BDT update

04/12/2024

Method

Approach 1

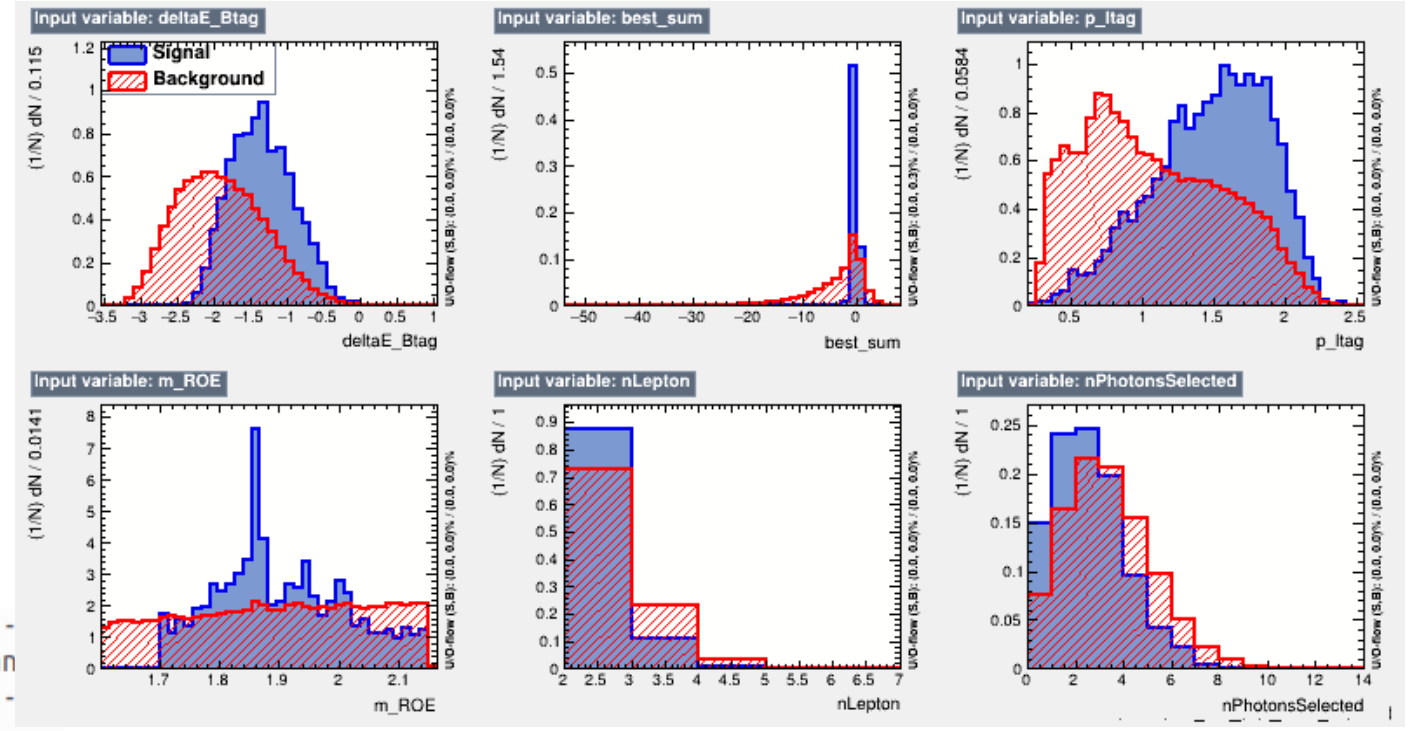
- Training (70%) and testing (30%) on the 1.0 M $\tau \rightarrow \pi$ sample.
- Applying on the 8.9 M $\tau \rightarrow$ generic signal and ten streams (07 training, 03 testing) of MC.
- Estimated the N_{sig} and N_{bg}.

Approach 2

- Training (70%) and testing (30%) on 8.9M $\tau \rightarrow$ generic sample.
- Applying on the 4.4 M $\tau \rightarrow$ generic signal and ten streams (07 training, 03 testing) of MC.
- Estimated the N_{sig} and N_{bg}.

Approach 1

- Training (70%) and testing (30%) on the 1.0 M $\tau \to \pi$ sample.
- Applying on the 8.9 M $\tau \to$ generic signal and ten streams (07 training, 03 testing) of MC.
- Estimated the N_{sig} and N_{bg}.



