Searching for soft leptons in compressed spectra with a Boosted Decision Tree

Alyssa Horne, Marcus Snedeker, Joel W. Walker, Bhaskar Dutta, Jason Kumar, Tathagata Ghosh, Pearl Sandick, Patrick Stengel

Overview

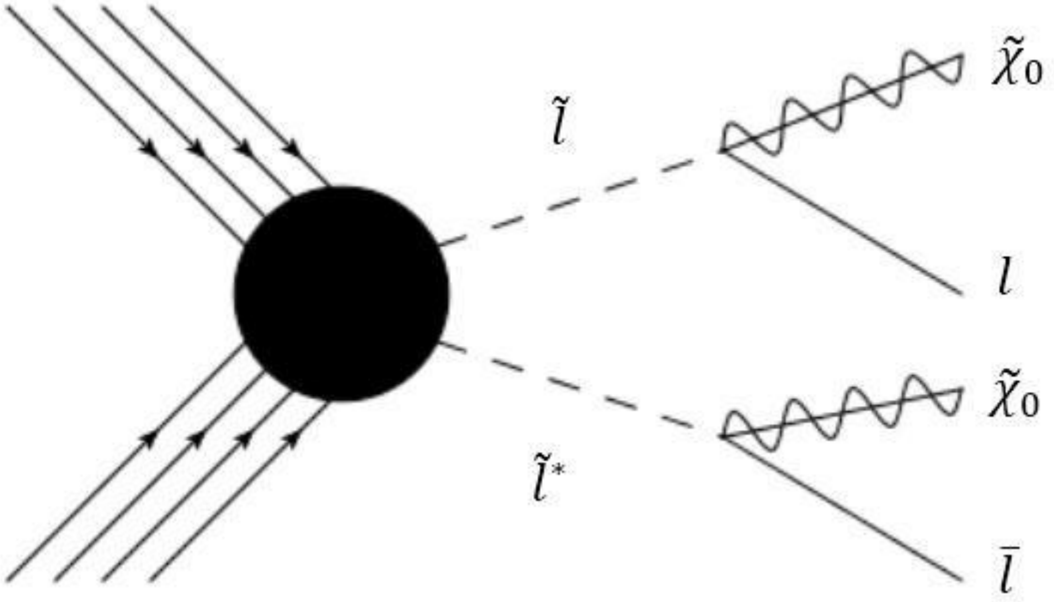
- Compressed Spectra
- Boosted Decision Tree
- Signal and Background Simulation
- Optimizations and Results

Compressed Mass Spectra

- We consider scenarios with new physics pair production followed by leptonic decay to a nearly mass-degenerate invisible daughter, e.g. SUSY slepton to neutralino decay.
- In this case, not much residual energy is available to the visible system.
- We require an ISR jet to boost the soft leptons.
- Monojet + dilepton + Missing E_T

-Baer et al. (arXiv:1409.7058v2),

-Dutta et al. (arxiv: 1706.05339)



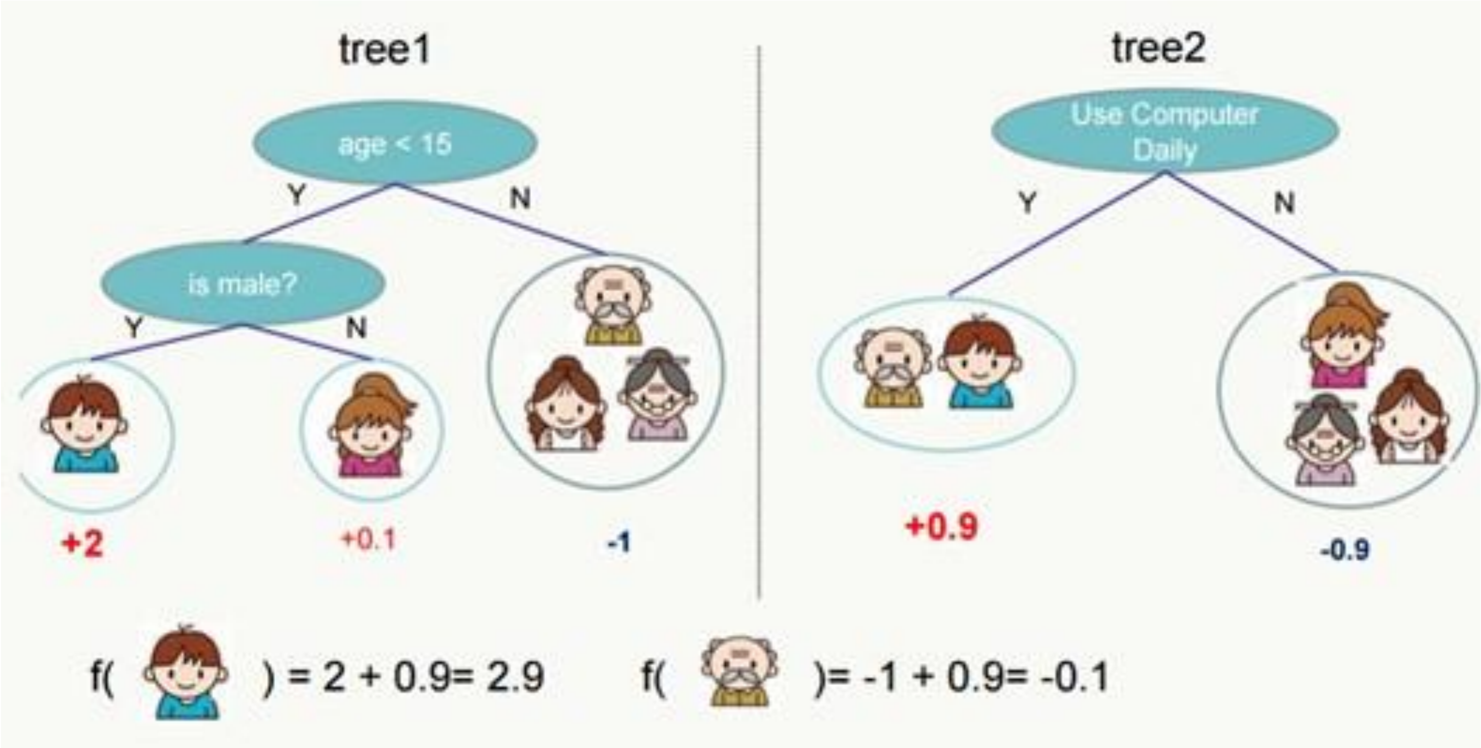
*Ashen and Kebur PHENO 2017

Primary Event Selection

- Require opposite-sign dimuon pair
- Require one and only one jet with PT > 30 GeV
- Require at least 30 GeV of MET
- B-Jet Veto
- Hadronic Tau Veto

Tool of Choice:
Boosted Decision Tree (BDT)

- ***Decision*** – Optimized binary event selection on one feature at a time
- ***Tree*** – Hierarchy of decision forks
- ***Boosted*** – Successive trees iteratively concentrate on events misclassified by prior trees



XGBoost by Tianqi Chen et al. xgboost.ai

Why BDT?

- It performs event selection cuts in an optimized fashion with curated high-level variables.
- Results can be clearly interpreted, since they amount to nothing more than sequential cuts.
- Systematic algorithmic approach eliminates the bias of hand-selected cuts.
- Iterative branching more fully explores the space of possible event selections.
- Sequential tree generation allows for self-improvement via continued training.
- Domain knowledge is incorporated via expert selection of features likely to discriminate.

Signal and Background Simulation

- We have simulated signal for a range of smuon masses between 110 and 300 GeV with neutralino mass gaps from 30 to 60 GeV.

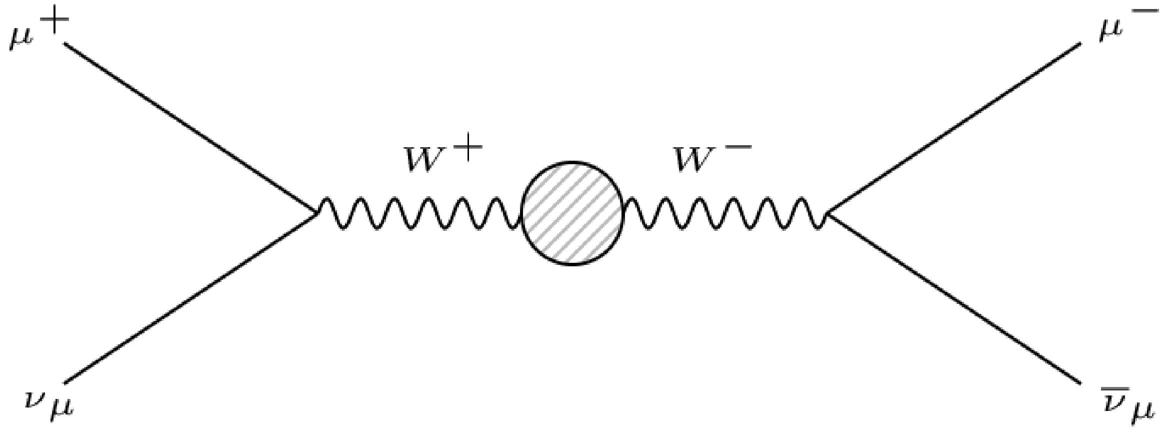
- We have simulated a variety of SM backgrounds, which are grouped into the following five categories representing similar event topologies and cross sections (note that all jets are inclusive):
  - [t jjj] Single top and three jets
  - [t t-bar jj; t W jj] Top, anti-top and two jets, as well as top with W and two jets
  - [ZZ jj; ɣ Z jj; W Z jj; W W jj; W jjj] Vectors
  - [ττ jj] Ditau and two jets
  - [$\ell\ell$ jjj] Two light leptons and three jets

