Deeply Learned Preselection of Dijet Higgs Decays at Future Lepton Colliders

Wenxing Zhang (TDLI, SJTU)
So Chigusa(Berkley), Shu Li(TDLI, SJTU), Yuichiro Nakai (TDLI, SJTU), Yufei Zhang (TDLI, SJTU), Jiaming Zheng (TDLI, SJTU),
in preparation

TDLI, Shanghai Jiao Tong University

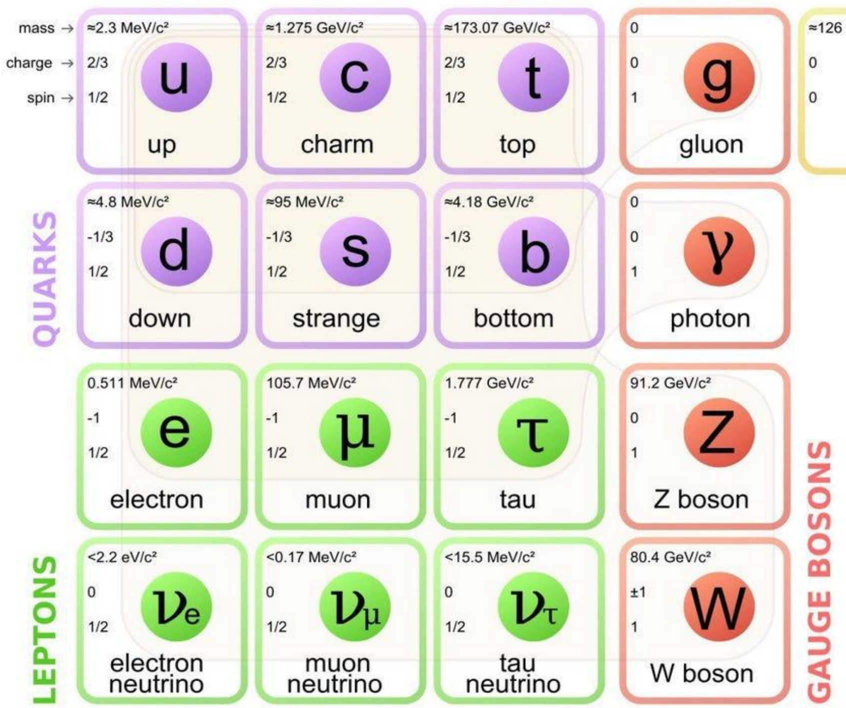
Outline

- Precise measurement of light Yukawa coupling in standard model.

- CEPC in Higgs coupling measurement.

- Preselection at CEPC.
  - Cut-based selection
  - Boosted Decision Tree algorithm
  - Fully-connected neural network algorithm

Yukawa couplings in SM



Why should we test yukawa coupling?

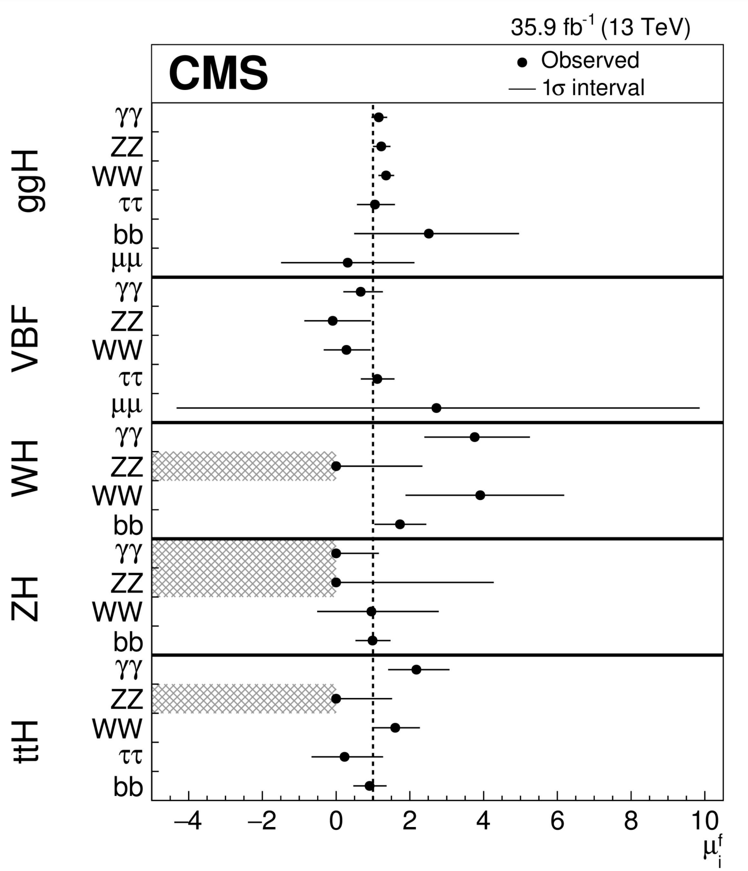
- $m_f \propto y_f v$.
- Potential for new scalar/physics.
- New physics contributions to light fermions can easily dominate over the SM predictions.

