Updates of the global analysis

Gang Li, Libo Liao, Peixun Shen,
Weimin Song, Shudong Wang,
Zhaoling Zhang

CEPC detector plenary, June 16, 2021

Previous

Updates

- ☑ **Statistical method**
- ☑ **Efficiency matrix**
- ☑ **Combination**
- ☑ **Results**

Classification of 9 signals

Higgs -> cc, bb, mm, $\tau\tau$, gg, $\gamma\gamma$, ZZ, WW, γZ

1 2 3 4 5 6 7 8 9

$$\begin{pmatrix} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \\ n_6 \\ n_7 \\ n_8 \\ n_9 \end{pmatrix} = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} & \epsilon_{14} & \epsilon_{15} & \epsilon_{16} & \epsilon_{17} & \epsilon_{18} & \epsilon_{19} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} & \epsilon_{24} & \epsilon_{25} & \epsilon_{26} & \epsilon_{27} & \epsilon_{28} & \epsilon_{29} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{23} & \epsilon_{34} & \epsilon_{35} & \epsilon_{36} & \epsilon_{37} & \epsilon_{38} & \epsilon_{39} \\ \epsilon_{41} & \epsilon_{42} & \epsilon_{33} & \epsilon_{44} & \epsilon_{45} & \epsilon_{46} & \epsilon_{47} & \epsilon_{48} & \epsilon_{49} \\ \epsilon_{51} & \epsilon_{52} & \epsilon_{43} & \epsilon_{54} & \epsilon_{55} & \epsilon_{56} & \epsilon_{57} & \epsilon_{58} & \epsilon_{59} \\ \epsilon_{61} & \epsilon_{62} & \epsilon_{53} & \epsilon_{64} & \epsilon_{65} & \epsilon_{66} & \epsilon_{67} & \epsilon_{68} & \epsilon_{69} \\ \epsilon_{71} & \epsilon_{72} & \epsilon_{63} & \epsilon_{74} & \epsilon_{75} & \epsilon_{76} & \epsilon_{77} & \epsilon_{78} & \epsilon_{79} \\ \epsilon_{81} & \epsilon_{82} & \epsilon_{73} & \epsilon_{84} & \epsilon_{85} & \epsilon_{86} & \epsilon_{87} & \epsilon_{88} & \epsilon_{89} \\ \epsilon_{91} & \epsilon_{92} & \epsilon_{83} & \epsilon_{94} & \epsilon_{95} & \epsilon_{96} & \epsilon_{97} & \epsilon_{98} & \epsilon_{99} \end{pmatrix} \begin{pmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \\ N_5 \\ N_6 \\ N_7 \\ N_8 \\ N_9 \end{pmatrix}$$

- **N: signals generated**
- **n: observed candidates**
- **Determination of N or B is an inverse problem**

Method: solving B by minimizing the χ^2

$$\boxed{\chi^2 = \sum_i \frac{(\sum \epsilon_{ij} N_j - n_i)^2}{\sigma_{n_i}^2} + \frac{(\sum_l N_l - N)^2}{\sigma_N^2}}$$

Based on N
instead of n, multinomial

$$\sum_i B_i = 1$$

$$\mathbf{\Sigma}^N = N_t^e \begin{pmatrix} B_1(1-B_1) & -B_1B_2 & \ldots & -B_1B_m \\ -B_2B_1 & B_2(1-B_2) & \ldots & -B_2B_m \\ \vdots & \vdots & \ddots & \vdots \\ -B_mB_1 & -B_mB_2 & \ldots & B_m(1-B_m) \end{pmatrix}$$

Covariance of 9 Bs has a simple quantitive property

$$\mathbf{\Sigma}^{\mathbf{B}} = \mathbf{J}_{BN}\mathbf{E}^{-1}\mathbf{\Sigma}^{n}\left(\mathbf{E}^{T}\right)^{-1}\mathbf{J}_{BN}^{T}$$

$$\boxed{\Sigma_B \propto \frac{1}{N^2|E|^2}}$$

Statistics

Performances

Combining multiple tagging modes

$$\chi^2 = \chi^2_{ee} + \chi^2_{\mu\mu} + \chi^2_{\tau\tau} + \chi^2_{q\bar{q}}$$