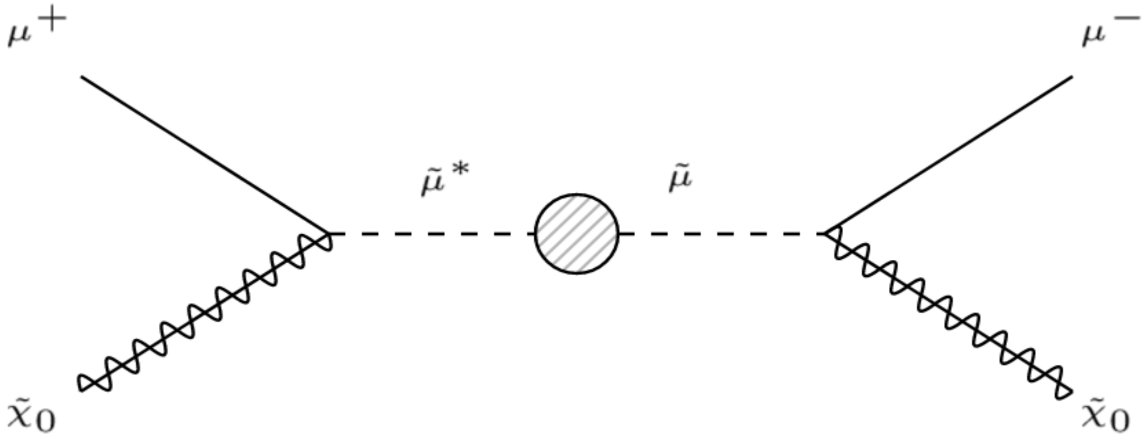
Signal/Background Discrimination Challenges

- One particularly difficult BG to control is the topologically identical WWj process.

Background (WWj)
- Spin 1 parent (W boson)
- (Near) massless daughter (neutrino)

Signal (slepton to neutralino pair)
- Spin 0 parent (smuon)
- Massive daughter (~100 GeV)





*A. Fernando and K. Fantahun PHENO 2017

Features Presented for Training

- MT2: minimal parent mass consistent with visible features, MET, and invisible mass hypothesis
  - We compute MT2 with both a massless daughter and a 100 GeV daughter hypothesis.
  - Useful for discriminating neutrino versus neutralino decay products.
  - arXiv:1412.0618 + PRD.
- Cos(θ*): Cosine of leptonic polar angle in frame with equal leptonic pseudorapidities.
  - Useful for discriminating angular features imprinted via spin of the decaying parent.
  - hep-ph/0511115 + JHEP
- Ditau mass: designed to tag leptonic tau decays, by associating MET with colinear neutrinos.
  - Nucl. Phys. B 297, 221 (1988).
- Elementary kinematic scale and angular variables
  - MET, HT, PTs, masses
  - $\Delta\eta$, $\Delta\varphi$ between particles and MET, as applicable

Questions for BDT Analysis

- Can we meaningfully improve upon prior analysis with 1-dimensional, hand-selected cuts?
- What is the optimal approach to training against a variety of distinct backgrounds?
- How strong should our prior event selections be?

BDT Binary Classification

- We train on 2/3 of the Monte Carlo and reserve 1/3 for test validation.
- We assign training labels of 0 and 1 to background and signal, respectively.
- Once trained, the BDT returns a continuous classification score in the range of 0 to 1 for each validation event.
- We can adjust our signal classification threshold in order to optimize significance.
- We employ the new BDT analysis MInOS.

  - please see presentation by Joel. W. Walker

Measures of Signal Significance

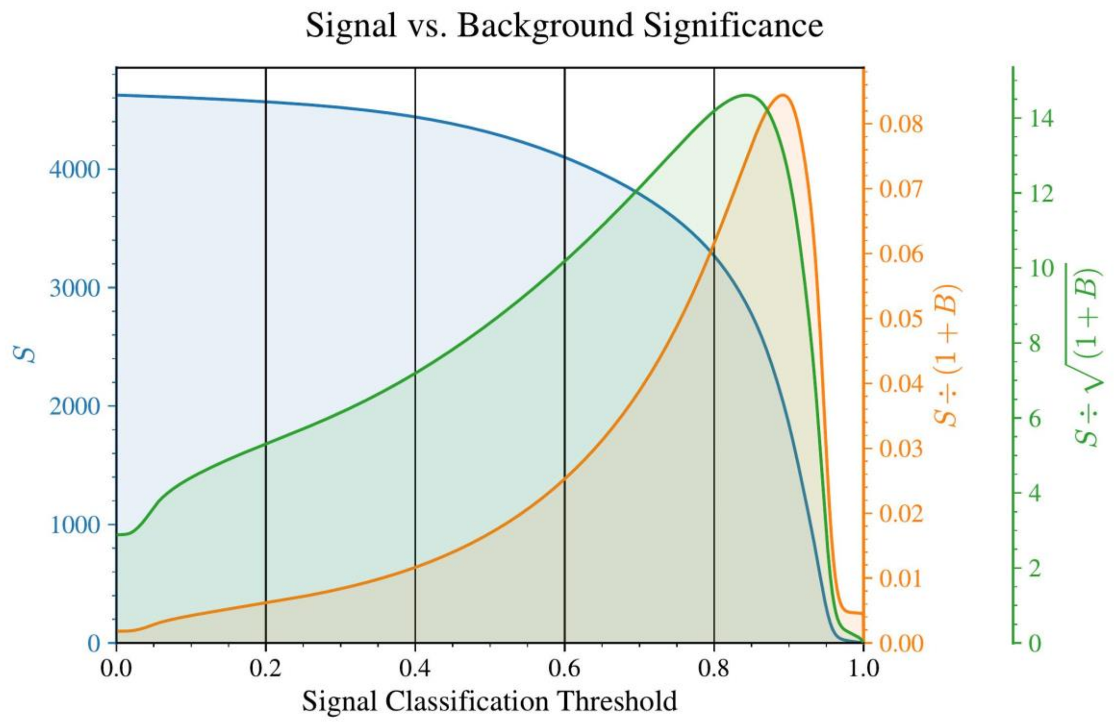
- *S*: the number of truth-level signal events with BDT classification above threshold.
- *B*: the number of truth-level background events with BDT classification above threshold.

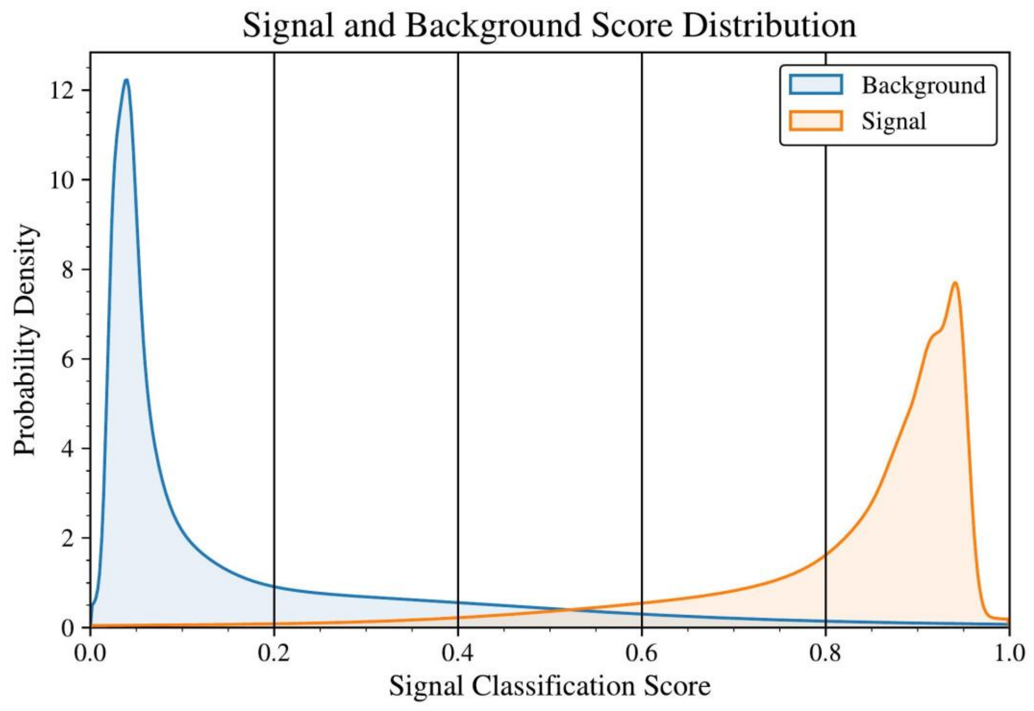
- $\frac{S}{1+B}$: Signal to background ratio is indicative of resiliency to systematic errors.
- $\frac{S}{\sqrt{1+B}}$: Statistical significance measures signal in units of background fluctuation.

- Note: Statistical significance always scales as a square root of luminosity. However, systematic errors limit the extent to which additional data is practically beneficial.