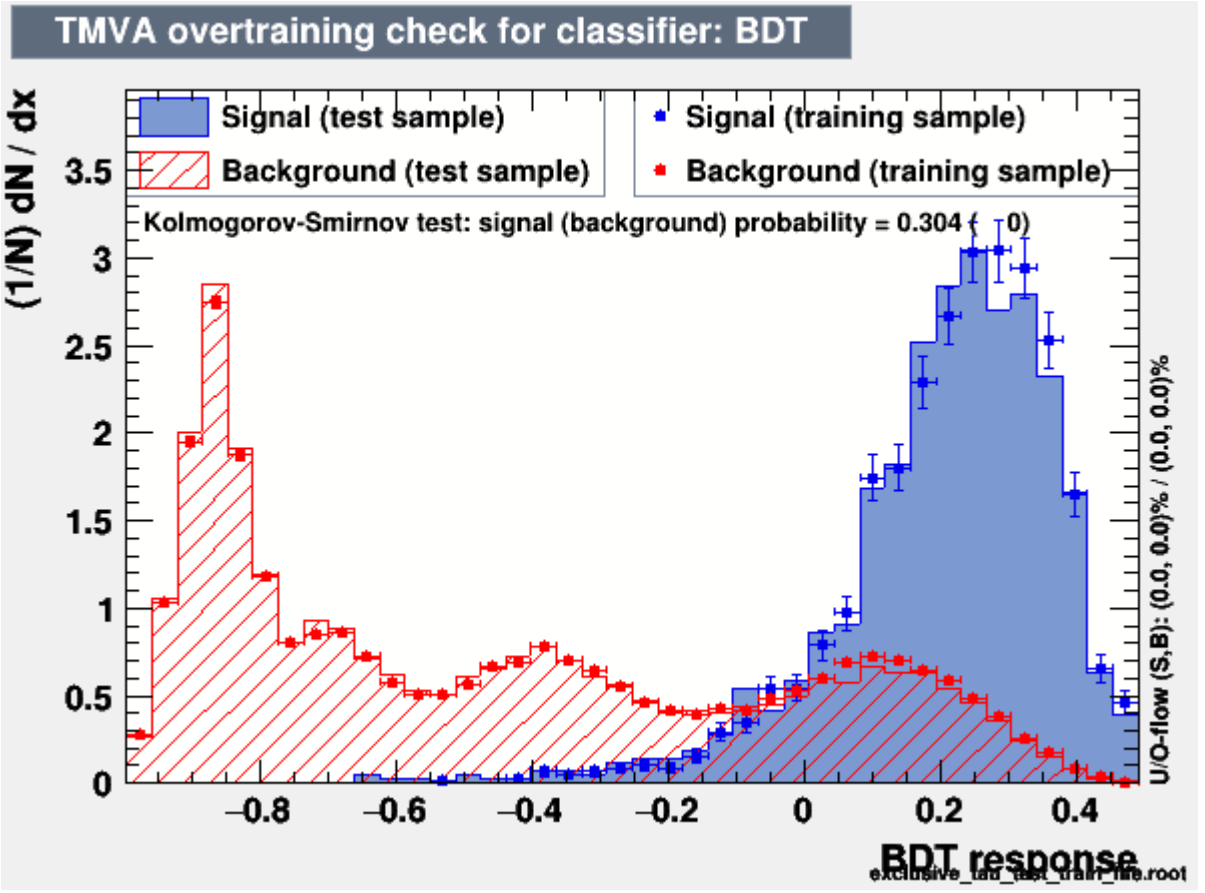
Var. importance

```
-------------------------------------------
Rank : Variable         : Variable Importan
-------------------------------------------
   1 : best_sum         : 3.530e-01
   2 : deltaE_Btag      : 2.586e-01
   3 : m_ROE            : 1.414e-01
   4 : p_ltag           : 1.345e-01
   5 : nPhotonsSelected : 7.128e-02
   6 : nLepton          : 4.118e-02
```

Approach 1

- Training (70%) and testing (30%) on the 1.0 M $\tau \rightarrow \pi$ sample.
- Applying on the 8.9 M $\tau \rightarrow$ generic signal and ten streams (07 training, 03 testing) of MC.
- Estimated the N_{sig} and N_{bg}.

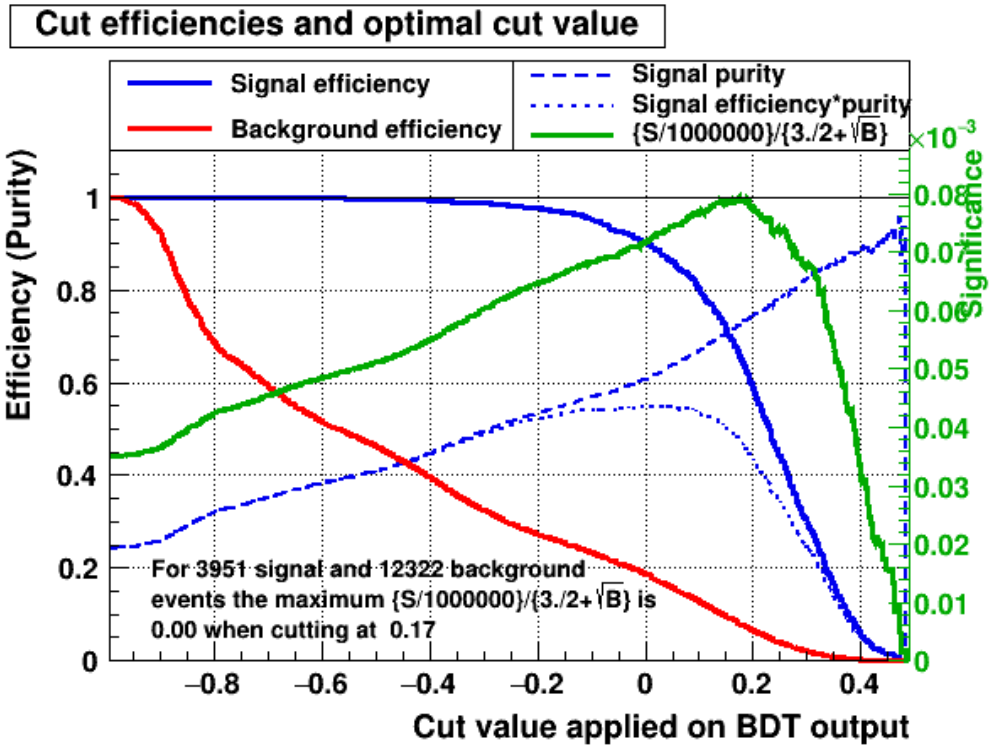
Now quite better agreement between train and test signal, background.



