Photon conversion identification with machine learning approach

Pavel Gordeev

Motivation

- Particle identification is important in almost any high-energy physics analysis, but in some measurements such identification becomes crucial

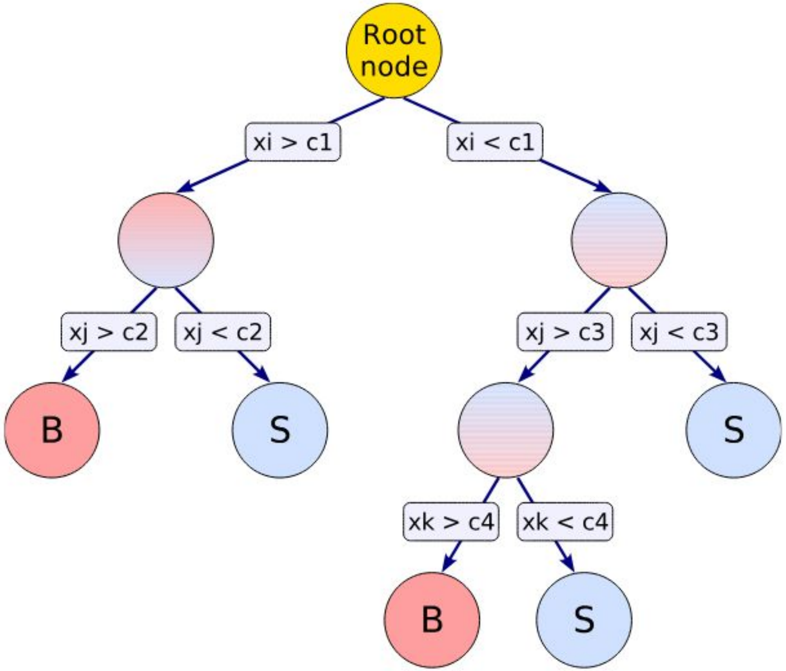
- Such analyses are the measurement of direct photon spectra and correlations, where the signal is comparable with possible contamination

- In this presentation, we will discuss improvements in particle identification in MPD detector that can be achieved by applying machine learning approach for particle identification in MPD tracking system

Boosted Decision Trees (BDT)

- BDT are widely used in HEP
- The training starts with the **root node**, where an initial splitting criterion for the full training sample is determined
- At each node, the split is determined by finding the variable and corresponding cut value **that provides the best separation** between signal and background
- The leaf nodes are classified as signal or background according to the class the majority of events belongs to

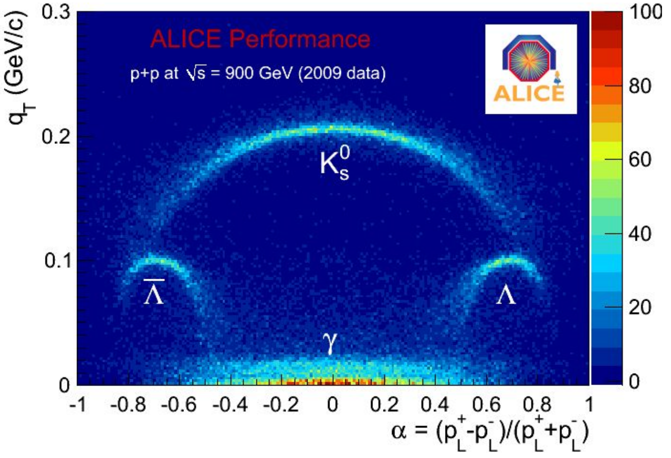
One decision tree



Training sample

Variables for training

- **N_clu** - number of TPC clusters
- **χ^2** - obtained from Kalman filter
- **η_1-2** - difference of pseudorapidity of tracks
- **DCA** - Distance of Closest Approach to PV for tracks
- **DCA_daug** - DCA between positively and negatively charged tracks
- **CPA** - Cosine of Pointing Angle
- **R** - conversion radius, distance from PV to SV
- **n_dE/dx** - PID of tracks based on specific loss in TPC, number of σ from electron/positron line
- **M_inv** - invariant mass of track pair
- Armenteros-Podolanski variables - **q_T** and **α**
- **|cosΨ|** - cosine of angle between pair plane and magnetic field (for Dalitz decays reduction)