Unknowns: 9 Bs (only 8 independent) and a cross section
Knowns: 36 ns

The key ingredient is efficiency matrix
To be determined with ML techniques

Preliminary performance :
Averaged accuracy ~ 87%

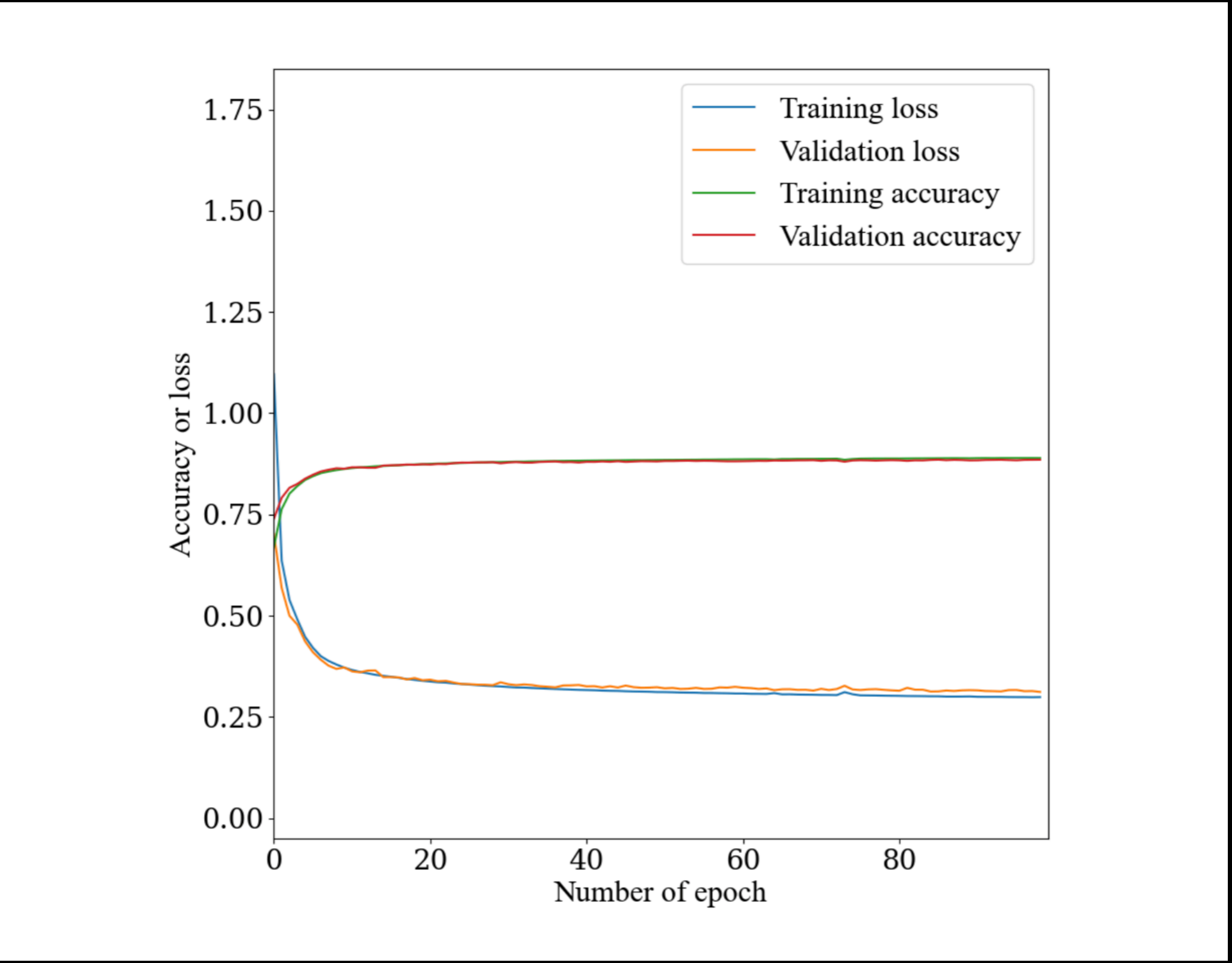
Updates

4x9 channels — 4 ML tasks

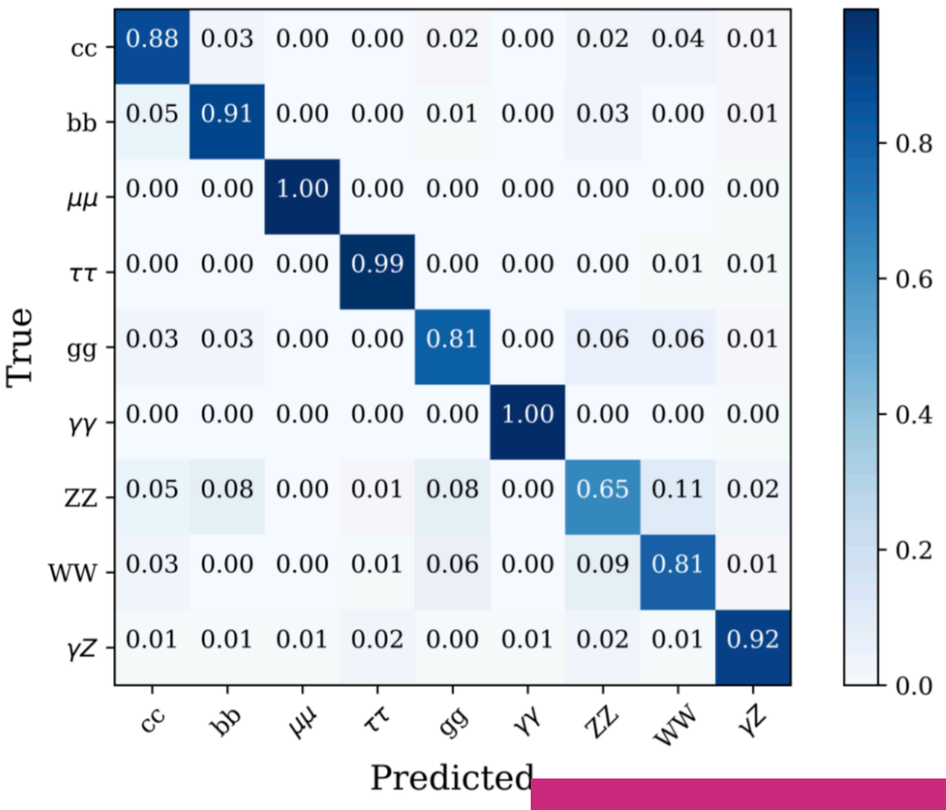
$$e^+e^- \rightarrow ZH, Z \rightarrow \mu^+\mu^-, e^+e^-, \tau^+\tau^-, q\bar{q}$$

- 400 k evts for each channel: $cc, bb, \mu\mu, \tau\tau, gg, \gamma\gamma, ZZ, WW, \gamma Z$
- Train: validation: test = 8:1:1
- Fast simulation
  - momenta of tracks and energies of neutrals smeared according to CEPC_v4
  - Ideal PID and Impact parameters
- The SM background left out temporally

