INSTITUTE FOR COLLIDER PARTICLE PHYSICS
UNIVERSITY OF THE WITWATERSRAND

ATLAS EXPERIMENT

The Search for Dark Photons in $H \rightarrow Z\gamma_D$ using Machine Learning Algorithms

Roja T. Rimer , Karabo Tau, Rachid Mazini, Thabo Lepota, Arnab Laha, Mukesh Kumar, Edward Nkadimeng

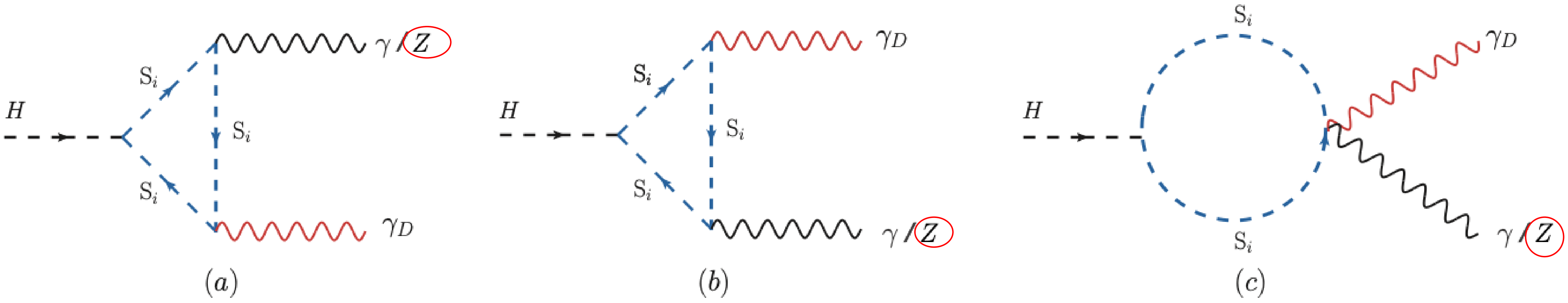
University of the Witwatersrand

NRF National Research Foundation

iThemba LABS Laboratory for Accelerator Based Sciences

Dark Photon

- Hypothetical gauge boson similar to the ordinary photon
- If detected => Hints about Dark Sector Interactions
- In this project it is considered as a Massless Particles
- Hard to detect! – weakly interacting
- Models predict it production via the Higgs portal mechanism



- Several ways that Higgs that can produce a dark photon
- They typically produce $\gamma\gamma_D$, $\gamma\gamma$,$\gamma_D Z$ or $\gamma_D\gamma_D$
- $\gamma\gamma_D$ in the most common method used to search for Dark Photons
- Here we are trying to search using $\gamma_D Z$
  - ➢ New search at the LHC

Higgs Decay to a Dark Photon

Very Preliminary: $H \to Z\gamma_D$

$$BR_{\gamma\gamma} = BR_{\gamma\gamma}^{SM} \frac{\left(1+\chi\sqrt{r_{\gamma\gamma}}\right)^2}{1+r_{\gamma_D\gamma_D}BR_{\gamma\gamma}^{SM}}$$

$$BR_{\gamma\gamma_D} = BR_{\gamma\gamma}^{SM} \frac{r_{\gamma\gamma_D}}{1+r_{\gamma_D\gamma_D}BR_{\gamma\gamma}^{SM}}$$

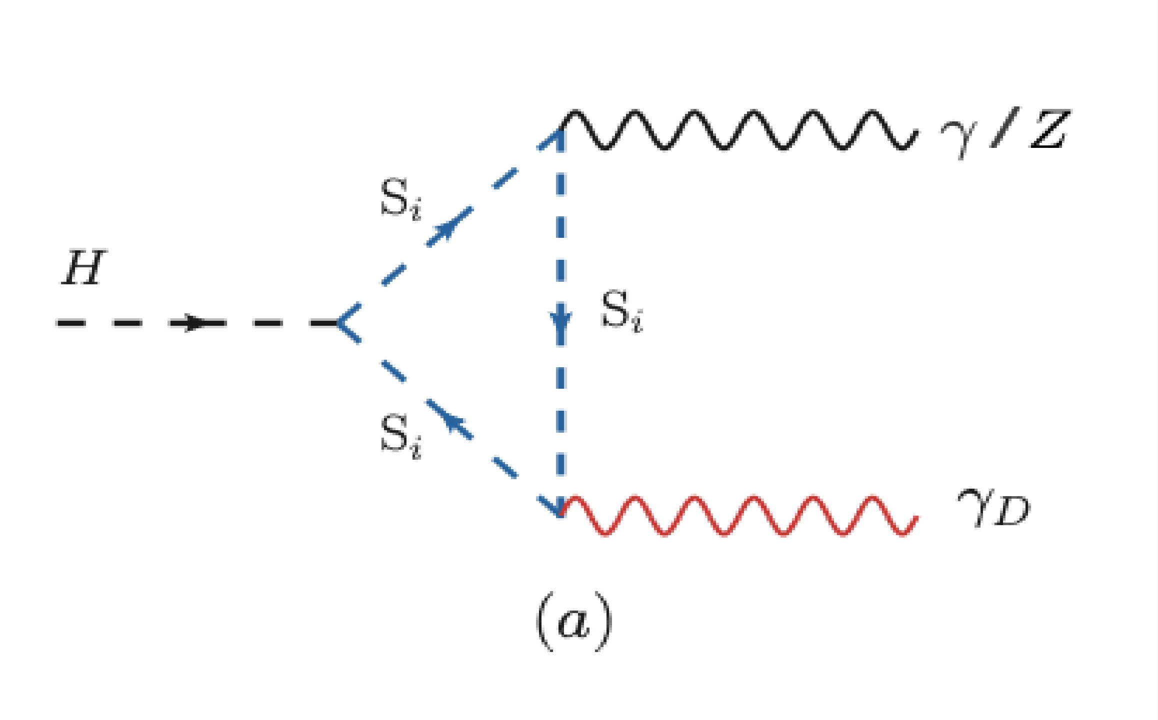
$$BR_{Z\gamma_D} = BR_{\gamma\gamma}^{SM} \frac{r_{Z\gamma_D}}{1+r_{\gamma_D\gamma_D}BR_{\gamma\gamma}^{SM}}$$

$$BR_{\gamma_D\gamma_D} = BR_{\gamma\gamma}^{SM} \frac{r_{\gamma_D\gamma_D}}{1+r_{\gamma_D\gamma_D}BR_{\gamma\gamma}^{SM}}$$

$$\mathrm{r}_{\gamma\gamma_D} = 2X^2\left(\frac{\alpha_D}{\alpha}\right)^2 \qquad \mathrm{r}_{\gamma_D\gamma_D} = X^2\left(\frac{\alpha_D}{\alpha}\right)^2$$

$$\mathrm{r}_{Z\gamma_D} = 2X^2R_{Z\gamma}^2\left(\frac{\alpha_D}{\alpha}\right)^2 \qquad r_{\gamma\gamma} = X^2$$

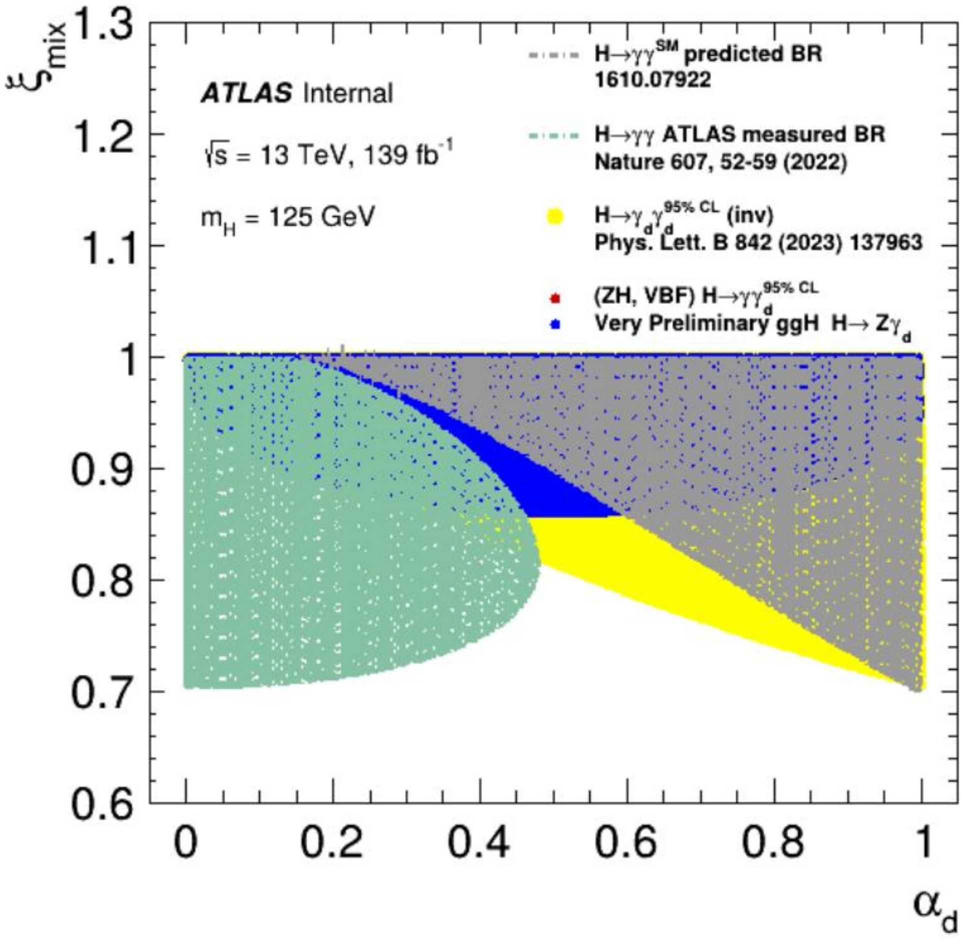
$$X = \frac{\zeta^2}{3F(1-\zeta^2)} \qquad \zeta = \frac{\Delta}{\bar{\mathrm{m}}^2}$$



