



Department of Physics, Shandong University

# Compressed EWK study(ISRC1N2)

Chengxin Liao  
[liaocx@ihep.ac.cn](mailto:liaocx@ihep.ac.cn)

May, Wed 14, 2025



# Tasklist

- PAR slides already upload(some figures need to update)
- Prepare graduation defense in next week(no exactly date)
- BSc thesis: <https://www.overleaf.com/project/674e7119837a2580151a0868>  
(Fill the blank area left before, submit before graduation defense)



山东大学  
SHANDONG UNIVERSITY

# Backup



# Top(HH) Binary

== 2 medium tau  
== 0 lepton  
METtrig && MET  $\geq$  200  
OS  
nBaseJet  $\geq$  1  
Jet pt  $>$  100GeV  
Mtt\_reco  $<$  40GeV && Mtt\_reco  $>$  130GeV

$\geq$  1 bTag(improve top events)

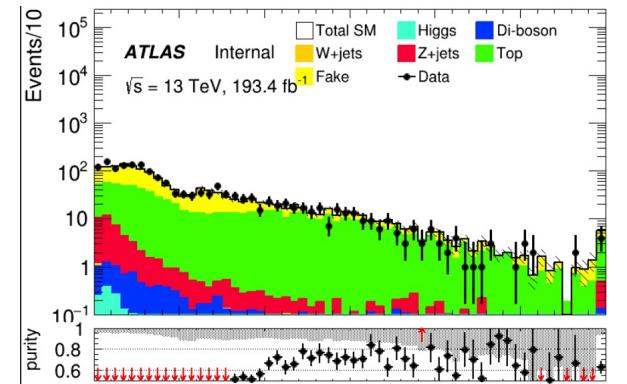
mtx\_tau1  $\geq$  200(to improve top purity)

CR: C1N2 score  $\leq$  -0.8

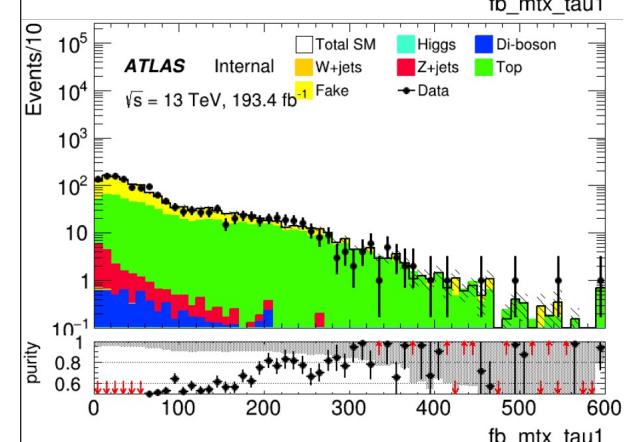
VR: -0.8  $\leq$  C1N2 score  $\leq$  0.7

Same with pre-selection

CR after score cut



VR after score cut



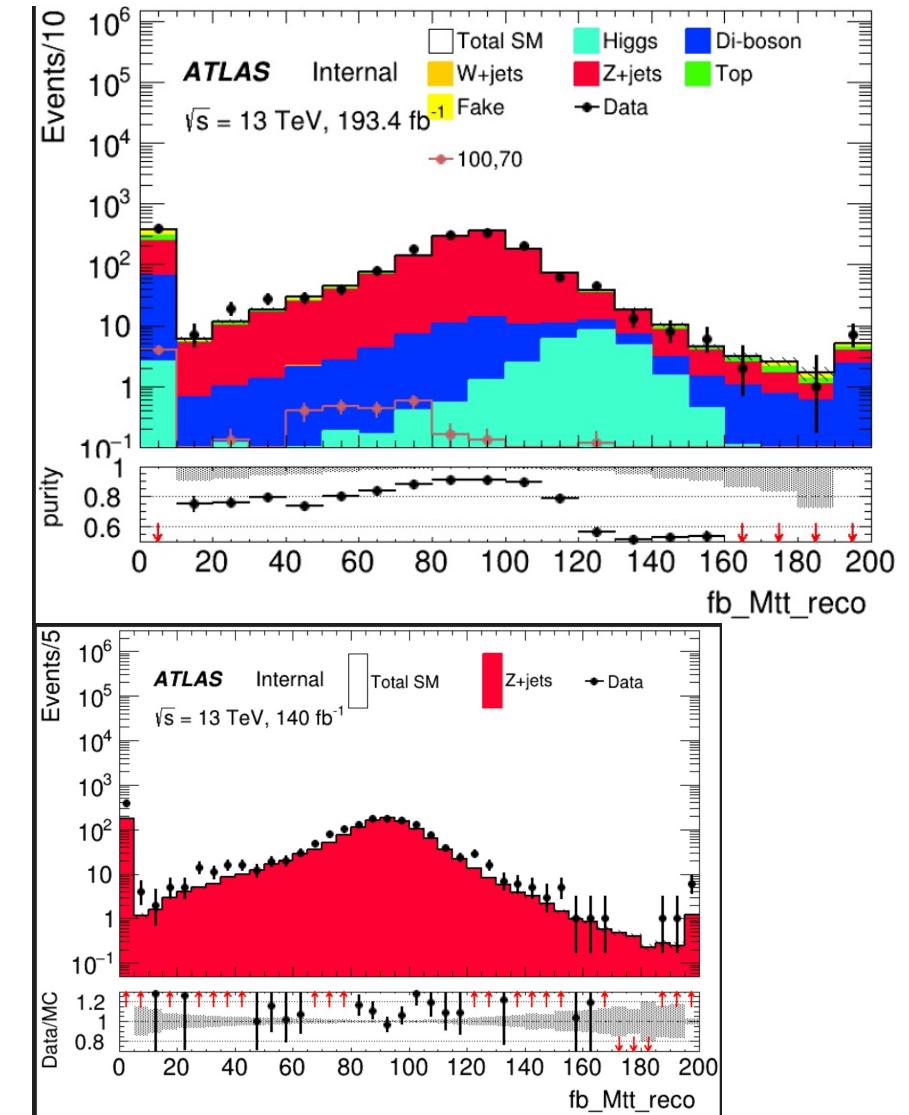
Region	TotalBkg	Top	purity	Data	Data/Bkg
CR	290.841+-7.459	206.202+-4.807	0.709	264	0.91
VR	162.58+-5.438	131.125+-3.869	0.806	161	0.99

# Zjets(LH) Binary

== 1 medium tau  
 >= 1 lepton  
 METtrig && MET >= 200  
 OS  
 bVeto  
 nBaseJet >= 1  
 Jet pt > 100GeV  
 dRtt <= 0.6(improve purity of Zjets)  
 C1N2 score <= 0.7(Othogonal with SR)  
 CR: Mlt reco >= 80 && Mlt reco <= 110  
 VR: (Mlt reco >= 40 && Mlt reco < 80) || (Mlt reco > 110 && Mlt reco < 130)

Same with pre-selection

Region	TotalBkg	Zjets	purity	Data	Data/Bkg
CR	917.505+-6.668	823.435+-4.028	0.897	908	0.99
VR1	592.091+-5.617	518.684+-3.244	0.876	635	1.07
VR2	128.617+-2.384	87.882+-1.217	0.683	120	0.93
VR	720.708+-6.102	606.566+-3.464	0.842	745	1.034



# Top(LH) Binary

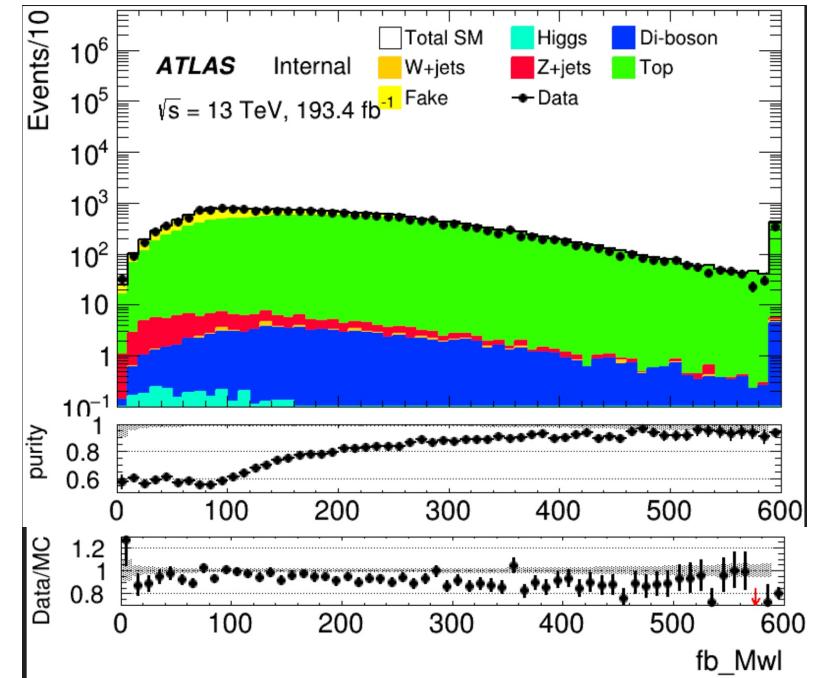
$\geq 1$  medium tau  
 $\geq 0$  lepton  
 METtrig && MET  $\geq 200$   
 OS  
 nBaseJet  $\geq 1$   
 Jet pt  $> 100\text{GeV}$   
 Mtt\_reco  $< 40\text{GeV}$  && Mtt\_reco  $> 130\text{GeV}$