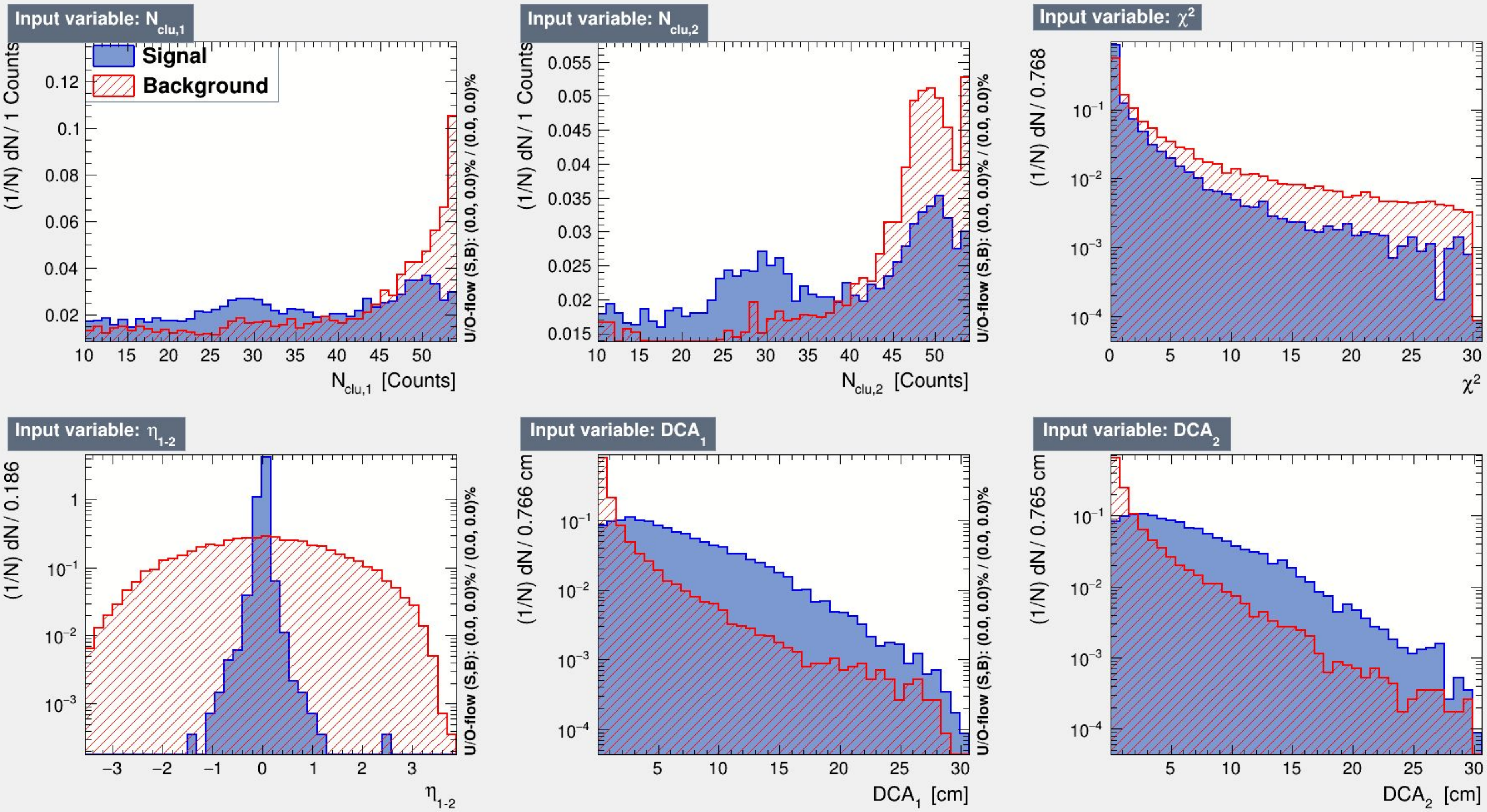
UrQMD, Bi-Bi, √s_NN=9.2 GeV

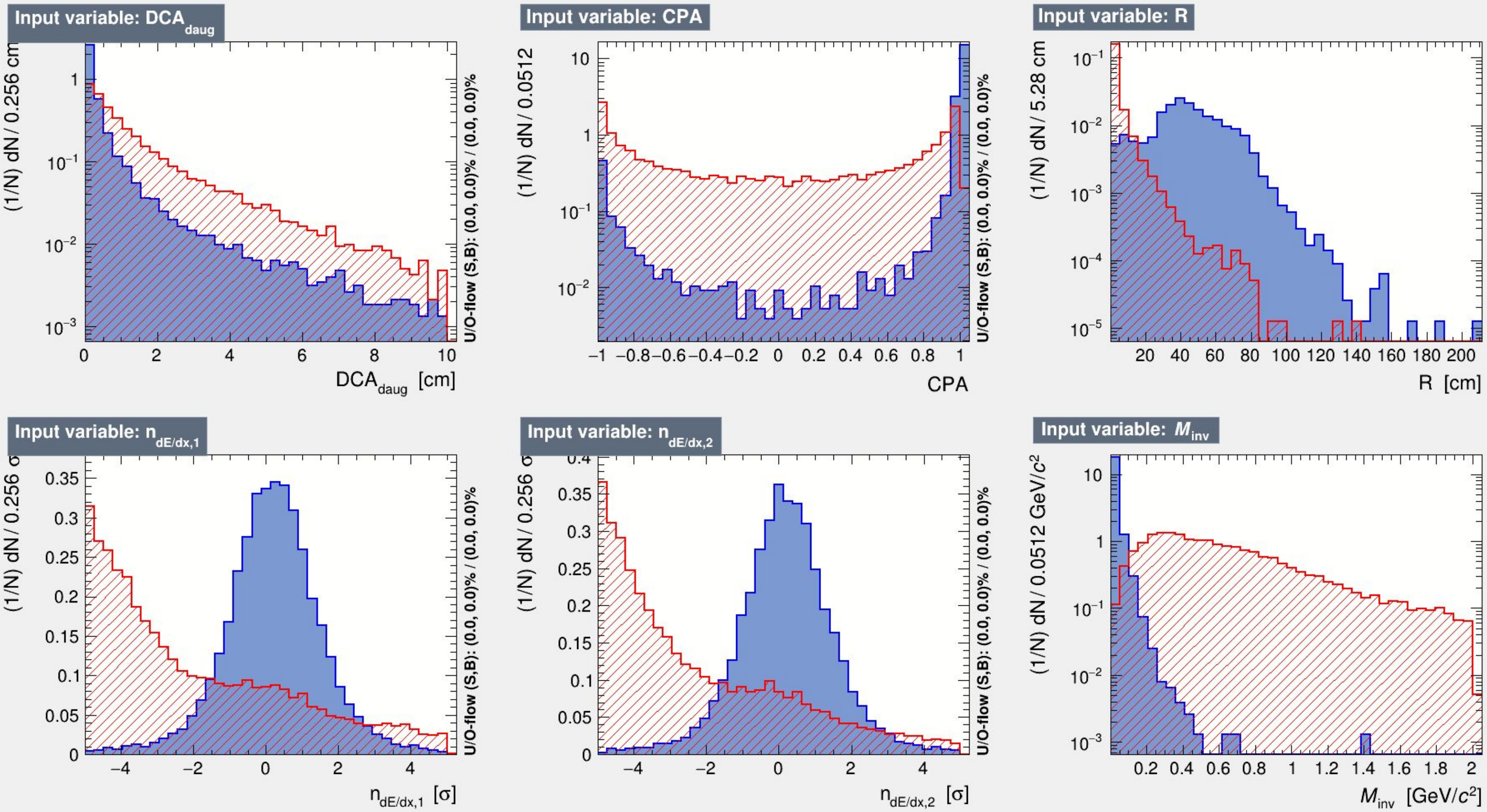
Event selection

- |V_z|<100 cm
- 0%<centrality<90%

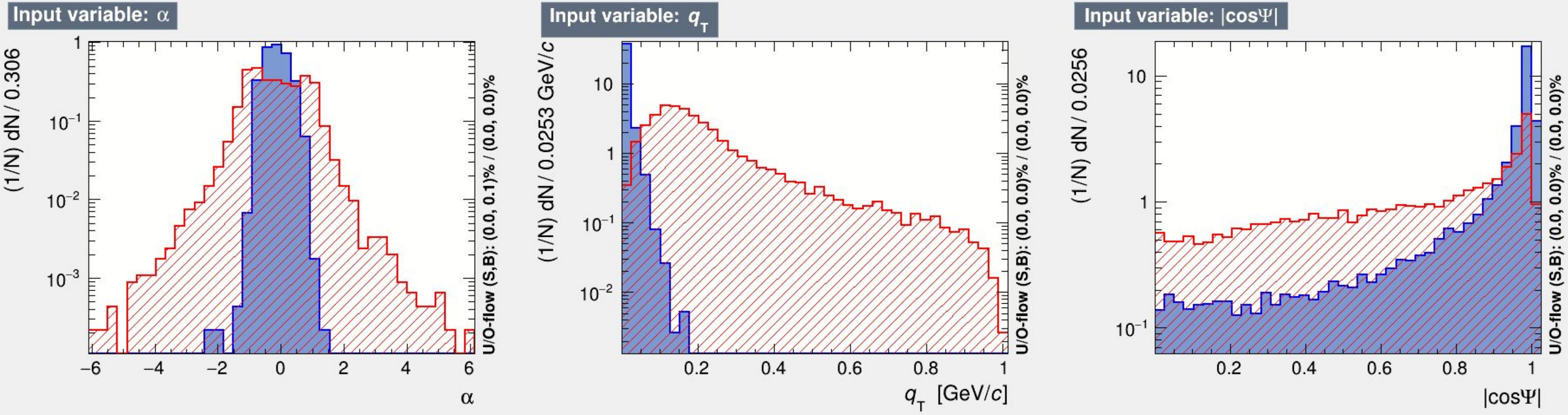
Preselections while tree writing:

- M_inv<2 GeV/c^2
- q_T<1 GeV/c
- χ^2<30
- DCA_daug<10 cm
- DCA_1<30 cm
- DCA_2<30 cm
- p_T,1<15 GeV/c
- p_T,2<15 GeV/c


Input variable: $N_{clu,1}$
Signal
Background
(1/N) dN / 1 Counts
$N_{clu,1}$ [Counts]
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: $N_{clu,2}$
(1/N) dN / 1 Counts
$N_{clu,2}$ [Counts]
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: χ^2
(1/N) dN / 0.768
χ^2
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: η_{1-2}
(1/N) dN / 0.186
η_{1-2}
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: DCA_1
(1/N) dN / 0.766 cm
DCA_1 [cm]
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: DCA_2
(1/N) dN / 0.765 cm
DCA_2 [cm]
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%



Input variable: DCA_daug
(1/N) dN / 0.256 cm
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
DCA_daug [cm]
Input variable: CPA
(1/N) dN / 0.0512
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
CPA
Input variable: R
(1/N) dN / 5.28 cm
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
R [cm]
Input variable: n_dE/dx,1
(1/N) dN / 0.256 σ
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
n_dE/dx,1 [σ]
Input variable: n_dE/dx,2
(1/N) dN / 0.256 σ
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
n_dE/dx,2 [σ]
Input variable: M_inv
(1/N) dN / 0.0512 GeV/c^2
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
M_inv [GeV/c^2]



