中国科学院高能物理研究所
Institute of High Energy Physics, Chinese Academy of Sciences

CEPC

Measurement of branching ratios of Higgs hadronic decays

Yanping Huang, Xinchou Lou, Gang Li,
Gangcheng Lu,Jinfei Wu

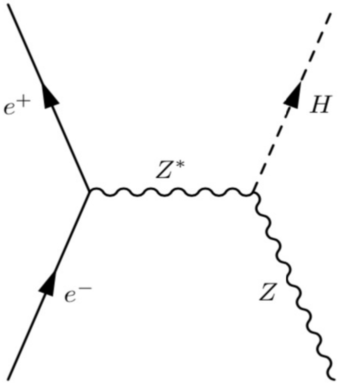
IHEP,CAS

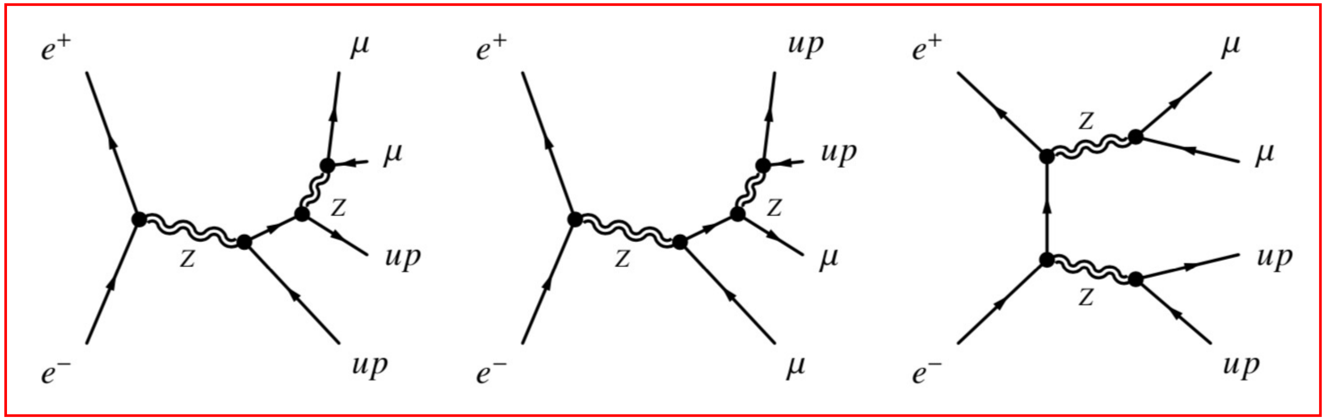
Introduction

- Branch ratios of Higgs hadronic decays : measured in LHC, but limited by the large background.
- CEPC has advantages to perform more precise measurement,
  - Small background
  - Tunable initial energy
- In the previous study at CEPC, the branch ratios of $H \to b\bar{b}/c\bar{c}/gg$ is measured with a 3D-fit method. Previous Study
- To improve the measurement and include more decay channels, a **machine learning** technique and **matrix method** are introduced in this study.
  - $H \to b\bar{b}/c\bar{c}/gg/ww^*/zz^*$

Introduction

- Measurement of decay branching ratios of $H \to b\bar{b}/c\bar{c}/gg/ww^*/zz^*$ in associated with $z \to \mu^+\mu^-$at the CEPC.
- MC samples : e2e2h_bb, e2e2h_cc, e2e2h_gg, e2e2h_ww, e2e2h_zz, zz_sl0mu_down and zz_sl0mu_up

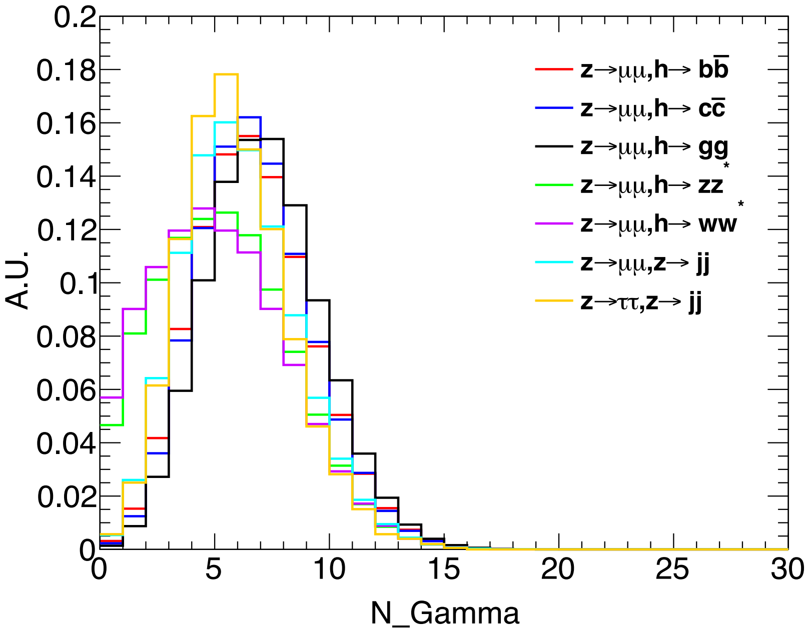


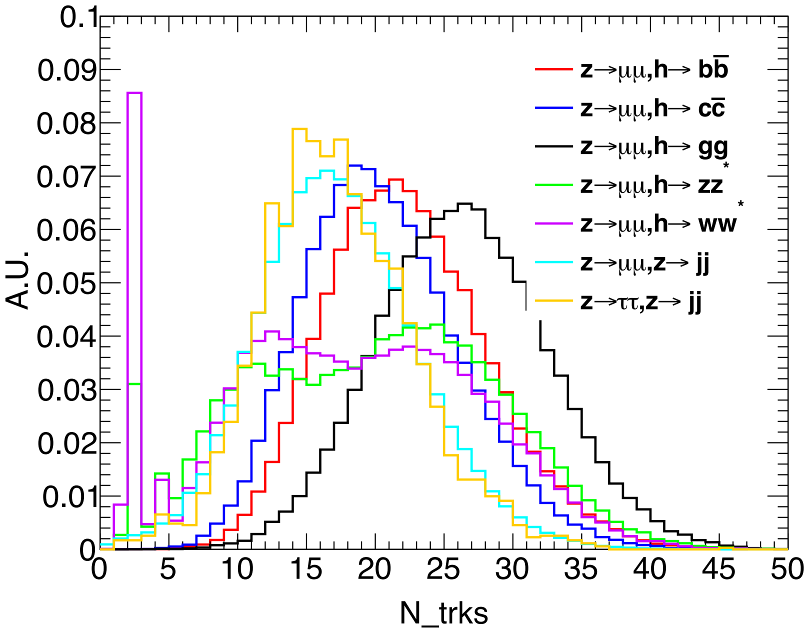


- For each sample, 60K full simulation events are used.

