Domain Adversarial Learning for Reducing Training Bias in ttH(bb) search at ATLAS

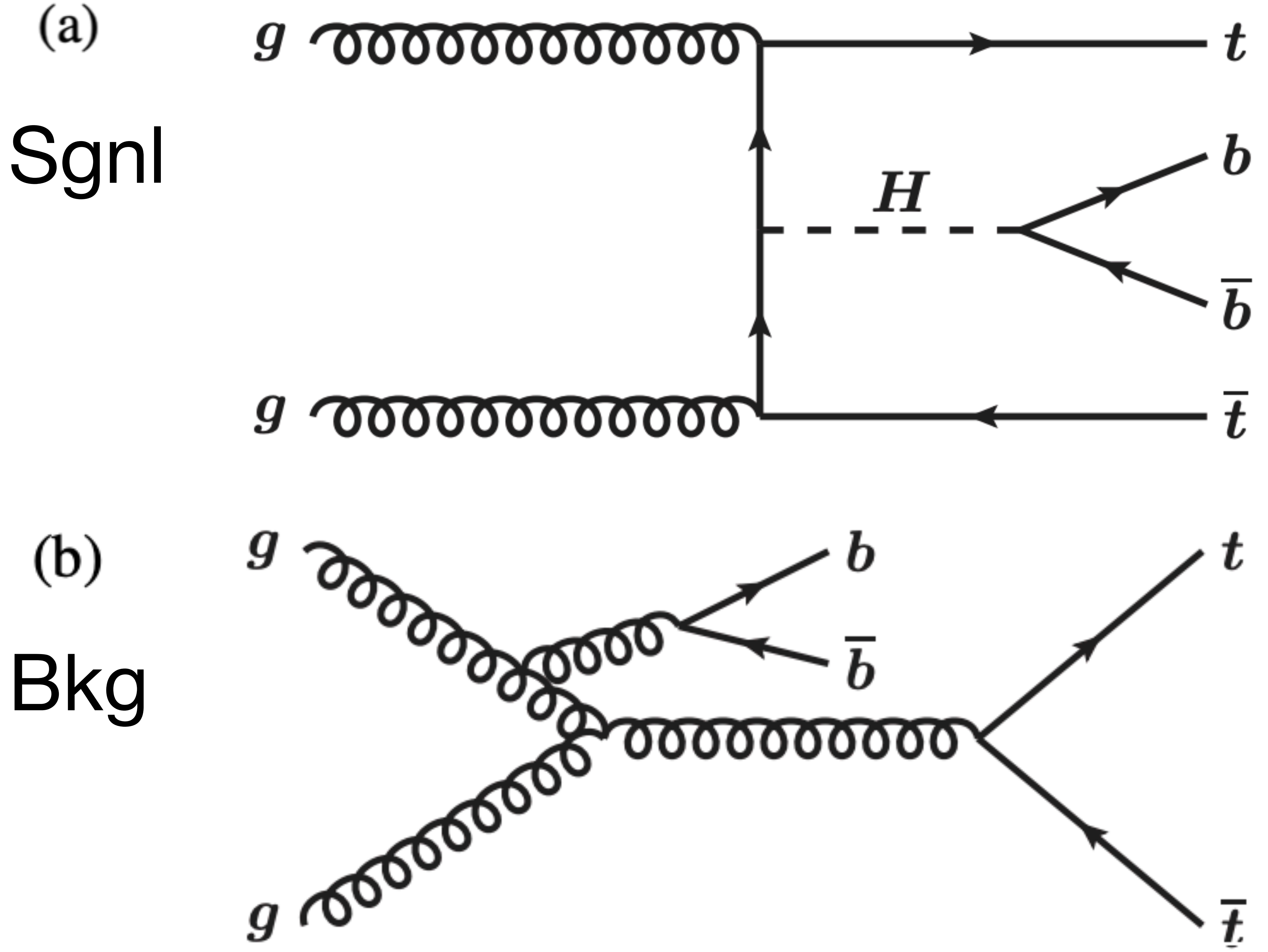
Fatkhullin I. F.
Supervisors: Glaysher P., Katzy J., Pollard C.

ATLAS, DESY
2019

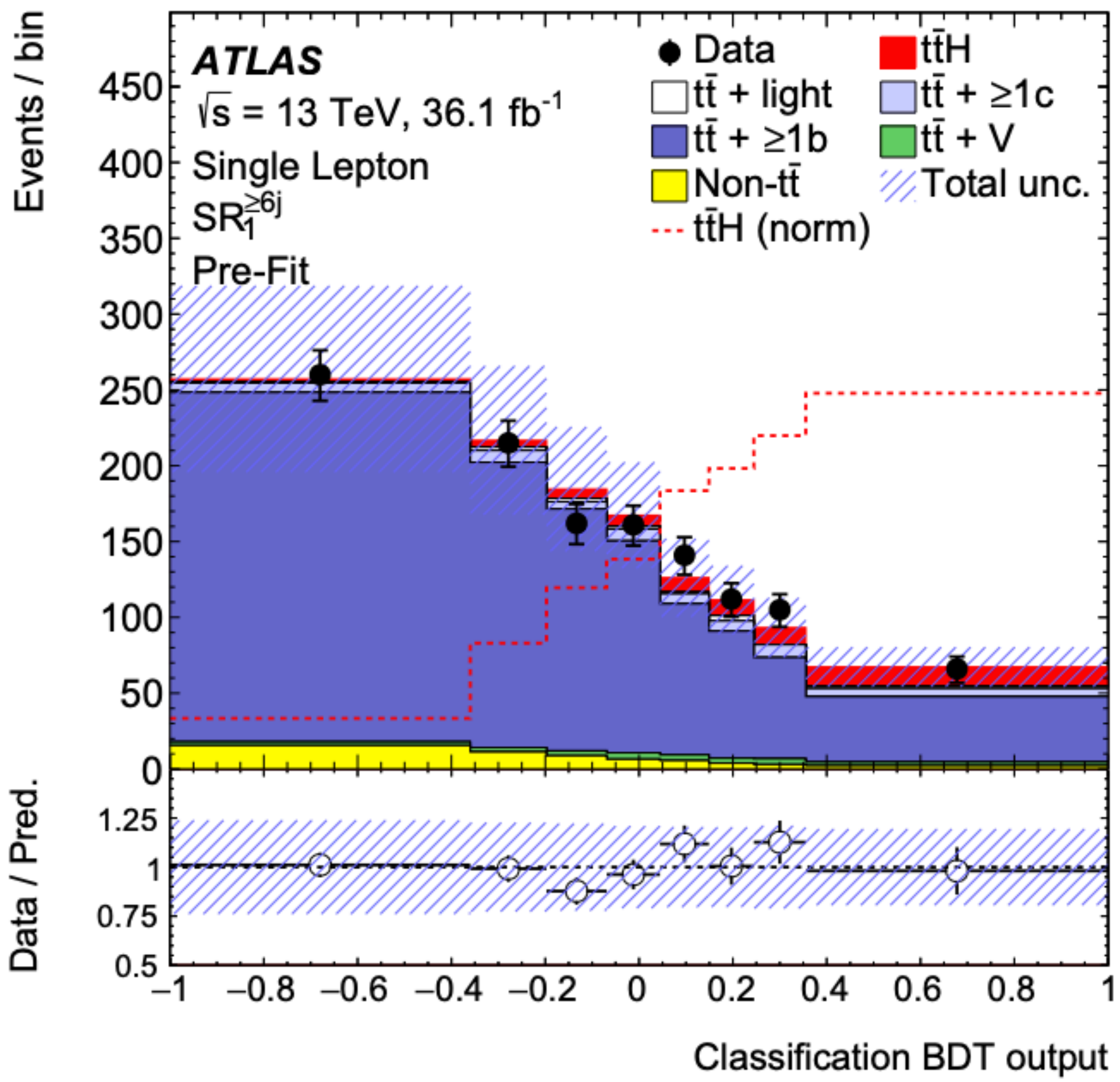
Outline

- Introduction
- Classification of ttH(bb) vs ttbb bkg
- Idea of Adversarial Domain Adaptation
- Results

ttH(bb) Production Search



Feynman diagrams for a) ttH(bb) production b) tt+bb background

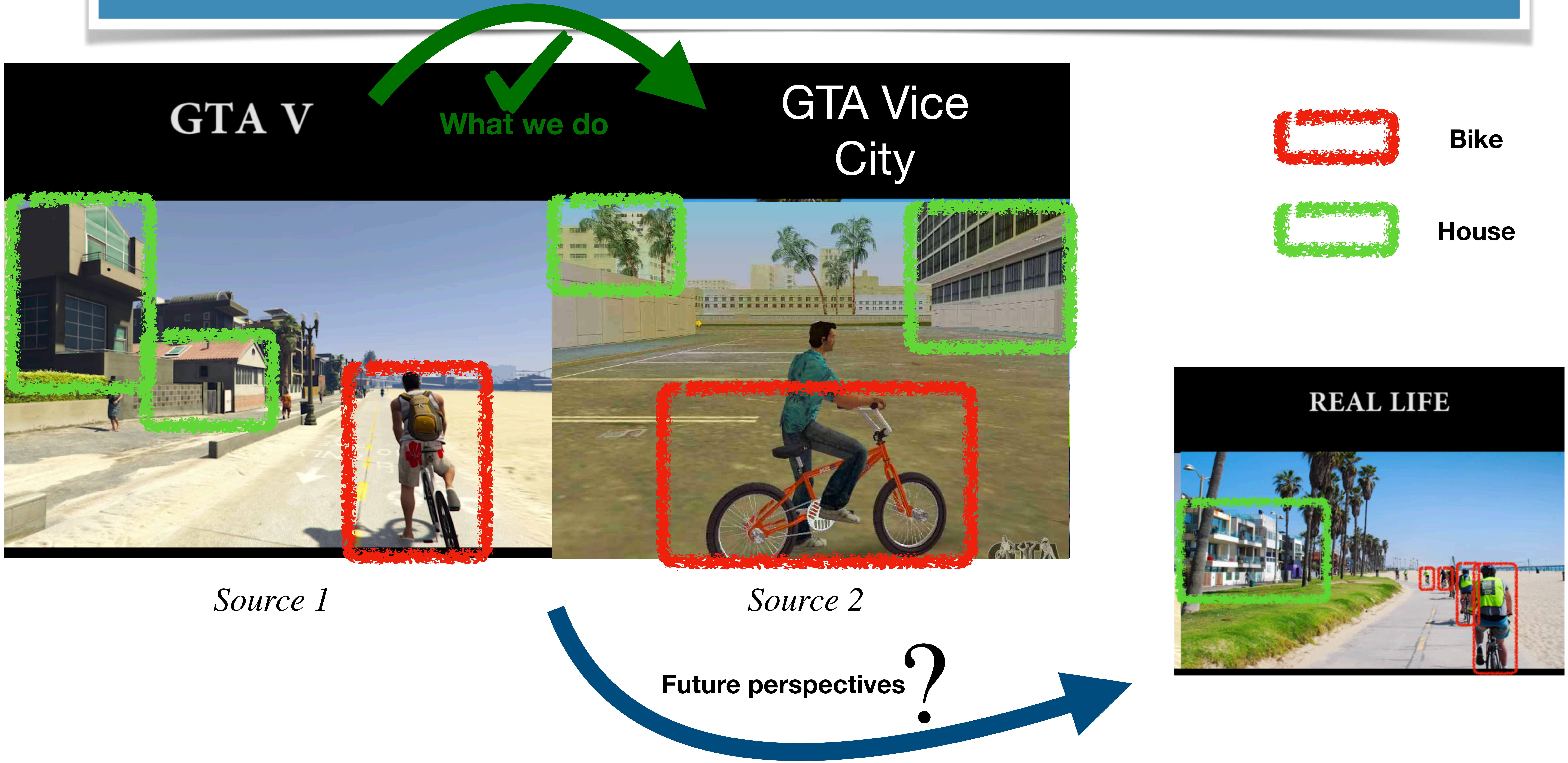


- *Systematic uncertainties come because of different simulations*
- *The goal is to minimise this uncertainty while preserving good classification performance*

