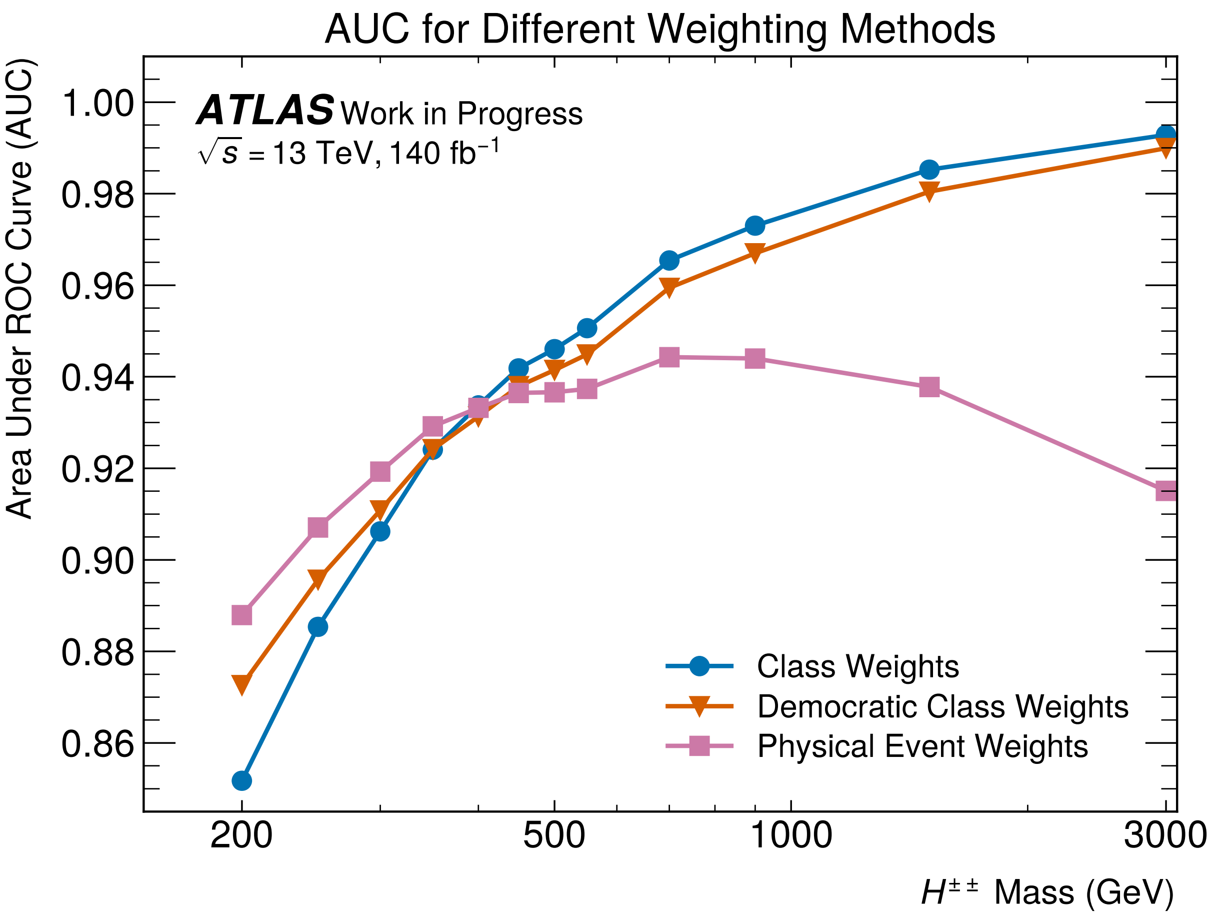
- *Can we improve performance at low mass by modifying how we combine signal samples?*
- Class weights method depends on number of events per mass point sample and distribution of simulated mass points

Two new approaches:

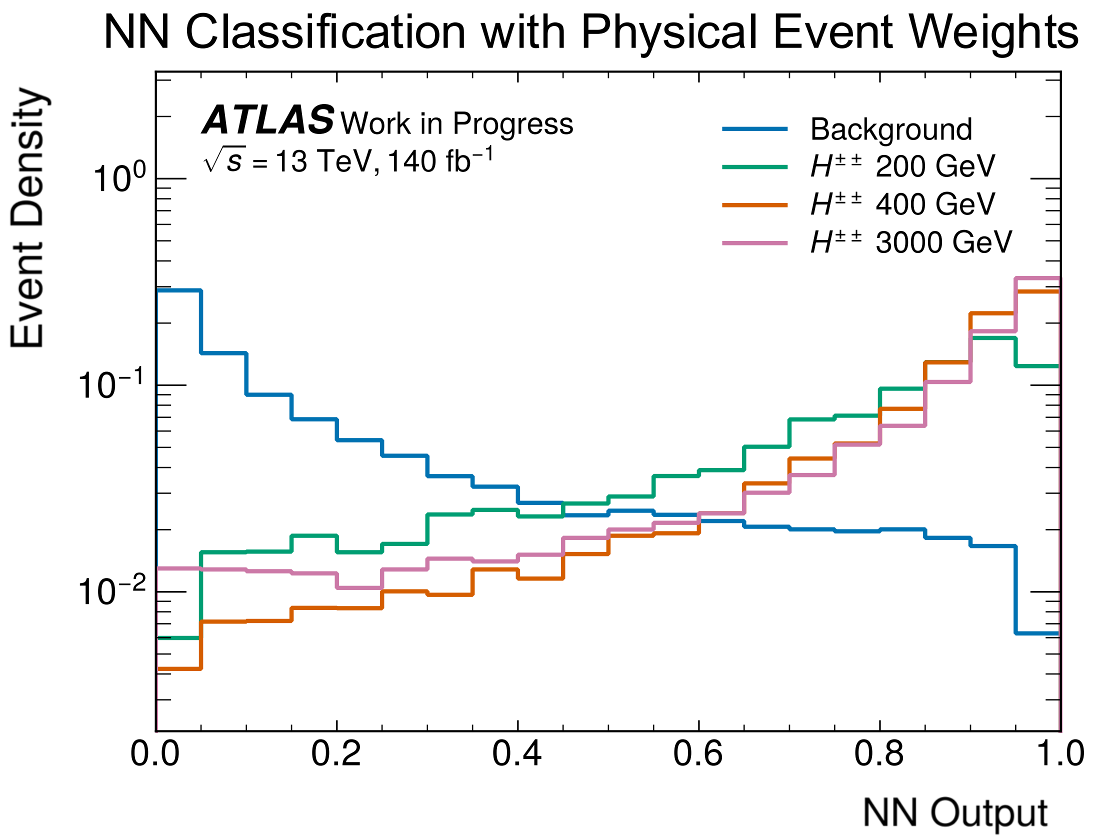
1. Democratic Signal Class Weights
    - Assign signal weights so that each signal sample has equal sum of weights
    - Eliminates dependence on number of events per mass point sample
2. Physical Event Weights
    - Depends on Monte Carlo weights and cross section
    - Encourages network to focus on events which are more likely to be found in data

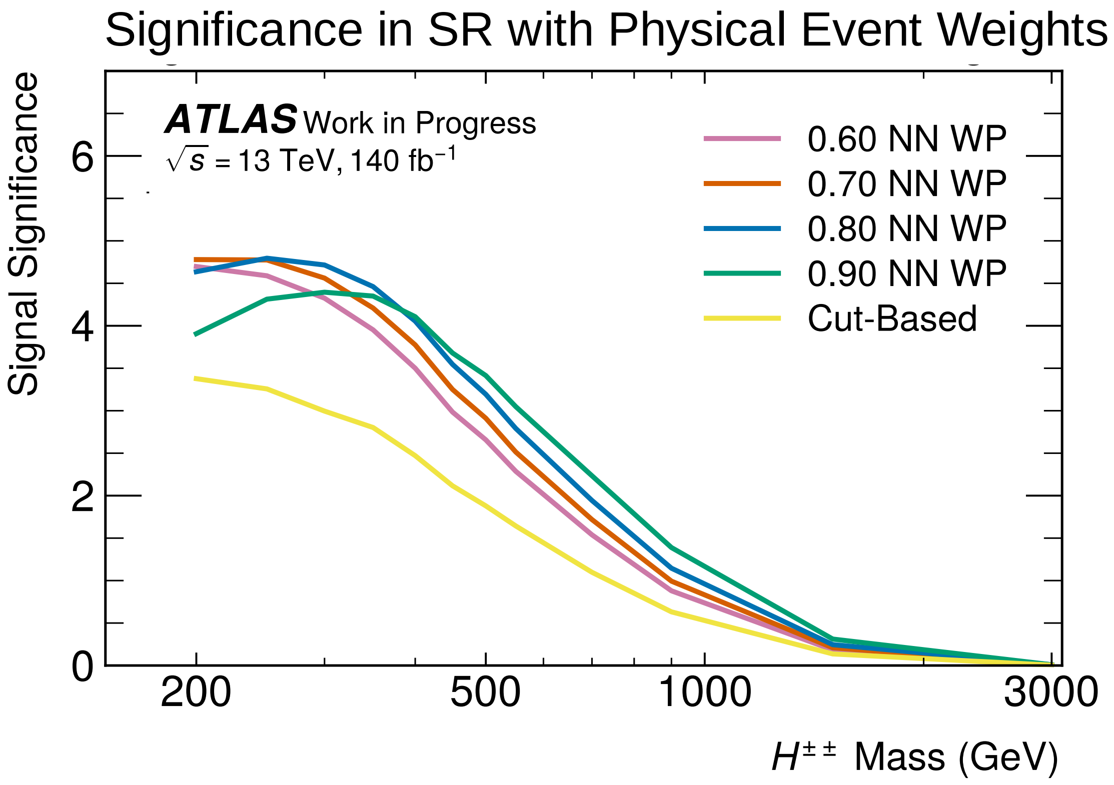
Event Weighting Results

- Democratic class weights slightly improve performance at low mass
  - High mass samples have more events
- Physical event weights significantly improve performance at low mass
  - Cross-section decreases monotonically with mass
- Physical event weights also reduce performance at high mass
  - We expect lower signal significance in this mass range



Physical Event Weights Performance





Conclusion

- Features that are important for NNs may not be for cuts-based analysis
- Ensemble training enables interpolation between mass points
- Class and event weighting methods can significantly impact NN outcomes
- Simple NNs can out-perform cuts-based selections and offer a rich environment for exploring ML behaviour