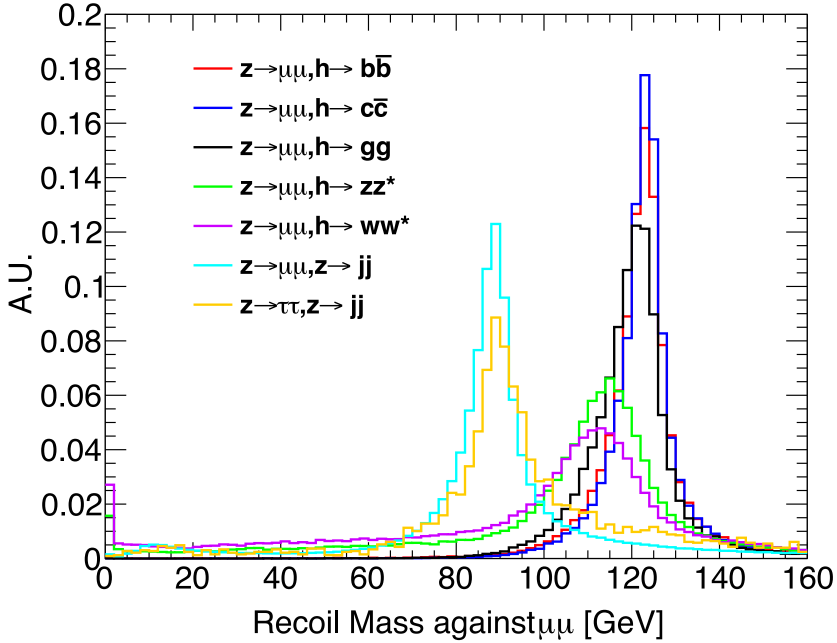
Pre-selection

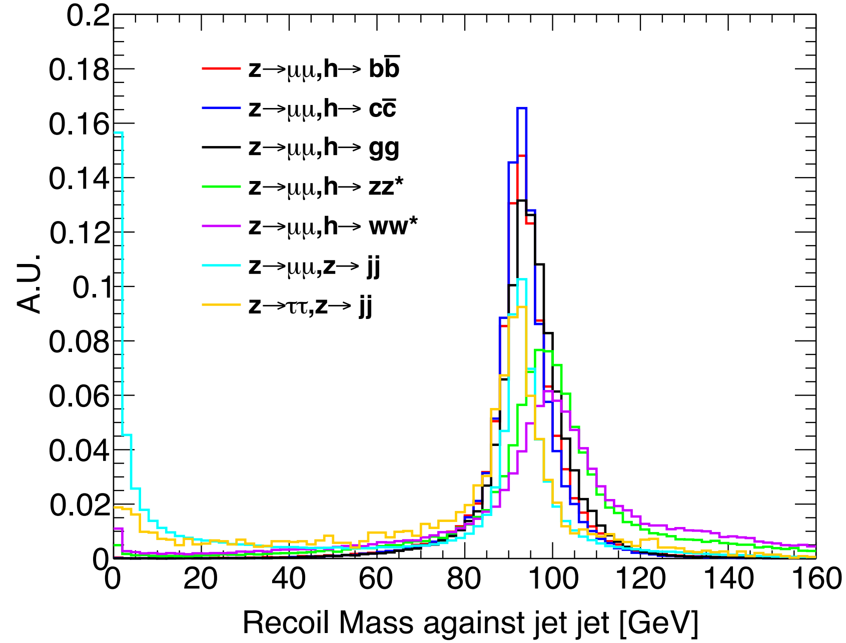
- To select the signal, a simple pre-selection is required:
  - $N_{jet} = 2$
  - $N_{\mu} = 2$
- After the pre-selection, the distributions of some variables are shown:





Pre-selection





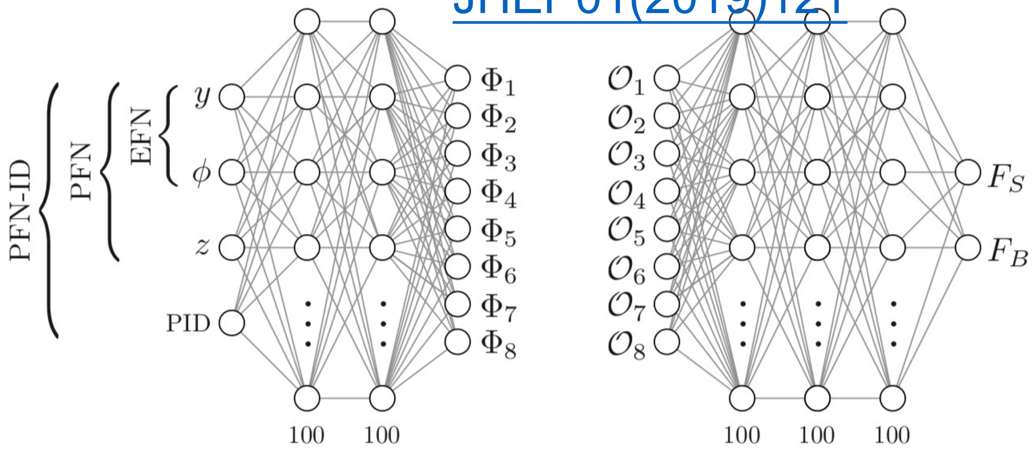
- From these distributions, a cut-based method including several variables cannot separate the different channels very well.
- A machine learning technique is introduced to improve the performance.

Particle Flow Network

- PFN : Particle Flow Networks is a model architectures designed for learning from collider events.