- Minimal Model
  - 2 Scalar Messengers that act as a bridge between SM and the dark sector
  - Contribute through quantum loop processes
    - Generate both $H \to \gamma\gamma_D$ and $H \to Z\gamma_D$
- Where
  - $R_{Z\gamma}$ - parametrizes the ratio between the amplitudes of and $H \to Z\gamma_D$ over the $H \to \gamma\gamma_D$ one
  - θ_W - is the Weinberg angle
  - Q^i - the electric charge in unit of e,
  - Y^i - the corresponding hypercharge of the scalar fields contributing to the effective coupling

Very Preliminary: $H \to Z\gamma_D$

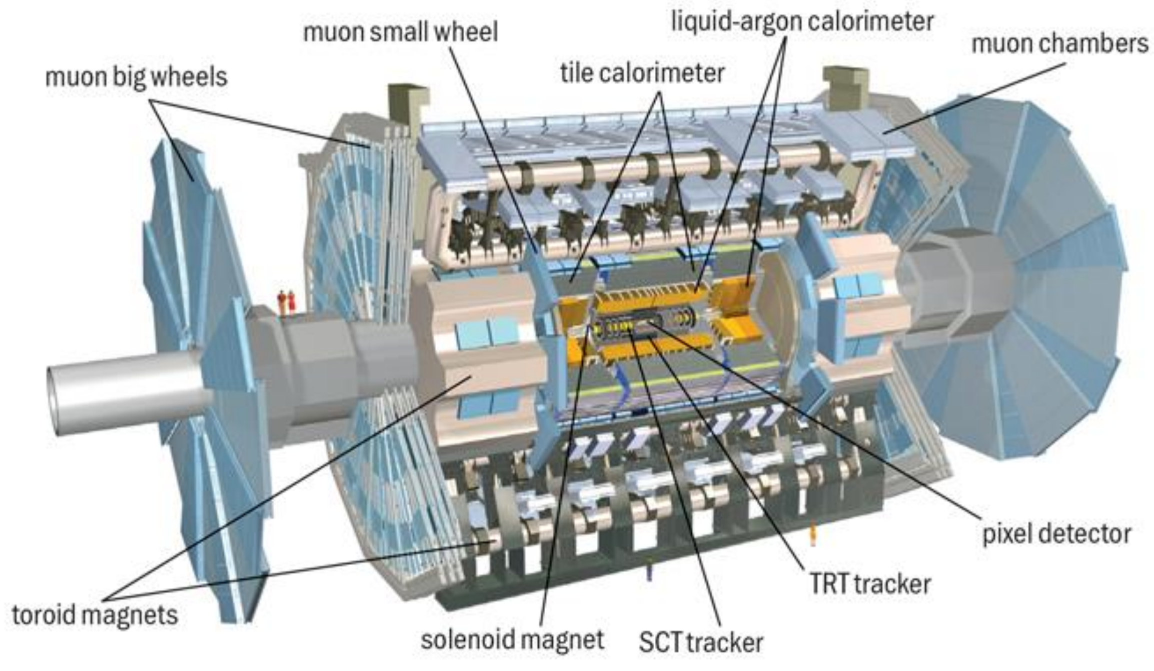
- $R_{Z\gamma}$ - depends on the nature and quantum numbers of the scalar messengers running in the loops.
  - $R_{Z\gamma}$ could range from 0.045 up to 1.3.
- A model-independent approach could also be used for BR($H \to Z\gamma_D$) interpretation.
  - Assuming BR($H \to Z\gamma_D$) < 1% and an average $R_{Z\gamma}$ value



Final State and Data

$$H \rightarrow \gamma_{\mathrm{D}} Z$$
$$\mathrm{Z} \rightarrow \mathrm{l}^{+}\mathrm{l}^{-}$$

- Final state is $\mathrm{l}^{+}\mathrm{l}^{-}$ + MET
  - l = e/μ and MET from γ_{D}
- Monte Carlo Samples used
  - ➢ Signals
    - ❖ggH
    - ❖VBFH
  - ➢ Backgrounds
    - ❖llvv
    - ❖ttbar
    - ❖Z+jets



ATLAS and the LHC

- LHC
    - Proton-proton collider
    - Center of Mass Energy: 13-14TeV
    - Very high luminosity
        - Luminosity
            - measurement of the performance of the collider
            - Determines the rate of collisions
- ATLAS
    - One of the multipurpose detectors at the LHC
    - High precision tracking system
    - Compact hydronic electromagnetic
- calorimeters
    - High coverage muon spectrometer
    - High efficiency
        - Particle identification
        - Energy resolution
        - MET measurements
    - Widely used for
        - Standard Model precision measurements
        - New physics searches

Machine Learning

- Main challenge: distinguishing a rare dark-photon signal
- Machine learning techniques:
  - Rather than relying solely on traditional cut-based methods
- Boosted Decision Trees
  - Operated via a combination of several decision trees
  - Commonly used in High Energy Physics
  - Two main types
    1. ADABoost
    2. GradBoost
- Recurrent Neural Networks
  - Designed for sequential data
  - Can take time to run
  - Variable order matters

Pre-selections and Input Variables

Pre-selections
Exactly two leptons $(e^+e^-/\mu^+\mu^-)$
Leading Transverse Momentum $(p_T) > 27GeV$ Sub-leading Transverse Momentum $(p_T) > 15GeV$ Electron Pseudorapidity (η) cut: $\lvert\eta\rvert < 2.47$ excluding the $1.37 < \lvert\eta\rvert < 1.52$ calorimeter crack region Muon Pseudorapidity (η) cut: $\lvert\eta\rvert < 2.5$
Jet Transverse Momentum $\left(p_T{}^{\mathrm{Jet}}\right) > 30GeV$ Jet Pseudorapidity (η) cut: $\lvert\eta\rvert < 4.5$
Dilepton invariant mass: $65\mathrm{GeV} < \mathrm{m_{ll}} < 110\mathrm{GeV}$
Missing Transverse Energy: $\mathrm{MET} > 15\mathrm{GeV}$
$N_{jet} \leq 2$ $N_{b_{Jet}} = 0$

Input Variables
$\Delta R\ lepton\ (l^+l^-)$
Missing transverse Energy Significance MET_{sig}
$\Delta\phi(Z, MET)$
$\Delta\phi(\text{Leading Jets}, MET)$
Sub-leading Lepton Transverse Momentum
Number of Jets
Dilepton Invariant Mass
$\frac{P_{T_{l1}} + P_{T_{l2}}}{\mathrm{m_{t_{ll}}}}$
$\frac{MET}{\mathrm{m_{t_{ll}}}}$

TMVA Configuration

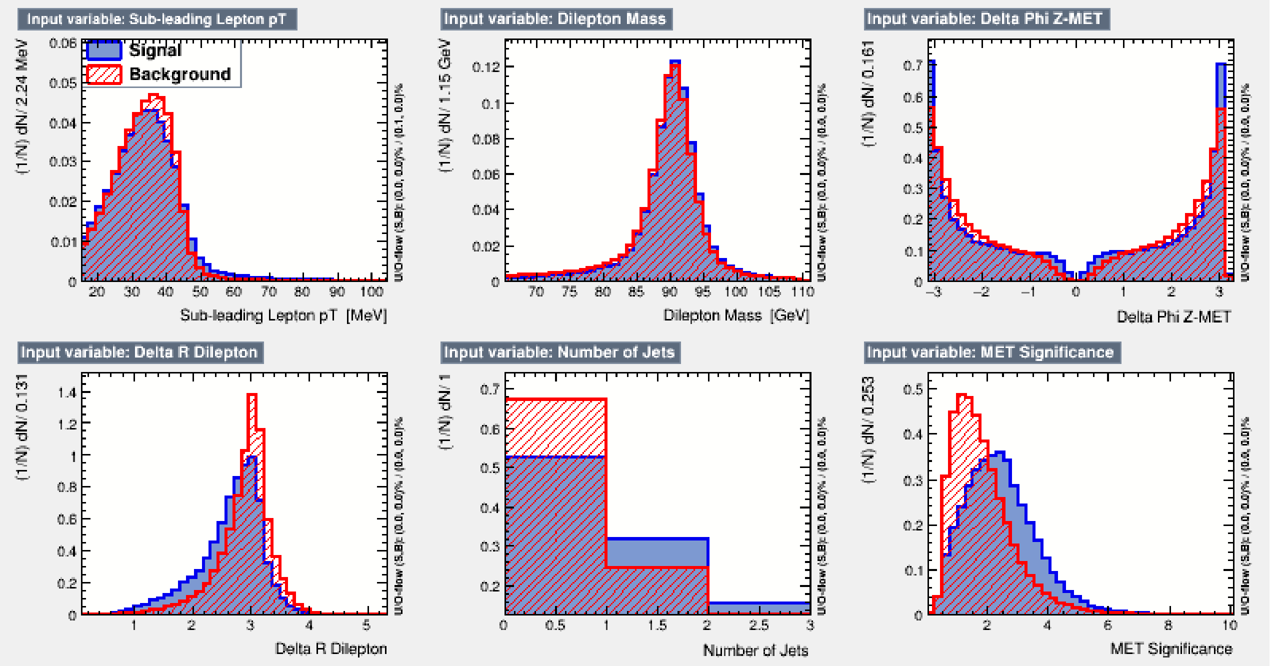
- Toolkit: **ROOT TMVA** using `TMVA(Toolkit for Multivariate Analysis). Factory and TMVA . DataLoader`
- Analysis type: **Classification**
- Event split: 75% training and 25% testing

