SEARCH FOR HEAVY RESONANCES USING DEEP NEURAL NETWORK

Shivam Raj

Supervised by: - Dr. Prolay Kr. Mal

Master's Thesis Presentation

6th May 2022

Overview

- *Standard Model*
- *Vector-Like quark(VLQ)*
- *Compact Muon Solenoid(CMS)*
- *Signal and backgrounds events*
- *machine learning(DNN)*
- *Output from DNN*
- *Output from present CMS analysis(BDT)*
- *Comparison of the BDT and DNN outputs*
- *Summary and Conclusion*

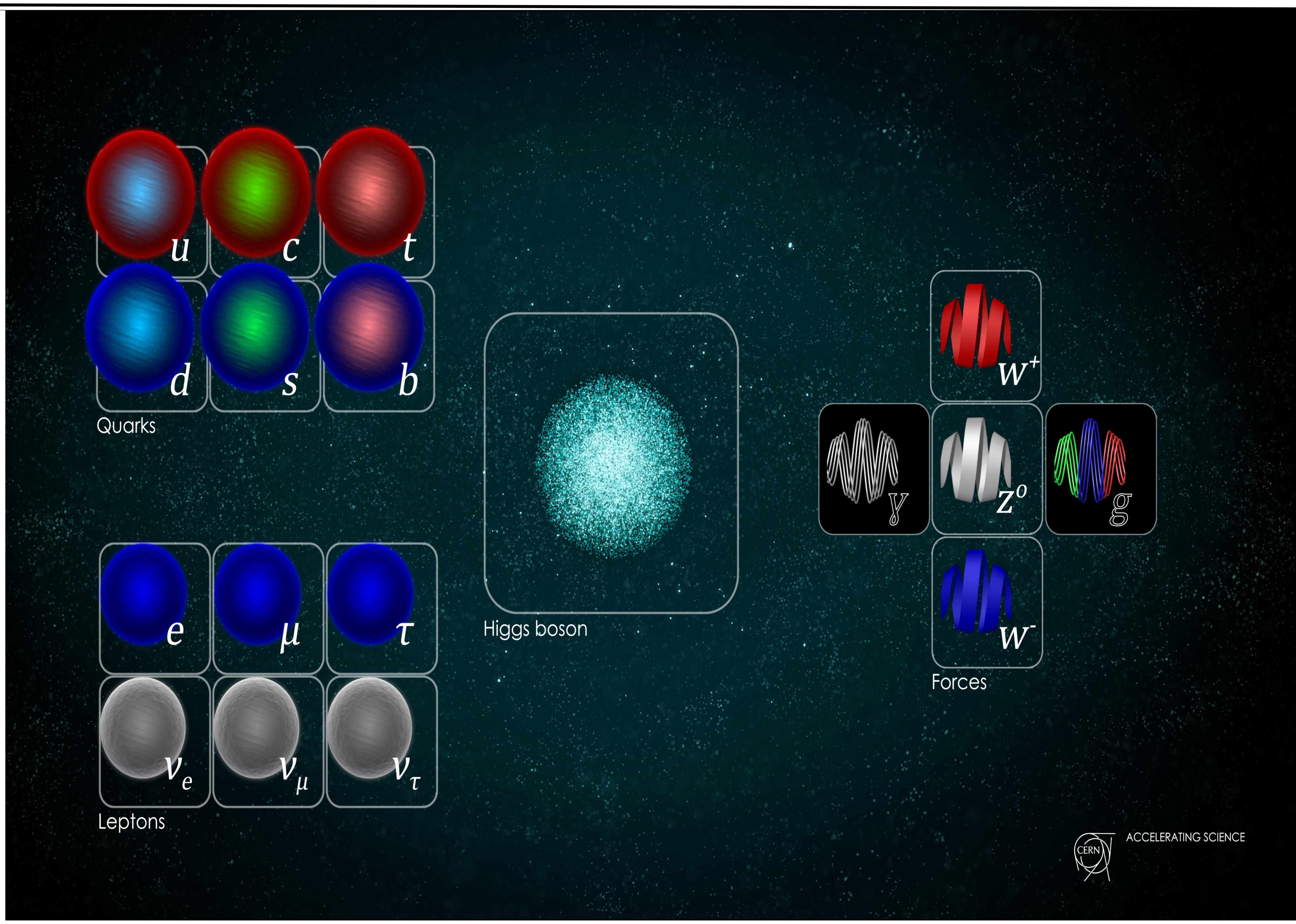
Standard Model

- The Standard Model(SM) of particle physics is the theory describing the electroweak and strong interactions.
- SM has only chiral fermions i.e. the left handed and right handed fermionic fields transform differently under the SU(2) gauge transformations.
- SM with charged fermions gain mass via coupling to the Higgs boson(Brout-Englert-Higgs) BEH mechanism.
- SM successfully predicted the existence of quarks, neutrino, and Higgs Boson.
- SM also leaves some phenomena unexplained,

  Presence of dark matter and dark energy.

  Neutrino Oscillations.

  Mass hierarchy problems.



- Some theoretical models predict the existence of heavy quarks which are not Chiral in nature.

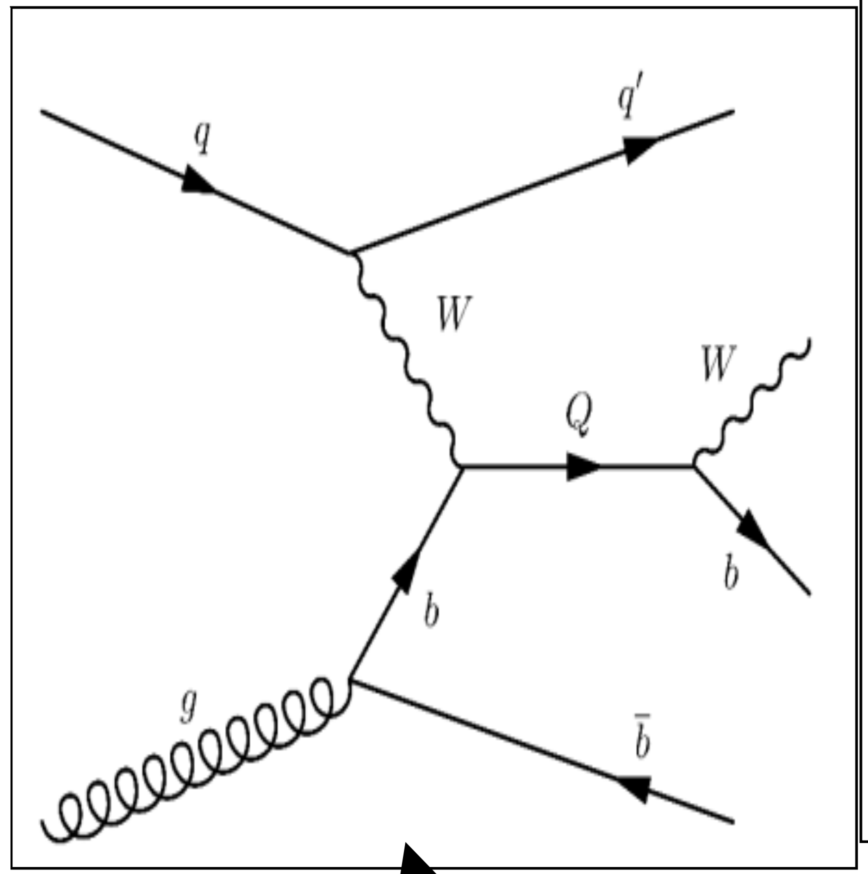
Several theories have been proposed to explain all the phenomena, one of them is the existence of vector-like quark

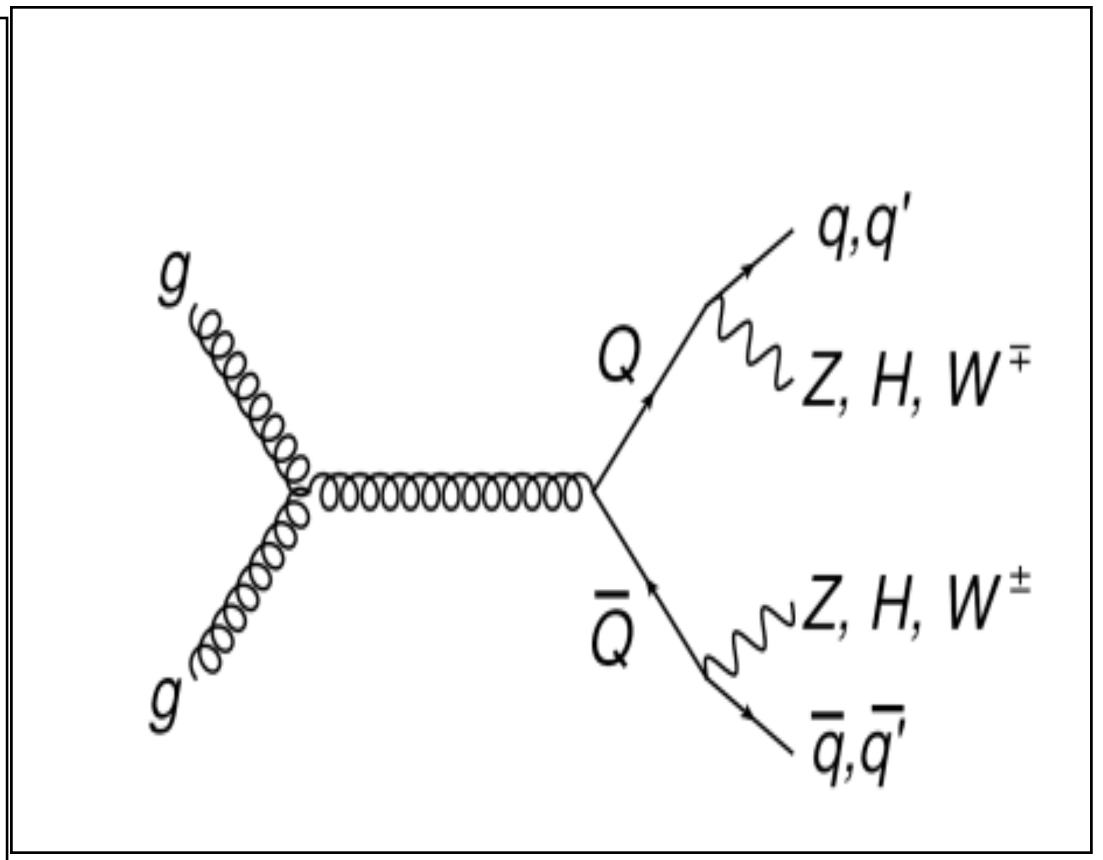
Vector-Like Quark(VLQ)

- VLQ have both left- and right-handed coupling to charged currents.
- These quarks are heavier partner of top and bottom quark.
- T' has charge +2e/3 while B' quark has charge -e/3
- It is a colored spin 1/2 fermions.
- The vector-like top quark partner T′ has two generation modes
  - **Single production through weak interactions**
  - pair production through strong interactions.
- The VLQ pairs to SM quarks q and weak bosons V ∈ {W,Z,H}, via QqV vertices

$$T' \longrightarrow bW, T' \longrightarrow \boxed{tZ, T'} \longrightarrow tH$$
$$B' \longrightarrow tW, B' \longrightarrow bZ, B' \longrightarrow bH$$

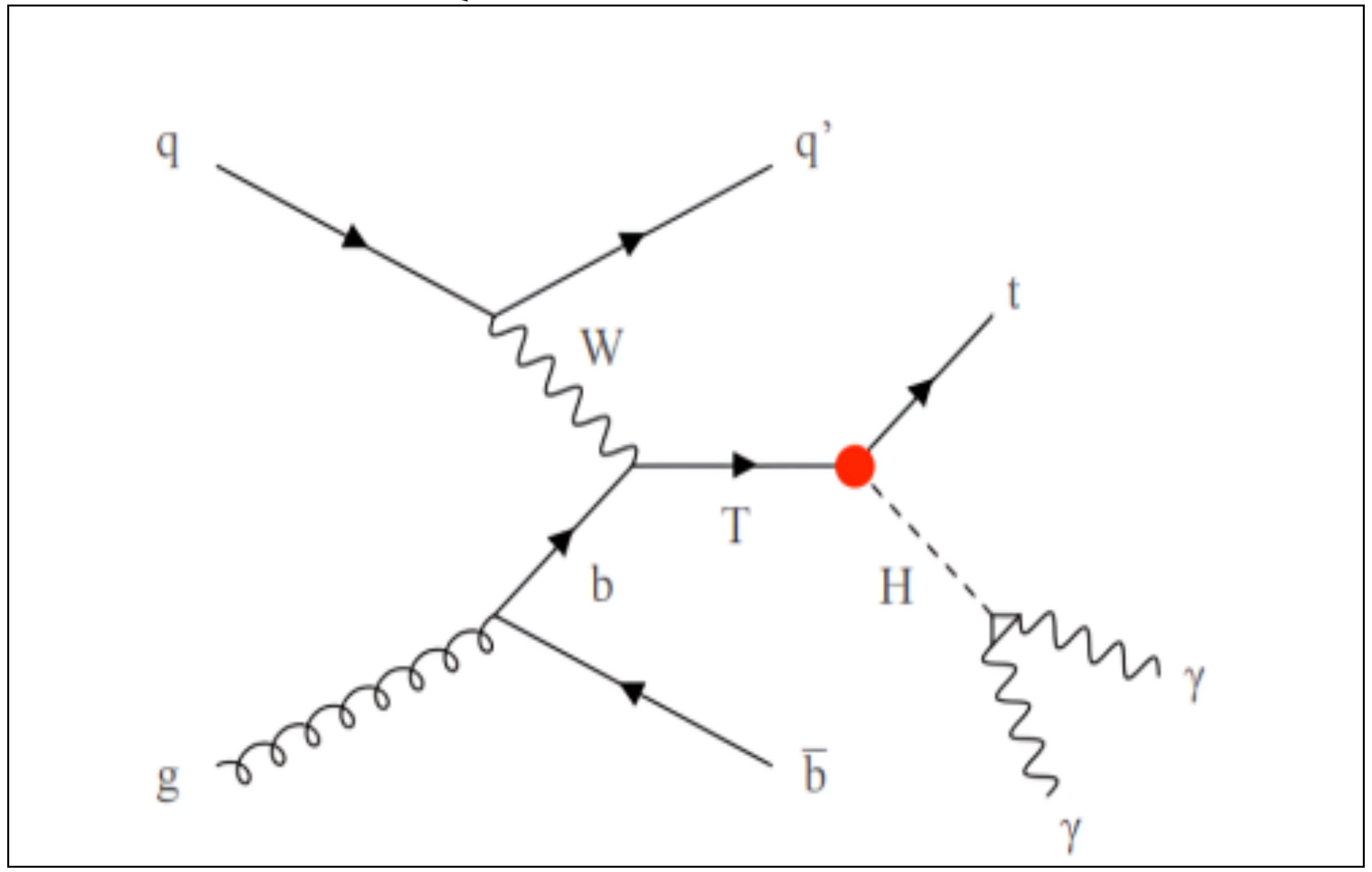
- **In this analysis, we will look for VLQ T' in the mass range from [600,1200] Gev**.



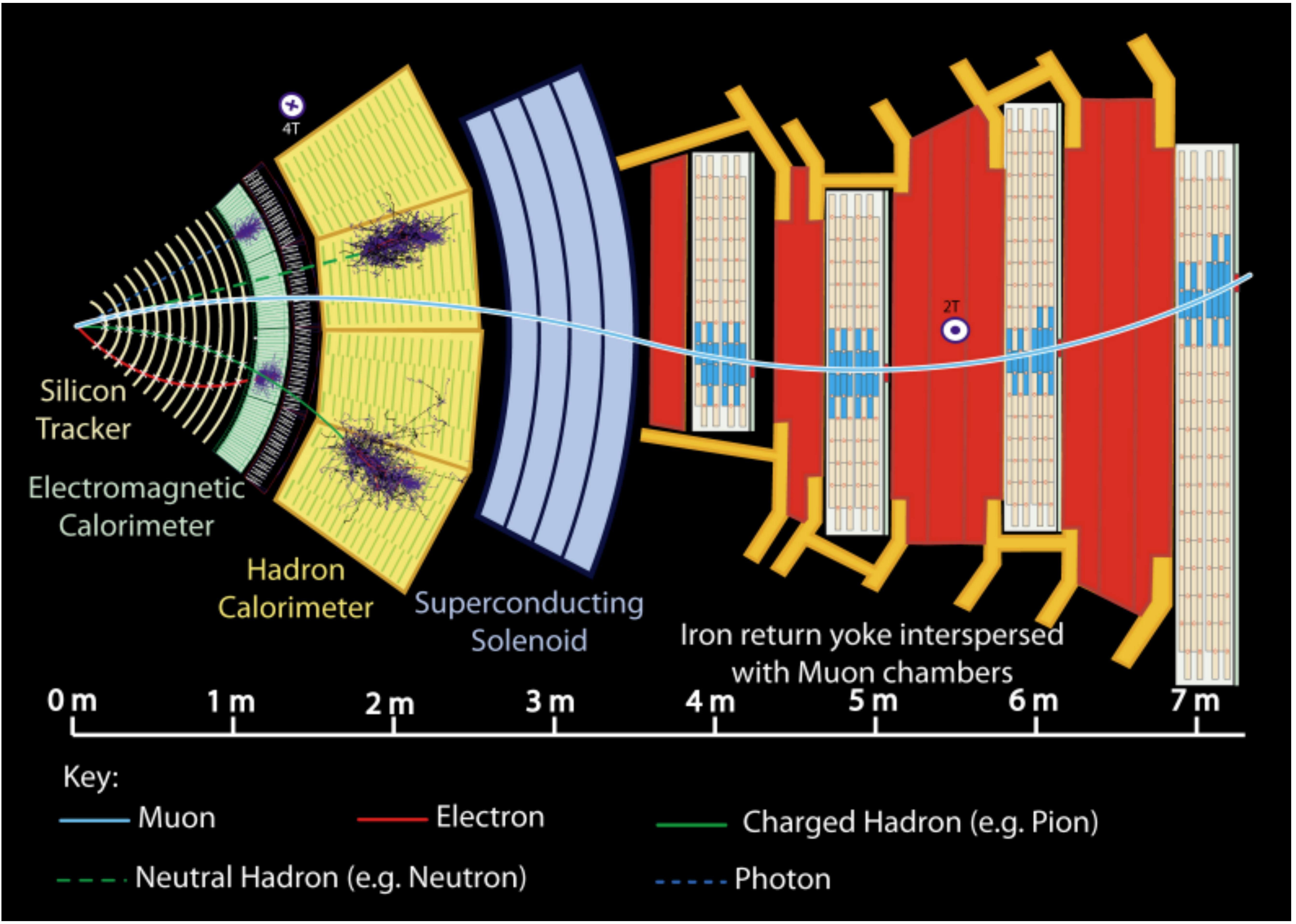


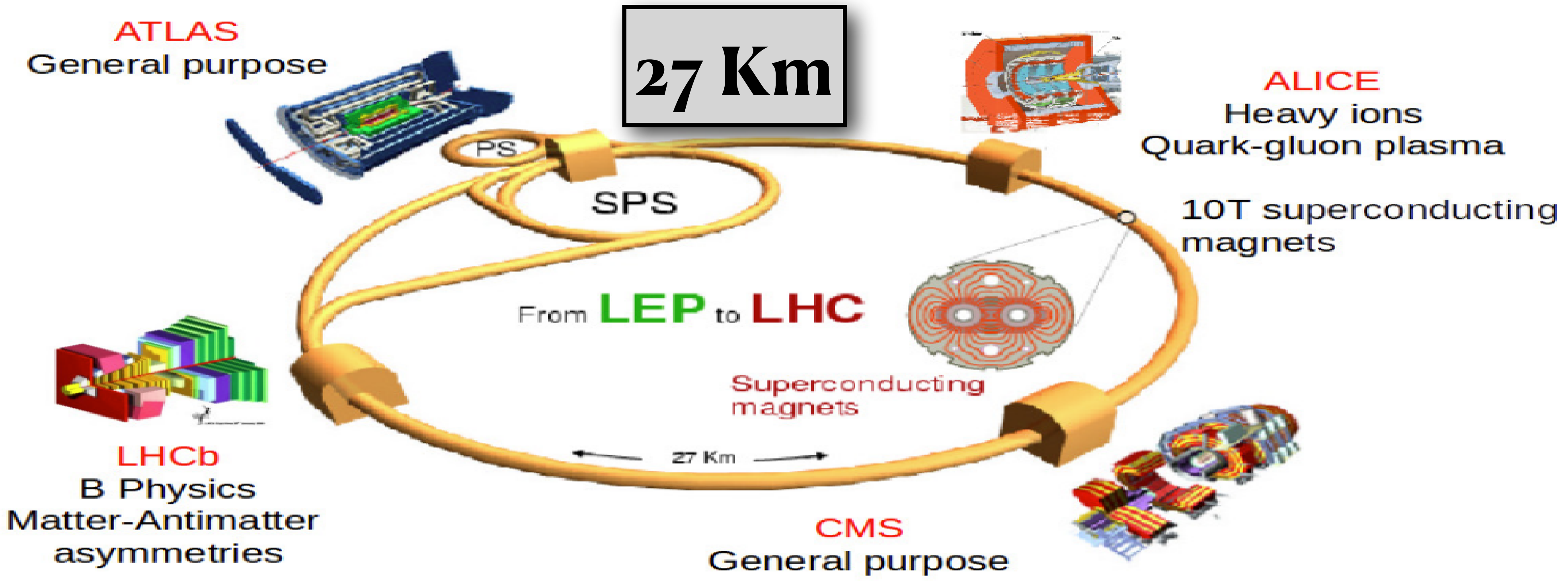
Single

Pair Production



Compact Muon Solenoid(CMS)





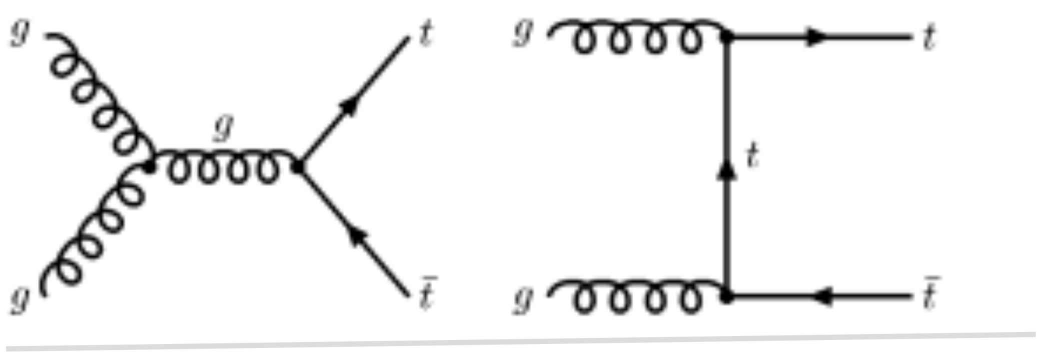
- Electromagnetic Calorimeter (ECAL) measures the energy of electrons and photons by stopping them completely.
- Hadrons, which are composite particles made up of quarks and gluons, fly through the ECAL and are stopped by the outer layer called the Hadron Calorimeter (HCAL).
- Charged particle trajectories are measured by the silicon pixel and strip sub detectors, covering $0< \phi <2\pi$ in azimuth and $|\eta| < 2.5$, where the pseudo rapidity η is defined as $\eta = -ln[\tan \theta/2]$, with θ being the polar angle of the trajectory of the particle with respect to the counterclockwise-beam direction.