BDT results for 30 GeV Mass Gap

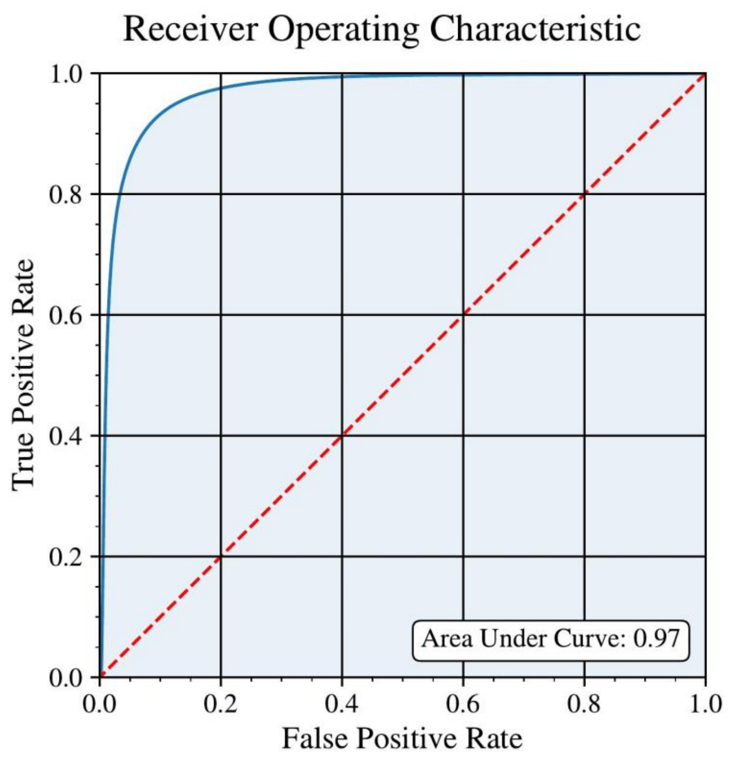
- Signal events are reduced as the classification threshold increases.
- The signal to background ratio and signal significance grow, as background is cut more rapidly.
- This is also apparent in the probability density.

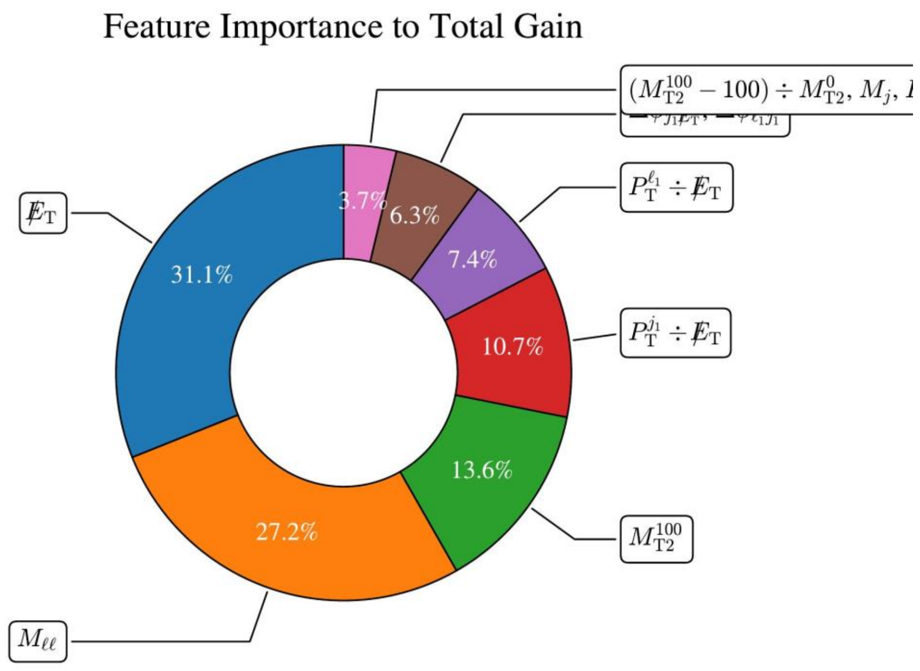




More results for 30 GeV Mass Gap

- The ROC curve likewise shows excellent signal to background discrimination.
- However, this is an illusion, because the majority of training has been against simple high cross section backgrounds.
- The feature importance reveals a concentration on missing energy and the Z-boson peak.





Secondary Event Selection

Applying more "obvious" cuts at the outset can free the BDT to focus on more subtle features.

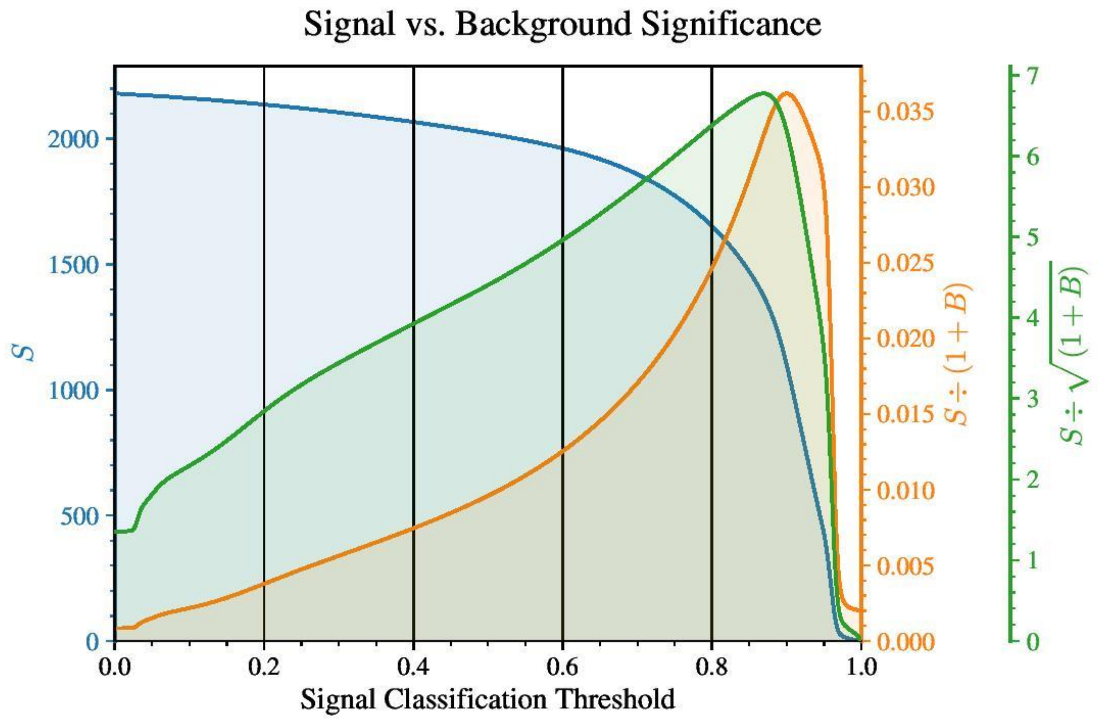
We can do much better by prior assumption of cuts listed to the right, which were previously identified as universally beneficial.

-Dutta et al. (arxiv: 1706.05339)

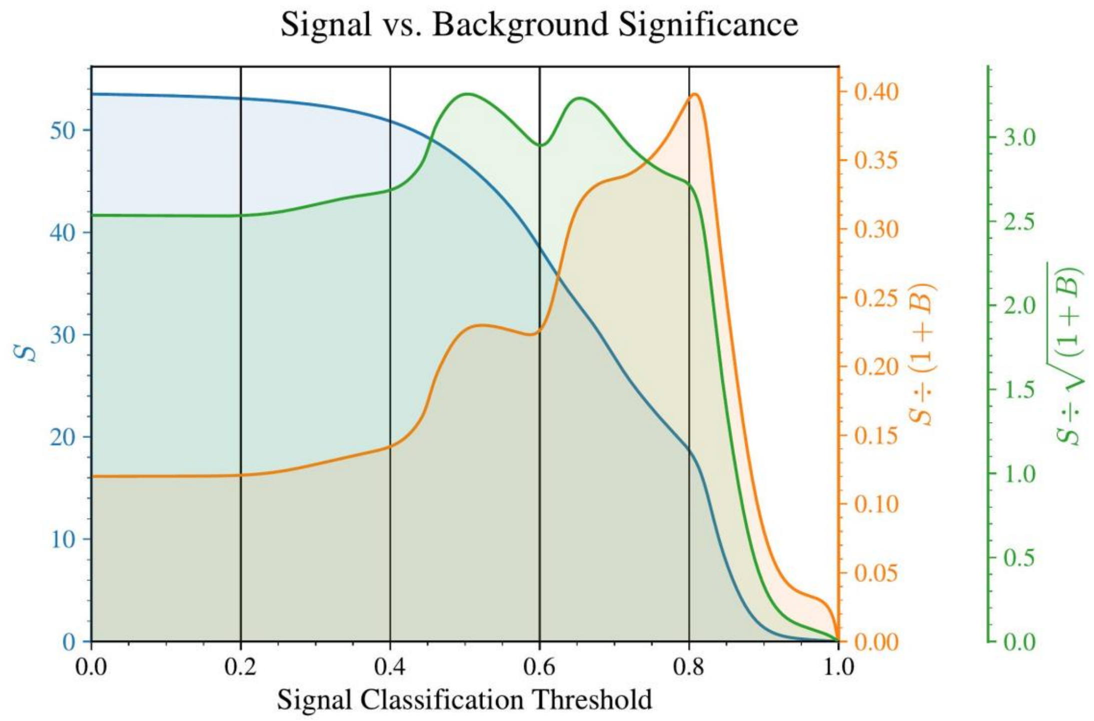
- Dilepton mass exclusion for Mz ± 10 GeV
- $\cos(\theta^*) < 0.5$
- Ditau mass > 125 GeV
- MET > 125 GeV
- Monojet PT > 125 GeV

Primary vs. Secondary Event Cuts for 60 GeV

- The signal to background ratio has improved by an order of magnitude.
- The signal significance has declined…
- These cuts are already so hard that the significance was reasonable prior to action from the BDT!