JHEP01(2019)121

$$PFN = F(\sum_{i=1}^{M} \Phi(p_i))$$

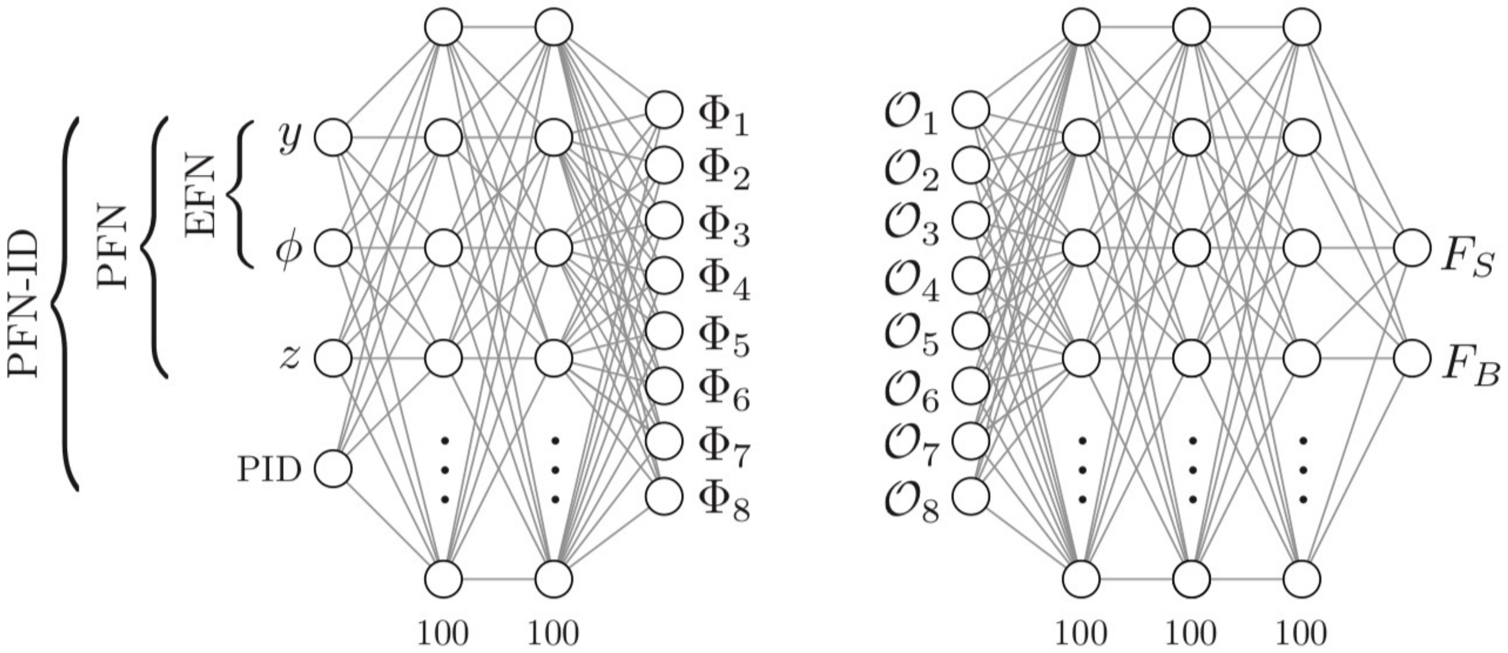


- p_i : the information of particle i, such as four-momentum, charge, or flavor.

- **Advantage**
  - use all info at particle level,
  - remove the impact from jet clustering and e/γ isolation,
  - Enlarge the size of input,

Config of PFN

- Training variables
  - Energy, momentum, $cos\theta$, ϕ, PDGID, D0, Z0 for each particle in a jet.
- Network architecture



3 layers : 100, 100, 256 nodes   3 layers : 100, 100, 100 nodes

Activation function : ReLU for each dense layer, softmax for output layer

Methodology

- Let's consider a simple example, only $H \to b\bar{b}$ and $H \to c\bar{c}$.

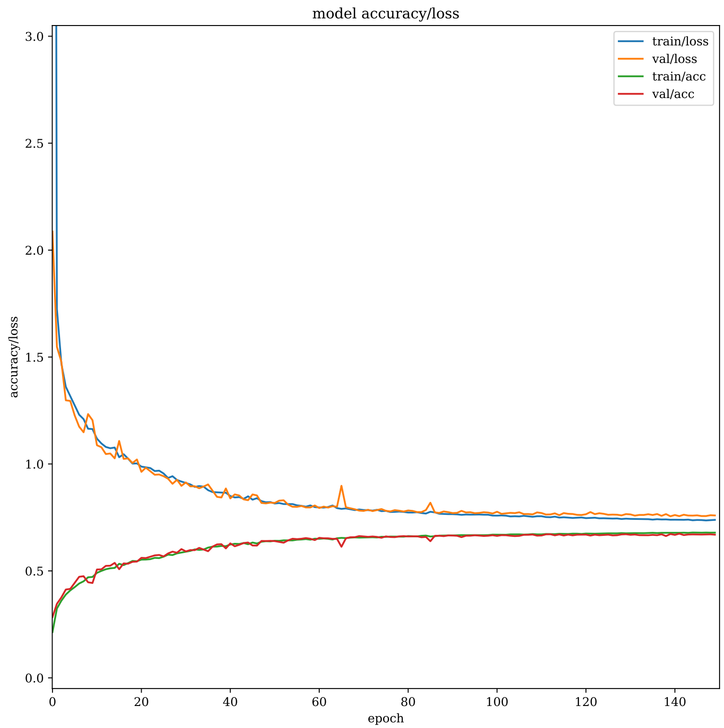
$$\begin{pmatrix} n_b \\ n_c \end{pmatrix} = \begin{pmatrix} \epsilon_{bb} & \epsilon_{bc} \\ \epsilon_{cb} & \epsilon_{cc} \end{pmatrix} \begin{pmatrix} N_b \\ N_c \end{pmatrix} \qquad n = EN$$

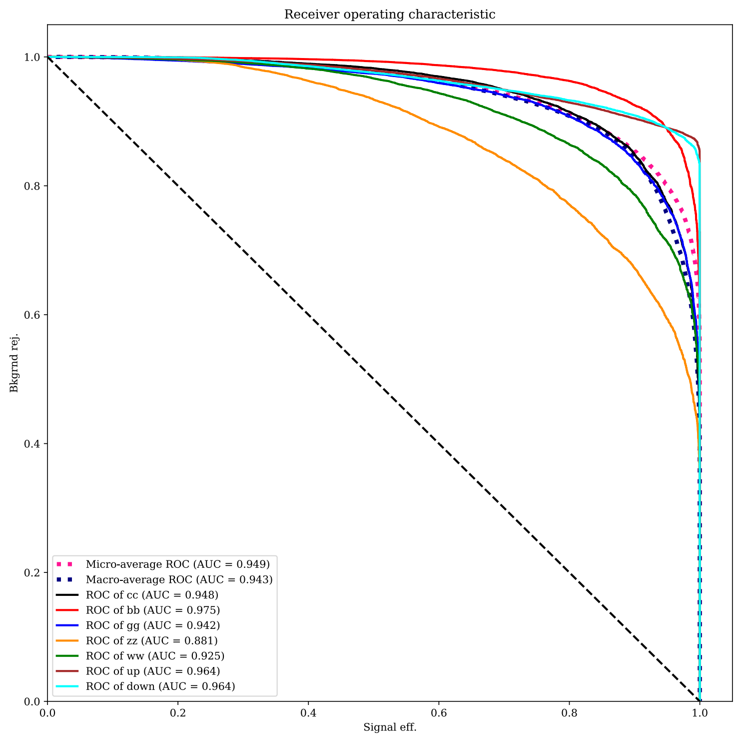
- n_i : the observed number of events of i class,
- N_i : the production number of events of i class,
- ϵ_{ij} : the rate of state i reconstructed to be state j.

- If we can measure the matrix E, then $N = E^{-1}n$
- The PFN is used to extract the matrix.