- Higgs rare decay: BR($h \to V\gamma$)$\sim 10^{-6}$.
- Flavor violation (eg.$h \to \tau\mu$).
- Jet flavor tagging with ML method

For process $i \to H \to f$, the signal strength is

$$\mu_i^f = \frac{\sigma_i \times BR^f}{(\sigma_i \times BR^f)_{SM}} = \mu_i \times \mu_f$$

Signal Strength

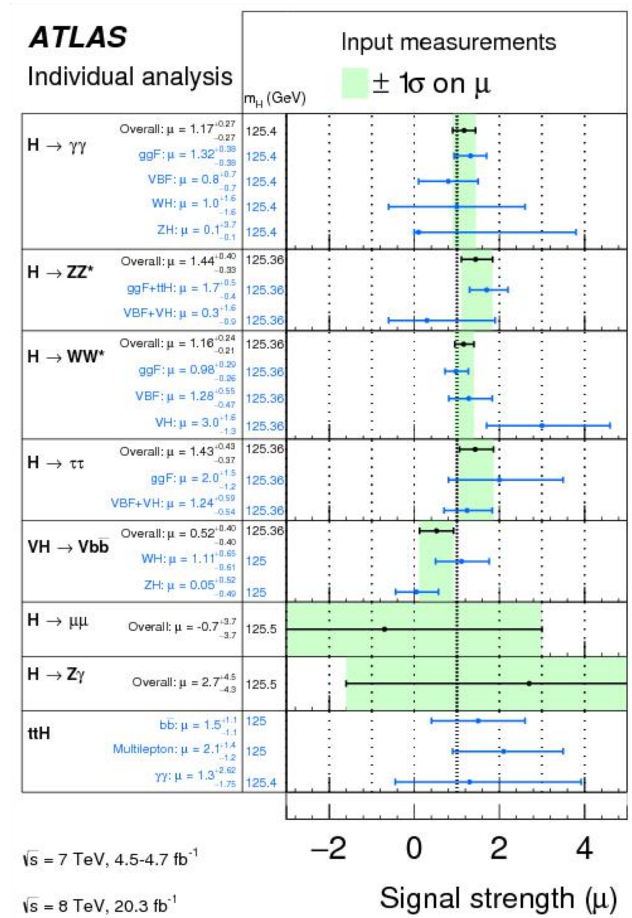


arXiv: 1809.10773

The light fermion signal strength are[1]