Input variable: α
(1/N) dN / 0.306
U/O-flow (S,B): (0.0, 0.1)% / (0.0, 0.0)%
α
Input variable: q_T
(1/N) dN / 0.0253 GeV/c
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
q_T [GeV/c]
Input variable: |cosΨ|
(1/N) dN / 0.0256
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
|cosΨ|

Correlation Matrix (signal)

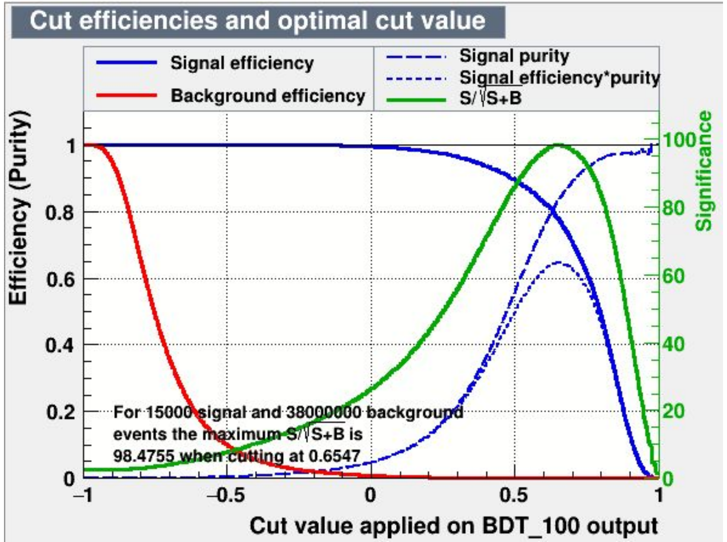
Linear correlation coefficients in %

	$N_{clu,1}$	$N_{clu,2}$	χ^2	$\eta_{1\text{-}2}$	DCA_1	DCA_2	DCA_{daug}	CPA	R	$n_{dE/dx,1}$	$n_{dE/dx,2}$	M_{inv}	α	q_T	$\lvert\cos\Psi\rvert$
$\lvert\cos\Psi\rvert$	-7	-7	-5	4	-8	-7	-2	-3	-1	1	1	18	3	22	100
q_T	-25	-26	26	2			27	-29	-2	-5	-5	90	4	100	22
α	-3		3	-1	2	-1	1		1	1	2		100	4	3
M_{inv}	-23	-22	23	11		1	22	-24	-1	-4	-6	100		90	18
$n_{dE/dx,2}$	-2	6	-3		5	1	-7	3	7	9	100	-6	2	-5	1
$n_{dE/dx,1}$	6	-3	-5		-1	6	-9	4	8	100	9	-4	1	-5	1
R	-21	-22	1		35	37	-14	25	100	8	7	-1	1	-2	-1
CPA	8	10	-13	-1	1		-27	100	25	4	3	-24		-29	-3
DCA_{daug}	-18	-16	60	-2	7	7	100	-27	-14	-9	-7	22	1	27	-2
DCA_2	-3	-39	3	-1	34	100	7		37	6	1	1	-1		-7
DCA_1	-39	-3	3	-1	100	34	7	1	35	-1	5		2		-8
$\eta_{1\text{-}2}$	-1		-2	100	-1	-1	-2	-1				11	-1	2	4
χ^2	-18	-16	100	-2	3	3	60	-13	1	-5	-3	23	3	26	-5
$N_{clu,2}$	32	100	-16		-3	-39	-16	10	-22	-3	6	-22		-26	-7
$N_{clu,1}$	100	32	-18	-1	-39	-3	-18	8	-21	6	-2	-23	-3	-25	-7

Training result

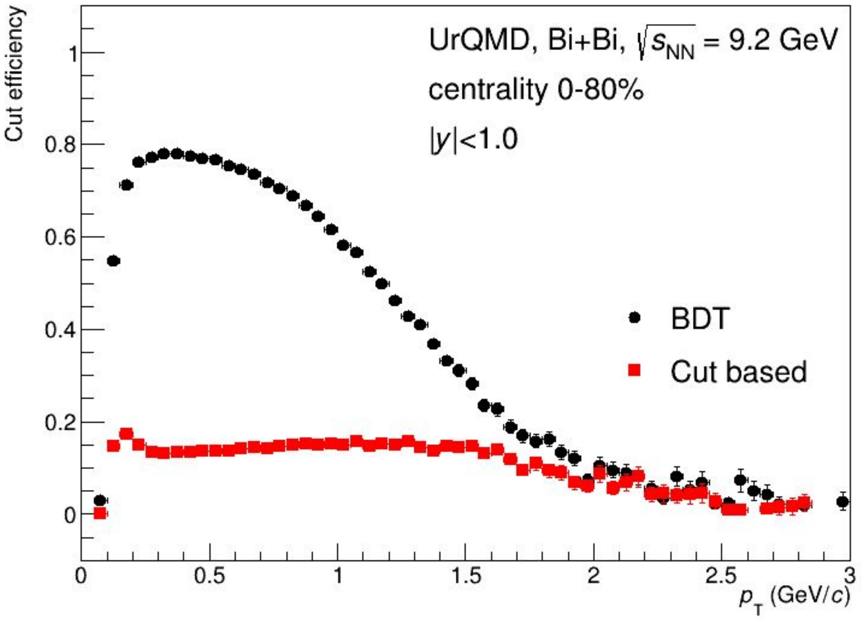
- For training: S=15’000 and B=15’000
- For testing: S=15’000 and B=38’000’000
- In the data sample we have ~2500 background to 1 real conversion photon
- The results of the training are: weight file, BDT response plot and optimal selection, variable ranking

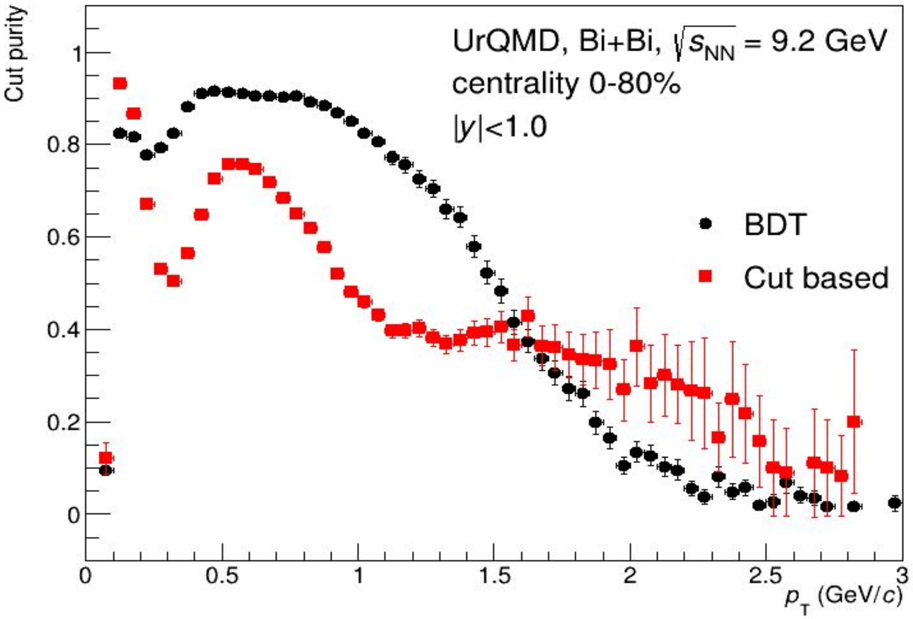
```
Ranking result (top variable is best ranked)
-------------------------------------------
Rank : Variable   : Variable Importance
-------------------------------------------
   1 : mass       : 1.348e-01
   2 : qt         : 9.360e-02
   3 : cpa        : 8.976e-02
   4 : dEdx2      : 8.416e-02
   5 : dEdx1      : 7.222e-02
   6 : R          : 7.028e-02
   7 : etadiff    : 6.443e-02
   8 : abscospsi  : 6.072e-02
   9 : dca1       : 5.382e-02
  10 : dDCA       : 5.217e-02
  11 : dca2       : 4.924e-02
  12 : ncl2       : 4.918e-02
  13 : chi2       : 4.666e-02
  14 : alpha      : 4.489e-02
  15 : ncl1       : 3.403e-02
```



