1 Introduction

1.1 Motivation

Goal of this project is the improvement in the distinction process between $t\bar{t}\gamma$ and $t\bar{t}$ events in the signal region. Depending on the success, the results can be further used for other projects, which would improve the simulations and their predictions regarding the standard model (SM) and beyond standard model (BSM).

1.2 Approach

An attempt was perfomed through a multivariate analysis and by developing and training models. At the beginning different variables were analysed and compared and the most promising were used for training. Type of models, denoted MVA in the following, used are boosted decision tree (BDT) and mutilayer perceptron (MLP). The results of the MVAs for three different variable categories were created and in the course of this project analysed and presented.

2 Background information

2.1 ATLAS detector and data

The ATLAS detector is part of the Large Hadron Collider, a particle accelerator, which was built by the research organisation CERN in Genf, Siwtzerland. LHC has a circumference of 27km and is up to 175m deep. Besides ATLAS, there are other experiments as well, such as CMS, ALICE, LHCb. ATLAS can be described by three sub-detectors. All detectors surround the beam pipe, where the particles are accelerated and collided against eachother.

In Figure 1, a sketch of the detectorsystem can be seen. The inner detecor is a tracking chamber for charged partices. It is surrounded by a superconducting solenoid with a magnetic field of 2T used to divert the particle trajectory to measure the momentum. The calorimeter is made up of alternating layers. One is a dense absorber material, which is used to prompt electromagnetic and hadronic showers (ECAL, HCAL). The other layer is of active material used for energy measurement. The muon spectrometer (MS) is positioned in the outermost layer since muons may pass inner layers as well as the calorimeters. It is surrounded by a magnetic field and the diversion is measured by a multiple layer of high precision tracking chambers.
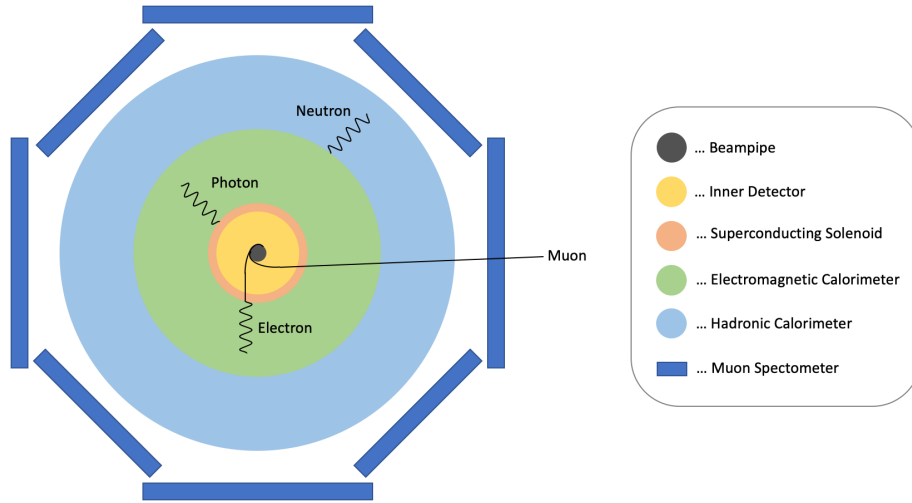
Figure 1: detector system draft

The data are filtered by a two-step triggersystem. The hardware-based system (L1) is triggered when muons, electrons, photons and jets are detected. This reduced the data amount from 40MHz to 100kHz. The software-based system (L2) is triggered by the different energies, type and numbers of particles:

- lepton has 30/35 GeV

- at least 3 jets were detected

- at least one of the jets have a b-tag

- exactly one photon

This reduced the data amount from 100kHz auf 1kHz. The data used during this project was recorded in 2016 at a centre-of mass energy of 13TeV. The signal and background processes were modelled using Monte Carlo generators and passed through a detector simulation (DOUBLE CHECK).

2.2 Proton-Proton-collision and $t\bar{t}\gamma$

The protons are accelerated in bunches, each containing 10^{11} particles, through different beam pipes in opposite directions and then collided with each other. The $t\bar{t}\gamma$ events were analysed, meaning top quark pair events with additional photons.

The top quark is the heaviest quark in the known elementary particles and is the only quark so far whose properties can be analysed, since the lifetime is very short and it

forms no bound state before it decays. By that the electromagnetic coupling, therefore the number of radiated photons after interaction with other charged particles, can be measured. This is a test of the standard model, since it is a direct measurment of the charge of the top quark and shows possible anomalous structure of its electromagnetic interaction.

During the pp-process a top quark pair is produced. The top quark then decays in a W-boson and a b-quark. The W-boson disintegrates into a charged lepton and a corresponding neutrino or a pair of up-type quark and down-type antiquark. The remaining quarks decomposes into jets. This process can be seen in Figure 2.
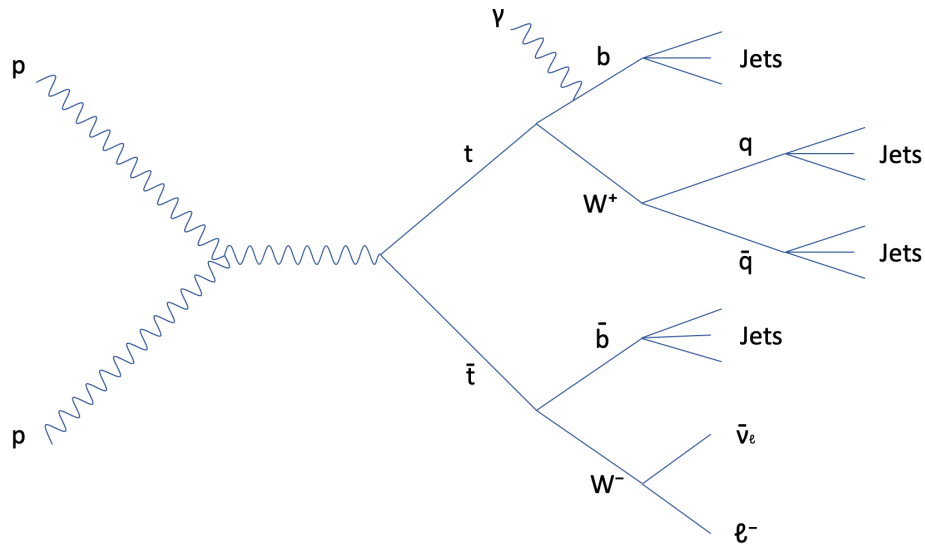
Figure 2: PP process draft