Loss and accuracy
Take ee mode as an example

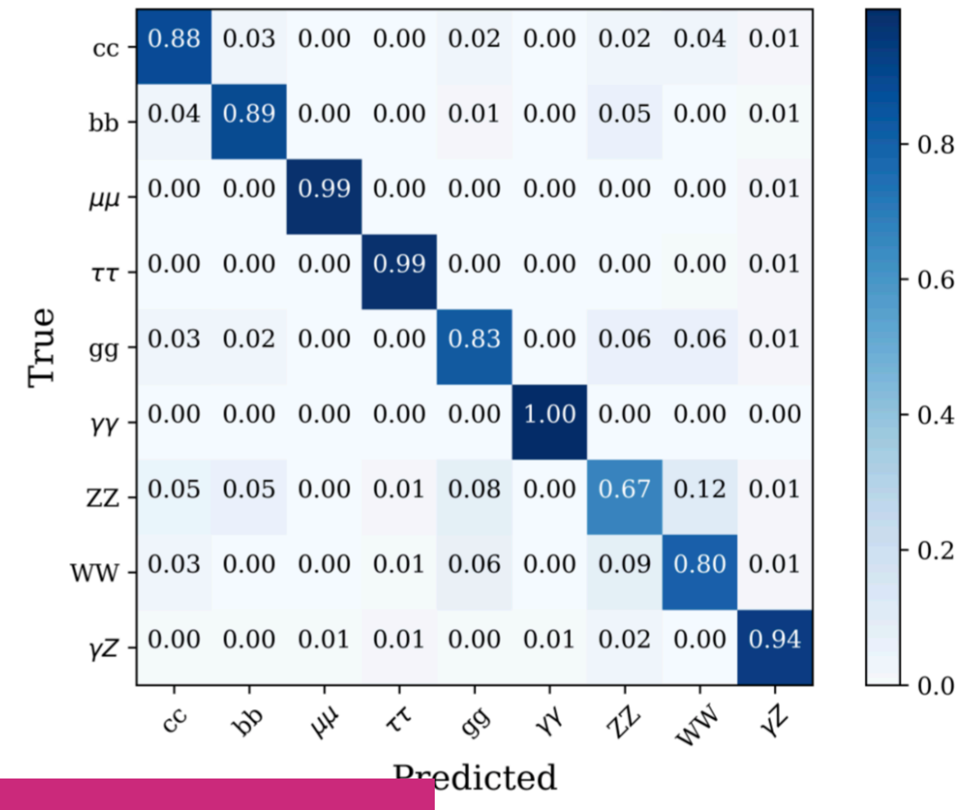


Converged without obvious over training

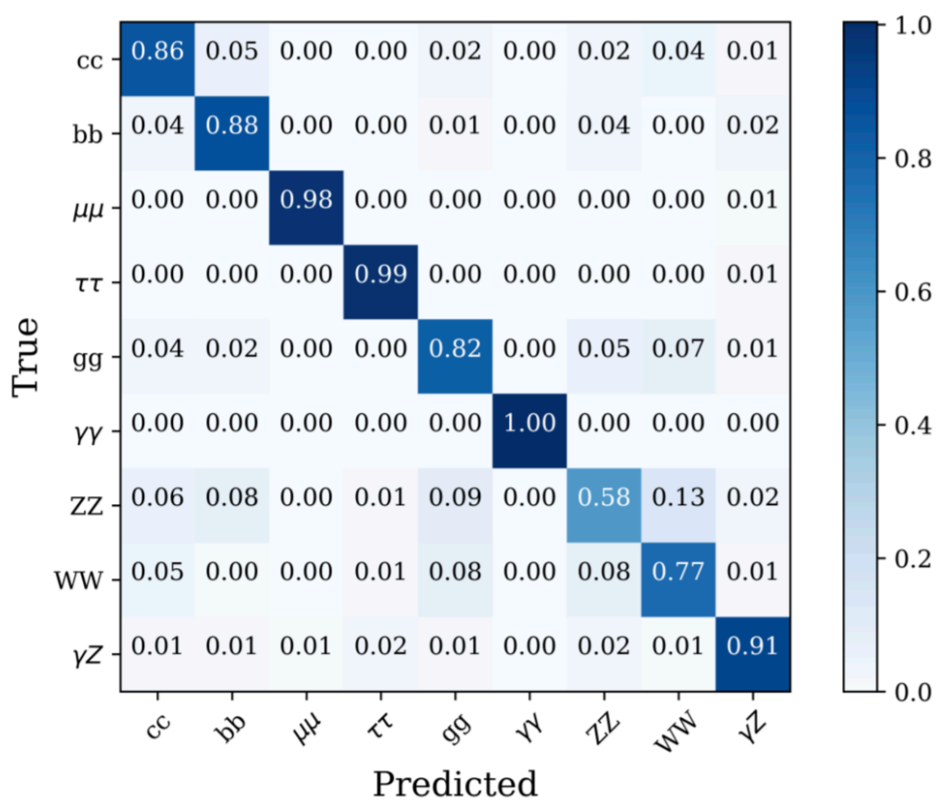


(a)

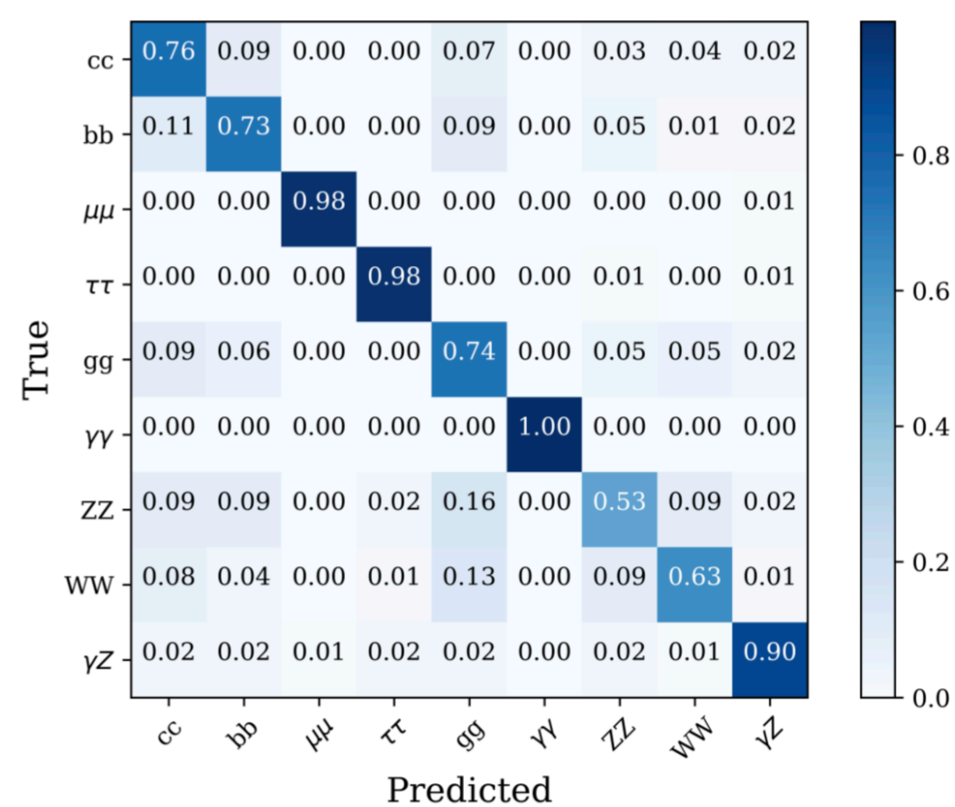
4 efficiency matrices



(b)