Primary



Secondary

Merged vs. Split Training

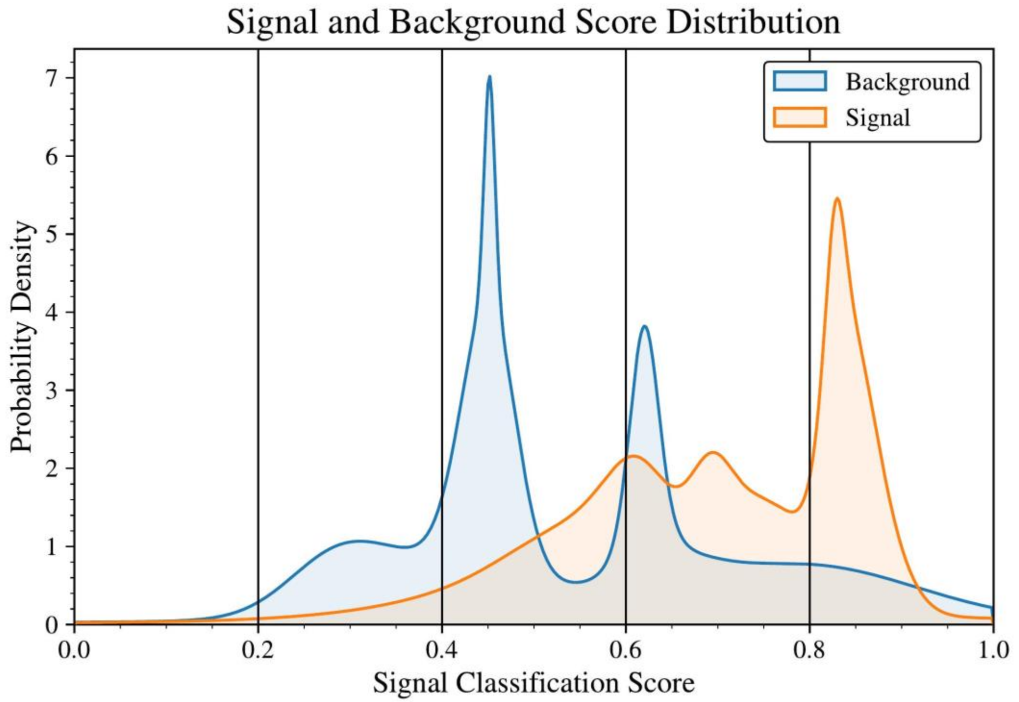
Because the five background categories differ from signal in unique ways, and have distinct cross sections, we tried training individually against each.

The ensemble of trainings was recombined into a single score via a cross section weighted algebraic mean.

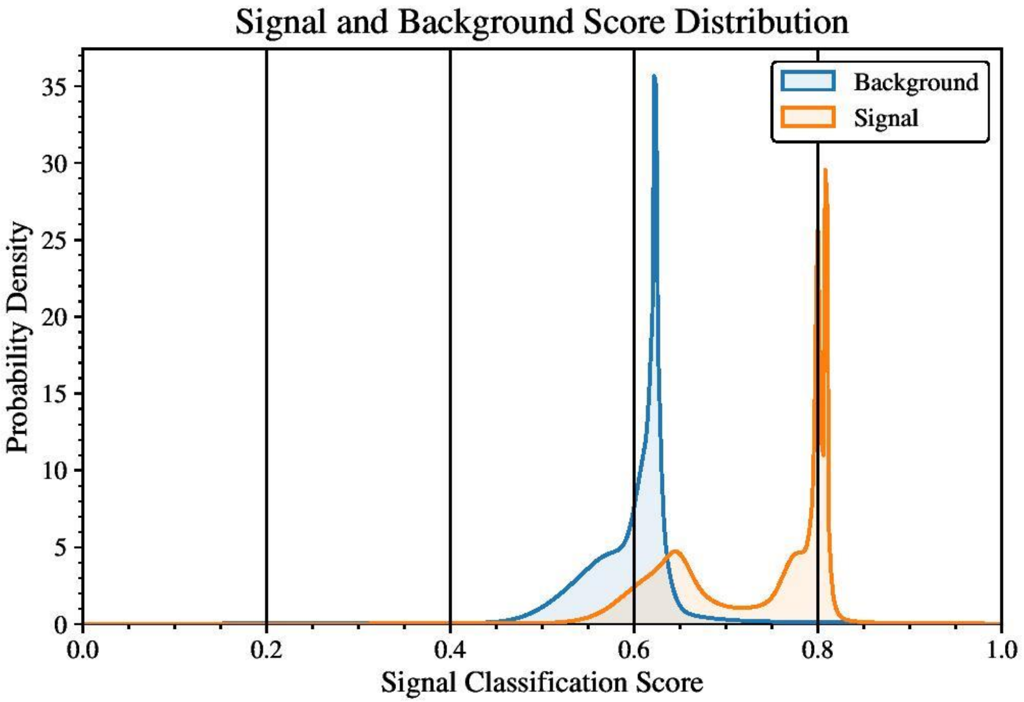
This yielded the best results of all attempted approaches.

Merged vs. Split Training for 60 GeV Mass Gap

- The distinction between signal and background is much sharper for split training.
- Note: Hardness of the secondary cuts leads to discretization. This is to be addressed by additional Monte Carlo.