3 Challenge in categorisation

In Figure 3, the transverse momentum of the photon is visible. The background makes up nearly 50% of the signal region. These are made up by particles, which were misconstructed or not detected due to the limited detector's acceptence. Other reasons are photons not being radiated from the $t\bar{t}$ process, but rather from an initial charged patron, an intermediate top quark or any charged final state particles or decay products. Further sources could be hadrons misreconstructed as electrons or hadrons and electrons misidentified as photons.
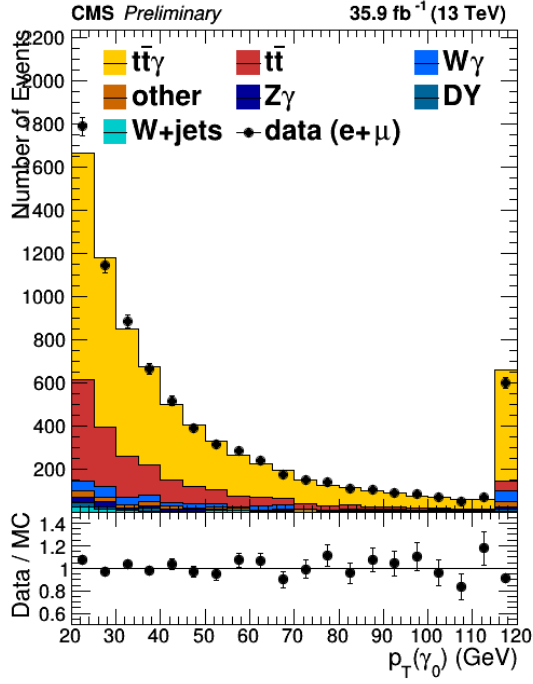
Figure 3: Distribution of Pt Photon
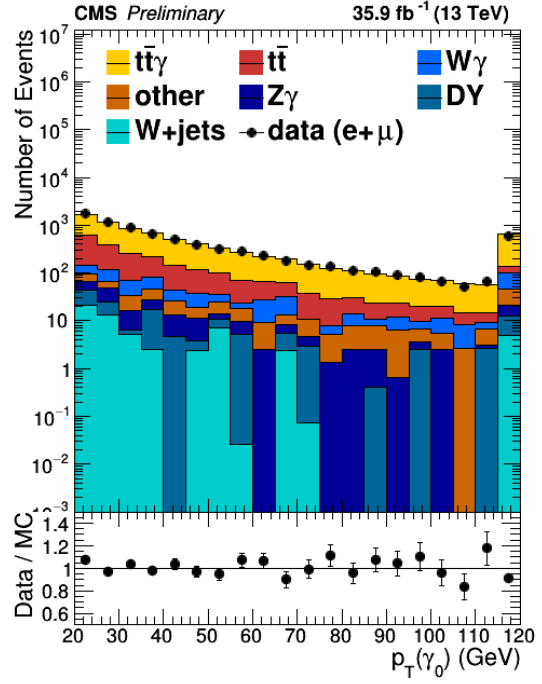
Figure 4: Distribution of Pt Photon (log)

3.1 Current status

First a neural network was developed to seperate the hadronic-fake photons, which are hadrons misidentified as photons and made up the majority of the background signal. The result was then used as an input for the neural network created to split the $t\bar{t}\gamma$ signal and the background further. What still needs to be further investigated is the possibility of improvement in the discrimination between $t\bar{t}\gamma$ and $t\bar{t}$ events in the signal region through MVAs.

4 Analysis process

4.1 List of variables

The analysed variables are listed below and they are categorized by the type of particles.

- B-jet: variables of jets containing b-hadrons (b-tag)

    - Bj0_eta
    - Bj0_phi
    - Bj0_pt
    - Bj1_eta

    - Bj1_phi

    - Bj1_pt

    - nBTagGood

4

- Jet: variables of all jets

  - JetGood0_eta
  - JetGood0_phi
  - JetGood0_pt
  - JetGood1_eta
  - JetGood1_phi
  - JetGood1_pt
  - nJet
  - nJetGood
  - JetGood0_neEmEF
  - JetGood0_chEmEF
  - JetGood0_neHEF
  - JetGood0_chHEF

- Lepton: variables of electron, muon

  - LeptonTight0_eta
  - LeptonTight0_phi
  - LeptonTight0_pt
  - nElectronTight
  - nLeptonTight
  - nMuonTight
  - LeptonTight0_pfRelIso03_all
  - LeptonTight0_pfRelIso03_chg
  - LeptonTight0_pfRelIso03_ne

- Combination: variables of two particle types

  - leptonJetdR
  - tightLeptonJetdR
  - ltight0GammadPhi
  - ltight0GammadR
  - PhotonLepdR
  - PhotonJetdR
  - JetGood0_btagDeepB
  - JetGood1_btagDeepB

- Photon:

  - nHighPTPhotons
  - nPhotonGood
  - PhotonGood0_eta
  - PhotonGood0_mvaID
  - PhotonGood0_phi
  - PhotonGood0_pt
  - PhotonGood0_sieie
  - PhotonGood0_hoe
  - PhotonGood0_r9
  - PhotonGood0_pfRelIso03_all
  - PhotonGood0_pfRelIso03_chg
  - PhotonGood0_pfRelIso03_ne

- Others:

  - MET_phi
  - MET_pt
  - ht
  - m3
  - mT

The pre- or suffix of the name reveal, what information is saved. The description of variables, which don't follow this logic, are listed at the end.

Prefix and suffix:

- _eta: pseudorapidity ($\eta = -ln(tan(\frac{\vartheta}{2}))$; wheras ϑ is the polar angle)

- _phi: azithmuthal opening angle, around z-axis ($\Delta\phi$)

- _pt: transverse momentum of variable

- dR: distance between two objects (ΔR)

- n: number of particles

Other:

- ht

- m3

- mT

- PhotonGood0: number of photons

- MVA_ID: currently used as a parameter to differentiate between good and bad photon (??)

- MET: Missing transverse momentum

4.2 Analysis of variables

The variables were analysed by comparing the distribution of events with an uniform scaling. A uniform scaling was needed to have a better view of relative differences between signal and background. If the variable shows a different behaviour for each type of signal, this could be an indication of discriminatory power.