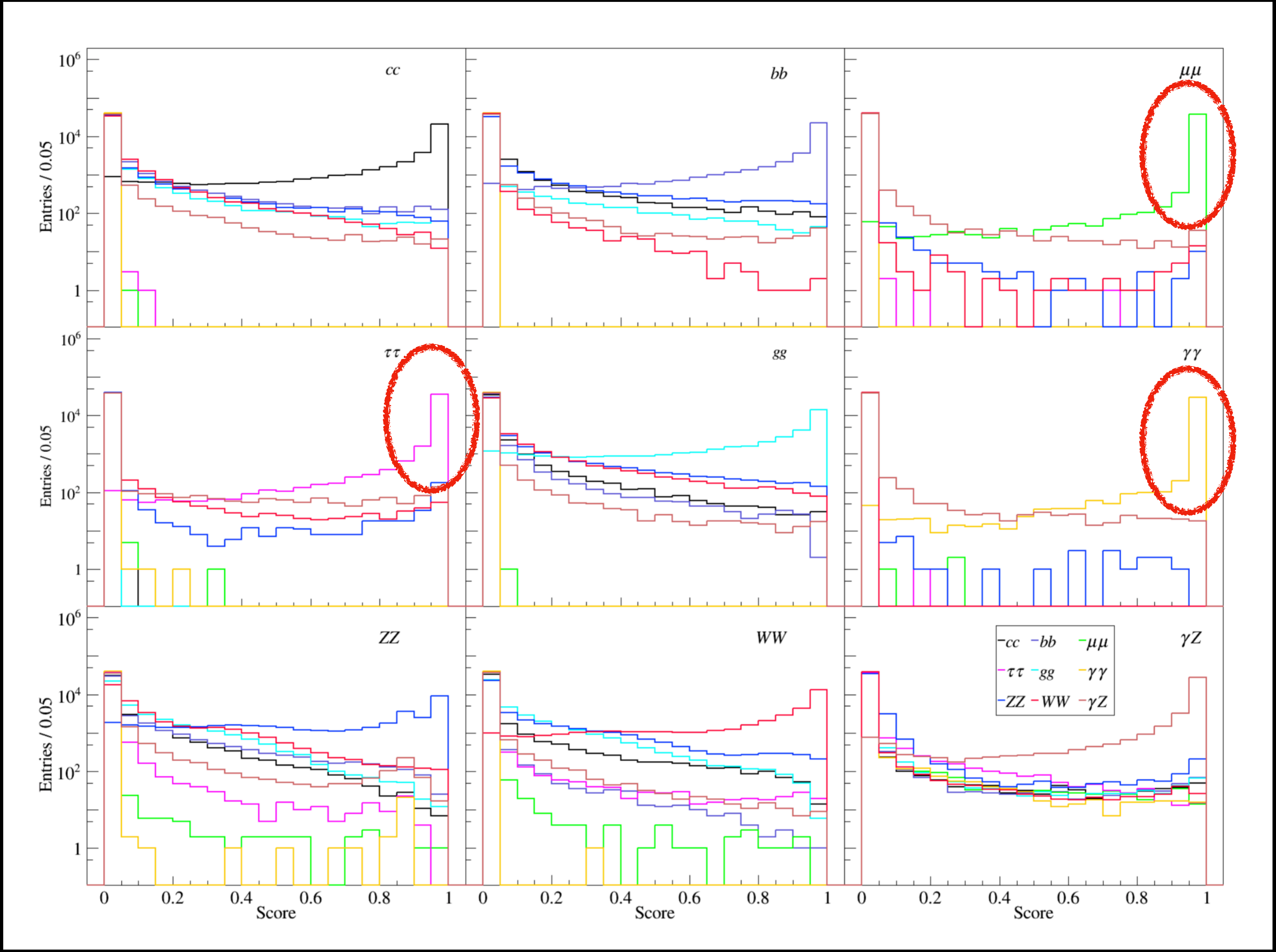


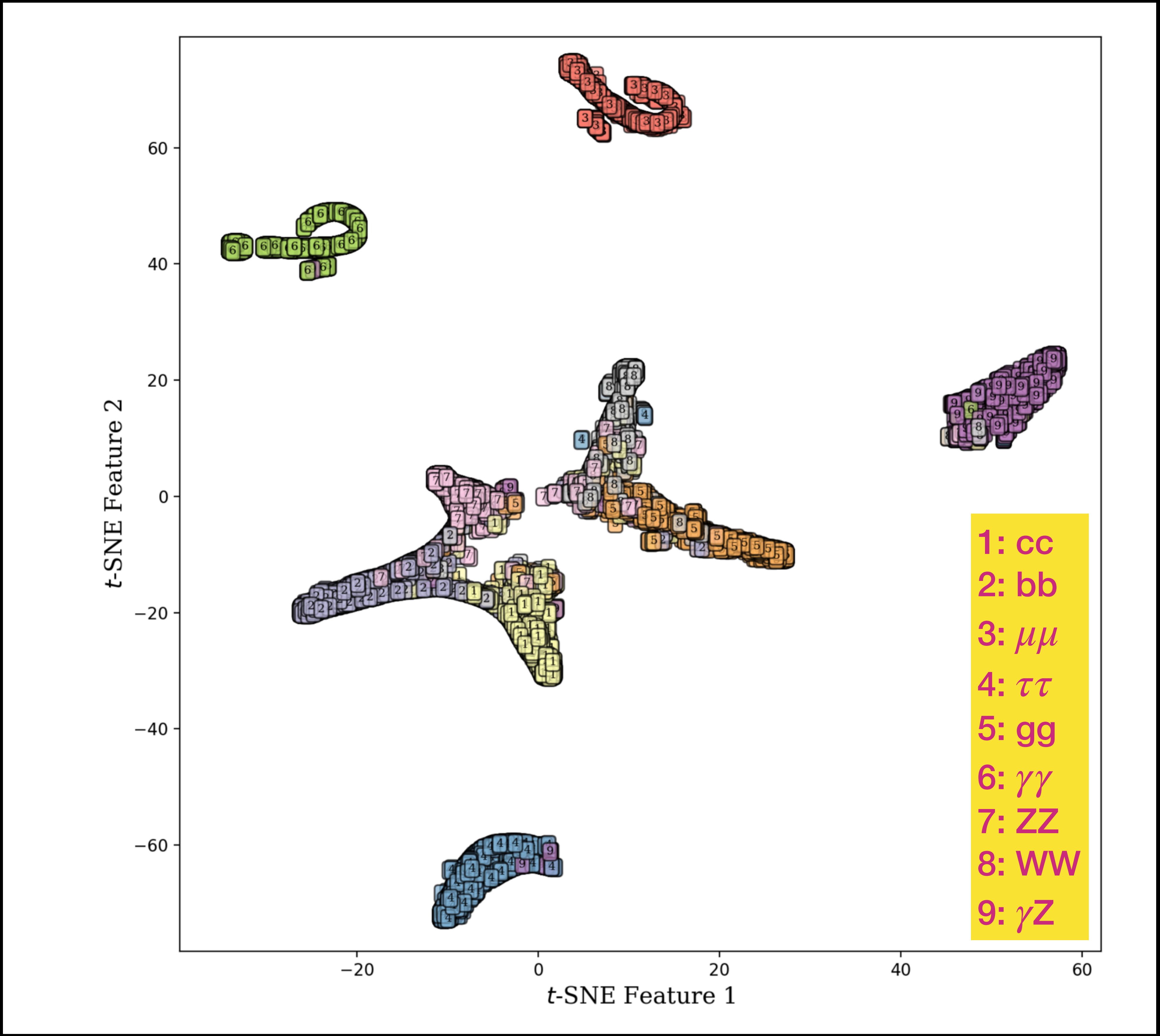
(c)



(d)

Likelihood distribution of predictions
ee example

