

Department of Physics, Shandong University

Compressed EWK study(ISRC1N2)

Chengxin Liao
liaocx@ihep.ac.cn

Feb, Mon 24, 2024



Hyperparameters optimization(LH)

Input(LH-Channel):

Sample:

Sig: ISRC1N2(mass_C1 = 100GeV, mass_N2 = 70GeV)->21225 entries

Bkg: 1703476 entries

All input data(C1N2_100_70 and Bkg) already passed pre-selection

```
Signal -- training events      : 12735
Signal -- testing events       : 4245
Signal -- training and testing events: 16980
Background -- training events   : 1022092
Background -- testing events    : 340692
Background -- training and testing events: 1362784
```

Strategy:

method: BDTG

Separate sig(bkg) into five folders, one for test, the other three for train, and last one for validation set, then traverse all possibilities.

Pre-Selection

lep-had channel: $nTaus \geq 1, nLeps \geq 1$

pass MET trigger; $MET \geq 200$

$1 \leq nBaseJet \leq 8$

b - Veto

OS

Hyperparameters optimization(LH)

Variables(30):

Obj kinematics

nBase_Jet
mt_lep
e_lep(energy of tau2)

Angular correlations

dPhitt
dRtt
dRt1x
dPhiMin_xj
dPhiMax_tj

Event kinematics

Mll(Invariant Mass of tau1 and tau2)
METsig
MT2_50
Mwh(Invariant Mass of tau1 and MET)
Mwl(Invariant Mass of tau2 and MET)
MCT(Transverse Mass Squared)
Proj_j(Projection of pt jet on zeta)
Proj_tt(Projection of tau1+tau2 on zeta)
mtx_tau
Mtx_lep

ht_tau
mt_quad_sum
mt_sum
frac_MET_tau1
frac_MET_tau2
frac_MET_tt
frac_MET_sqrtHT_40
frac_jet_tau1
frac_jet_tau2
frac_jet_tt
[MT_tauamin
pt_Vframe

High importance at shiyi's feature

Note:

zeta is bisector direction of tau1 and tau2[PhyUtils::bisector(tau1, tau2)]

Hyperparameters optimization(LH)

Grid Search:

Ntrees: 200, 300, 400, 500

Max Depth: 6, 8, 10, 12

MinNodeSize: 1%, 2%, 3%

Learning Rate: 0.01, 0.05, 0.1

Binned significance: $Z = \sqrt{2((s_i + b_i) \log\left(1 + \frac{s_i}{b_i}\right) - s_i)}$

Show top Zn

	Model Name	Binned Significance	Max Zn	Max Zn Bin
12	400_8_1_001	15.6795	4.31391	192
79	400_10_1_001	15.6755	4.26908	192
92	400_12_1_001	15.6890	4.21178	192
77	400_10_2_001	15.3196	4.11376	191
52	500_10_1_001	15.8304	4.11162	194
13	500_12_1_001	15.8210	4.05346	194
120	400_12_1_01	16.0665	4.02939	199
113	500_10_3_001	15.3232	4.02306	192
0	300_12_1_005	16.1734	4.01739	198
139	400_12_1_005	16.2126	4.00753	199
123	500_12_3_001	15.3067	4.00343	192
118	500_12_1_01	16.0441	4.00080	199
24	500_8_1_01	15.9307	3.99007	199
133	500_8_3_001	15.3061	3.97695	192
97	400_6_3_001	15.0010	3.97216	190
26	300_10_1_01	16.0095	3.96339	199
136	300_12_1_01	16.0204	3.94916	199
128	400_8_3_001	14.9962	3.93255	190
107	200_12_1_005	16.0375	3.93002	197
45	400_12_2_001	15.2724	3.92019	191
88	400_12_3_001	14.9991	3.91396	190