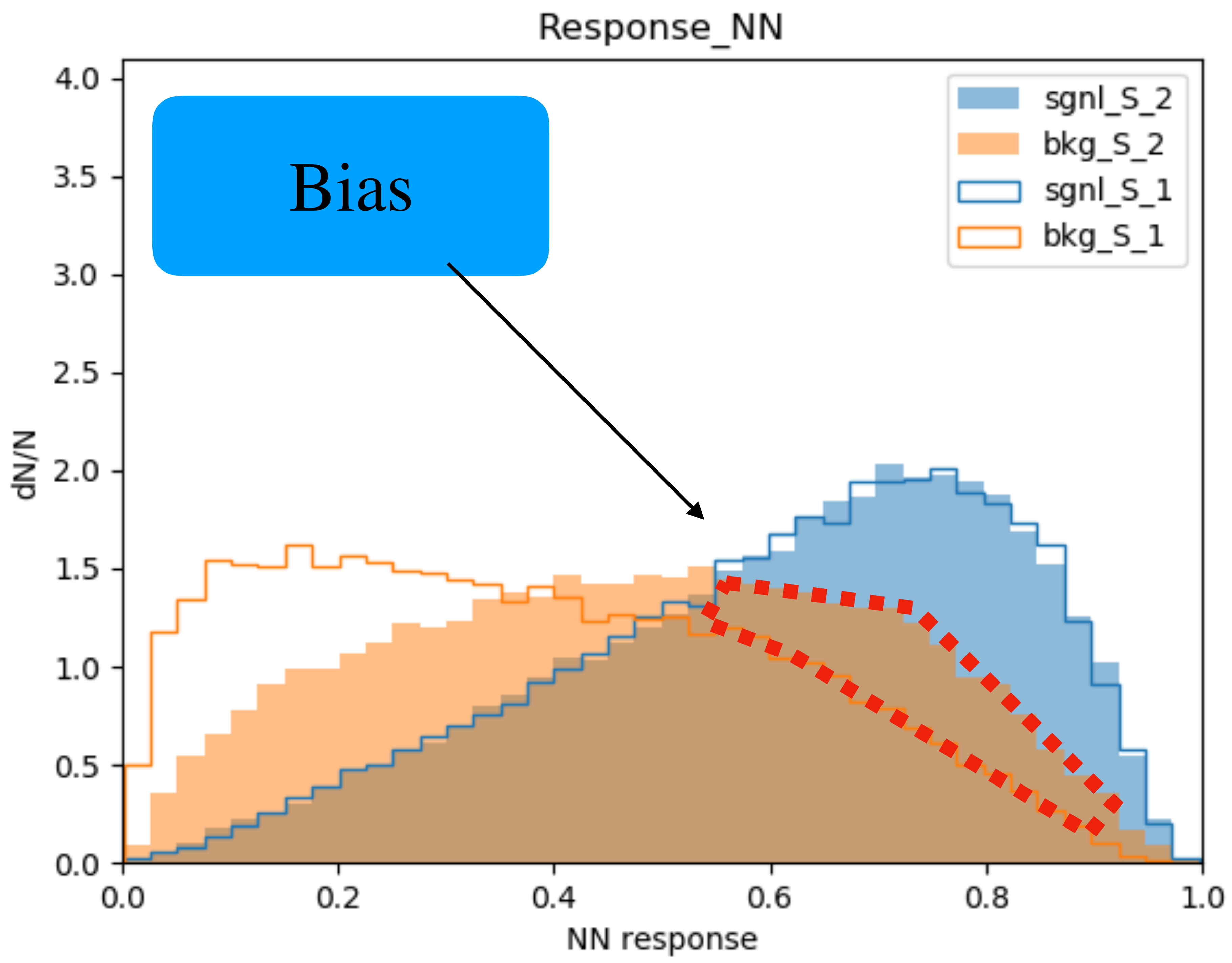
Domain Adaptation

Source 1 *Source 2*

Domain Adaptation



Motivation

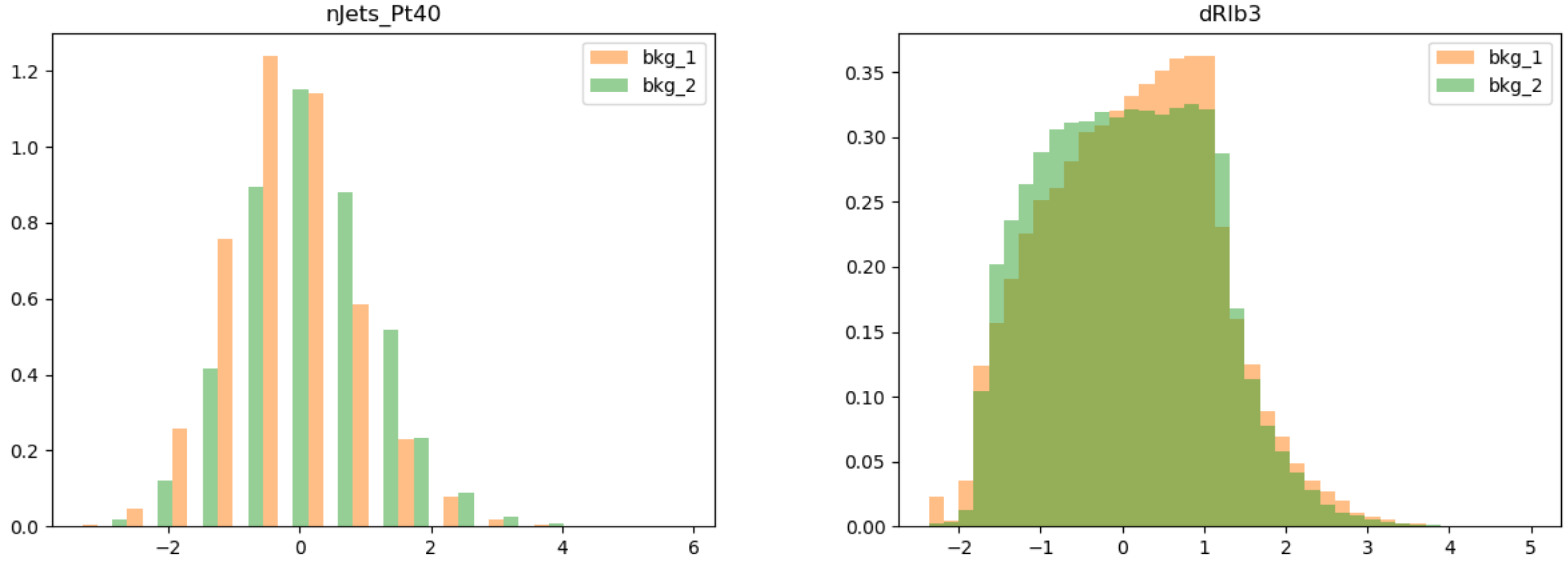


Performance of baseline Feed Forward NN with 3 hidden layers

- We have one signal and two different background simulations
- It's been noted that if we train NN classifier on data with one background simulation and apply it to the test data with another background, its performance is reduced significantly
- This happens because backgrounds are slightly different
- Domain adaptation techniques [1] are hoped to reduce the training bias of NN and show good performance

[1] *A theory of learning from different domains, Shai Ben-David et.al., 2010*

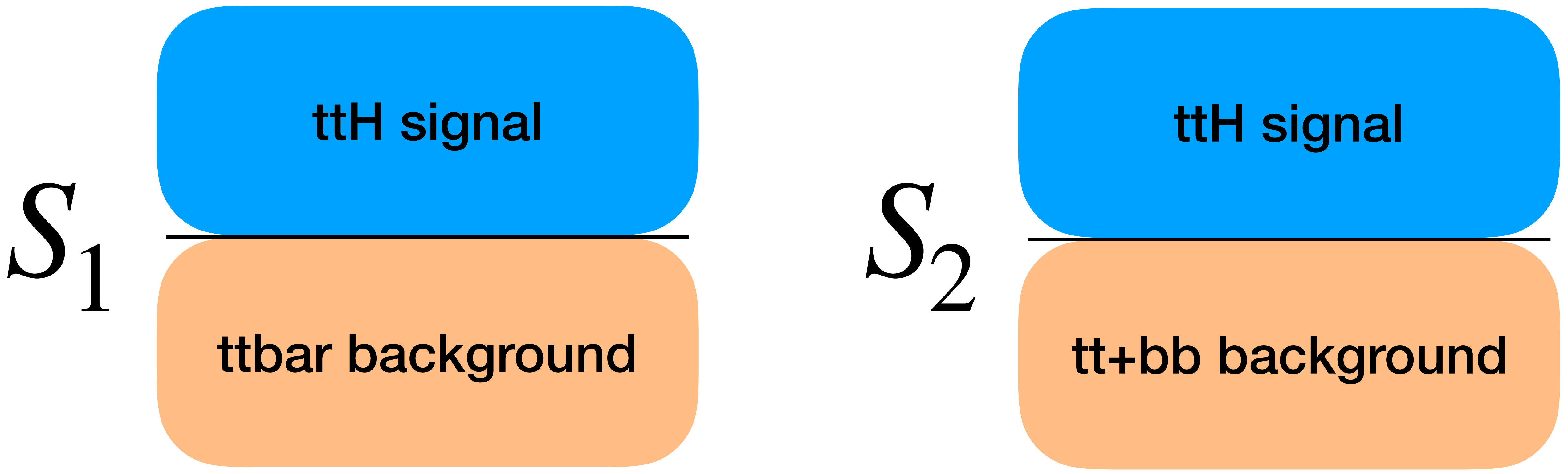
Background distributions



- *bkg_1 = ttbar background: MadGraph / Pythia6*
- *bkg_2 = tt+bb background: Powheg Pythia8*
- *Two background simulations are similar in distribution shapes, but there is some difference.*

Notations

- $\mathscr{X}$ an input space (40 variables)
- $\mathscr{Y} = \{0,1\}$ a set of classes (if an event comes from <u>signal</u> or <u>background</u>)
- We define <u>*domain*</u> as a pair $< \mathscr{D}, f >$ consisting of distribution $\mathscr{D}$ on $\mathscr{X}$ and a labelling function f on $\mathscr{X} \times \mathscr{Y}$.
- S_1 and S_2 denote two datasets obtained from domains <u>Source 1</u> $< \mathscr{D}_{S_1}, f_{S_1} >$ and <u>Source 2</u> $< \mathscr{D}_{S_1}, f_{S_1} >$

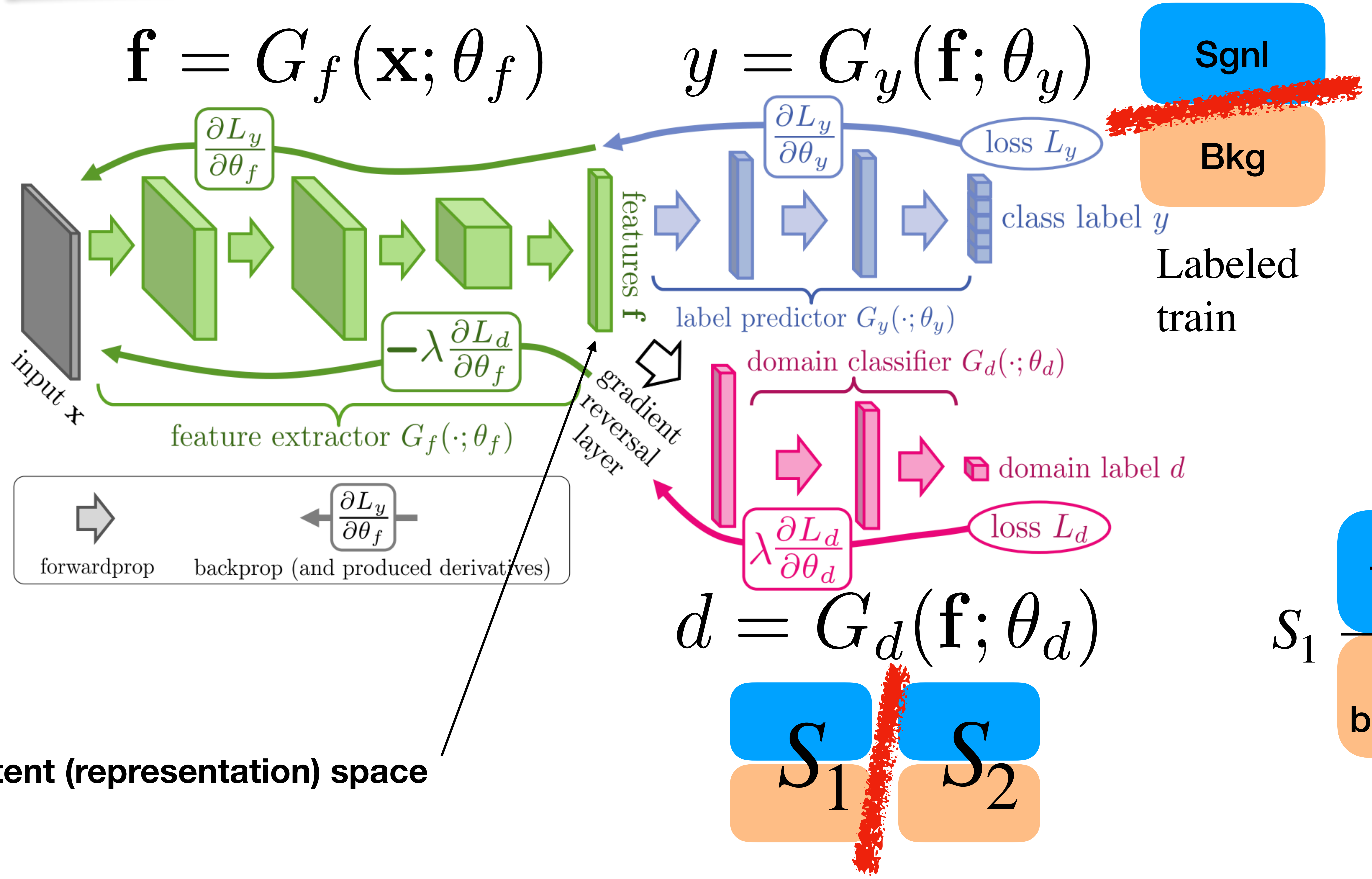


Source 1 *Source 2*

- ttH signal: MadGraph/Herwig6
- ttbar background: MadGraph/Pythia6
- tt+bb background: Powheg Pythia8