Figure 5 to 8 show some examples of variables, where the distributions are almost identical and were therefore not further analysed. A full list of these variables are listed in the Annex, chapter 9.1.
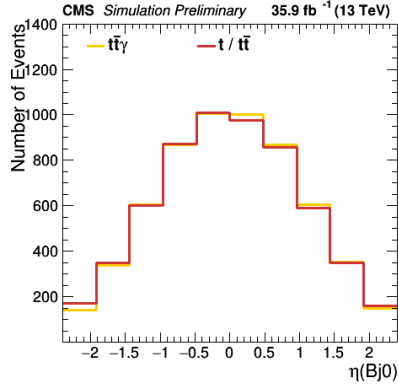
Figure 5: Bj0_eta
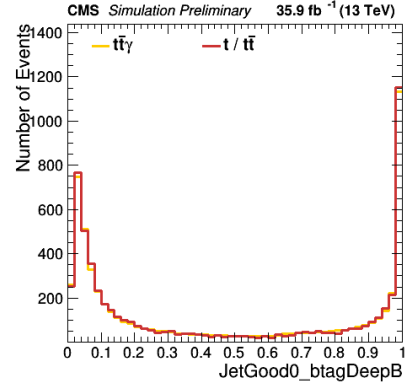
Figure 6: JetGood0_btagDeepB
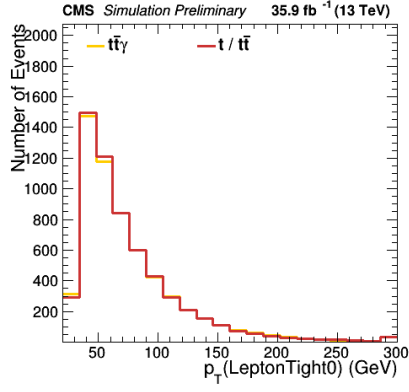
Figure 7: LeptonTight0_pt
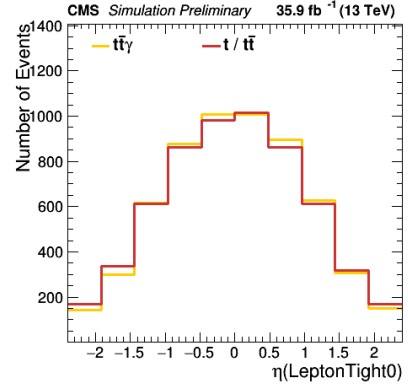
Figure 8: LeptonTight0_eta

In Figure 9 to 19 there are some differences visible in the signal region. These were candidates for the first set of selected variables.
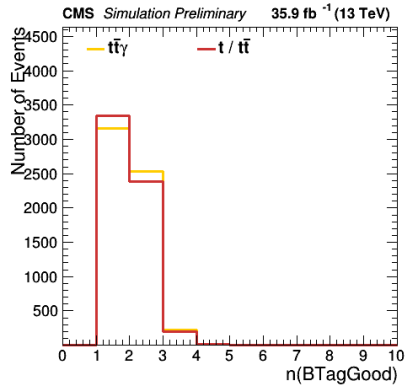
Figure 9: nBTagGood
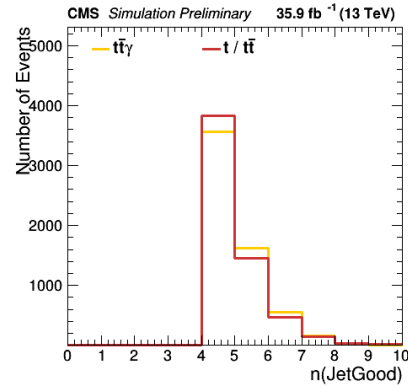
Figure 10: nJetGood
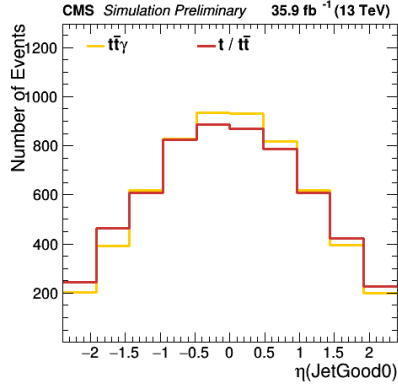
Figure 11: JetGood0_eta
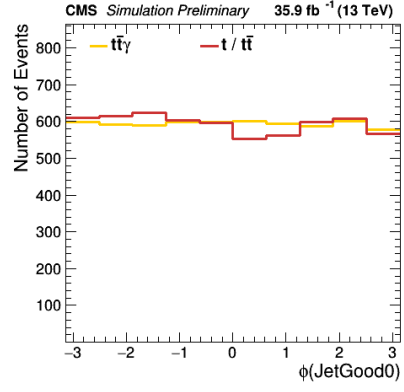
Figure 12: JetGood0_phi
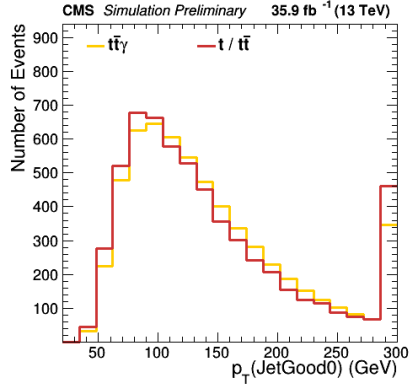
Figure 13: JetGood0_pt
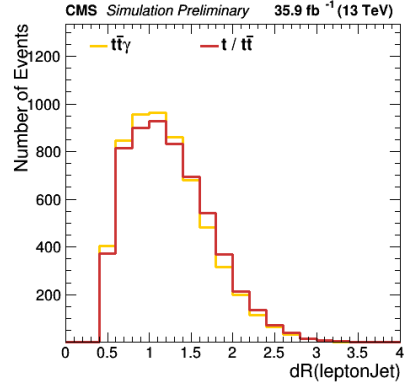
Figure 14: leptonJetdR
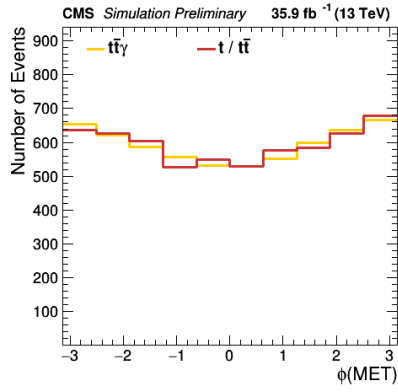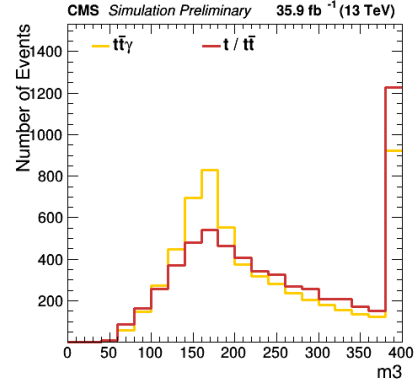