Shiyi's result of LH channel

Top Sig

	hy	sig	zn
400_10_2_0.05	15.3225	3.72536	
300_11_1_0.05	15.3127	3.87694	
500_10_2_0.05	15.3099	3.60778	
400_6_1_0.05	15.3075	3.91373	
500_8_1_0.05	15.2990	3.58389	
400_8_2_0.05	15.2980	3.74427	
300_6_1_0.05	15.2929	4.09837	
500_8_2_0.05	15.2891	3.63322	
200_11_1_0.05	15.2849	3.92924	
300_11_2_0.05	15.2804	3.85617	
400_11_2_0.05	15.2780	3.68484	
300_8_1_0.05	15.2753	3.82506	
300_10_1_0.05	15.2733	3.71921	
400_11_1_0.05	15.2701	3.60863	
500_6_1_0.05	15.2593	3.84429	
200_6_1_0.1	15.2559	3.90950	
400_12_1_0.05	15.2554	3.58328	
500_10_1_0.05	15.2493	3.49410	

Top Zn

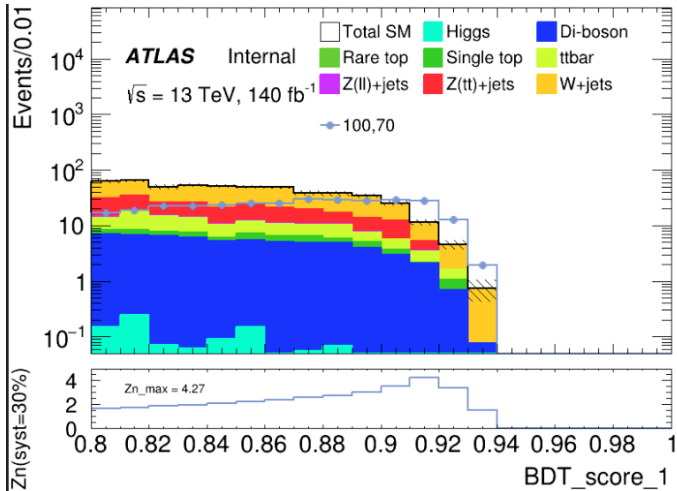
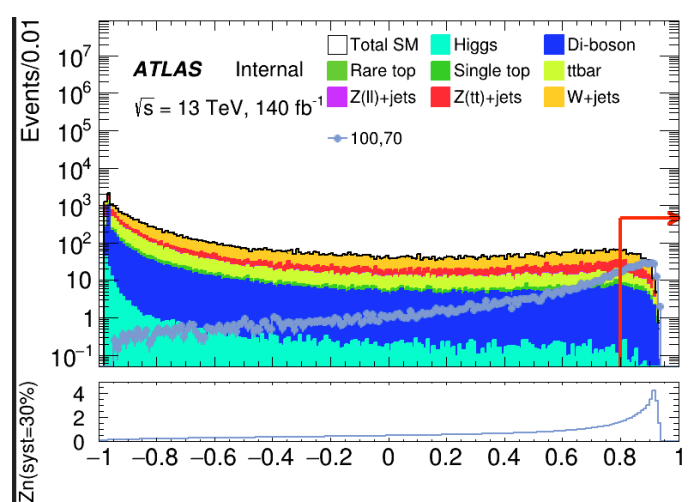
	hy	sig	zn
200_6_3_0.05	15.0164	4.29022	
200_6_1_0.05	15.0755	4.10077	
300_6_1_0.05	15.2929	4.09837	
200_10_2_0.05	15.1606	4.09228	
200_12_2_0.05	15.1803	4.04800	
200_8_2_0.05	15.0857	4.01373	
200_8_3_0.05	14.9662	4.01324	
200_6_2_0.05	14.9743	3.94396	
200_11_1_0.05	15.2849	3.92924	
300_6_2_0.05	15.1858	3.91508	
400_6_1_0.05	15.3075	3.91373	
200_6_1_0.1	15.2559	3.90950	
200_12_1_0.05	15.2279	3.90380	
400_8_1_0.01	14.6829	3.90189	
300_11_1_0.05	15.3127	3.87694	
200_8_1_0.05	15.1285	3.85623	
300_11_2_0.05	15.2804	3.85617	
500_6_1_0.05	15.2593	3.84429	

400_10_1_001, 15.6755, 4.26908, 192, 200
 400_10_1_001, 15.4762, 3.53693, 96, 100
 400_10_1_001, 15.1985, 3.40439, 49, 50
 400_10_1_001, 15.3013, 3.53693, 39, 40
 400_10_1_001, 14.8172, 3.40439, 25, 25
 400_10_1_001, 15.06, 3.53693, 20, 20
 400_10_1_001, 13.9532, 1.6563, 10, 10