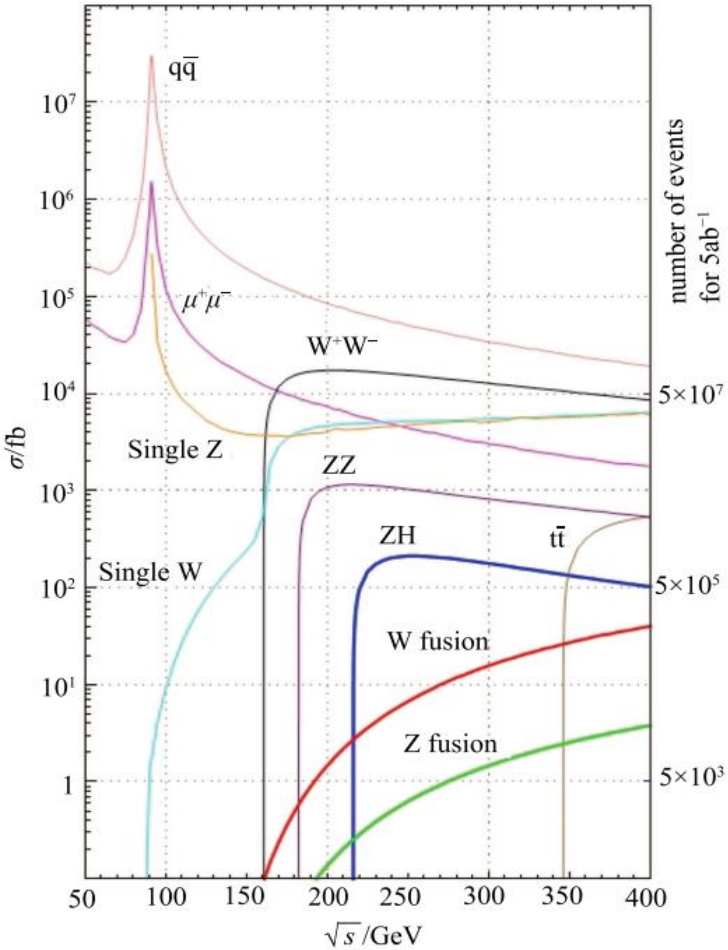
$$\mu_{cc} \lesssim 110, \quad \mu_{ss} \lesssim 7.2 \times 10^8$$

[1] arXiv:1811.09636

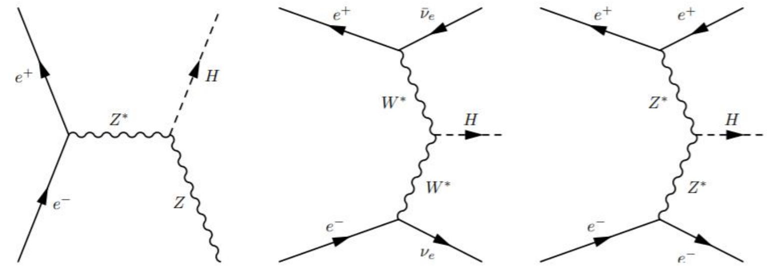


arXiv:1507.04548

CEPC



arXiv:1810.09037



Process	Cross section	Events in 5.6 ab^{-1}
Higgs boson production, cross section in fb		
$e^+e^- \to ZH$	204.7	1.15×10^6
$e^+e^- \to \nu_e\bar{\nu}_e H$	6.85	3.84×10^4
$e^+e^- \to e^+e^- H$	0.63	3.53×10^3
Total	212.1	1.19×10^6
Background processes, cross section in pb		
$e^+e^- \to e^+e^-(\gamma)$ (Bhabha)	850	4.5×10^9
$e^+e^- \to q\bar{q}(\gamma)$	50.2	2.8×10^8
$e^+e^- \to \mu^+\mu^-(\gamma)$ [or $\tau^+\tau^-(\gamma)$]	4.40	2.5×10^7
$e^+e^- \to WW$	15.4	8.6×10^7
$e^+e^- \to ZZ$	1.03	5.8×10^6
$e^+e^- \to e^+e^- Z$	4.73	2.7×10^7
$e^+e^- \to e^+\nu W^- / e^-\bar{\nu} W^+$	5.14	2.9×10^7

Preselection

- Missing mass M_{miss}

$$m_h \leq M_{miss} = \sqrt{E^2 - p_T^2},$$

- p_T, p_L and N_{chd}
- Y_{ij}

  Y_{ij} is the threshold of y-value for clustering jets from $i \to j$.