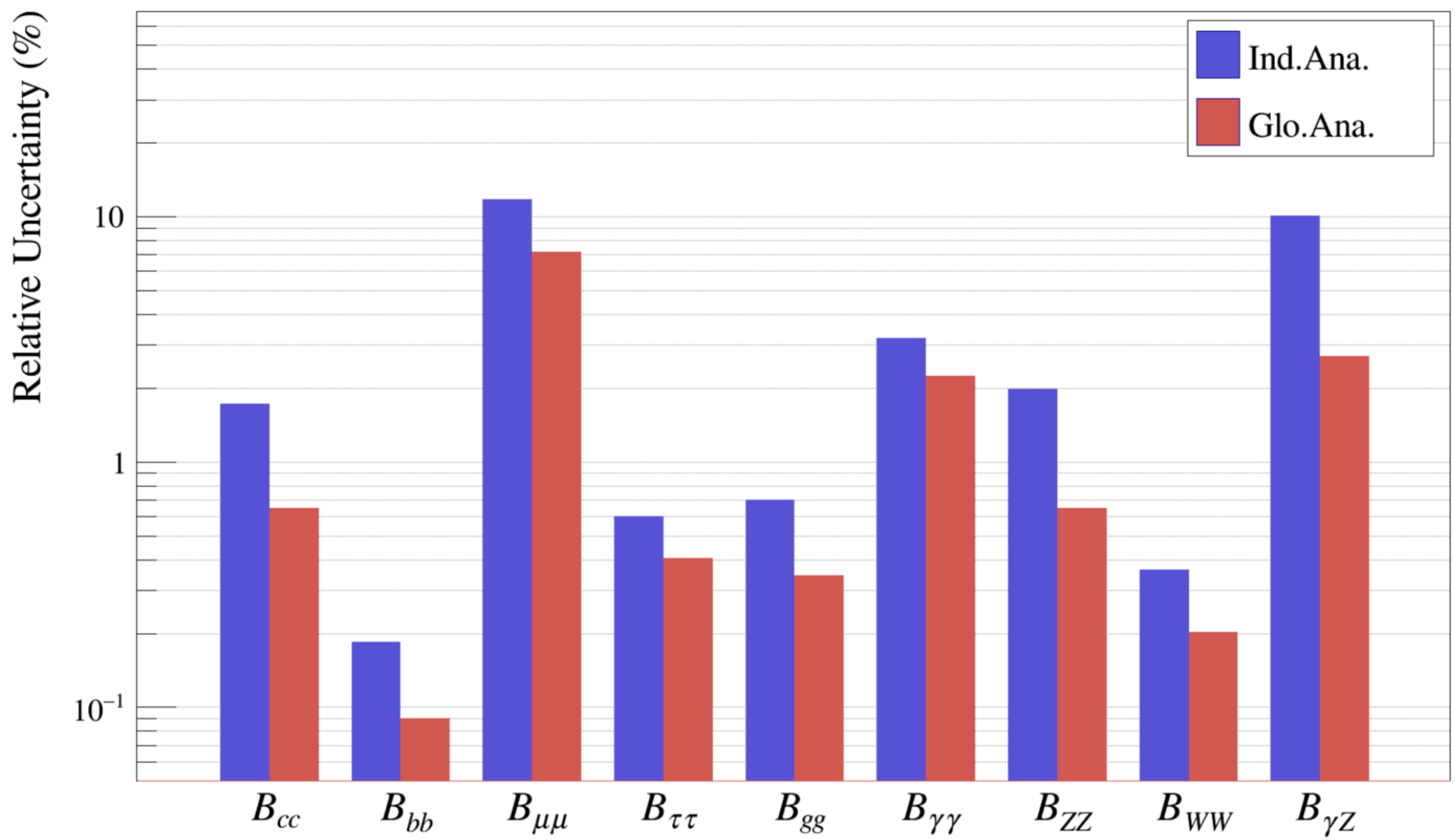




Dimensionality reduction

Maaten&Hinton. Visualizing data using t-SNE. Journal of Machine Learning Research (2008).