$\geq 1$  bTag(improve top events)  
 C1N2 score  $\leq 0.7$ (Orthogonal with SR)

CR:  $450 < M_{inv}(l, MET) < 560$

VR:  $400 < M_{inv}(l, MET) < 450$

Same with pre-selection



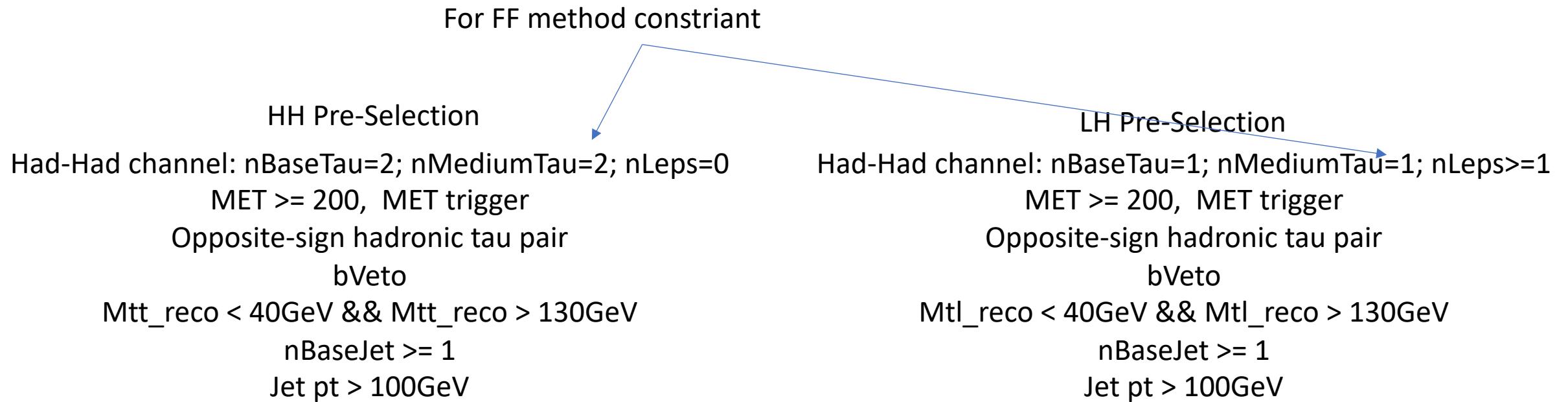
Try to define by using C1N2 score

Region	TotalBkg	Top	purity	Data	Data/Bkg
CR	888.156+11.505	830.914+9.676	0.935	785	0.884
VR	902.627+11.837	828.424+9.679	0.912	792	0.878

# Introduction

Reference point: C1 mass = 100GeV, N2 mass = 70 GeV

SR optimization: BDT method, optuna to auto-optimize hps



# Binary class(HH)

Hyperparameters: Ntrees = 200, MaxDepth = 6, MinNodeSize = 2%, Learning rate = 0.03(initial setting)

## Feature engineering:

Select a simple model and put all features into model, choose Top 30 vars based on importance list, drop high correlated vars

## Final feature list:

: Rank	: Variable	: Variable Importance
: 1	: fb_dEtatt	: 5.153e-02
: 2	: fb_dRtt	: 4.318e-02
: 3	: fb_dRMax_xt	: 4.248e-02
: 4	: fb_dPhitt	: 4.228e-02
: 5	: fb_MIA	: 4.205e-02
: 6	: fb_METsig	: 3.979e-02
: 7	: fb_dPhizxe	: 3.972e-02
: 8	: fb_dPhiztt	: 3.942e-02
: 9	: fb_frac_MET_tau1	: 3.735e-02
: 10	: fb_dPhiMin_xt	: 3.513e-02
: 11	: fb_dPhiMin_tj1	: 3.512e-02
: 12	: fb_MT2_150	: 3.494e-02
: 13	: fb_frac_MET_MeffInc_40	: 3.474e-02
: 14	: fb_dRMin_tj	: 3.467e-02
: 15	: fb_eta_tau2	: 3.454e-02
: 16	: fb_frac_MET_tt	: 3.452e-02
: 17	: fb_frac_MET_Meff	: 3.408e-02
: 18	: fb_dPhit2x	: 3.277e-02
: 19	: fb_dPhiMax_xt	: 3.207e-02
: 20	: fb_dRt2x	: 3.131e-02
: 21	: fb_dPhit1x	: 3.089e-02
: 22	: fb_frac_MET_tau2	: 3.085e-02
: 23	: fb_Mll	: 2.960e-02
: 24	: fb_MET_Jet	: 2.734e-02
: 25	: fb_sum_cos_dphi	: 2.530e-02
: 26	: fb_pt_Vframe	: 2.272e-02
: 27	: fb_Pt_tt	: 1.912e-02
: 28	: fb_MstauA	: 1.881e-02
: 29	: fb_Proj_t1	: 1.594e-02
: 30	: fb_Proj_tt	: 1.427e-02
: 31	: fb_MCT	: 1.345e-02

Weight choose: no weight, abs(weight)

No weight have better performance  
but abs(weight) fit our analysis requirement

Split strategy: Separate entries by using mod 5, for Fake bkg, if separate follow sequence, all weighted entry will split into first fold

# Binary class(HH)

Hyperparameter tune:  
use optuna to auto-optmize

constraint:

average of AUC need to  $\geq 0.6$

penalty function:  $\text{score} = \text{test\_auc} - 0.3 * \text{auc\_gap}$  ( $\text{auc\_gap} = \text{abs}(\text{train\_auc} - \text{test\_auc})$ )  
 $\text{maximum}(\text{score})$

Class: C1N2, bkg

$\text{Test\_auc} = \sum \{\text{Test\_auc\_class}\}$   
 $\text{Train\_auc} = \sum \{\text{Train\_auc\_class}\}$

## Grid Search

Ntrees: [200, 300, 400]

MaxDepth: [4, 6, 8, 10]

MinNode: [1, 3, 5, 7]

Learning rate: [0.001, 0.005, 0.01, 0.05, 0.1]



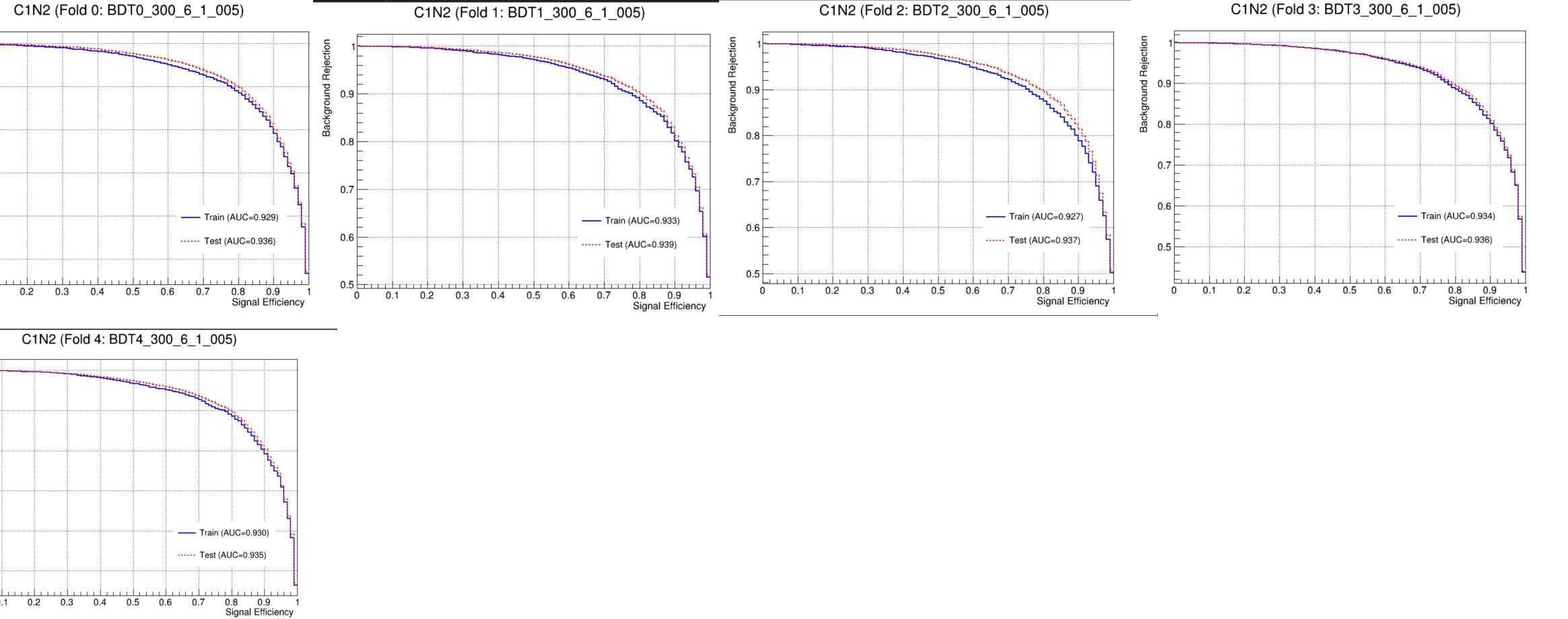
Best one: Ntree=300, MaxDepth=6, MinNode=1%, Learning Rate=0.05