Approach 1

- Training (70%) and testing (30%) on the 1.0 M $\tau \rightarrow \pi$ sample.
- Applying on the 8.9 M $\tau \rightarrow$ generic signal and ten streams (07 training, 03 testing) of MC.
- Estimated the N_{sig} and N_{bg}.

BDT > 0.17



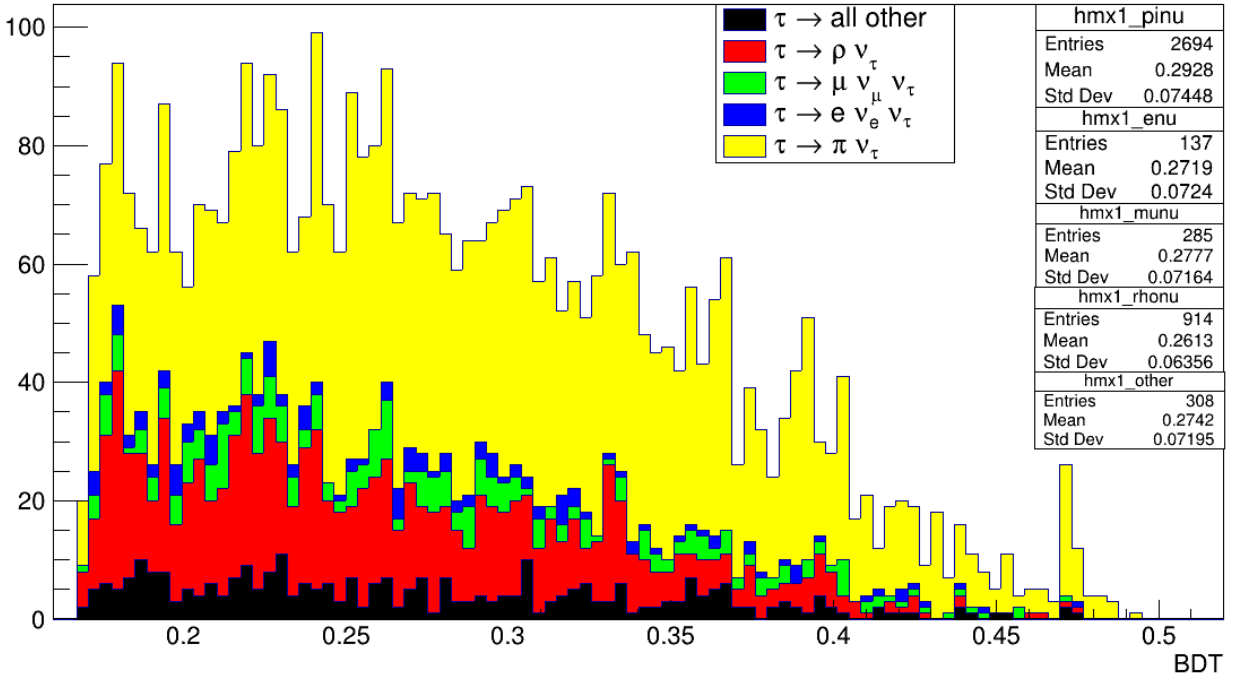

```
==========================================================================================================
Classifier   (  #signal, #backgr.)  Optimal-cut  (S/1000000)/(3./2+sqrt(B))      NSig      NBkg   EffSig   EffBkg
----------------------------------------------------------------------------------------------------------
       BDT:  (     3951,    12322)       0.1712   7.93917e-05  2654.005  1019.478   0.6717  0.08274
      BDTG:  (     3951,    12322)       0.6288   7.83965e-05  2247.235  737.9374   0.5688  0.05989
    Fisher:  (     3951,    12322)      -0.0114   5.00991e-05  3197.476  3884.158   0.8093   0.3152
       MLP:  (     3951,    12322)       0.6463   5.60126e-05  2757.365  2277.918   0.6979   0.1849
```

Approach 1

- Training (70%) and testing (30%) on the 1.0 M $\tau \rightarrow \pi$ sample.
- Applying on the 8.9 M $\tau \rightarrow$ generic signal and ten streams (07 training, 03 testing) of MC.
- Estimated the N_{sig} and N_{bg}.