Results



- **Improvement about 2**
- **Both have no the SM background**
- **Improved if**
  - **One suffers from background**
    - **cc, ZZ, γZ**
  - **One has large Bs**
    - **bb**
- $\tau\tau$ *and* $\gamma\gamma$ **not significant**
  - **Background low**
  - **B small**

The SM background

- Fair comparison between the individual and global analysis because both have no the SM background.
  - The uncertainties calculated with the generated numbers of events, insensitive to backgrounds.
  - Full efficiency matrix used
- For the individual analysis and CDR: roughly 2 times better. Because
  - The SM background
  - Full simulation
- Conclusion: Improvement understandable. The global analysis could keep its advantages when all the SM background and full simulation are used

Decay Mode	Ind.Ana.	Gl.Ana.	IP
$H \to c\bar{c}$	1.8%	0.65%	2.7
$H \to b\bar{b}$	0.19%	0.09%	2.1
$H \to \mu^+\mu^-$	12%	7.2%	1.7
$H \to \tau^+\tau^-$	0.61%	0.41%	1.4
$H \to gg$	0.7%	0.35%	2.0
$H \to \gamma\gamma$	3.3%	2.3%	1.4
$H \to ZZ$	2.0%	0.65%	3.0
$H \to W^+W^-$	0.37%	0.21%	1.7
$H \to \gamma Z$	11%	2.8%	3.9

Decay Mode	Ind.Ana.	CEPC CDR.	Ratio
$H \to c\bar{c}$	1.8%	3.3%	1.8
$H \to b\bar{b}$	0.19%	0.56%	2.9
$H \to \mu^+\mu^-$	12%	17%	1.4
$H \to \tau^+\tau^-$	0.61%	1.0%	1.6
$H \to gg$	0.7%	1.4%	2.0
$H \to \gamma\gamma$	3.3%	6.9%	2.1
$H \to ZZ$	2.0%	5.1%	2.6
$H \to W^+W^-$	0.37%	1.1%	3.0
$H \to \gamma Z$	11%	15%	1.4

Summary and plan

- A proof-of-principle study shows
  - Global analysis enhances the precision by a factor of ~2
  - Confusion matrix as a metric for detector optimization — single-purpose
- ArXiv: 2105.14997, submitted to JHEP
- Plan
  - ✓ Verify background effect
  - ✓ Factorize the impacts of detector performances
  - ✓ Setup a framework for detector optimization
  - ✓ Move to full simulation of CEPCSW