"Domain-Adversarial Training of Neural Networks" by Y. Ganin et. al. 2016

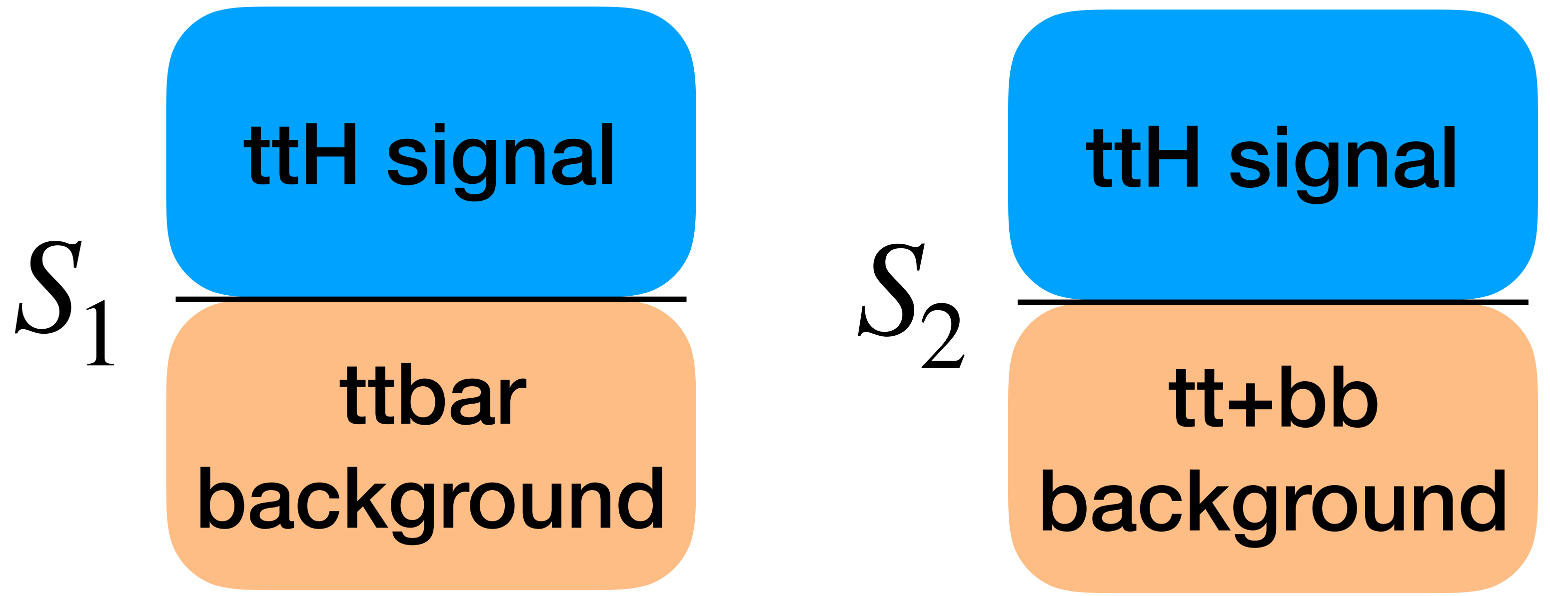
Main Idea



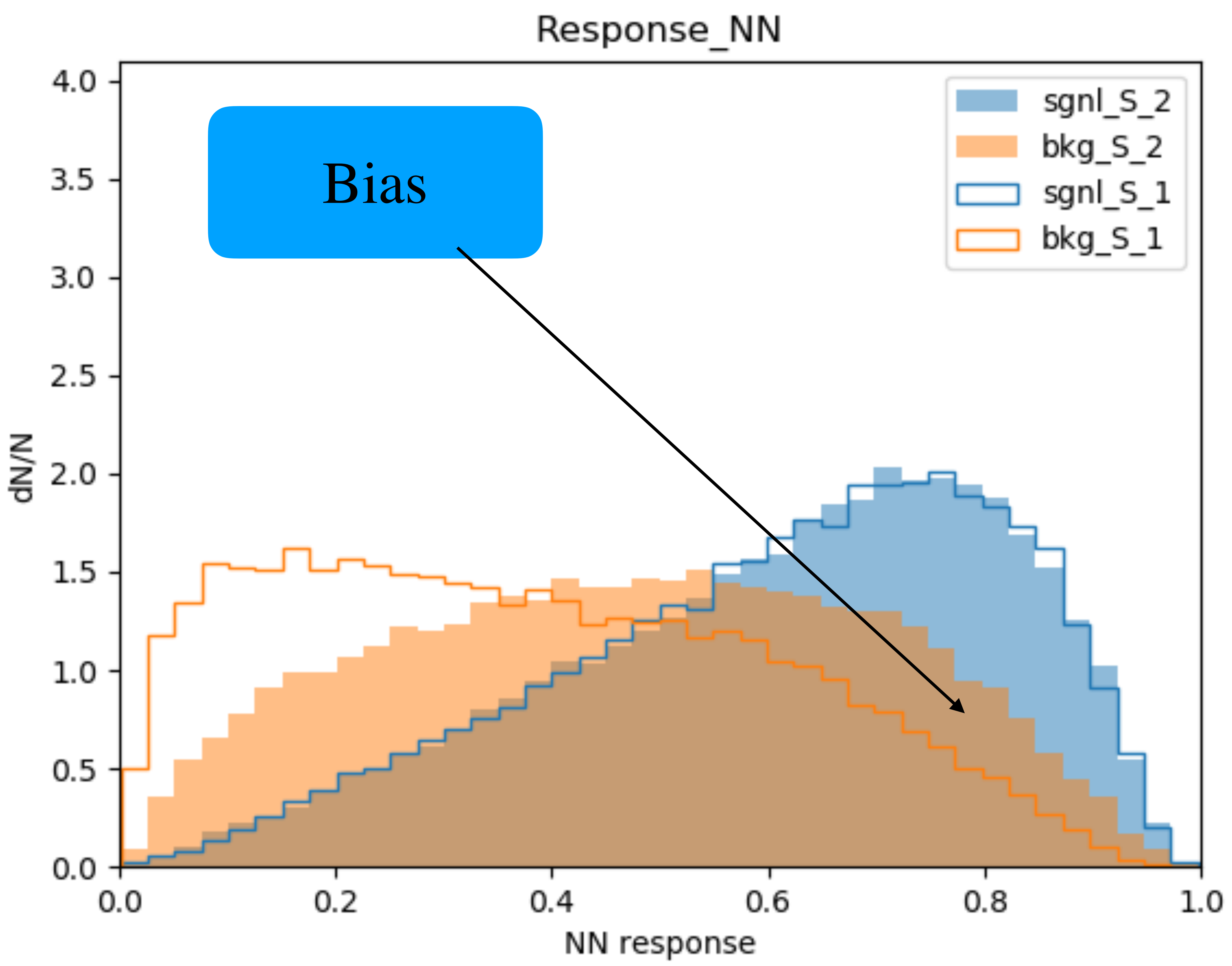
Latent (representation) space

Unlabeled train

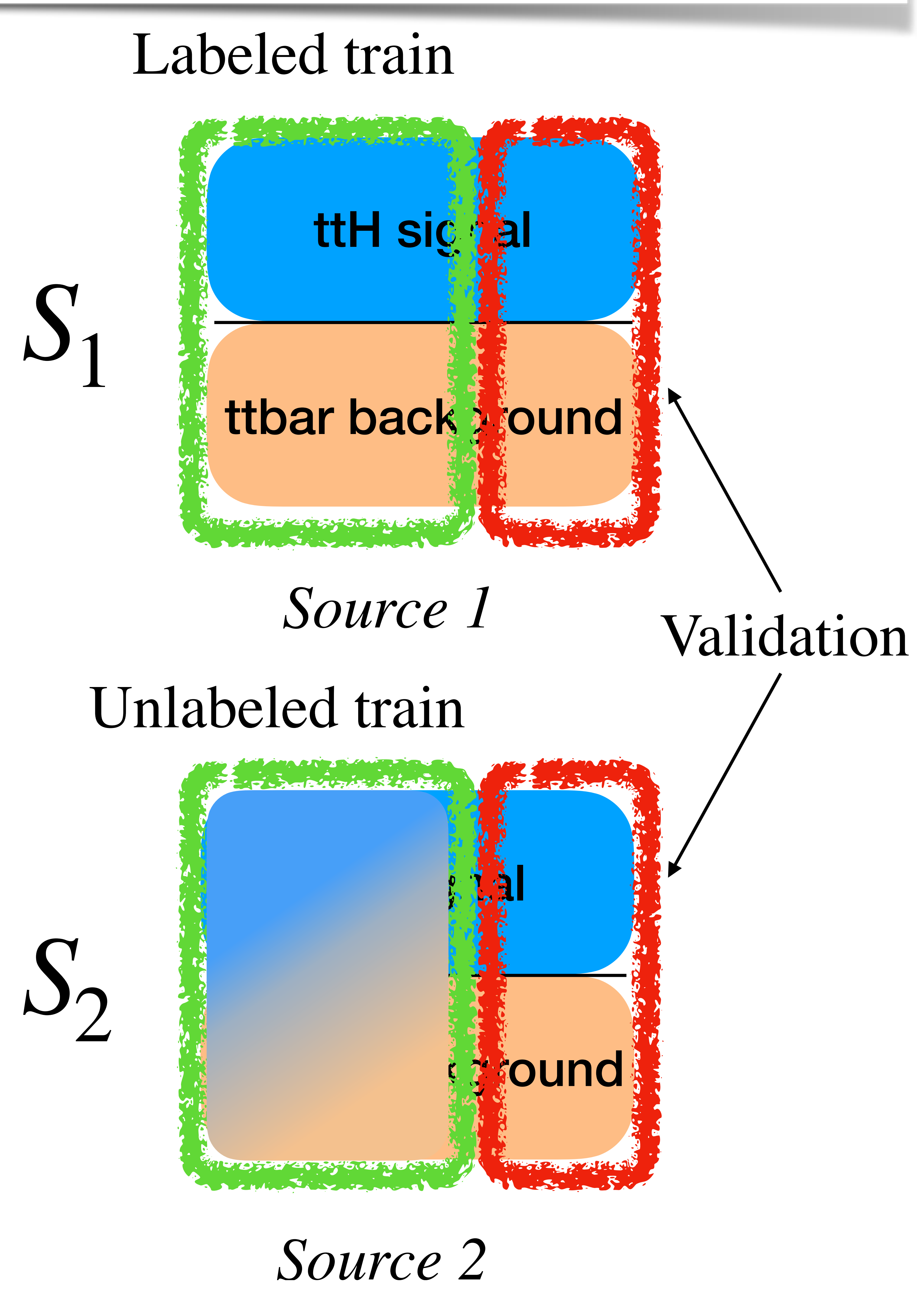
- Domain discriminator with gradient reversal layer is introduced
- Discriminator impacts on feature extractor so that it makes different source instances indistinguishable in the representation space