N_{pi}	2694	1
N_e	137	0.05
N_{mu}	285	0.10
N_{rho}	914	0.34
N_{others}	308	0.11

BDT score

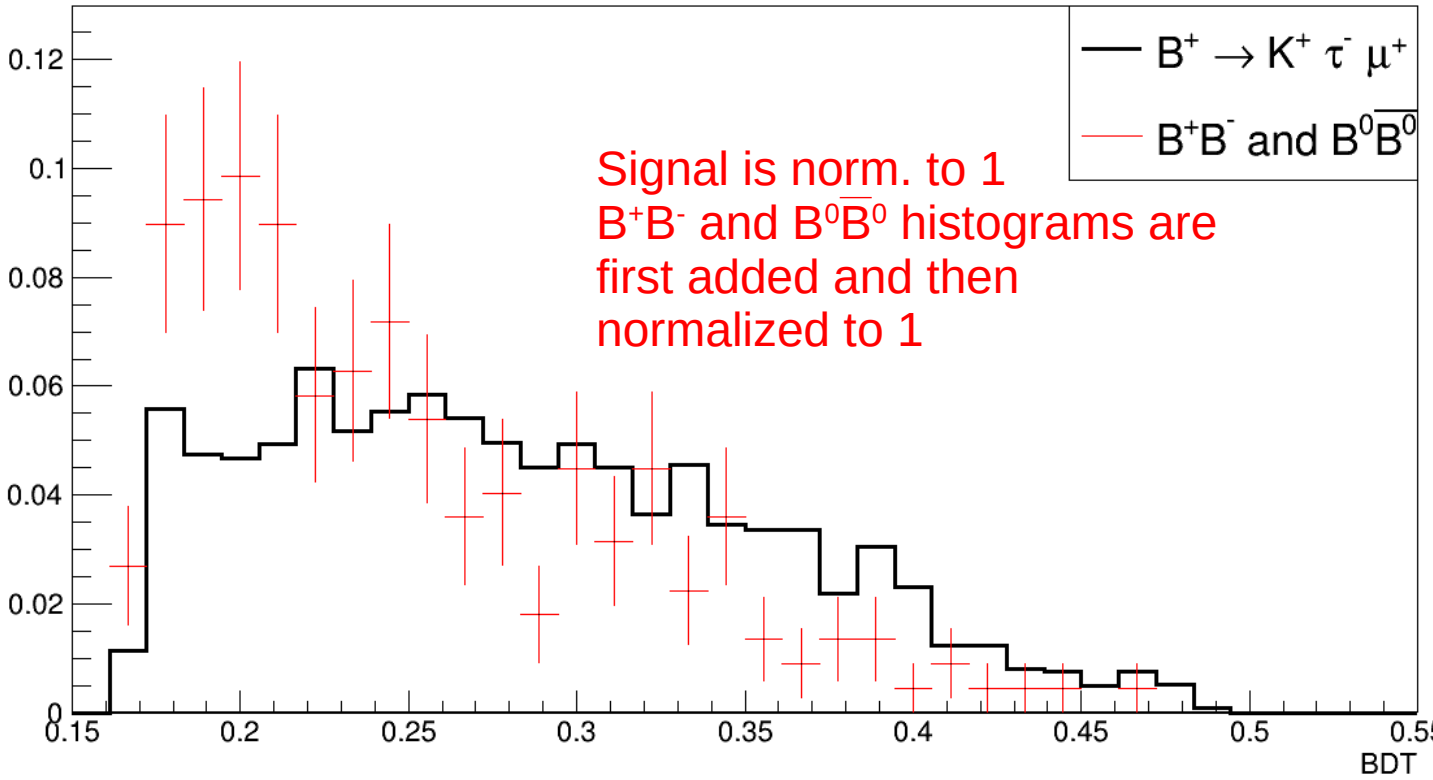


For 5×10^{-5} BF

$N_{sig} = 19$ & $N_{sig} = 12$ (for only pi mode)