400_12_1_001, 15.689, 4.21178, 192, 200
 400_12_1_001, 15.4949, 3.52564, 97, 100
 400_12_1_001, 15.2434, 3.52564, 49, 50
 400_12_1_001, 15.3089, 3.52196, 39, 40
 400_12_1_001, 14.8653, 3.52564, 25, 25
 400_12_1_001, 15.0506, 3.52196, 20, 20
 400_12_1_001, 13.9276, 1.6473, 10, 10

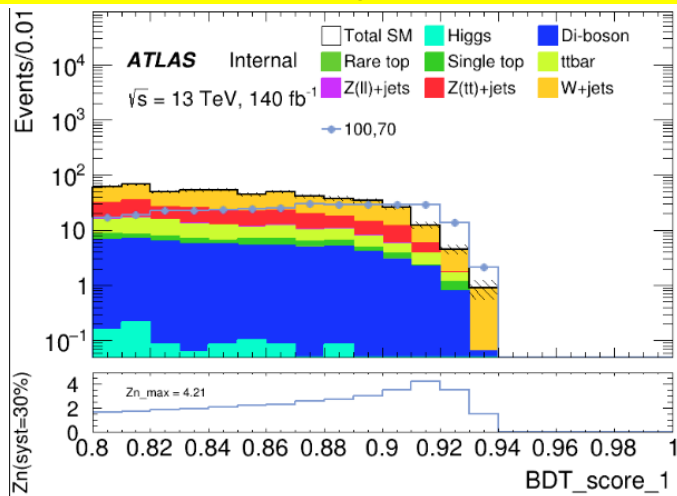
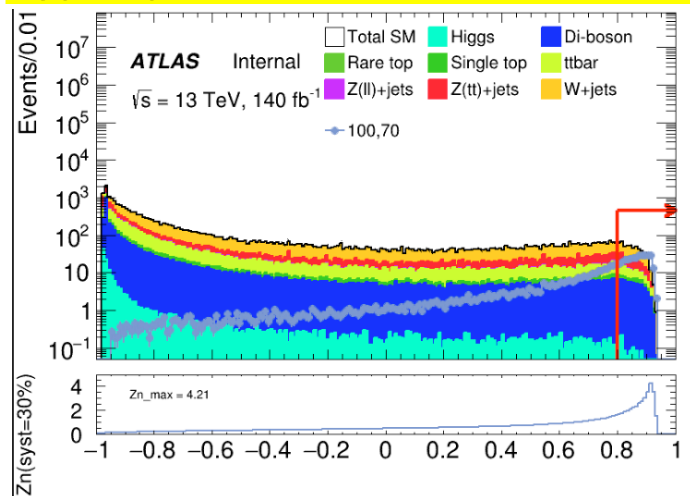
Performance of Model(LH)

hyper parameter: NTrees=400, learning rate=0.01, max depth=10, MinNodeSize=1%(default)



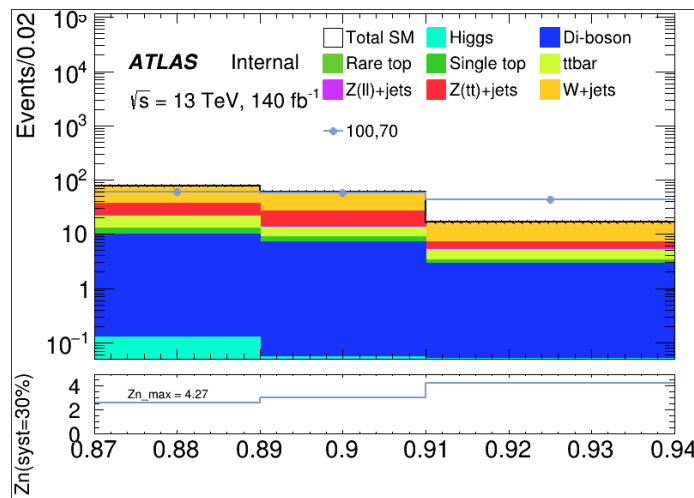
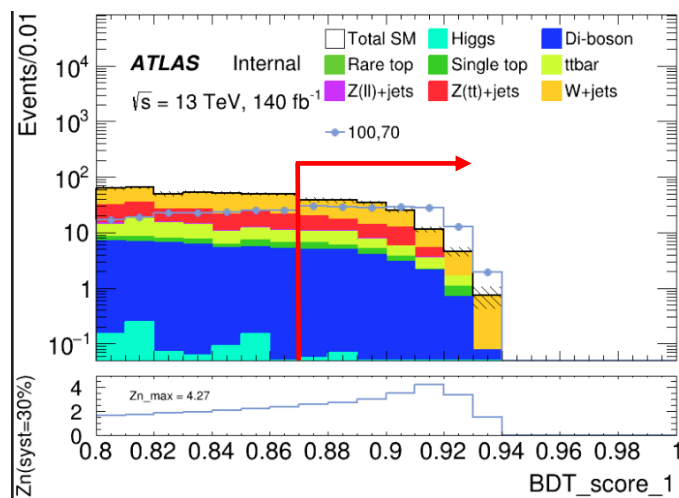
Cut at BDT_score = 0.8