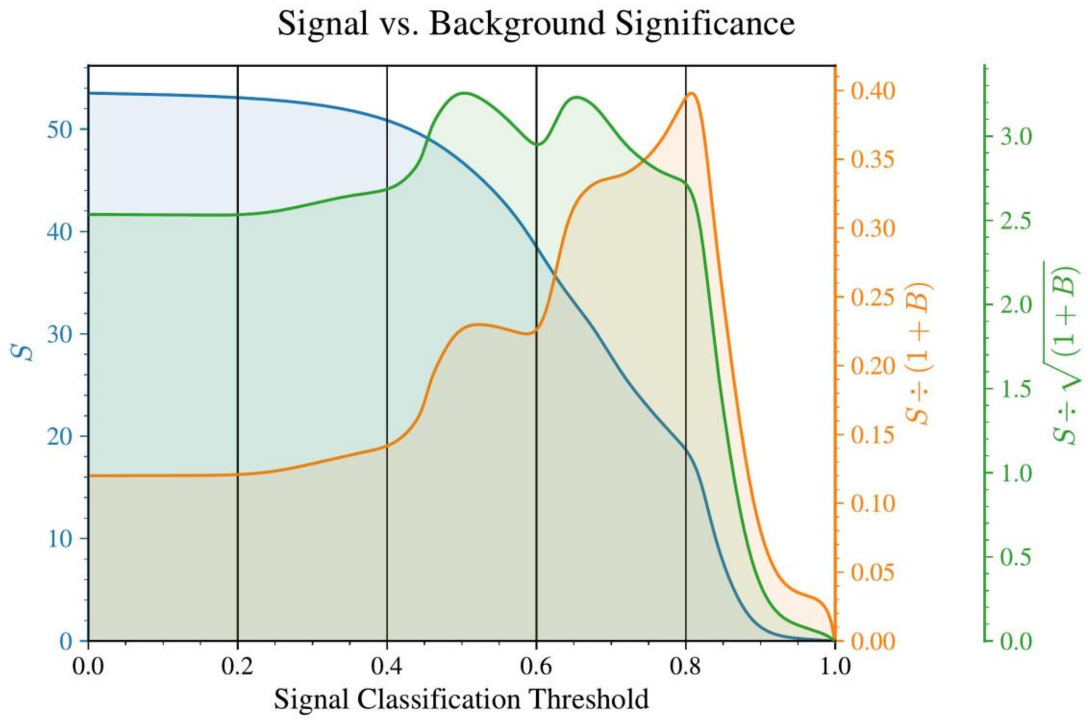
Merged



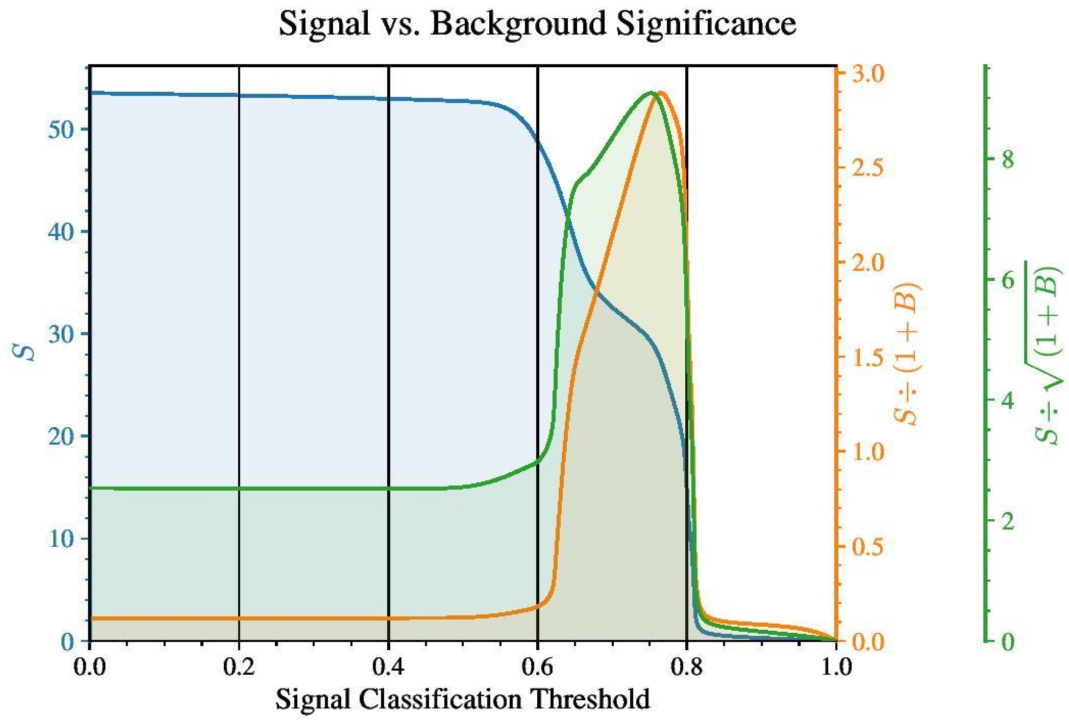
Split

More Merged vs. Split Results for 60 GeV Mass Gap

- Both signal to background ratio and statistical significance are meaningfully improved.



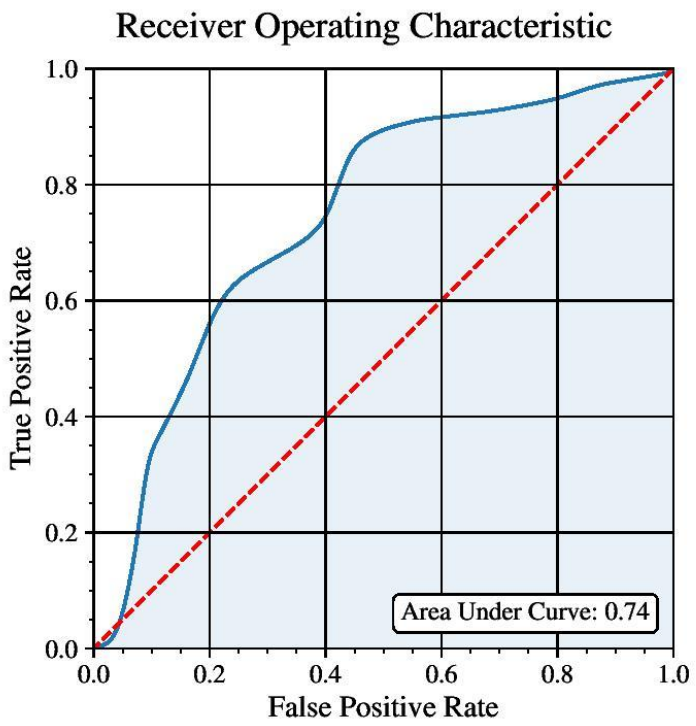
Merged



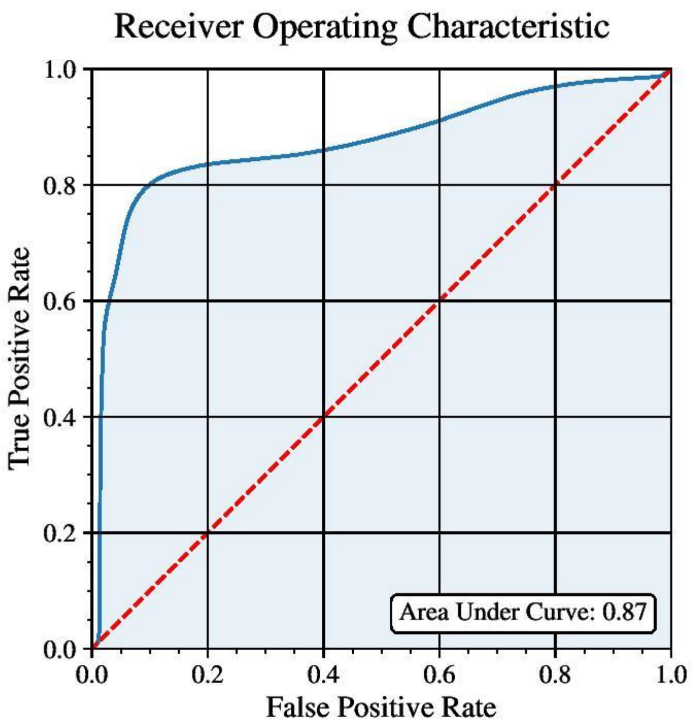
Split

More Merged vs. Split Results for 60 GeV Mass Gap

- … as is the ROC Area Under Curve.



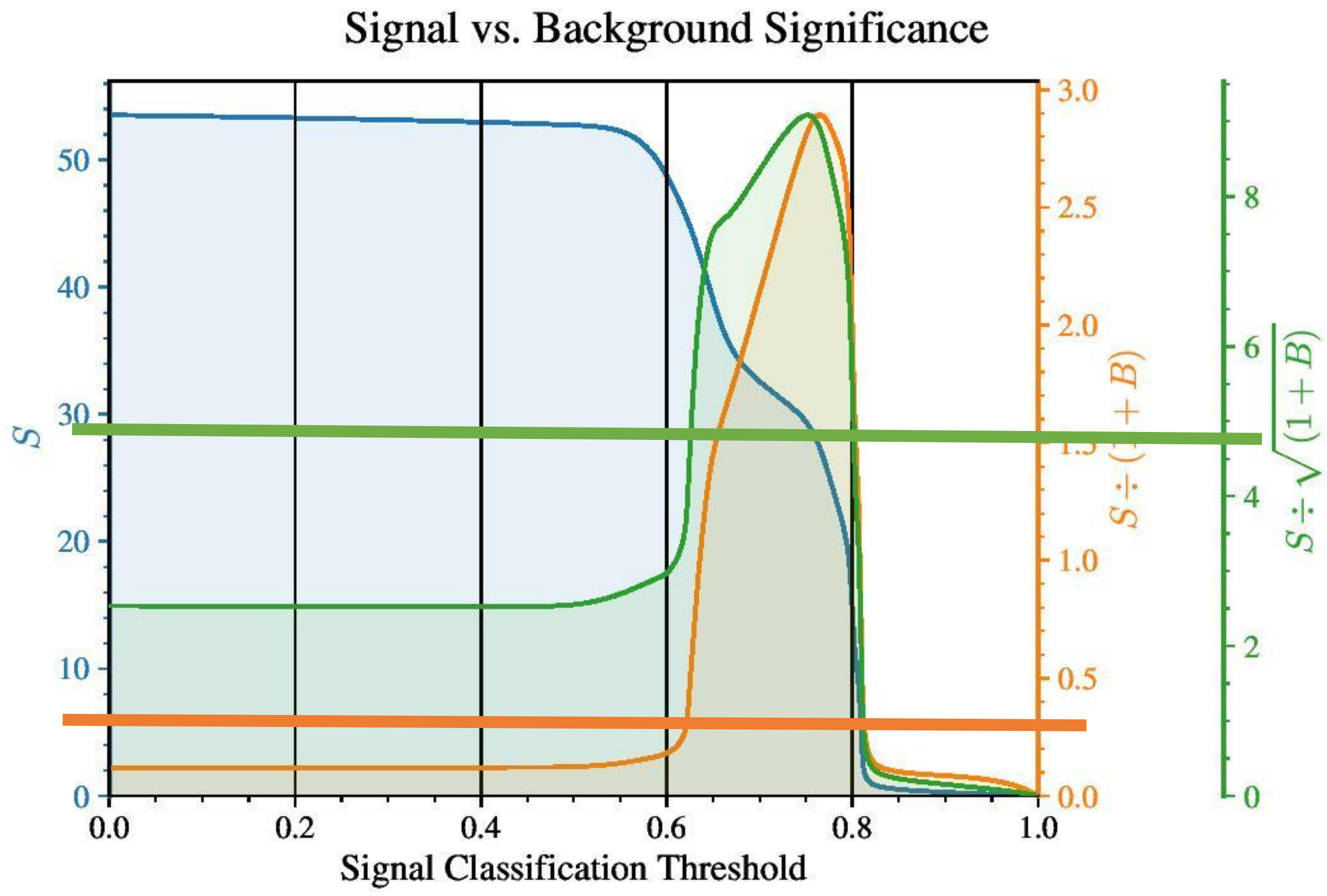
Merged



Split

Comparison to Prior Study for 60 GeV

- Green and orange lines are statistical significance and signal to background ratio, respectively, from the previous study.
- The BDT enhances signal visibility and resiliency to systematic errors