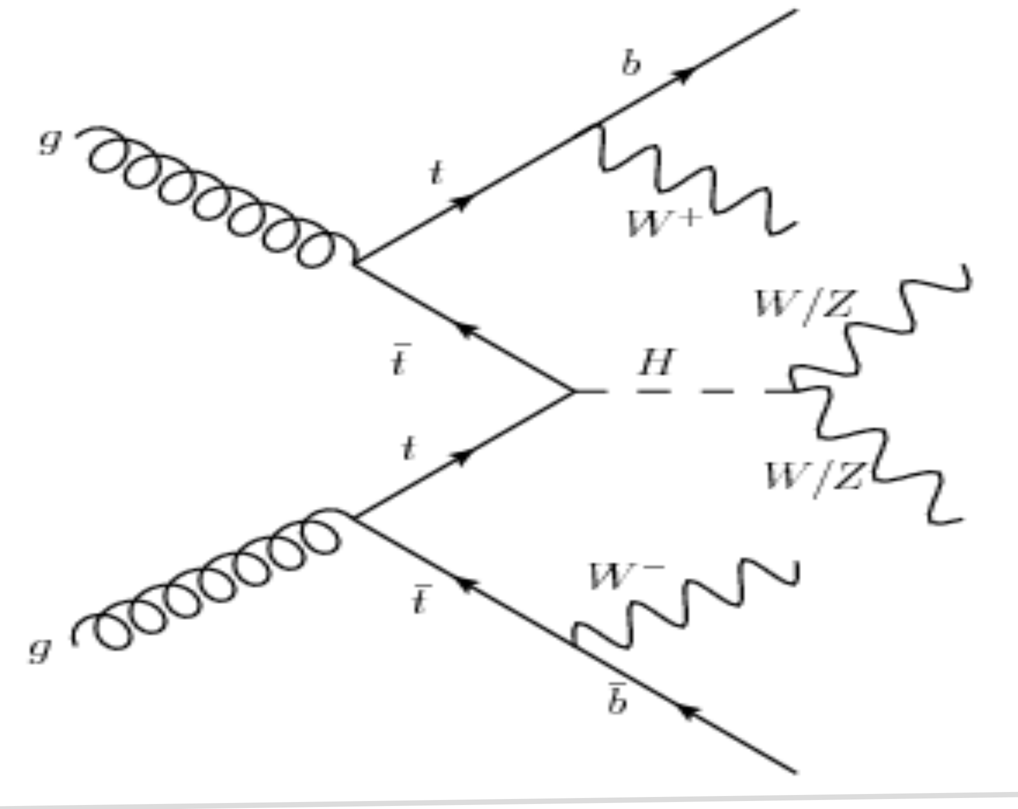
- Acts as a giant, high-speed camera, taking 3D "photographs" of particle collisions from all directions up to 40 million times per second.

Signal and Backgrounds

- All samples of the signal and backgrounds are simulated for a center of mass energy ($\sqrt{s}$) of 13 TeV.
- T' →tH($\rightarrow \gamma\gamma$) has similar final state topologies like other Higgs production mechanism.
- Machine learning techniques have been used to separate signal T' events from different background.
- Two type of background is taken: Standard Model Higgs(SMH) and Non-Resonant backgrounds(NRB).

ttgg

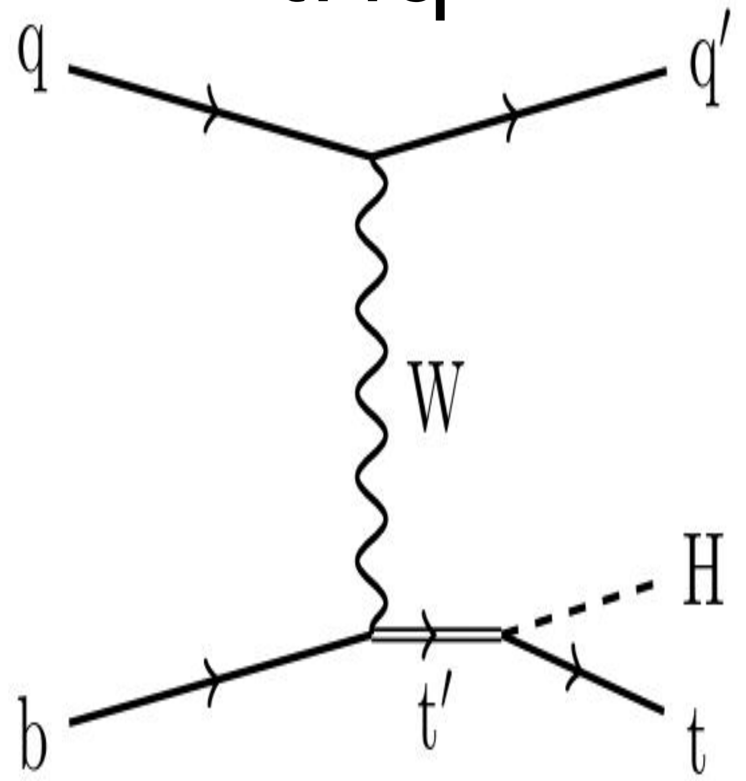




Possible background for T':

- SMH: $t\bar{t}H$($\rightarrow \gamma\gamma$), tH($\rightarrow \gamma\gamma$)q, VH($\rightarrow \gamma\gamma$), ggH($\rightarrow \gamma\gamma$), VBFH($\rightarrow \gamma\gamma$)

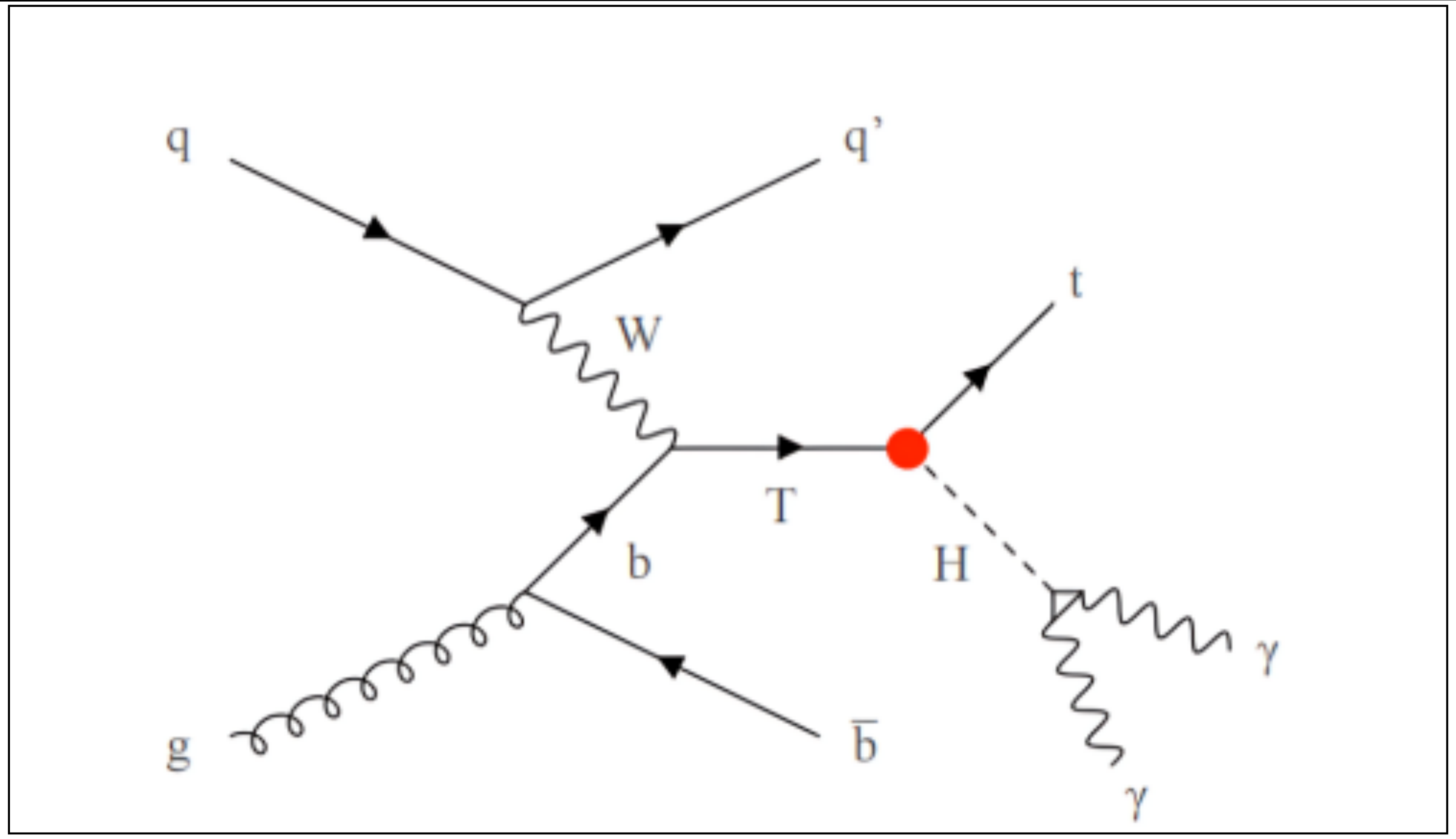
tHq



Total number of input variable taken for the DNN training is 42.

Deep Neural Network(DNN) is trained with T' events and SMH backgrounds ($t\bar{t}H$, tHq, VH, ggH, VBF) processes.

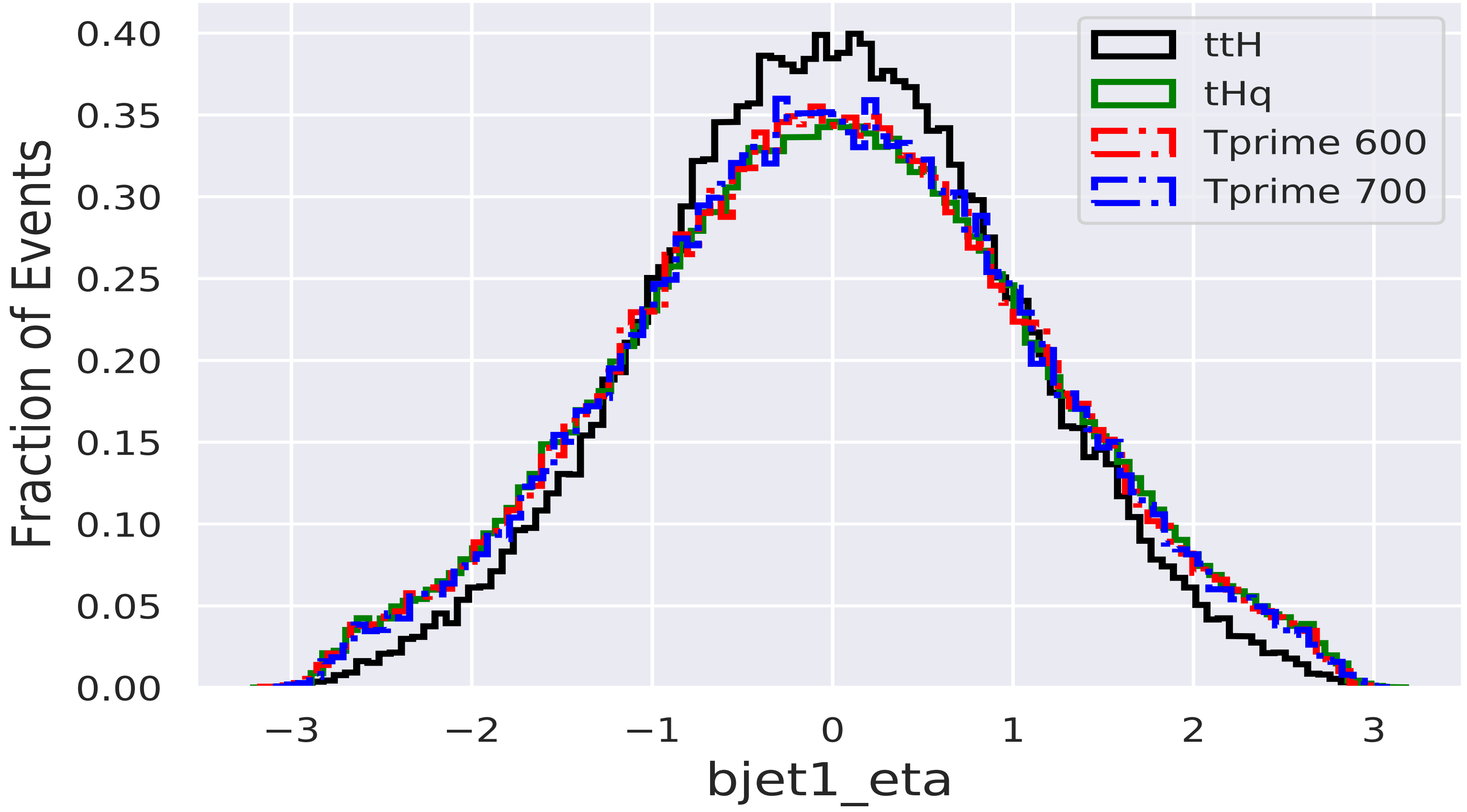
Variable Selection



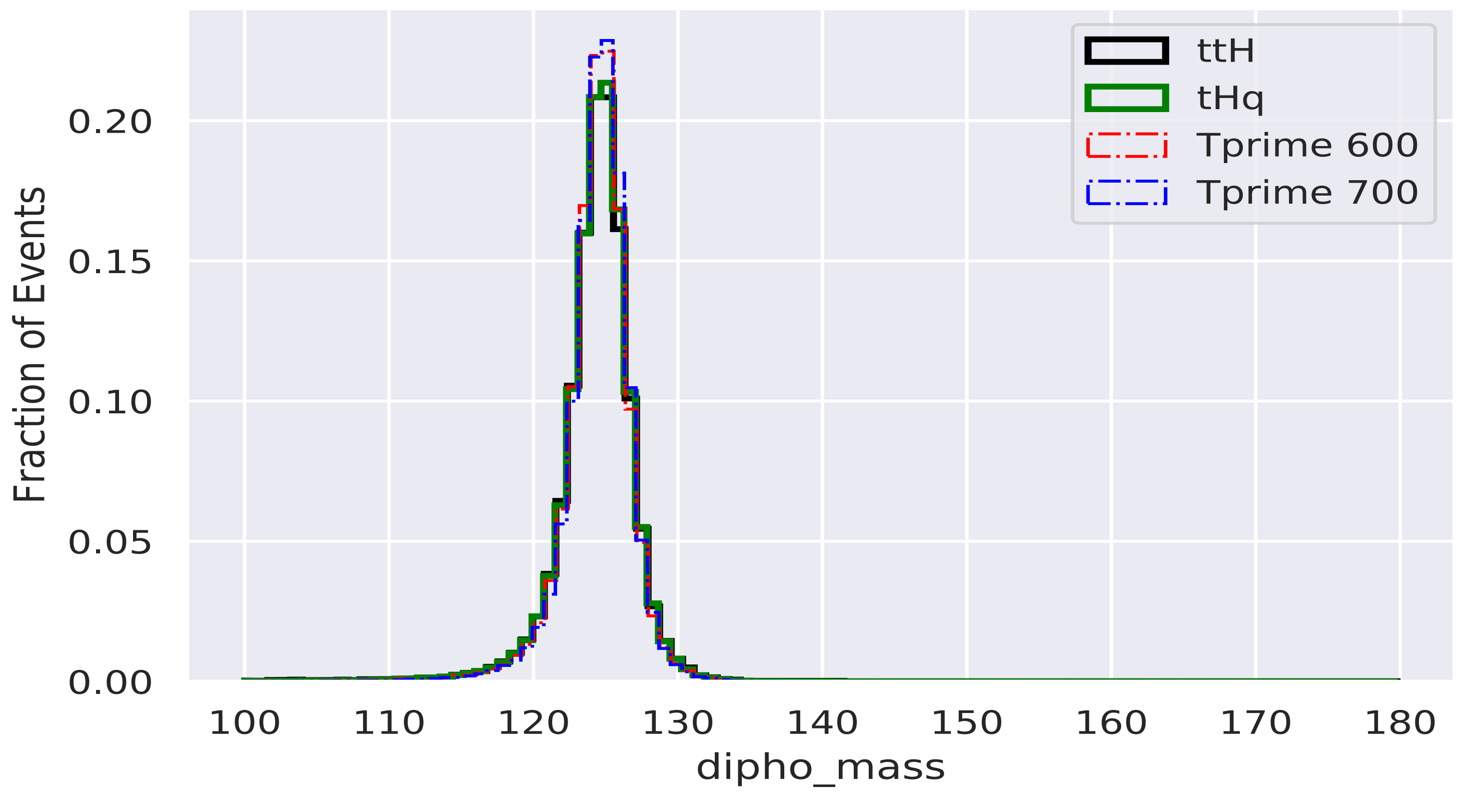
- leading (sub-leading) photon IDMVA
- leading (sub-leading) photon pT/mγγ
- leading lepton charge, pT, η
- Jet, b-jet and central jet(lηl < 1) multiplicity
- pT and lηl for leading two jets
- angular separation between: each photon and the forward jet, the leading lepton
- Leading b-jet and the forward jet.
- leading (sub-leading) pixel seed veto
- pT and b-tag score of the forward jet. The forward jet is the highest lηl jet in the event