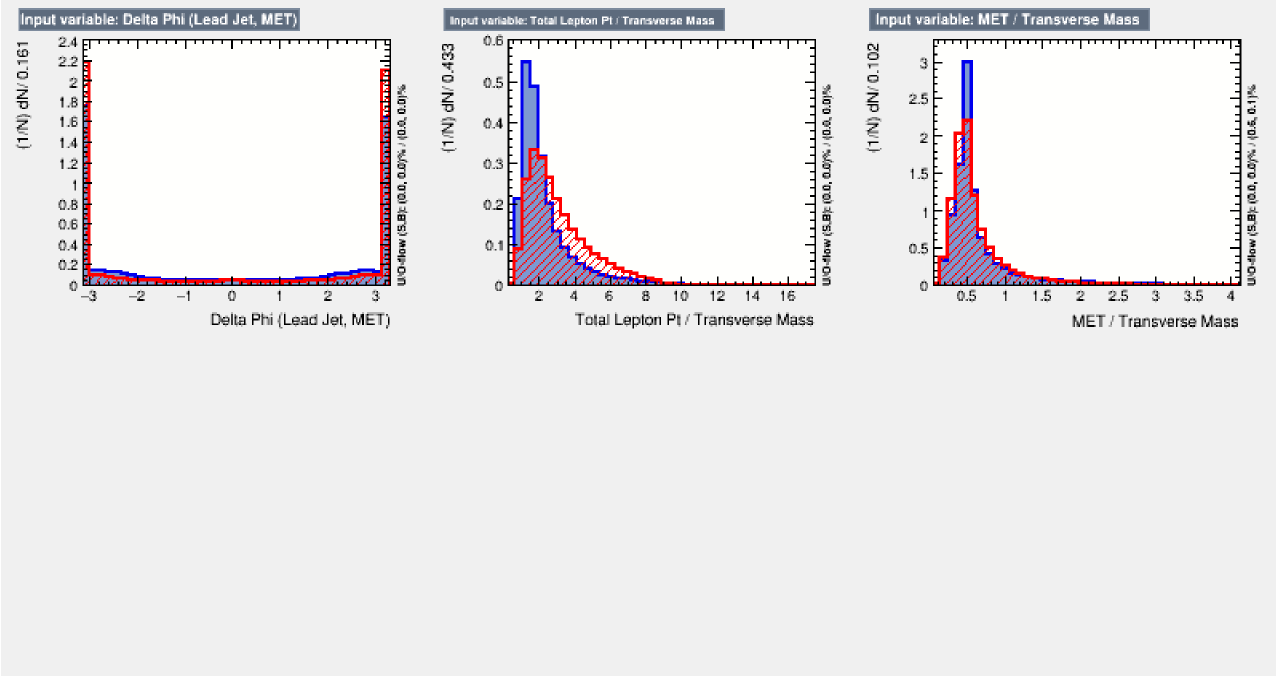
Note:

--BDT & RNN hyperparameters were optimized using the grid search method.

Distributions of input variables

- Both models are trained using the same events and same input variables

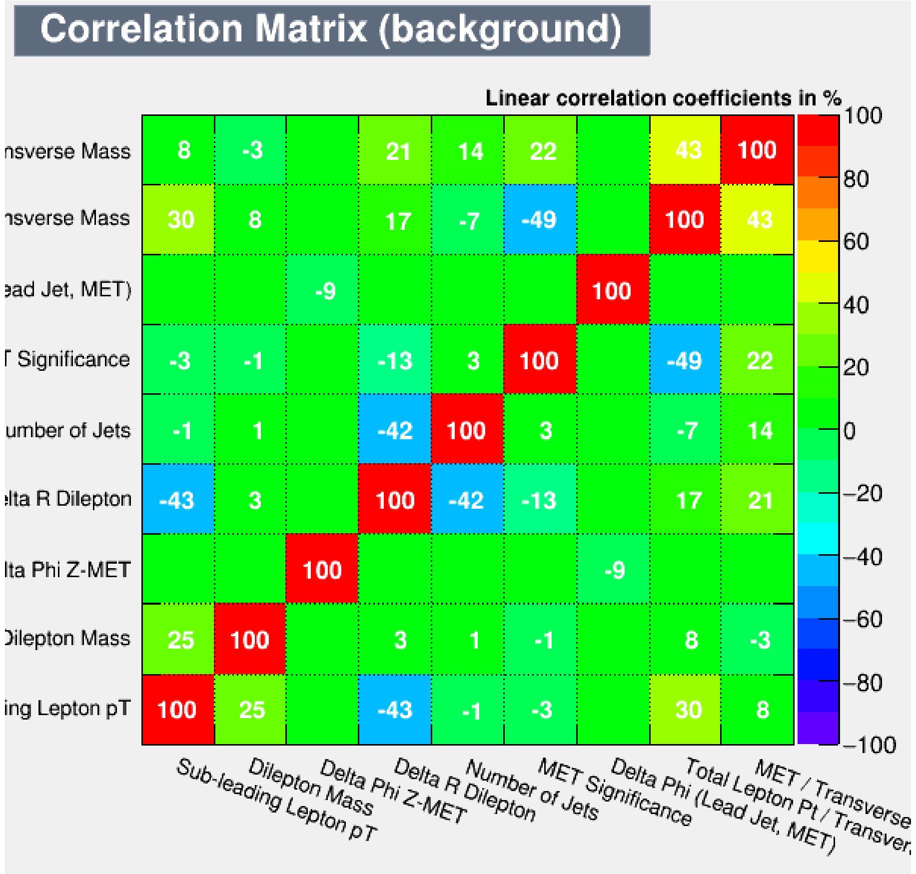




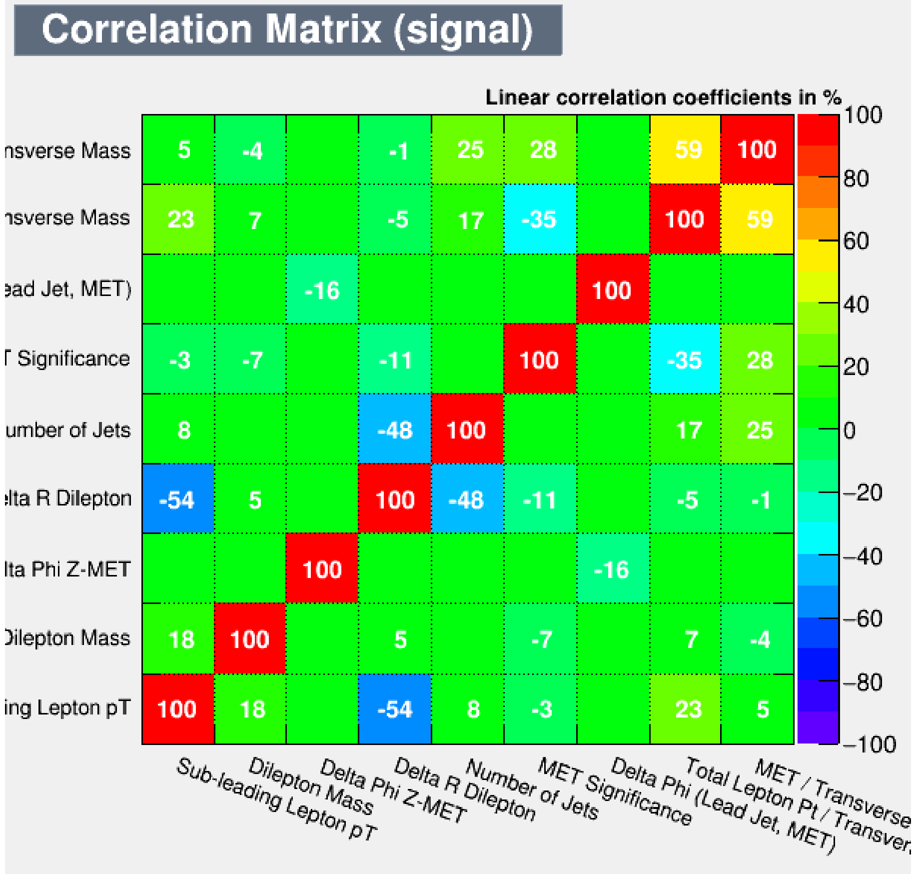
Correlation Matrix

Correlation Matrices for the BDT Training

- Identical to the RNN Matrices



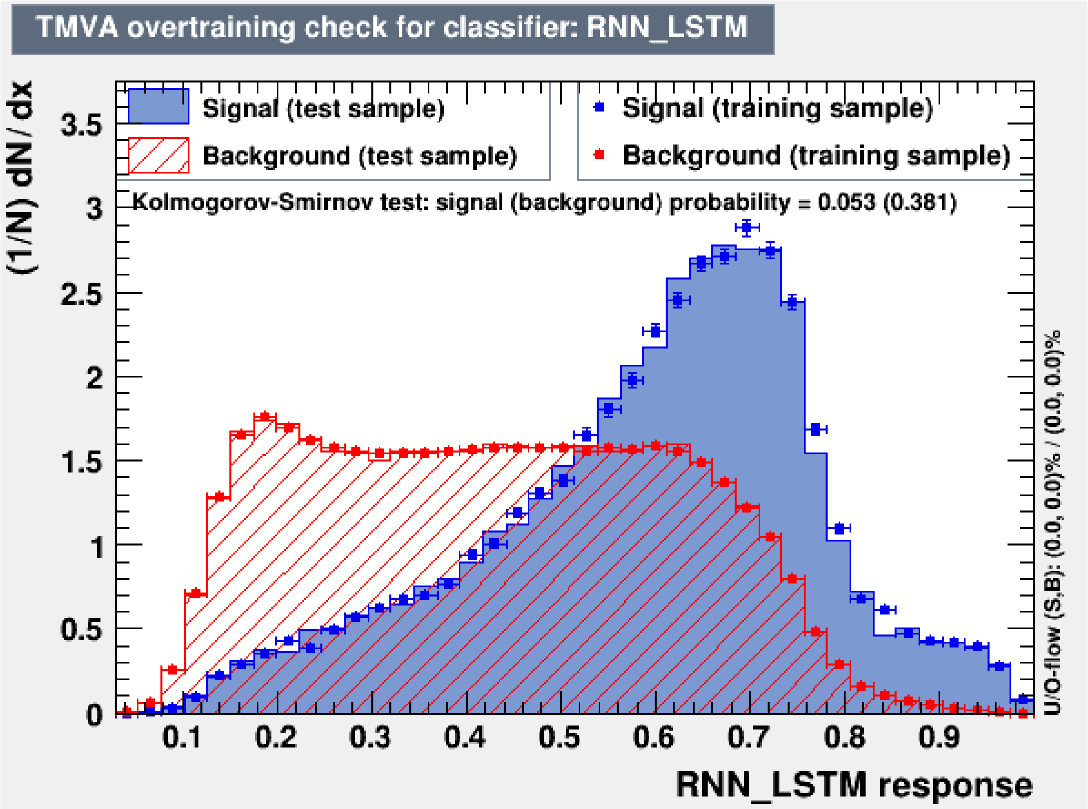
High correlation between $\frac{P_{T_{l1}}+P_{T_{l2}}}{m_t}$ and $\frac{MET}{m_t}$