The end

Thanks a lot

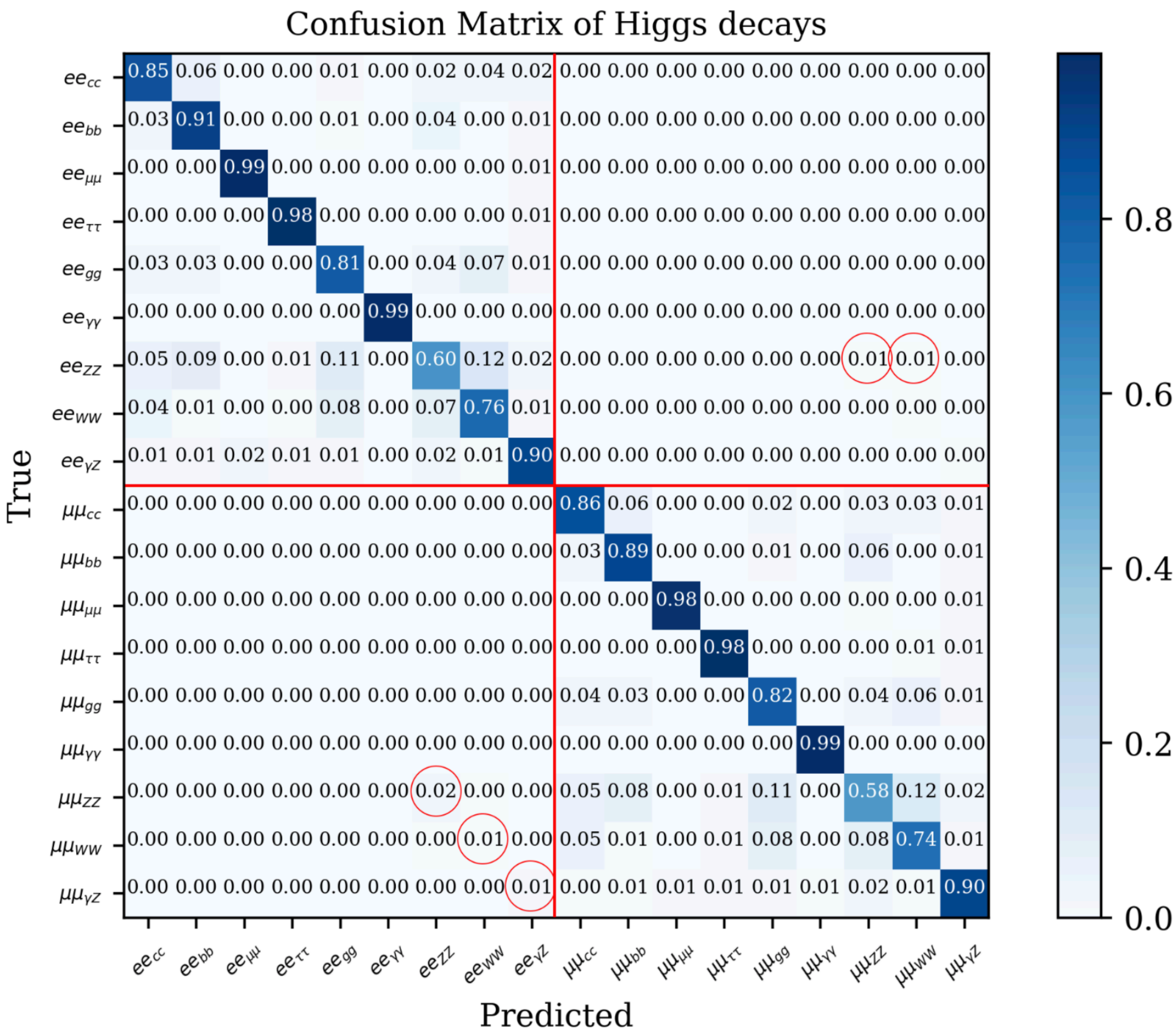


Figure 3. PFN_CONF $ee + \mu\mu$

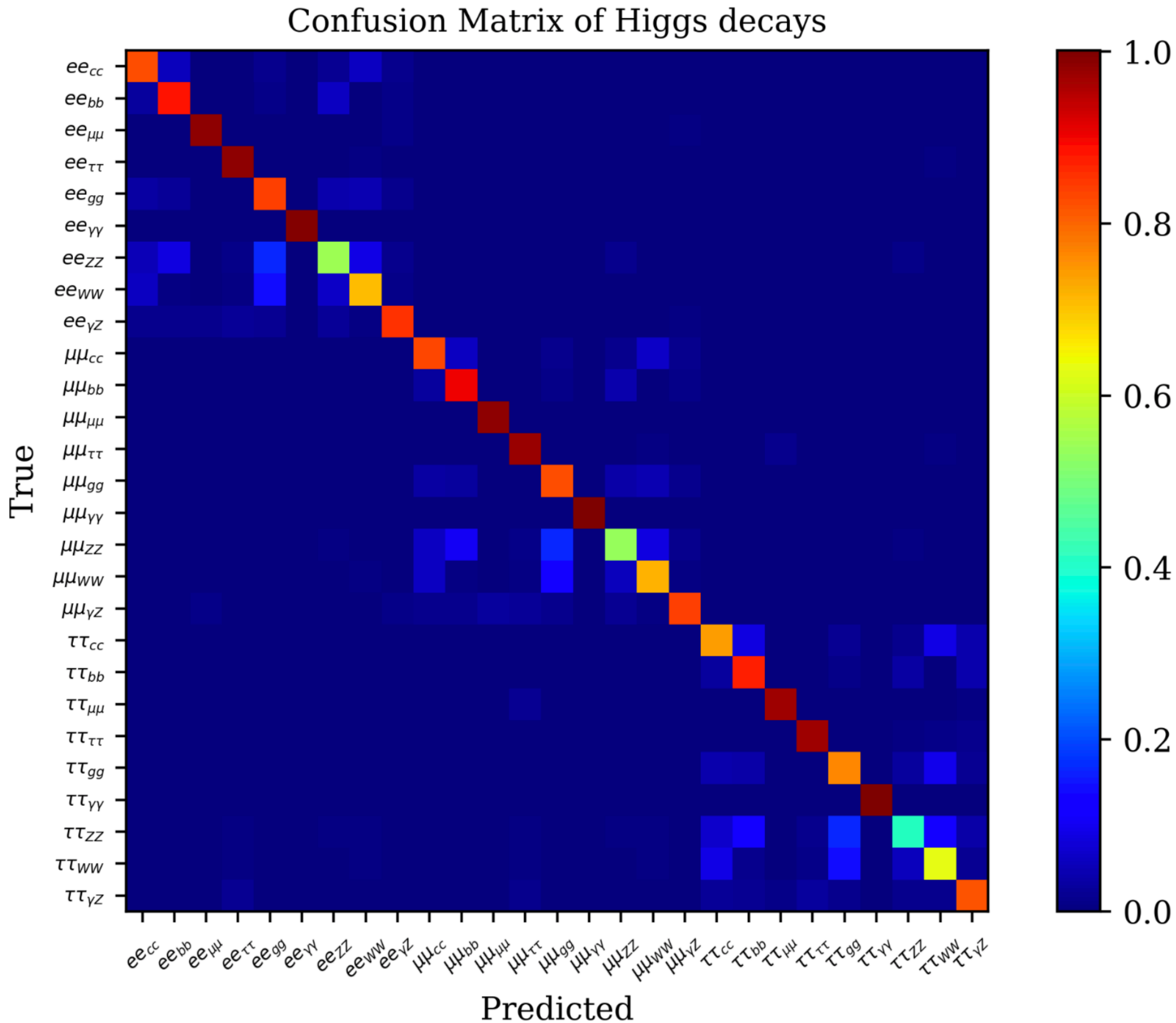


Figure 7. PFN_CONF $ee + \mu\mu + \tau\tau$