Figure 15: MET_phi

Figure 16: MET_pt

8
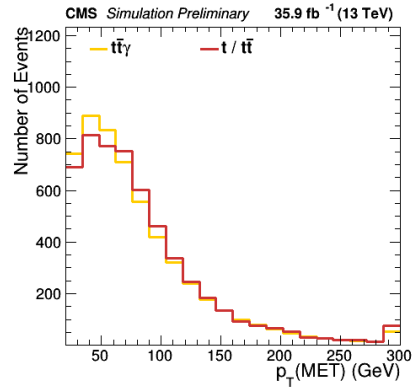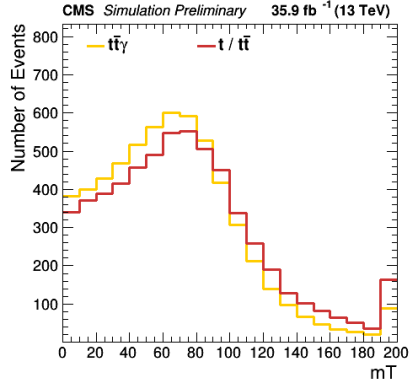
Figure 17: mT

Figure 18: m3
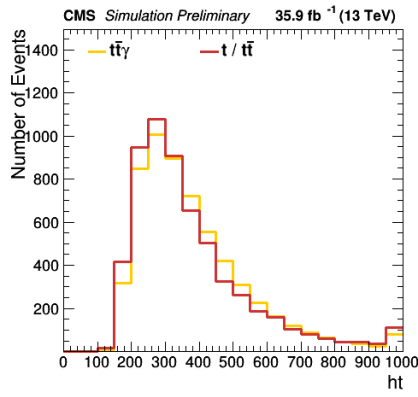
Figure 19: ht

The variable MVA_ID, seen in figure 20, shows quite a good distinction and is therefore used in the second selection.
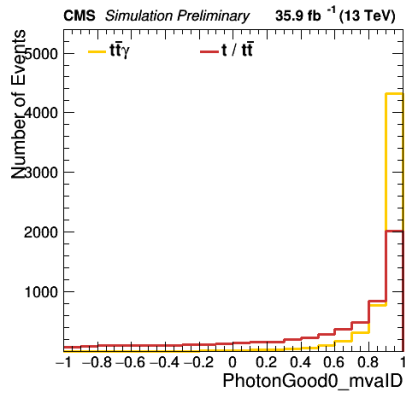
Figure 20: PhotonGood0_mvaID

Figure 21 to 31 show the photon variables. These show the most promising behaviour for discrimination between $t\bar{t}\gamma$ and $t\bar{t}$ events and were put into the third selection category.
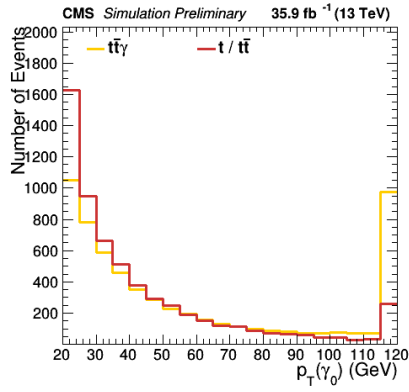
Figure 21: PhotonGood0_pt
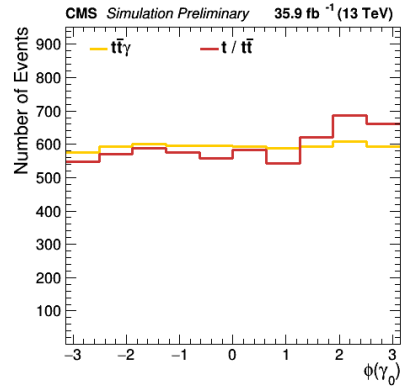
Figure 22: PhotonGood0_phi
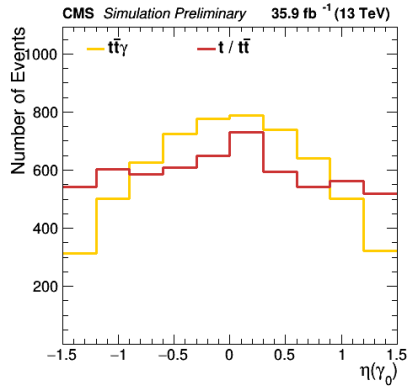
Figure 23: PhotonGood0_eta
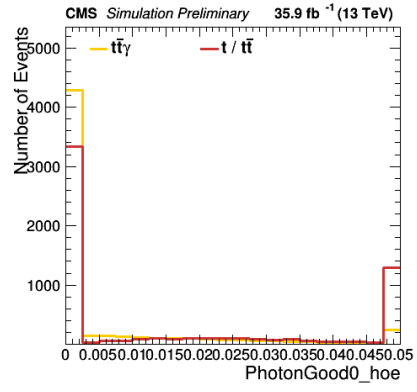
Figure 24: PhotonGood0_hoe
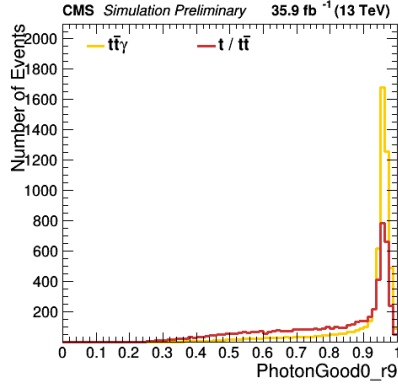
Figure 25: PhotonGood0_r9
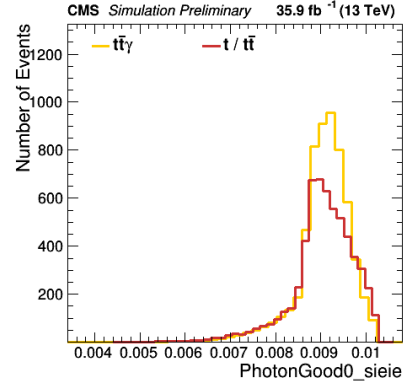
Figure 26: PhotonGood0_sieie
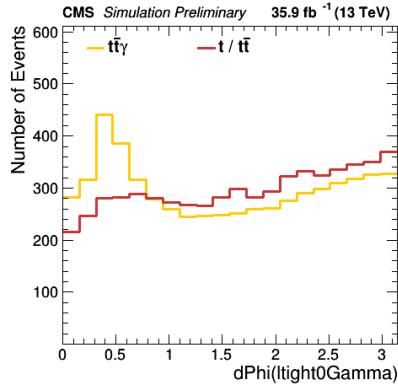
Figure 27: ltight0GammadPhi
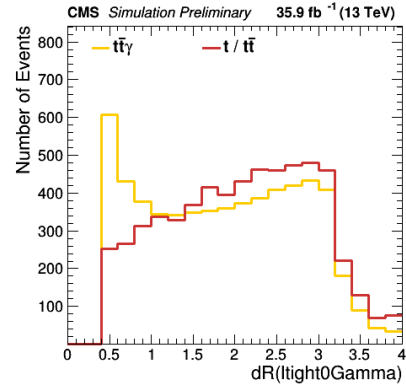
Figure 28: ltight0GammadR
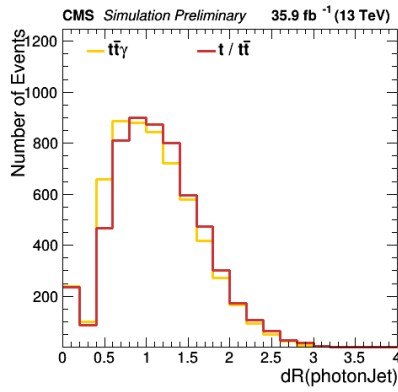
Figure 29: photonJetdR
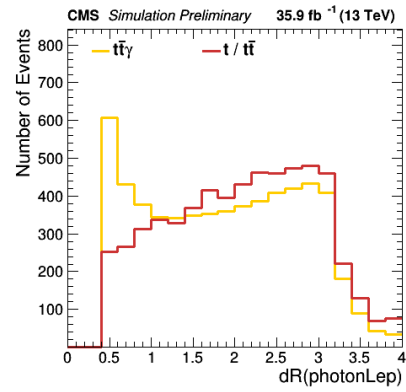
Figure 30: photonLepdR