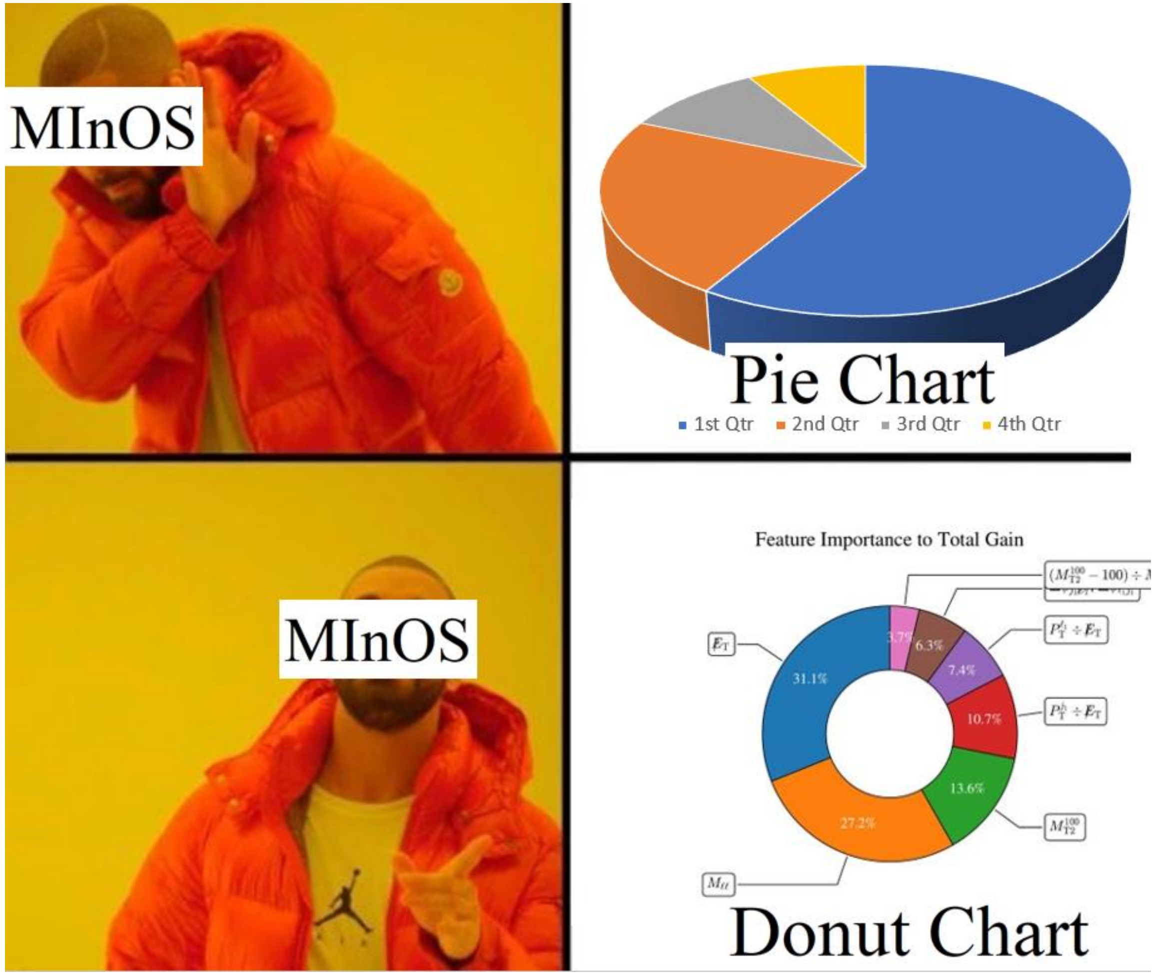


Results and Conclusions

- The BDT approach has provided substantial improvement relative to hand-selected 1-dimensional cuts.
- However, naïve application of BDT's may do much worse than careful manual event selection.
- Split training against specific backgrounds with recombination provides better results than single training.
- Prior implementation of obvious cuts (e.g. Z-Window) allows the BDT to focus on more subtle discrimination.

THANK YOU!

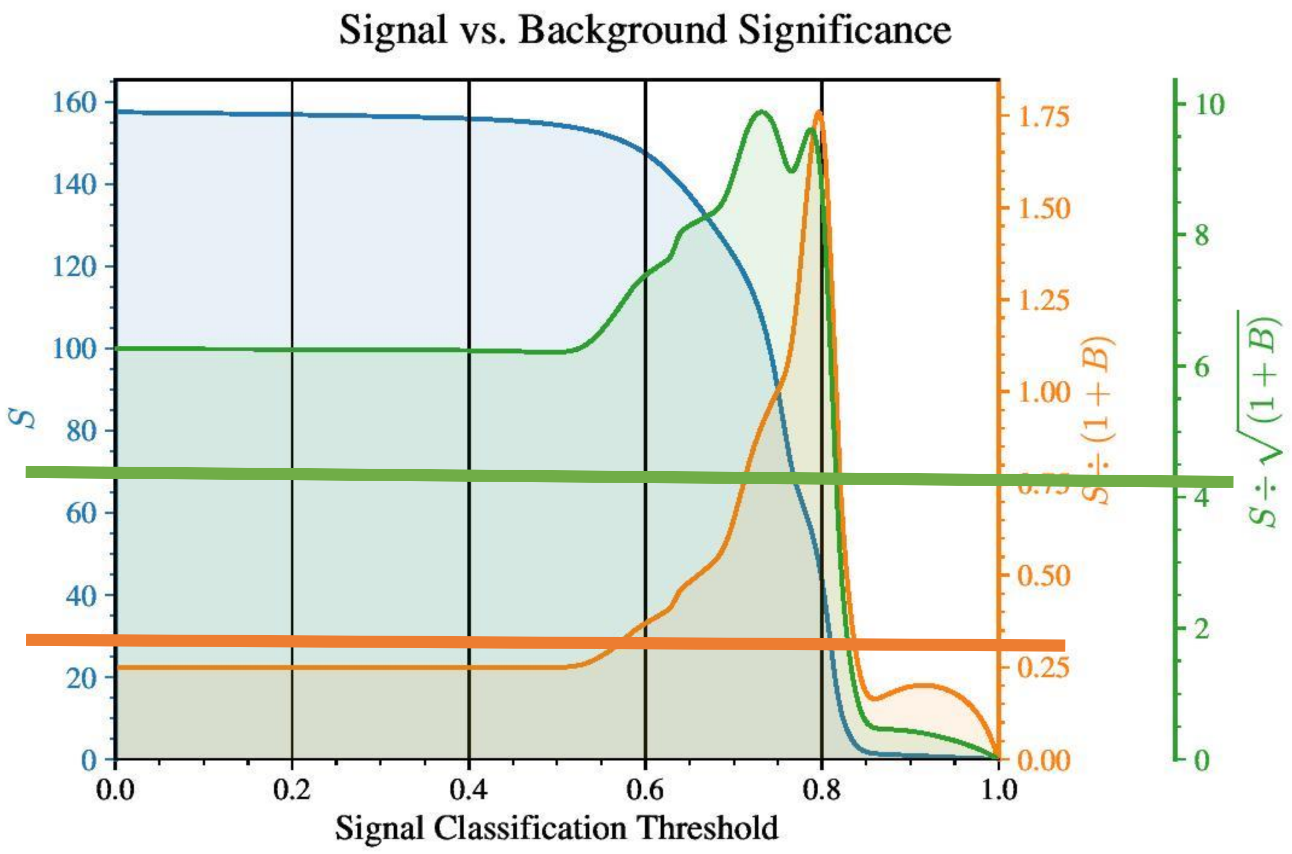
- To PHENO 2021
- To SHSU for permitting us to use their supercluster
- To the amazing Dr. Walker
- To the other professors on the project
- To Ashen and Kebur for their hard work on the original project
- Y'all for attending our talk!



Improvement for 30 GeV

- Green line is statistical significance from the previous study
- Orange line is the systematic Significance from the previous study

- Horizontal lines represent the values found in previous study
- Green line at 4.4
- Orange line at 0.30
- BDT preserves stronger Signal to BG ratio and resilient from systematic errors