Results

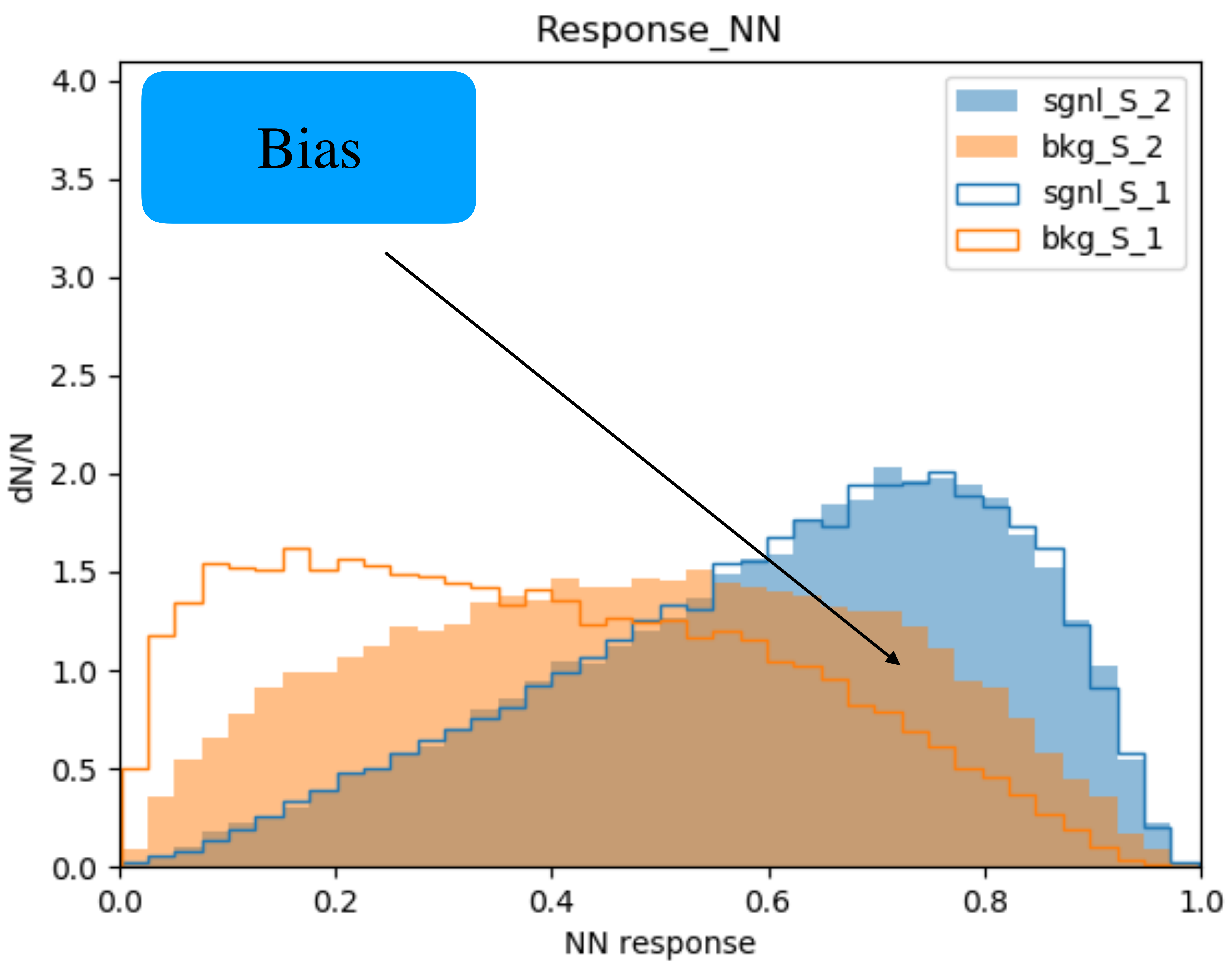


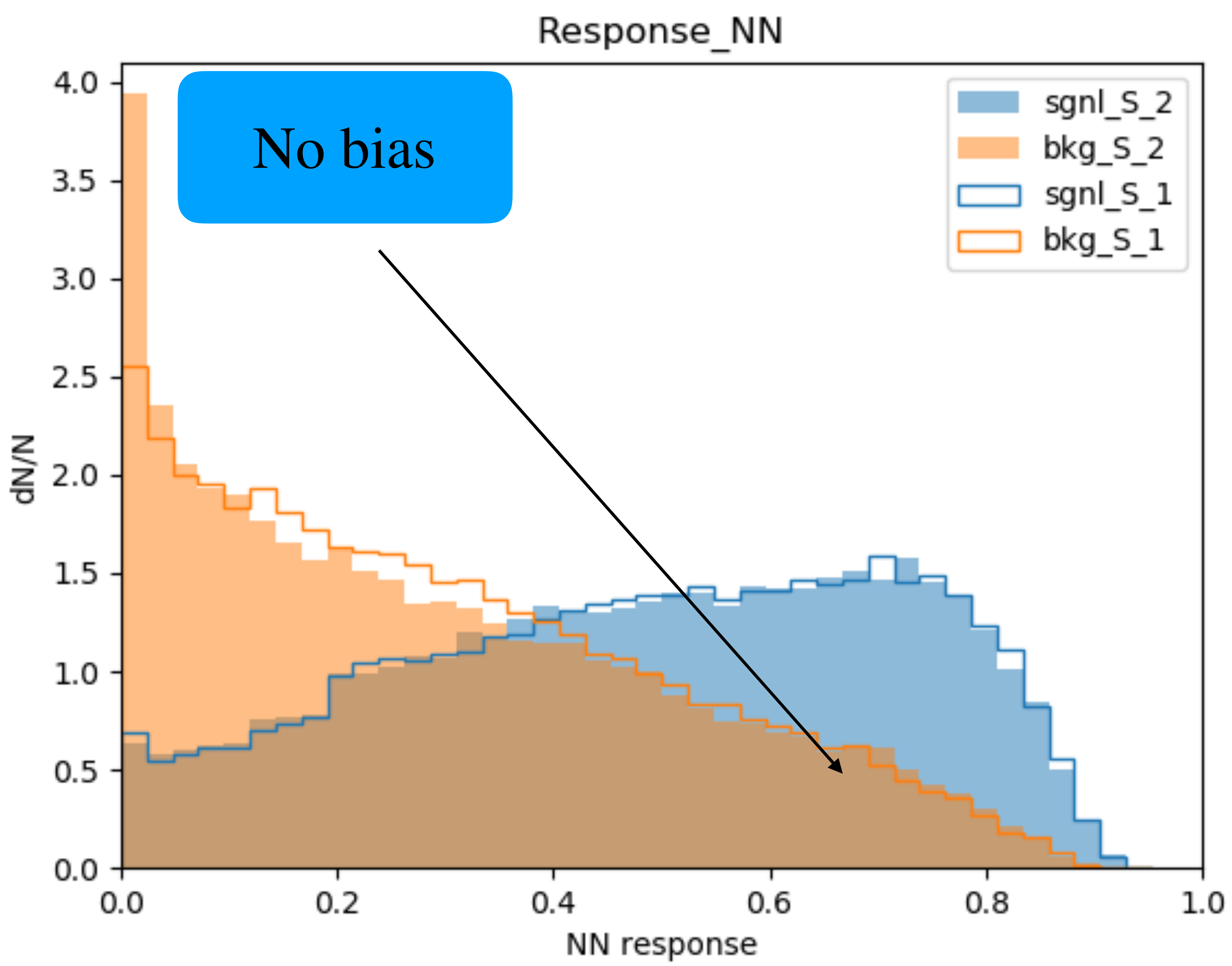
Response_NN
Bias
sgnl_S_2
bkg_S_2
sgnl_S_1
bkg_S_1
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
dN/N
0.0
0.2
0.4
0.6
0.8
1.0
NN response



Labeled train
S_1
ttH signal
ttbar background
Source 1
Validation
Unlabeled train
S_2
Source 2

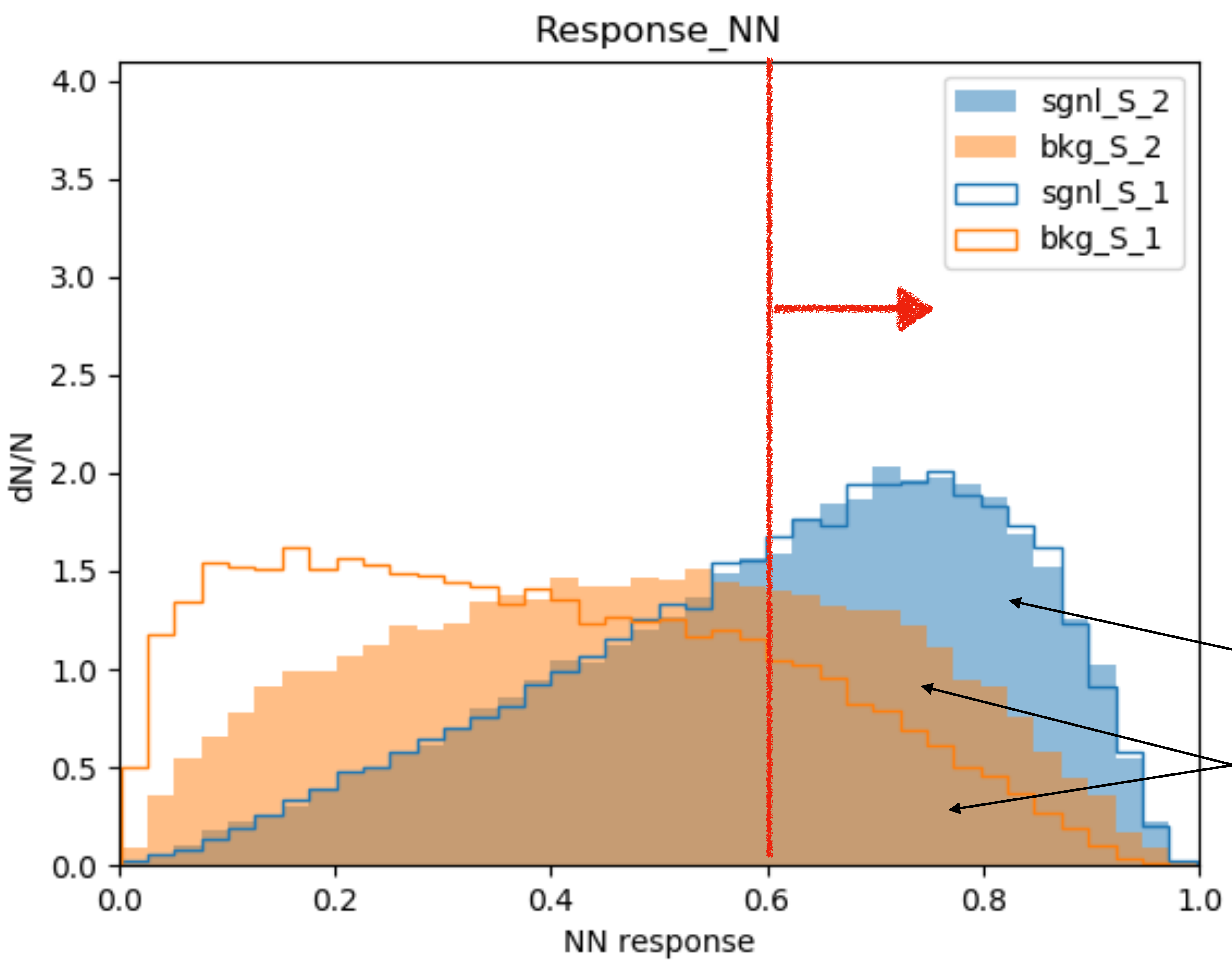
Results





- Discriminator importance parameter $\lambda = 0$
- Discriminator importance parameter $\lambda = 1.3$