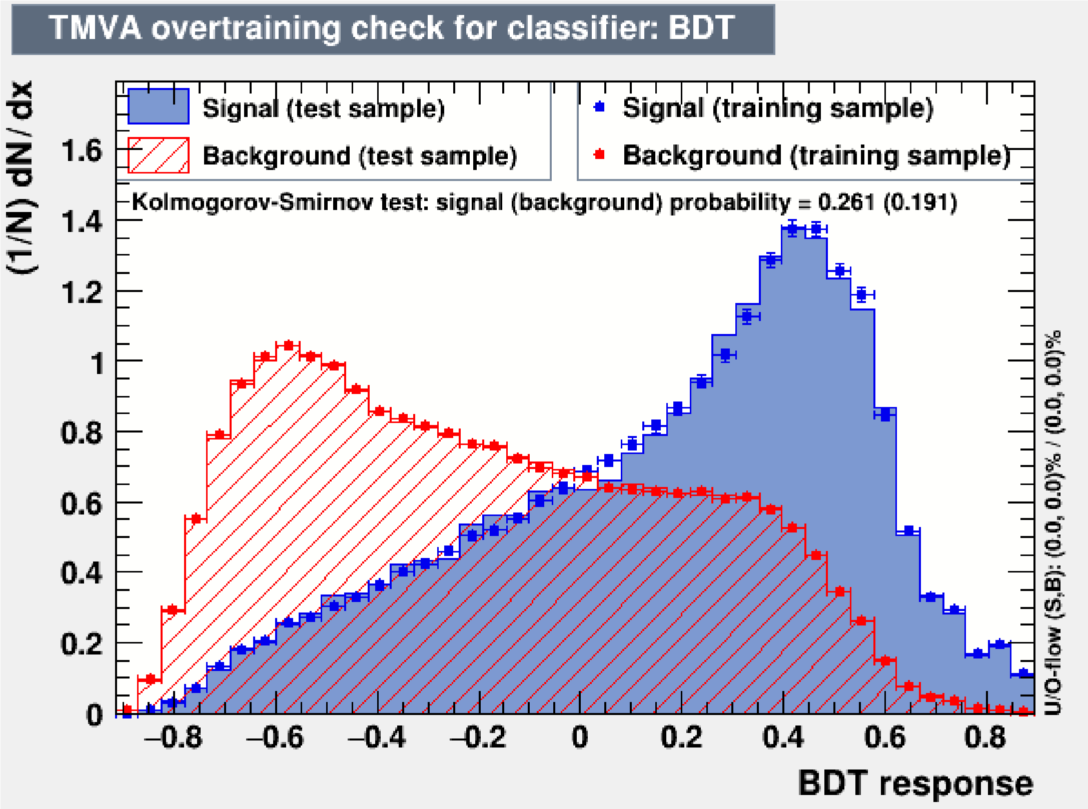


BDT vs RNN Comparison - Testing

Testing check comparison from the training stage

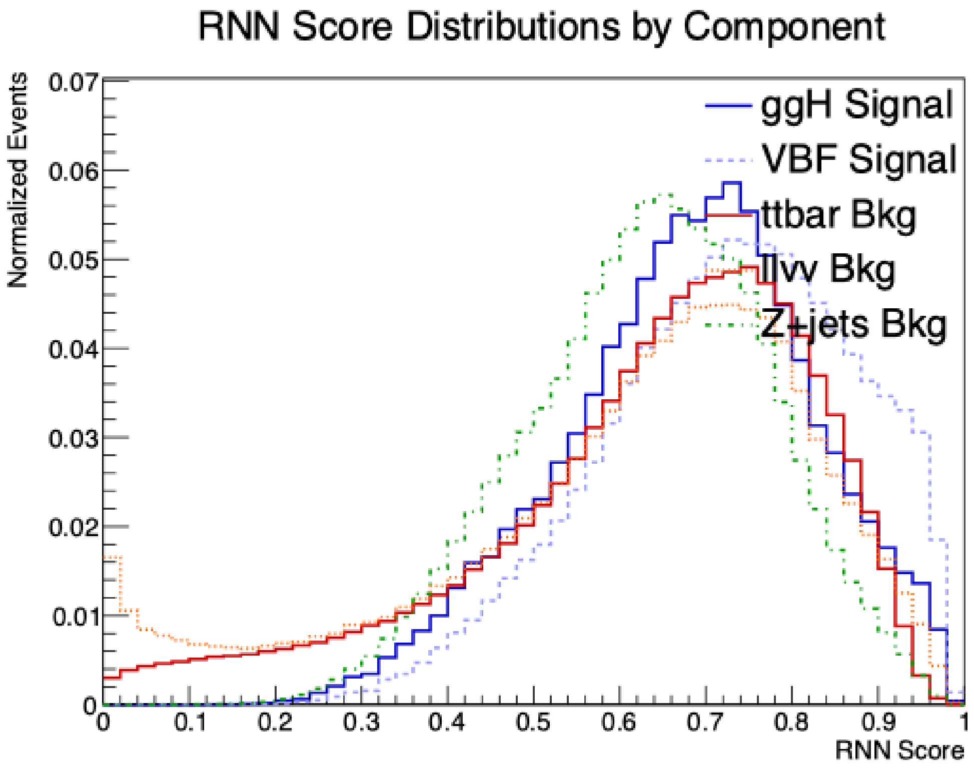
- Training and testing samples both follow each other within each method -> No overtraining

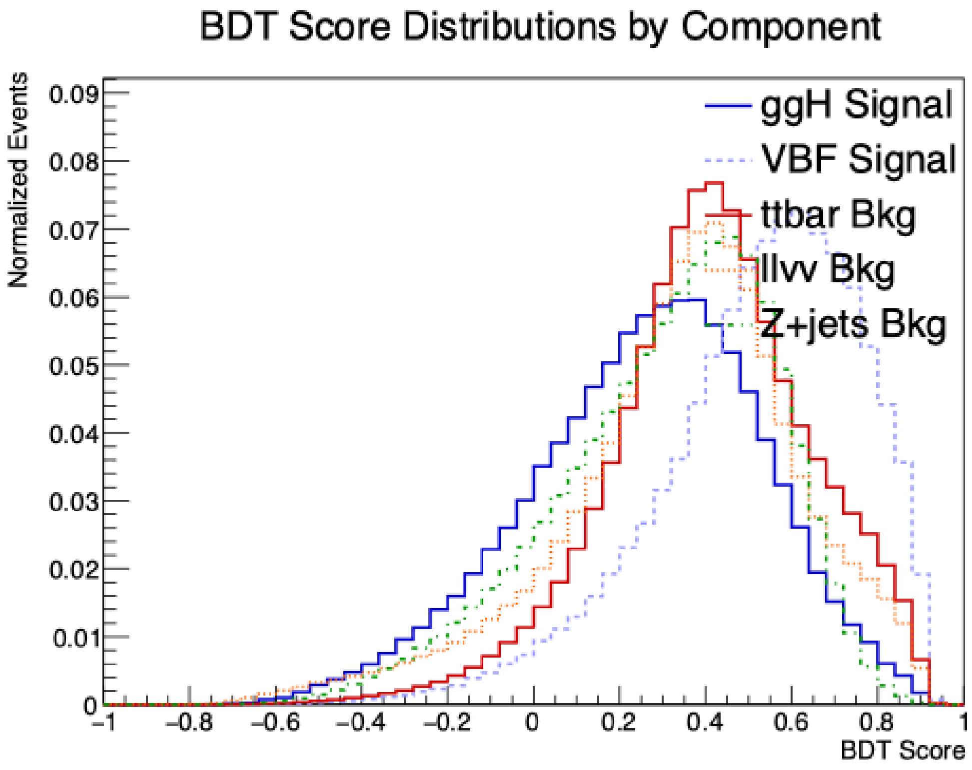




Score Distrubutions

Score Distributions after the training has been applied to the Monte Carlo Samples for the individual signal and background processes