$$y_{ij} = \frac{2min(E_i^2, E_j^2)(1 - cos\theta_{ij})}{E_{vis}^2}$$

- Di-jet mass M_{jj}

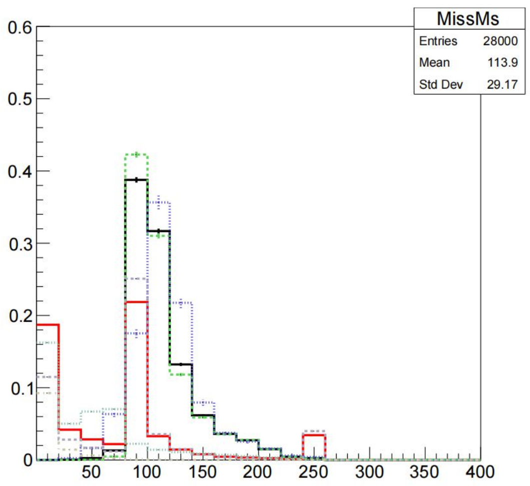
$$M_{jj} = \sqrt{(E_1 + E_2)^2 - (p_1 + p_2)^2},$$

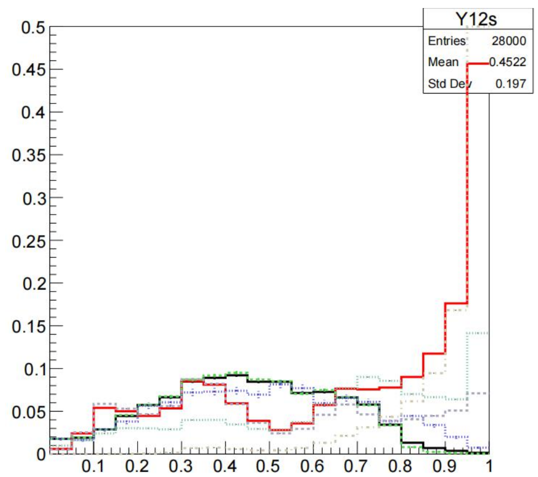
arXiv:1207.0300

CUT-Based Preselection

	$ee \to WW$	$ee \to q\bar{q}$	$ee \to ZZ$	$ee \to e\nu W$	$ee \to eeZ$	$ee \to ZH$	$ee \to WWH$
80GeV$\leq M_{miss} \leq$ 140GeV	452748	33876.3	83565.2	242224	241020	8854	1318
20GeV$\leq p_T \leq$ 70GeV	322171	4376.25	49099.1	169402	144559	8161	1069
$\lvert p_L \rvert \leq$ 60GeV	160461	4043.27	16086.3	83310.8	38178.3	7967	967
$N_{chd} \geq 10$	58722	4008.92	14072.8	4477.92	0	7772	943
$P_{max} \leq$ 30GeV	30168	2618.88	10951.1	446.69	0	6963	852
$Y_{23} \leq 0.02$	4807	2193.41	7545.75	108.919	0	4623	552
$0.2 \leq Y_{12} \leq 0.8$	3745	2008.42	4995.1	90.5619	0	4535	498
100GeV$\leq M_{jj} \leq$ 130GeV	1856	277.479	856.486	50.1762	0	4331	474

Table 1: The result is obtained with $\sqrt{s} = 250GeV$ and $\mathcal{L} = 250fb^{-1}$.