11
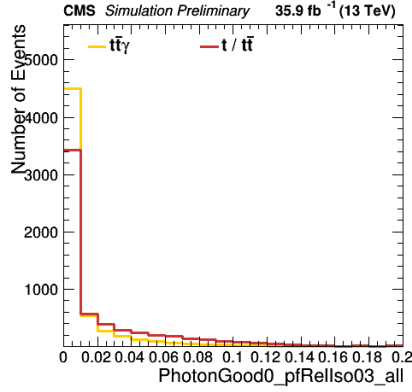
Figure 31: PhotonGood0_pfRelIso03_all

Variables, which were excluded because another related variable was used instead, are listed in the Annex, chapter 9.2.

5 MVA

Trying to split signal and background through a variable with a cut-off value faces the problem, that this criteria doesn't apply for every event. If it works for one event, it doesn't mean that it works for the next one. Therefore mutliple variables need to be used to distinguish between those two categories, so called multivariate analysis (MVA). As already mentioned in chapter 1.2, the chosen MVA-types are boosted decision trees (BDT) and multilayer perceptron (MLP).

5.1 BDT - Boosted Decision Trees

The decision tree starts with the root node containing the whole sample For a binary tree, the node is then split into two branches using a variable and a corresponding cut-off value. The cut-off value should be a value, which seperates the signal from the background the best. This process will be repeated for each branch for all relevant variables, including already used variables, until an end criteria is met. The final branch is called a leaf and this will be assigned to one of the two categories. End criterias could be a minimum size of a leaf, perfect separation, insignificant improvement after split or maximal tree depth. Boosted decision trees are built out of additional trees, so called weak classifieres. They include variables with a low discriminatory power and will be used to improve the main decision tree to a more stable model with a lower error rate.
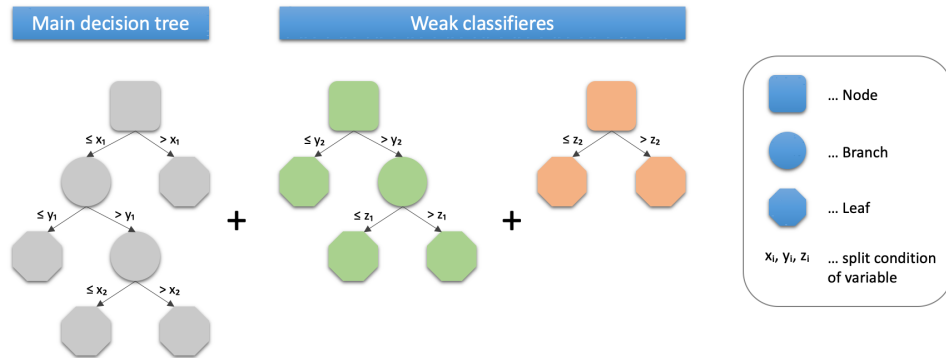
Figure 32: PP process draft

5.2 MLP - Multilayer Perceptron

The model is built out of multiple perceptrons, which are arranged in layers shown in Figure 34. The value of one perceptron is the sum of all weighted inputs. This value is then transformed through an activation function and a treshhold. The activation funtion is a linear function for the input- and outputlayer, while for the hiddenlayers it is usually the log-sigmoid transfer function, seen in figure 33. All layers have a treshhold, except the inputlayer.
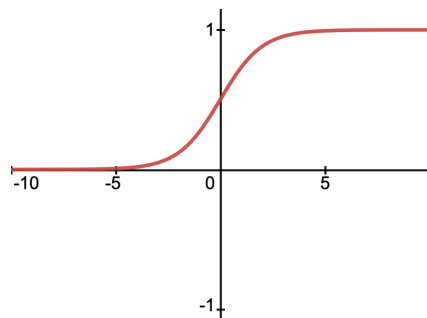
$$f(x) = \frac{1}{1 + e^{-x}}$$

Figure 33: log-sigmoid function

In the MLP, every perceptron of a layer is connected to every perceptron of the previous and next layer. Each connection has an assigned weight, so called weight coefficient. All coefficients of one layer can be written as a weight matrix. The first and layer are called

input- and outputlayer respectively, while the intermediate layers are the hiddenlayers. A MLP has a minimum of one hiddenlayer, therefore contains at least three layers.