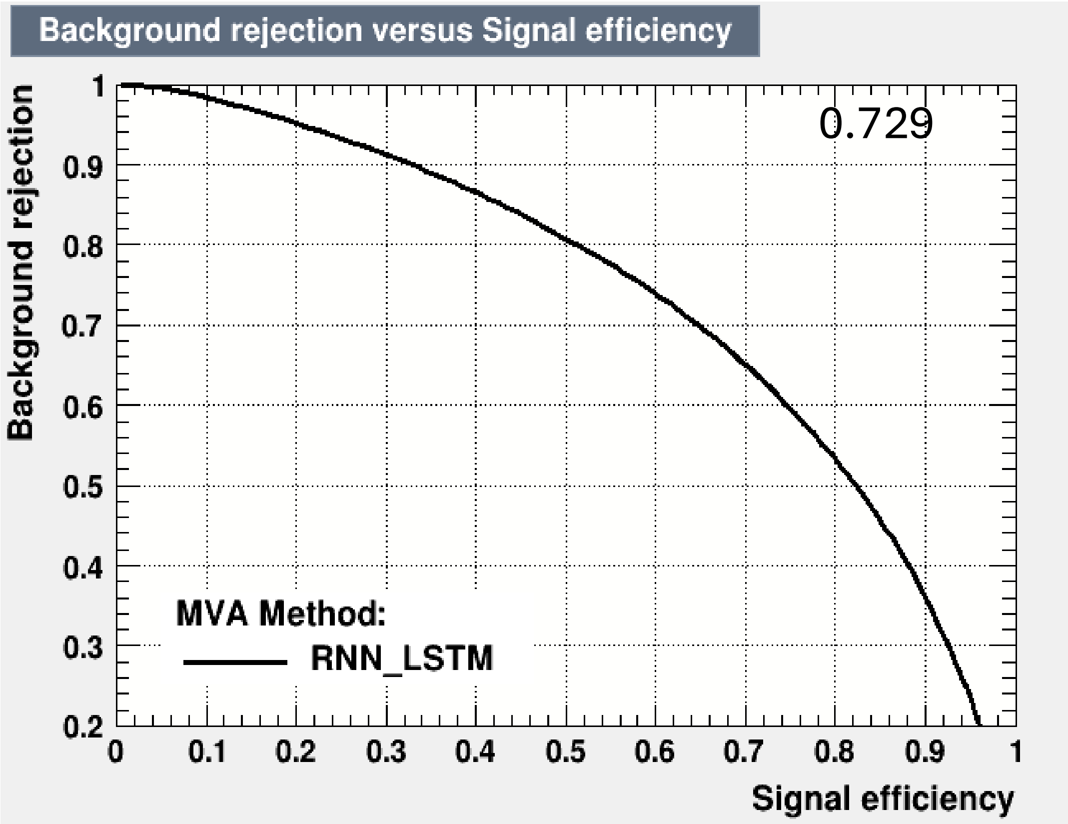


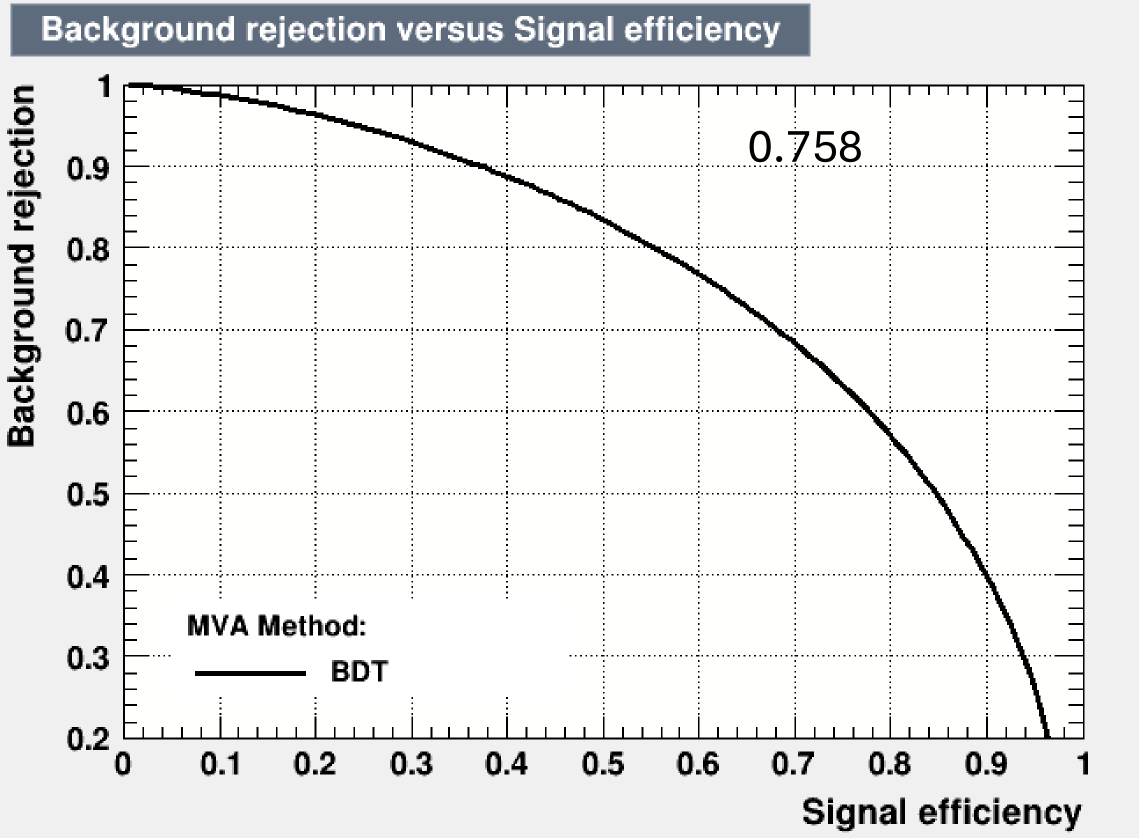


ROC

ROC curves are near identical

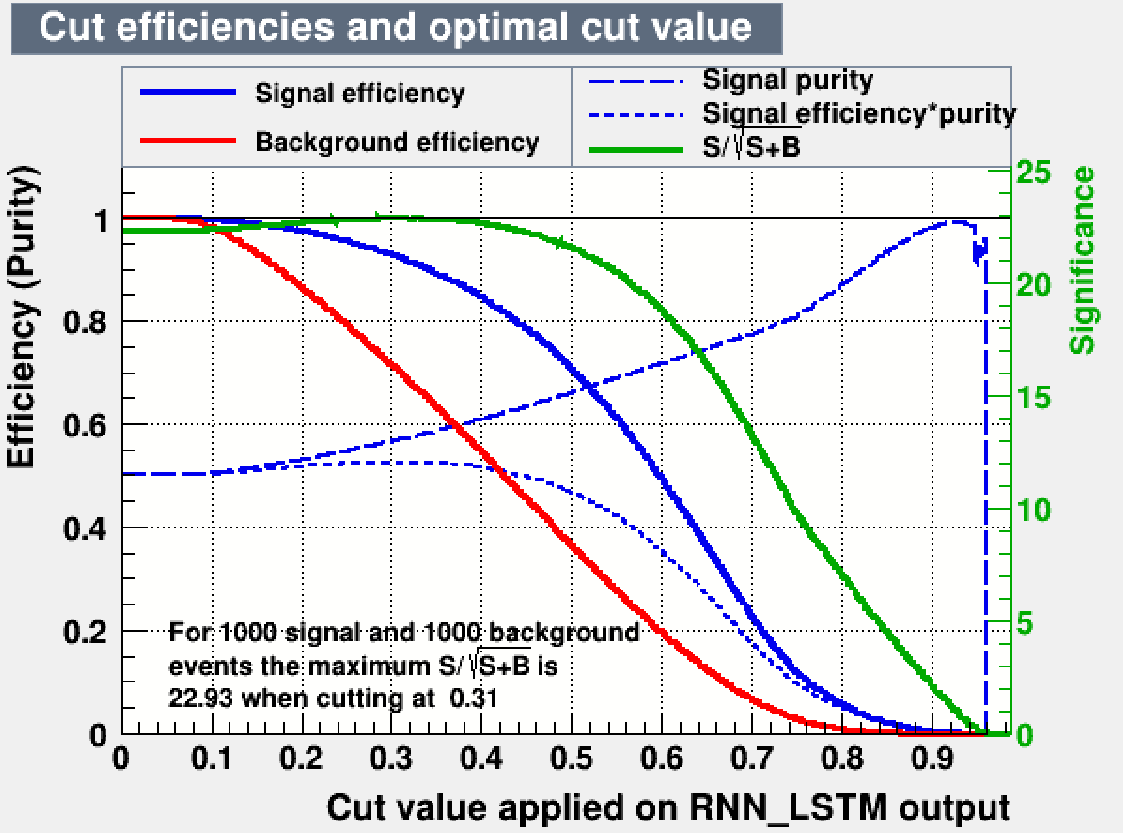
- Have different ROC values
- The inverse curves are also similar
- BDT slightly better (modest improvement in classification performance)