Approach 1

Background calculation

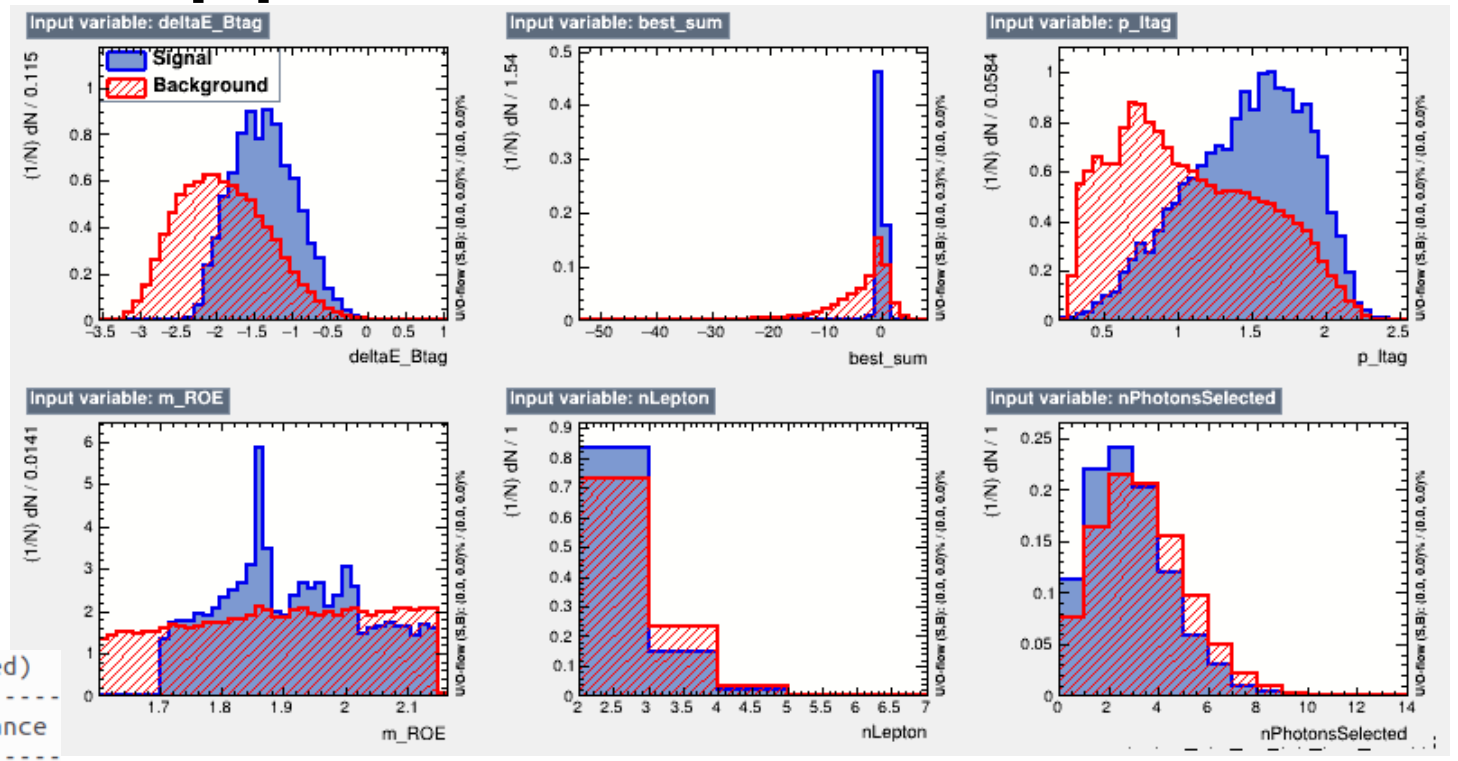


$N_{B+B-} = 150$

$N_{B0B0} = 73$

Approach 2

- Training (70%) and testing (30%) on 8.9M $\tau \rightarrow$ generic sample.
- Applying on the 4.4 M $\tau \rightarrow$ generic signal and ten streams (07 training, 03 testing) of MC.
- Estimated the N_{sig} and N_{bg}.