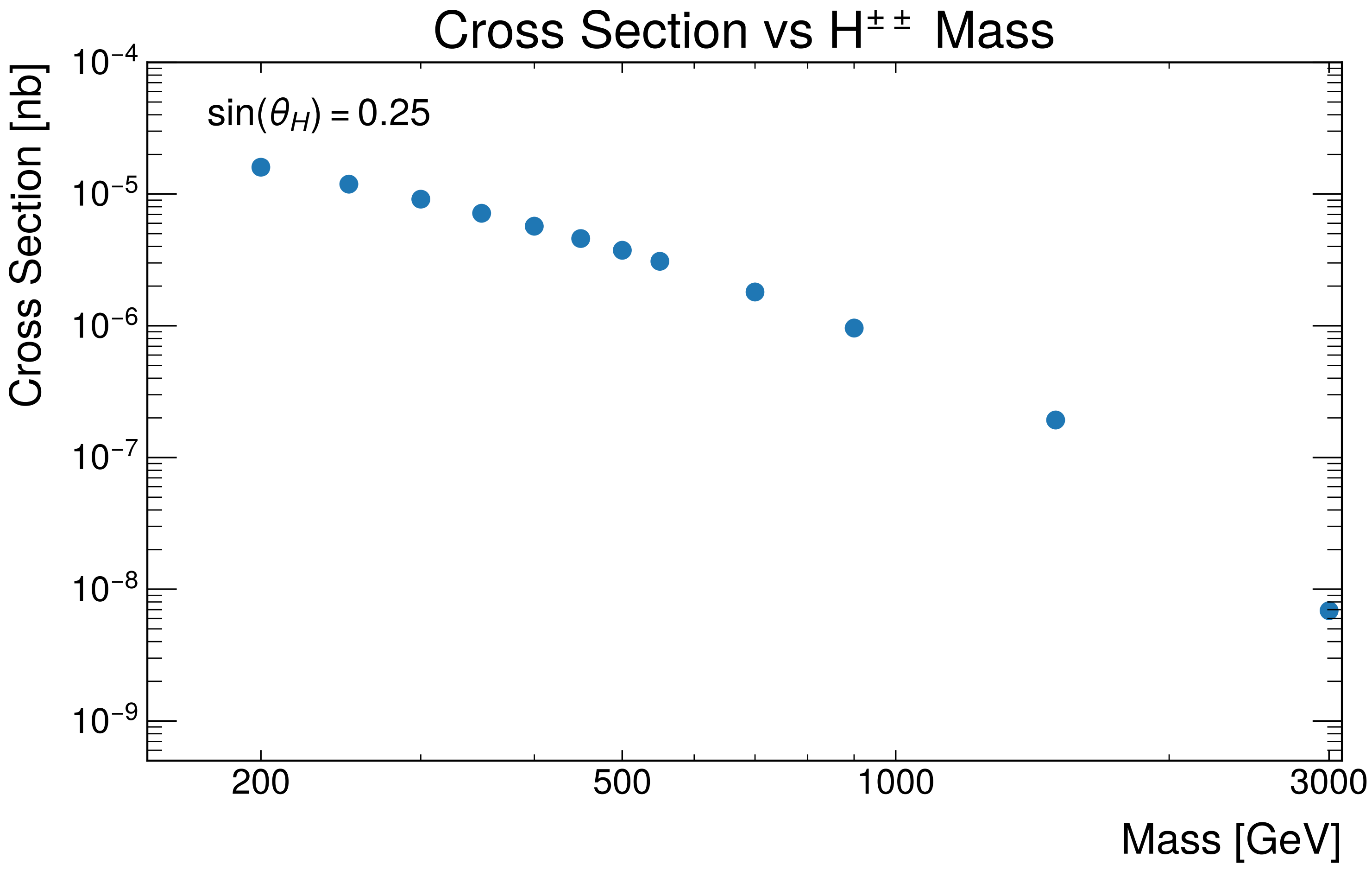
Thank you!

ATLAS EXPERIMENT

University of Victoria

NSERC CRSNG

Backup: $H^{\pm\pm}$ Cross-Section



Backup: Feature Formulas

- MET (or $E_{\text{miss}}^{\text{T}}$) is the magnitude of the $p_{\text{T}}^{\text{miss}}$ vector defined as

$$\boldsymbol{p}_{\text{T}}^{\text{miss}} = -\Bigg(\underbrace{\sum_{\substack{\text{selected}\\ \text{electrons}}} \boldsymbol{p}_{\text{T}}^{e} + \sum_{\substack{\text{accepted}\\ \text{photons}}} \boldsymbol{p}_{\text{T}}^{\gamma} + \sum_{\substack{\text{accepted}\\ \tau-\text{leptons}}} \boldsymbol{p}_{\text{T}}^{\tau} + \sum_{\substack{\text{selected}\\ \boldsymbol{\mu}}} \boldsymbol{p}_{\text{T}}^{\mu} + \sum_{\substack{\text{accepted}\\ \text{jets}}} \boldsymbol{p}_{\text{T}}^{\text{jet}}}_{\text{hard term}} + \underbrace{\sum_{\substack{\text{unused}\\ \text{tracks}}} \boldsymbol{p}_{\text{T}}^{\text{track}}}_{\text{soft term}}\Bigg).$$

  - Source: [The performance of missing transverse momentum reconstruction and its significance with the ATLAS detector using 140 fb−1 of s√=13 TeV pp collisions.]() (2024)