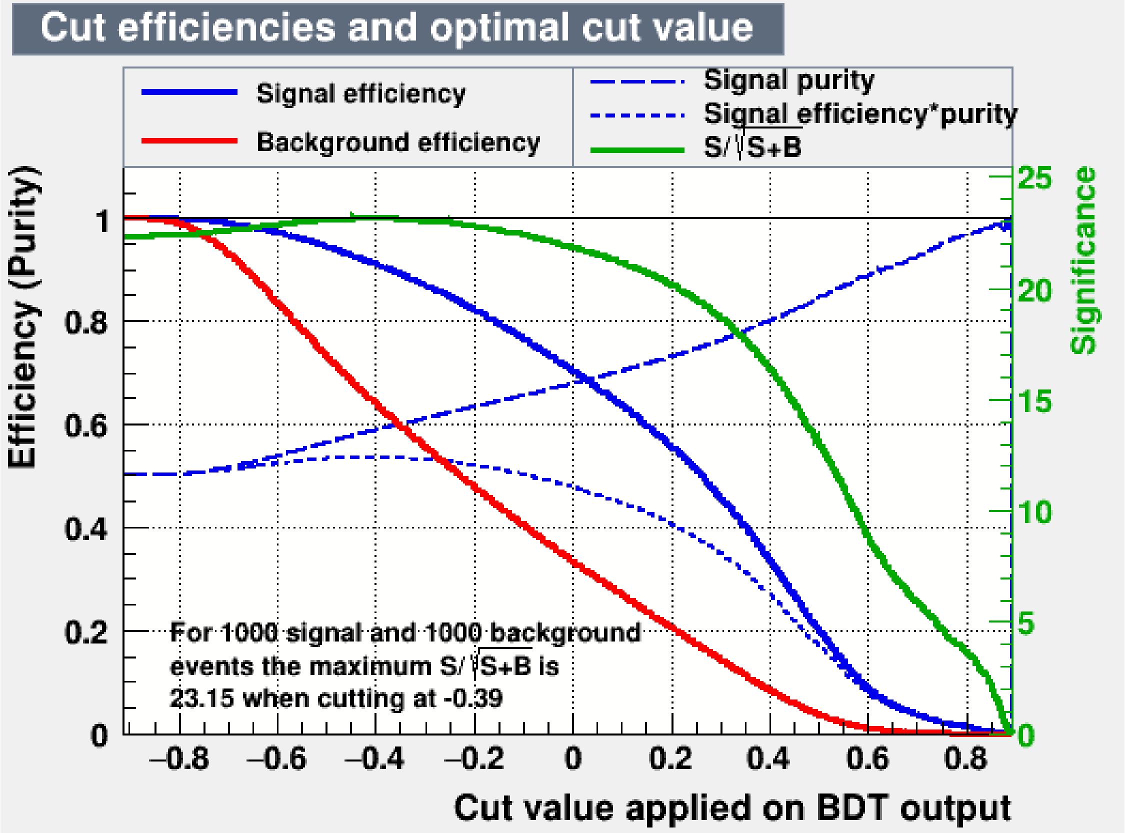




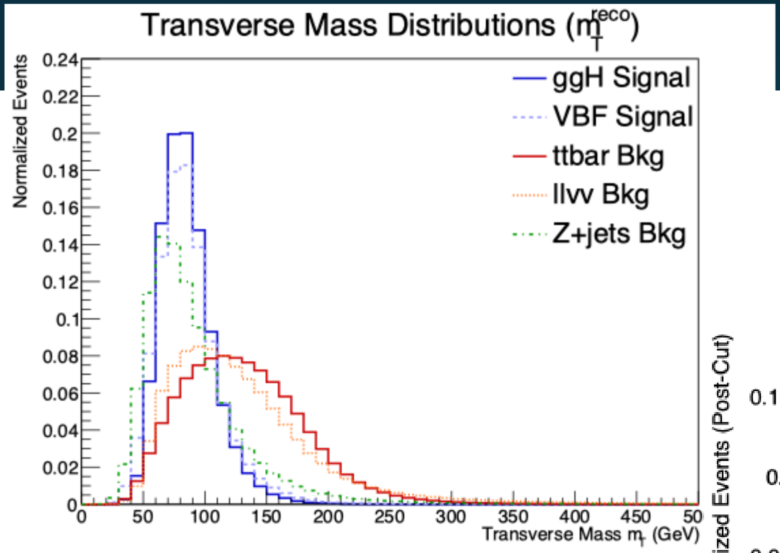
Efficiencies

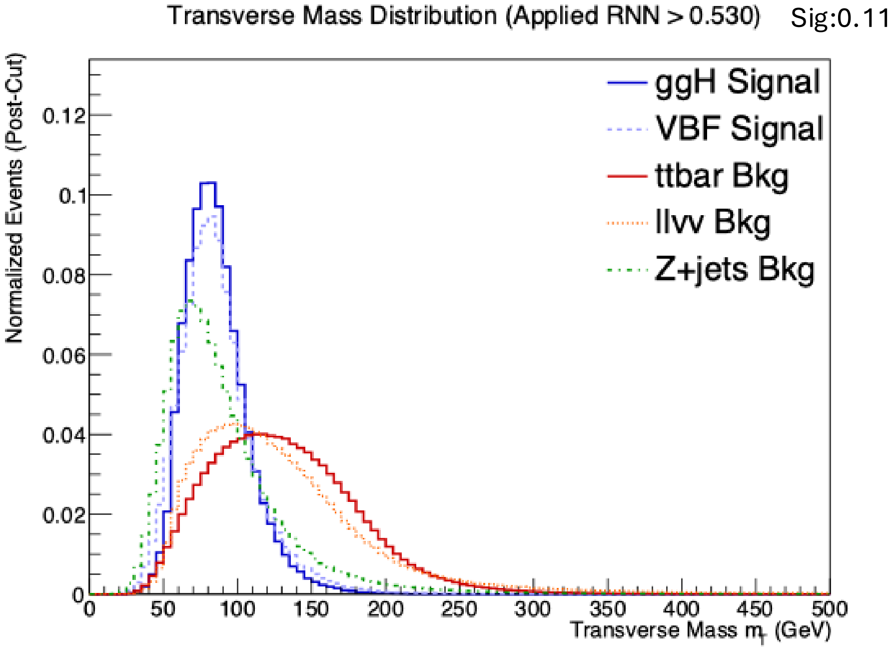
Comparison of the efficiencies as well as the Signal Significance from the training of the models
- Expected and close in value

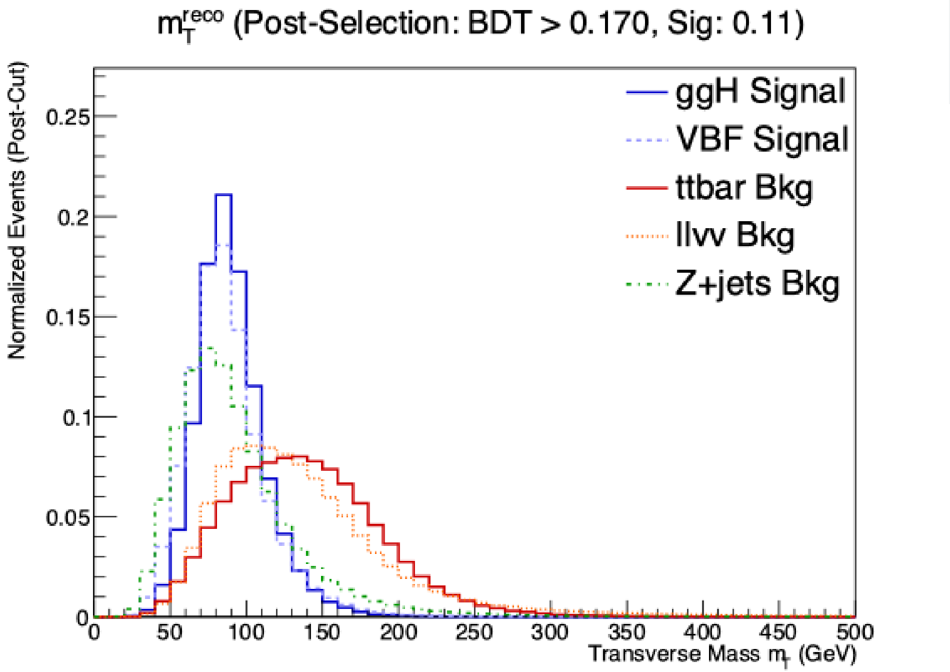




m_t^{reco} Distributions







- Signal processes remain concentrated around the expected m_T^{reco} region.
- Backgrounds are more broadly distributed after the classifier selection.
- The BDT produces a slightly cleaner signal-enriched sample, consistent with the ROC curve results.

Conclusions

- Both the BDT and RNN successfully distinguish the signal from the dominant SM backgrounds, with no evidence of significant overtraining.
- The BDT consistently provides a modest improvement in discrimination performance over the RNN, although both classifiers show comparable overall behaviour.
- These results are preliminary and demonstrate the potential of machine learning techniques for this search, while highlighting areas where further optimisation is still possible.

Future Work

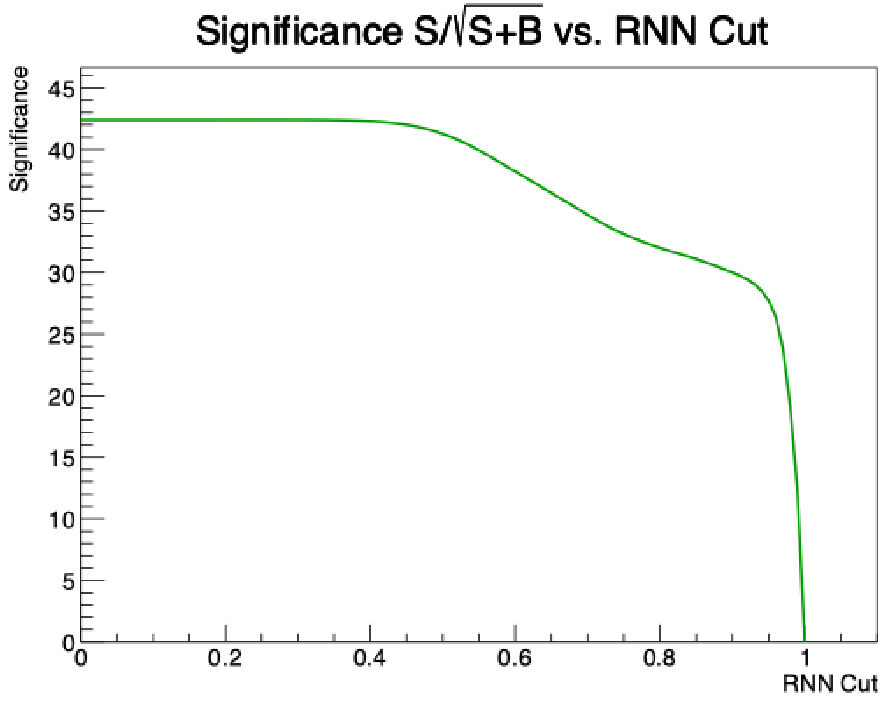
- Optimise the machine learning models and input variables.
- Develop dedicated background control regions.

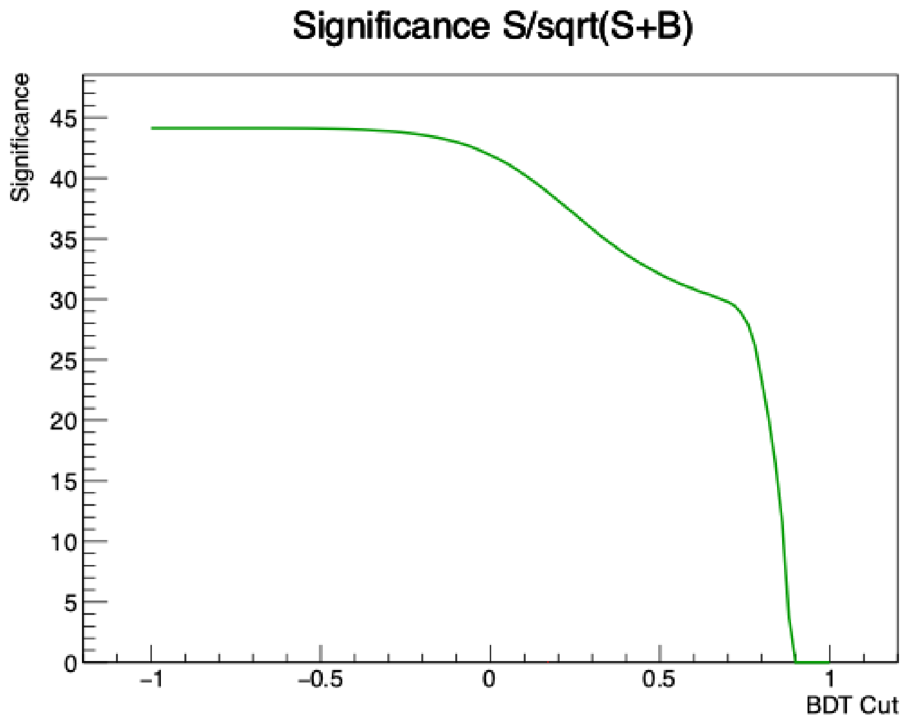
Discussion

Signal Significance

Comparison of the Signal Significance after the model has been applied to the Monte Carlo Samples