If the signal is only transmitted in one direction, therefore the MLP doesn't have loops, it is described as feed forward. The learning process of the model is called backpropagation algorithm. It learns by minimizing the error between network output and expexted output. Initially all weights are assigned random values, then every event will go through the network and the weights are adapted. Either all events will be put through the network first and then the values are adapted or are adjusted after each event, but for the latter the order of the events might be important. After one epoch the end criteria will be tested. If it fails, the whole learning process will be carried out again. If it is met, the algorithm is finished.
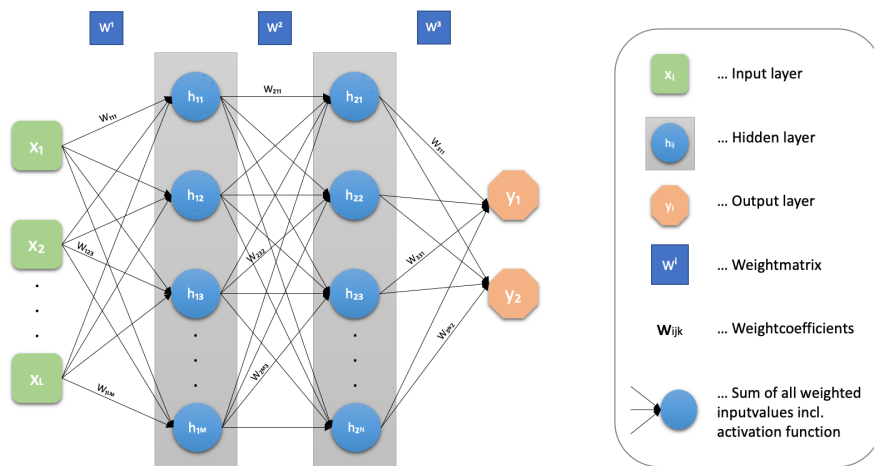
Figure 34: MLP draft

5.3 Training of MVA

The following table shows some of the default values used for the models.

- BDT
    - Minimum node size
    - Separation type
- MLP
    - Number of cycles

14

– Learning Rate

  – Decayrate

  – Neurontype

  – Estimatortype

  – Neuroninputtype

  – Trainingmethod

A full list of all parameters, configuration options as well as their default values are given in the annex, chapter xx.

These parameters were adapted to improve the result:

- BDT

  – Maximum depth

  – Number of trees

  – Number of Cuts

- MLP

  – Number of layers

The default values was used as basis. Additionally, the MVAs were trained with a lower and a higher configuration:

	BDT			MLP
Configuration	maximum depth	number of trees	number of cuts	number of layers
1	1	400	10	5
2	3	800	25	7
3	5	1000	50	10

The distribution and ROC-curves for the BDT- and MLP-models are shown for each selection and configuration in the following chapter. To avoid overfitting, each MVA is trained on one part of the sample (training sample) and tested on the rest (testing sample). If the results differ significantly, then this would be a sign of overfitting - the model only differentiates correctly on this specific sample. The results of the train- and test-sample can be seen in the distribution plots.

The ROC-curves show the background rejection versus signal efficiency. While discriminating between signal and background, the signal efficiency is the proportion of the correctly categorized signal events ($t\bar{t}\gamma$) and the background rejection is therefore the correctly assigned fraction of background events ($t\bar{t}$). There are two types of errors during the distinction process:

- a signal event is categorized as background, which is called false negative

- a background event is assigned to the signal category, calles false positive

Examples for different types of ROC-curves are displayed in figure 35. The green curve indicate a good discriminatory power. The orange curve show that the model assigns the events the wrong way around. If the ROC-curve follows closely the identity line, then the model doesn't have discriminatory power.
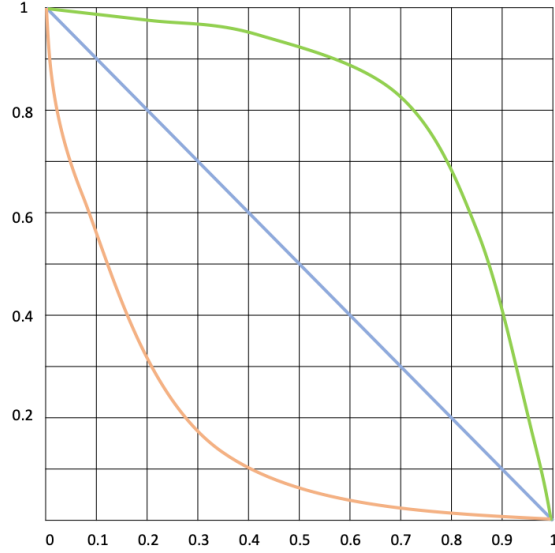
Figure 35: ROC-curve examples

5.4 Results

The first selection category were put into the models and are seen in figure 36 - 38. A slight improvement in the distinction between $t\bar{t}\gamma$- and $t\bar{t}$-events is visible.

In the second selection category the variable PhotonGood0mvaID was added (figure 39 - 41). The result improved significantly.

Now adding the photon variables for the third selection show a higher success in the distinguishing process. They can be seen in figure 42 - 44.