Significance Calculation

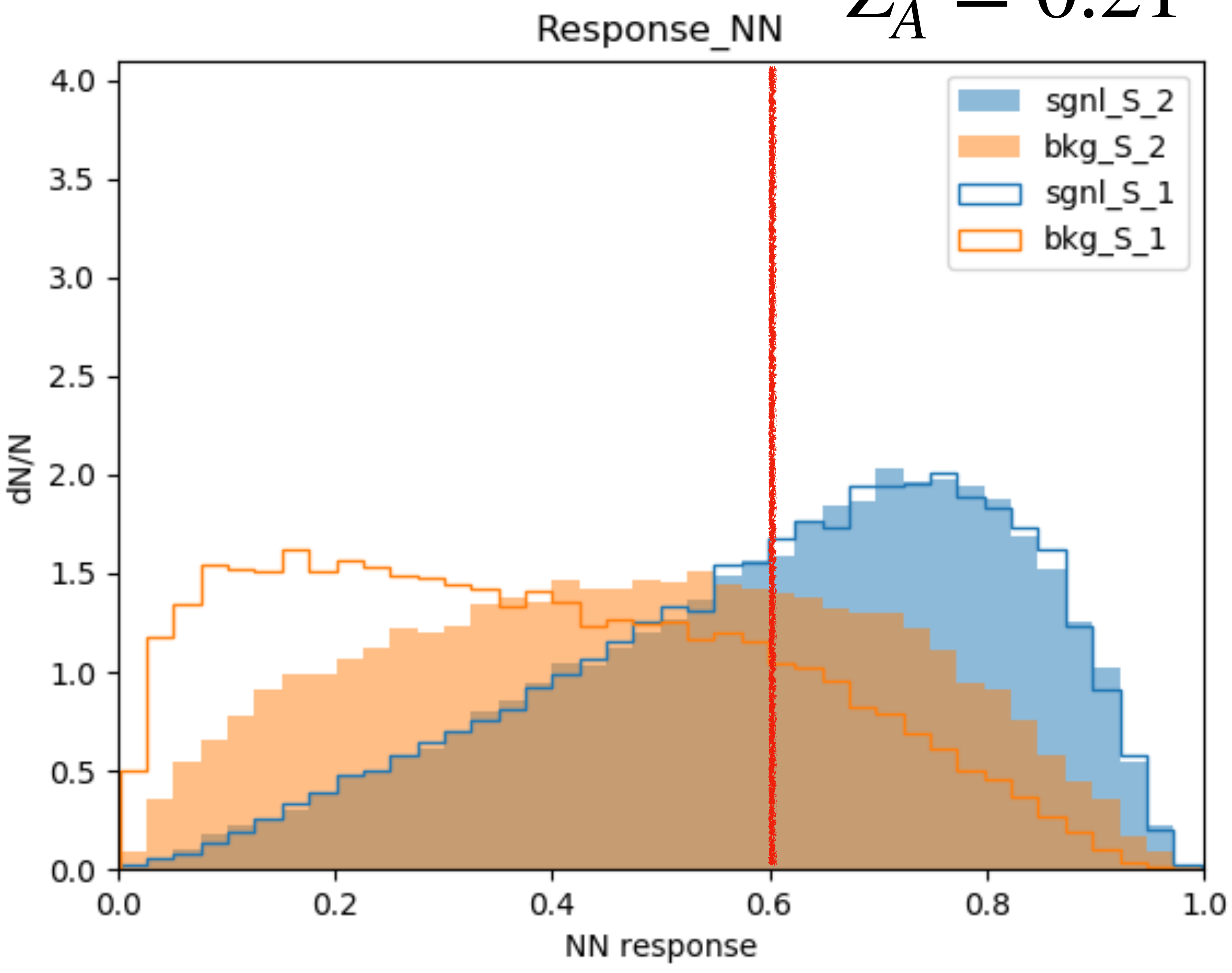


$$Z_A = \frac{s}{\sqrt{b + \sigma_b^2}}$$

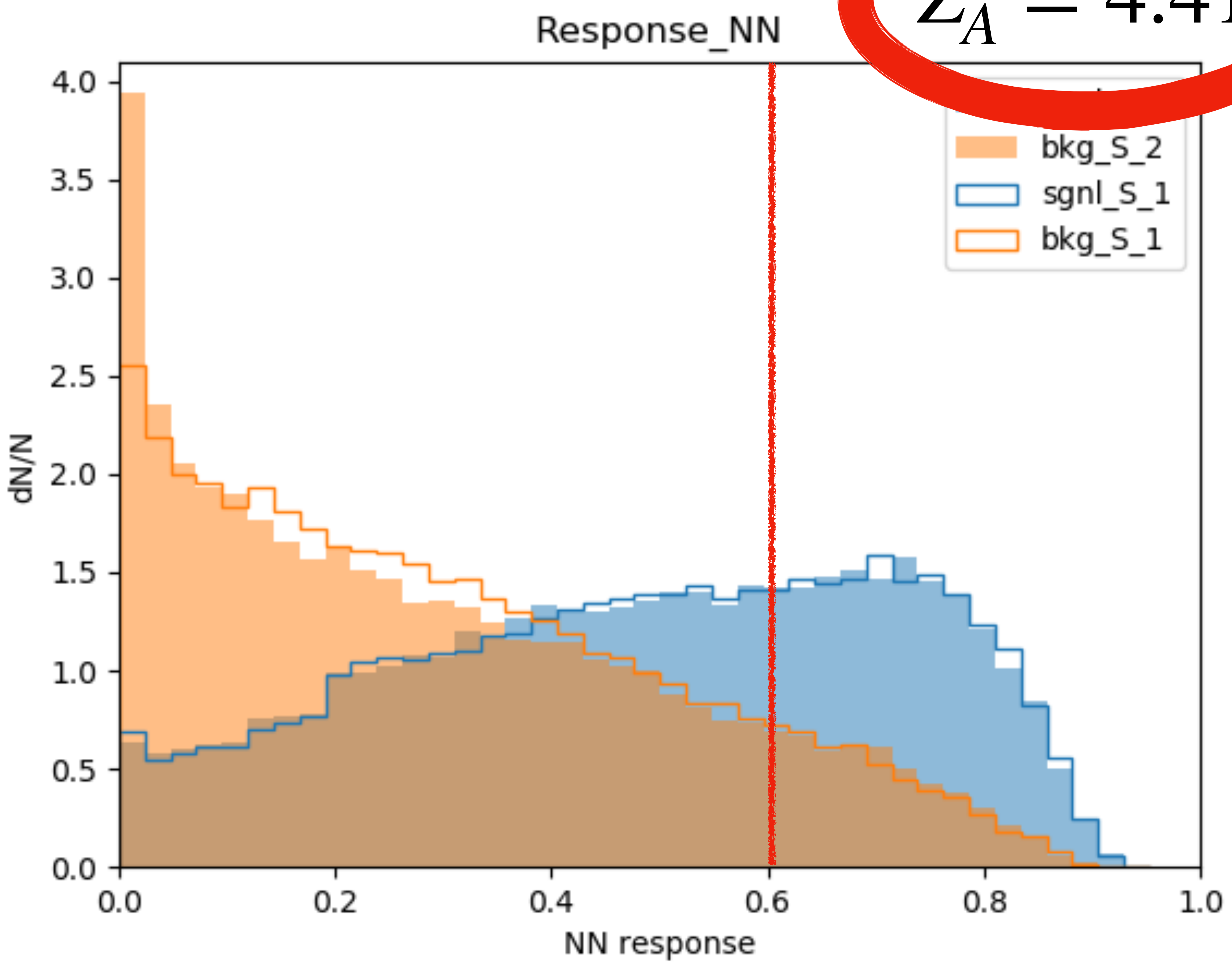
- s - signal events in the cut
- b - test bkg events in the cut
- σ_b^2 - difference btw test and train bkgs

Significance and Response

$Z_A = 0.21$



$Z_A = 4.41$

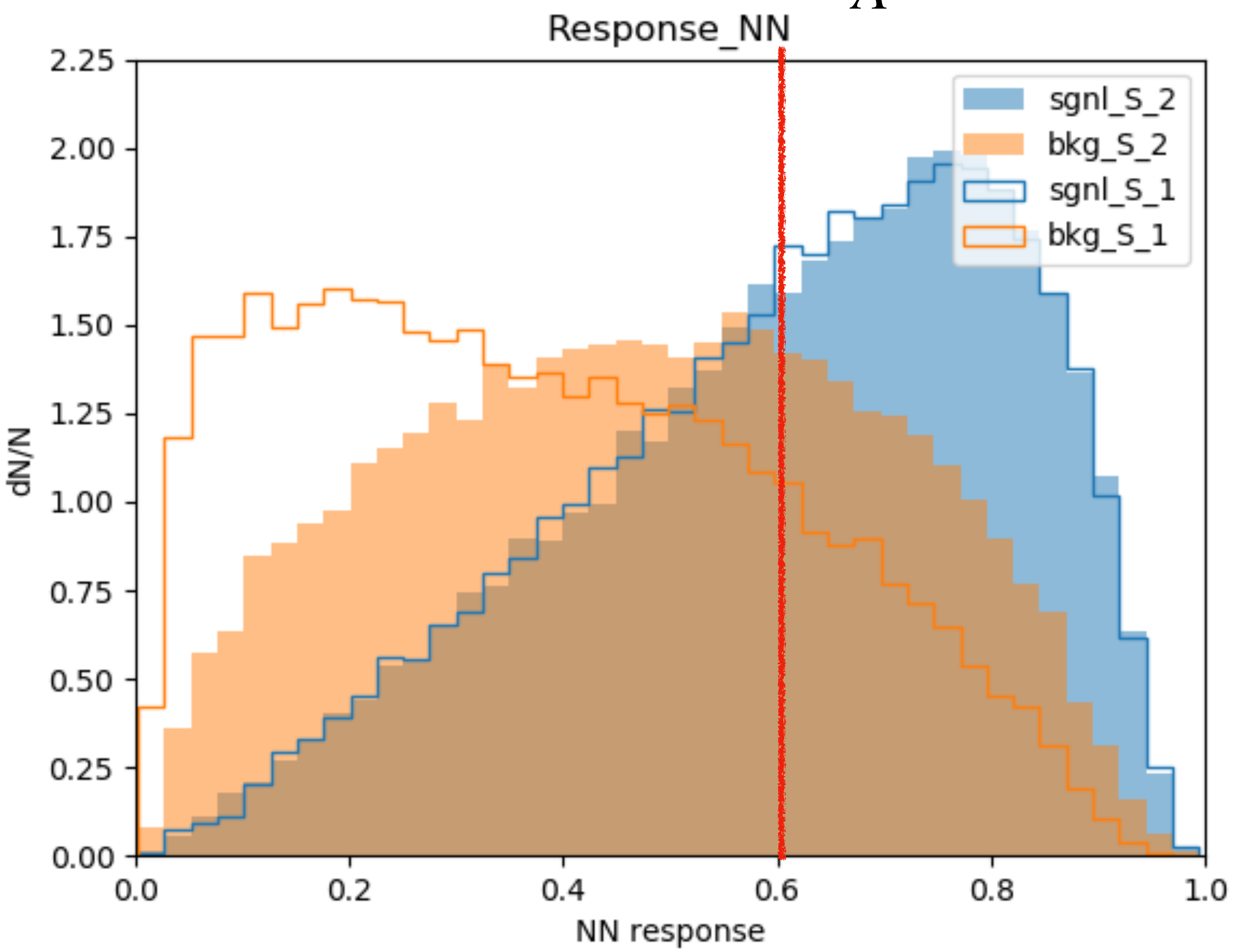


- Discriminator importance parameter $\lambda = 0$
- Discriminator importance parameter $\lambda = 1.3$