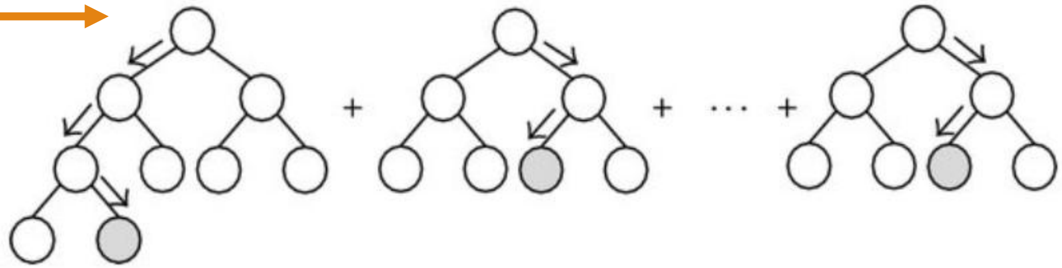




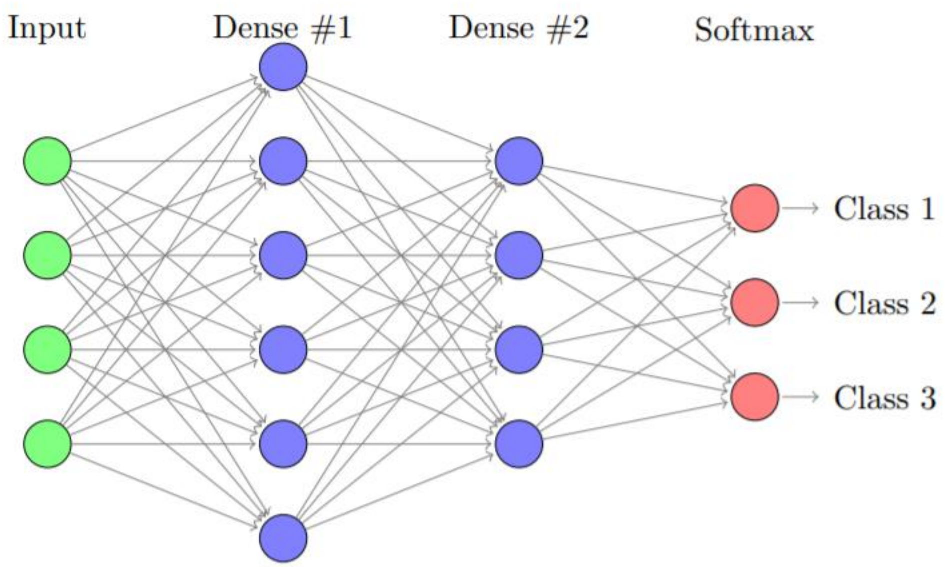
All the quantities are put into the BDT networks and fully-connected neural networks.