Between the three configuration combinations there is no significant difference. For all selection categories, the MLP model perfomed better than the BDT model. The third set of variables show the best result.
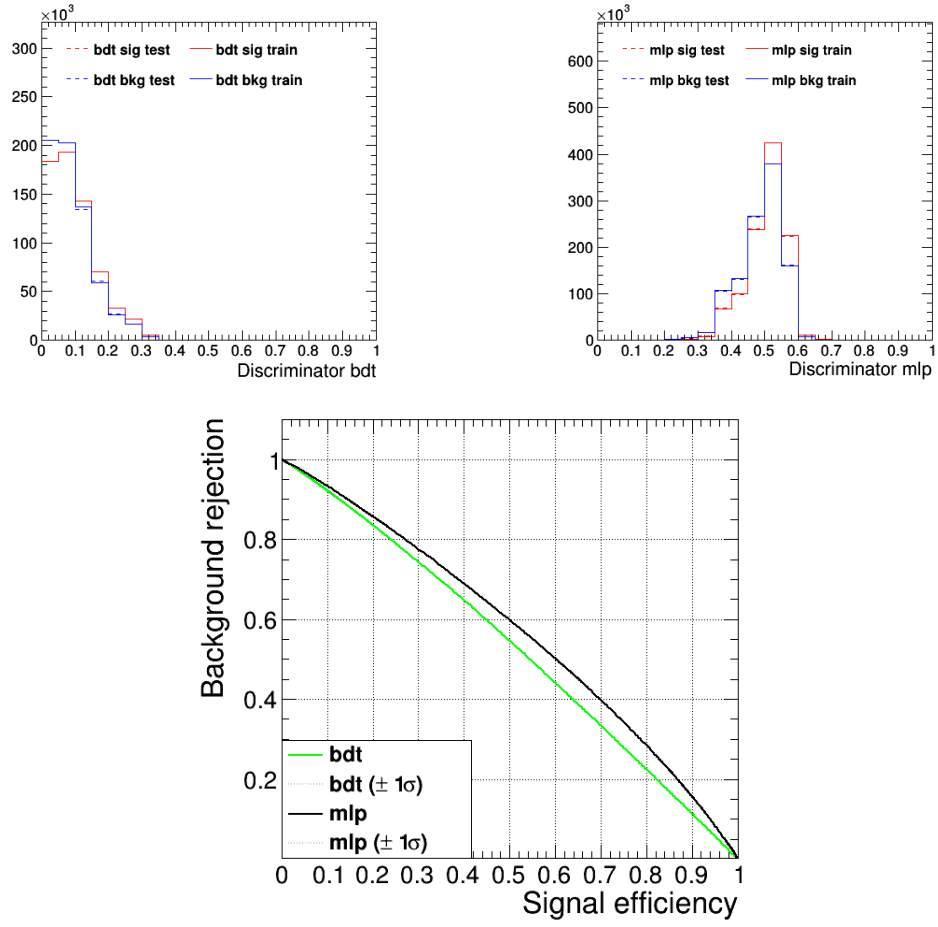
Figure 36: BDT-, MLP-distribution and ROC curve for variables from selection 1 with configuration 1
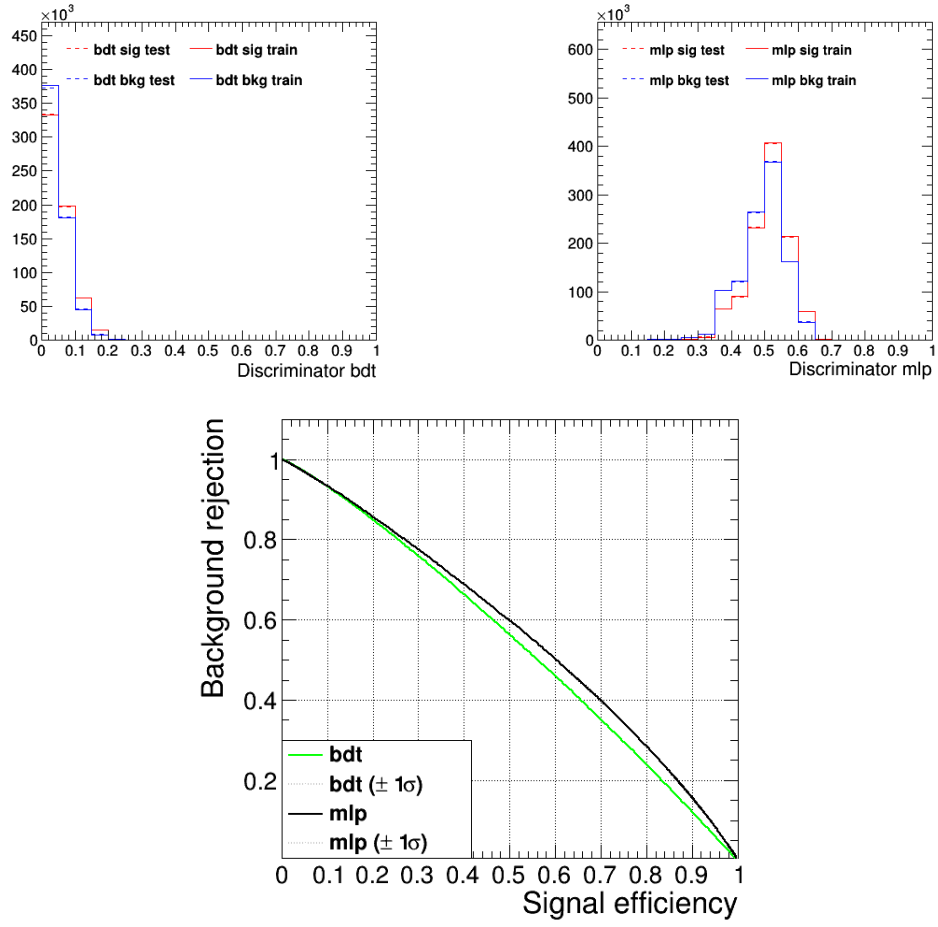
Figure 37: BDT-, MLP-distribution and ROC curve for variables from selection 1 with configuration 2