Variable Output

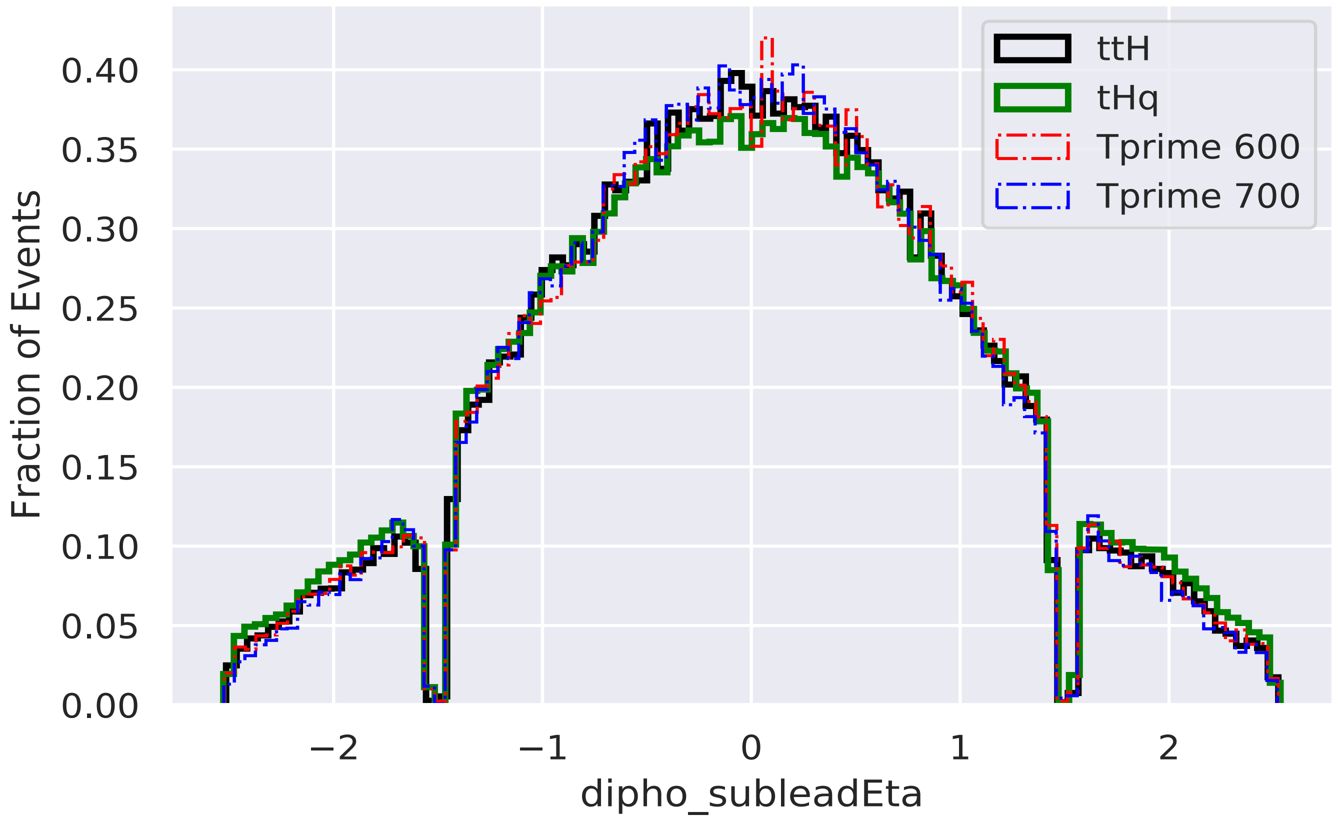
bjet_eta



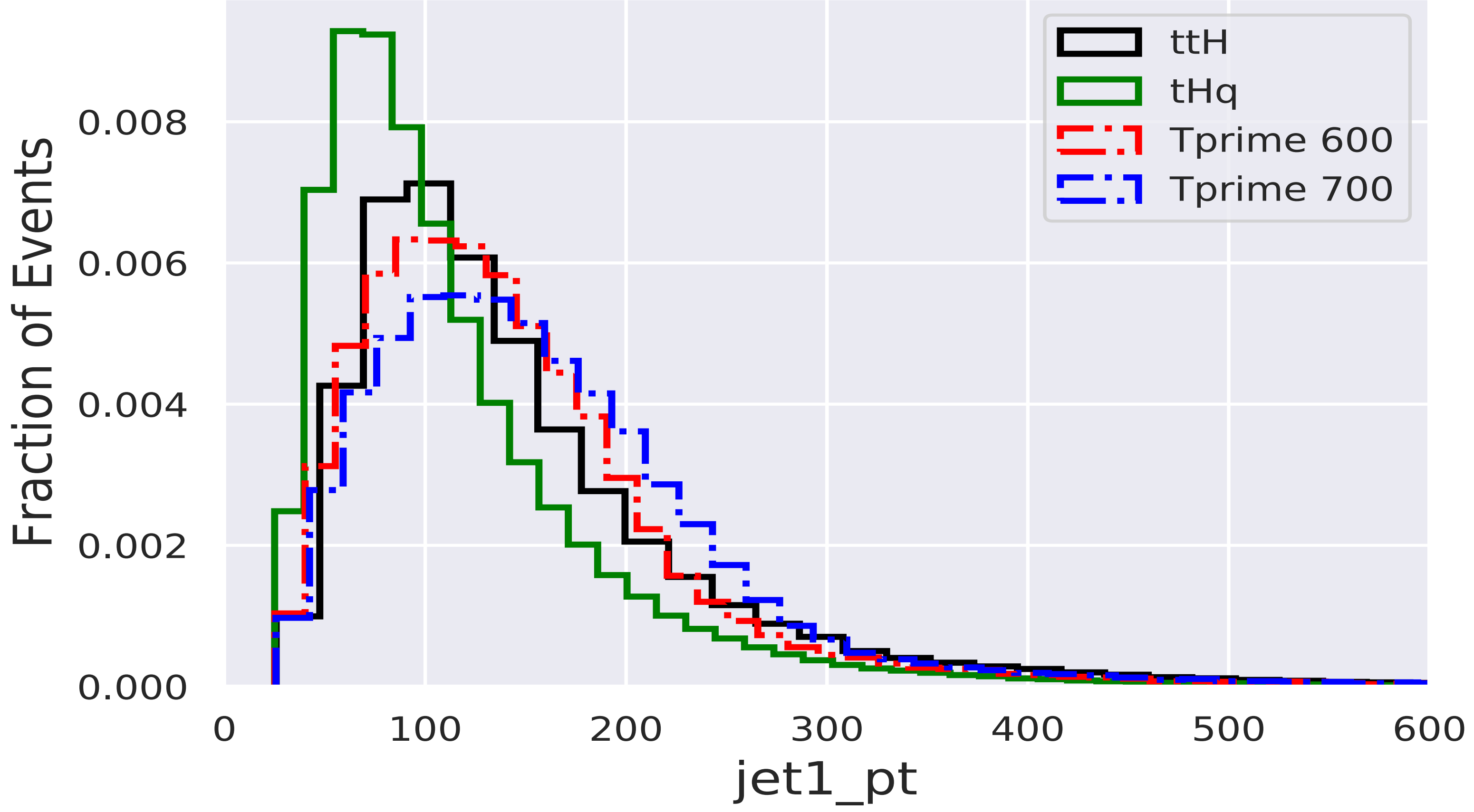
dipho_mass



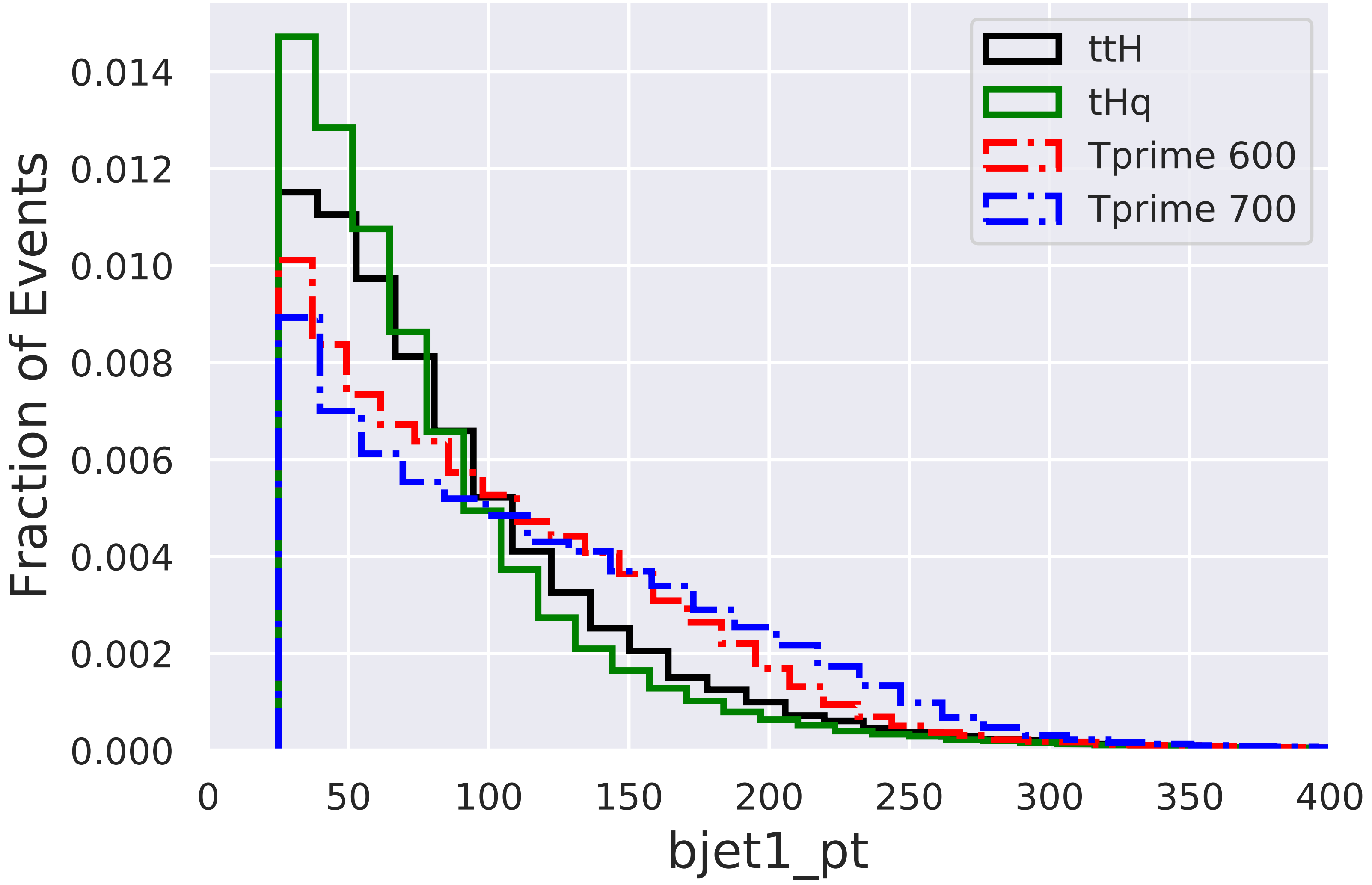
dipho_subleadeta



jet1_pt



bjet1_pt



lepton_leadPt

Input variable of signal and background cannot be separated, as shown in figure. To separate the signal and background we need machine learning techniques. Here, we will using Deep Neural Network.

Machine learning

edureka!

- Machine learning(ML) is a part of artificial intelligence.
- It is a process by which performance of system gets improved from previous experiences.
- ML allows softwares to become accurate without being explicitly programmed

Machine Learning
- Supervised
  - Classification
    - Multi
    - Binary Classification
  - Regression
- Unsupervised Learning
- Semi-Supervised Learning
- Reinforcement

Example of ML:-
1. Siri, Alexa: virtual personal assistance
2. Facebook(use GNN)
3. Email classification as spam or not spam

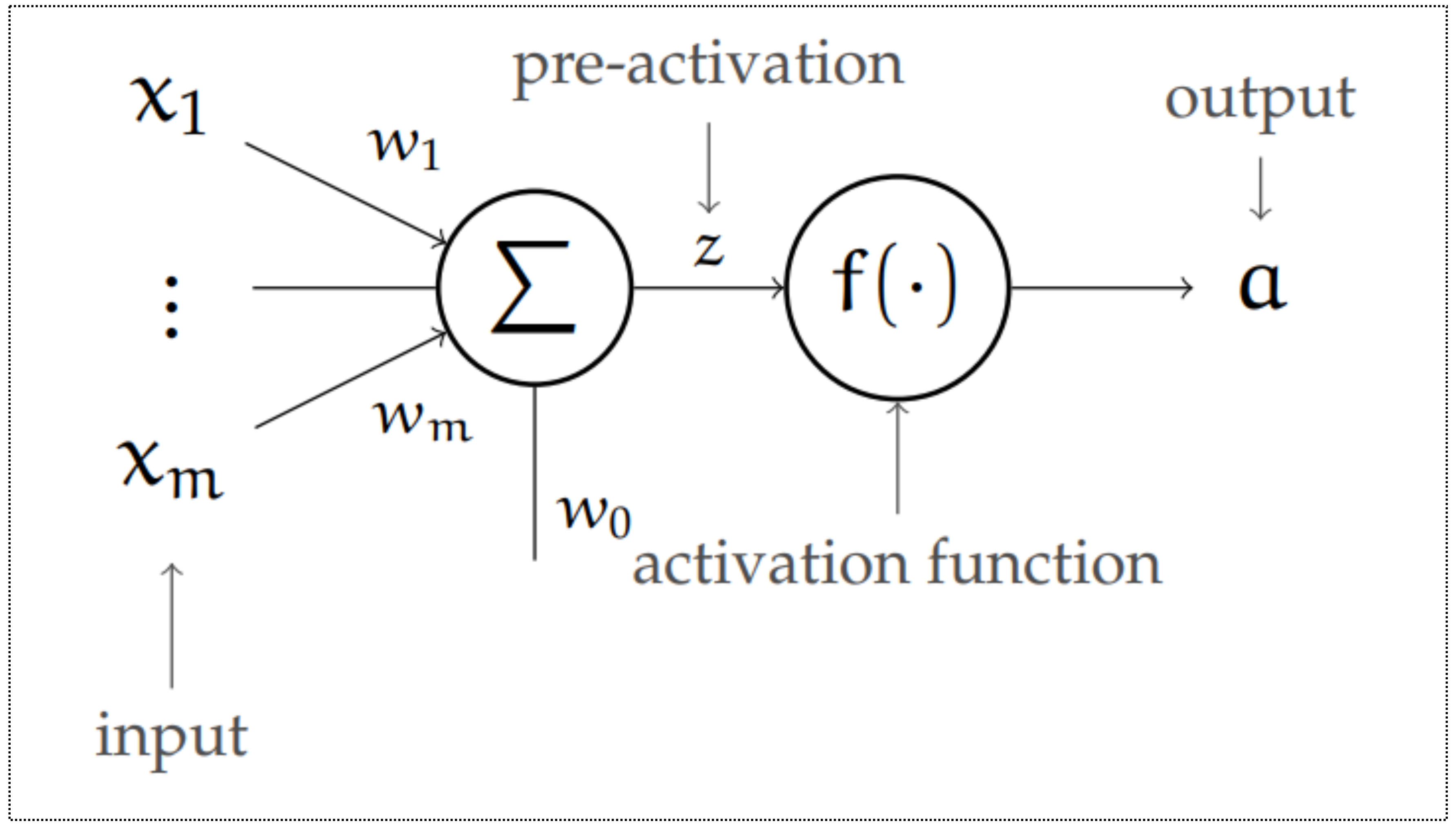
Supervised Learning:
- Uses labelled datas as input and output
e.g. Biometric Attendance

Classification
- Works on discrete class of datasets

Regression
- Works on continuous dataset.

Deep Neural Network(DNN)

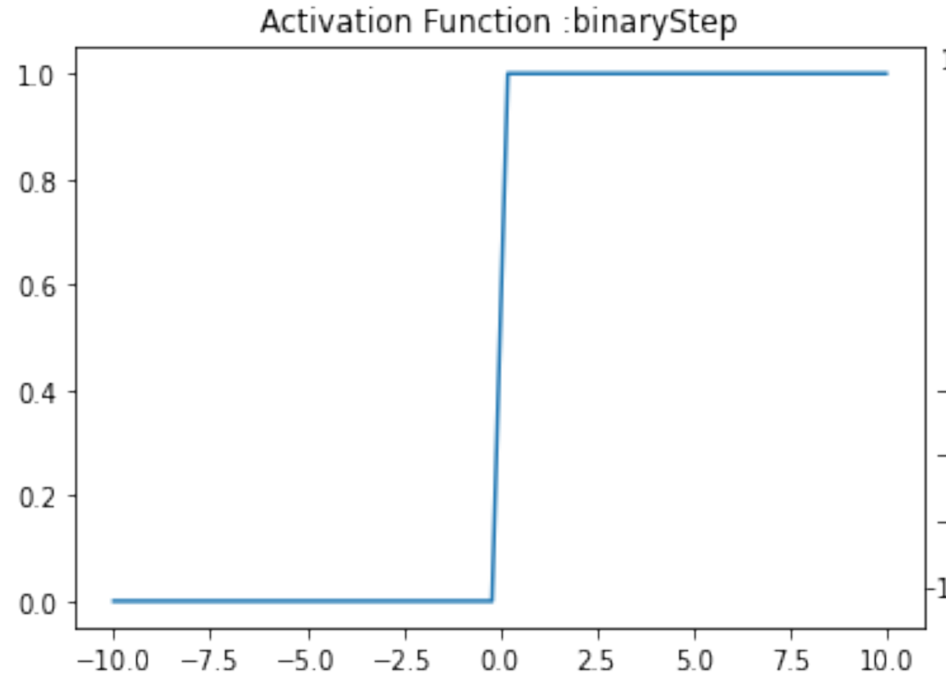
- It can be conceived as part of ANN, with similar structure of the human brain. Where neurons can be compared to nodes and connection between nodes as synapses.



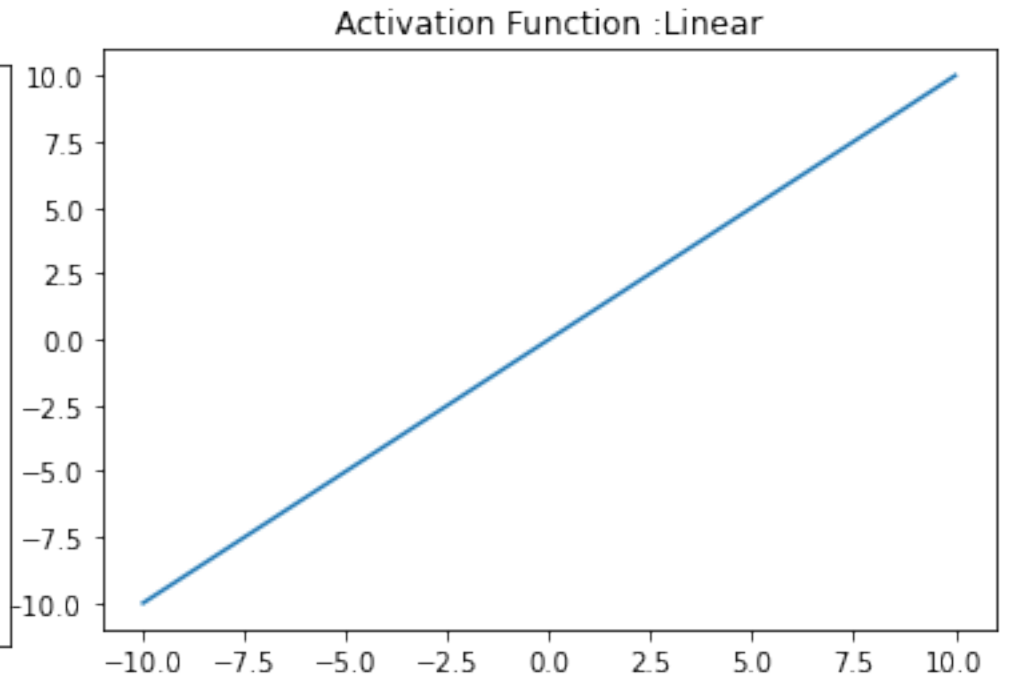
$$a = f(z) = f(\sum_{j=1}^{j=m} x_j w_j + w_0) = f(w^T x + w_0)$$

Three types of Activation Functions:-

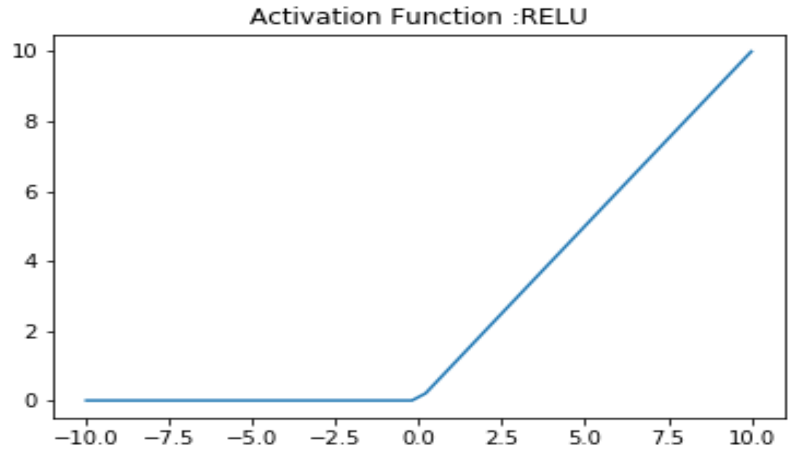
Binary

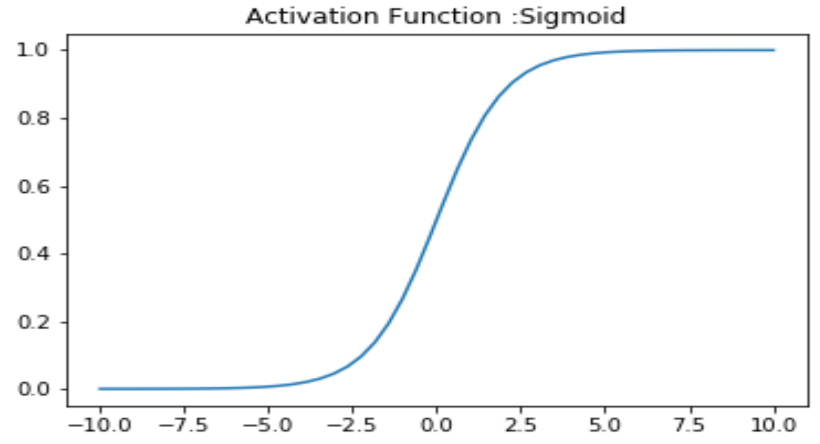


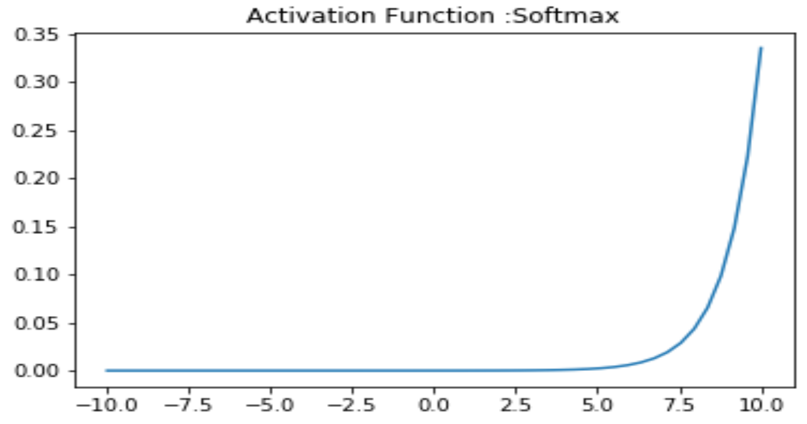
Linear

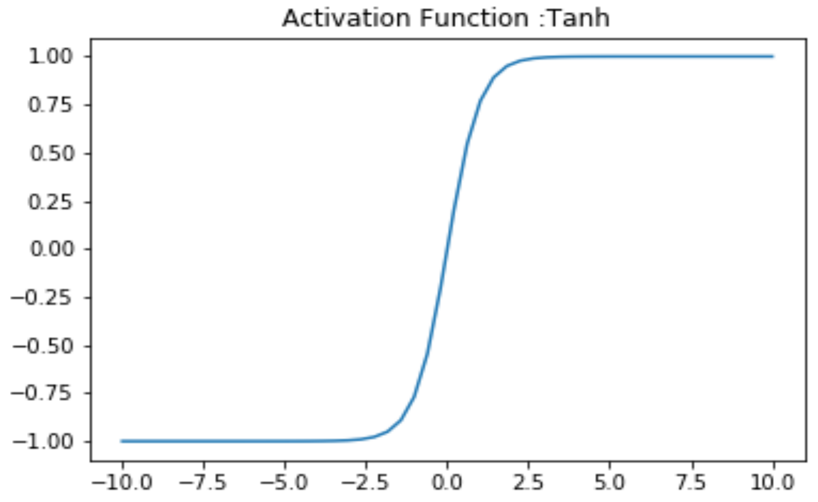


Non-Linear









Each dataset divided into training, validation, and testing.

- Single network(a input, hidden, and a output) is also known as perceptron, which only works for binary classification task.
- The Output depends on the type of activation function
- Deep means number of layer>2

Low level framework

High level framework

Keras

Error Estimation(DNN)

Loss Function

- Loss is calculated for the previously used function as:

$$J(W, W_0) = \sum_{i=1}^{n} L(f((x^i); W), y^i)$$

- For any model, the main goal is to minimize the loss function.
- Also known as cost function, objective function.

$$L(f(x^i; W), y) = Loss(A^L, y)$$

$$A^L = f^L(z^L) and z^L = W^{L^T} A^{L-1}$$

Back Propagation

- To minimize the loss function, we apply chain rule w.r.t the weights in the final layer:

$$\frac{\partial loss}{\partial w^L} = \frac{\partial loss}{\partial A^L} \cdot \frac{\partial A^L}{\partial Z^L} \cdot \frac{\partial Z^L}{\partial W^L}$$

- If we continue to apply chain rule we can write the loss as,

$$\frac{\partial \mathrm{loss}}{\partial Z^l} = \frac{\partial A^l}{\partial Z^l} \cdot W^{l+1} \cdot \frac{\partial A^{l+1}}{\partial Z^{l+1}} \cdot \ldots W^{L-1} \cdot \frac{\partial A^{L-1}}{\partial Z^{L-1}} \cdot W^L \cdot \frac{\partial A^L}{\partial Z^L} \cdot \frac{\partial \mathrm{loss}}{A^L}$$

Loss Optimization

- Mini Batches Gradient Descent

We have batch gradient descent,

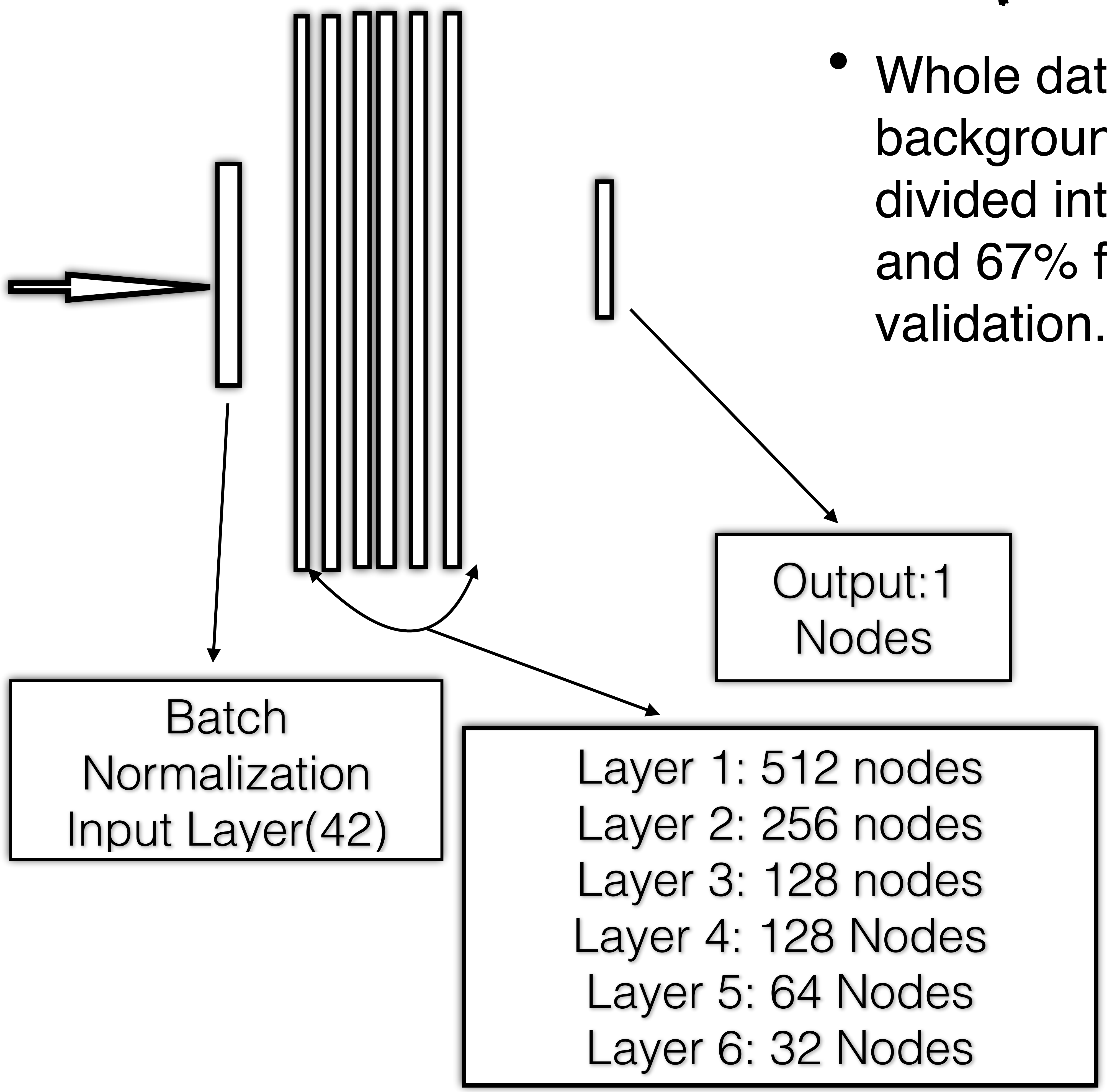
$$W \leftarrow W - \eta \frac{\partial J(W)}{\partial W}$$

η is the learning rate, which is equivalent to

$$W \leftarrow W - \eta \nabla_W \sum_{i=1}^{n} L(h(x^{(i)}; W), y^{(i)})$$

- Adadelta
- Adam
- SGD
- Adagrad
- RMSprop
- SGD with monentum

Deep Neural Network(DNN)



- Whole dataset (signal and background) have been divided into 33% for testing and 67% for the training and validation.