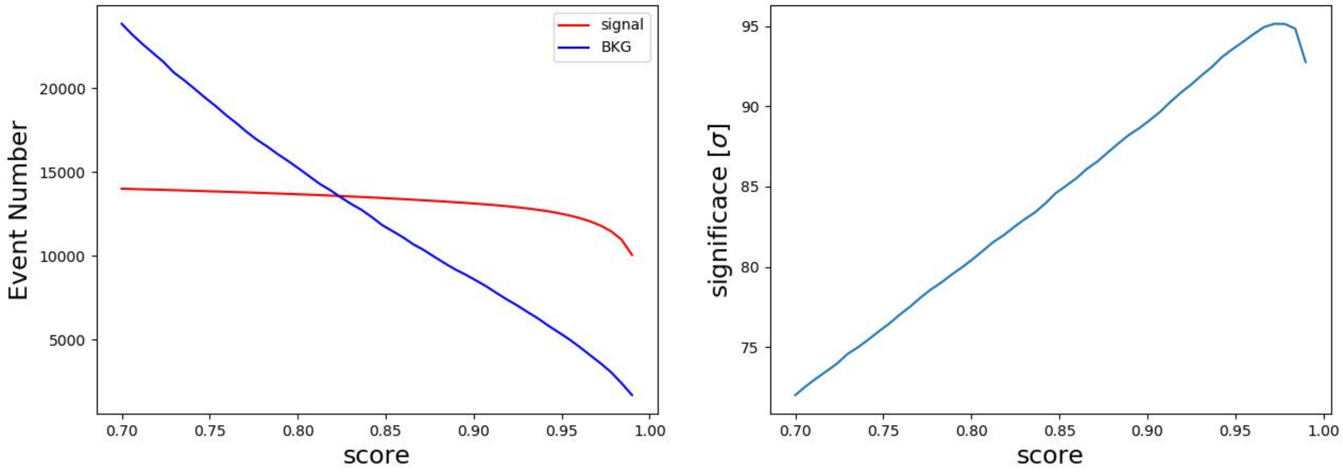
BDT & FCNN networks

BDT input, cut-based physics quantities…

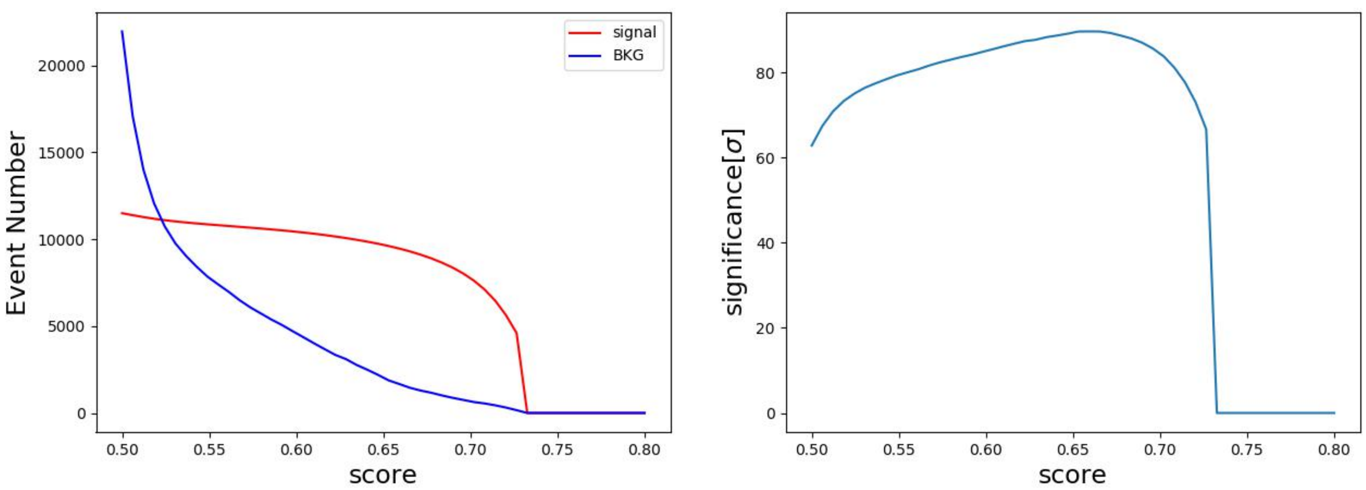


Input data is same with BDT

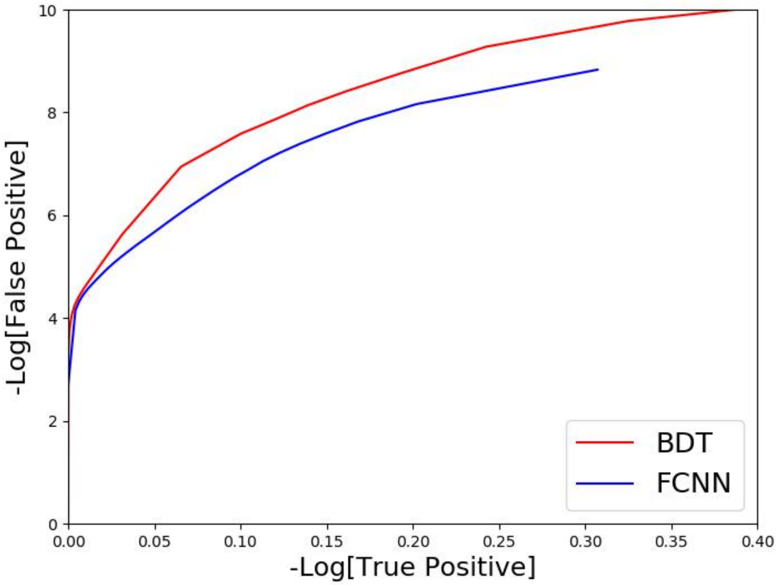




FCNN result



BDT result



	$ee \to WW$	$ee \to q\bar{q}$	$ee \to ZZ$	$ee \to e\nu W$	$ee \to eeZ$	$ee \to ZH$	$ee \to WWH$	Significance
BDT-only	67	195	399	9	0	6721	1044	84.6σ
FCNN-only	835	1101	419	109	1	9562	1423	94.7σ
CNN-only	7012	5407	432	1237	552	4776	747	38.9σ
Cut-based only	1856	277	856	50	0	4331	474	52.7σ
Cut+BDT	11	129	57	1	0	3744	402	62.9σ
Cut+FCNN	105	145	170	0	0	3941	408	63.0σ
Cut+CNN	320	219	486	0	0	3733	408	57.6σ

Table 1: Event number after cut-based selection. The result is obtained with $\sqrt{s} = 250GeV$ and $\mathcal{L} = 250fb^{-1}$.