Cut based method and BDT comparison

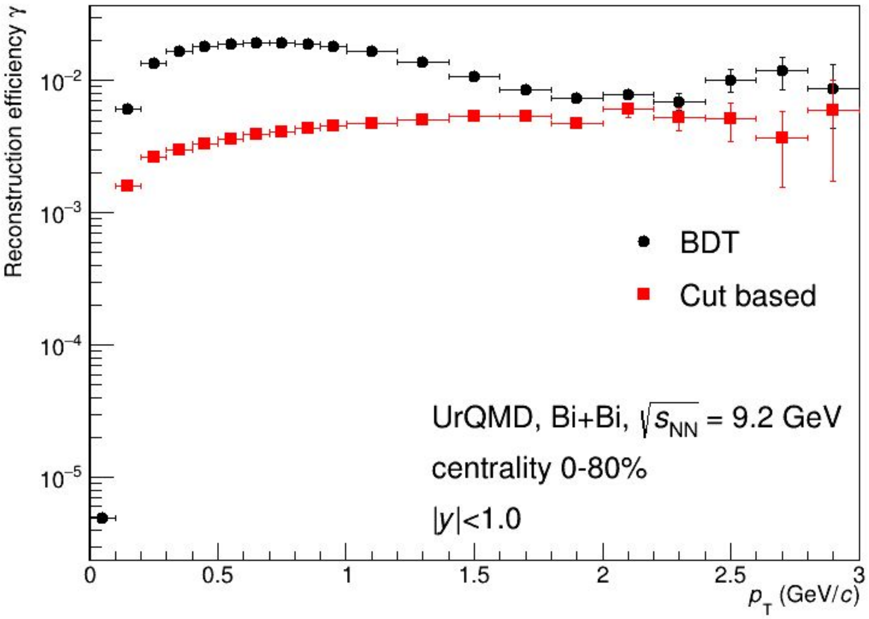
- Cut efficiencies and purities were calculated for Cut based method (with default values) and BDT method

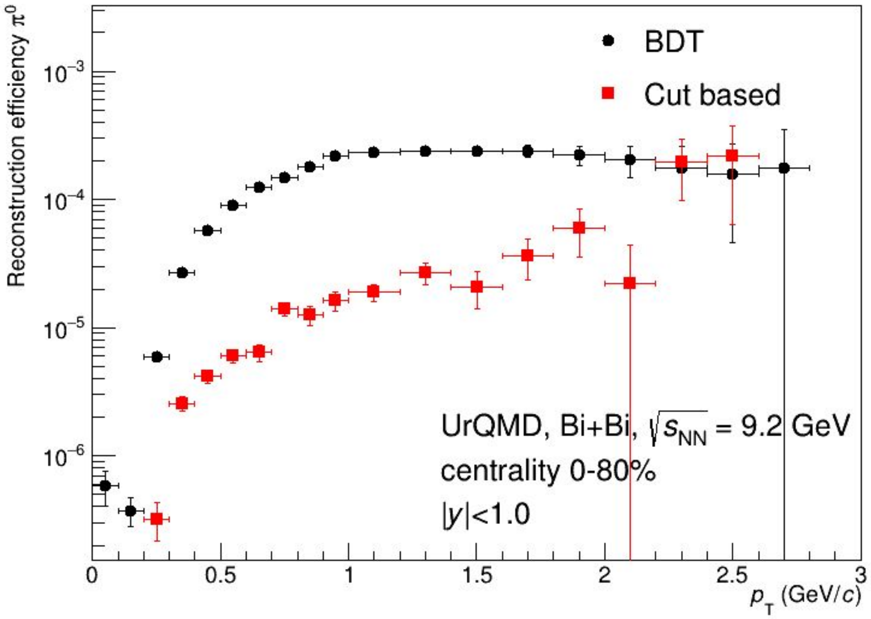




Cut based method and BDT comparison

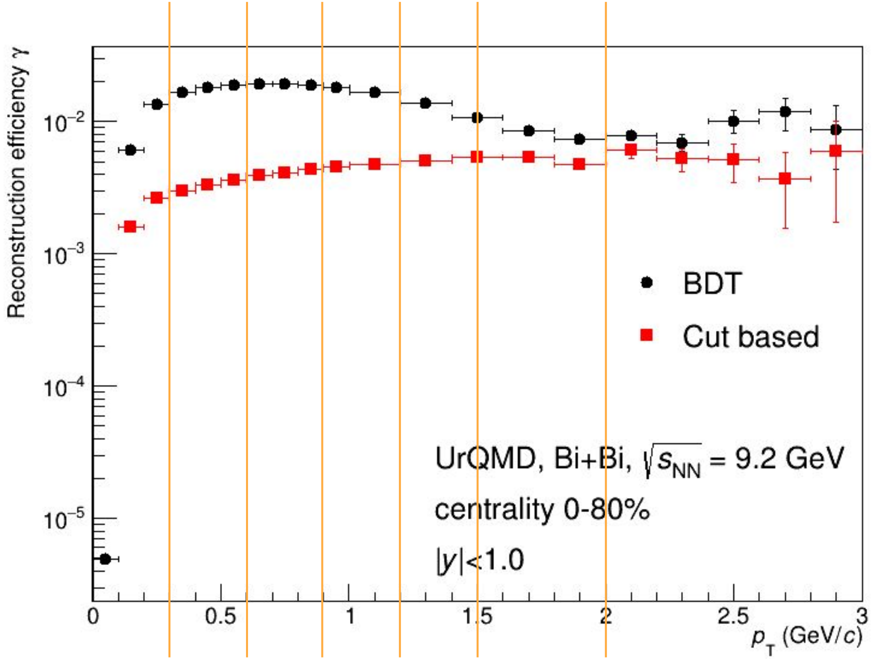
- Reconstruction efficiencies were also calculated for ɣ and pi0