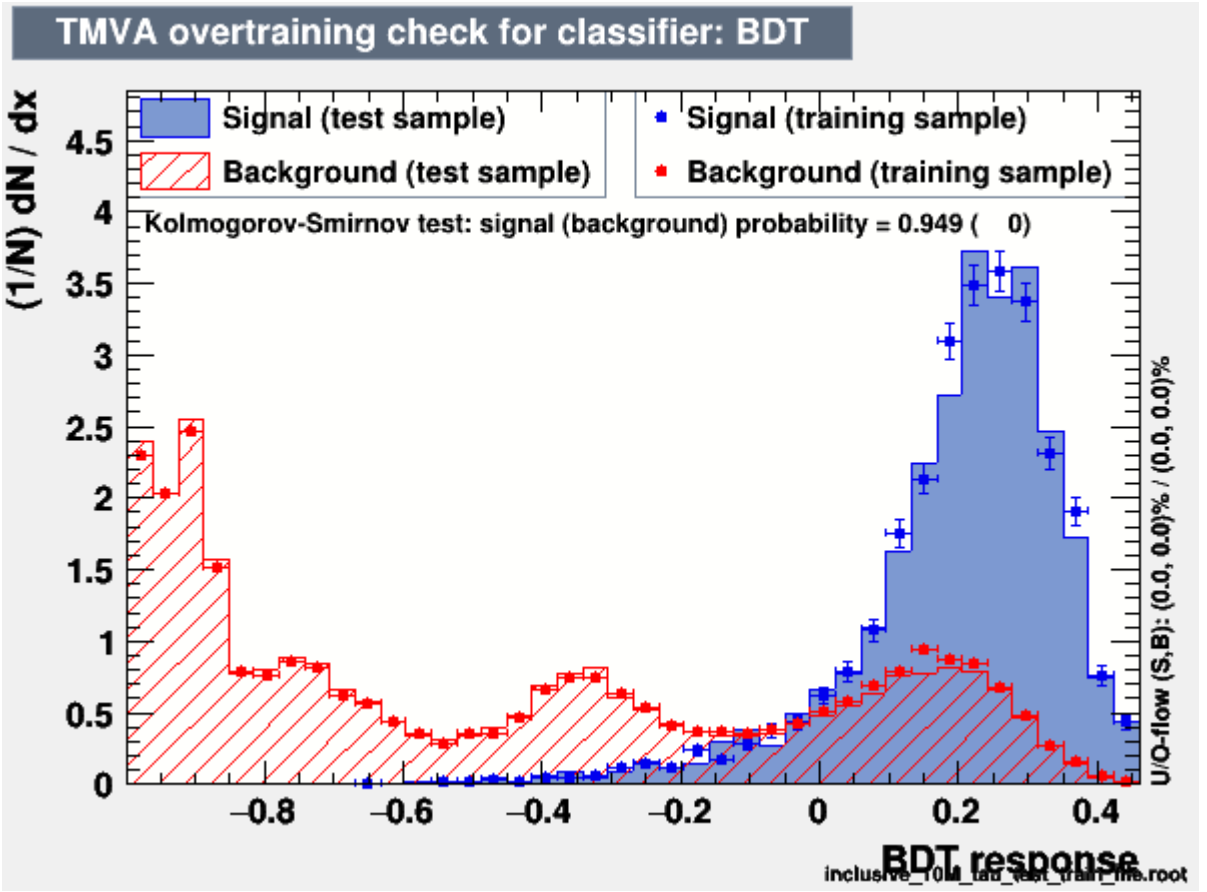
Var. importance

```
Ranking result (top variable is best ranked)
-----------------------------------------------
Rank : Variable         : Variable Importance
-----------------------------------------------
   1 : best_sum         : 4.116e-01
   2 : deltaE_Btag      : 2.937e-01
   3 : m_ROE            : 1.364e-01
   4 : p_ltag           : 9.655e-02
   5 : nPhotonsSelected : 4.987e-02
   6 : nLepton          : 1.195e-02
```

Approach 2

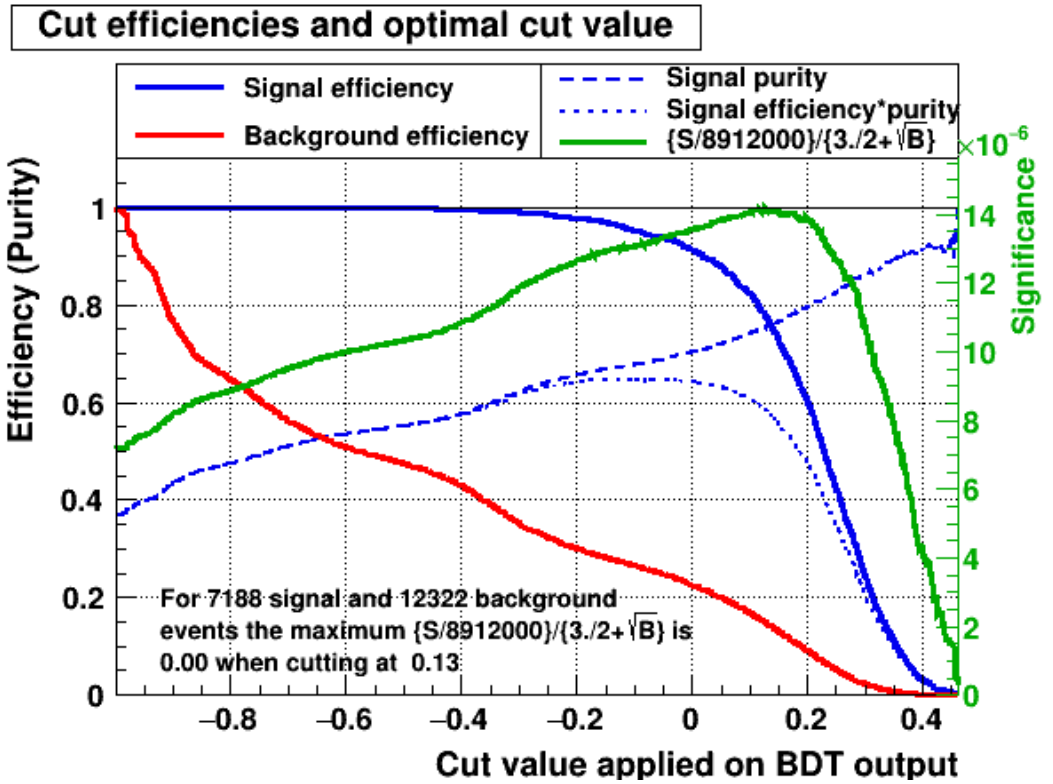
- Training (70%) and testing (30%) on 8.9M $\tau \rightarrow$ generic sample.
- Applying on the 4.4 M $\tau \rightarrow$ generic signal and ten streams (07 training, 03 testing) of MC.
- Estimated the N_{sig} and N_{bg}.