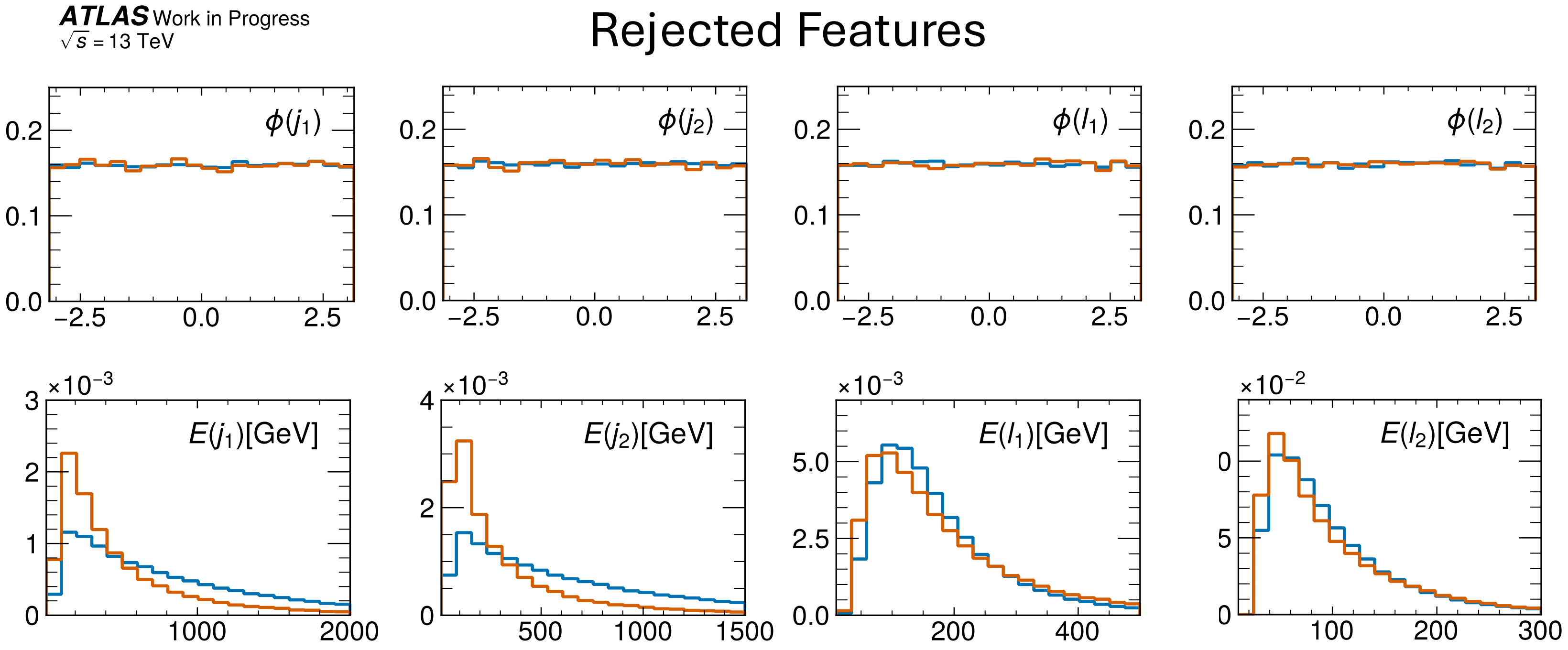
- Event centrality

$$\xi = \min\{\min(\eta_{l_1}, \eta_{l_2}) - \min(\eta_{j_1}, \eta_{j_2}), \max(\eta_{j_1}, \eta_{j_2}) - \max(\eta_{l_1}, \eta_{l_2})\}$$