There still have rooms to optimize for lr

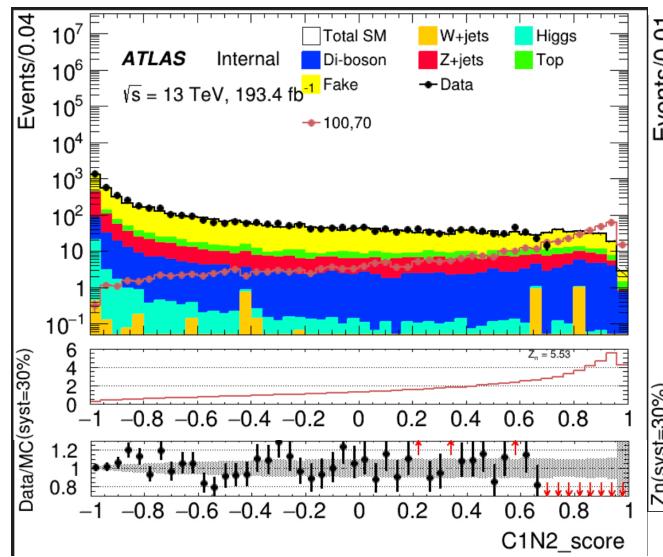
# Binary class(HH)

## Overfit Check

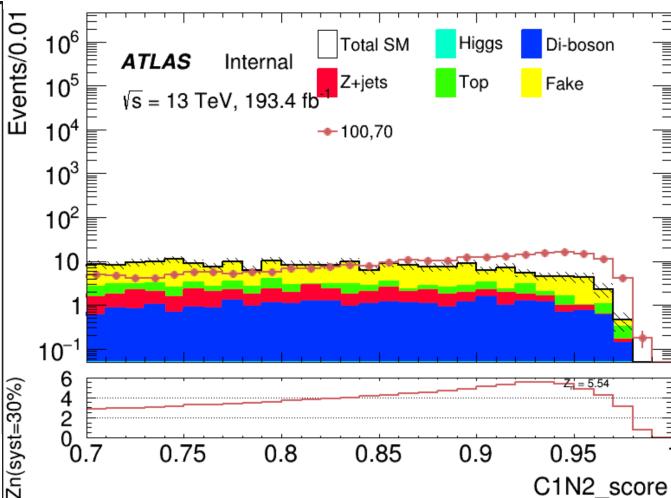


# SR(HH) Binary

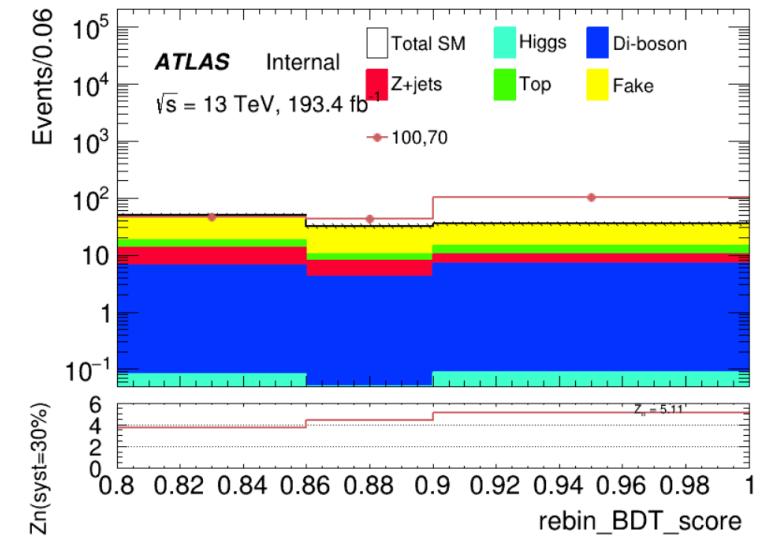
50 bins



Cut at 0.8



rebin



Sum Zn = 7.65

Bin Range	Zn	C1N2 (100_70) Yield $\pm$ Error	VV Yield $\pm$ Error	Top Yield $\pm$ Error	Fake Yield $\pm$ Error	Higgs Yield $\pm$ Error	Zjets Yield $\pm$ Error	Wjets Yield $\pm$ Error	Total Bkg Yield $\pm$ Error
[0.80,0.85]	3.65	$46.569 \pm 1.327$	$6.508 \pm 0.368$	$4.719 \pm 0.686$	$31.458 \pm 3.898$	$0.080 \pm 0.023$	$6.659 \pm 0.416$	$0.978 \pm 0.978$	$50.402 \pm 4.042$
[0.85,0.90]	4.37	$43.128 \pm 1.283$	$4.124 \pm 0.341$	$2.663 \pm 0.546$	$21.525 \pm 3.292$	$0.045 \pm 0.019$	$3.564 \pm 0.286$	$0.000 \pm 0.000$	$31.921 \pm 3.354$
[0.90,1.00]	5.11	$101.059 \pm 1.961$	$6.855 \pm 0.398$	$4.331 \pm 0.731$	$20.438 \pm 2.976$	$0.086 \pm 0.031$	$3.207 \pm 0.250$	$0.000 \pm 0.000$	$34.917 \pm 3.118$

# Zjets(HH) Binary

== 2 medium tau

== 0 lepton

METtrig && MET  $\geq 200$

OS

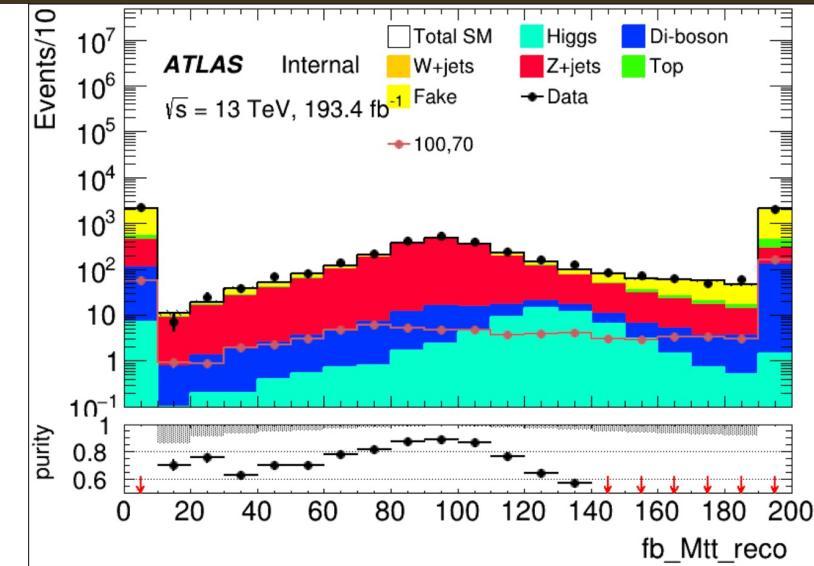
bVeto

nBaseJet  $\geq 1$

Jet pt  $> 100\text{GeV}$

C1N2 score  $\leq 0.7$ (Othogonal with SR)

Same with pre-selection



CR: Mtt reco  $\geq 80 \text{ && } \text{Mtt reco} \leq 110$

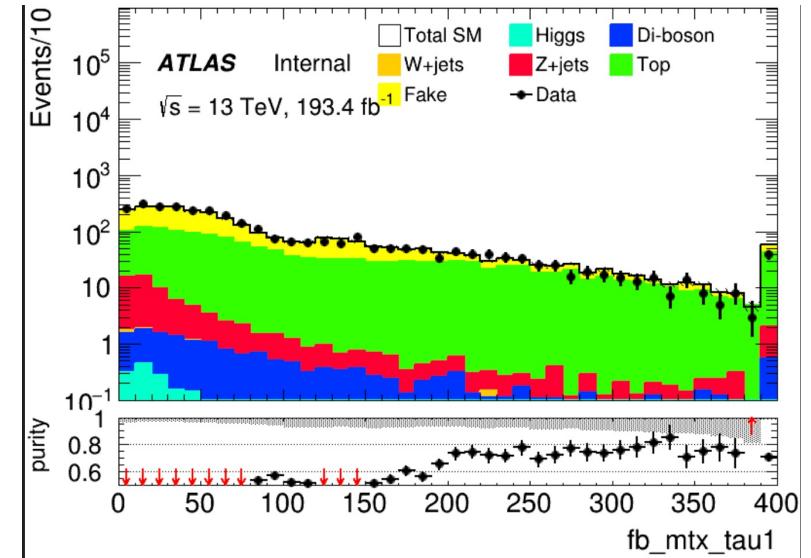
VR: (Mtt reco  $\geq 40 \text{ && } \text{Mtt reco} < 80$ ) || (Mtt reco  $> 110 \text{ && } \text{Mtt reco} < 130$ )