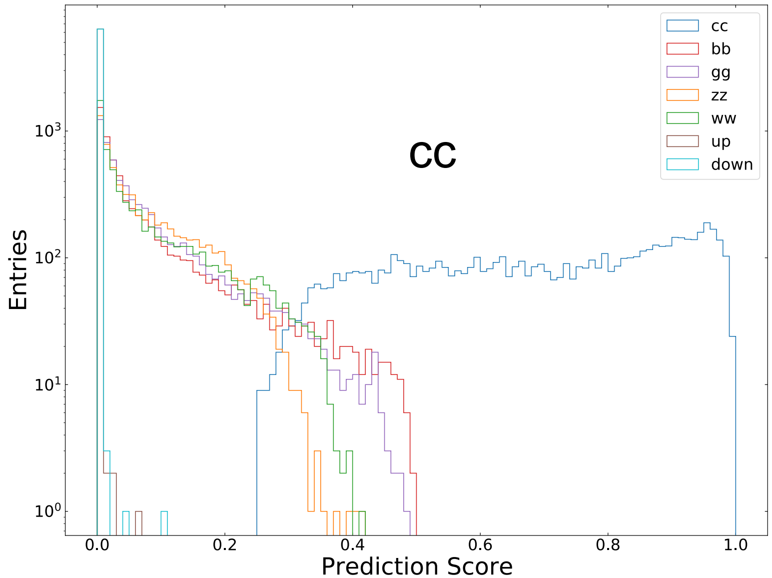
Results of full simulation

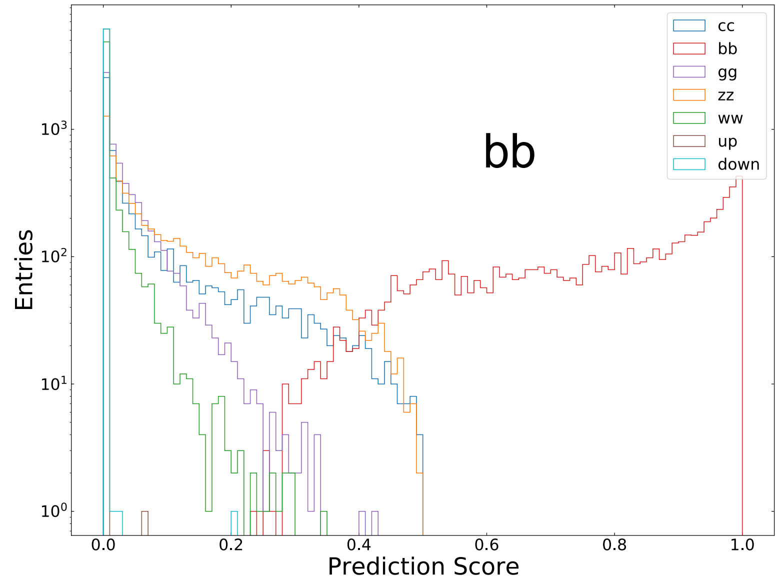
- Use full simulation sample, including $b\bar{b}/c\bar{c}/gg/ww^*/zz^*$, zz_sl0mu_down(label as down) and zz_sl0mu_up(label as up).
- Tiny difference between train and validation on loss.
- From the ROC curve, the separation power of $b\bar{b}$ is highest, zz^* is lowest.