Results

The deviation from SM predictions is defined as,

$$\sigma^2 \equiv \frac{(N_{\rm exp} - N_{\rm SM})^2}{N_{\rm SM}} \,,$$

The actural and SM branching ratio can be expressed as

$$\begin{aligned} N^{\rm exp} &= N_{jj}\left(\mathrm{Br}(h \to f\bar{f})\epsilon_{ff} + \mathrm{Br}(h \to \mathrm{else})\epsilon_{\rm bkg}^{h\to jj}\right) \\ &+ N_{\mathrm{non}-jj}\epsilon_{\rm bkg}^{\mathrm{non}\to jj} \,, \\ N^{SM} &= N_{jj}\left(\mathrm{Br}^{\rm SM}(h \to f\bar{f})\epsilon_{ff} + \mathrm{Br}^{\rm SM}(h \to \mathrm{else})\epsilon_{\rm bkg}^{h\to jj}\right) \\ &+ N_{\mathrm{non}-jj}\epsilon_{\rm bkg}^{\mathrm{non}\to jj} \,, \end{aligned}$$

where the the signal strength μ_{ff} is defined as

$$\mathrm{Br}(h \to s\bar{s}) = \mu_{ff}\mathrm{Br}^{\rm SM}(h \to s\bar{s}) \,,$$