Region	TotalBkg	Zjets	purity	Data	Data/Bkg
CR	1420.2+-10.9304	1221.51+-5.434	0.860	1559	1.09
VR1	821.998+-9.435	673.277+-4.073	0.819	904	1.10
VR2	465.375+-7.800	320.194+-2.908	0.688	523	1.12

# Top(HH) Binary

== 2 medium tau  
 == 0 lepton  
 METtrig && MET  $\geq$  200  
 OS  
 nBaseJet  $\geq$  1  
 Jet pt  $>$  100GeV  
 Mtt\_reco  $<$  40GeV && Mtt\_reco  $>$  130GeV  
  
 **$\geq 1$  bTag(improve top events)**  
**C1N2 score  $\leq$  0.7(Othogonal with SR)**

Same with pre-selection



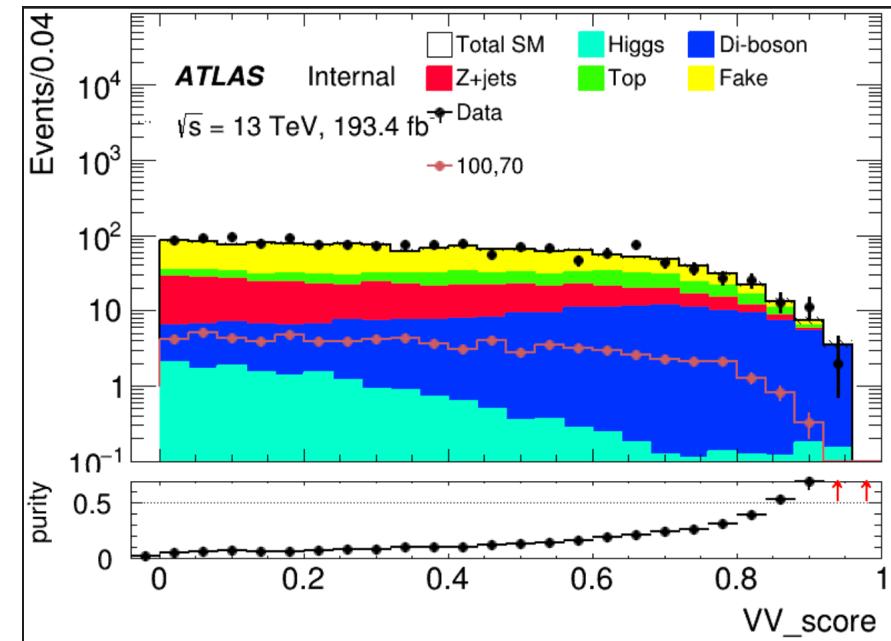
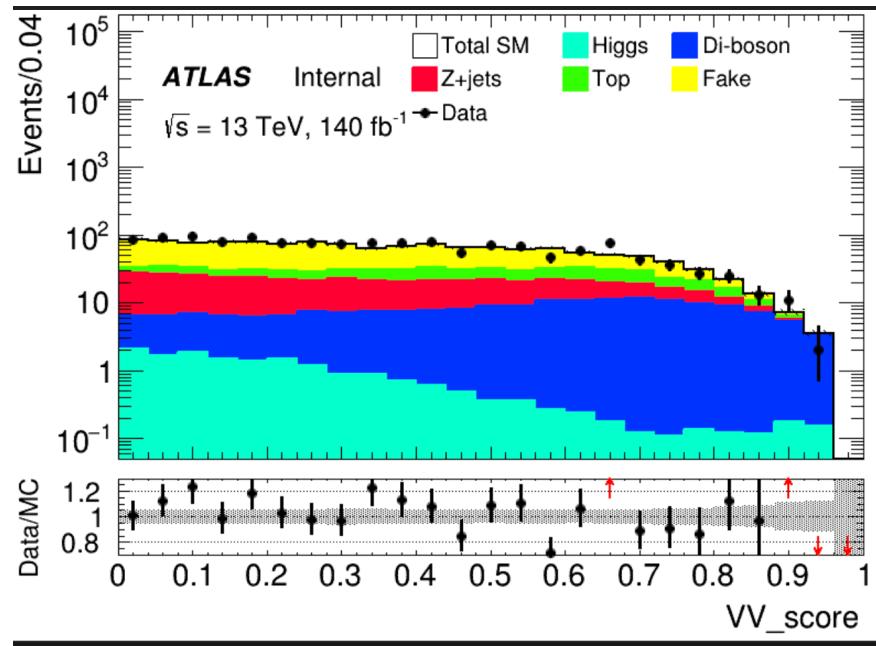
CR:  $270 < M_T(\tau, MET) < 400$   
 VR:  $230 < M_T(\tau, MET) < 270$

Huge overestimation

Region	TotalBkg	Top	purity	Data	Data/Bkg
CR	182.692+-5.738	140.981+-4.009	0.771	150	0.82
VR	140.228+-5.236	103.866+-3.445	0.740	134	0.94

# VV(HH) Binary

Pre-selection && C1N2\_score <= 0.7 && VV\_score >= 0.80



Region	TotalBkg	VV	purity	Data	Data/Bkg
VR	46.6004+-2.77	24.193+-0.816	0.519	51	1.108

# Binary class(LH)

Hyperparameters: Ntrees = 200, MaxDepth = 6, MinNodeSize = 2%, Learning rate = 0.03(initial setting)

## Feature engineering:

Select a simple model and put all features into model, choose Top 30 vars based on importance list, drop high correlated vars

### Final feature list:

: Rank	: Variable	: Variable Importance
:	1 : fb_frac_MET_tau2	: 8.270e-02
:	2 : fb_dRtt	: 6.684e-02
:	3 : fb_dPhitt	: 6.226e-02
:	4 : fb_frac_MET_tt	: 5.197e-02
:	5 : fb_frac_jet_tau2	: 5.179e-02
:	6 : fb_MT2_50	: 5.077e-02
:	7 : fb_dPhiMax_tj	: 4.779e-02
:	8 : fb_dPhiMin_xj	: 4.343e-02
:	9 : fb_mt_taumin	: 3.547e-02
:	10 : fb_Mll	: 3.511e-02
:	11 : fb_mtx_tau1	: 3.408e-02
:	12 : fb_nBaseJet	: 3.146e-02
:	13 : fb_frac_jet_tt	: 3.110e-02
:	14 : fb_mtx_tau2	: 2.941e-02
:	15 : fb_frac_MET_tau1	: 2.898e-02
:	16 : fb_METsig	: 2.824e-02
:	17 : fb_pt_Vframe	: 2.726e-02
:	18 : fb_Mwh	: 2.684e-02
:	19 : fb_Proj_j	: 2.678e-02
:	20 : fb_frac_MET_sqrtHT_40	: 2.560e-02
:	21 : fb_frac_jet_tau1	: 2.518e-02
:	22 : fb_MCT	: 2.254e-02
:	23 : fb_Mwl	: 2.185e-02
:	24 : fb_mt_quad_sum	: 2.165e-02
:	25 : fb_Proj_tt	: 2.038e-02
:	26 : fb_ht_tau	: 1.992e-02
:	27 : fb_e_tau2	: 1.819e-02
:	28 : fb_mt_sum_ttj	: 1.624e-02
:	29 : fb_mt_tau2	: 1.618e-02

Weight choose: no weight, abs(weight)

No weight have better performance  
but abs(weight) fit our analysis requirement

Split strategy: Separate entries by using mod 5, for Fake bkg, if separate follow sequence, all weighted entry will split into first fold

# Binary class(LH)

Hyperparameter tune:  
use optuna to auto-optmize

constraint:

average of AUC need to  $\geq 0.6$

penalty function:  $\text{score} = \text{test\_auc} - 0.3 * \text{auc\_gap}$  ( $\text{auc\_gap} = \text{abs}(\text{train\_auc} - \text{test\_auc})$ )  
 $\text{maximum}(\text{score})$

Class: C1N2, bkg

$\text{Test\_auc} = \sum \{\text{Test\_auc\_class}\}$   
 $\text{Train\_auc} = \sum \{\text{Train\_auc\_class}\}$

## Grid Search

Ntrees: [200, 300, 400]

MaxDepth: [4, 6, 8, 10]

MinNode: [1, 3, 5, 7]

Learning rate: [0.001, 0.005, 0.01, 0.05, 0.1]