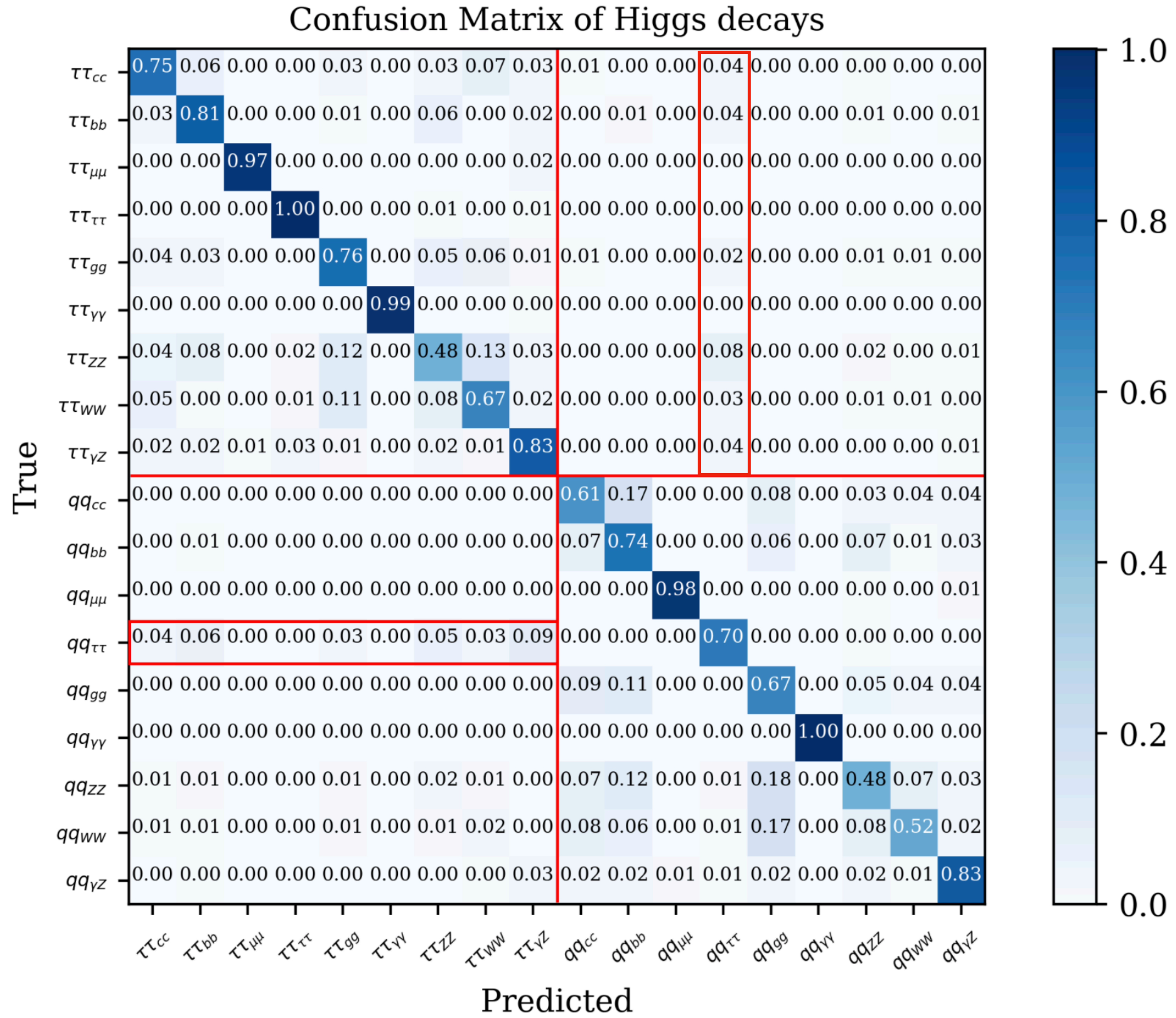


Figure 5. PFN_CONF $\tau\tau + qq$

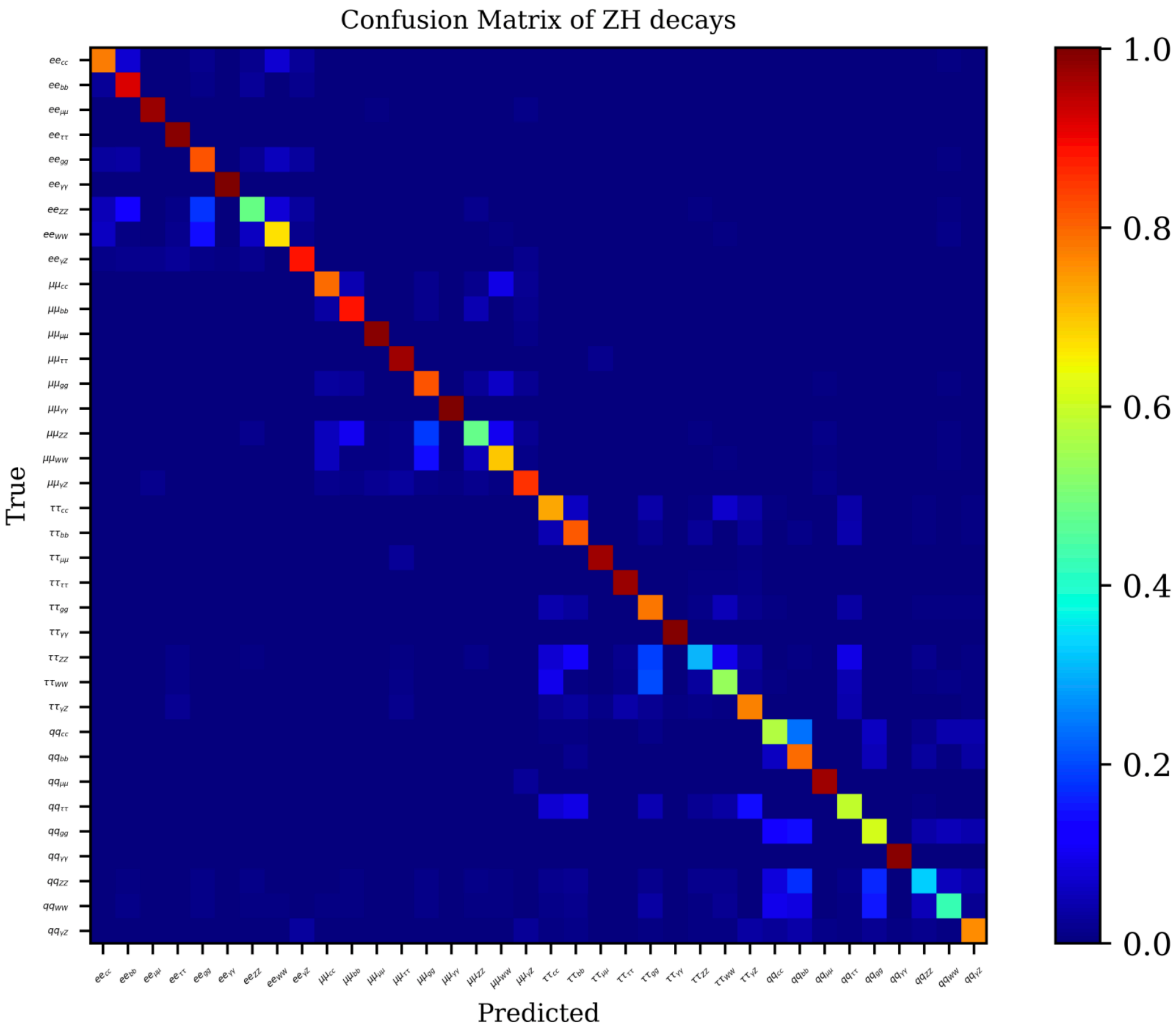


Figure 9. PFN_CONF $ee + \mu\mu + \tau\tau + qq$