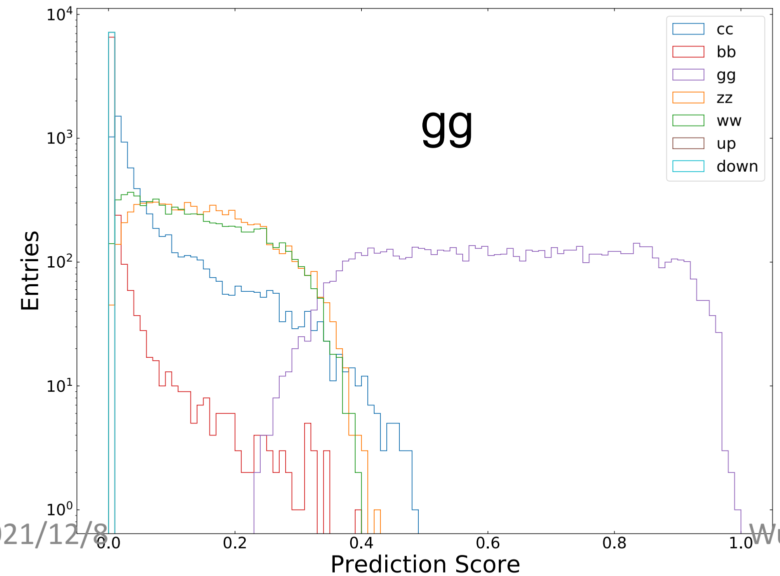


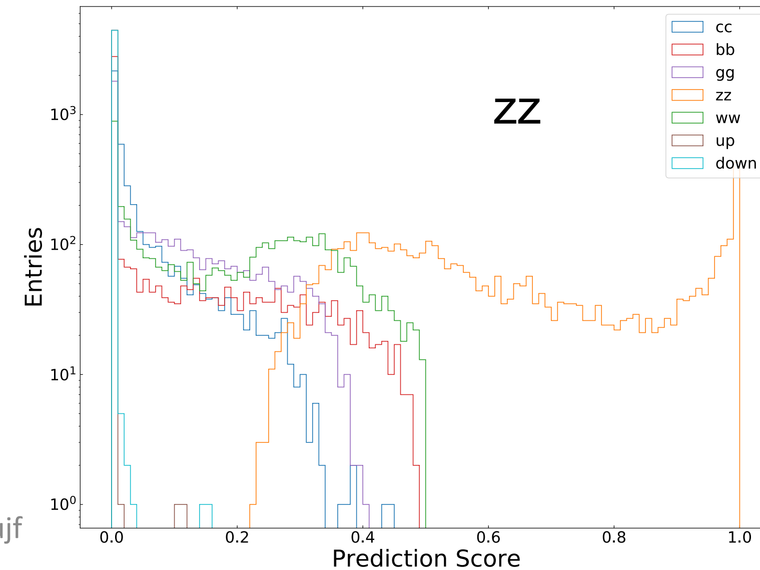


Results of full simulation

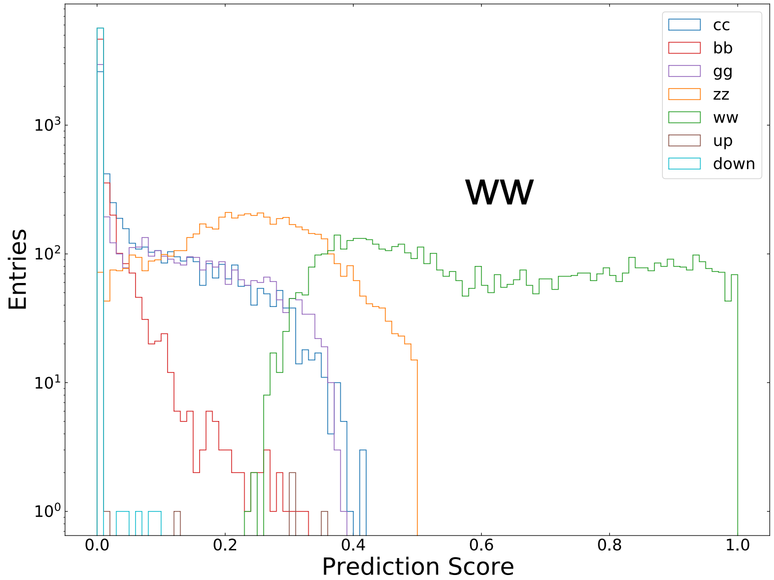


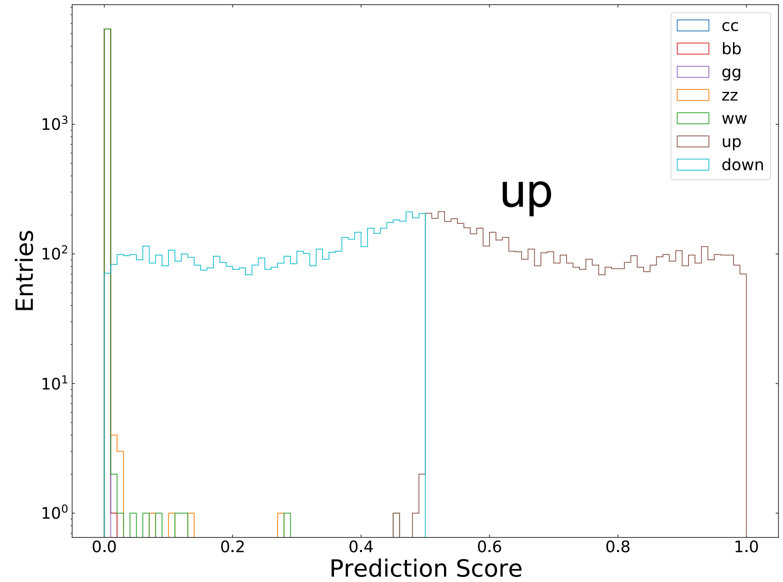


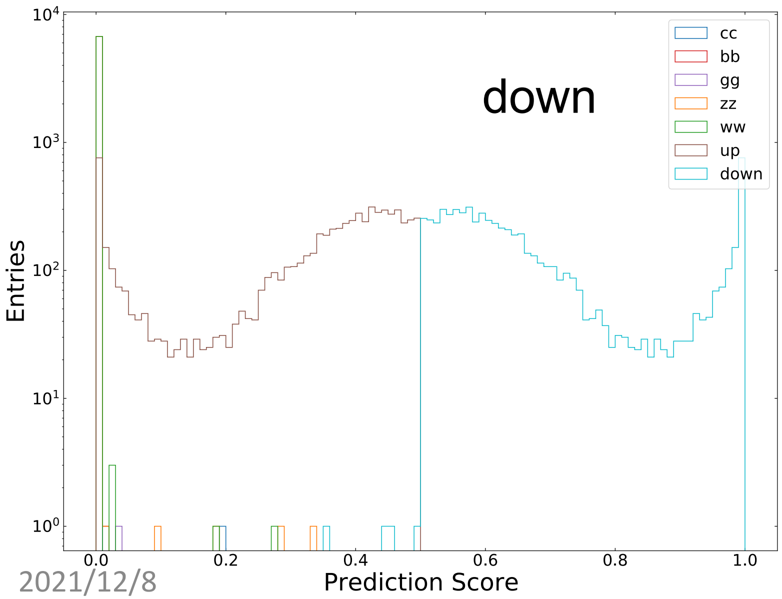




Results of full simulation



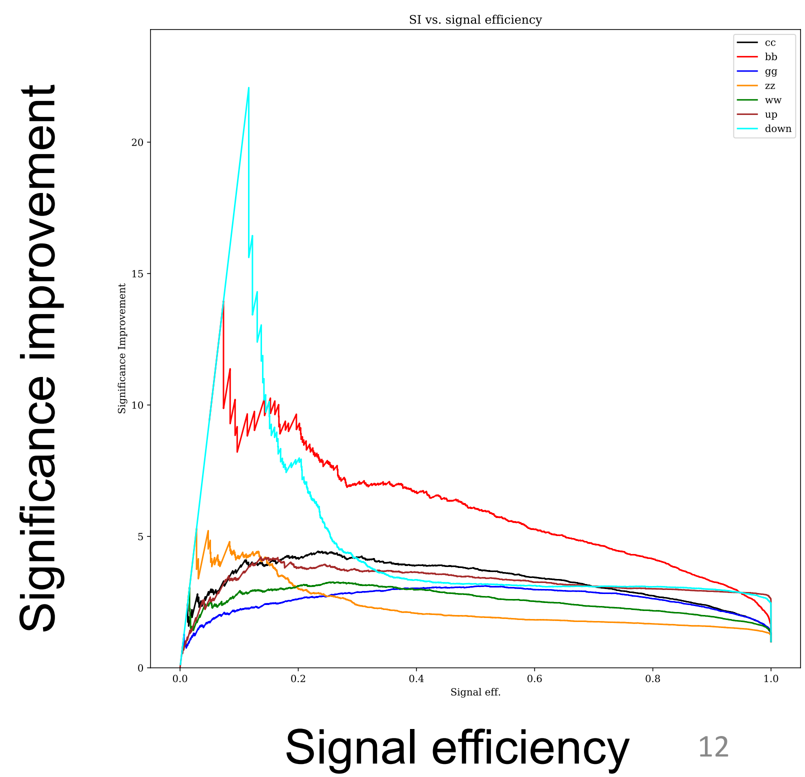




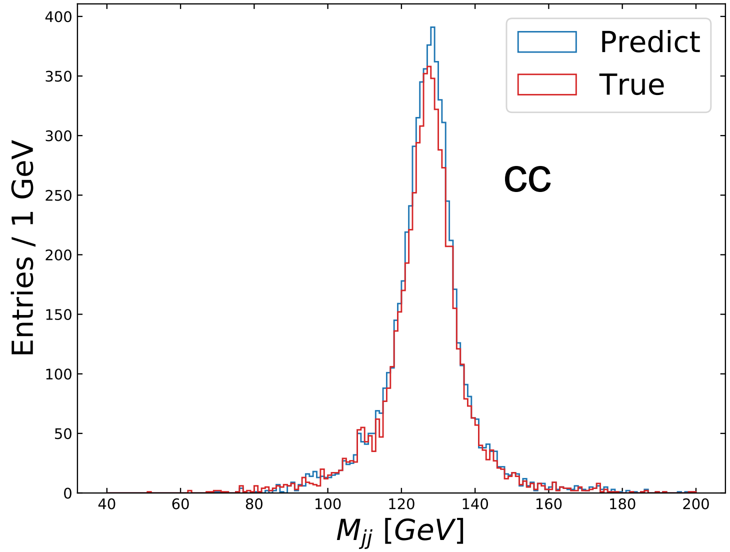
The good separation power among each class.

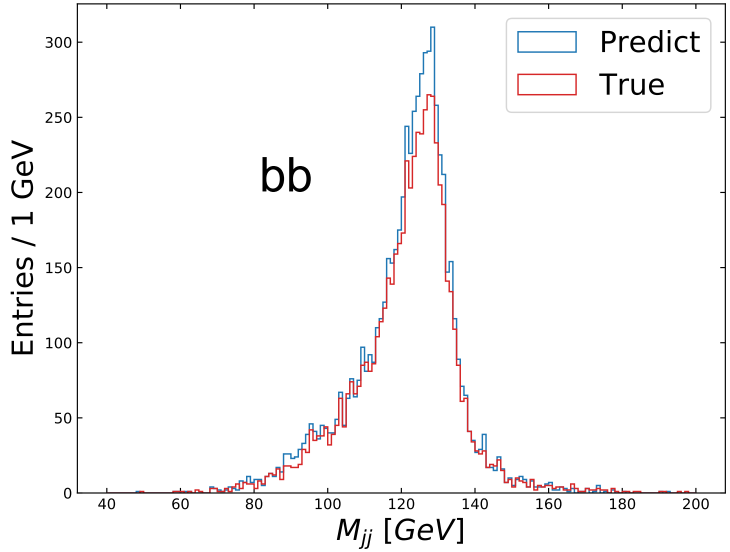
Results of full simulation

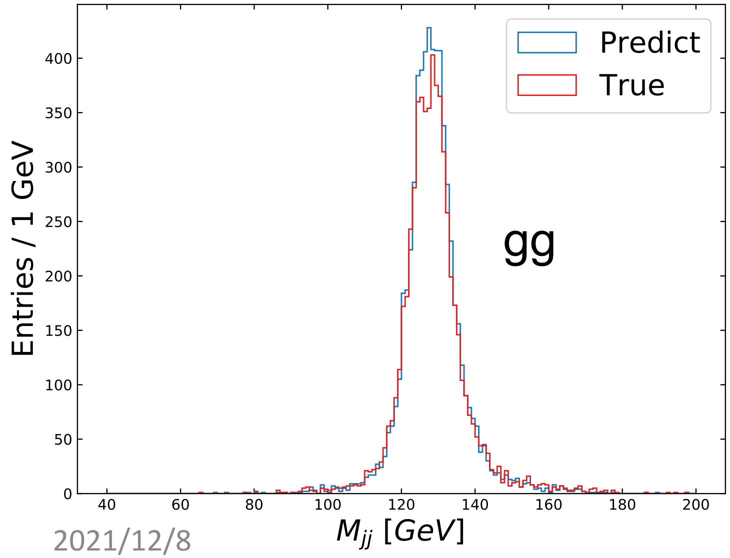
- For the classes of signal, the separation power of $b\bar{b}$ is highest, zz^* is lowest.
- The separation power between background is lower, but the separation power between signal and background is higher.