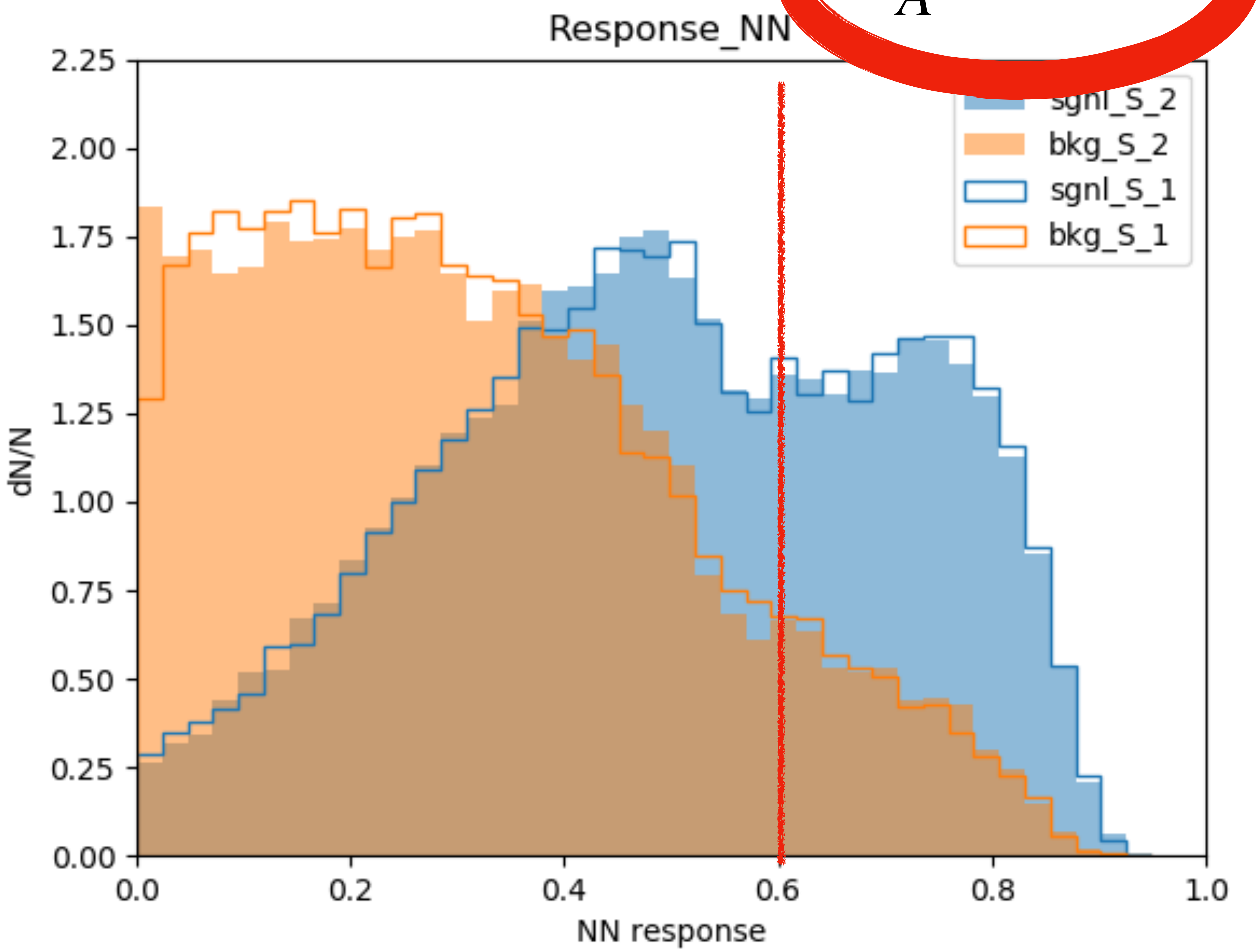
Testing on different NN settings

Significance and Response

$Z_A = 0.21$

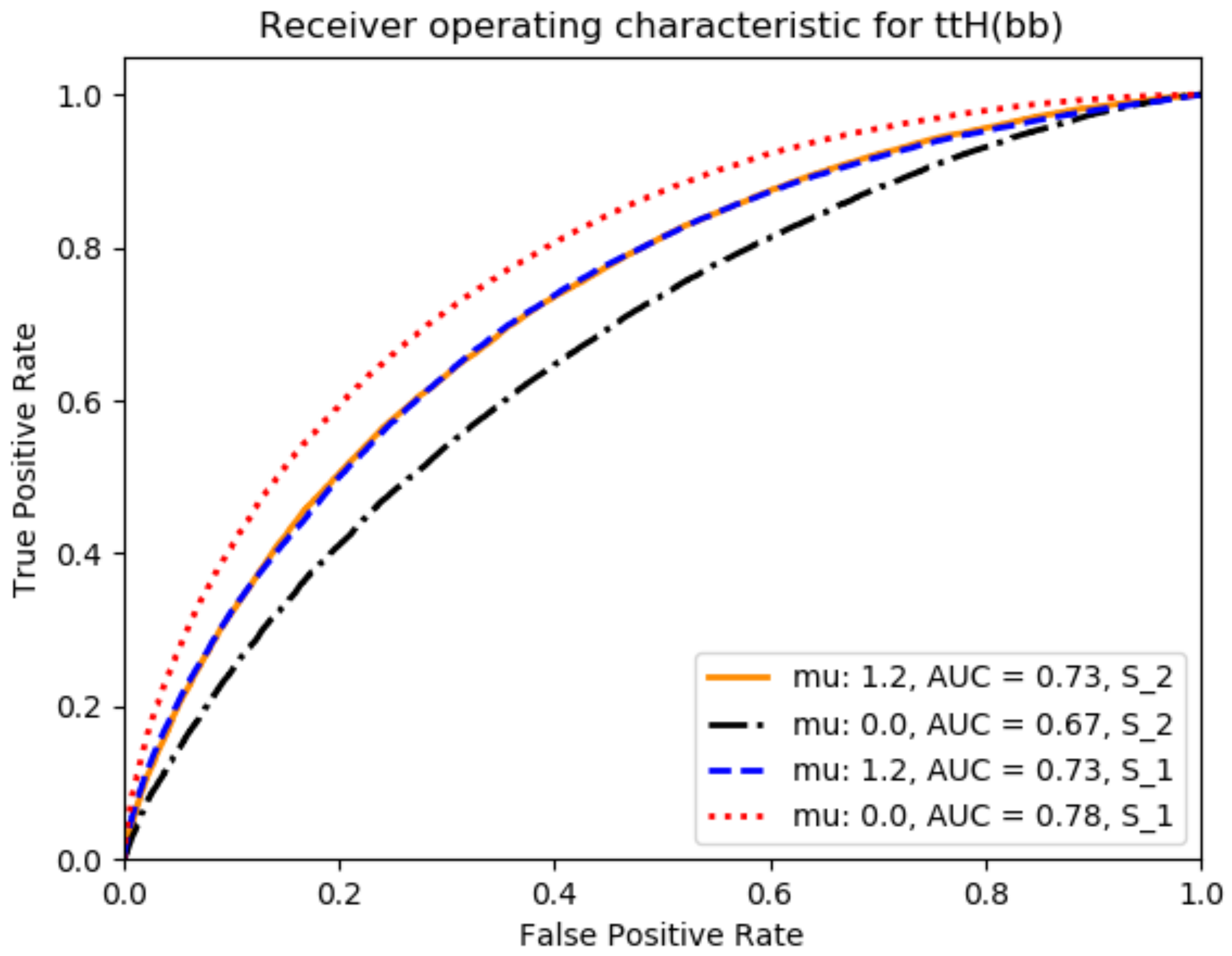


$Z_A = 6.42$

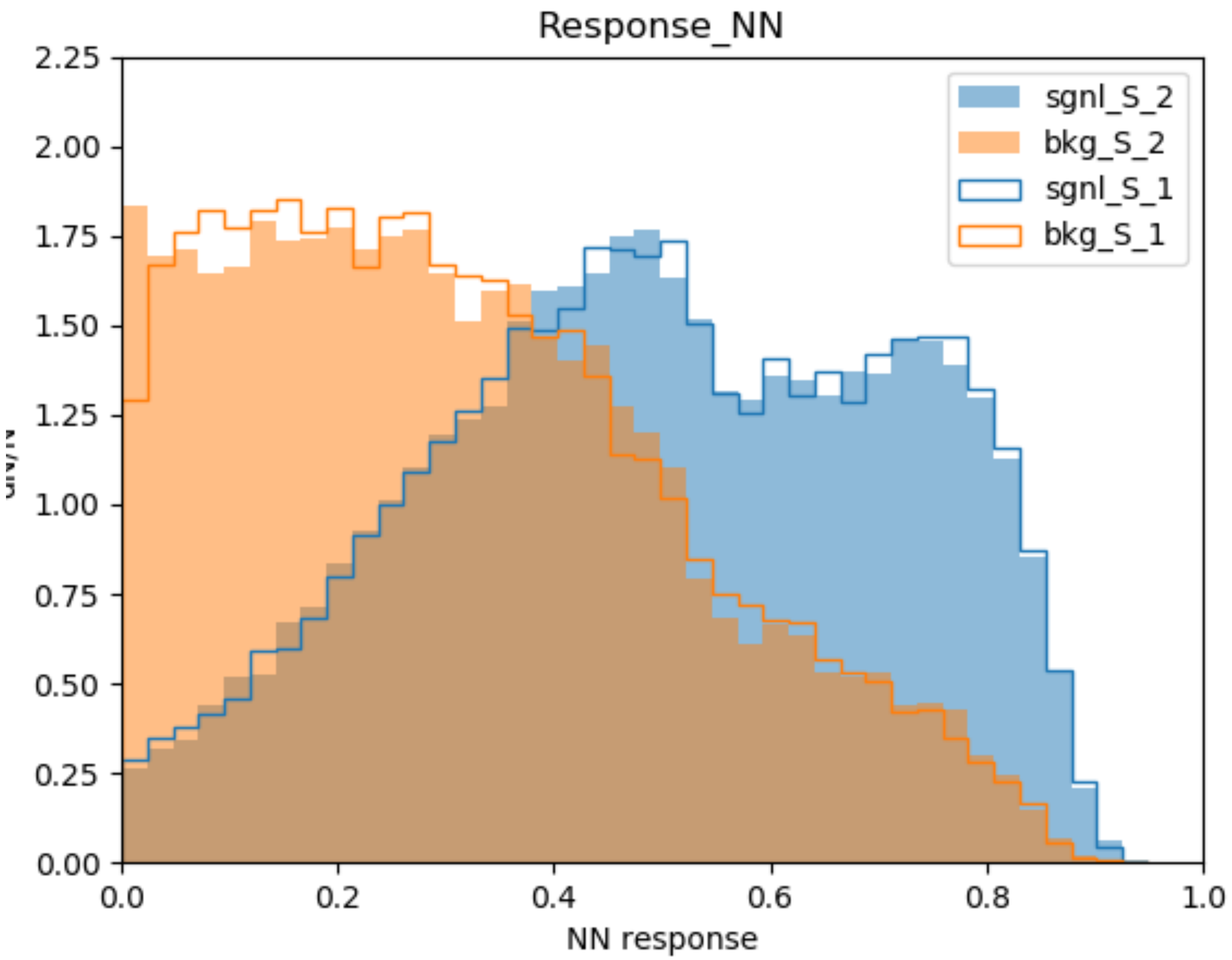


- Discriminator importance parameter $\lambda = 0$
- Discriminator importance parameter $\lambda = 1.2$

Results



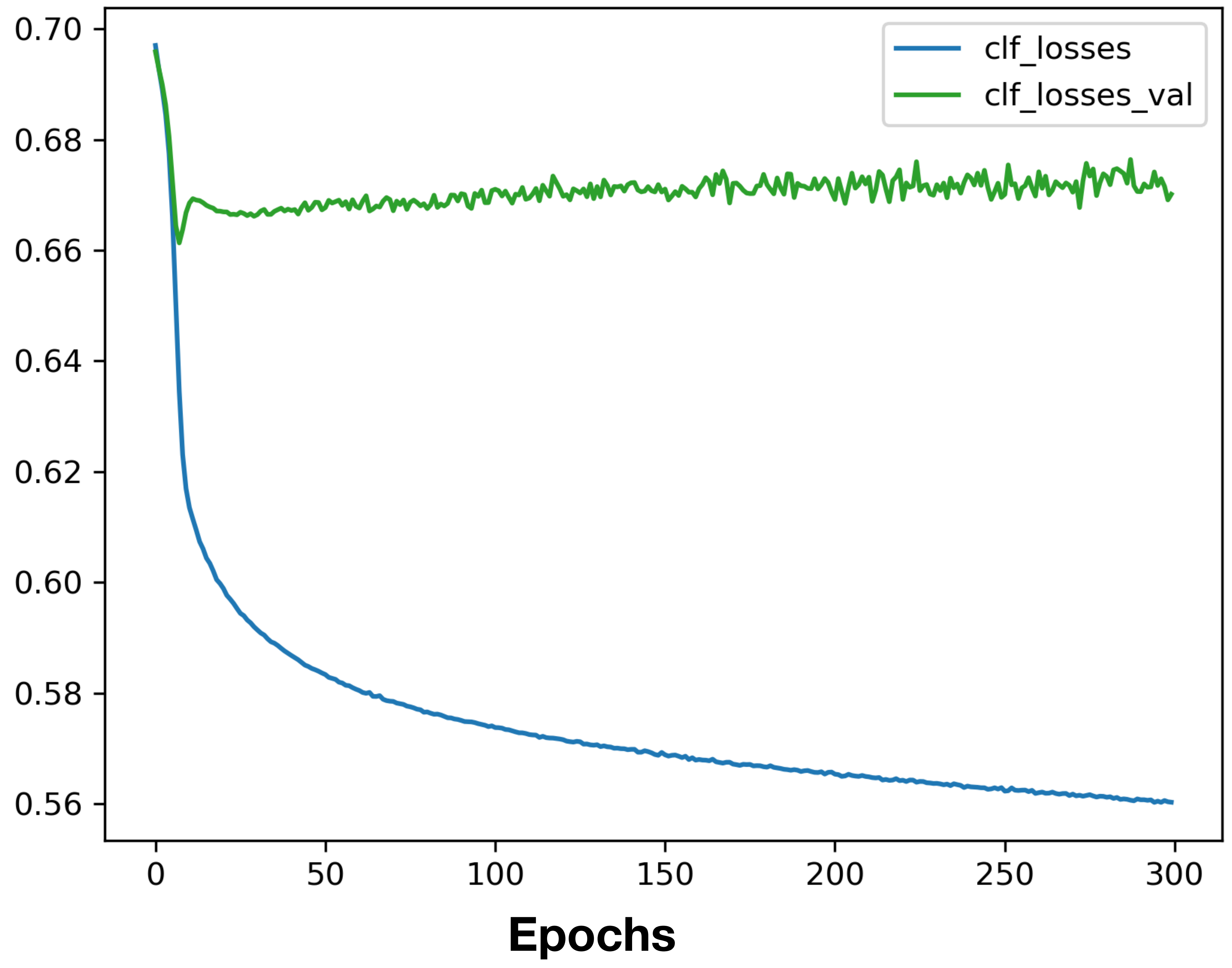
Receiver operating characteristic for ttH(bb)
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
mu: 1.2, AUC = 0.73, S_2
mu: 0.0, AUC = 0.67, S_2
mu: 1.2, AUC = 0.73, S_1
mu: 0.0, AUC = 0.78, S_1

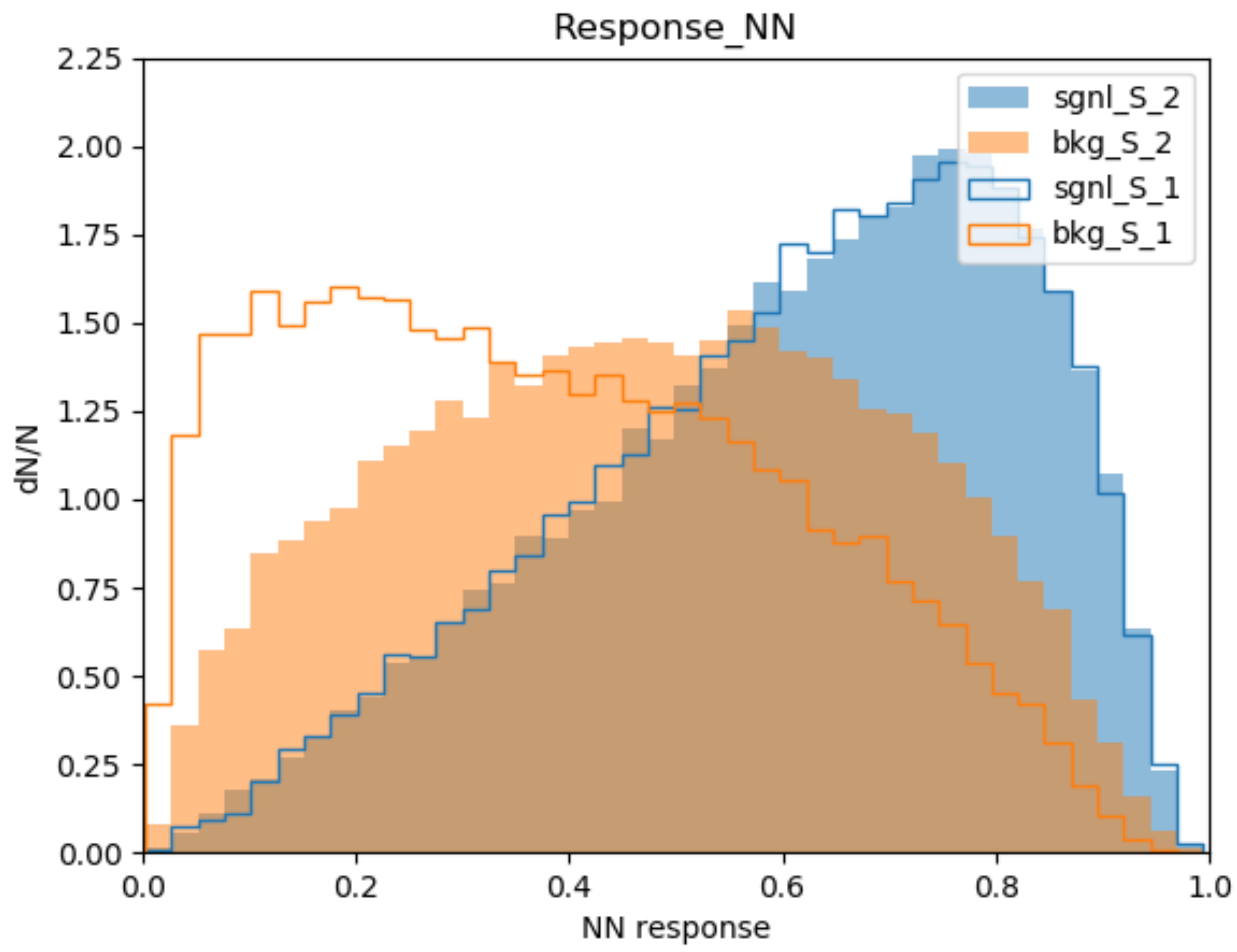


Response_NN
NN response
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
0.0
0.2
0.4
0.6
0.8
1.0
sgnl_S_2
bkg_S_2
sgnl_S_1
bkg_S_1