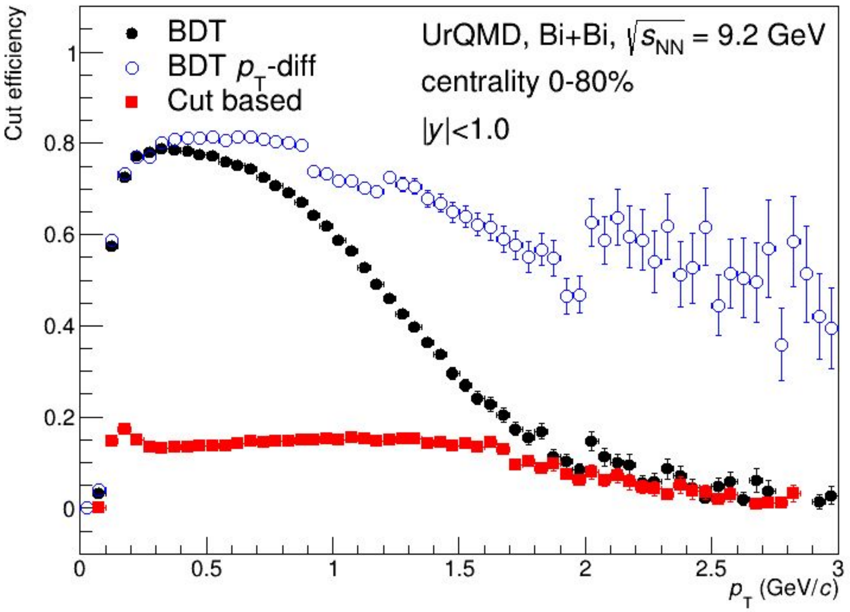
p_T - differential training

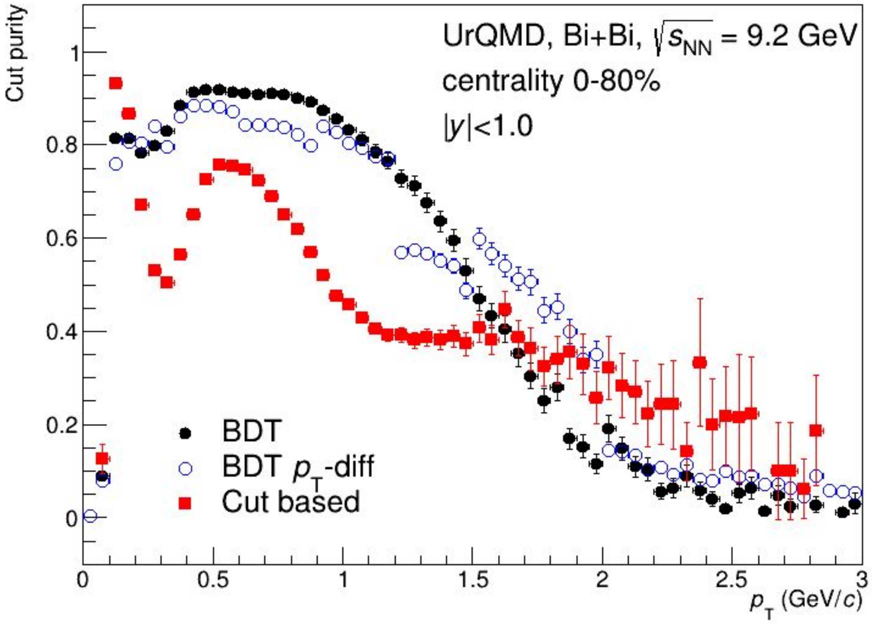
- Reconstruction efficiency for photons rapidly decreases from p_T = 0.7 GeV/c
- p_T differential training should solve this issue
- Selected p_T intervals for training:
- 0.0-0.3
- 0.3-0.6
- 0.6-0.9
- 0.9-1.2
- 1.2-1.5
- 1.5-2.0
- >2.0



Cut based method and BDT comparison

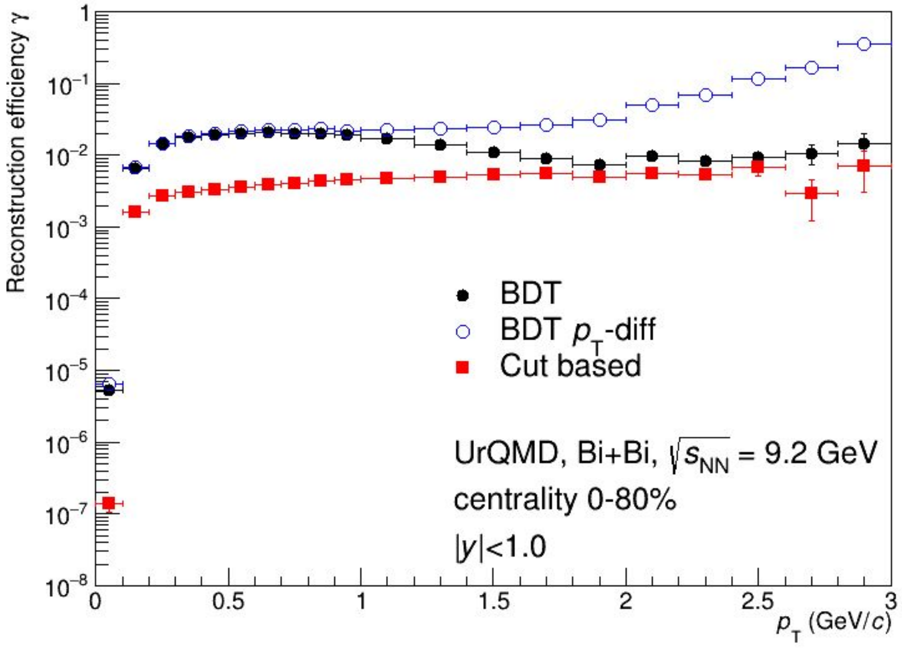
- Efficiency and purity have steps-like structure (need to be fixed)