- Activation Function:-
  - Hidden Layer- 'relu'
  - Output layer - 'Sigmoid'

- Kernel Initializer:- 'random_uniform'
- Kernel Regulizer:- l2 regulizer
- Optimizer:- 'Adam'
- Loss:- 'binary cross entropy'
- metrics = ['accuracy']

DNN training model

- Dropout of around 0.45 applied to each layers(to avoid overfitting).
- Batch Size:- 9000
- Total number of Epochs:- 100
- Verbose: 1

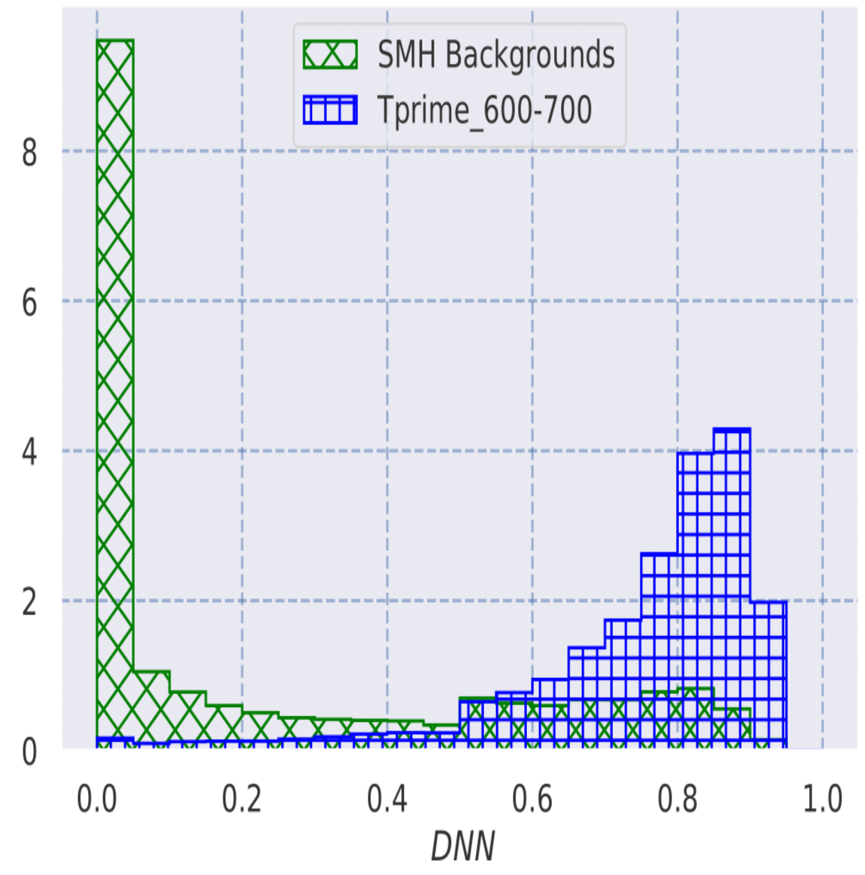
- Bias added to delay the triggering of the activation function

DNN Outputs

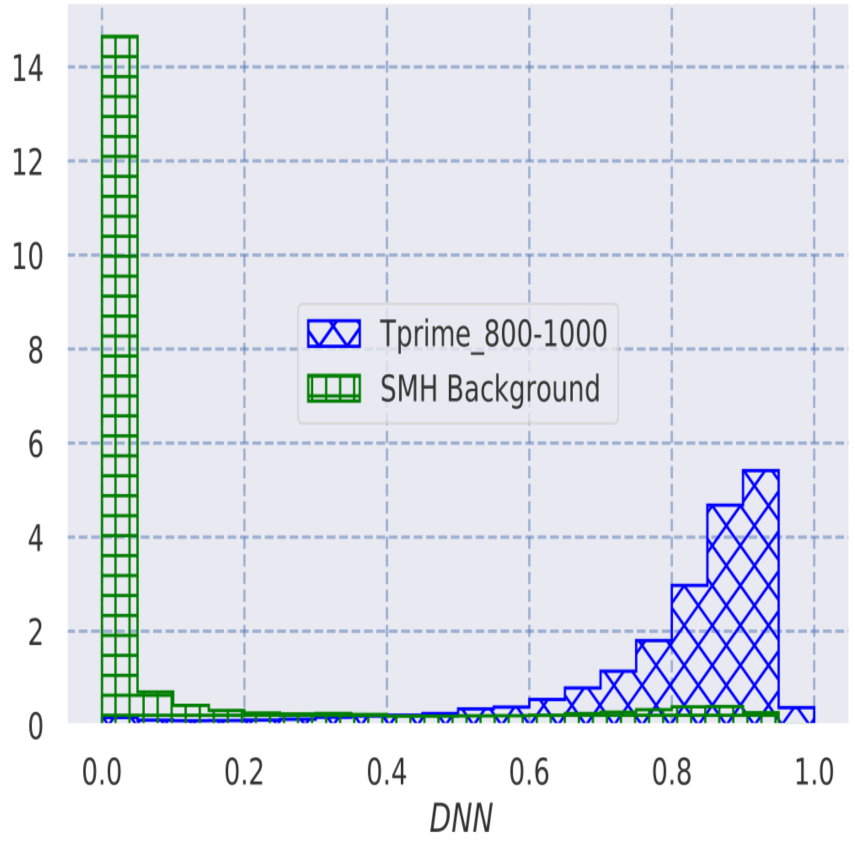
SM Higgs background training(DNN-SMH):

- Signal sample: $M_{T'} \in$ [600, 700], [800, 900], [1100, 1200]GeV
- Backgrounds: SM Higgs (ttH, tHq, VBF, VH, ggH)

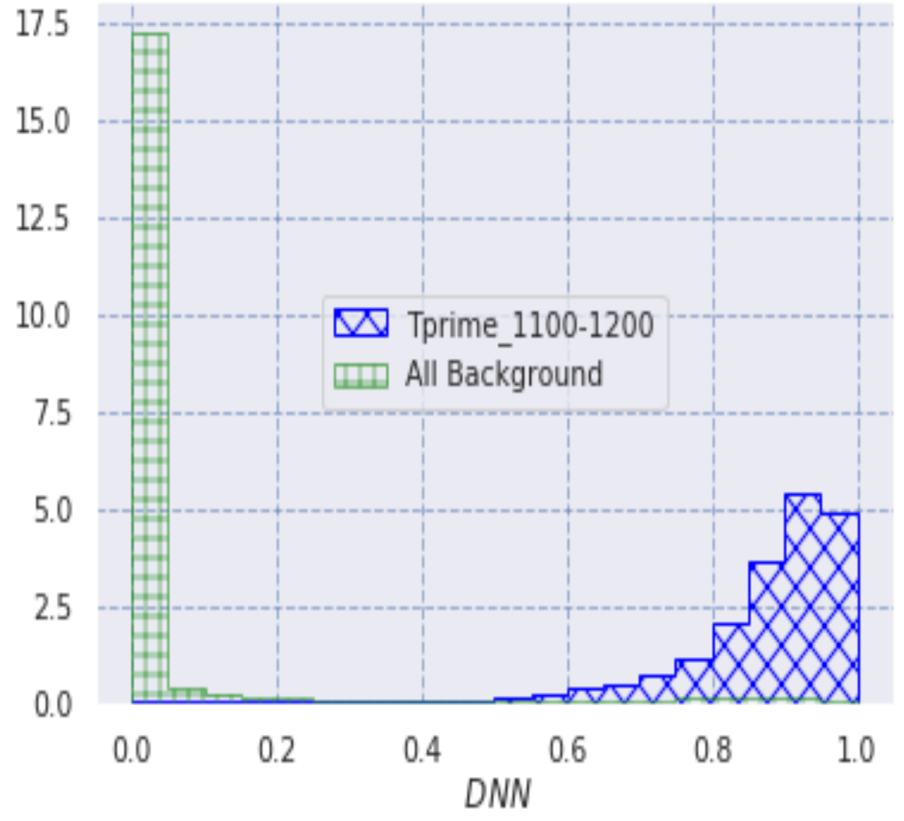
$M_{T'} \in [600,700]GeV$

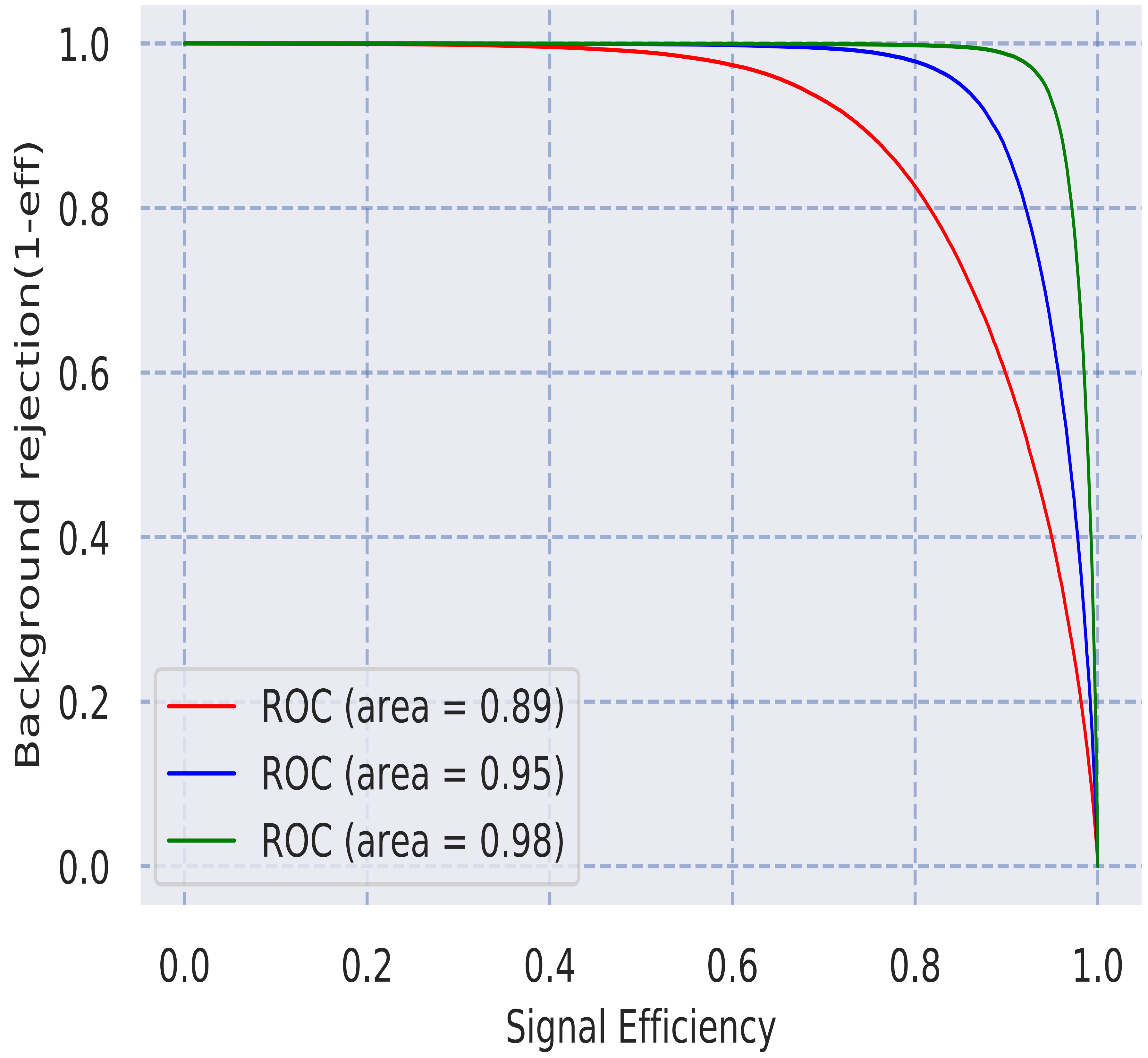


$M_{T'} \in [800,1000]GeV$



$M_{T'} \in [1100,1200]GeV$





Signal Selection vs Background rejection

The level of separation of signal and background increases with increase in the increasing mass of the T'

Analysis Strategy

Use of train test split → Training of DNN using SMH background → Testing on NRB, SMH, Data. → Ratio Plot → Scaled Plot → Cuts Applied → Limit Calculations

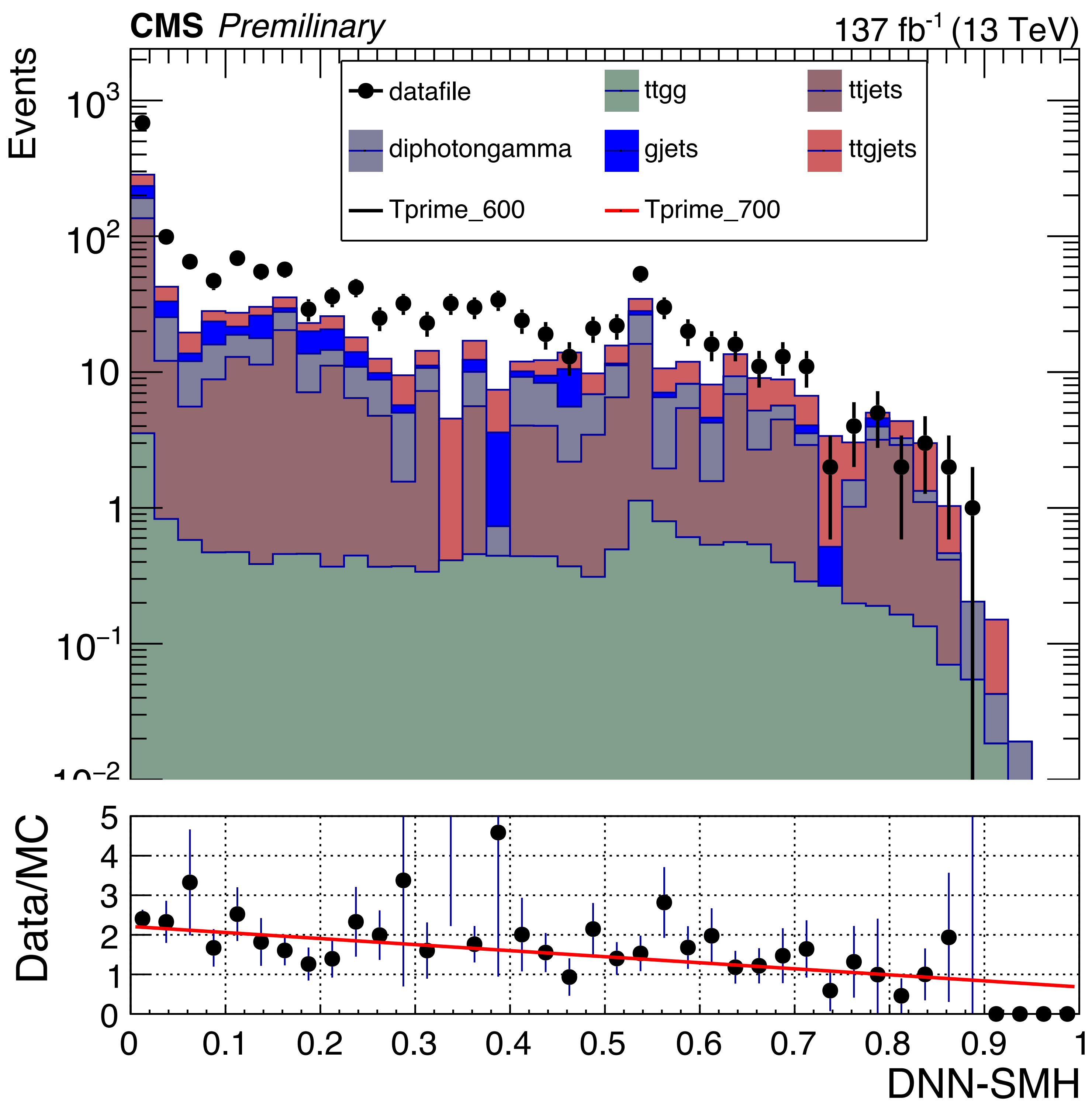
****The diphoton mass range, 115GeV $< M_{\gamma\gamma} <$ 135GeV is blinded for the data through out the analysis**

Separate trainings are performed in different T′ mass ranges

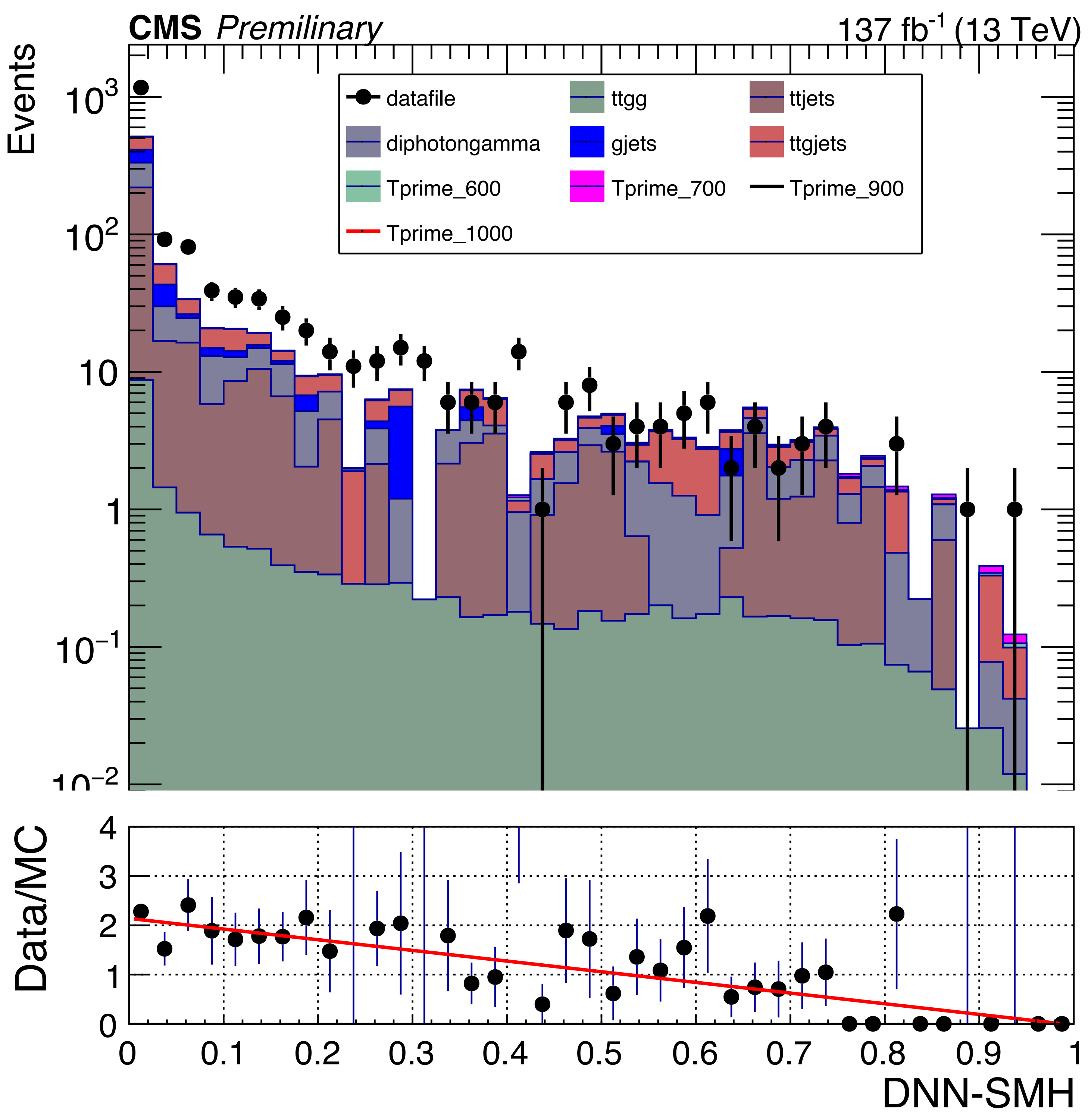
[600, 625, 650, 675, 700] [800, 900, 1000] [1100, 1200]

DNN Output Plot

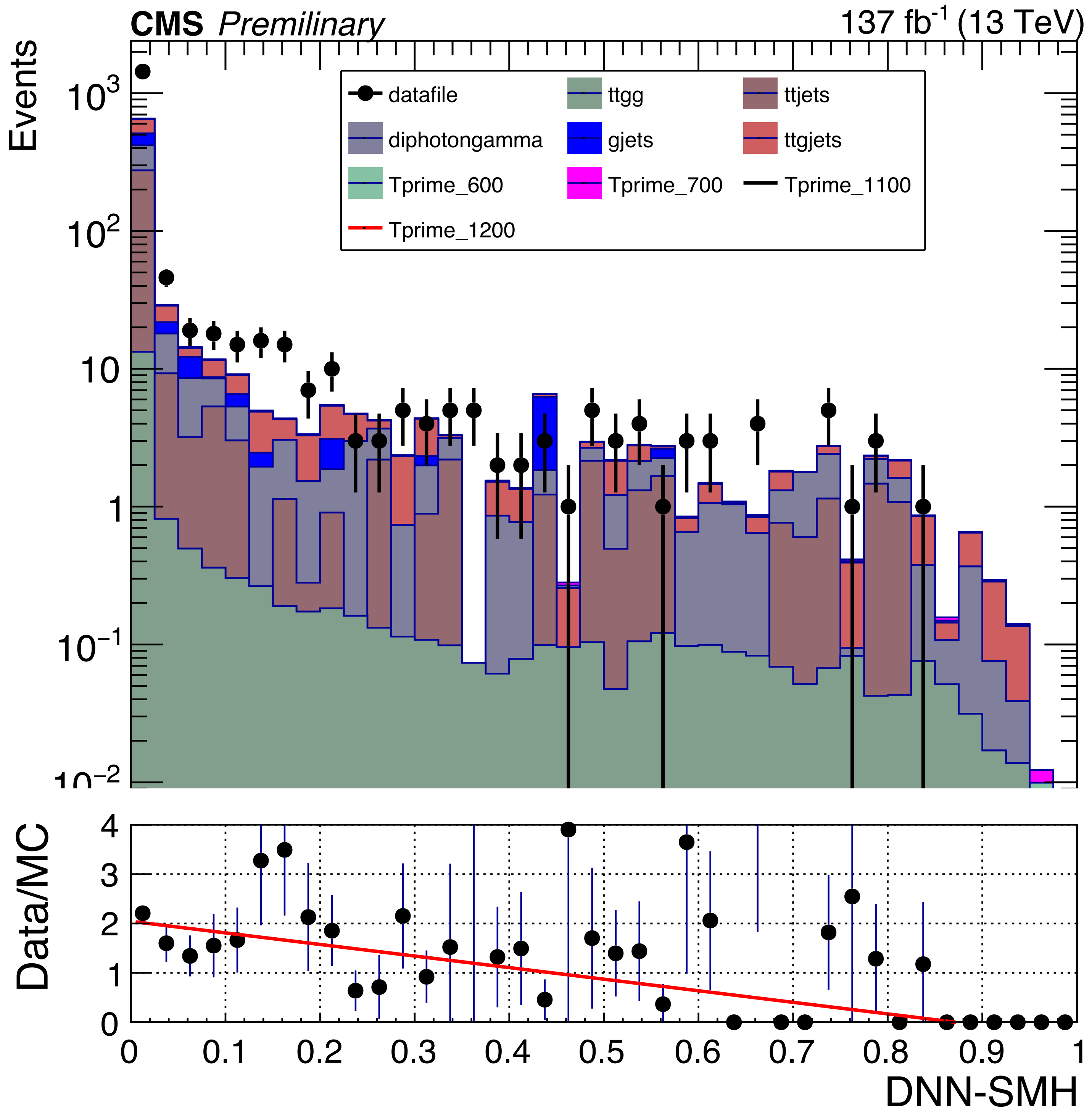
After testing the NRB backgrounds on the DNN model, the output along with the Run II data can be visualized as:



$M_{T'} \in [600,700] GeV$



$M_{T'} \in [800,1000] GeV$

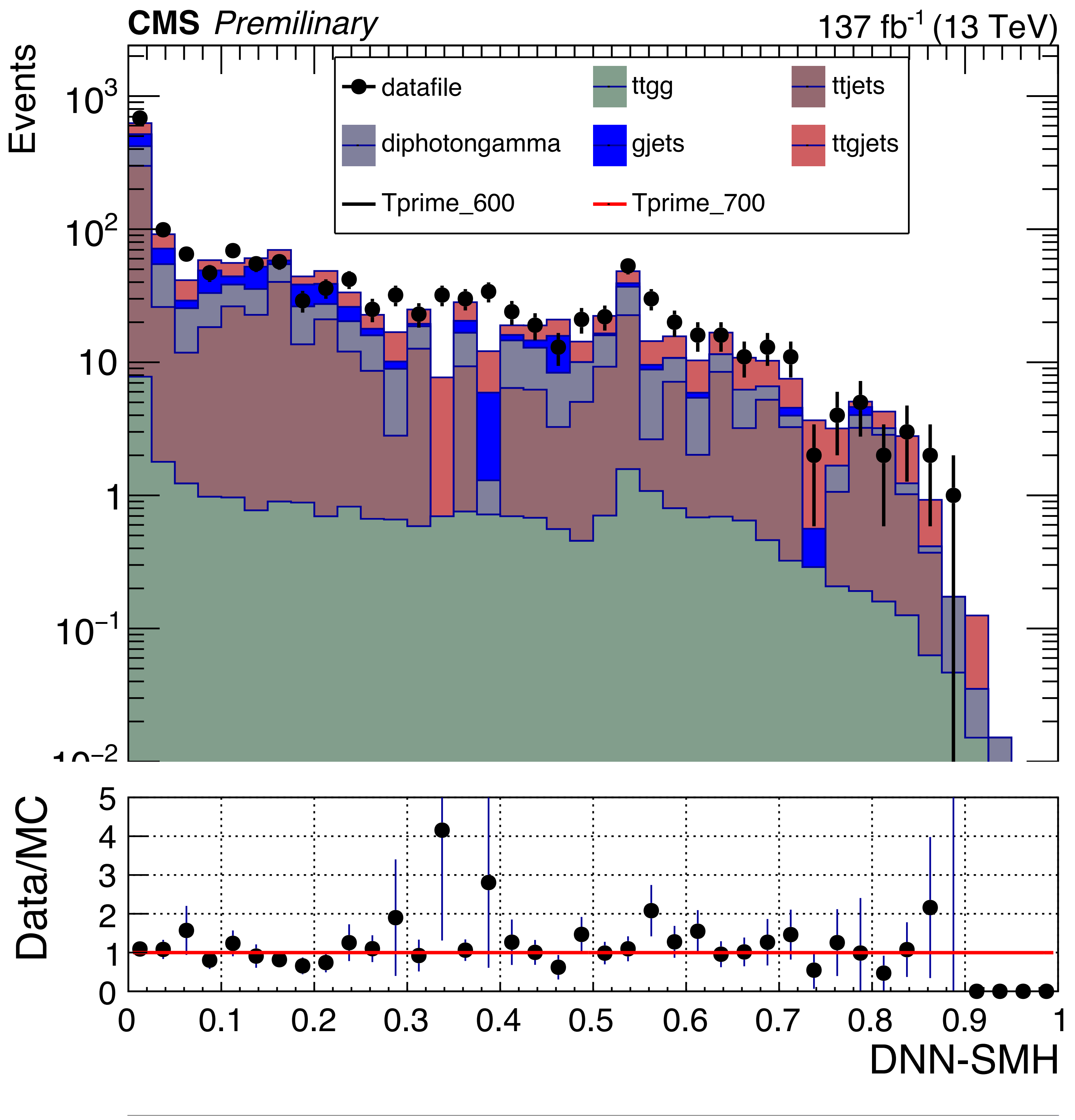


$M_{T'} \in [1100,1200] GeV$

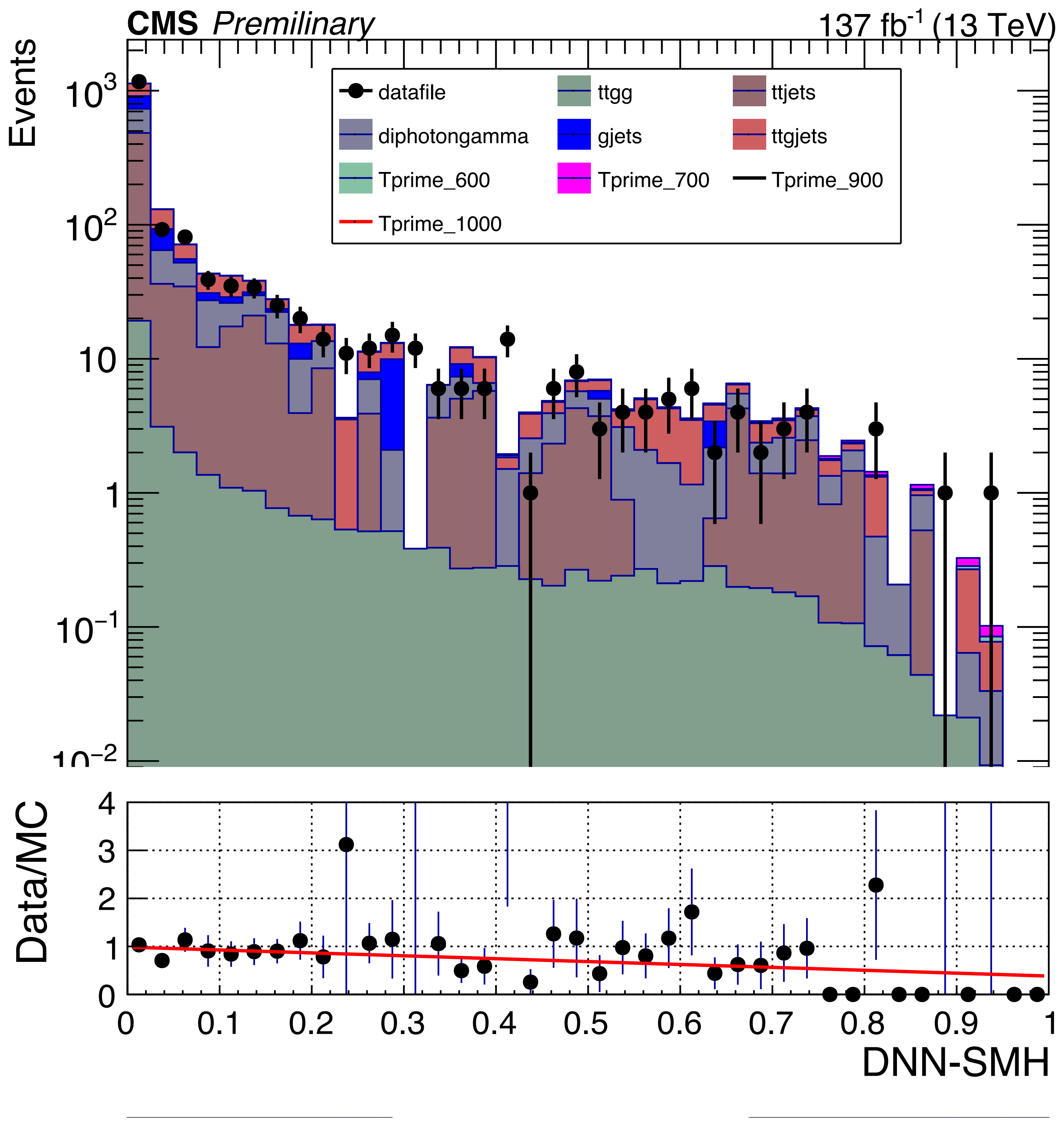
Slope	**-1.52828 +/- 0.342382**	**-2.16566 +/- 0.367164**	**-2.34415 +/- 0.477097**
Intercept	2.21165 +/- 0.155941	2.13988 +/- 0.130903	2.04436 +/- 0.122349

DNN Output Plot (after Scaling)

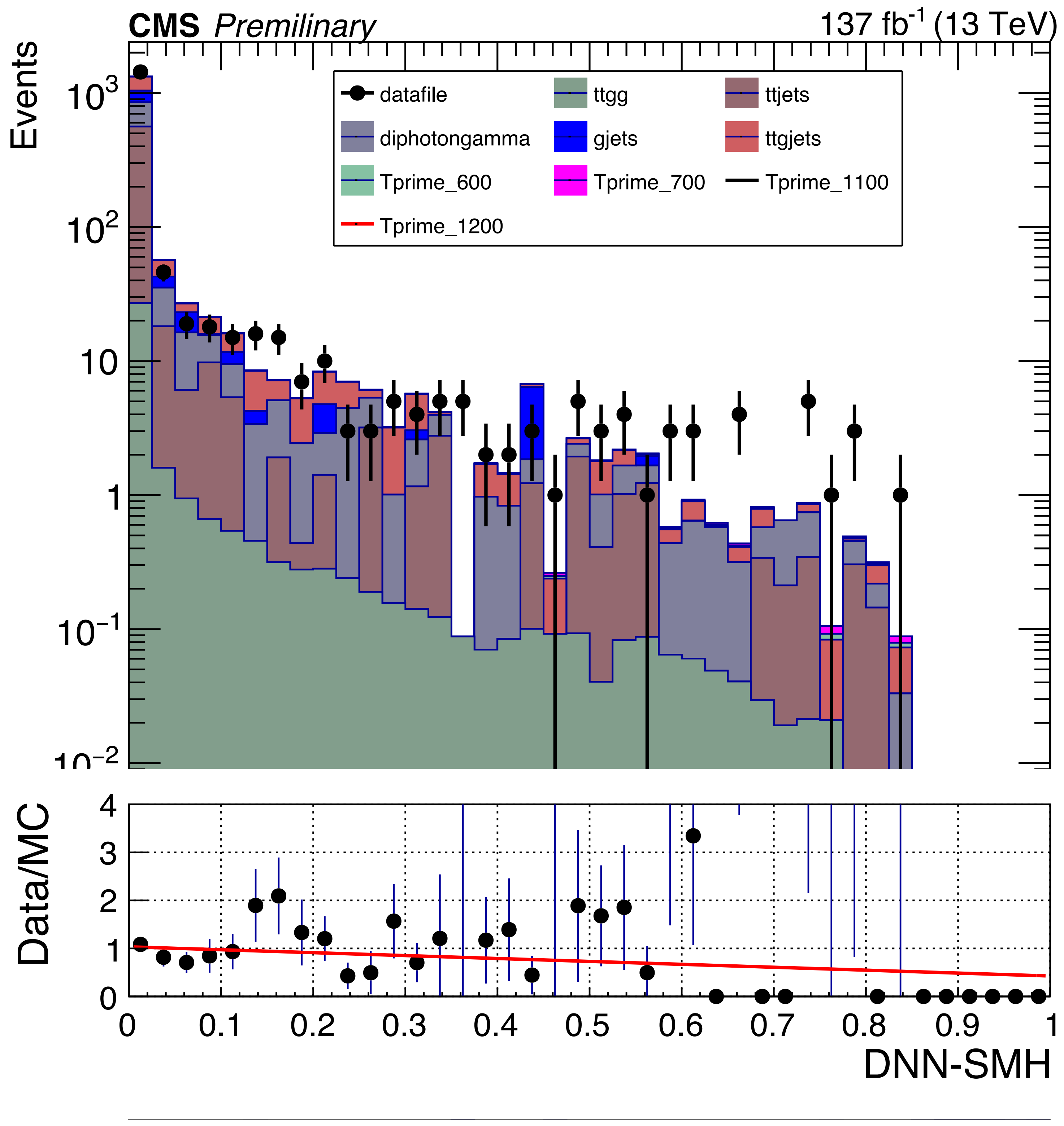
- Data and Monte Carlo mismatch have been scaled after linear fitting of the ratio plot.



$M_{T'} \in [600,700] GeV$

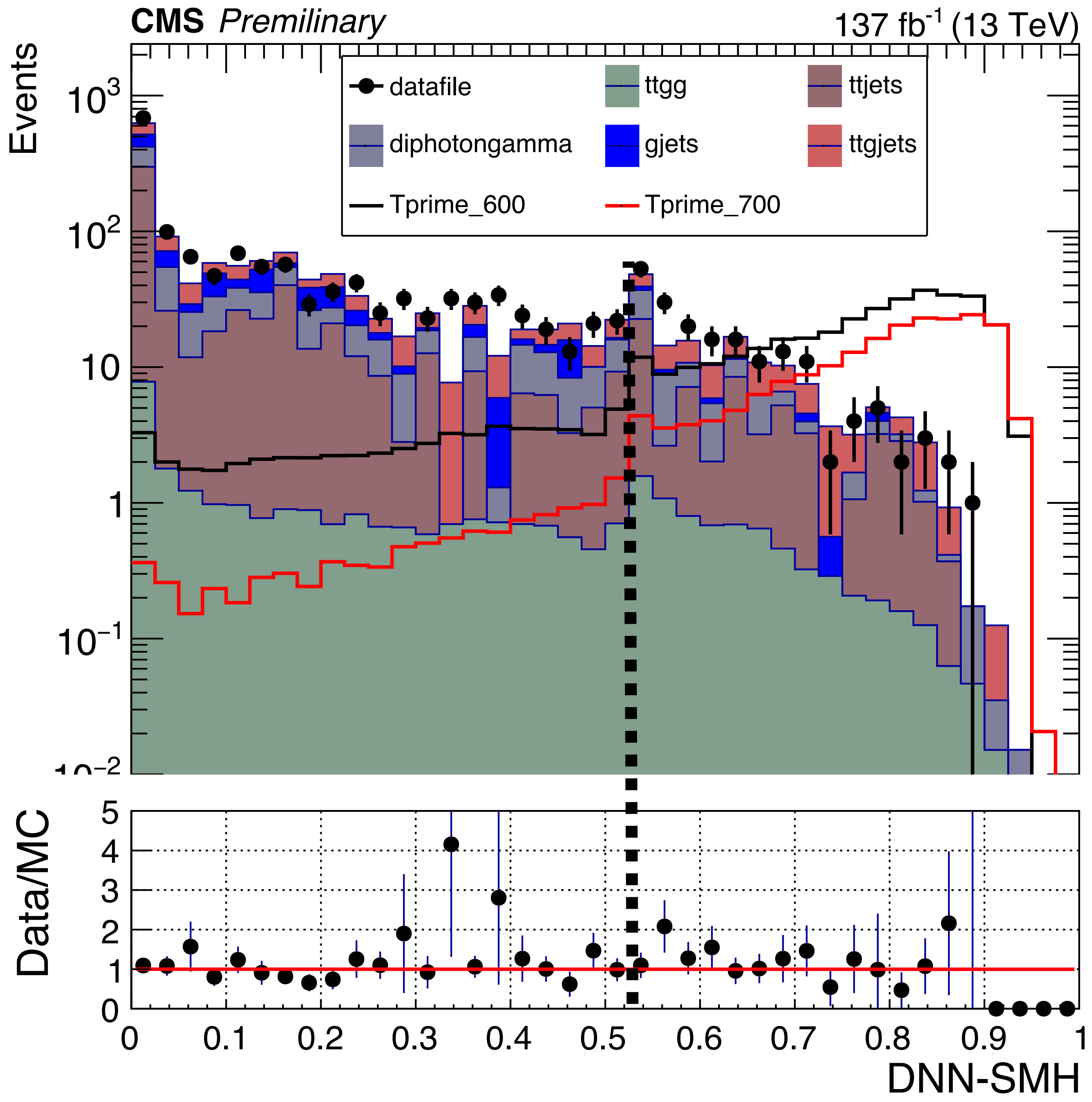


$M_{T'} \in [800,1000] GeV$

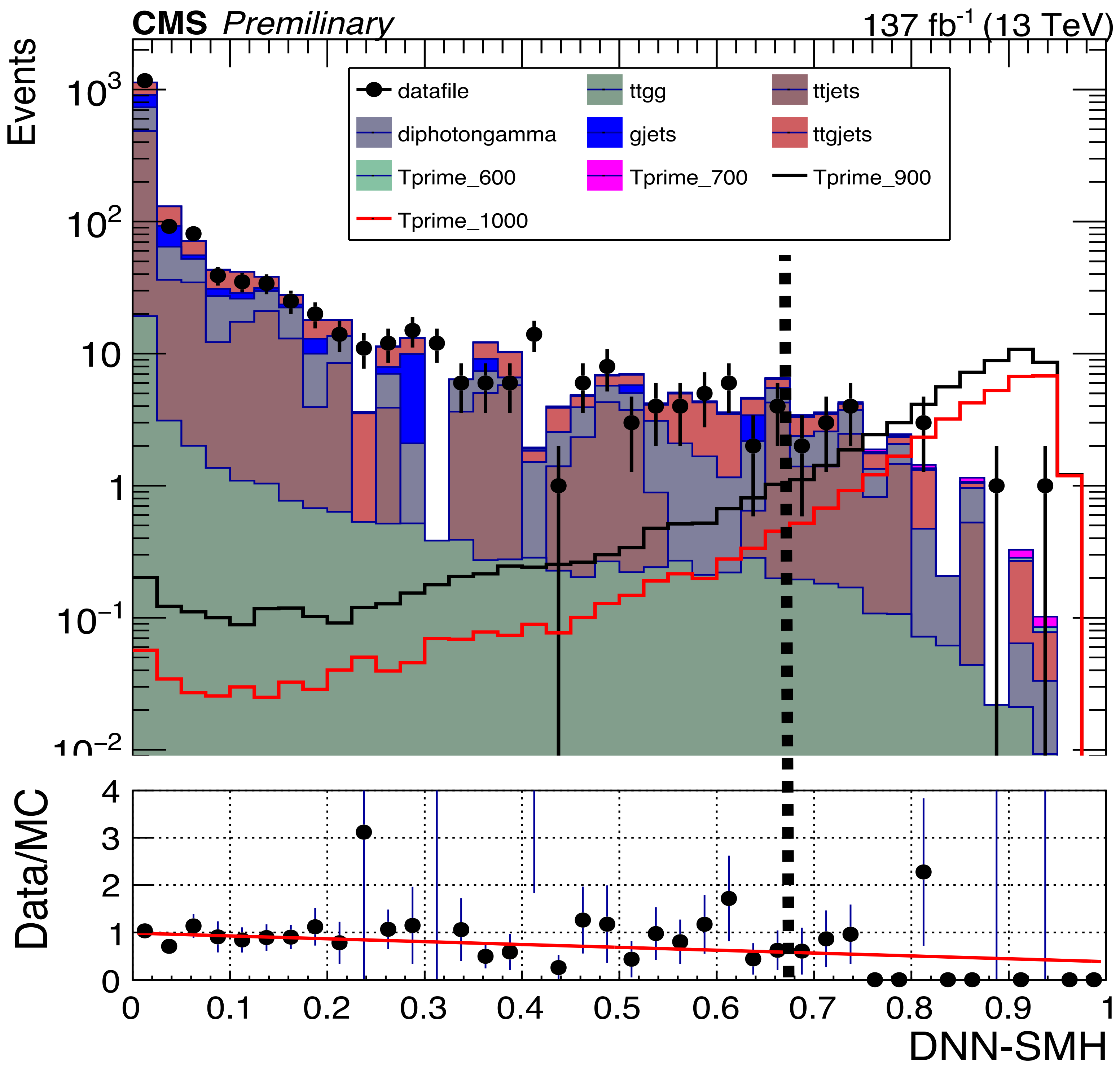


$M_{T'} \in [1100,1200] GeV$

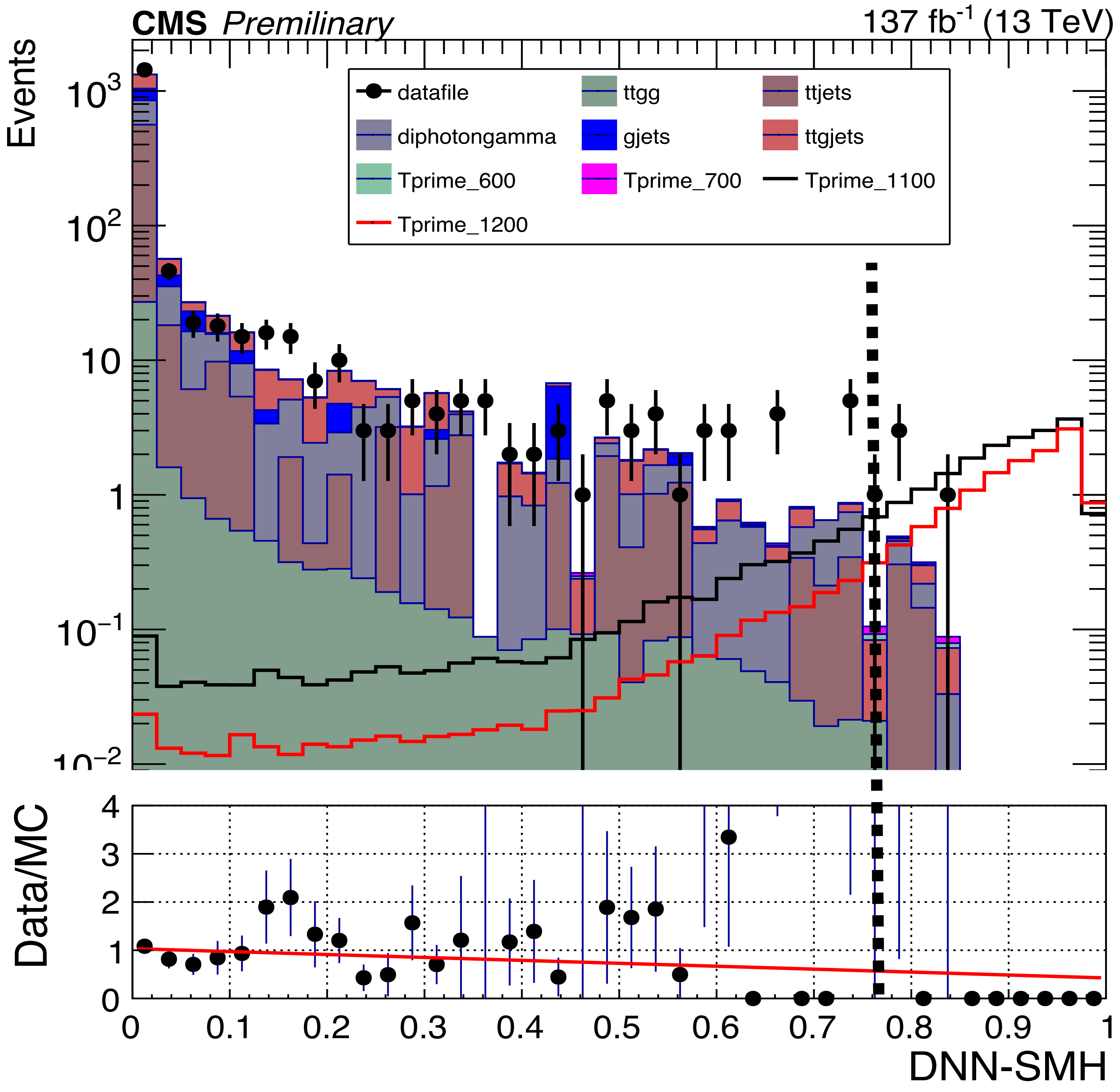
DNN Output Plot (after Scaling with Signal)