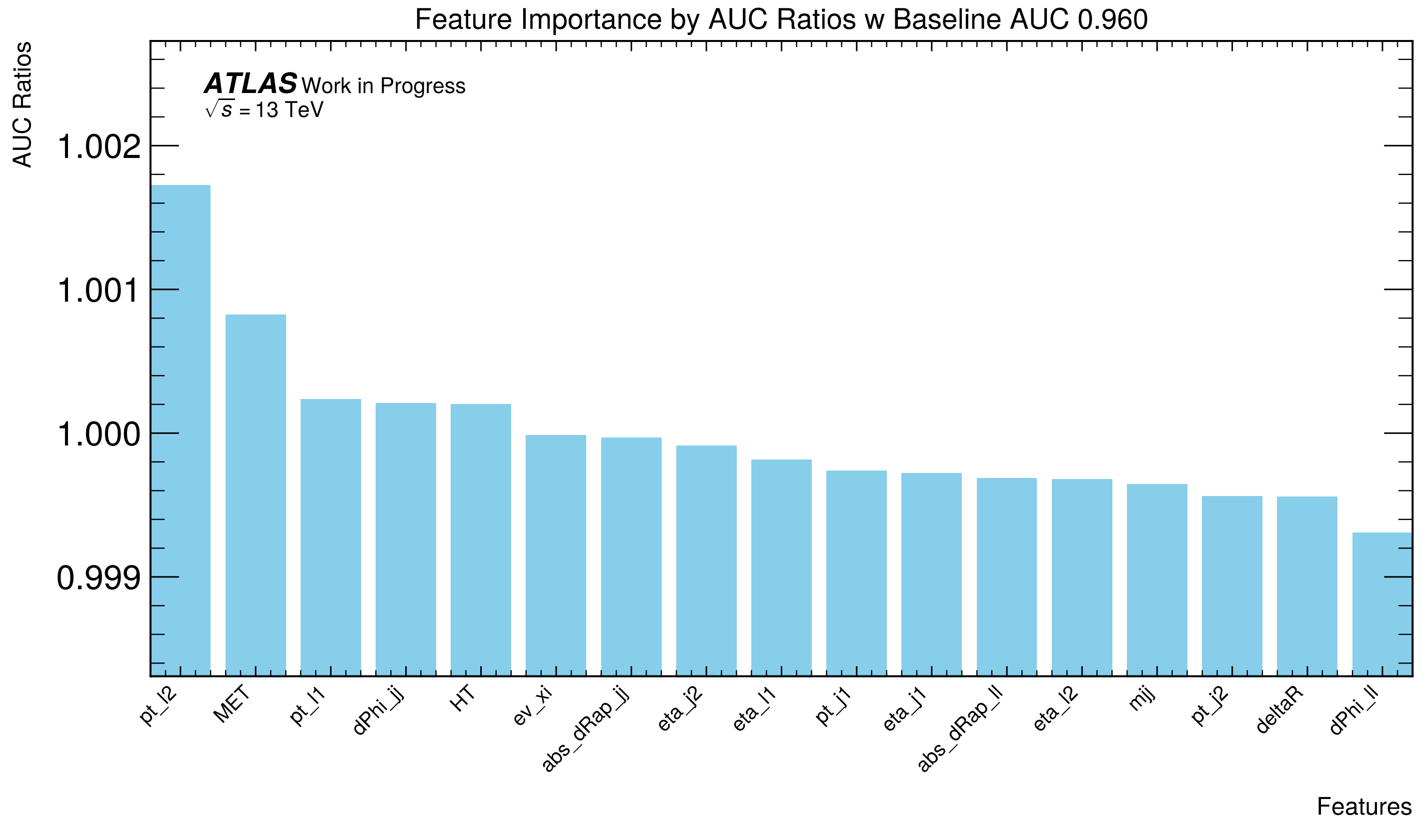
$H^{\pm\pm}$ (Signal)
ssWW & WZ EWK, QCD & INT (Background)

Backup: Rejected Features

- Additional redundant information can decrease network performance
  - Energy is highly correlated with p_T
  - Useful phi information was already encoded in $\Delta\phi_{jj}$ or $\Delta\phi_{ll}$