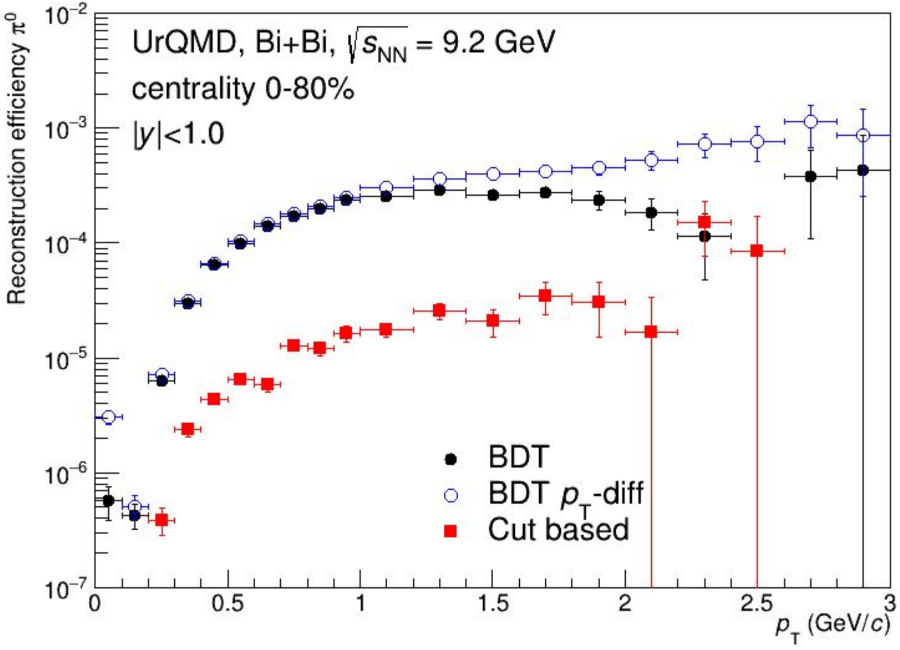




Cut based method and BDT comparison

- Reconstruction efficiency increased for high p_T with differential training

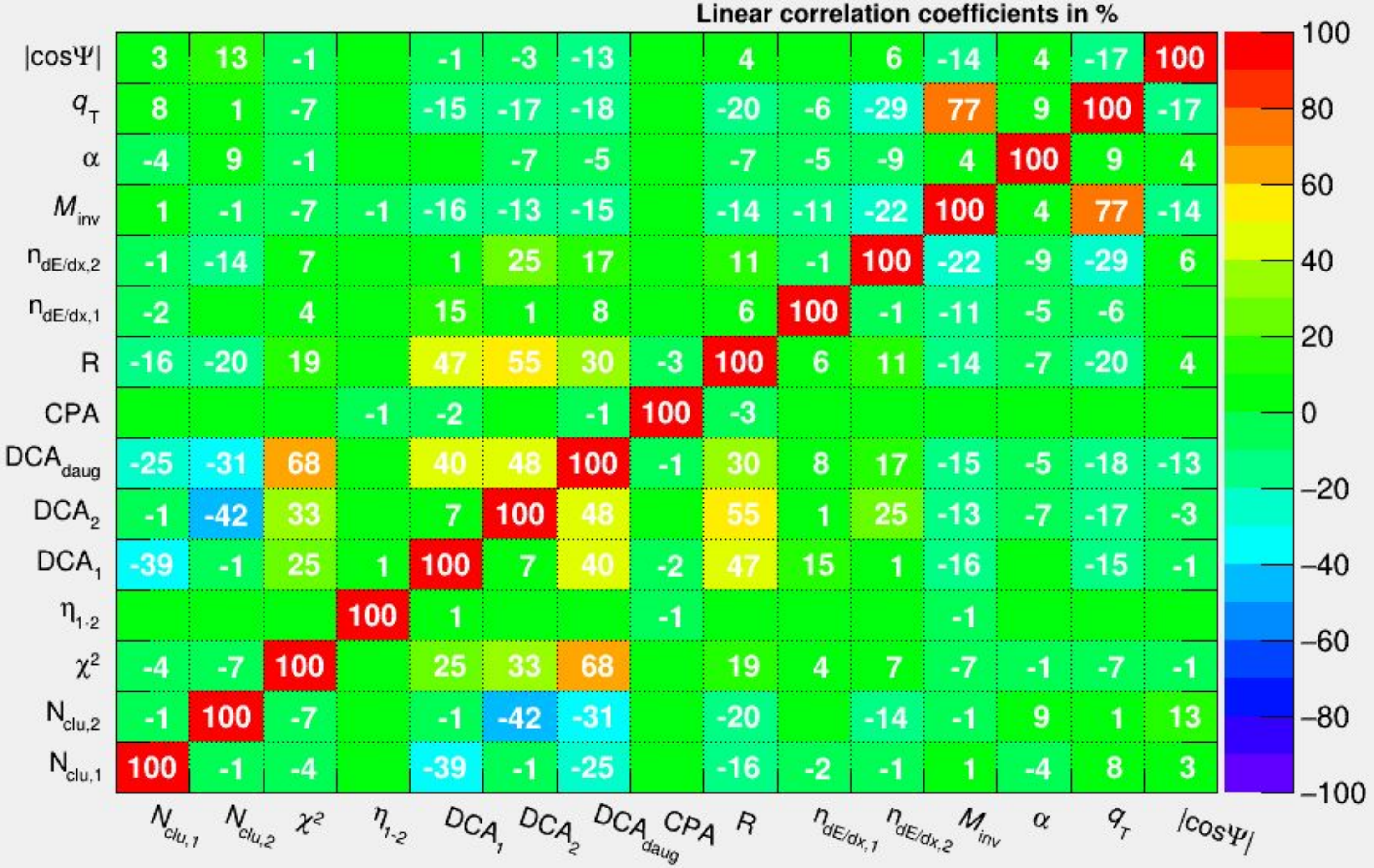


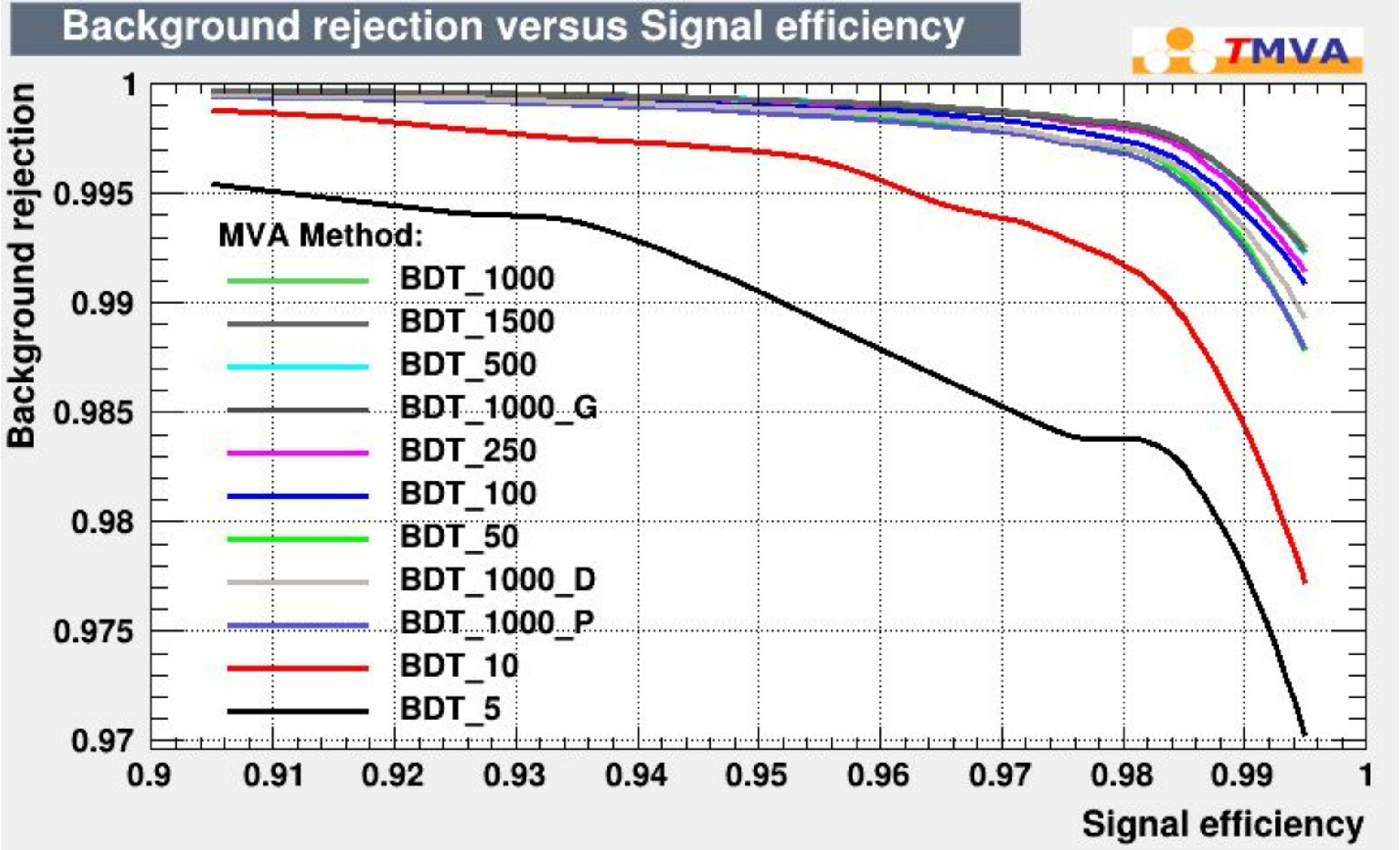


Conclusion

- Performance of BDT method is better than Cut based method, however default selections were not fully optimized
- p_T differential approach shows better reconstruction efficiency for higher p_T, but steps-like structure should be fixed
- BDT also have parameters that can be optimized (N_{trees}, etc.)