$M_{T'} \in [600,700]GeV$



$M_{T'} \in [800,1000]GeV$



$M_{T'} \in [1100,1200]GeV$

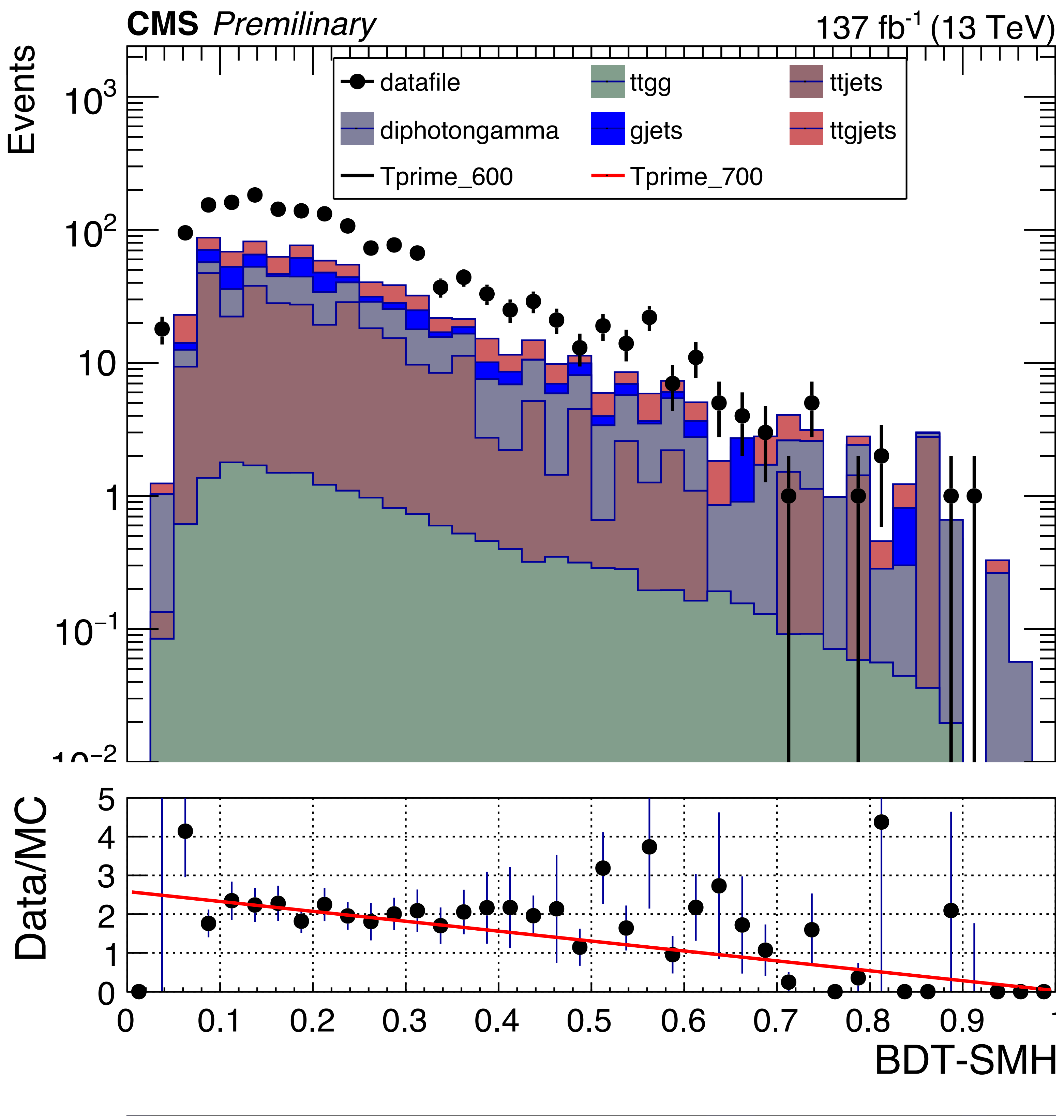
- After scaling the DNN output, for background rejections we cuts applied for different training at:

$M_{T'} \in [600,700]GeV$ > 0.54

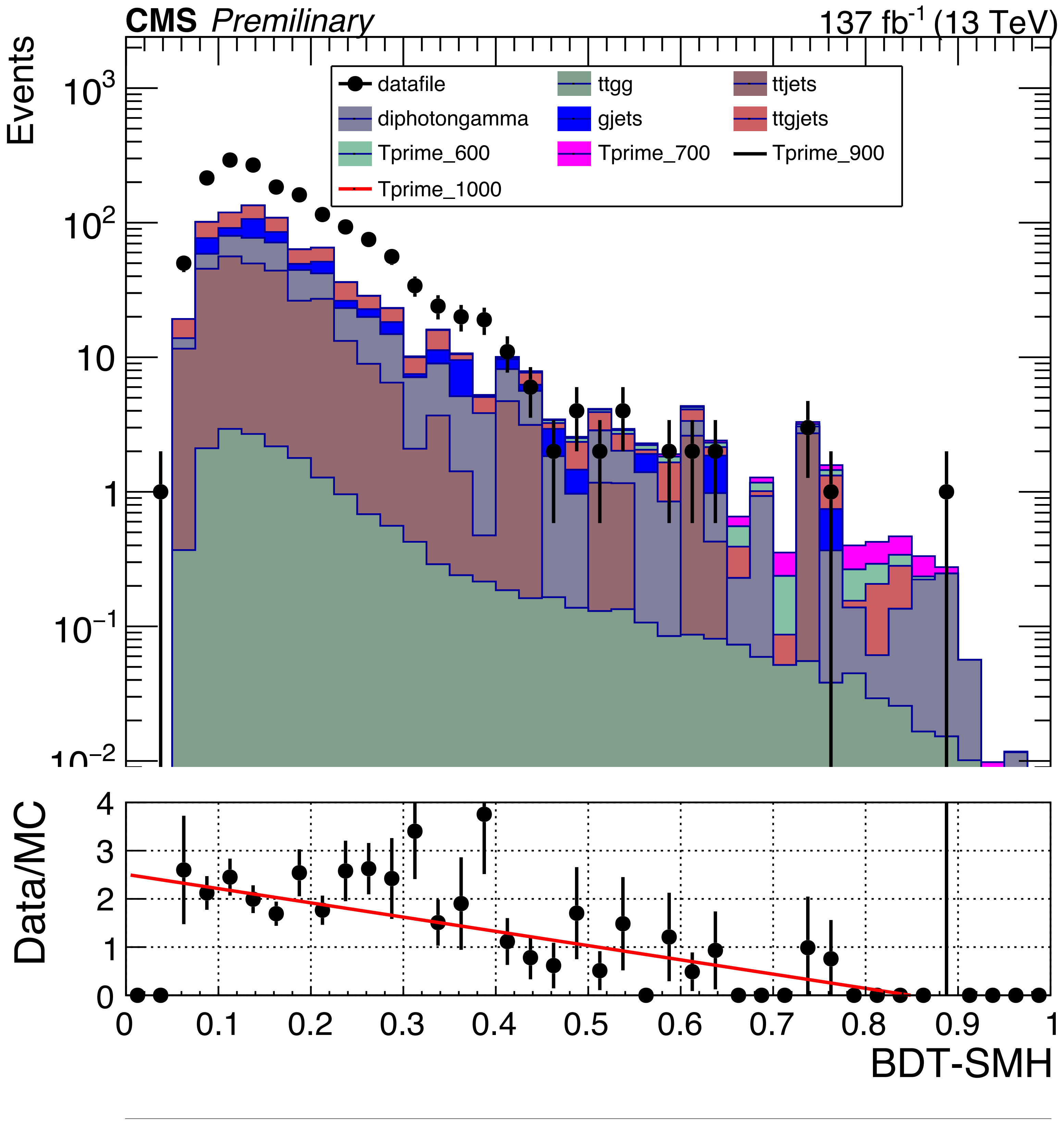
$M_{T'} \in [800,1000]GeV$ > 0.76

$M_{T'} \in [1100,1200]GeV$ > 0.80

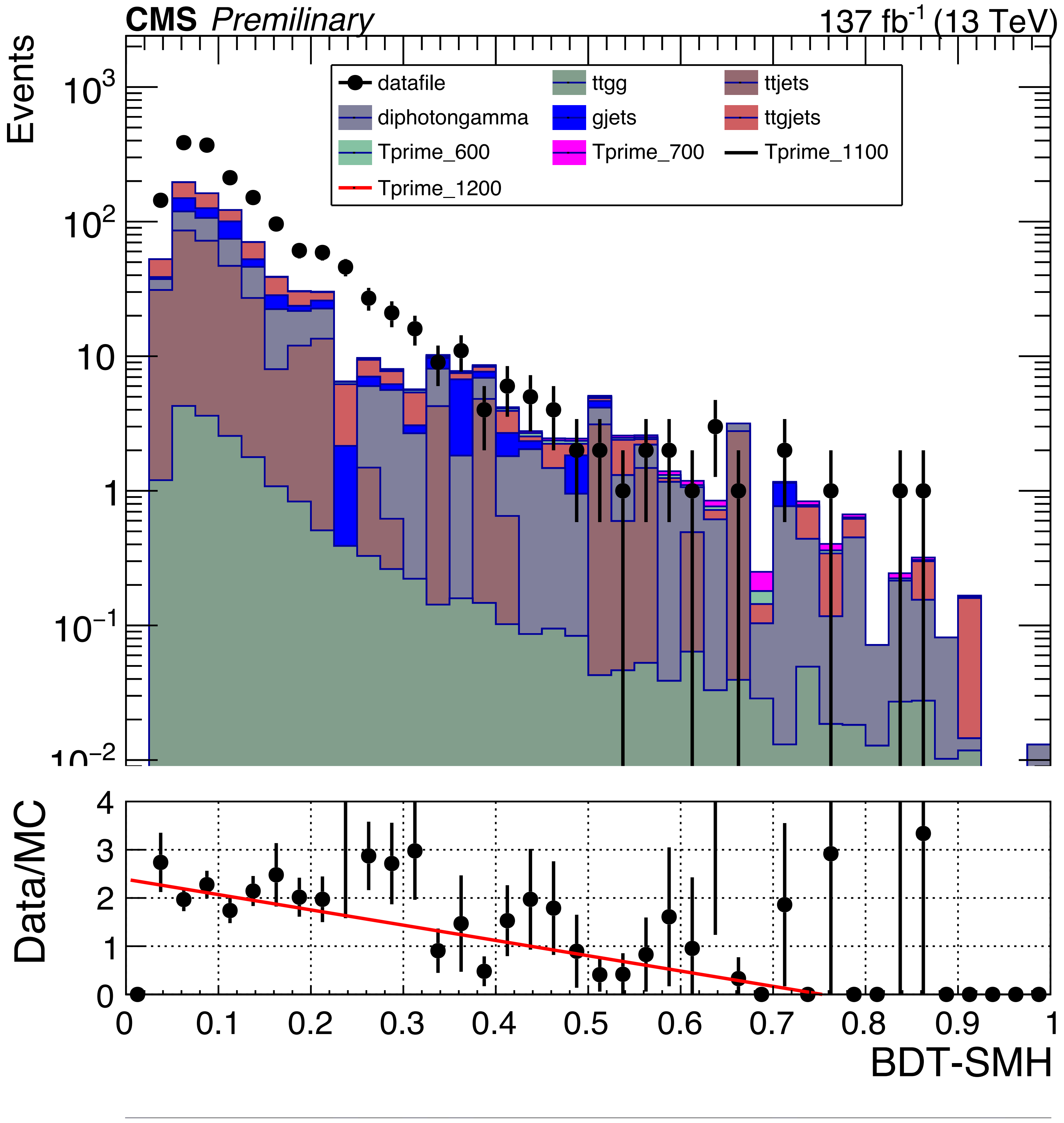
Present CMS analysis(BDT)



$M_{T'} \in [600,700] GeV$



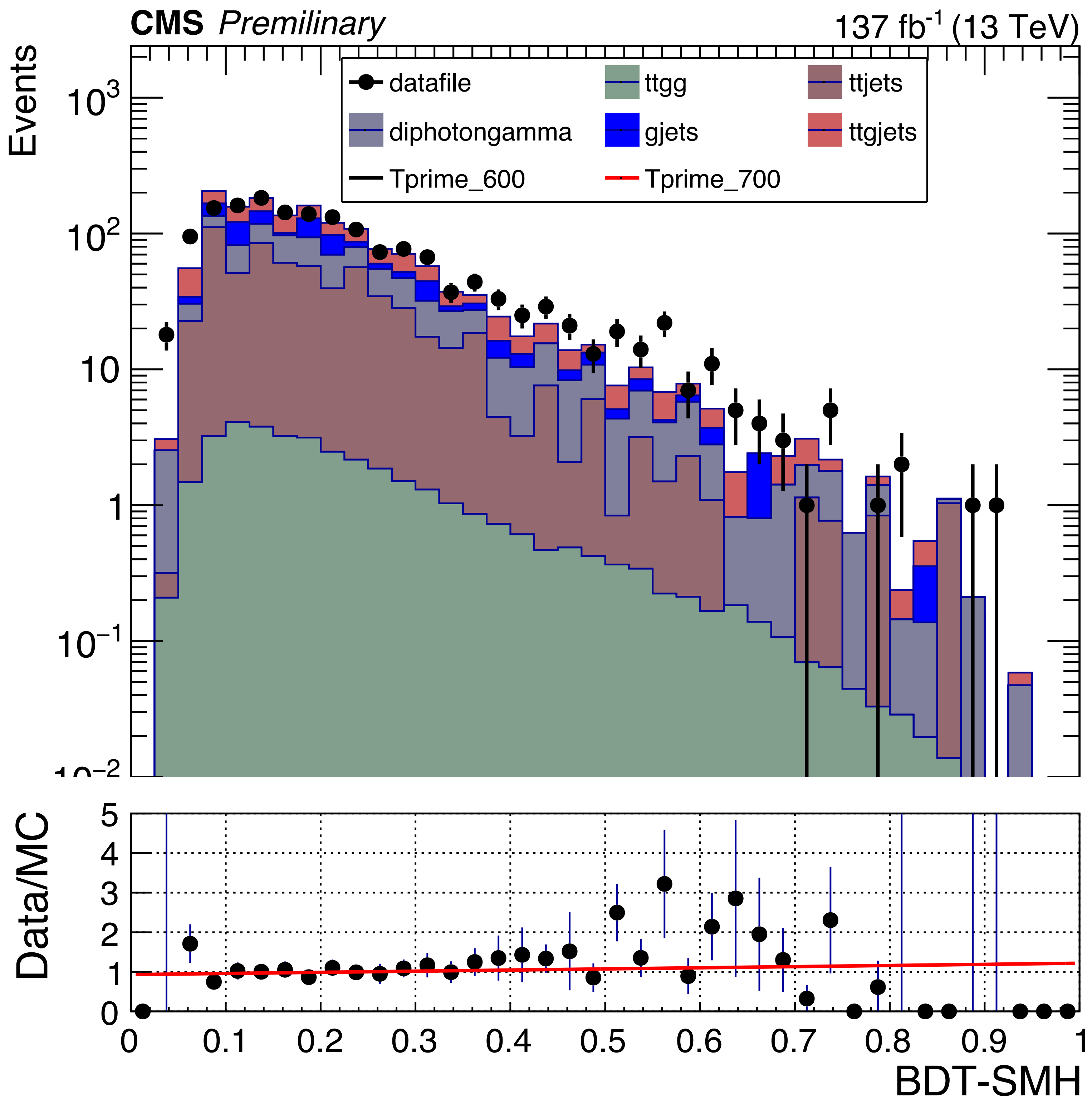
$M_{T'} \in [800,1000] GeV$



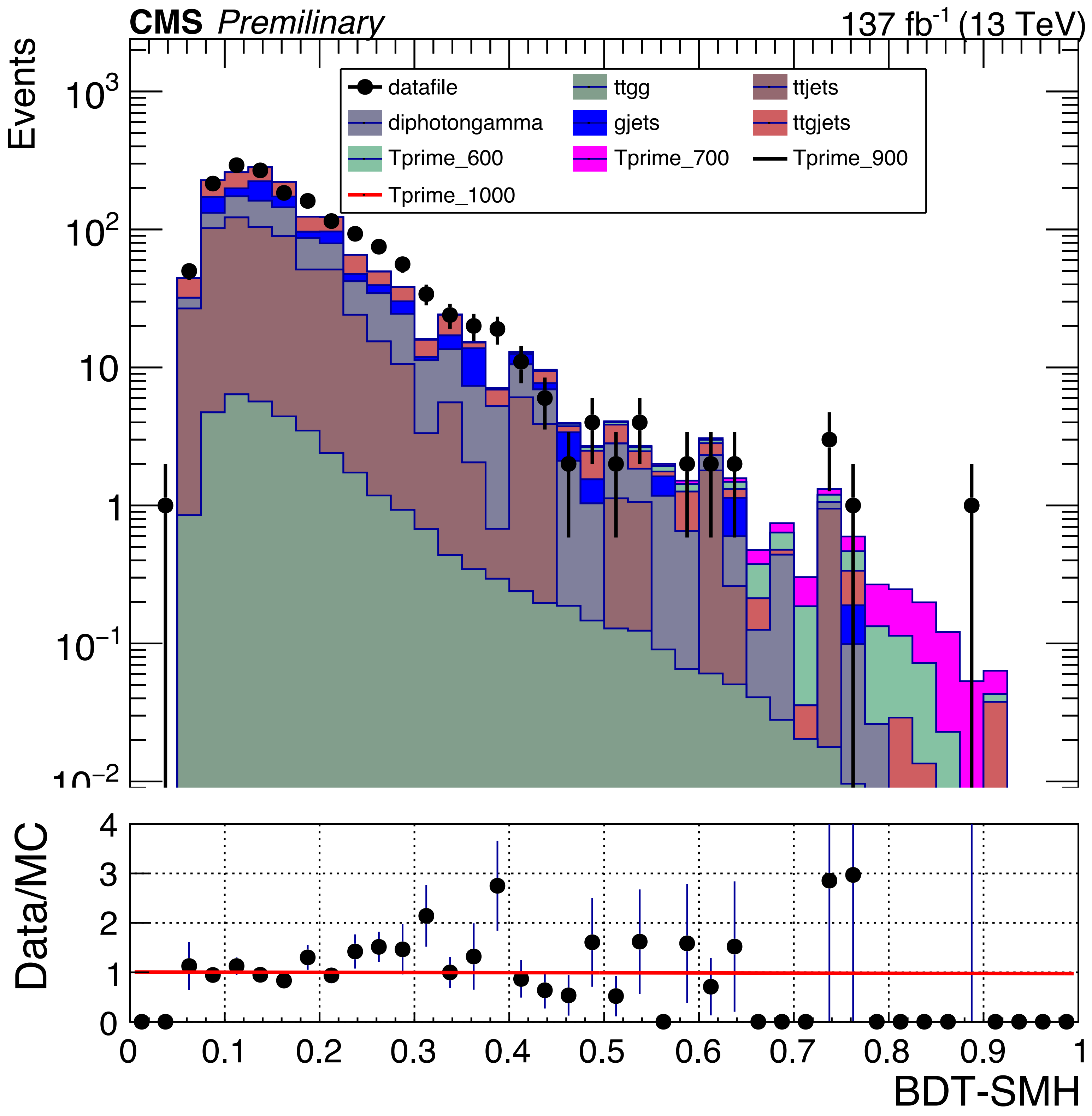
$M_{T'} \in [1100,1200] GeV$

Slope	**-2.55526 ± 0.406623**	**-2.95575 ± 0.537814**	**-3.16991 ± 0.483748**
Intercept	2.58306 ± 0.181075	2.50859 ± 0.182962	2.38726 ± 0.153541

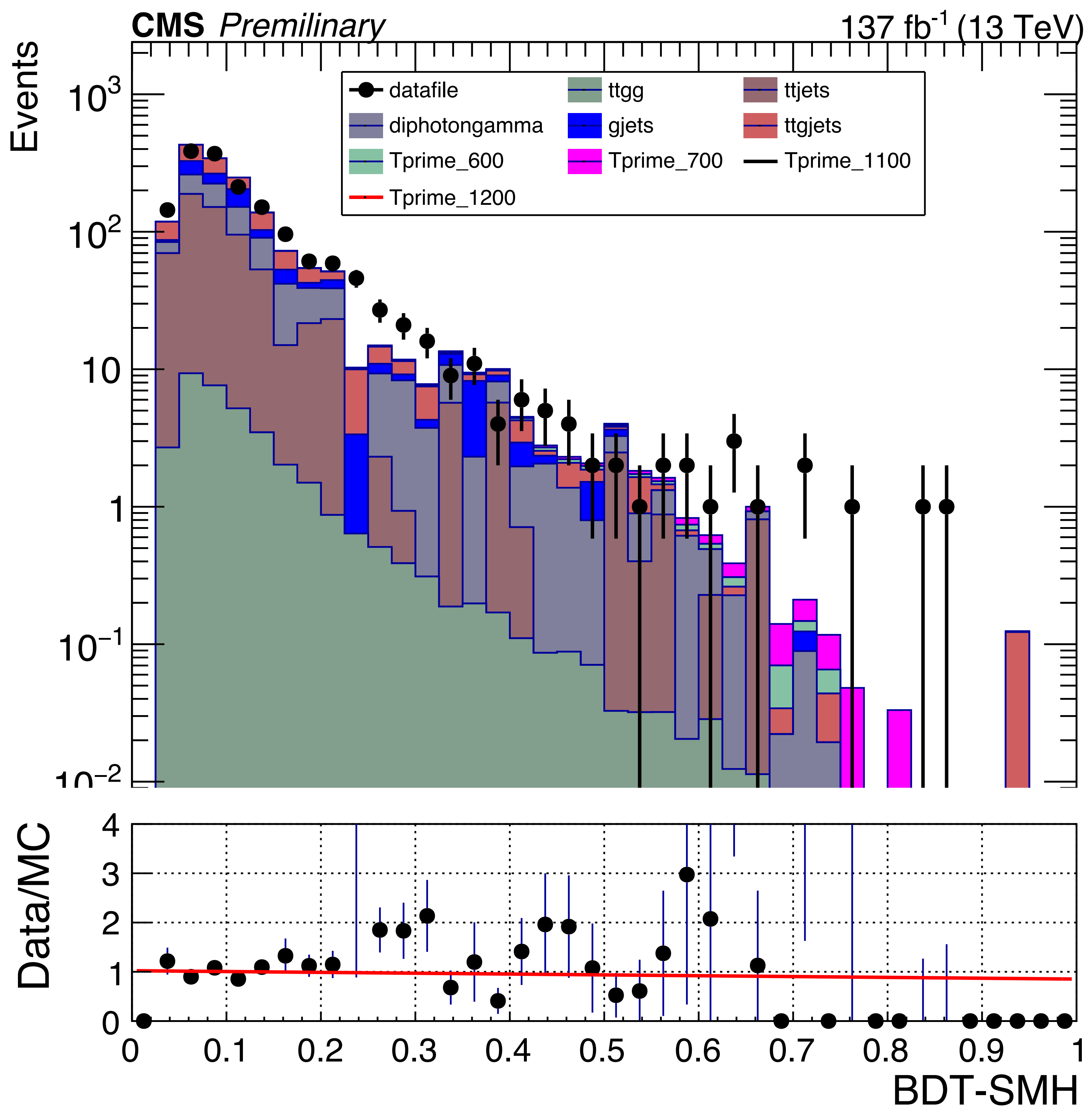
present CMS analysis(BDT) (after Scaling)