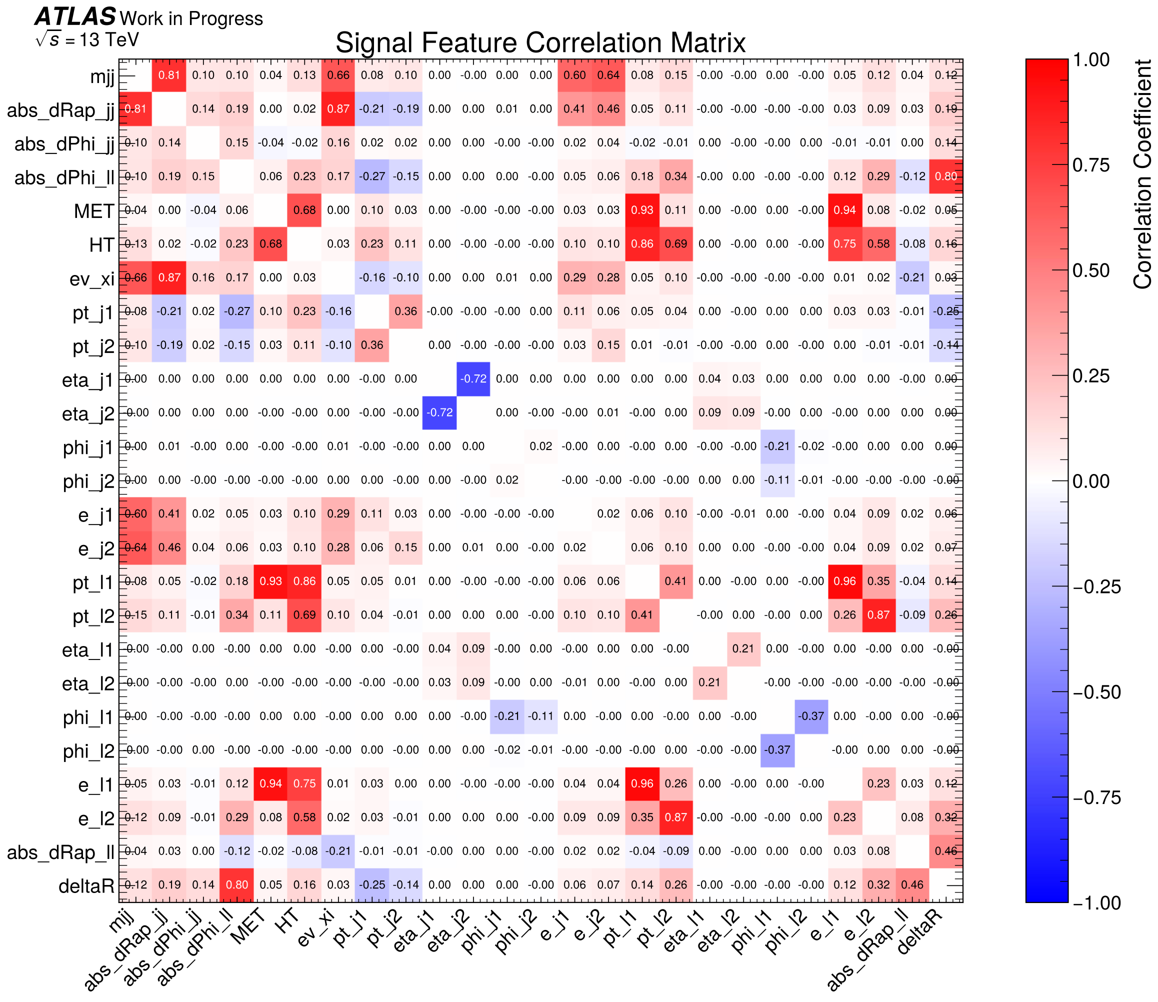


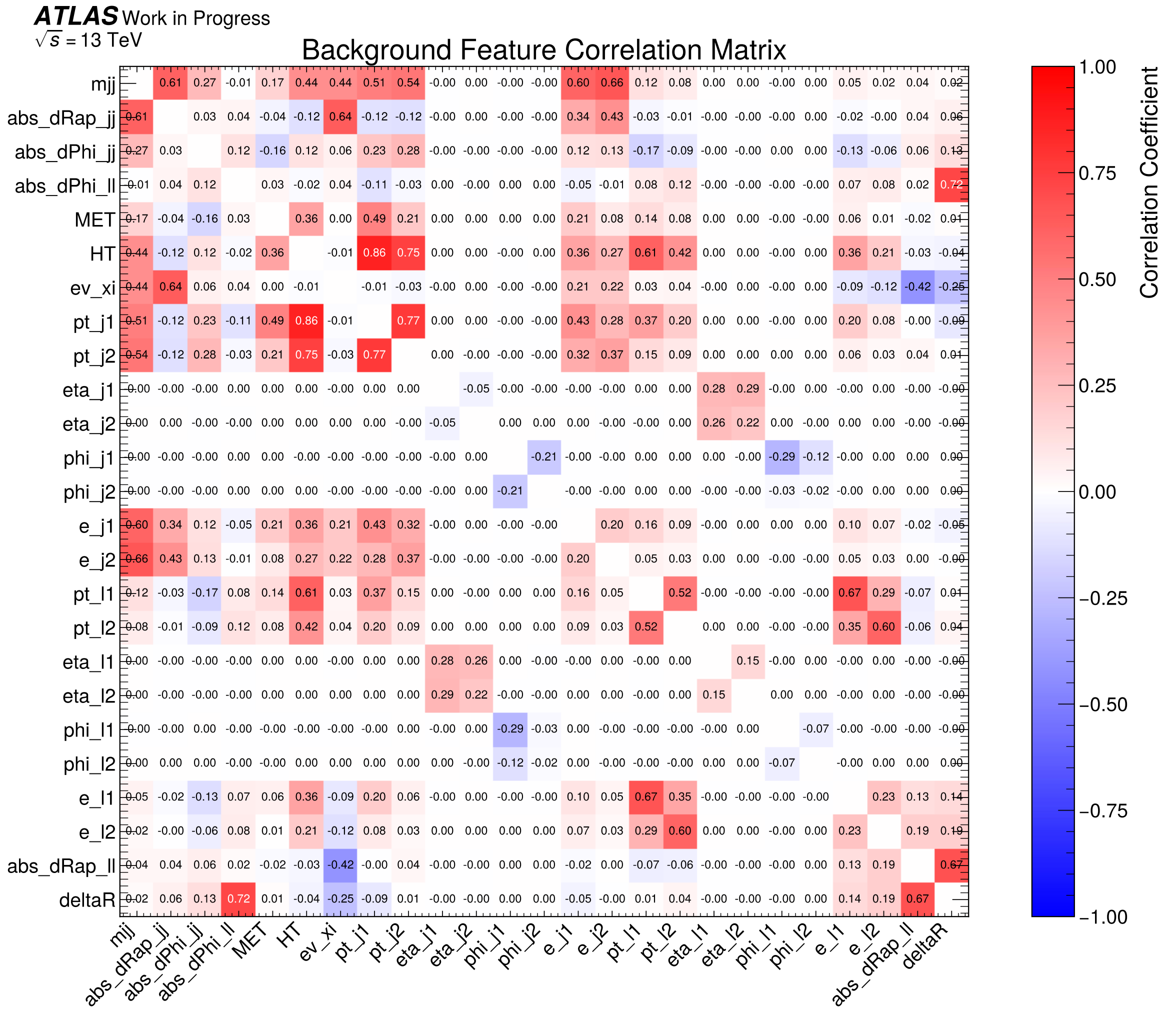
Backup: Feature Analysis



Iteration 1	Iteration 2	Iteration 3	Iteration 4
H_T	$p_T(l_1)$	$p_T(l_2)$	$p_T(l_2)$
$p_T(l_1)$	$p_T(l_2)$	$\eta(l_2)$	E_T^{miss}
$\eta(l_1)$	$\eta(l_1)$	$\Delta\phi_{ll}$	$p_T(l_1)$
$\lvert\Delta y_{jj}\rvert$	$\eta(l_2)$	$\eta(l_1)$	$\Delta\phi_{jj}$
$\eta(l_2)$	H_T	E_T^{miss}	H_T
$\Delta\phi_{ll}$	E_T^{miss}	$p_T(l_1)$	ξ
$\phi(j_1)$	$\Delta\phi_{ll}$	$\Delta\phi_{jj}$	$\lvert\Delta y_{jj}\rvert$
ξ	$\eta(j_2)$	H_T	$\eta(j_2)$
E_T^{miss}	ξ	$p_T(j_1)$	$\eta(l_1)$
$\phi(l_1)$	$\Delta\phi_{jj}$	m_{jj}	$p_T(j_1)$
$p_T(j_1)$	$\lvert\Delta y_{jj}\rvert$	$\eta(j_1)$	$\eta(j_1)$
$\Delta\phi_{jj}$	$\phi(j_1)$	ξ	$\boldsymbol{\lvert\Delta y_{ll}\rvert}$
$E(l_1)$	$p_T(j_1)$	$\lvert\Delta y_{jj}\rvert$	$\eta(l_2)$
$\phi(j_2)$	$\eta(j_1)$	$\eta(j_2)$	m_{jj}
$p_T(l_2)$	m_{jj}	$p_T(j_2)$	$p_T(j_2)$
$p_T(j_2)$	$p_T(j_2)$		$\boldsymbol{\Delta R_{ll}}$
$E(j_1)$	$\phi(j_2)$		$\Delta\phi_{ll}$
$\phi(l_2)$	$\phi(l_2)$		
m_{jj}	$\phi(l_1)$		
$\eta(j_1)$			
$E(l_2)$			
$\eta(j_2)$			
$E(j_2)$			

Backup: Feature Correlations