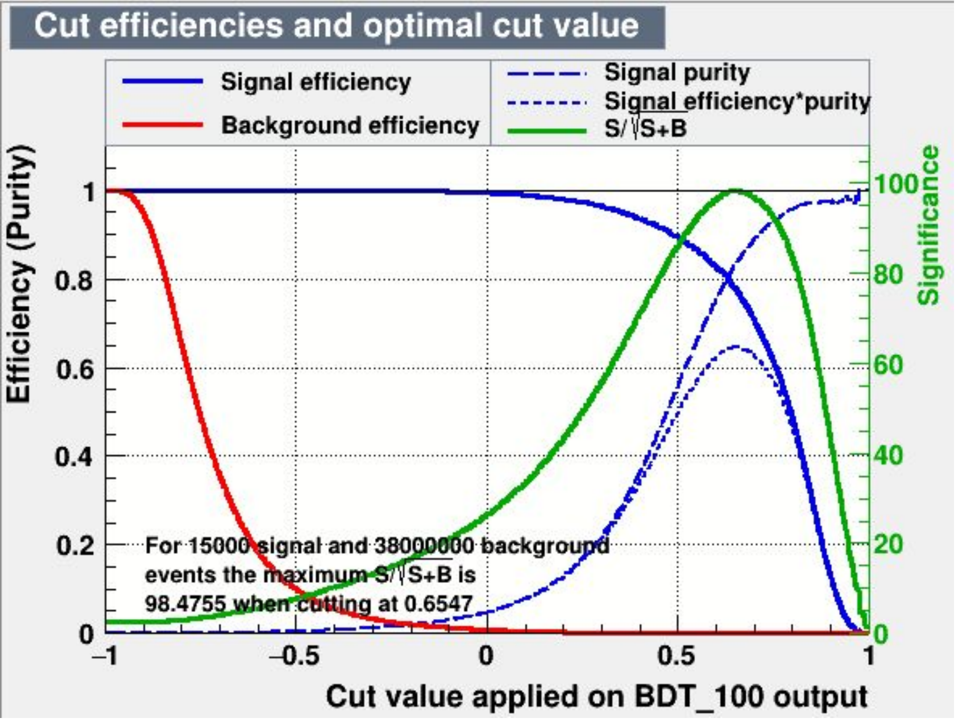
Correlation Matrix (background)

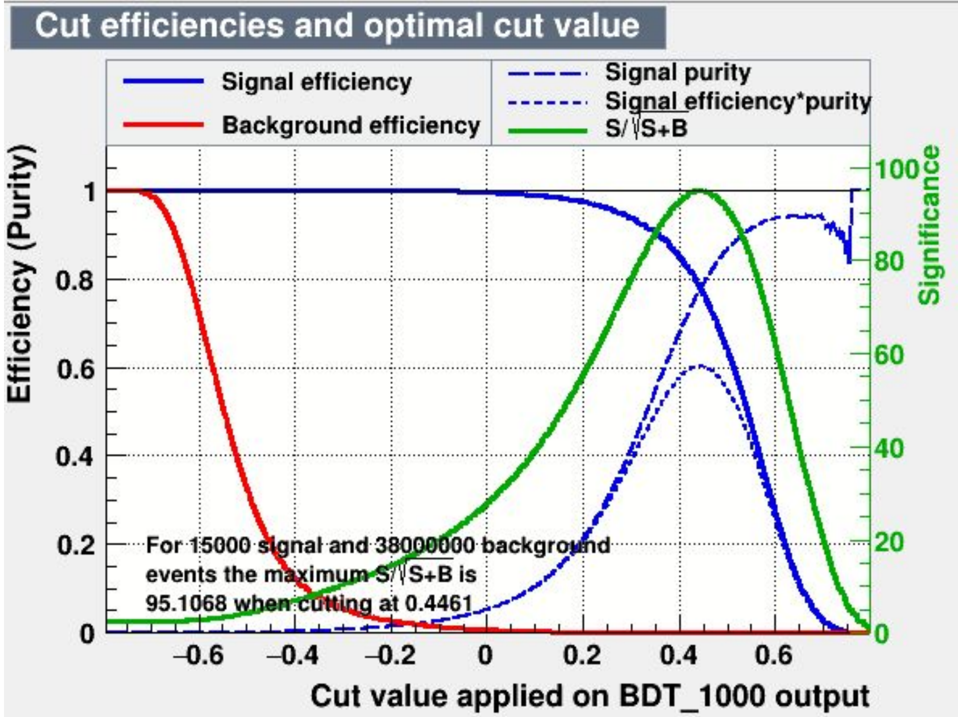




Background rejection versus Signal efficiency
TMVA
Background rejection
Signal efficiency
1
0.995
0.99
0.985
0.98
0.975
0.97
0.9
0.91
0.92
0.93
0.94
0.95
0.96
0.97
0.98
0.99
1
MVA Method:
BDT_1000
BDT_1500
BDT_500
BDT_1000_G
BDT_250
BDT_100
BDT_50
BDT_1000_D
BDT_1000_P
BDT_10
BDT_5

Classifier	(#signal,	#backgr.)	Optimal-cut	S/sqrt(S+B)	NSig	NBkg	EffSig	EffBkg
BDT_5:	(15000,	38000000)	0.6963	42.9949	13065	79274	0.871	0.002086
BDT_10:	(15000,	38000000)	0.8767	90.5816	10694	3244	0.7129	8.537e-05
BDT_50:	(15000,	38000000)	0.7413	97.9905	11412	2151	0.7608	5.661e-05
BDT_100:	(15000,	38000000)	0.6547	98.4755	11608	2287	0.7739	6.018e-05
BDT_250:	(15000,	38000000)	0.5184	96.7394	11966	3334	0.7977	8.774e-05
BDT_500:	(15000,	38000000)	0.4759	95.9206	11709	3192	0.7806	8.4e-05
BDT_1000:	(15000,	38000000)	0.4461	95.1068	11626	3317	0.7751	8.729e-05
BDT_1000_P:	(15000,	38000000)	0.3311	88.7779	10793	3987	0.7195	0.0001049
BDT_1000_D:	(15000,	38000000)	0.3338	88.6486	11006	4408	0.7337	0.000116
BDT_1000_G:	(15000,	38000000)	0.4474	94.0585	11787	3917	0.7858	0.0001031
BDT_1500:	(15000,	38000000)	0.4373	95.0268	11573	3259	0.7715	8.576e-05


Cut efficiencies and optimal cut value
Signal efficiency
Background efficiency
Signal purity
Signal efficiency*purity
S/√S+B
Efficiency (Purity)
Significance
For 15000 signal and 38000000 background
events the maximum S/√S+B is
98.4755 when cutting at 0.6547
Cut value applied on BDT_100 output


Cut efficiencies and optimal cut value
Signal efficiency
Background efficiency
Signal purity
Signal efficiency*purity
S/√S+B
Efficiency (Purity)
Significance
For 15000 signal and 38000000 background
events the maximum S/√S+B is
95.1068 when cutting at 0.4461
Cut value applied on BDT_1000 output

```
Testing efficiency compared to training efficiency (overtraining check)
--------------------------------------------------------------------------------------------
DataSet              MVA              Signal efficiency: from test sample (from training sample)
Name:                Method:          @B=0.01              @B=0.10            @B=0.30
--------------------------------------------------------------------------------------------
dataset              BDT_1000       : 1.000 (1.000)        1.000 (1.000)      1.000 (1.000)
dataset              BDT_1500       : 1.000 (1.000)        1.000 (1.000)      1.000 (1.000)
dataset              BDT_500        : 0.999 (1.000)        1.000 (1.000)      1.000 (1.000)
dataset              BDT_1000_G     : 1.000 (1.000)        1.000 (1.000)      1.000 (1.000)
dataset              BDT_250        : 0.997 (1.000)        1.000 (1.000)      1.000 (1.000)
dataset              BDT_100        : 0.997 (1.000)        1.000 (1.000)      1.000 (1.000)
dataset              BDT_50         : 0.992 (0.994)        1.000 (1.000)      1.000 (1.000)
dataset              BDT_1000_D     : 0.994 (1.000)        1.000 (1.000)      1.000 (1.000)
dataset              BDT_1000_P     : 0.992 (0.997)        1.000 (1.000)      1.000 (1.000)
dataset              BDT_10         : 0.983 (0.987)        1.000 (1.000)      1.000 (1.000)
dataset              BDT_5          : 0.952 (0.962)        1.000 (1.000)      1.000 (1.000)
--------------------------------------------------------------------------------------------
```

```
Evaluation results ranked by best signal efficiency and purity (area)
----------------------------------------------------------------------
DataSet       MVA
Name:         Method:            ROC-integ
dataset       BDT_1000         : 1.000
dataset       BDT_1500         : 1.000
dataset       BDT_500          : 1.000
dataset       BDT_1000_G       : 1.000
dataset       BDT_250          : 1.000
dataset       BDT_100          : 1.000
dataset       BDT_50           : 1.000
dataset       BDT_1000_D       : 1.000
dataset       BDT_1000_P       : 1.000
dataset       BDT_10           : 0.999
dataset       BDT_5            : 0.997
----------------------------------------------------------------------
```