hyper parameter: NTrees=400, learning rate=0.01, max depth=12, MinNodeSize=1%(default)



Performance of Model(LH)

hyper parameter: NTrees=400, learning rate=0.01, max depth=10, MinNodeSize=1%(default)



Apply BDT score cut at 0.87

Root of square sum of Zn of each bin: 5.8479

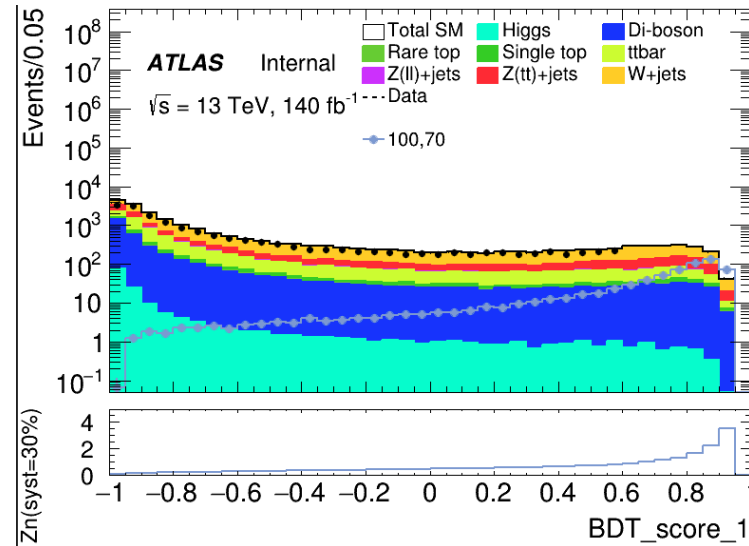
bin	max Zn	C1N2ISR (100,70)	bkg	Higgs	OtherTop	SingleTop	TopPair	VV	Wjets	Zlljets	Zttjets
(0.87-0.89)	2.59868	59.238+- 1.484	76.648+- 5.530(7.21%)	0.126+- 0.034	0.033+- 0.021	2.619+- 0.548	8.311+- 1.141	9.569+- 0.520	39.971+- 5.074	0.760+- 0.130	15.259 +-1.718
(0.89-0.91)	3.03656	57.663+- 1.447	59.803+- 3.946(6.59%)	0.053+- 0.020	0.078+- 0.030	1.761+- 0.420	4.401+- 0.823	6.851+- 0.399	33.586+- 3.367	0.453+- 0.128	12.620 +-1.792
(0.91-0.94)	4.26908	42.715+- 1.251	16.632+- 1.683(10.11%)	0.005+- 0.004	0.006+- 0.004	0.450+- 0.202	1.819+- 0.532	2.858+- 0.249	9.733+- 1.536	0.039+- 0.020	1.722+- 0.298

Performance of Model(LH)

hyper parameter: NTrees=400, learning rate=0.01, max depth=10, MinNodeSize=1%(default)

Here suppose to have a fig to show the distribution
but I can't login computing platform with Vscod
and even can't use vim at terminal

BDT score distribution of Validation set



BDT score distribution of test set and data
(Blind with events with score > 0.6)