*The absolute value of $\Delta\phi_{ll}$ and $\Delta\phi_{jj}$ was only used for the correlation matrices. In training, $\Delta\phi_{ll}$ and $\Delta\phi_{jj}$ were used with no absolute value.

Backup: Event Weighting Equations

Class Weights:

$$w_s = \frac{N_s+N_b}{2 \cdot N_s}, w_b = \frac{N_s+N_b}{2 \cdot N_b}$$

Democratic Signal Class Weights:

$$w_{s,i} = \frac{\sum_i^n N_{s,i}}{n \cdot N_{s,i}} \frac{N_s+N_b}{2 \cdot N_s}, w_b = \frac{N_s+N_b}{2 \cdot N_b}$$

Physical Event Weights:

1. For each event the initial weight is $\mathrm{MC\ Weight} \times \mathrm{Scale\ Factors} \times \sigma \times \mathcal{L}/\sum w_i$ where $\sum w_i$ is the sum of weights of events from that signal or background sample
2. The 1st percentile event weight is scaled to 0, the 99th percentile event weight is scaled to 1
   - Weights below the 1st percentile are set to 0, weights above the 99th percentile are set to 1
3. Multiply signal and background weights by f_s and f_b respectively to equalize the sum of weights of signal and background

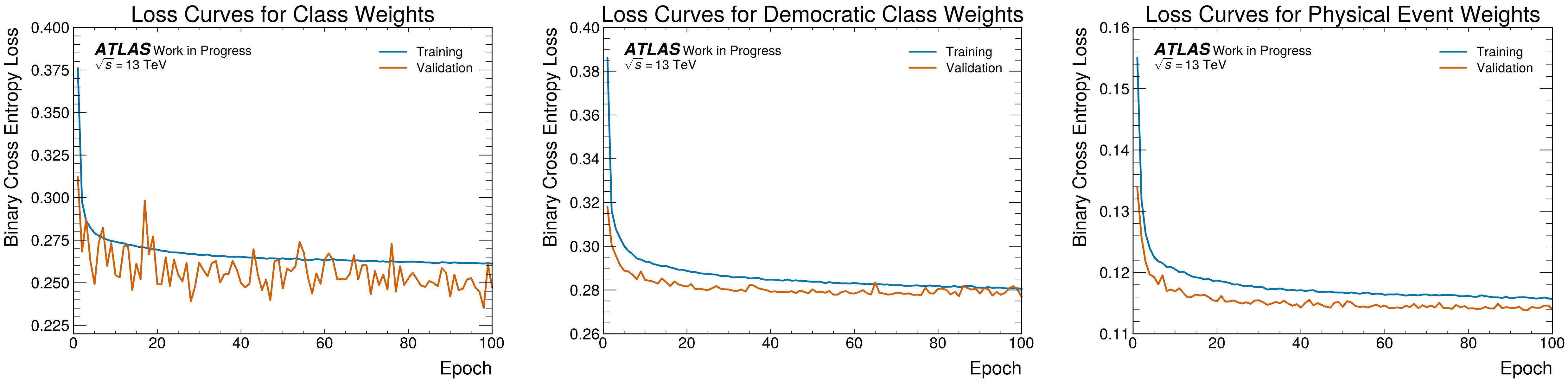
$$f_s = \frac{\sum w_s + \sum w_b}{2 \cdot \sum w_s}, f_b = \frac{\sum w_s + \sum w_b}{2 \cdot \sum w_b}$$