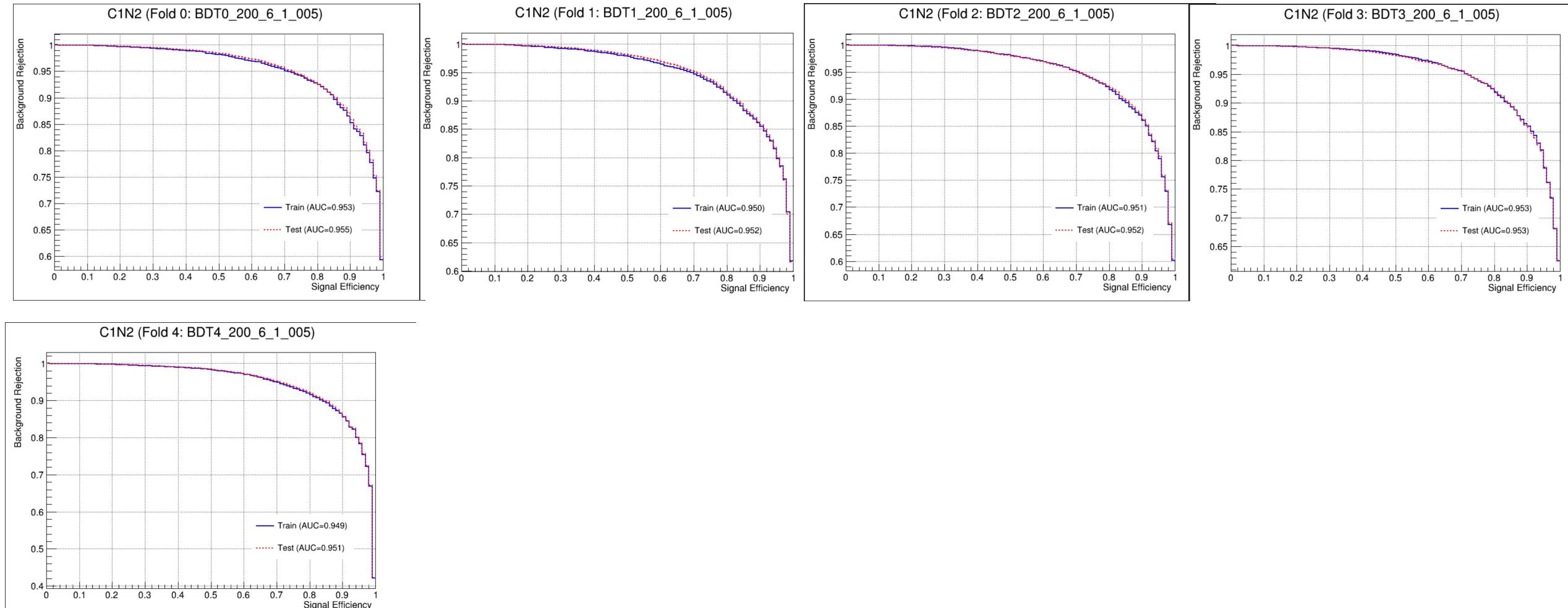
Best one: Ntree=200, MaxDepth=8, MinNode=1%, Learning Rate=0.05



There still have rooms to optimize for lr

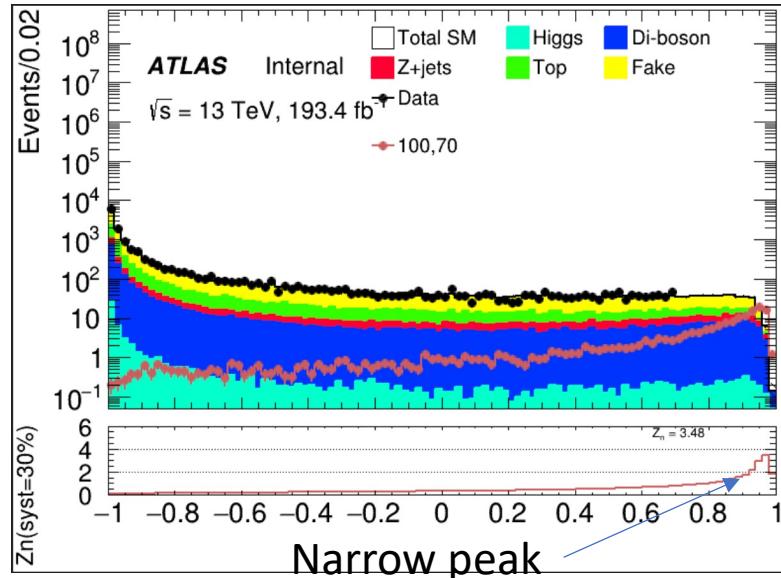
# Binary class(LH)

## Overfit Check

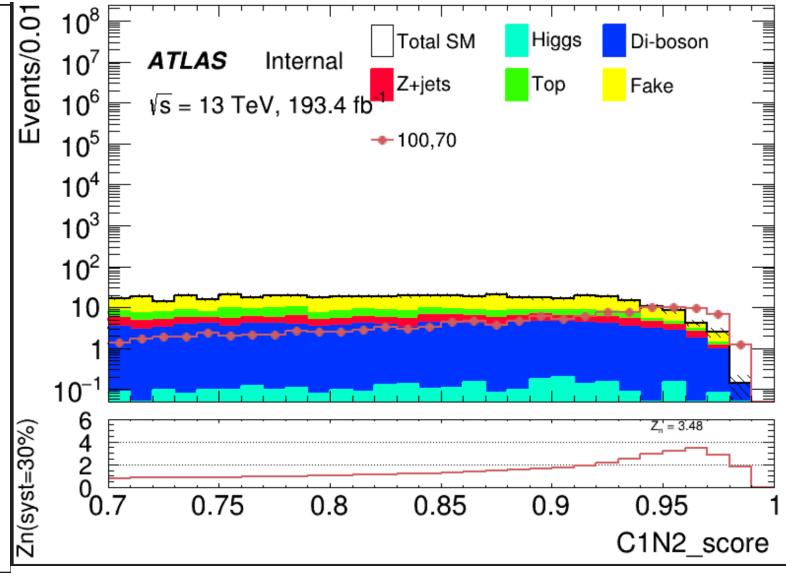


# SR(LH) Binary

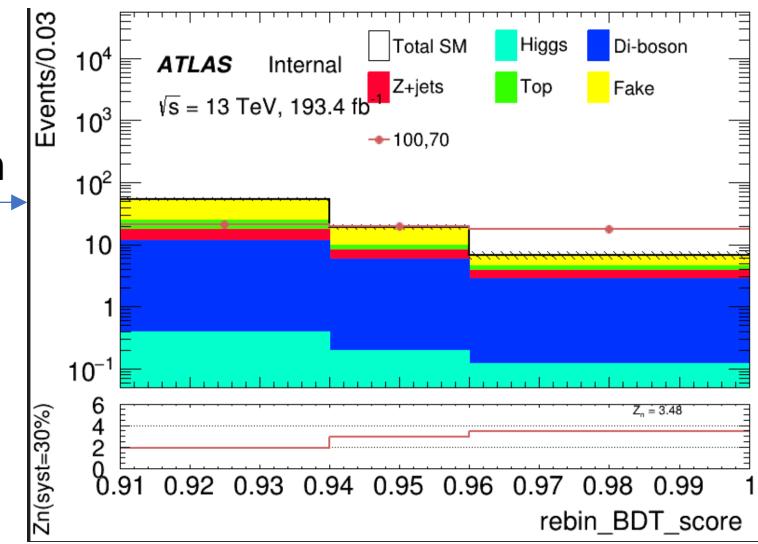
100 bins



Cut at 0.91



rebin



Sum  $Z_n = 4.934$

Bin Range	$Z_n$	C1N2 (100_70) Yield $\pm$ Error	VV Yield $\pm$ Error	Top Yield $\pm$ Error	Fake Yield $\pm$ Error	Higgs Yield $\pm$ Error	Zjets Yield $\pm$ Error	Wjets Yield $\pm$ Error	Total Bkg Yield $\pm$ Error
[0.91,0.94]	1.90	$21.447 \pm 0.901$	$11.105 \pm 0.565$	$7.432 \pm 0.956$	$28.625 \pm 3.295$	$0.376 \pm 0.057$	$5.515 \pm 0.357$	$0.173 \pm 0.142$	$53.226 \pm 3.497$
[0.94,0.96]	2.94	$19.678 \pm 0.874$	$5.550 \pm 0.367$	$1.524 \pm 0.441$	$9.404 \pm 1.929$	$0.189 \pm 0.052$	$2.251 \pm 0.222$	$0.000 \pm 0.000$	$18.918 \pm 2.018$
[0.96,1.00]	3.47	$17.543 \pm 0.820$	$2.666 \pm 0.263$	$0.578 \pm 0.237$	$2.400 \pm 1.009$	$0.118 \pm 0.040$	$1.026 \pm 0.118$	$0.000 \pm 0.000$	$6.788 \pm 1.067$

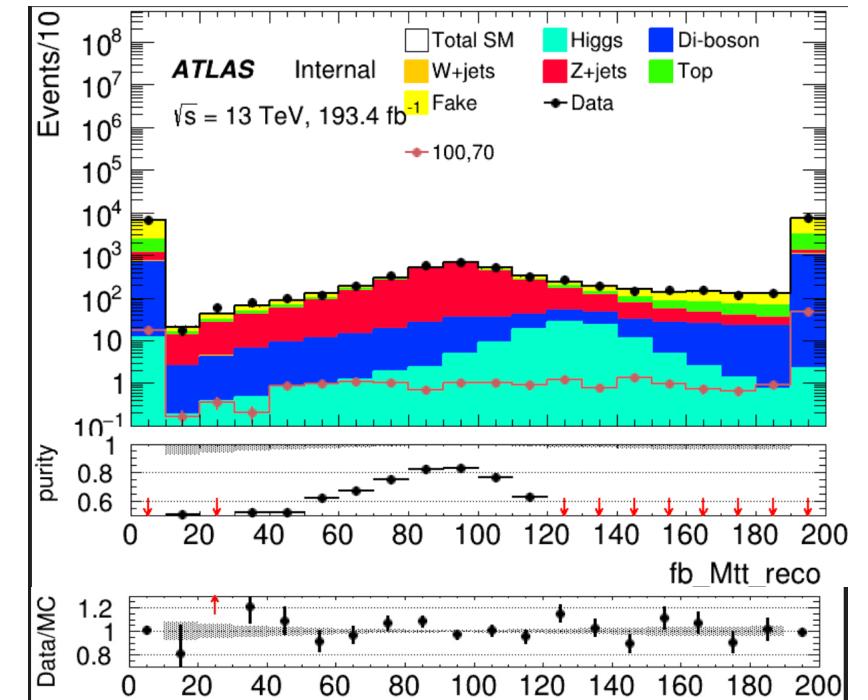
# Zjets(LH) Binary

== 1 medium tau  
 >= 1 lepton  
 METtrig && MET >= 200  
 OS  
 bVeto  
 nBaseJet >= 1  
 Jet pt > 100GeV  
 C1N2 score <= 0.7(Othogonal with SR)