Training

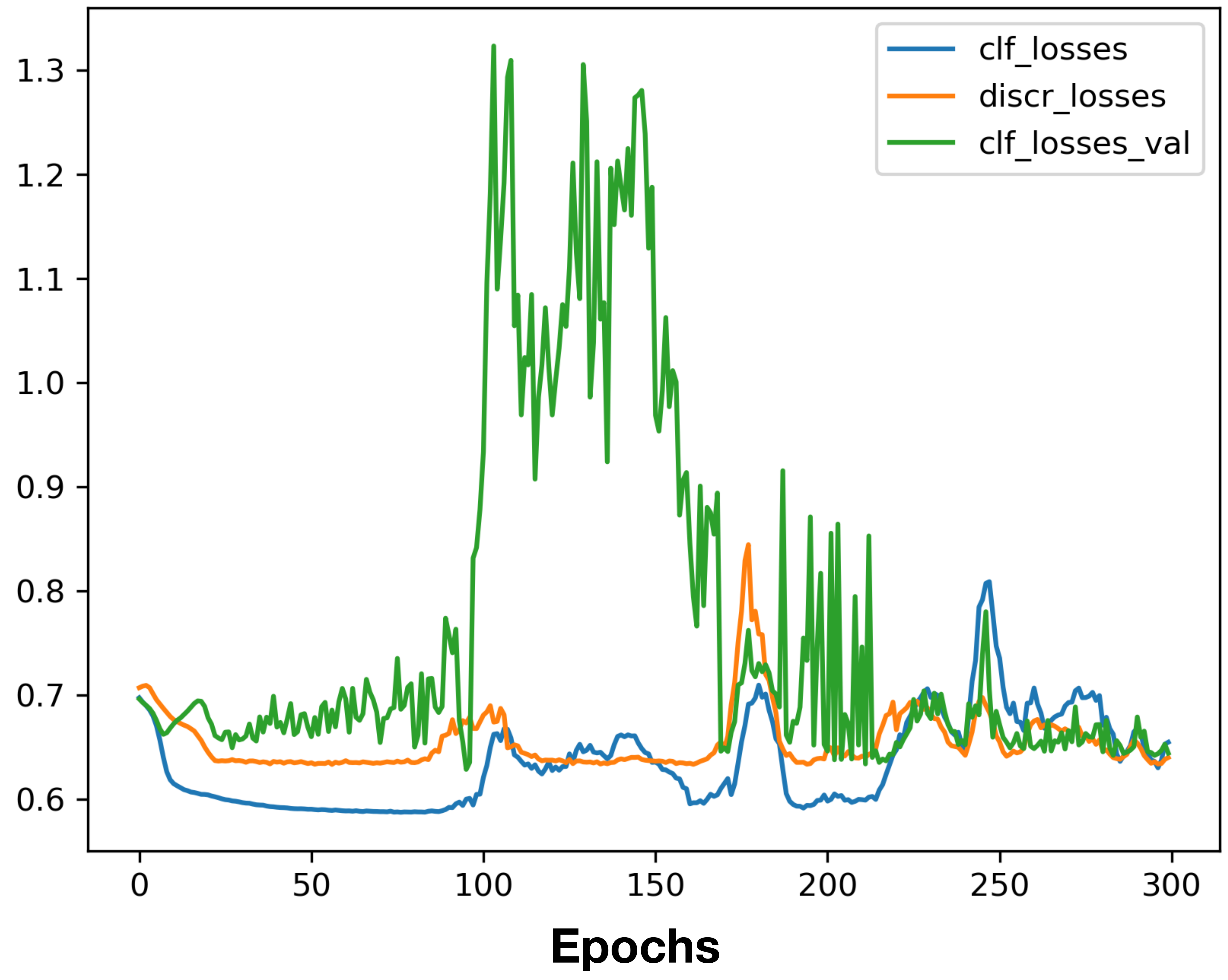
- $\lambda = 0$ Training plot

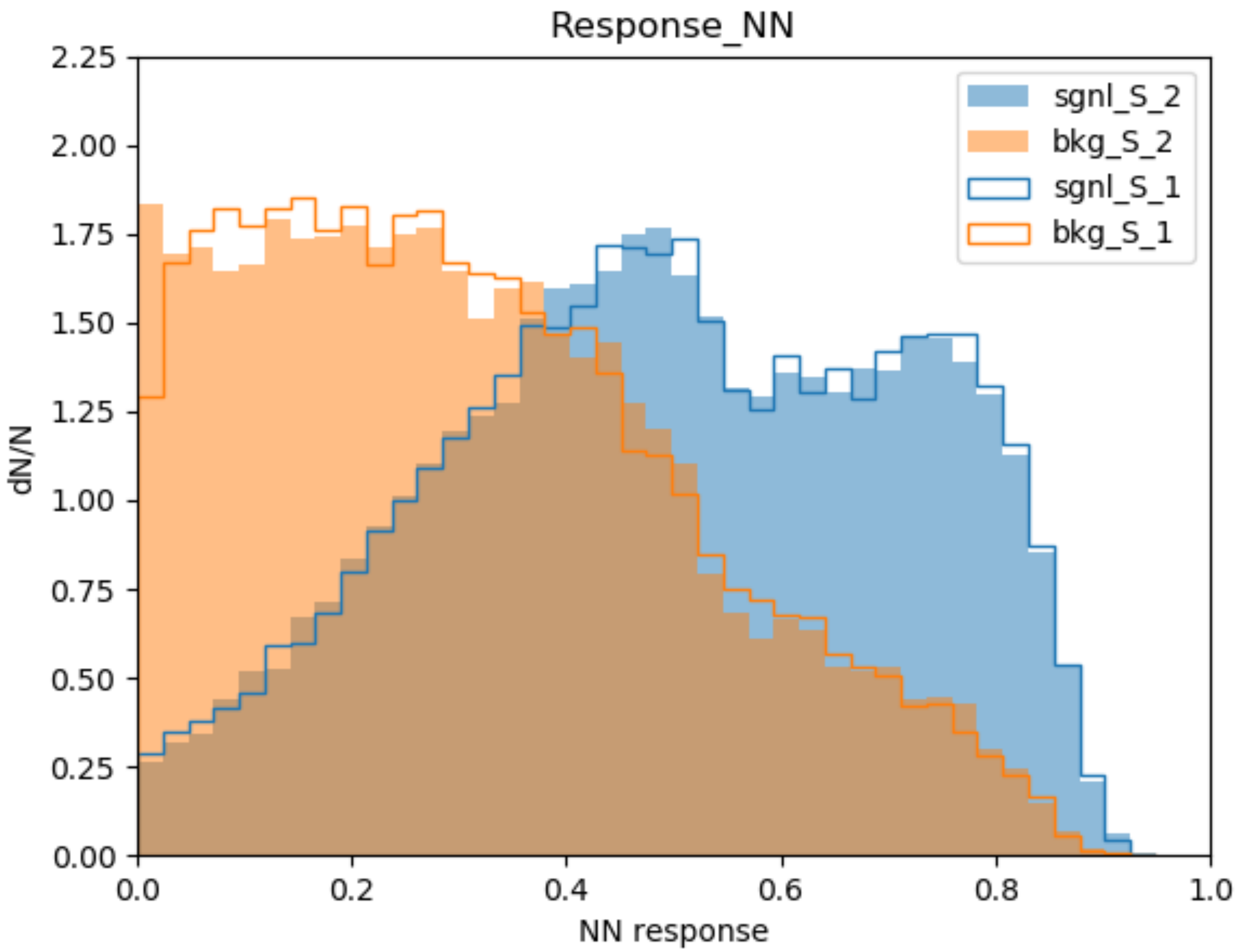




Training

- $\lambda = 1.20$ Training plot





Results

Table 1

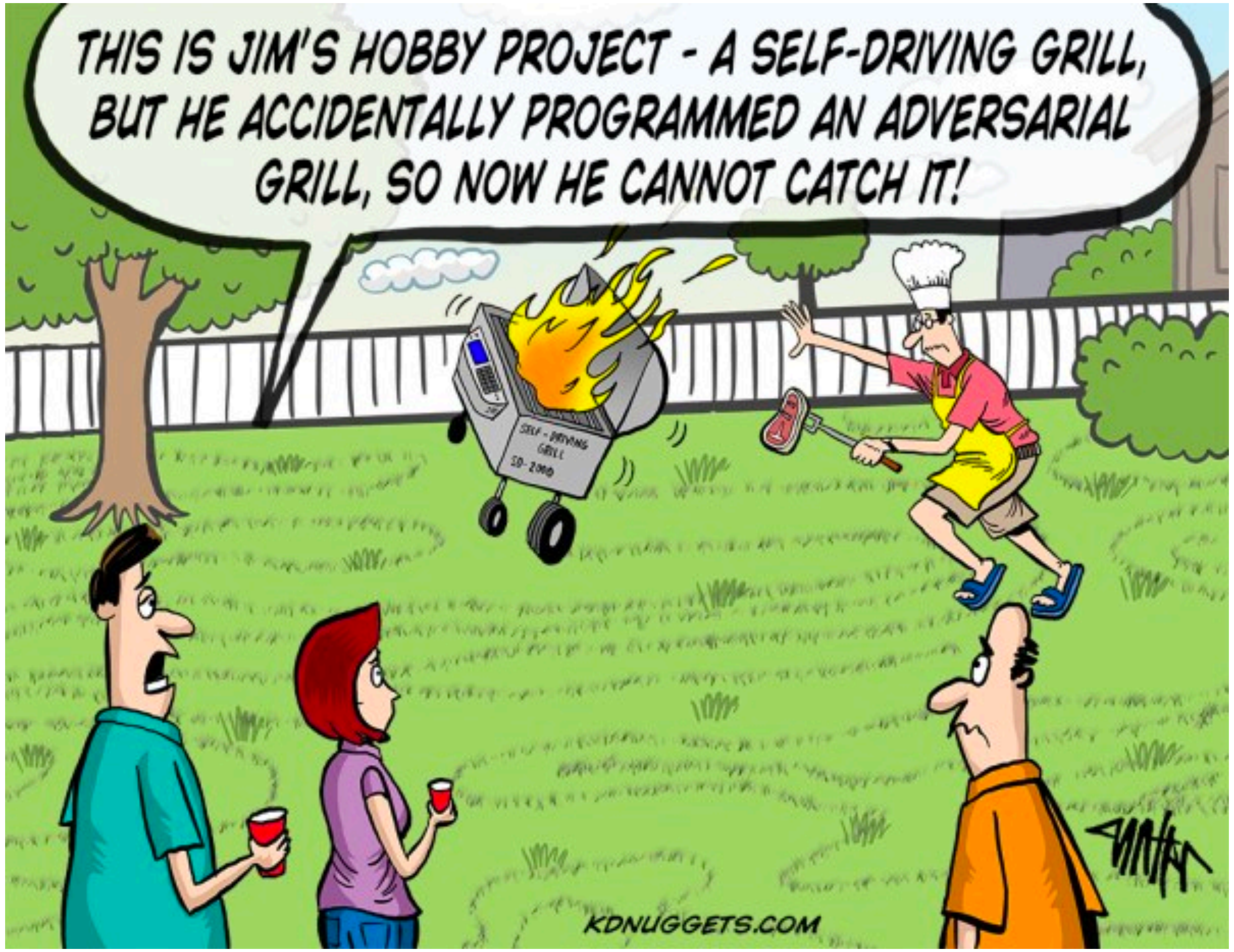
#Neurons in hidden layer	\lambda	Accuracy (test)	AUCROC (test)	Significance
[35, 30, 15]	0.000	0.621	0.668	0.212
[35, 30, 15]	1.300	0.652	$\boxed{0.729}$	$\boxed{\underline{4.408}}$
[40, 25, 25]	0.000	0.623	0.671	0.215
[40, 25, 25]	1.245	0.642	$\boxed{0.729}$	$\boxed{\underline{3.953}}$
[45, 30, 25]	0.000	0.622	0.671	0.214
[45, 30, 25]	1.201	0.649	$\boxed{0.729}$	$\boxed{\underline{6.417}}$