Performance of PFN

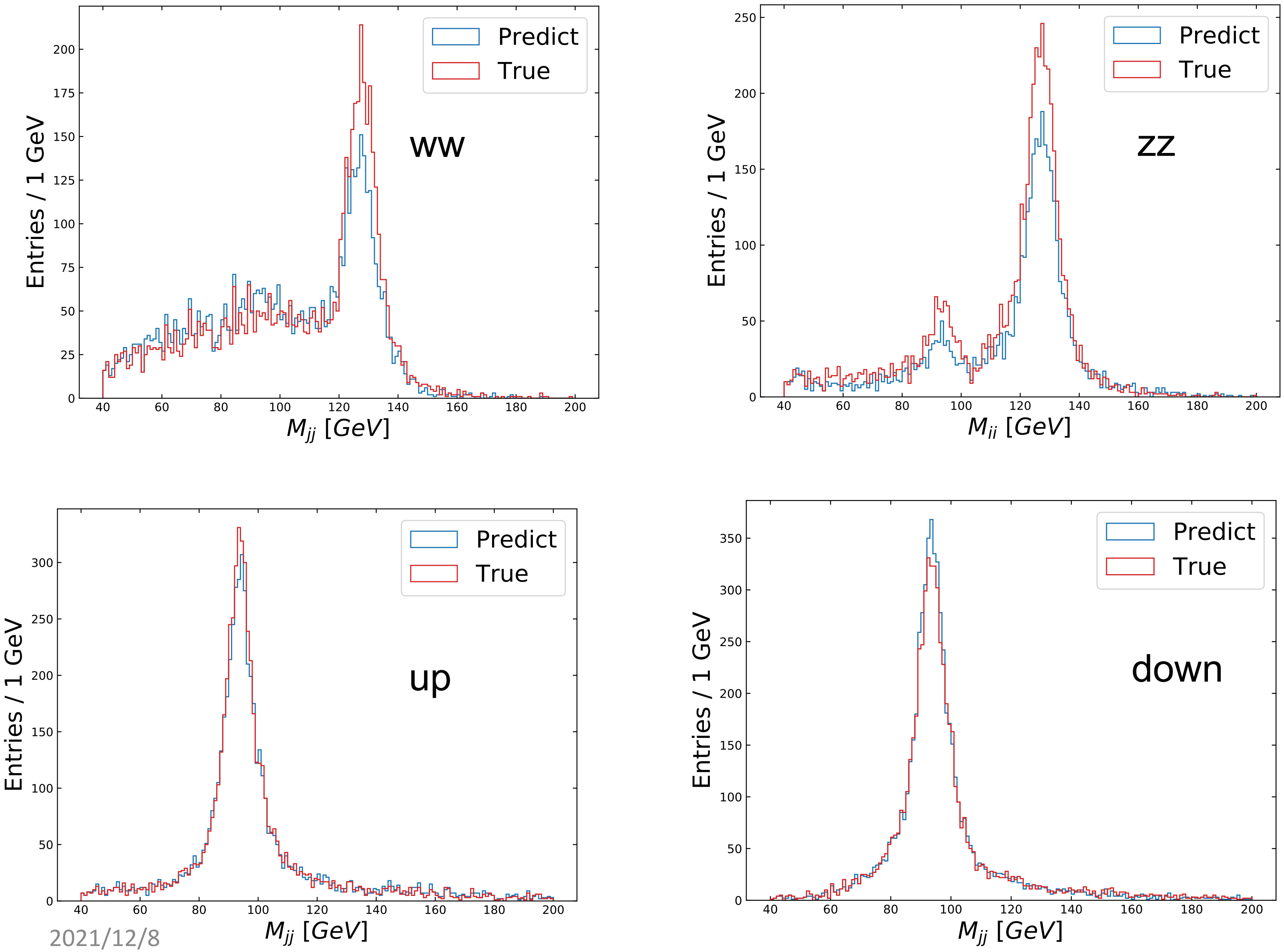






- Show the comparison of true and prediction on the test sample(10% of total MC samples).
- The predictions of cc, bb and gg are better, the difference between predict and true is small.

Performance of PFN



Results

- Use test samples(10% of MC events) to perform the study.
- Scale the MC events according to the cross-section × integrated lumi(5.6 ab^{-1})

	$c\bar{c}$	$b\bar{b}$	gg	zz	ww	up	down
n	1272±36	21435±146	3689±61	8822±94	11709±334	66245±257	105853±325
$\hat{N}$	1079±33	21389±146	3177±56	14189±119	107436±328	72711±270	97784±313
N	1089±33	21539±147	3079±55	14430±119	108045±329	72729±270	98448±314

n : observed number of events of each channel,
$\hat{N}$: the true number of events of each channel,
N : the number of events of each channel, calculated from observed number.

Summary

- Use PFN to classify the hadronic decay channels of Higgs, including $H \rightarrow b\bar{b}/c\bar{c}/gg/ww^*/zz^*$.
- The separation power between background is lower, but the separation power between signal and background is higher.
- Potential improvement with respect to previous study.
  - Good separation power of each class from PFN.
  - Smaller global uncertainty from the matrix method.
- Next to do
  - Include more backgrounds and try to improve the performance of PFN.
  - Extract the branch ratio of Higgs hadronic decay channels
  - Estimate the systematic uncertainty.