Same with pre-selection

CR: Mlt reco >= 80 && Mlt reco <= 110  
 VR: (Mlt reco >= 40 && Mlt reco < 80) || (Mlt reco > 110 && Mlt reco < 130)

Region	TotalBkg	Zjets	purity	Data	Data/Bkg
CR	2048.51+-12.873	1600.7+-6.313	0.781	2064	1.0078
VR1	1238.27+-10.537	915.681+-4.726	0.739	1296	1.046
VR2	746.998+-9.369	385.547+-3.280	0.516	774	1.037



# Top(LH) Binary

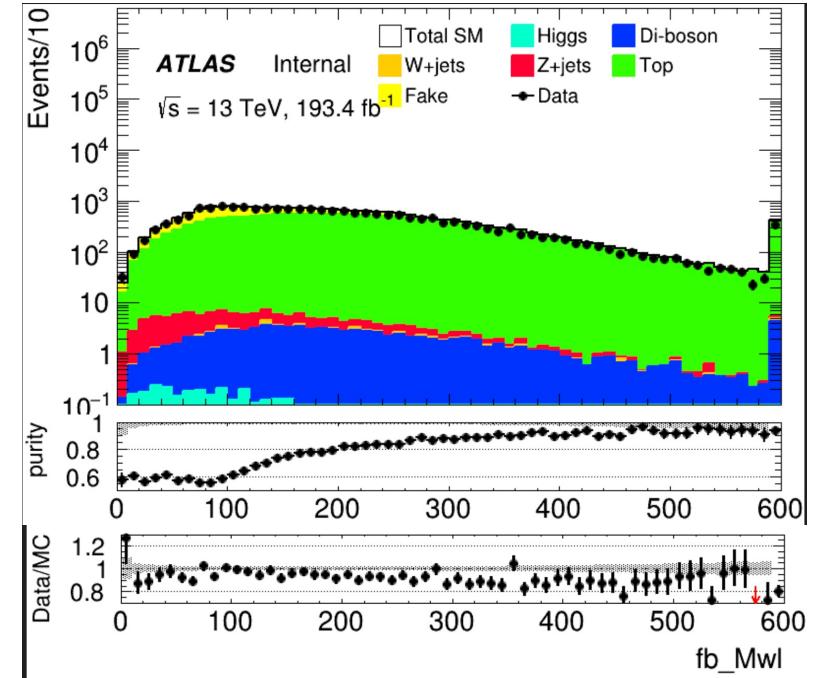
$\geq 1$  medium tau  
 $\geq 0$  lepton  
 METtrig && MET  $\geq 200$   
 OS  
 nBaseJet  $\geq 1$   
 Jet pt  $> 100\text{GeV}$   
 Mtt\_reco  $< 40\text{GeV} \&\& \text{Mtt}_\text{reco} > 130\text{GeV}$

$\geq 1$  bTag(improve top events)  
 C1N2 score  $\leq 0.7$ (Orthogonal with SR)

CR:  $300 < M_{inv}(l, MET) < 550$

VR:  $250 < M_{inv}(l, MET) < 300$

Same with pre-selection

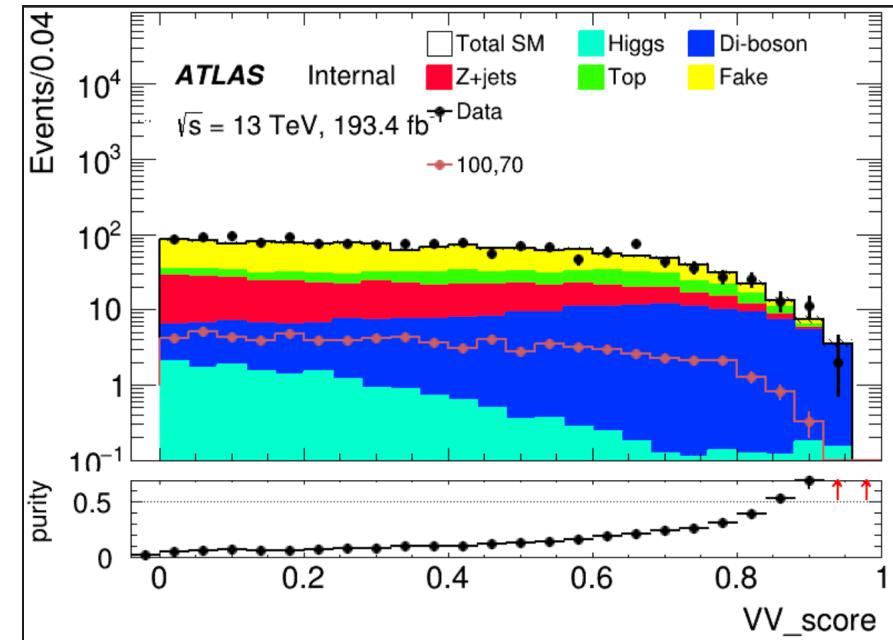
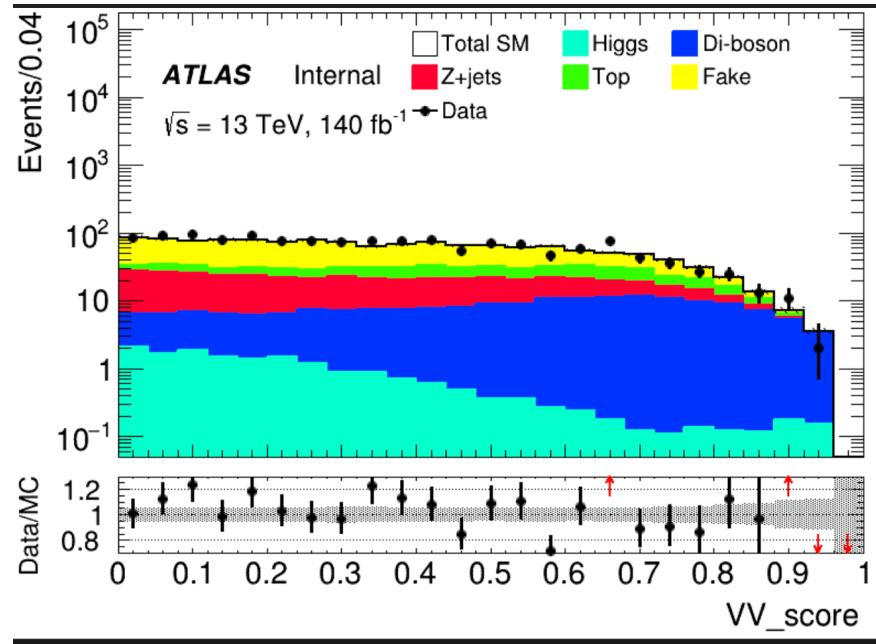


Huge overestimation

Region	TotalBkg	Top	purity	Data	Data/Bkg
CR	4657.04+27.0028	4224.32+21.804	0.907	4152	0.891
VR	2916.32+22.0316	2532.7+-16.8751	0.868	2700	0.925

# VV(LH) Binary

Pre-selection && C1N2\_score <= 0.7 && VV\_score >= 0.80



Region	TotalBkg	VV	purity	Data	Data/Bkg
VR	78.6819+2.26	53.8201+1.124	0.68402	72	0.92

# Production issues

Task ID	Task name	N files total	N files done	N files failed	%	Status (JEDI)	Duration, days	Task logged status	Jobs failure, %	Top job errors, count [component:code] "sample message" [log example]
44282337	user.liaoc.542937.MGPy8EG_A14N23LO_C1N2ISR_100p0_0p0_2TFilt_run3.mc20d.r14860_p6284_stauhh_no_done_2/	11	0	0	0	exhausted	4.79	mc20_13TeV:mc20...	0 -	
44282315	user.liaoc.542971.MGPy8EG_A14N23LO_C1N2ISR_180p0_80p0_2TFilt_run3.mc20a.r13167_p6284_stauhh_no_done_2/	2	0	0	0	exhausted	4.79	mc20_13TeV:mc20...	0 -	
44281469	user.liaoc.542937.MGPy8EG_A14N23LO_C1N2ISR_100p0_0p0_2TFilt_run3.mc20d.r14860_p6284_stauhh_no_done_1/	11	0	0	0	exhausted	4.89	mc20_13TeV:mc20...	0 -	
44281451	user.liaoc.542971.MGPy8EG_A14N23LO_C1N2ISR_180p0_80p0_2TFilt_run3.mc20a.r13167_p6284_stauhh_no_done_1/	2	0	0	0	exhausted	4.89	mc20_13TeV:mc20...	0 -	