$M_{T'} \in [600,700] GeV$

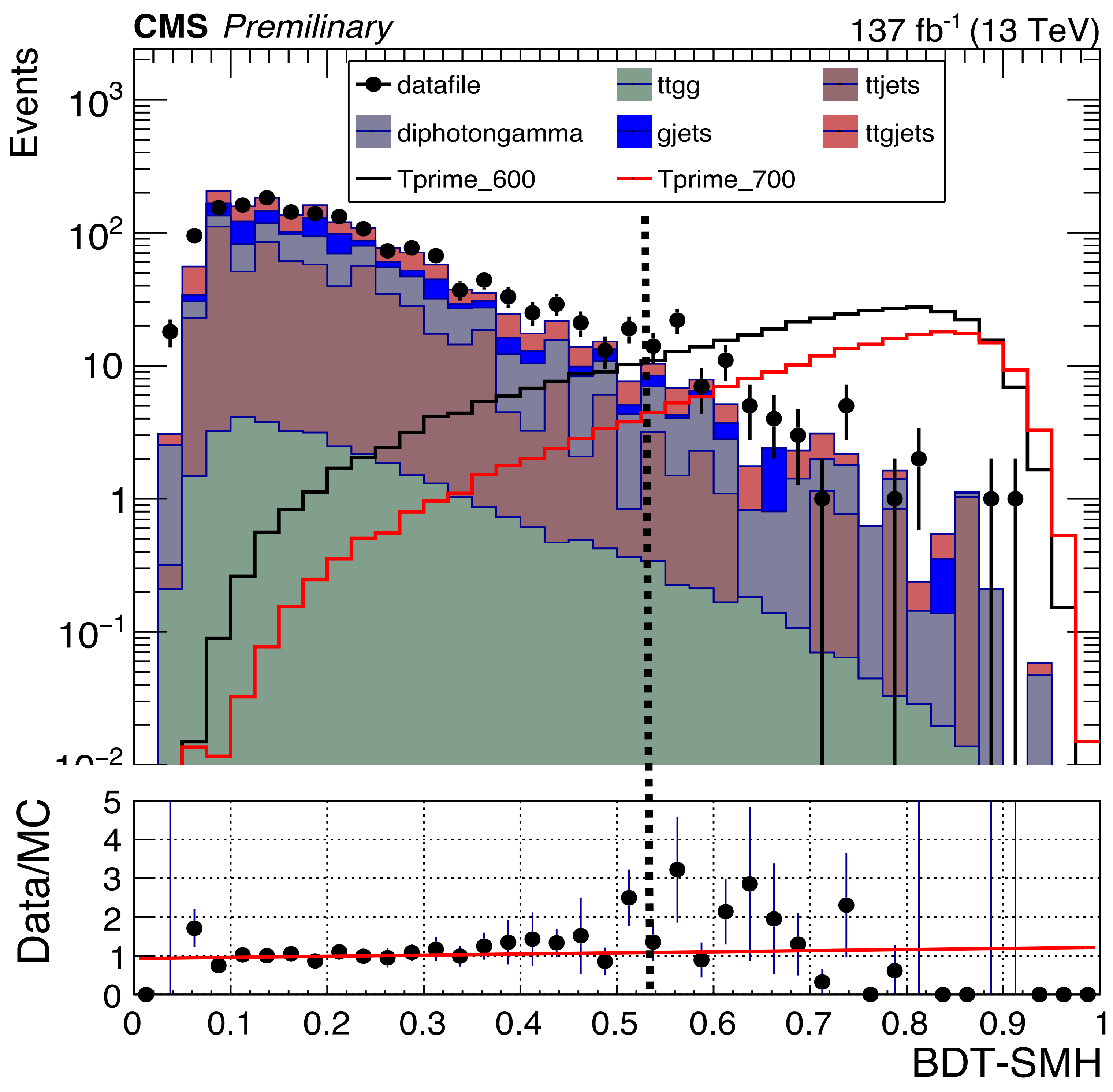


$M_{T'} \in [800,1000] GeV$

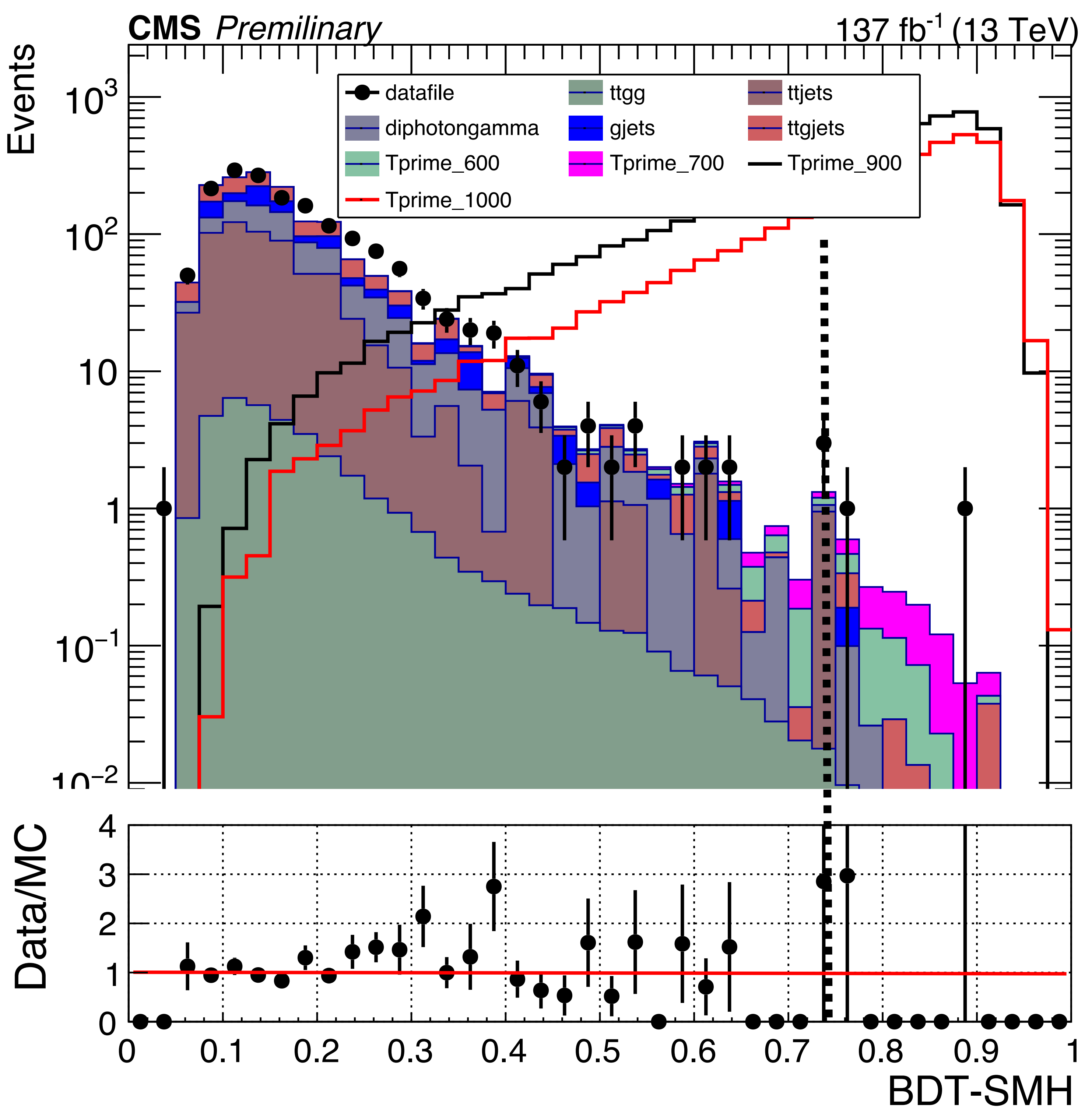


$M_{T'} \in [1100,1200] GeV$

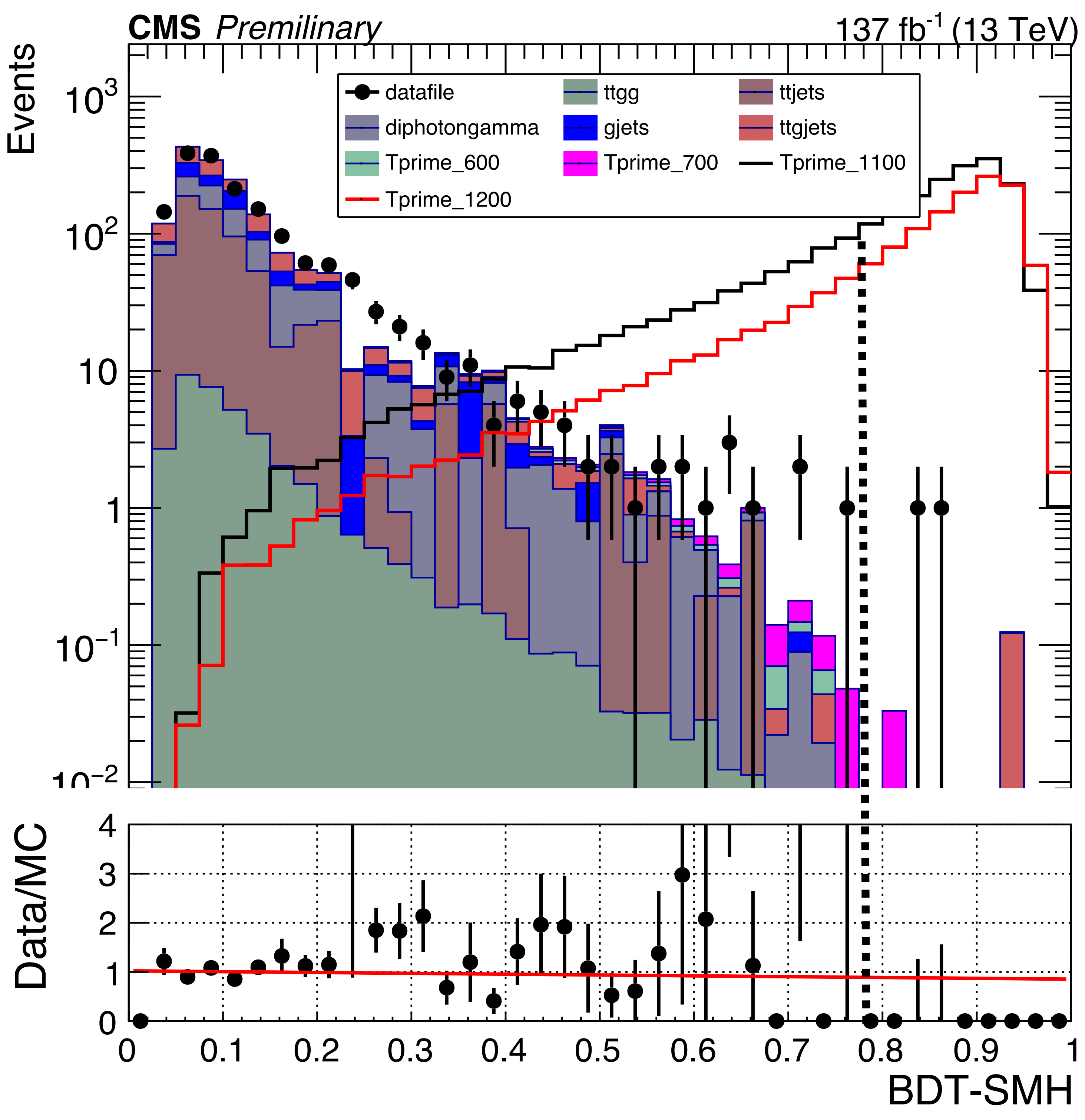
BDT Output Plot (after Scaling with Signal)



$M_{T'} \in [600{,}700] GeV$



$M_{T'} \in [800{,}1000] GeV$



$M_{T'} \in [1100{,}1200] GeV$

- After scaling the BDT output, as similar to DNN, for background rejections we applied cuts for different training at:

$M_{T'} \in [600{,}700] GeV$	0.54
$M_{T'} \in [800{,}1000] GeV$	0.76
$M_{T'} \in [1100{,}1200] GeV$	0.80

Probability of T' events

The probability of obtaining of getting T' events:

$$p(T'events) = \frac{L\sigma(pp \longrightarrow T')}{L\sigma(pp)}$$

Probability to observe n_T^{obs} T′ events out of N p-p collision

$$P(n_H^{obs}) = \frac{N!}{n_{T'}^{obs}(N - n_{T'}^{obs})!} p^{n_{T'}^{obs}} (1-p)^{N-n_{T'}^{obs}}$$

$$forN \longrightarrow \infty, p^{n_T^{obs}} = Poiss(n_{T'}^{obs}, \lambda) = \frac{e^{-\lambda}\lambda^{n_{obs}^{T'}}}{n_{T'}^{obs}!}$$

$$\lambda = Np = L\sigma(pp)\frac{L\sigma(pp \longrightarrow T')A_{eff}}{L\sigma(pp)} = n_{T'}^{exp}$$

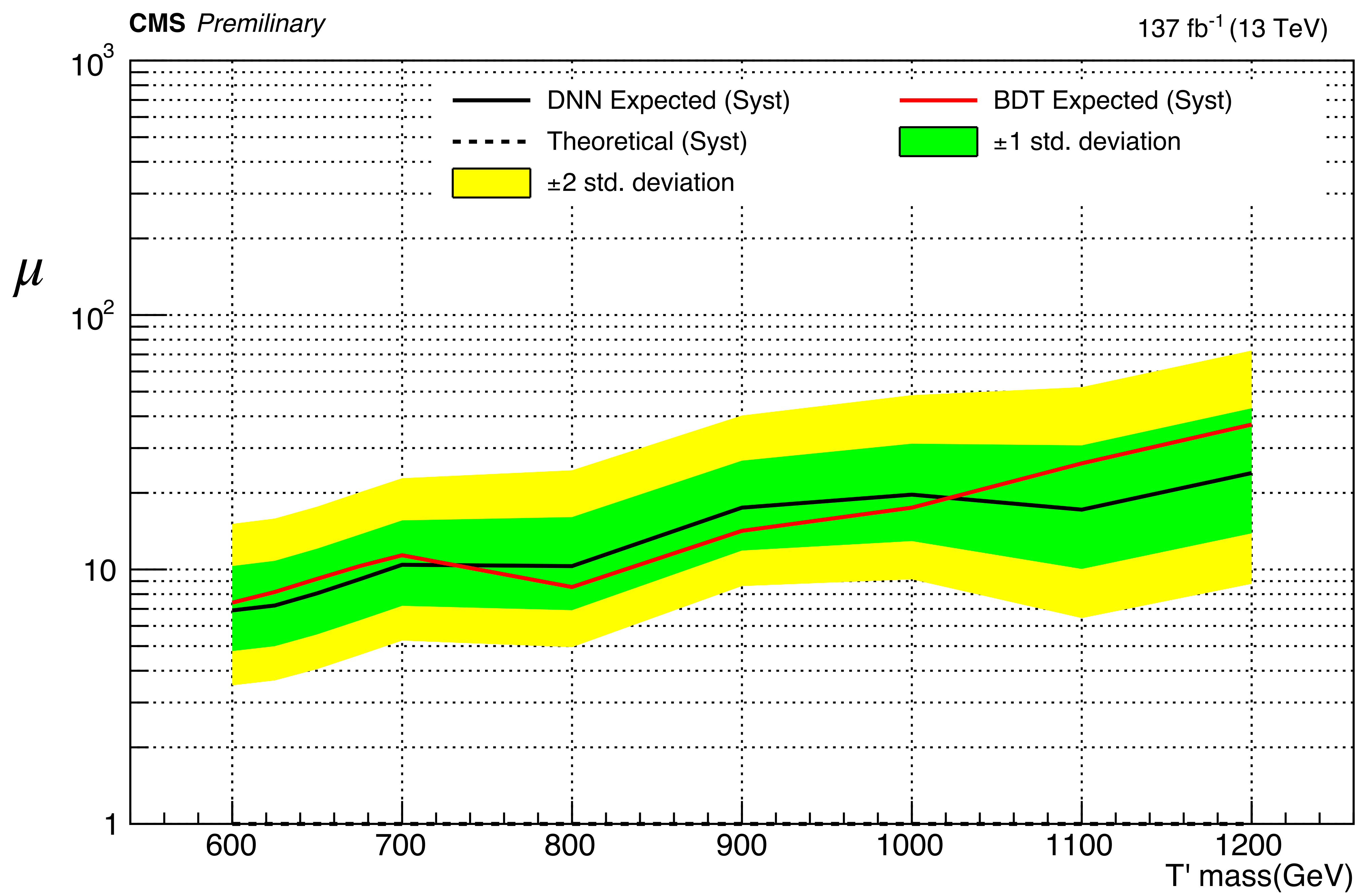
Or a counting experiment

$$n_{T'}^{exp} = \mu s + b, \mu = \frac{\sigma_{obs}(T')}{\sigma_{SM}(T')}$$

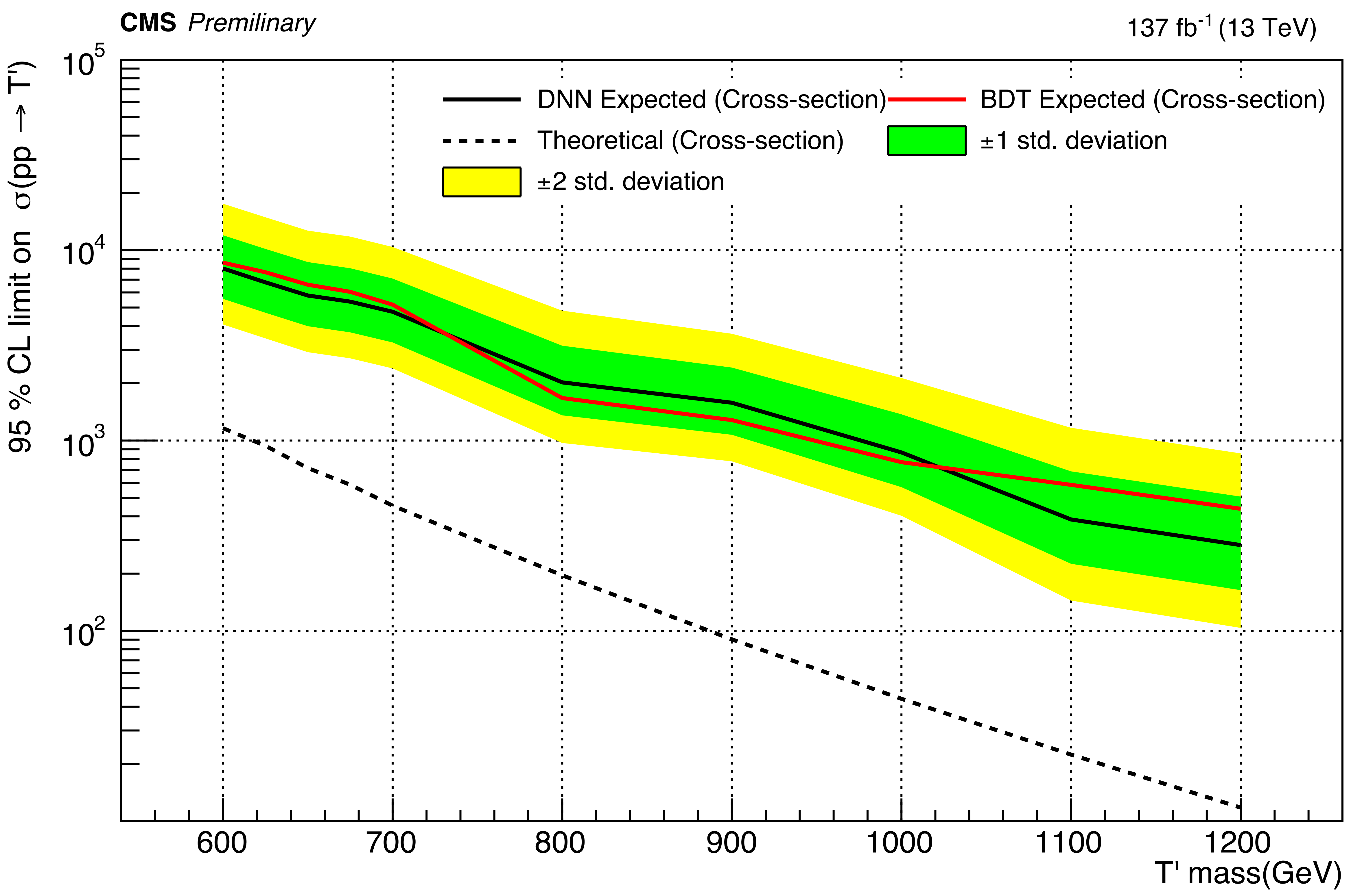
Samples	Total number of events	Number of events for counting
Signal		
Tprime 800	101761	61834
Tprime 900	100962	81552
Tprime 1000	105546	92489
SMH		
ttgg	55847	18016
ttjets	1320	642
diphotongamma	8323	2882
gjets	228	86
ttgjets	9441	3375
NRB		
tHq	338874	89800
ttH	92486	19000
vbfH	1199	235
vH	8080	1174
ggH	529	111

Comparison DNN and BDT

Using the number of events after scaling all the signal and background, with Higgs Combined Tools, the expected value of signal strength has been calculated.



Expected 95% CL limit on signal strength(μ)



Expected 95% CL limit on Cross-Section

Summary and Conclusion

- Deep Neural Network techniques have been explored in this thesis project.

- Extensively studied multivariate techniques to separate arbitrary signal (T') from the standard model Higgs(SMH) backgrounds processes using the CMS detector simulation and the reconstruction program, considering 137fb^{-1} of integrated luminosity at $\sqrt{s}$ = 13TeV.

- The performance of DNN has been well evaluated over the Non Resonant background processes.

- The expected limits at 95% CL on T' production processes have been extracted at each T' mass in the range [600,1200] using the DNN based selection criteria, as well as for the present CMS analysis(BDT).

- DNN-based analysis sensitivity is yet to reach the exclusion potential.