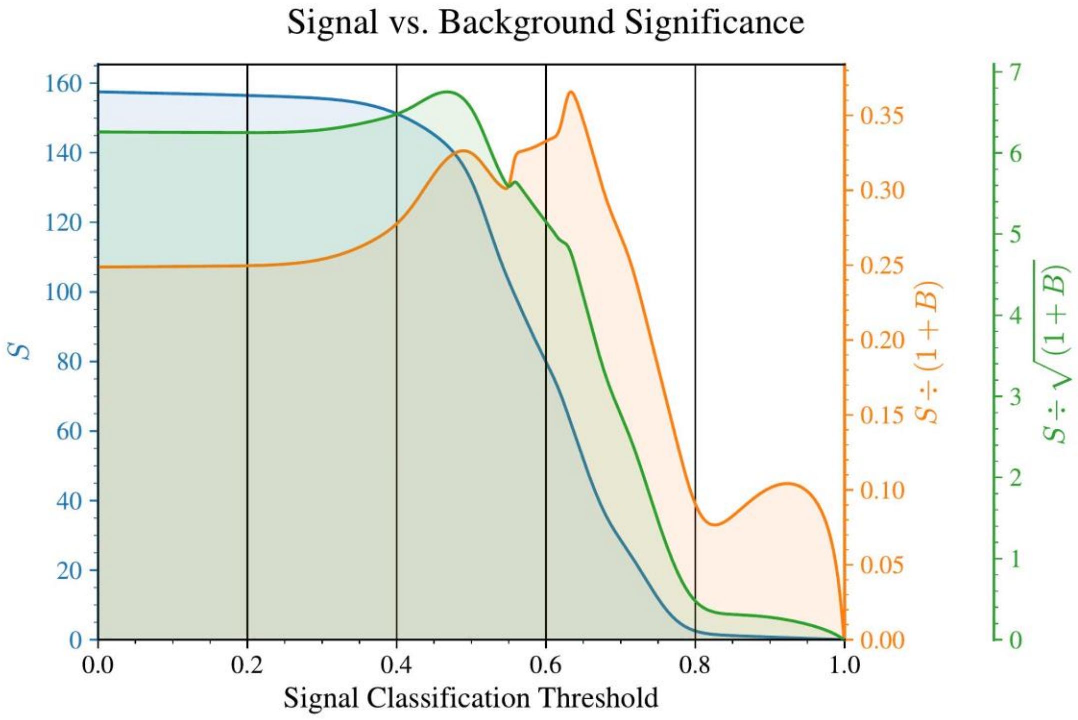


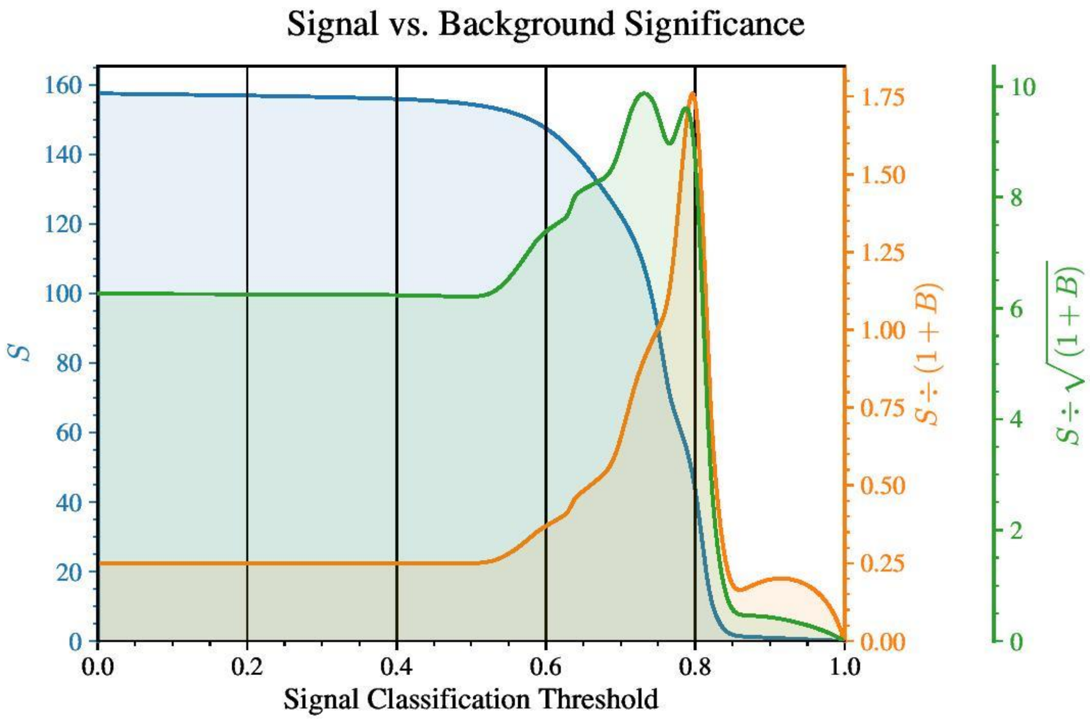
Tertiary cuts for smaller mass gaps

Selection	$t\bar{t}jj$	$ZZjj$	$WZjj$	$WWjj$	S^{110}_{10}	S^{110}_{20}	S^{110}_{30}	S^{110}_{40}	S^{110}_{50}	S^{110}_{60}
$M_{T2}^{WW} < 1$ GeV	2.4×10^{-1}	3.9×10^{-3}	7.0×10^{-2}	8.6×10^{-1}	2.7×10^{-1}	3.0×10^{-1}	2.8×10^{-1}	2.1×10^{-1}	2.0×10^{-1}	1.9×10^{-1}
$0.8 < P_T^j \div \not{E}_T < 1.4\}$	1.9×10^{-1}	3.3×10^{-3}	3.6×10^{-2}	6.0×10^{-1}	2.7×10^{-1}	2.7×10^{-1}	2.1×10^{-1}	1.5×10^{-1}	1.1×10^{-1}	1.2×10^{-1}
$\Delta\phi(\not{E}_T, j) \div \pi > 0.95$	1.4×10^{-1}	2.7×10^{-3}	2.3×10^{-2}	4.4×10^{-1}	2.6×10^{-1}	2.3×10^{-1}	1.6×10^{-1}	1.0×10^{-1}	6.7×10^{-2}	6.6×10^{-2}
$\Delta\phi(\ell_1, j) \div \pi > 0.5$	7.8×10^{-2}	2.2×10^{-3}	1.9×10^{-2}	3.7×10^{-1}	2.5×10^{-1}	2.1×10^{-1}	1.4×10^{-1}	8.9×10^{-2}	6.0×10^{-2}	6.0×10^{-2}
Events at $\mathcal{L} = 300$ fb^{-1}	23.4	0.7	5.7	111.3	75.0	63.0	42.0	26.7	18.0	18.0
$S \div B$	-	-	-	-	0.53	0.45	0.30	0.19	0.13	0.13
$S \div \sqrt{B}$	-	-	-	-	6.3	5.3	3.5	2.2	1.5	1.5
Poisson Significance	-	-	-	-	6.0	5.2	3.6	2.5	1.9	1.9

Merged vs Multiple Training Secondary Event Cuts for 30GeV

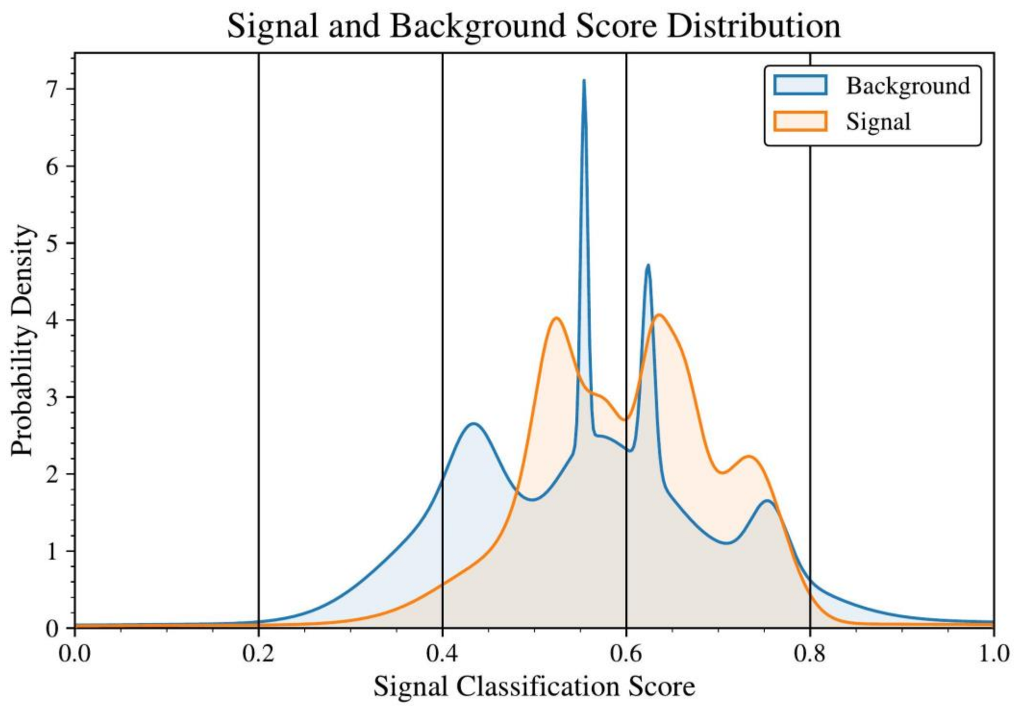
- Right is algebraic and the systematic has improved from roughly .038 to about 2.9