Now quite better agreement between train and test signal, background.



Approach 2

- Training (70%) and testing (30%) on 8.9M $\tau \rightarrow$ generic sample.
- Applying on the 4.4 M $\tau \rightarrow$ generic signal and ten streams (07 training, 03 testing) of MC.
- Estimated the N_{sig} and N_{bg}.



BDT optimal cut

BDT > 0.13

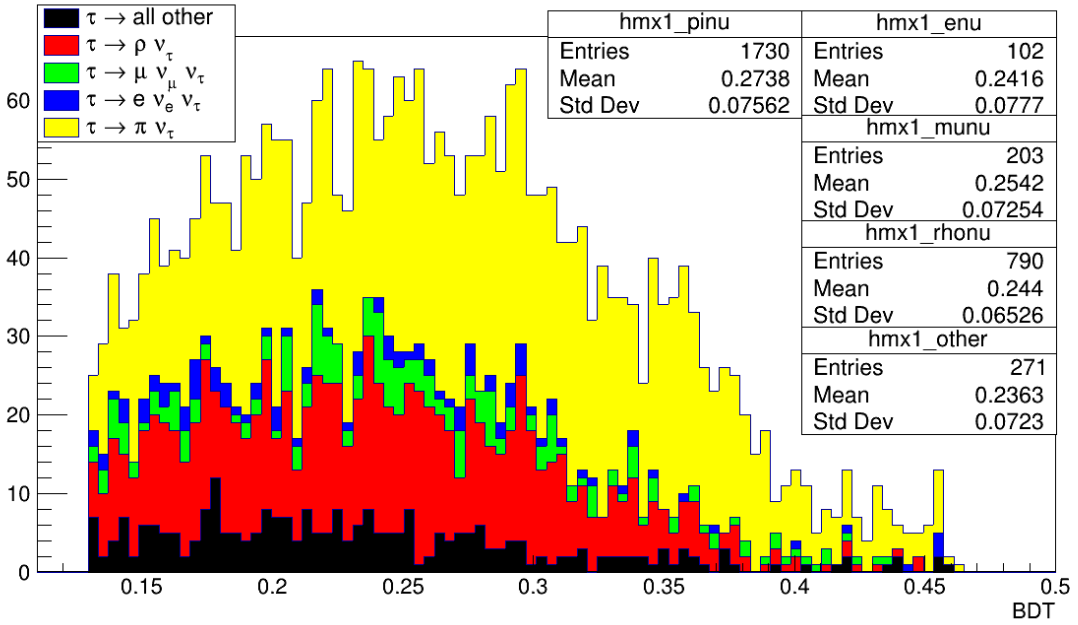
Classifier	(#signal, #backgr.)	Optimal-cut	(S/8912000)/(3./2+sqrt(B))	NSig	NBkg	EffSig	EffBkg
BDT:	(7188, 12322)	0.1314	1.42e-05	5541.028	1788.033	0.7709	0.1451
BDTG:	(7188, 12322)	0.3586	1.39815e-05	5501.02	1818.89	0.7653	0.1476
Fisher:	(7188, 12322)	-0.0942	9.7947e-06	6411.189	5176.326	0.8919	0.4201
MLP:	(7188, 12322)	0.6007	1.0543e-05	5507.688	3262.473	0.7662	0.2648

Approach 2

- Training (70%) and testing (30%) on 8.9M $\tau \rightarrow$ generic sample.
- Applying on the 4.4 M $\tau \rightarrow$ generic signal and ten streams (07 training, 03 testing) of MC.
- Estimated the N_{sig} and N_{bg}.