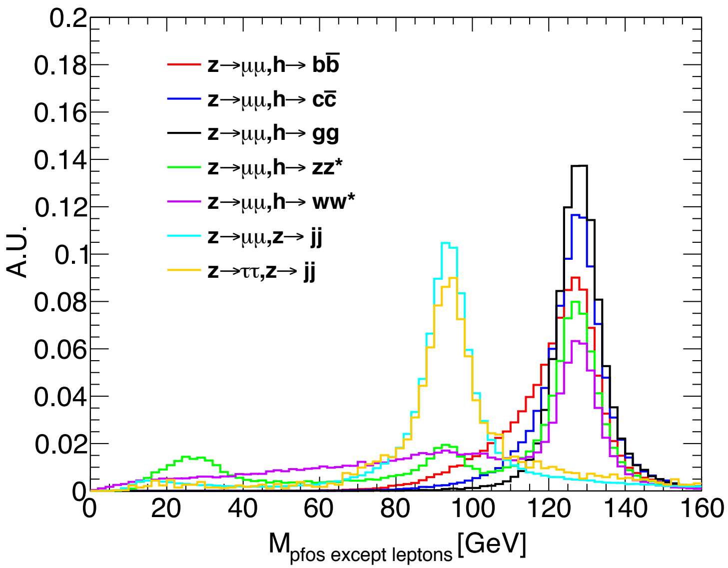
Backup

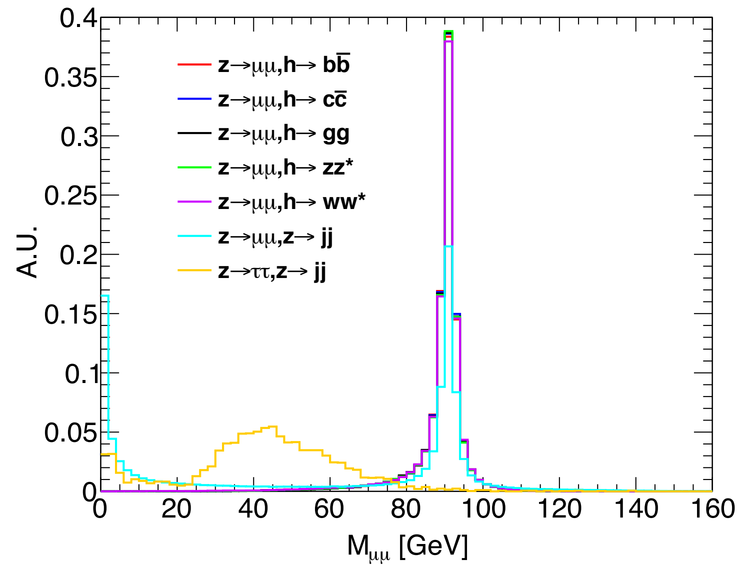
Table 2. Uncertainties on $\sigma^{b\bar{b}}_{l^+l^-H}$, $\sigma^{c\bar{c}}_{l^+l^-H}$ and $\sigma^{gg}_{l^+l^-H}$.

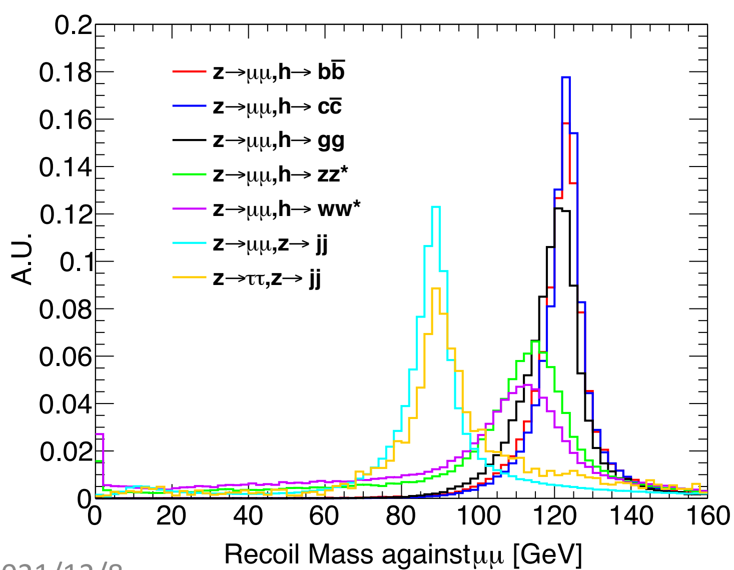
Higgs boson production	$\mu^+\mu^-H$			e^+e^-H		
Higgs boson decay	$H\to b\bar{b}$	$H\to c\bar{c}$	$H\to gg$	$H\to b\bar{b}$	$H\to c\bar{c}$	$H\to gg$
statistic uncertainty	1.1%	10.5%	5.4%	1.6%	14.7%	10.5%
fixed background	−0.2% / +0.1%	+4.1% / −4.2%	7.6%	−0.2% / +0.1%	+4.1% / −4.2%	7.6%
event selection	+0.7% / −0.2%	+0.4% / −1.1%	+0.7% / −1.7%	+0.7% / −0.2%	+0.4% / −1.1%	+0.7% / −1.7%
flavor tagging	−0.4% / +0.2%	+3.7% / −5.0%	+0.2% / −0.7%	−0.4% / +0.2%	+3.7% / −5.0%	+0.2% / −0.7%
combined systematic uncertainty	+0.7% / −0.5%	+5.5% / −6.6%	+7.6% / −7.8%	+0.7% / −0.5%	+5.5% / −6.6%	+7.6% / −7.8%