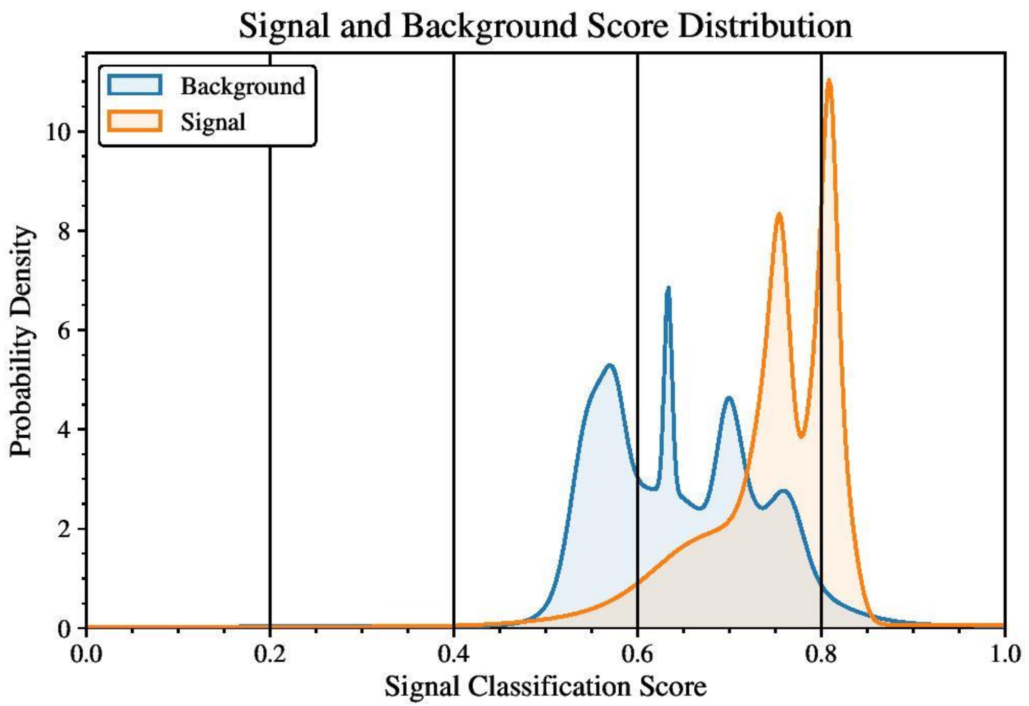




Merged vs Multiple Training Secondary Event Cuts for 30GeV

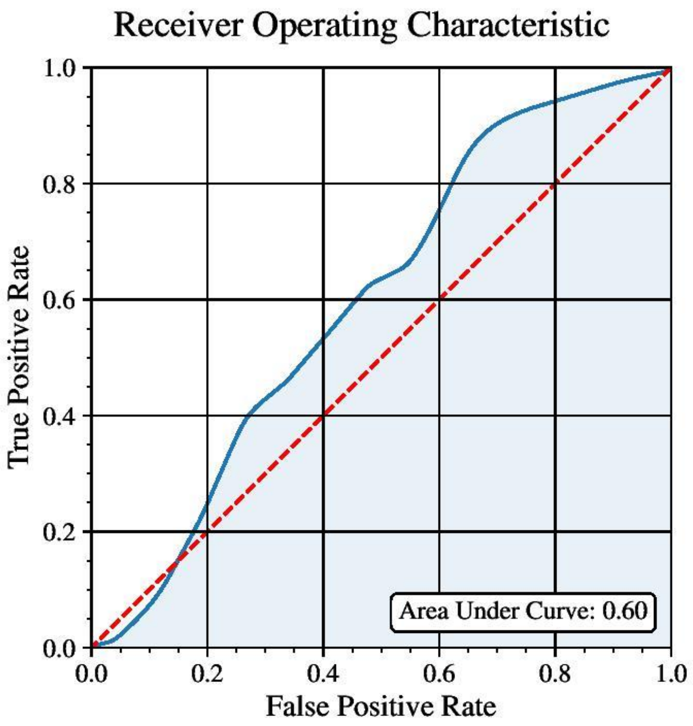
- Right is algebraic and smaller separation between Signal and BG because more nontrivial selections were inputted
- Clear distinction apearing with multiple training

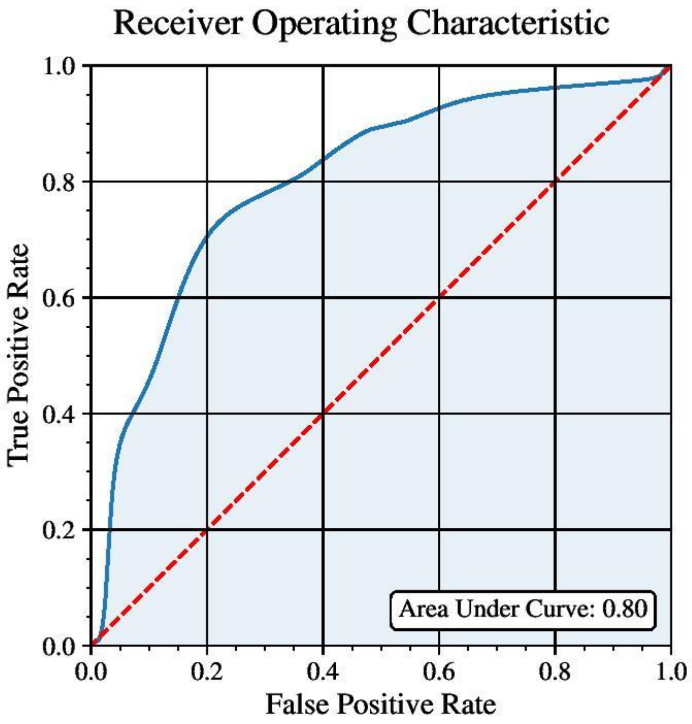




Multiple BG Training vs One BG Training 30 GeV

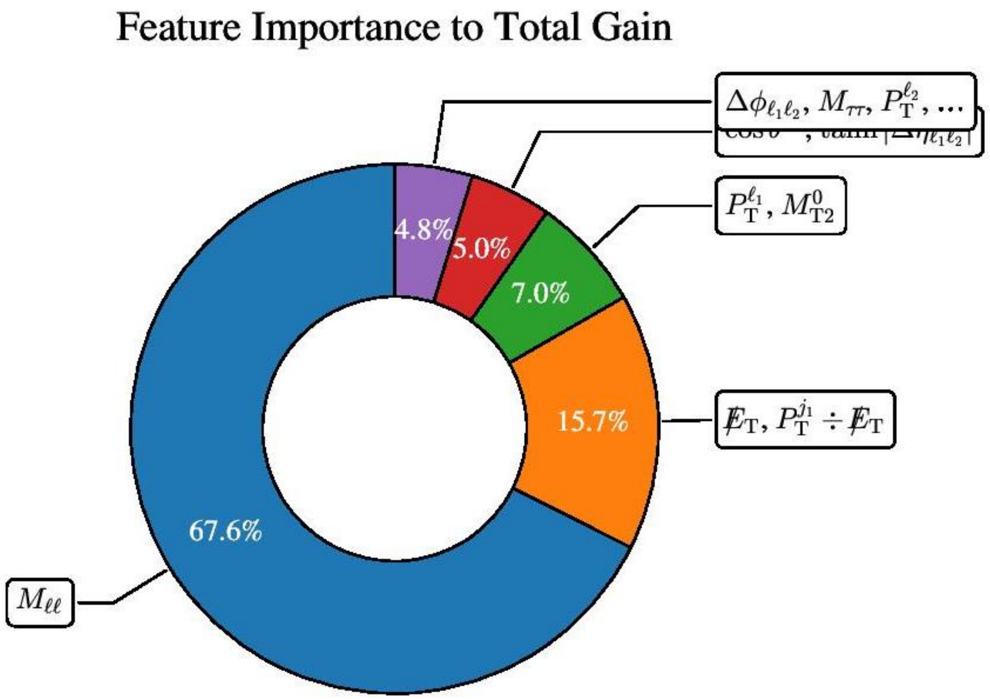
- The left is the merged training.
- The right is an algebraic sum of 5 individual trainings.

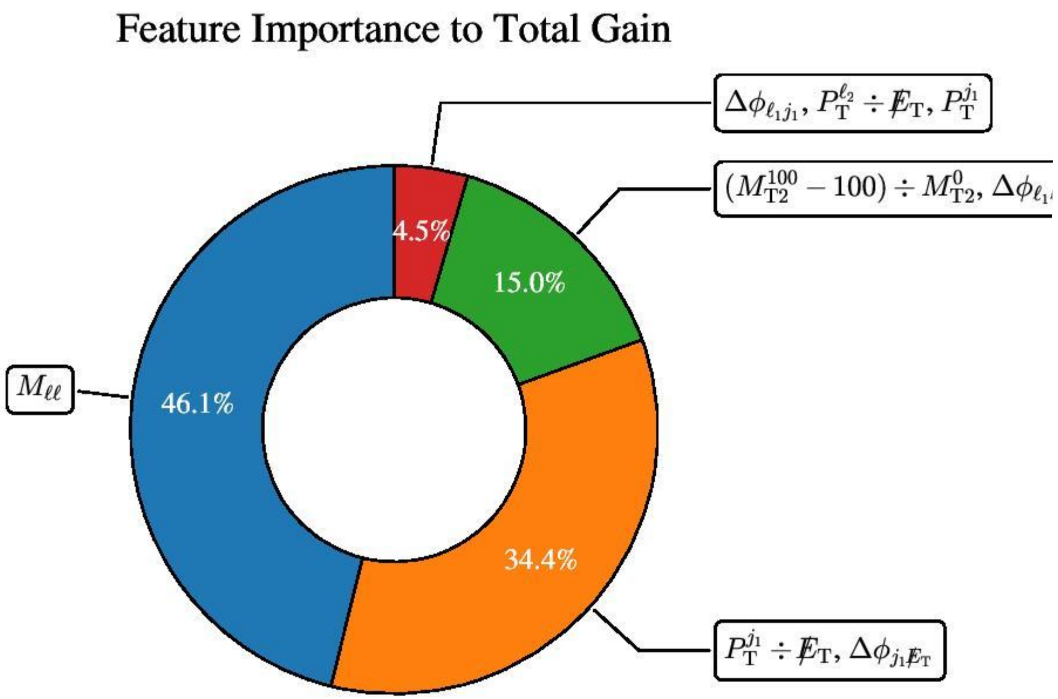




Feature Importance of Preliminary and Secondary

- DiVector Importance at 60 GeV
- Left is the Preliminary





Extra, Extra, Read all about it

- [[Add extra stuff here]]