- Changes from the training but they both seem to follow a similar pattern

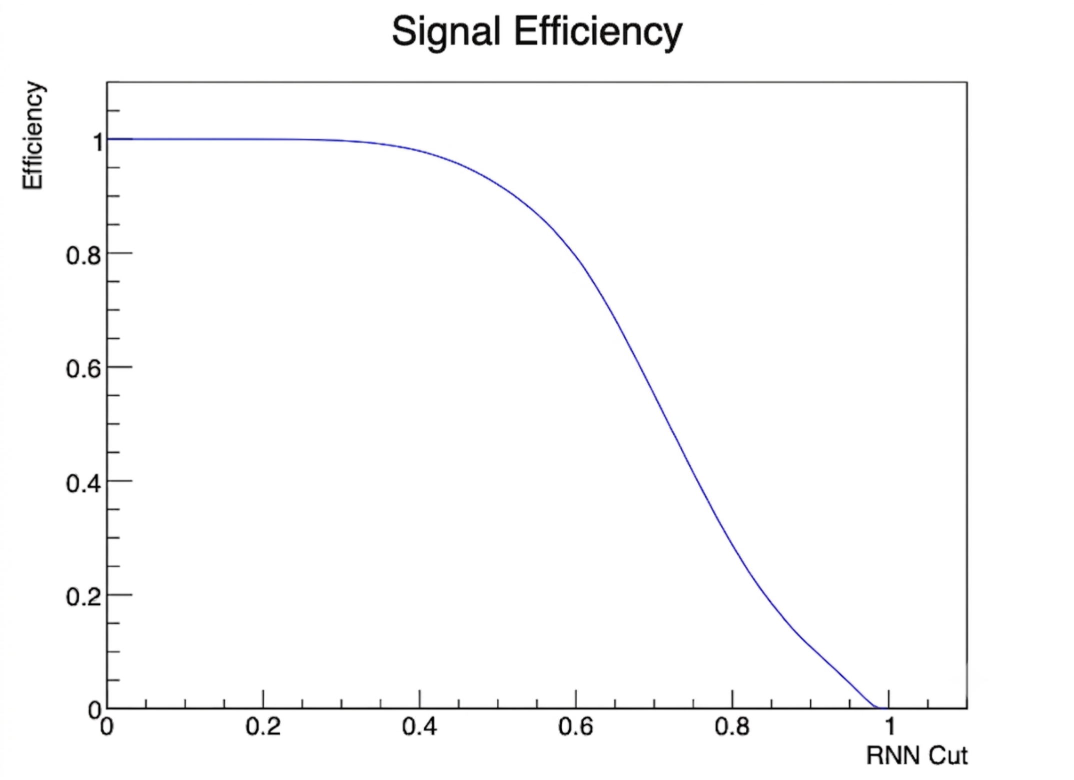


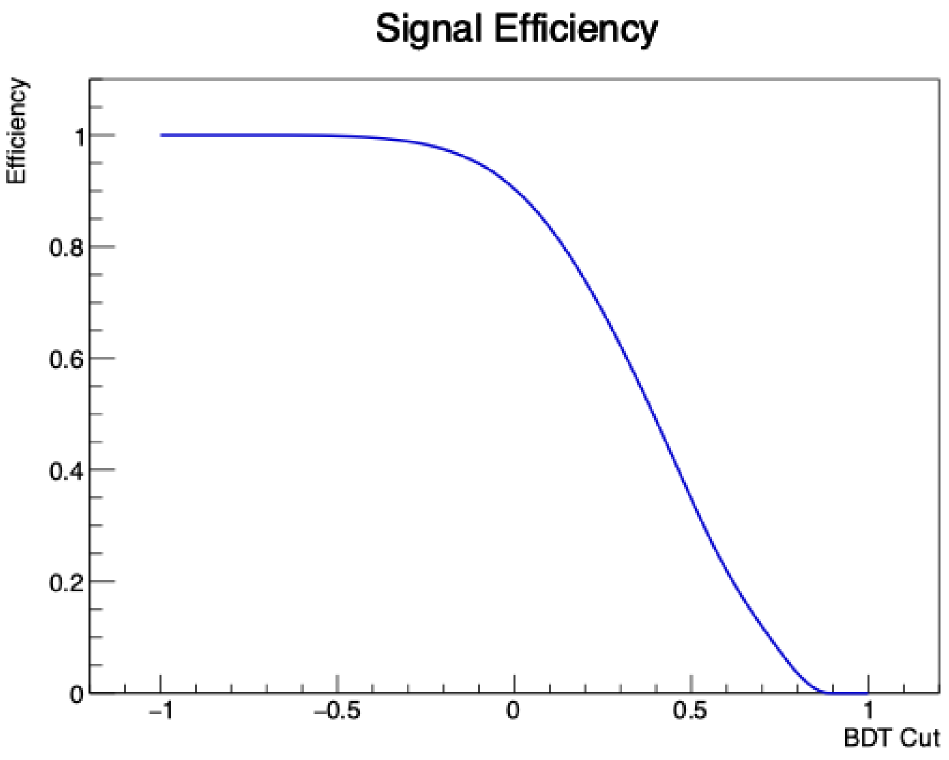


Signal efficiency

Comparison of the Signal Efficiency after the model has been applied to the MC samples

- Changes from the training but they both have the same despite being of different cuts





Configurations

- Depth = 4
- Input = (1, 9, 1)
- Batch size = 4000
- Loss function = C
- Layer 0 RECURRENT
  - Layer:
    - NInput = 9,
    - NState = 64,
    - NTime = 1
    - Output = (4000 , 1 , 64)
- Layer 1 RESHAPE
  - Layer
    - Input = (1 , 1 , 64)
    - Output = (1 , 4000 , 64)
- Layer 2 DENSE
  - Layer:
    - (Input = 64 , Width = 32)
    - Output = (1 , 4000 , 32)
    - Activation Function = Relu
- Layer 3 DENSE
  - Layer:
    - (Input = 32 , Width = 1)
    - Output = (1 , 4000 , 1)
    - Activation Function = Identity

BDT

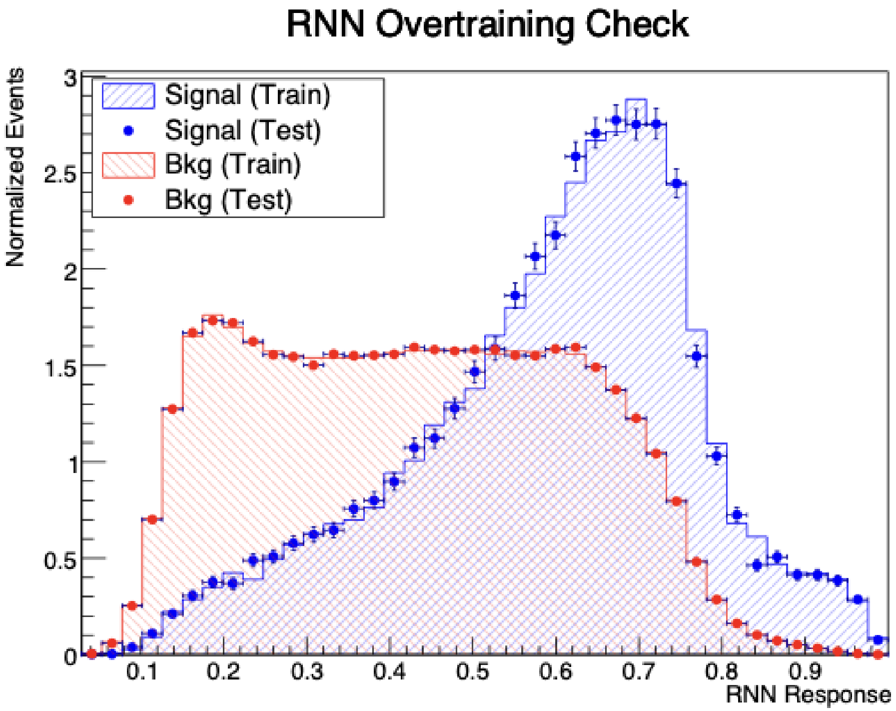
- NTrees=400
- MinNodeSize=2.5%
- MaxDepth=3
- BoostType=Grad
- Shrinkage=0.10
- nCuts=20
- DoEstTransition=1

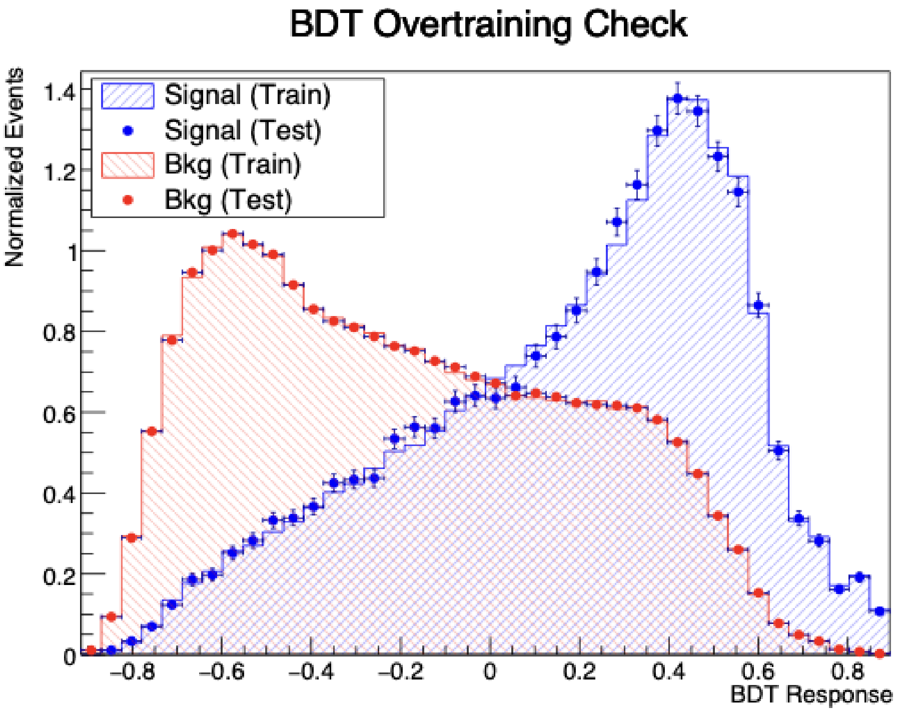
RNN

Stability

Stability checks of the models

- Checking on the overtraining again after the model has now been applied to the data
  - Testing and training are still relatively following each other -> No overtraining