- Adadelta optimiser with 300 epochs and batch size of 5000 events

Future work

- Develop an appropriate stopping criterion
- Understand how the network architecture and discriminator importance parameter λ result on the performance
    - Add additional layers to discriminator part
- Play with signal/bkg ratio

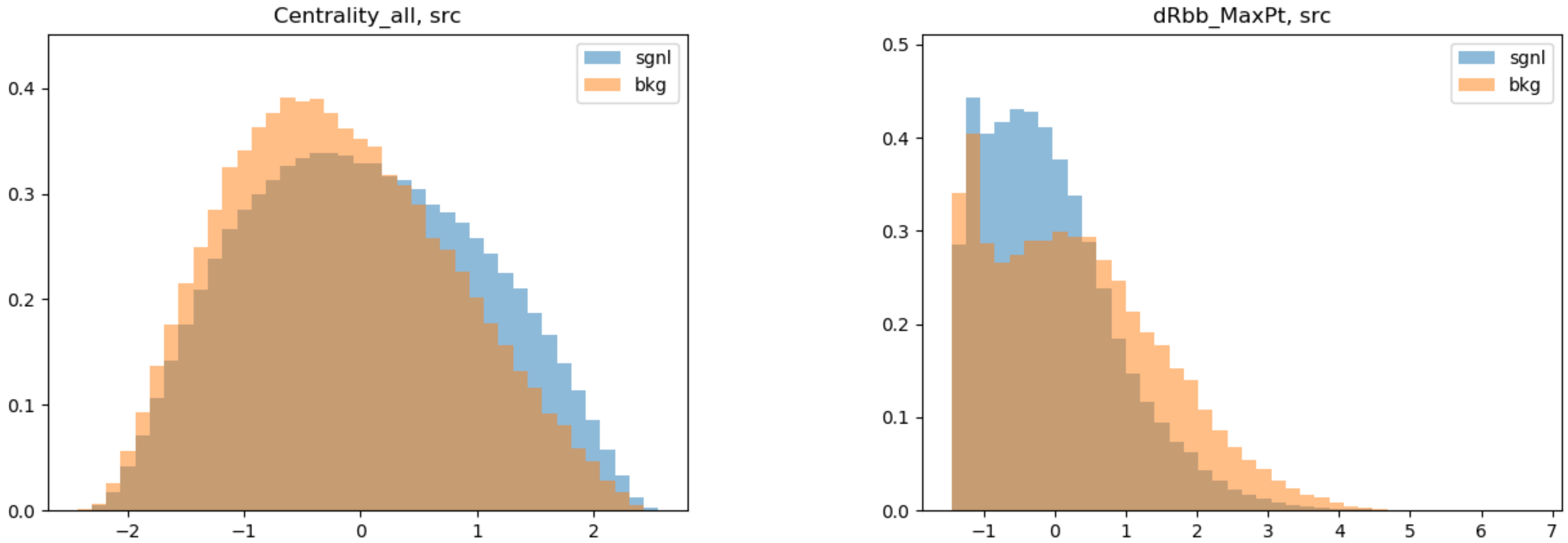
Questions?


THIS IS JIM'S HOBBY PROJECT - A SELF-DRIVING GRILL, BUT HE ACCIDENTALLY PROGRAMMED AN ADVERSARIAL GRILL, SO NOW HE CANNOT CATCH IT!
SELF-DRIVING GRILL
SD-2000
KDNUGGETS.COM

Variables description

dRbb_avg	average dR of all b-jet pairs
dRbb_MaxPt	dR of the b-jet pair with the highest sum of p_T
dRbb_MaxM	dR of the b-jet pair with the highest invariant mass
dRlb1-dRlb3	dR of the charged lepton and the b-jet with the 1st-3rd largest p_T
dRlbb_MindR	dR of the charged lepton and total b-jet pair system which has the smallest dR
dRlj_MindR	minimum dR between the charged lepton and any jet
Mbb_MaxM	maximum invariant mass of any b-jet pair
Mbb_MindR	invariant mass of b-jet pair which has the smallest dR
Mbj_MaxPt	invariant mass of two jets with the largest p_T sum, where exactly one of the jets is a b-jet
Mjjj_MaxPt	invariant mass of any three jets with the largest p_T sum
pT_lep	transverse momentum of the charged lepton
HT_jets	sum of transverse momentum of all jets
HT_all	sum of transverse momentum of all jets and the charged lepton
nJets_Pt40	number of jets with $p_T \geq$ 40 GeV
nbTag	number of b-jets
nHiggsbb30	number of b-jet pairs with an invariant mass within 30 GeV of the Higgs boson mass of 125 GeV
MET	missing transverse energy
dEtajj_MaxdEta	largest difference in longitudinal angle η of any two jets
Centrality_all	ratio of momentum sum over the energy sum of all objects
Hi_all, H2_jets	1st-5th Fox Wolfram transverse moment [9] of all objects

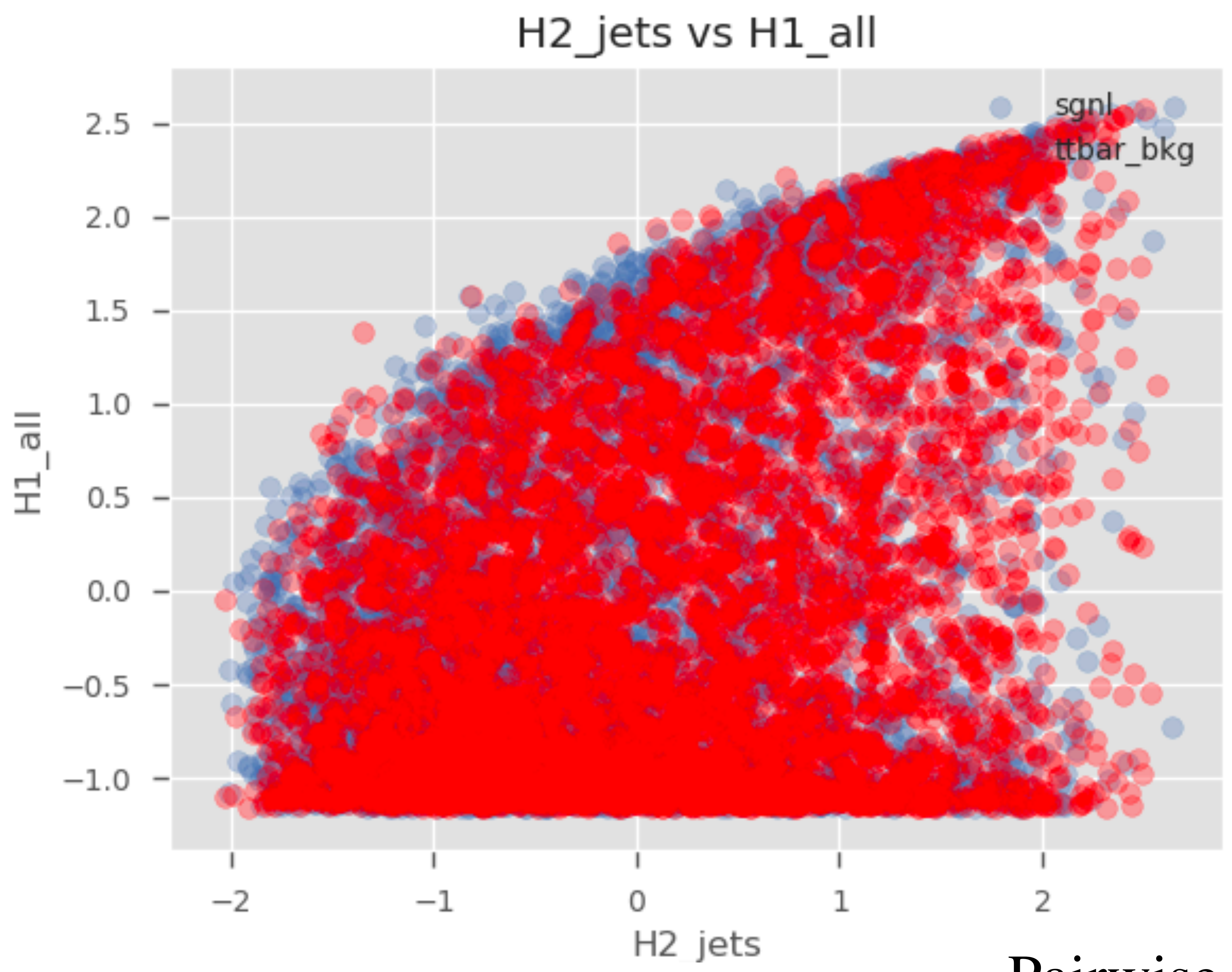
Signal vs Background classification is hard task

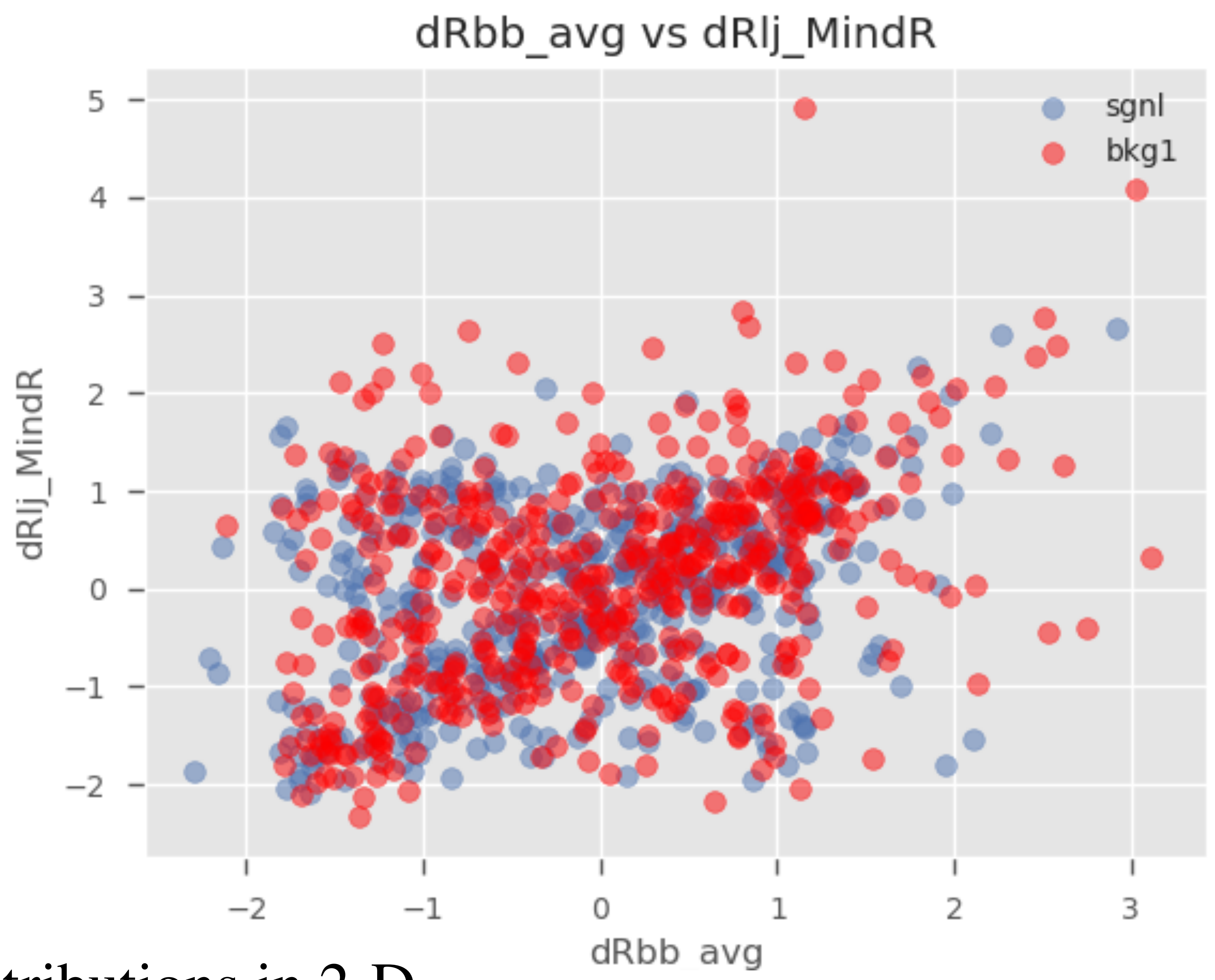


Distributions of variables

- The problem is <u>hard</u> in terms of classification
- Distributions most of the variables look <u>similar</u> for signal and background
- One <u>cannot separate</u> them neither in one nor in two dimensions

Signal vs Background classification is hard task





Pairwise distributions in 2-D

- One cannot separate signal vs background in two dimensions as well
- Principal Component Analysis (PCA) and Linear Discriminate Analysis (LDA) transforms into 2D were unsuccessful