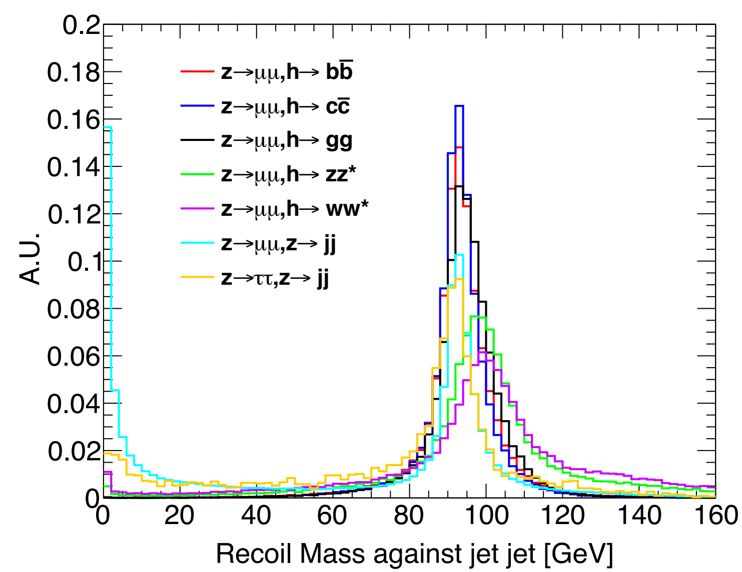
Backup

Pre-selection





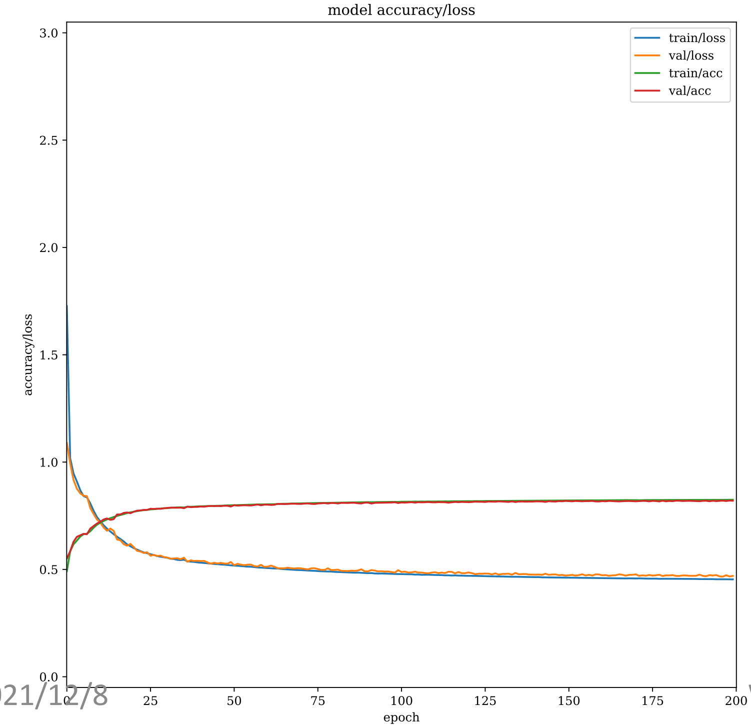


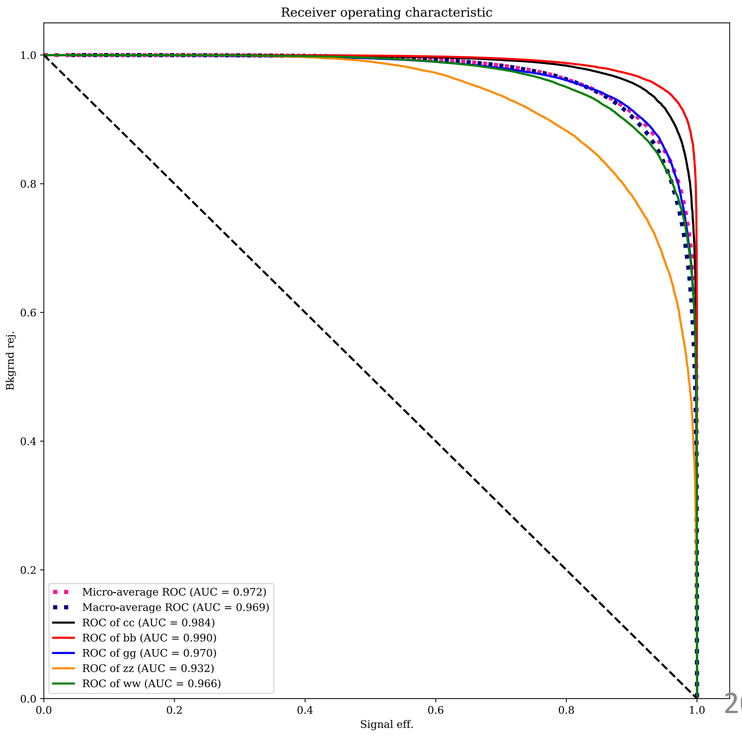


Results of fast simulation

- Fast simulation sample : only has $b\bar{b}/c\bar{c}/gg/ww^*/zz^*$.
- Tiny difference at loss between train and validation.
- From the ROC curve, the separation power of $b\bar{b}$ is highest, zz^* is lowest.

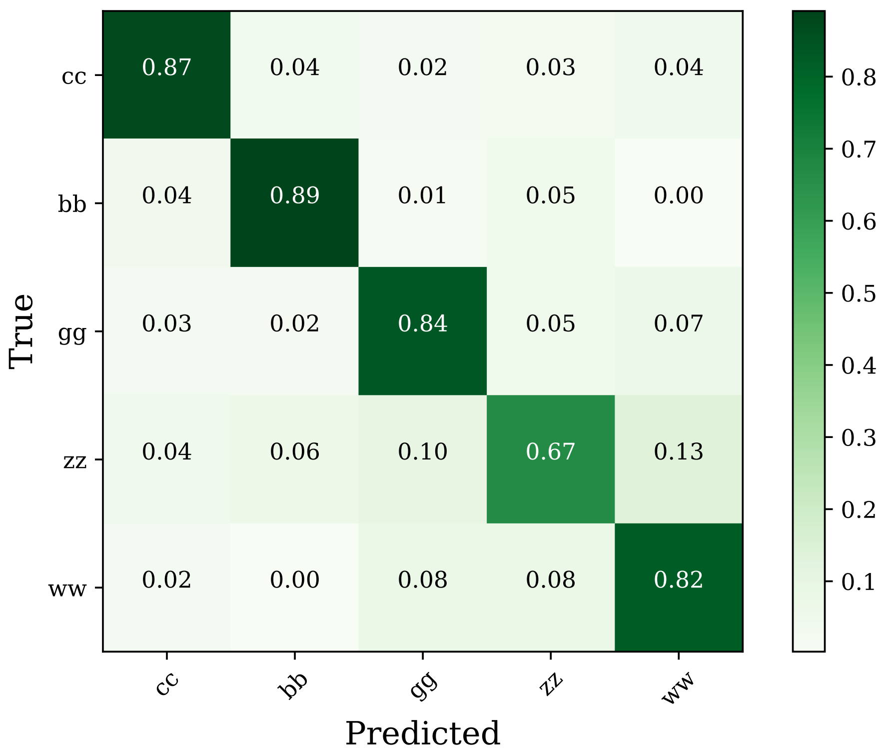
https://github.com/Wujinfei/HiggsHadron-PFNs-gpu.git

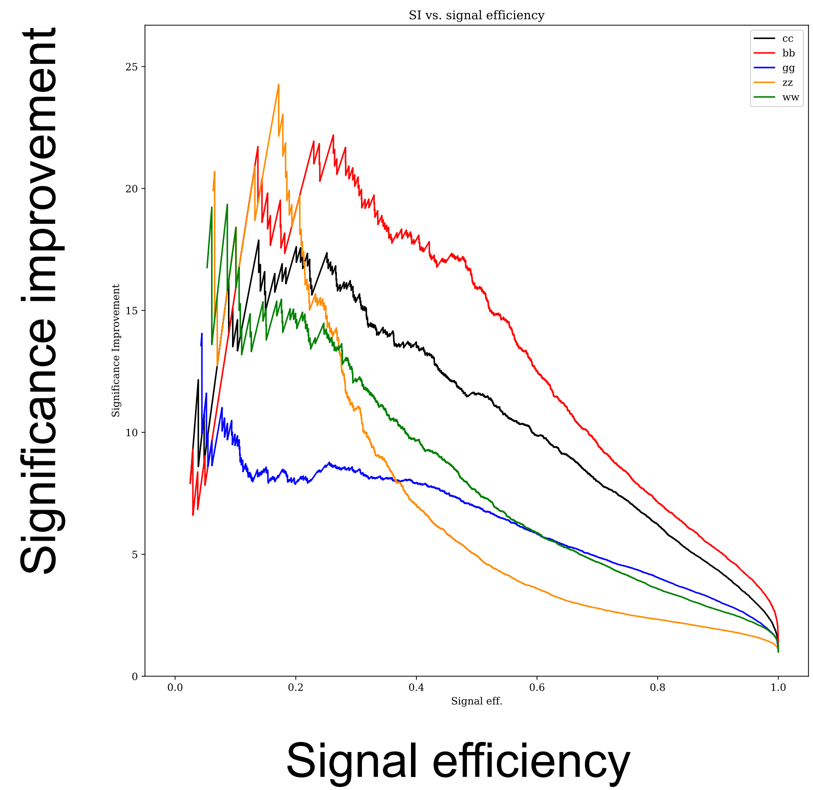




Results of fast simulation

- The performance of PFN on fast simulation is good, except the zz^* calss.





Comparison between fast and full simulation

- Why is the performance of full simulation worse than fast simulation:
  - Fast simulation has larger statistic than full simulation.
  - Maybe due to the reconstruction is not perfect.
  - Fewer training epochs of full simulation.

- Possible ways to improve the training performance
  - Include more input variables,
  - Generate more full simulation samples.