p_T - diff

		S tree	B tree	S train	B train	S test	B test
pt0	<0.3	69’539	96’721’034	15’000	15’000	1’000	1’400’000
pt1	0.3-0.6	108’357	275’641’881	15’000	15’000	1’000	2’600’000
pt2	0.6-0.9	79’975	480’041’325	15’000	15’000	1’000	6’000’000
pt3	0.9-1.2	20’019	284’515’986	15’000	15’000	1’000	14’200’000
pt4	1.2-1.5	6’049	129’631’686	5’000	5’000	1’000	21’400’000
pt5	1.5-2.0	2’757	67’519’977	1’700	1’700	1’000	24’500’000
pt6	>2.0	2’957	28’242’052	1’900	1’900	1’000	9’500’000

0.00-0.30

Classifier	(#signal,	#backgr.)	Optimal-cut	S/sqrt(S+B)	NSig	NBkg	EffSig	EffBkg
BDT_5:	(1000,	1400000)	0.6609	18.4241	835	1219	0.835	0.0008707
BDT_10:	(1000,	1400000)	0.8807	21.5665	713	380	0.713	0.0002714
BDT_25:	(1000,	1400000)	0.8111	25.5044	754	120	0.754	8.571e-05
BDT_50:	(1000,	1400000)	0.7377	25.6242	761	121	0.761	8.643e-05
BDT_100:	(1000,	1400000)	0.5899	25.8562	794	149	0.794	0.0001064
BDT_250:	(1000,	1400000)	0.5013	25.6488	760	118	0.76	8.429e-05
BDT_500:	(1000,	1400000)	0.4086	25.273	800	202	0.8	0.0001443

0.30-0.60

Classifier	#signal	#backgr.	Optimal-cut	S/sqrt(S+B)	NSig	NBkg	EffSig	EffBkg
BDT_5:	1000	2600000	0.6731	17.0036	873	1763	0.873	0.0006781
BDT_10:	1000	2600000	0.8645	24.6615	719	131	0.719	5.038e-05
BDT_25:	1000	2600000	0.8309	25.4956	775	149	0.775	5.731e-05
BDT_50:	1000	2600000	0.7585	25.8118	788	144	0.788	5.538e-05
BDT_100:	1000	2600000	0.6801	26.2401	792	119	0.792	4.577e-05
BDT_250:	1000	2600000	0.5809	26.1513	809	148	0.809	5.692e-05
BDT_500:	1000	2600000	0.5189	26.1649	817	158	0.817	6.077e-05

0.60-0.90

Classifier	#signal	#backgr.	Optimal-cut	S/sqrt(S+B)	NSig	NBkg	EffSig	EffBkg
BDT_5:	1000	6000000	0.6865	11.3898	871	4977	0.871	0.0008295
BDT_10:	1000	6000000	0.8619	22.5848	736	326	0.736	5.433e-05
BDT_25:	1000	6000000	0.8539	24.3422	755	207	0.755	3.45e-05
BDT_50:	1000	6000000	0.7631	24.8194	769	191	0.769	3.183e-05
BDT_100:	1000	6000000	0.6696	25.2709	813	222	0.813	3.7e-05
BDT_250:	1000	6000000	0.6172	25.4006	798	189	0.798	3.15e-05
BDT_500:	1000	6000000	0.5592	25.1344	782	186	0.782	3.1e-05

0.90-1.20

Classifier	(#signal, #backgr.)	Optimal-cut	S/sqrt(S+B)	NSig	NBkg	EffSig	EffBkg
BDT_5:	(1000, 14200000)	0.6423	6.80652	857	14996	0.857	0.001056
BDT_10:	(1000, 14200000)	0.8583	19.8638	654	430	0.654	3.028e-05
BDT_25:	(1000, 14200000)	0.8511	21.8873	739	401	0.739	2.824e-05
BDT_50:	(1000, 14200000)	0.7877	22.9439	747	313	0.747	2.204e-05
BDT_100:	(1000, 14200000)	0.7256	22.4872	700	269	0.7	1.894e-05
BDT_250:	(1000, 14200000)	0.6046	22.3752	778	431	0.778	3.035e-05
BDT_500:	(1000, 14200000)	0.5763	21.8762	747	419	0.747	2.951e-05

1.20-1.50

Classifier	(#signal, #backgr.)	Optimal-cut	S/sqrt(S+B)	NSig	NBkg	EffSig	EffBkg
BDT_5:	(1000, 21400000)	0.6469	8.73559	817	7930	0.817	0.0003706
BDT_10:	(1000, 21400000)	0.8545	13.9663	730	2002	0.73	9.355e-05
BDT_25:	(1000, 21400000)	0.8695	16.7643	682	973	0.682	4.547e-05
BDT_50:	(1000, 21400000)	0.8461	19.5016	628	409	0.628	1.911e-05
BDT_100:	(1000, 21400000)	0.7487	19.1221	707	660	0.707	3.084e-05
BDT_250:	(1000, 21400000)	0.7104	19.9571	687	498	0.687	2.327e-05
BDT_500:	(1000, 21400000)	0.6619	19.6531	699	566	0.699	2.645e-05

1.50-2.00

Classifier	#signal	#backgr.	Optimal-cut	S/sqrt(S+B)	NSig	NBkg	EffSig	EffBkg
BDT_5:	1000	24500000	0.6809	3.59991	856	55685	0.856	0.002273
BDT_10:	1000	24500000	0.8295	8.89618	706	5592	0.706	0.0002282
BDT_25:	1000	24500000	0.9273	12.8359	506	1048	0.506	4.278e-05
BDT_50:	1000	24500000	0.9575	13.805	397	430	0.397	1.755e-05
BDT_100:	1000	24500000	0.8835	14.9138	502	631	0.502	2.576e-05
BDT_250:	1000	24500000	0.7859	13.9182	607	1295	0.607	5.286e-05
BDT_500:	1000	24500000	0.7710	13.9886	592	1199	0.592	4.894e-05

>2.00

Classifier	#signal	#backgr.	Optimal-cut	S/sqrt(S+B)	NSig	NBkg	EffSig	EffBkg
BDT_5:	1000	9500000	0.6695	4.92438	760	23059	0.76	0.002427
BDT_10:	1000	9500000	0.8357	7.24165	643	7241	0.643	0.0007622
BDT_25:	1000	9500000	0.8517	8.64029	555	3571	0.555	0.0003759
BDT_50:	1000	9500000	0.7273	9.78958	637	3597	0.637	0.0003786
BDT_100:	1000	9500000	0.6660	9.77289	659	3888	0.659	0.0004093
BDT_250:	1000	9500000	0.5899	9.67643	689	4381	0.689	0.0004612
BDT_500:	1000	9500000	0.5801	9.58559	638	3792	0.638	0.0003992