100\_0\_mc20d, 180\_80\_mc20a

Submit them twice but still failed

# Bkg decay mode

Wjets:  $W \rightarrow e/\mu\text{on} + \nu$

$W \rightarrow \tau + \nu$ (can contribute true  $\tau_{\text{had}}$ )  
jet misidentified to a fake tau

Zjets:  $Z \rightarrow ll/\tau\tau\tau\tau$

jet misidentified to fake tau

Top:  $\text{top} \rightarrow W+b$ , W can contribute a true  $\tau_{\text{had}}$   
b-quark is a source of fake

VV:  $W/Z$

LH channel:  $\geq 1\tau, \geq 1\text{lep}$

Wjets: W contribute lep, jets misidentified to fake

Zjets:

SingleTop: W contribute lep, b-quark misidentified to fake

VV:

HH channel:  $\geq 2\tau, == 0\text{lep}$

Wjets: W contribute  $\tau_{\text{had}}$ , plus a fake tau

Zjets:  $Z \rightarrow \tau\tau\tau\tau(\text{had})$  or 2 fake tau

SingleTop: W contribute a  $\tau_{\text{had}}$ , plus a fake tau

VV:

# Multiclass(HH)

Hyperparameters: Ntrees = 200, MaxDepth = 6, MinNodeSize = 2%, Learning rate = 0.03(initial setting)

## Feature engineering:

Select a simple model and put all features into model, choose Top 30 vars based on importance list, drop high correlated vars

## Final feature list:

: Rank	: Variable	: Variable Importance
:		
1	: fb_dRtt	: 8.238e-02
2	: fb_dRMax_xt	: 7.068e-02
3	: fb_METsig	: 6.205e-02
4	: fb_frac_MET_tt	: 6.050e-02
5	: fb_dPhi1x	: 5.751e-02
6	: fb_MIA	: 5.460e-02
7	: fb_mt_taumin	: 5.411e-02
8	: fb_Asy_tt	: 5.363e-02
9	: fb_dEtat2j	: 4.903e-02
10	: fb_MET_Soft	: 4.737e-02
11	: fb_Asy_EH	: 4.625e-02
12	: fb_Mll	: 4.447e-02
13	: fb_frac_MET_MeffInc_40	: 4.317e-02
14	: fb_eta_jet2	: 4.282e-02
15	: fb_eta_jet1	: 4.229e-02
16	: fb_transSphericity	: 4.140e-02
17	: fb_dRMax_jets	: 4.086e-02
18	: fb_frac_MET_Meff	: 4.007e-02
19	: fb_m_jet1	: 3.421e-02
20	: fb_nJets30	: 3.260e-02
:		

Weight choose: no weight, abs(weight)

No weight have better performance  
but abs(weight) fit our analysis requirement

Split strategy: Separate entries by using mod 5, for Fake bkg, if separate follow sequence, all weighted entry will split into first fold

# Multiclass(HH)

Hyperparameter tune:  
use optuna to auto-optimize

constraint:

average of AUC need to  $\geq 0.6$

penalty function:  $\text{score} = \text{test\_auc} - 0.3 * \text{auc\_gap}$  ( $\text{auc\_gap} = \text{abs}(\text{train\_auc} - \text{test\_auc})$ )  
 $\text{maximum}(\text{score})$

After check some models, find C1N2 result is great, so the constraint and AUC calculation only in VV and Other bkg

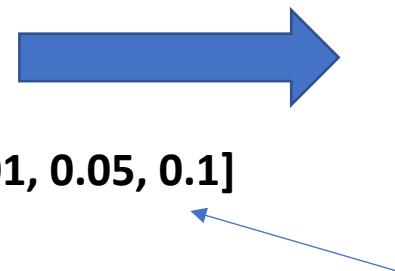
## Grid Search

Ntrees: [200, 300, 400]

MaxDepth: [4, 6, 8, 10]

MinNode: [1, 3, 5, 7]

Learning rate: [0.001, 0.005, 0.01, 0.05, 0.1]