Ranking

```
--- RNN_LSTM Variable Ranking ---
1. dphi_zmet : 0.0230081
2. Njets : 0.0222421
3. dphi_lead_jet_met : 0.00815718
4. deltaR_ll : 0.00713503
5. pt_over_mt : 0.00700777
6. met_sig : 0.00663855
7. met_over_mt : 0.00281547
8. lep_pt_sub : 0.00199826
9. lep_mass : 0.00012246
```

```
: ----------------------------------------------------
: Rank : Variable          : Variable Importance
: ----------------------------------------------------
:    1 : met_sig           : 2.034e-01
:    2 : pt_over_mt        : 1.832e-01
:    3 : met_over_mt       : 1.444e-01
:    4 : deltaR_ll         : 1.241e-01
:    5 : Njets             : 1.049e-01
:    6 : dphi_zmet         : 1.033e-01
:    7 : lep_mass          : 7.129e-02
:    8 : dphi_lead_jet_met : 5.098e-02
:    9 : lep_pt_sub        : 1.461e-02
: ----------------------------------------------------
```

Variable rankings -> Note that these rankings are not the same between the two models

Application of training on testing data

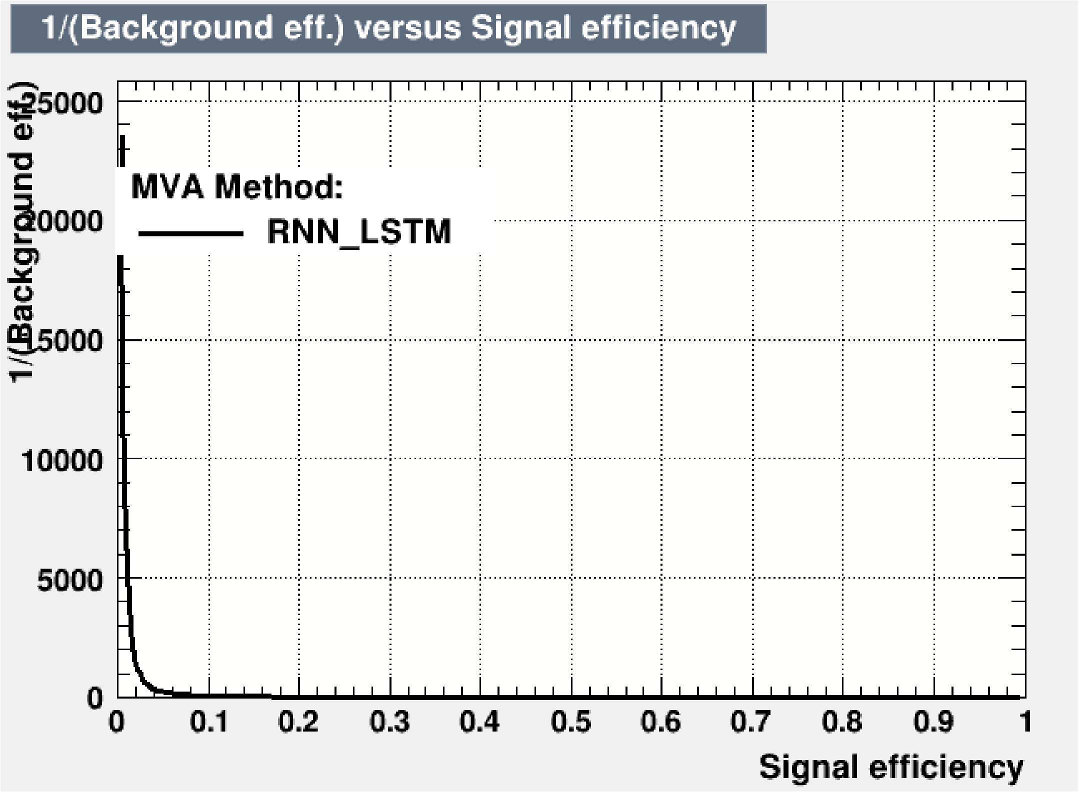
```
=====================================================
Optimal RNN Cut Value (Maximizing S/sqrt(B)): 0.52
Maximum Significance (S/sqrt(B)): 0.107625
-----------------------------------------------------
Yields at Optimal Cut (Raw Counts):
  Total Signal (S): 275761
  Total Background (B): 45384930
=====================================================
```

```
=====================================================
Optimal BDT Cut Value (Maximizing S/sqrt(B)): 0.14
Maximum Significance (S/sqrt(B)): 0.112246
-----------------------------------------------------
Yields at Optimal Cut (Raw Counts):
  Total Signal (S): 194125
  Total Background (B): 26992410
=====================================================
```

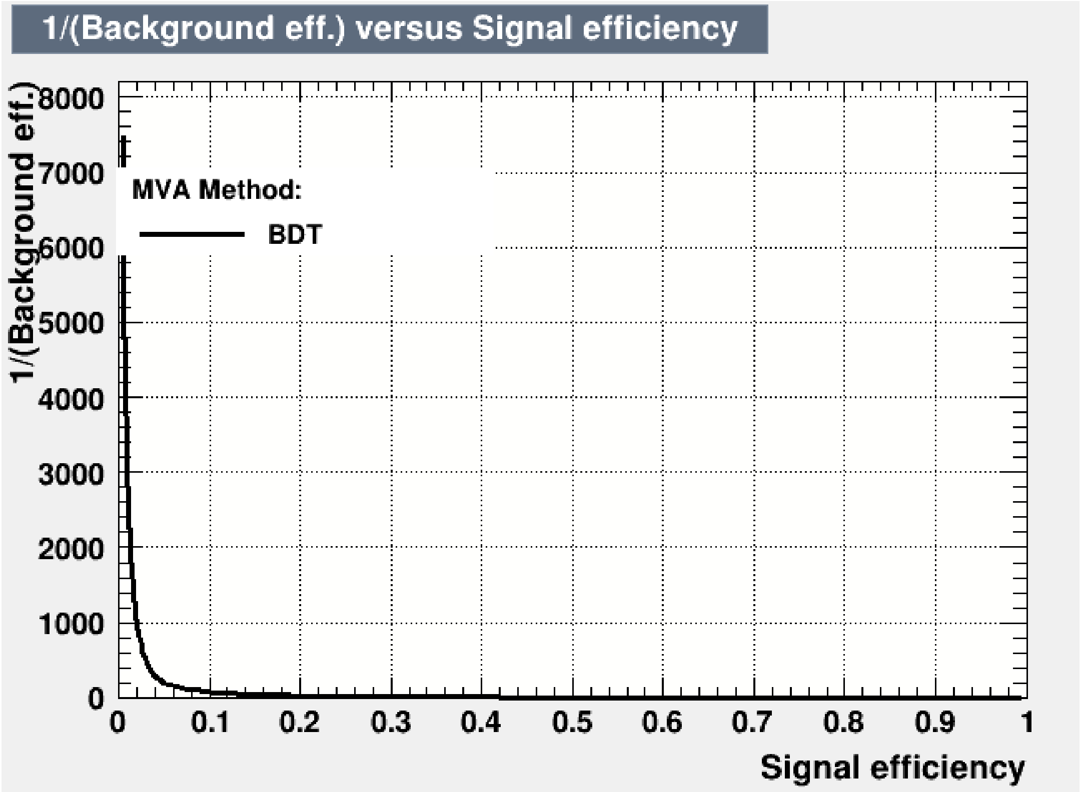
RNN

- Recurrent Neural Networks
  - ➢ It is a Sequential type of Neural Network
  - ➢ Struggles with longer sequences (Vanishing Gradient)
  - ➢ Can take time to run
  - ➢ Variable order matters

Inverse ROC



1/(Background eff.) versus Signal efficiency
1/(Background eff.)
25000
20000
15000
10000
5000
0
MVA Method:
RNN_LSTM
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Signal efficiency



1/(Background eff.) versus Signal efficiency
1/(Background eff.)
8000
7000
6000
5000
4000
3000
2000
1000
0
MVA Method:
BDT
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Signal efficiency

ADA versus GRAD

ADABoost

- Works via reweighing
- More popular one
- Not super robust – harder to look at outliers

GradBoost

- Based on the constraints of the ADABoost
- Minimizes the deviations between the model response and the actual true values
- Uses a loss function => Overtraining!

Definitions

- Delta R lepton – separation of the eta and azimuthal (phi) planes
- MET Significance – statistical significance of the MET
- Invariant Mass – mass of the particle system (mass before the decay took place)
- Dilepton Mass – invariant mass of two leptons
- SubLeading Lepton Transverse Momentum – transverse momentum of the second highest lepton
- Longitudinal Momentum Component of the dark photon the z component of the momentum of the dark photon found via solving for it with the quadratic eqn.
- Psuedorapidity - the angle at which a particle travels relative to the collider beam pipe (invariant under longitudinal Lorentz boosts)

Cuts

- 2 leptons – Z boson decay
- Leading Transverse Momentum – less likely to be from background
- Subleading Transverse Momentum – ensures leptons are energetic enough to be reliably measured
- Electron eta cut – electron outside of 2.47 harder to detect and physical gap between 1.37 and 1.52v leading to poor energy resolution
- Muon eta cut – reliable coverage of the muon
- Dilepton invariant mass cut – consistent with decay of Z-boson
- MET cut - in order to reject processes with
- Number of jets cut – Higgs decay does not produce many jets
- Jet eta – realiable coverage inside this range
- Jet Momentum – ensures that the jets are energetic enough and will be realiably measured

Exactly two leptons ($e^+e^-/\mu^+\mu^-$)
Leading Transverse Momentum (p_T‘)>27GeV
Sub-leading Transverse Momentum (p_T‘)>15GeV
Electron pseudorapidity (η) cut: $\lvert\eta\rvert < 2.47$ excluding the $1.37 < \lvert\eta\rvert < 1.52$ calorimeter crack region.
Muon pseudorapidity (η) cut: $\lvert\eta\rvert < 2.7$
Dilepton invariant mass: 65Gev< m_{ll} <110 GeV
MET >15 GeV
$N_{jet} \leq 2$
$p_T^{jet} > 30$ GeV
$\lvert\eta\rvert < 4.5$