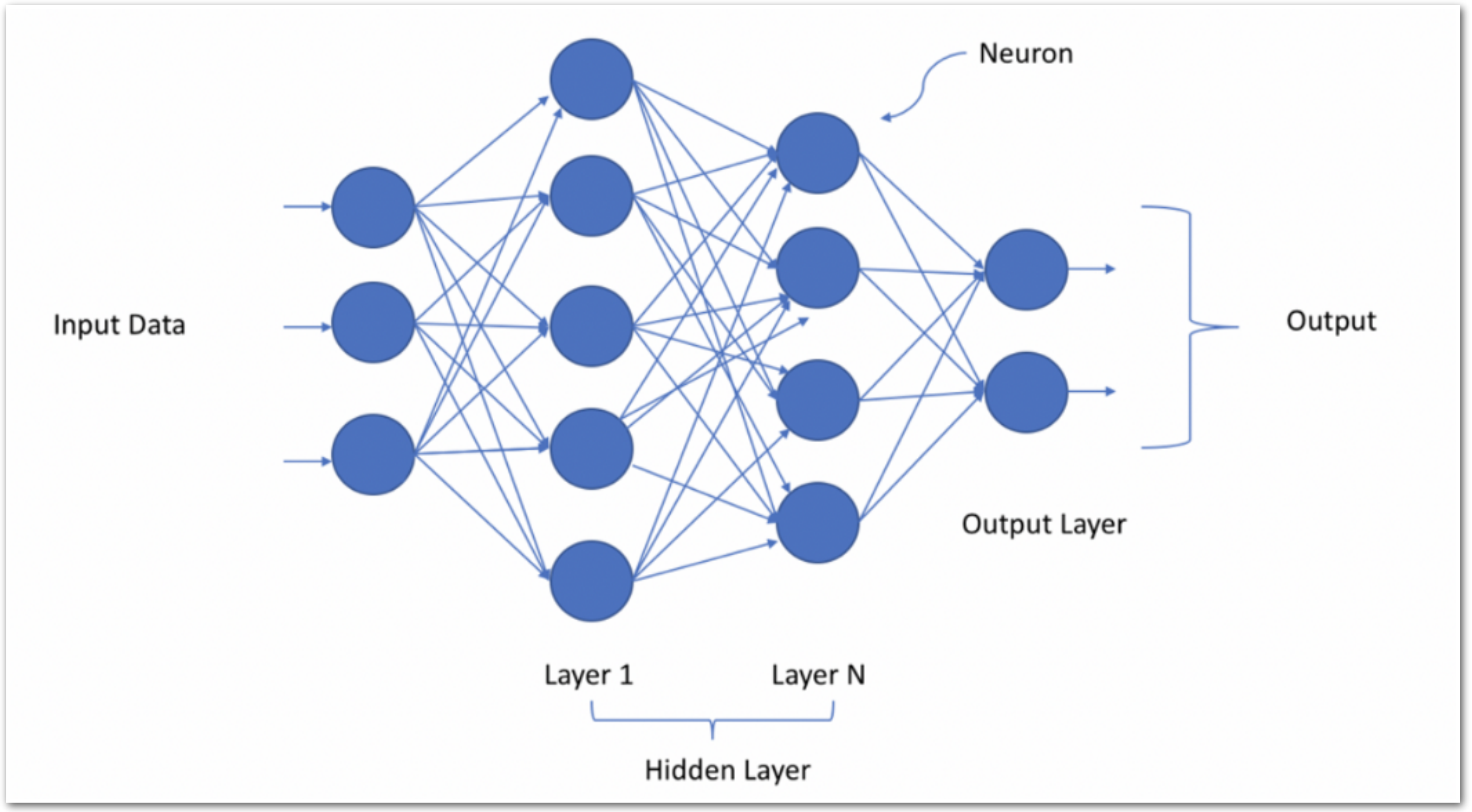
Thank you

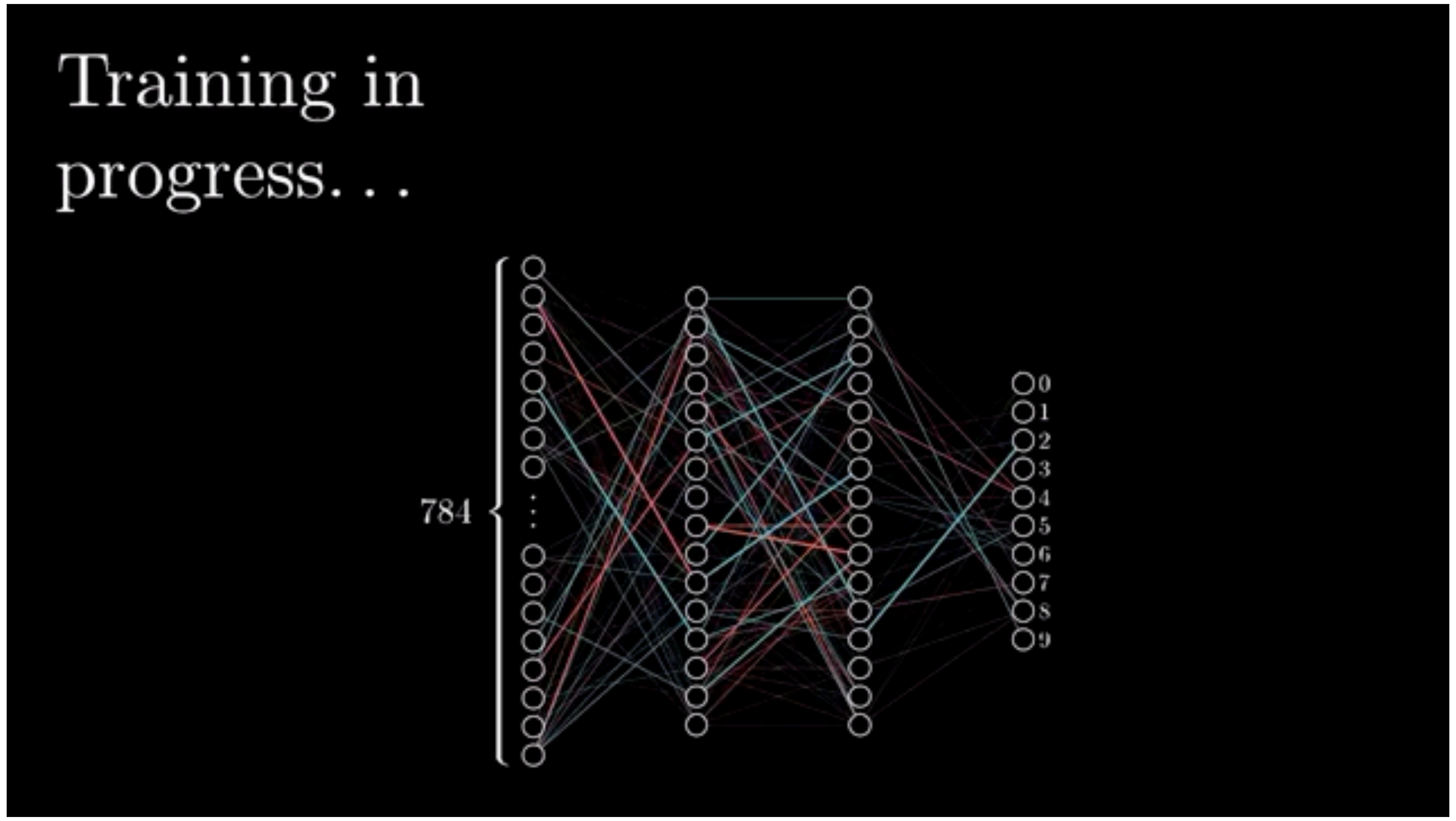
Backup Slides

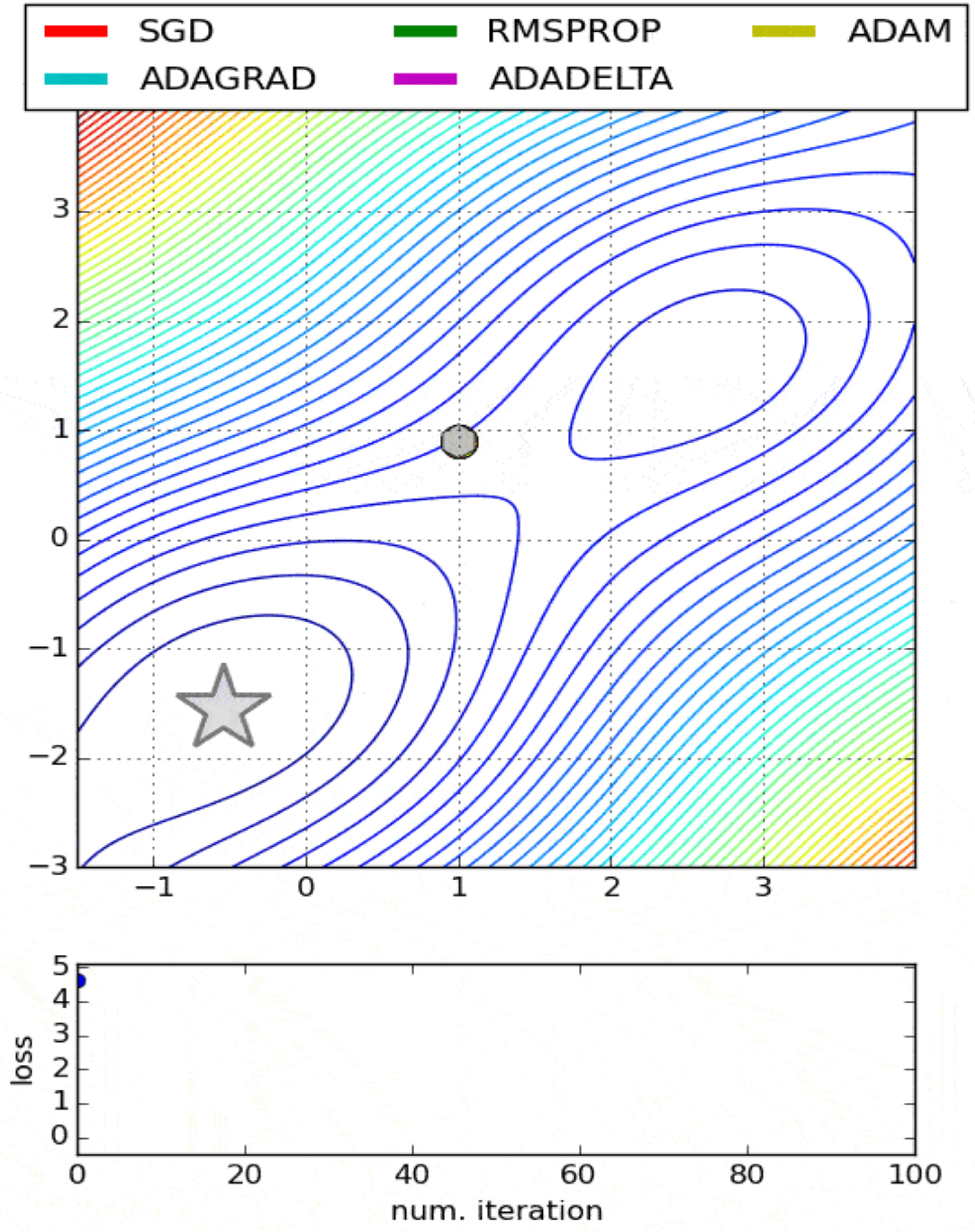
CMS


CMS Detector
Pixels
Tracker
ECAL
HCAL
Solenoid
Steel Yoke
Muons
SILICON TRACKER
Pixels (100 x 150 μm²)
~1m² ~66M channels
Microstrips (80-180μm)
~200m² ~9.6M channels
CRYSTAL ELECTROMAGNETIC CALORIMETER (ECAL)
~76k scintillating PbWO4 crystals
PRESHOWER
Silicon strips
~16m² ~137k channels
STEEL RETURN YOKE
~13000 tonnes
SUPERCONDUCTING SOLENOID
Niobium-titanium coil carrying ~18000 A
HADRON CALORIMETER (HCAL)
Brass + plastic scintillator
~7k channels
FORWARD CALORIMETER
Steel + quartz fibres
~2k channels
MUON CHAMBERS
Barrel: 250 Drift Tube & 480 Resistive Plate Chambers
Endcaps: 473 Cathode Strip & 432 Resistive Plate Chambers
Total weight : 14000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T

Deep Neural Network(DNN)





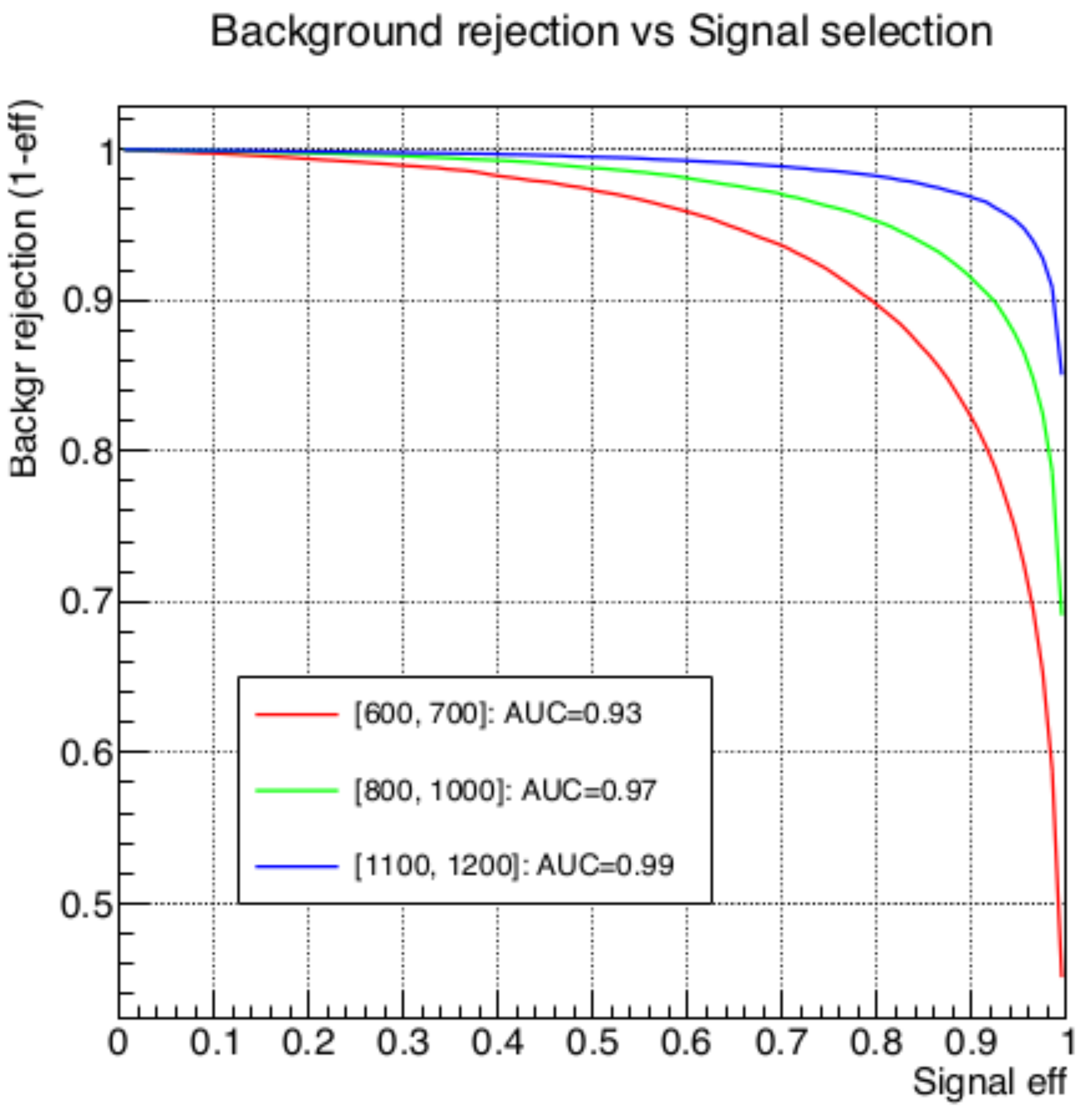


BDT Output

Training Configuration:

Boosted Decision Tree:
Gradient Boosting
Number of Trees: 1000
Tree Depth:- 2
Training Sample:- 50%
Testing Sample:-50%

Model	Training Accuracy (%)	Testing Accuracy (%)
Tprime[600-700GeV]	82.29	81.81
Tprime[800-1000GeV]	89.62	89.25
Tprime[1100-1200GeV]	94.76	94.37
Tprime[600-1200GeV]	87.32	87.14



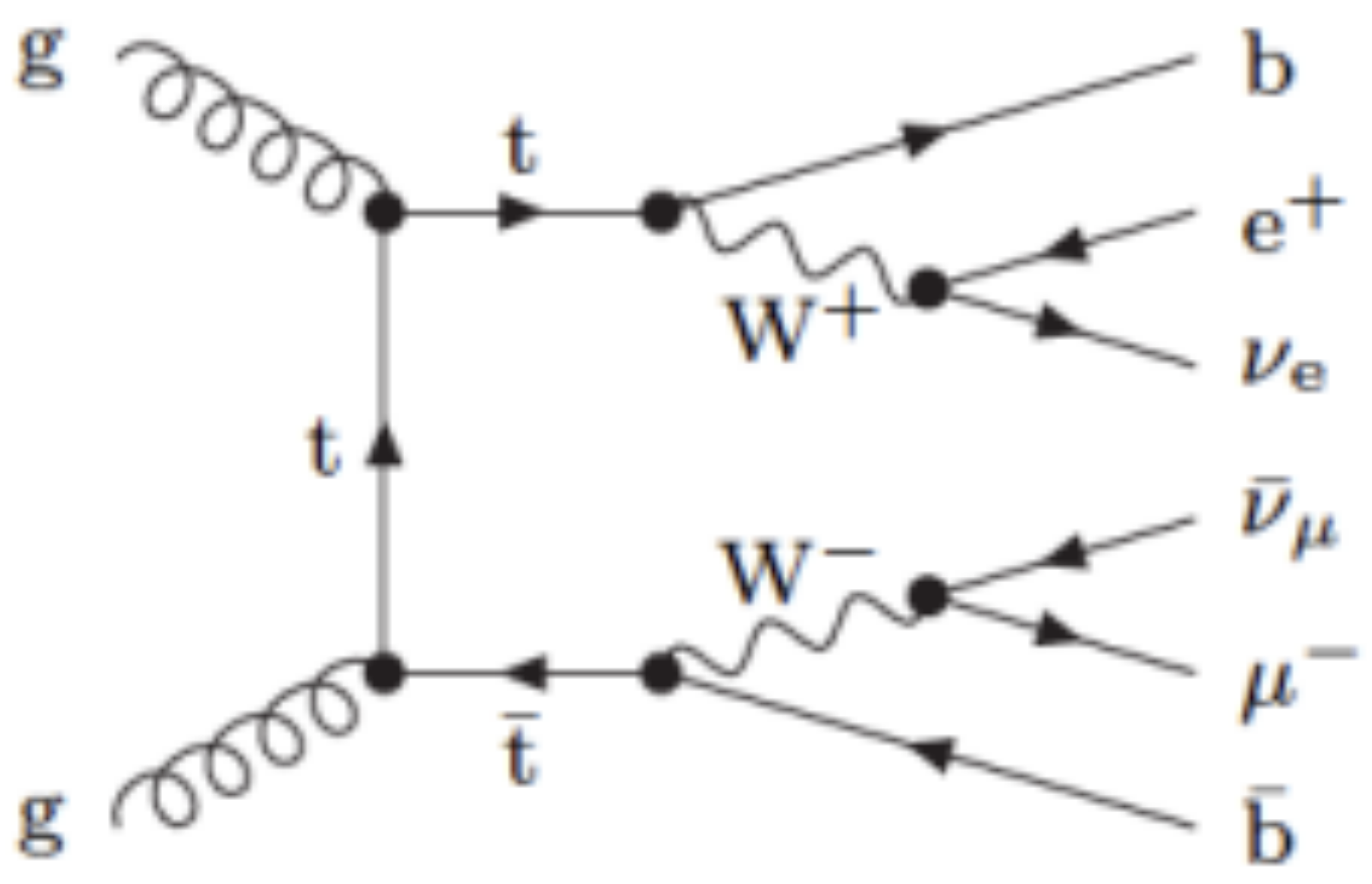
Vector-Like Quark(VLQ)

- Search for decays to 3rd gen quark and SM boson
- Also potential for decays to quark + DM

How do we search for VLQs?

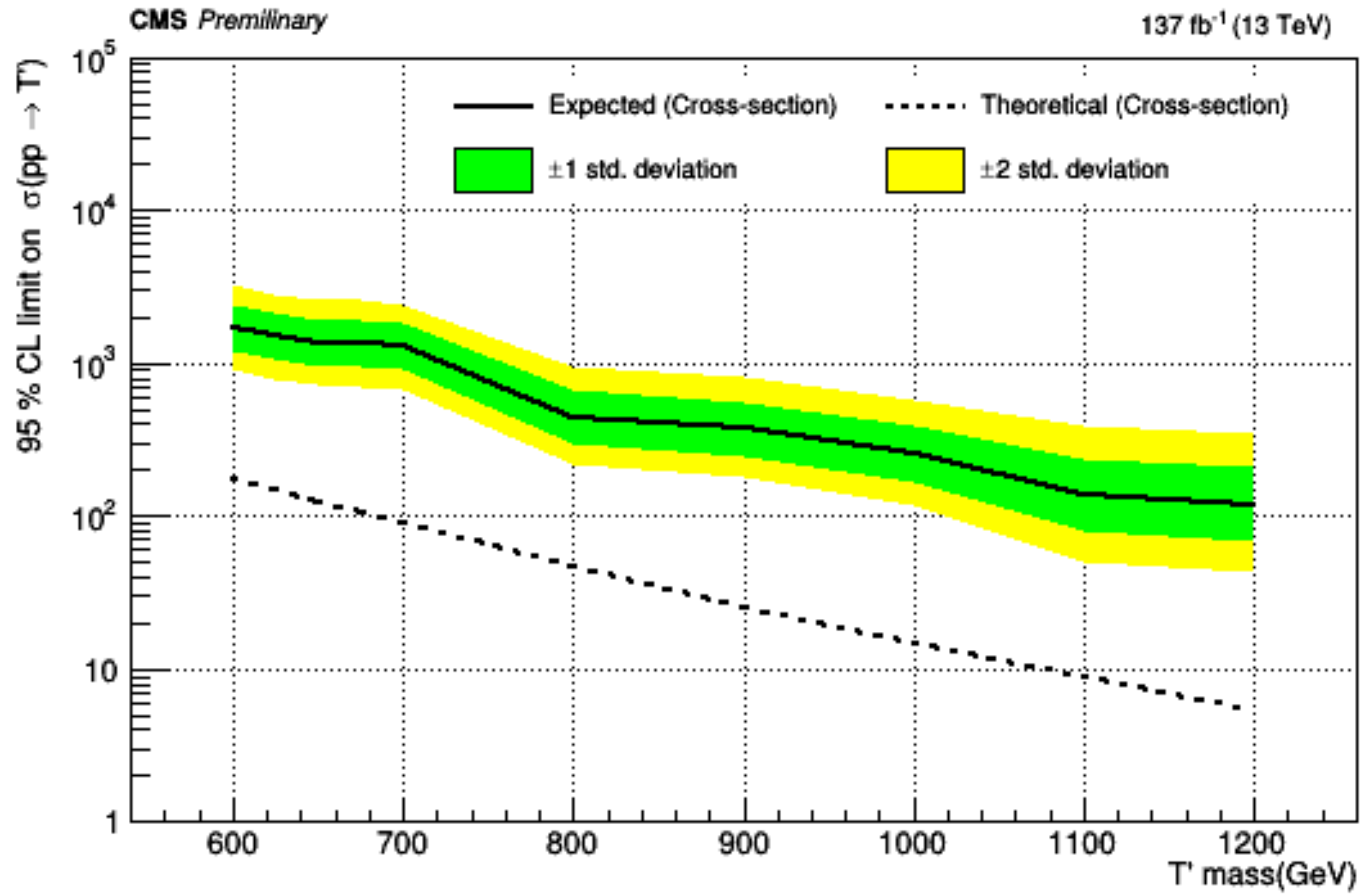
Exploit high particle masses! Many boosted decays simplify event topologies

Identify signal: reconstruct mass or rely on high pT (signal >> Standard Model)

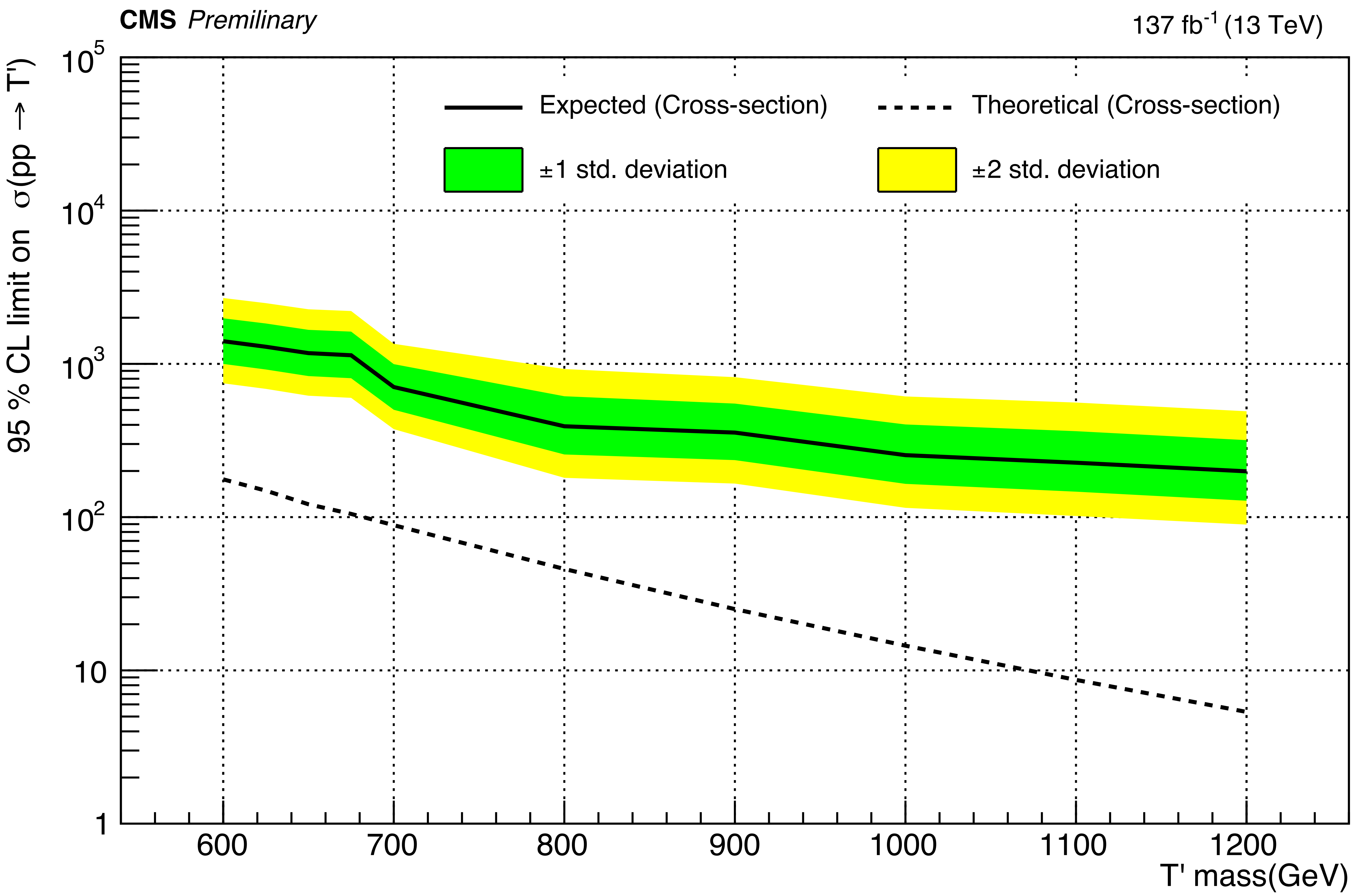


ttgg

Exclusion Limit Plot



Limit on cross section with DNN Output



Limit on cross section with BDT Output

From above plot, we conclude that our expected limit is ~10 times higher than theoretical and our search will only be sensitive if we have around ~10 times more data.