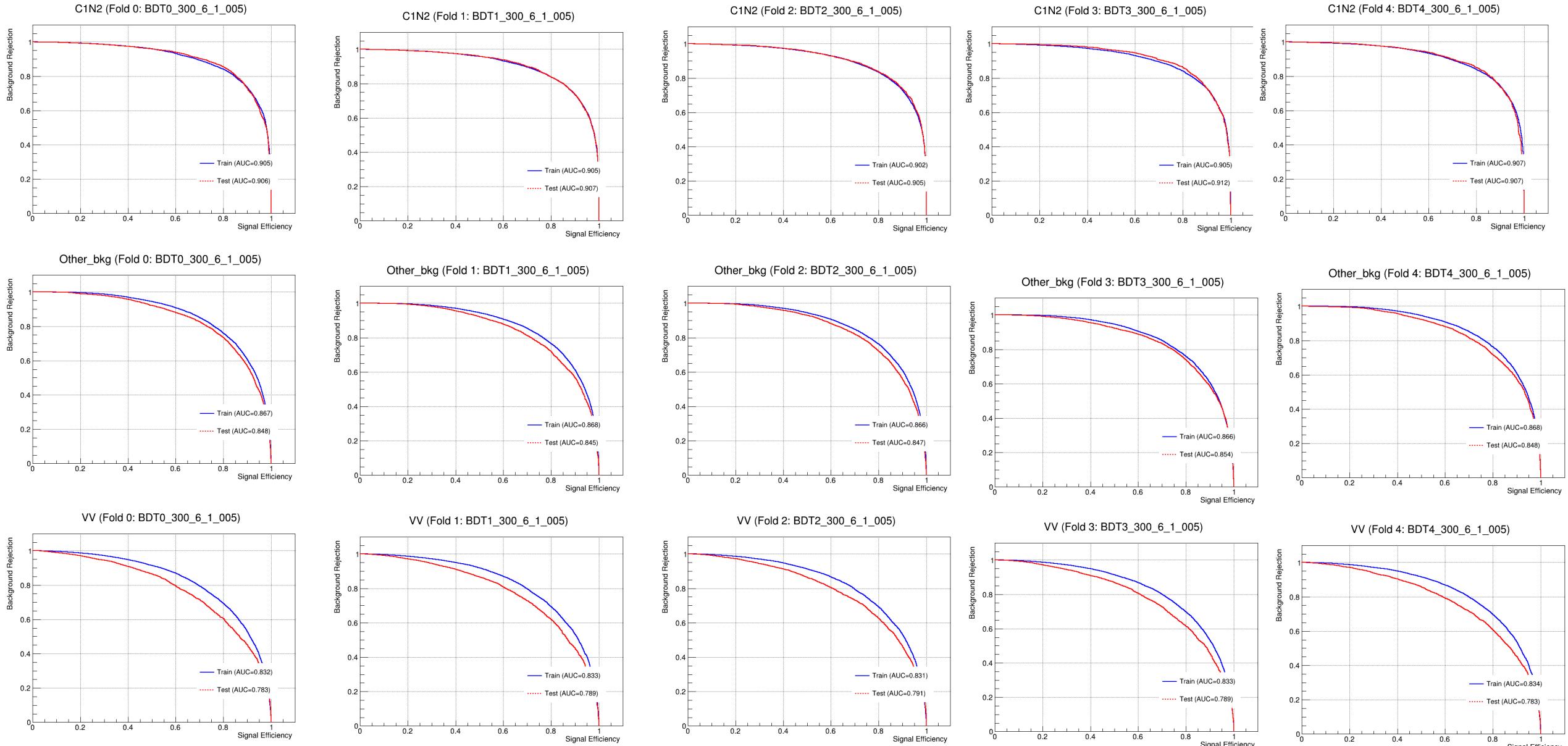


Best one: Ntree=300, MaxDepth=6, MinNode=1%, Learning Rate=0.05

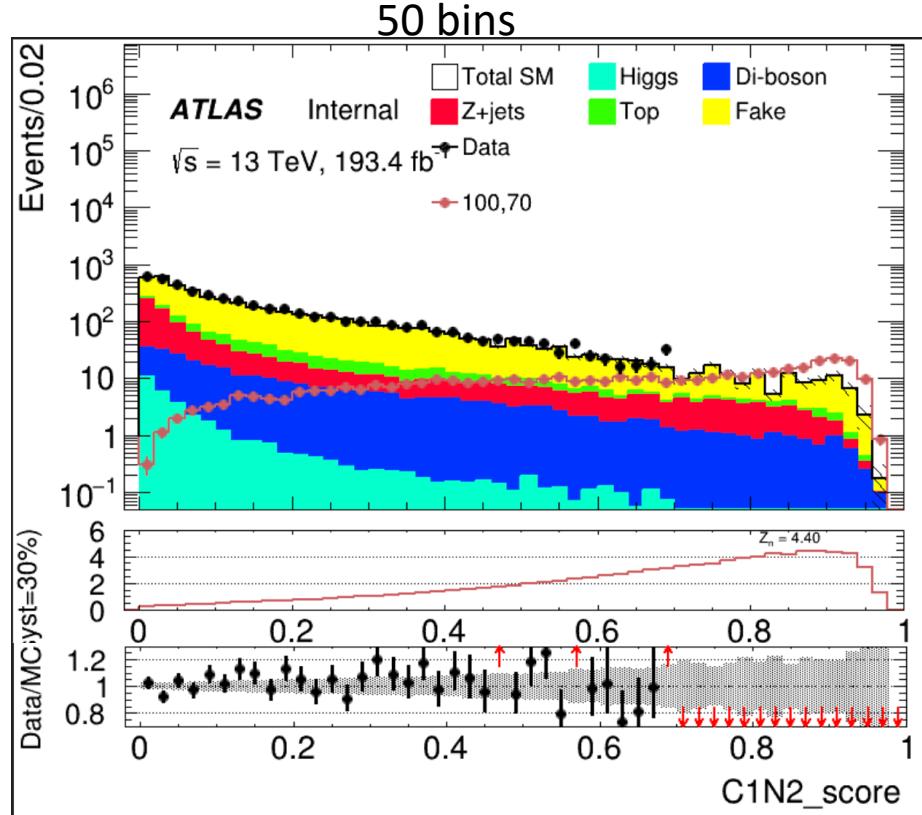
There still have rooms to optimize for lr

# Multiclass(HH)

# OverFit Check



# SR(HH) Multi

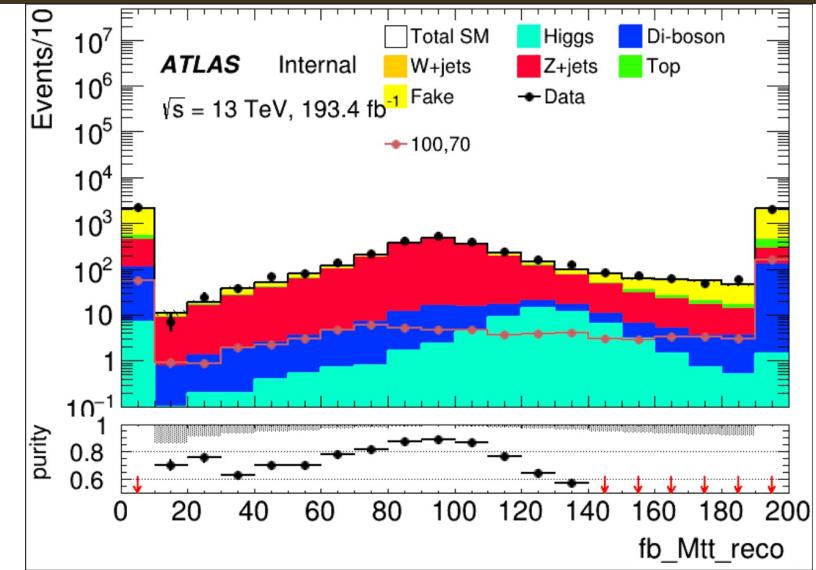


Each index stand for a region, and index 10 stands for sum

# Zjets(HH)

== 2 medium tau  
 == 0 lepton  
 METtrig && MET  $\geq 200$   
 OS  
 bVeto  
 nBaseJet  $\geq 1$   
 Jet pt  $> 100\text{GeV}$   
 C1N2 score  $\leq 0.7$ (Othogonal with SR)

Same with pre-selection



CR: Mtt reco  $\geq 80 \text{ && } \text{Mtt reco} \leq 110$

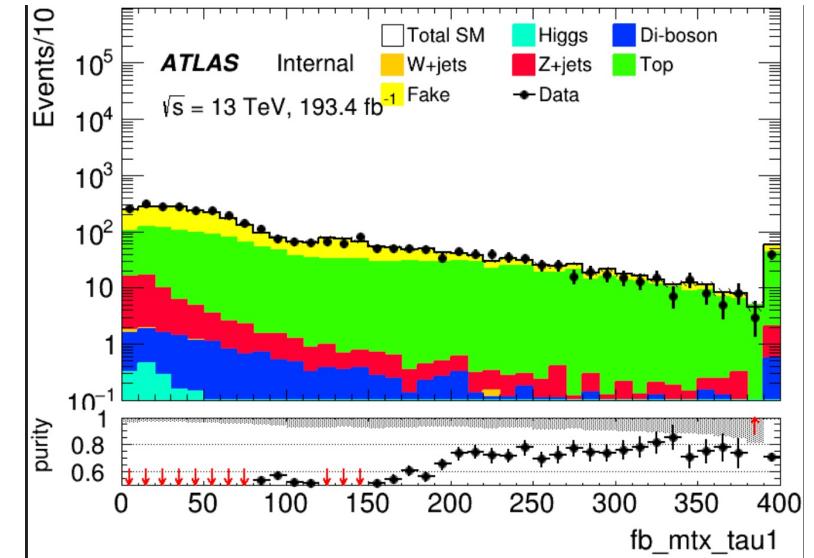
VR: (Mtt reco  $\geq 40 \text{ && } \text{Mtt reco} < 80$ ) || (Mtt reco  $> 110 \text{ && } \text{Mtt reco} < 130$ )

Region	TotalBkg	Zjets	purity	Data	Data/Bkg
CR	1433.93+-11.003	1232.36+-5.476	0.859	1571	1.096
VR1	824.938+-8.468	674.307+-4.082	0.817	904	1.097
VR2	469.252+-7.834	321.877+-2.913	0.685	524	1.117

# Top(HH)

== 2 medium tau  
 == 0 lepton  
 METtrig && MET  $\geq 200$   
 OS  
 nBaseJet  $\geq 1$   
 Jet pt  $> 100\text{GeV}$   
 Mtt\_reco  $< 40\text{GeV} \&\& \text{Mtt}_\text{reco} > 130\text{GeV}$   
  
 $\geq 1$  bTag(improve top events)  
 C1N2 score  $\leq 0.7$ (Orthogonal with SR)

Same with pre-selection



CR:  $270 < M_T(\tau, \text{MET}) < 400$

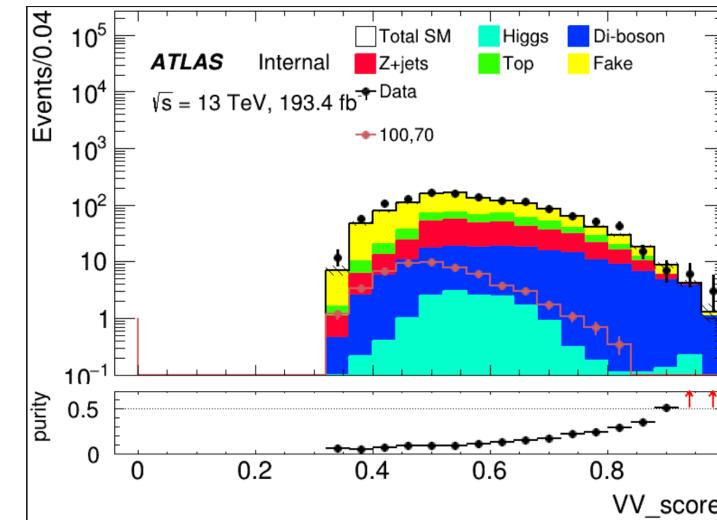
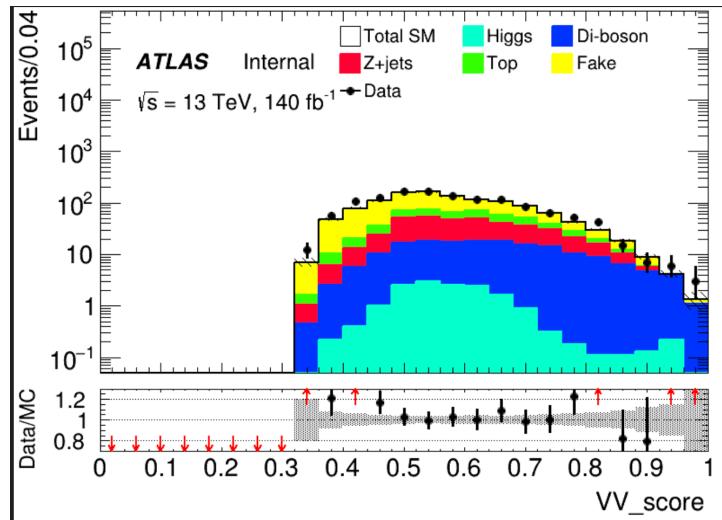
VR:  $230 < M_T(\tau, \text{MET}) < 270$

Huge overestimation

Region	TotalBkg	Top	purity	Data	Data/Bkg
CR	182.692+5.738	140.981+4.009	0.771	150	0.82
VR	140.228+5.236	103.866+3.445	0.740	134	0.94

# VV(HH)

Pre-selection && C1N2\_score <= 0.7 && Max(score) == VV\_score && VV\_score >= 0.85



Region	TotalBkg	VV	purity	Data	Data/Bkg
VR	32.4944+-2.09	15.4902	0.476	31	0.954