N_{pi}	1730	1
N_e	102	0.05
N_{mu}	203	0.11
N_{rho}	790	0.46
N_{others}	271	0.16

BDT score

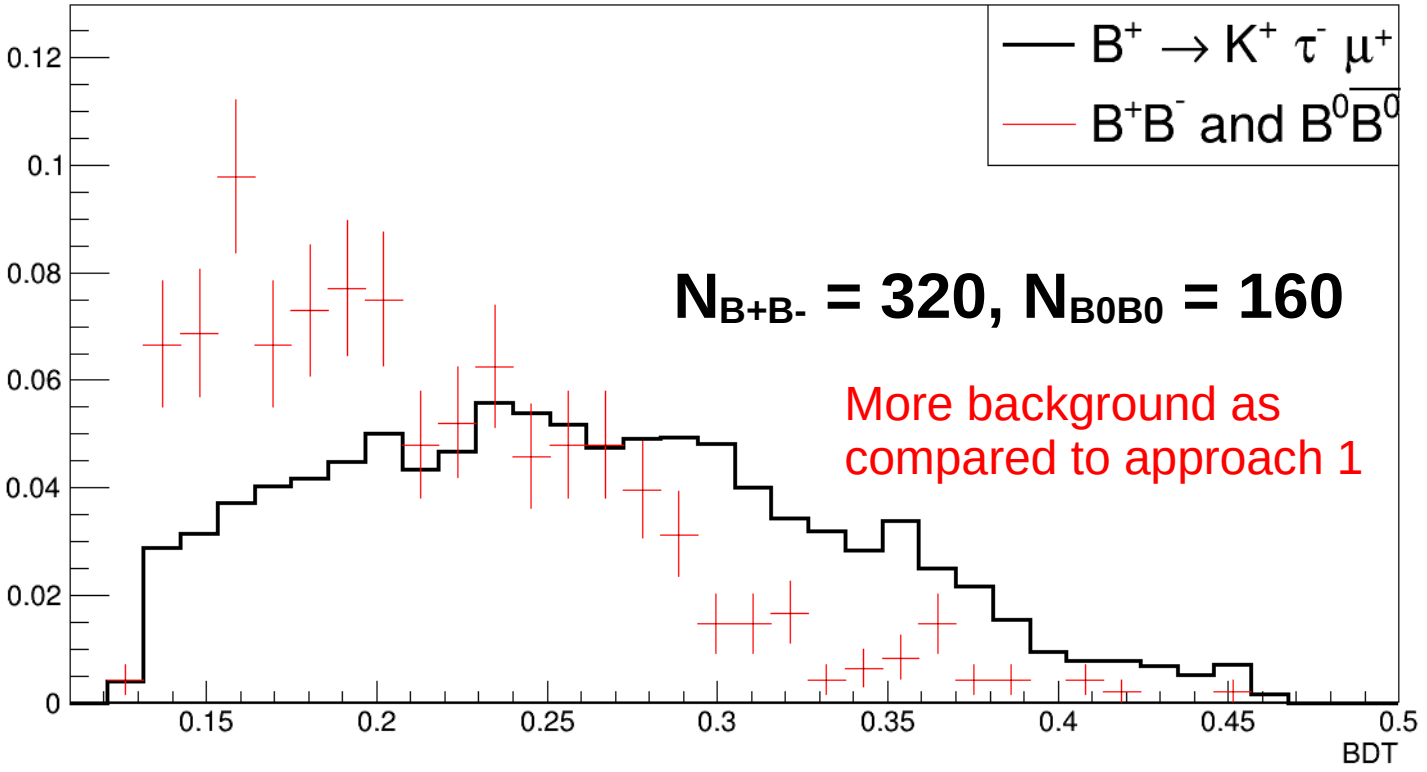


For 5×10^{-5} BF

$N_{sig} = 27$ & $N_{sig} = 15$ (for only pi mode)

Approach 2

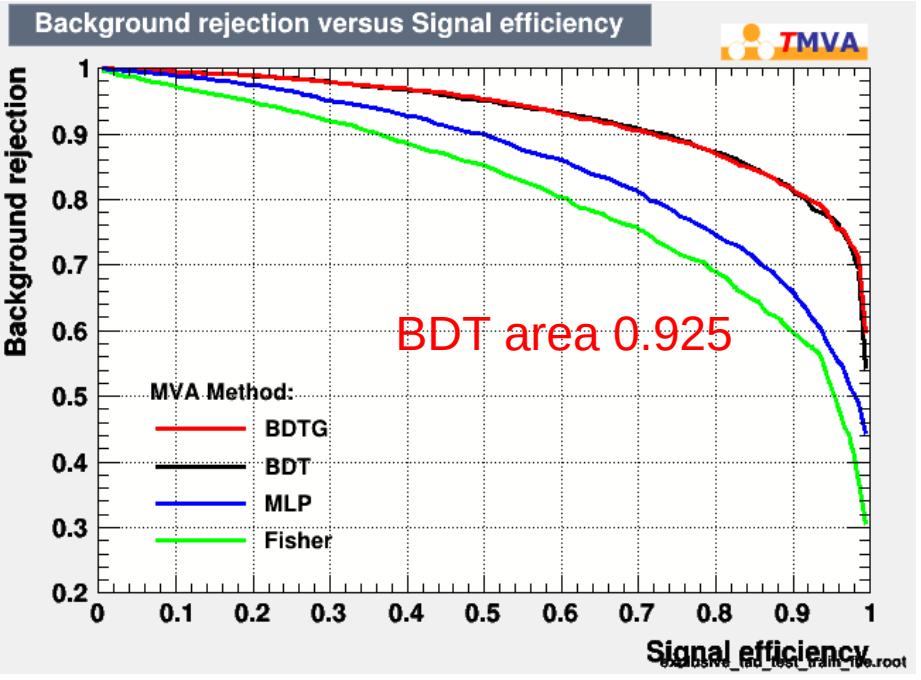
Background calculation



Back up

ROC (test train samples)

Approach 1



Approach 2