Backup: Loss Function & Training Curves

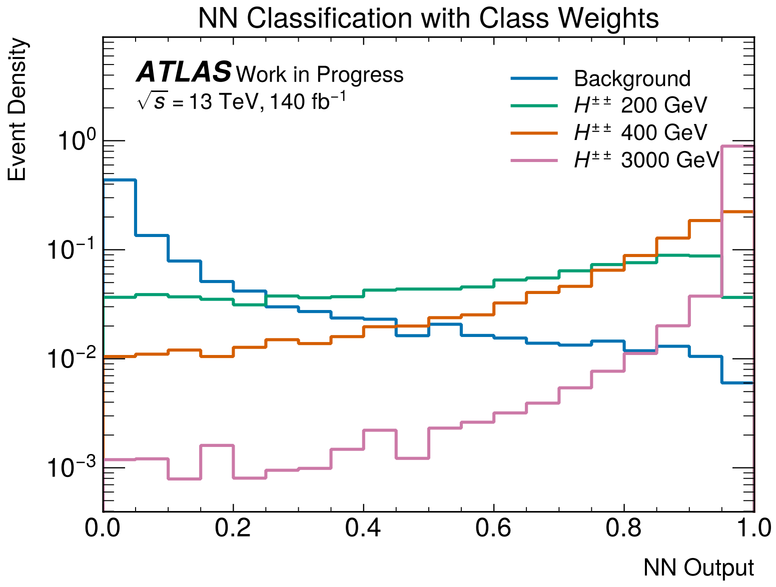
- Loss Function in use is Binary Cross Entropy with class/event weights w_i

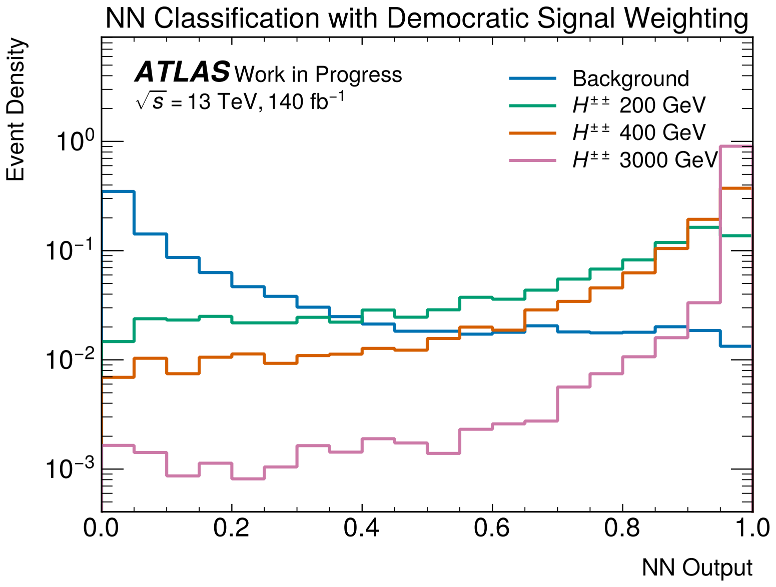
$$BCE_q(p) = -\sum_{i}^{N} \frac{w_i}{\sum_i w_i} \cdot \left[(1 - y_i) \cdot \log\big(1 - p(y_i)\big) + y_i \cdot \log\big(p(y_i)\big)\right]$$

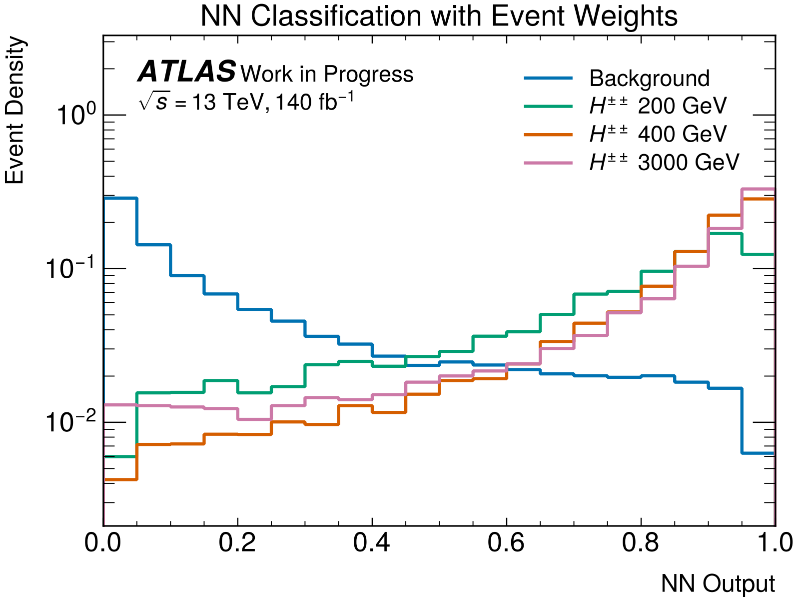
- Training Set: 60%, Validation Set: 20%, Test Set: 20%



Backup: Classification Distributions







Backup: Basic ssWW SR Selection

Selection
Exactly two same-sign signal leptons with $p_T > 27$ GeV ($\lvert\eta\rvert < 1.37$ in ee channel)
At least 2 jets with p_T > 25 GeV
$m_{ll} > 20$ GeV
Z peak removal (m_{ee} more than 15 GeV separated from m_Z)
Jet 1 p_T > 65 GeV & jet 2 p_T > 35 GeV
$MET > 30$ GeV
Pass 3 lepton veto (less than 3 baseline leptons)