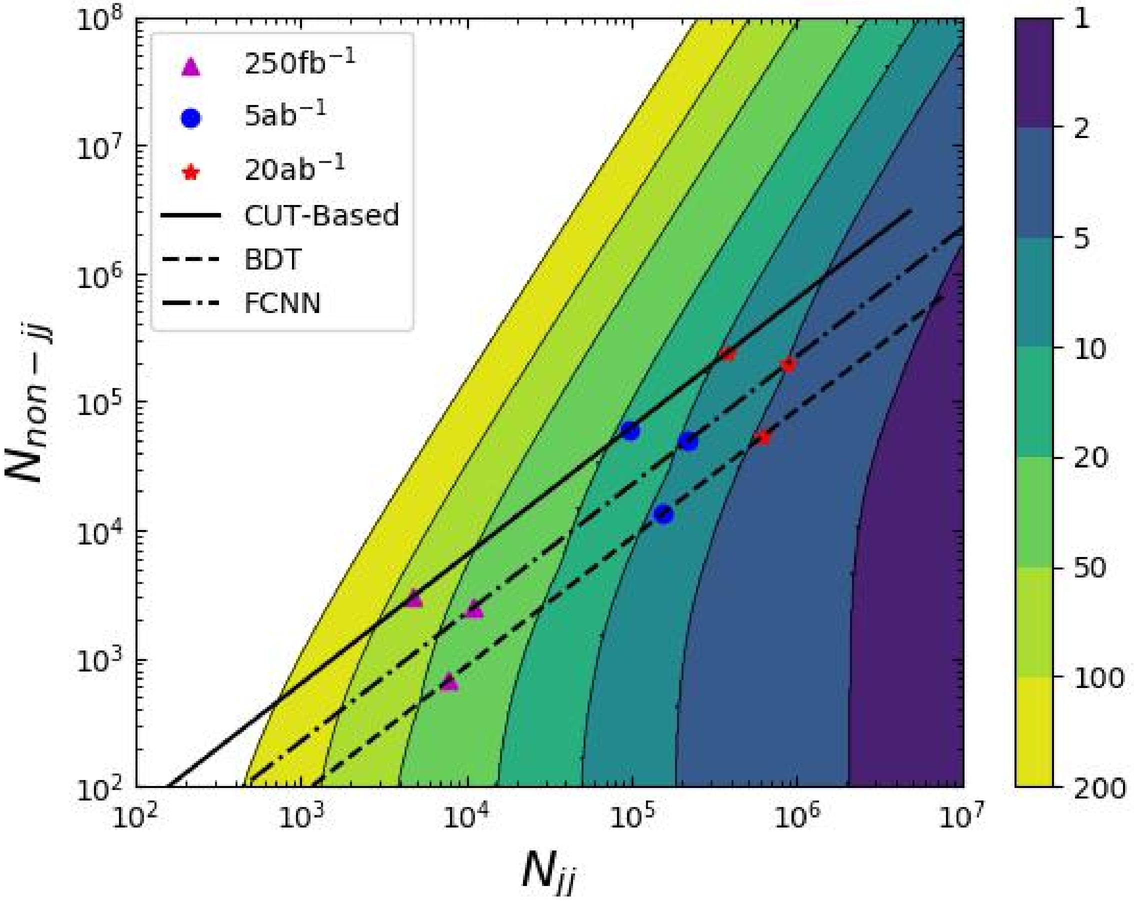
$$\mathrm{Br}(h \to else) = 1 - \mu_{ff}\mathrm{Br}^{\mathrm{SM}}(h \to f\bar{f})\,,$$
$$\mathrm{Br}^{\mathrm{SM}}(h \to else) = 1 - \mathrm{Br}^{\mathrm{SM}}(h \to f\bar{f})\,,$$

so that

$$N_{\exp} - N_{\mathrm{SM}} = N_{jj}\mathrm{Br}^{\mathrm{SM}}(h \to s\bar{s}) \cdot (\mu - 1)\left(\epsilon_{ss} - \epsilon_{\mathrm{bkg}}^{h\to jj}\right)\,.$$

Substitute the equation in the significance σ expression, we have

$$\sigma^2 \equiv \frac{N_{jj}\mathrm{Br}^{\mathrm{SM}}(h \to s\bar{s}) \cdot (\mu - 1)\left(\epsilon_{ss} - \epsilon_{\mathrm{bkg}}^{h\to jj}\right)}{N_{jj}\left(\mathrm{Br}^{\mathrm{SM}}(h \to f\bar{f})\epsilon_{ff} + \mathrm{Br}^{\mathrm{SM}}(h \to \mathrm{else})\epsilon_{\mathrm{bkg}}^{h\to jj}\right) + N_{\mathrm{non}-jj}\epsilon_{\mathrm{bkg}}^{\mathrm{non}\to jj}}$$
$$= F(N_{jj}, N_{\mathrm{non}-jj})$$



250fb^{-1}
5ab^{-1}
20ab^{-1}
CUT-Based
BDT
FCNN
N_{non-jj}
N_{jj}
10^2
10^3
10^4
10^5
10^6
10^7
10^8
1
2
5
10
20
50
100
200

μ_{ss} at CEPC

Summary

- We propose a method to improve the efficiency of preselection in Higgs signal searches at CEPC.
- For this propose we developed three machine learning algorithms including boosted decision tree algorithm, fully-connected neural networks and convolutional neural networks.
- Among all these algorithms, we found the fully-connected neural networks gives the best prediction on Higgs signals. Using such algorithms, we improve the signal strength of s-tagging events from cut-based result, $\mu_{ss} \sim 200$, to FCNN result $\mu_{ss} \sim 50$.