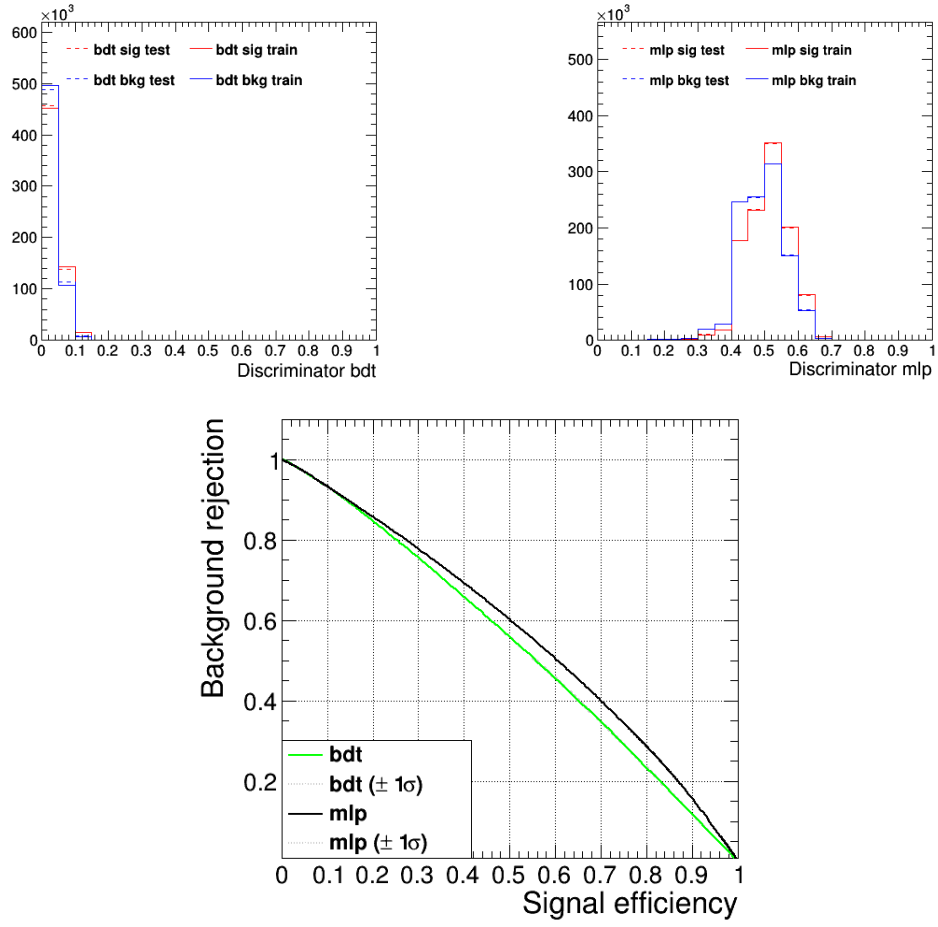
Figure 38: BDT-, MLP-distribution and ROC curve for variables from selection 1 with configuration 3
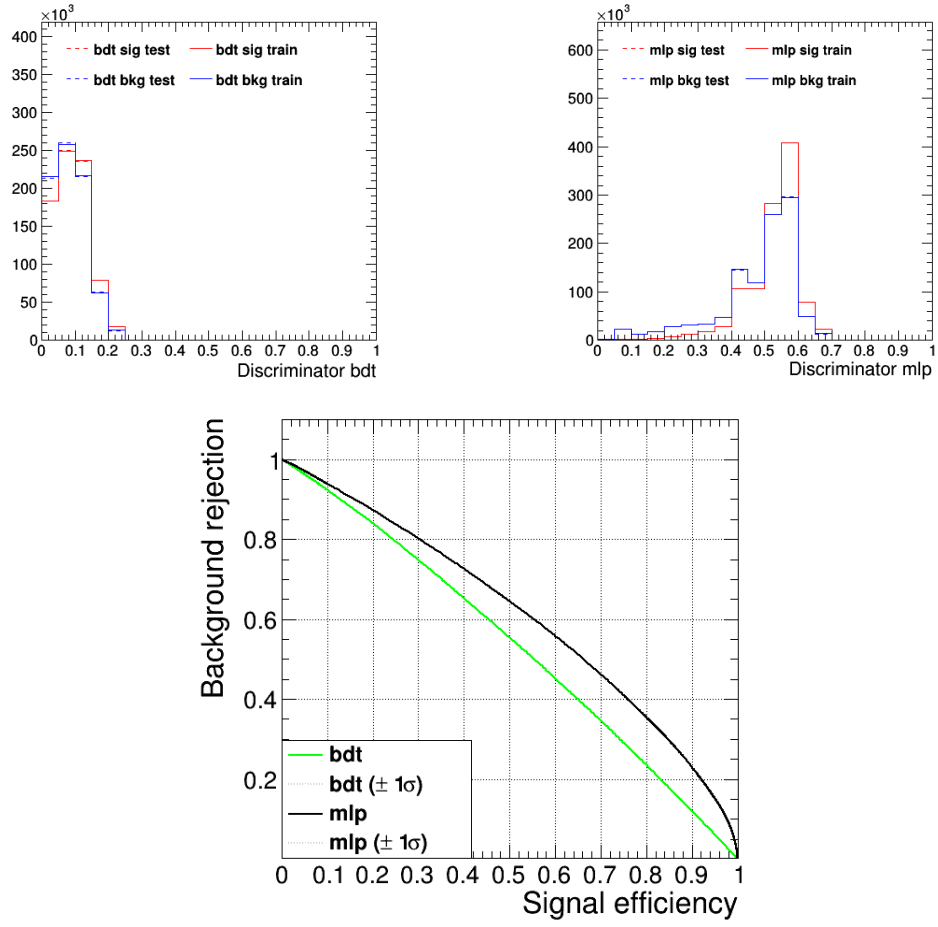
Figure 39: BDT-, MLP-distribution and ROC curve for variables from selection 2 with configuration 1
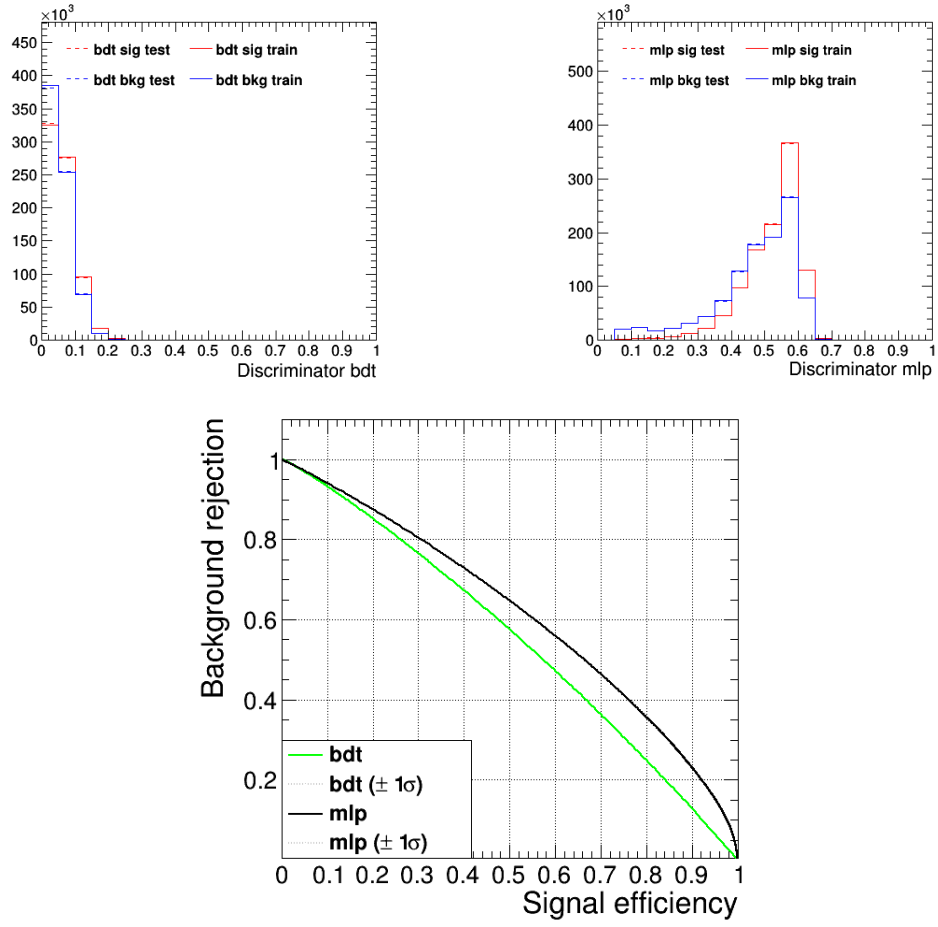
Figure 40: BDT-, MLP-distribution and ROC curve for variables from selection 2 with configuration 2
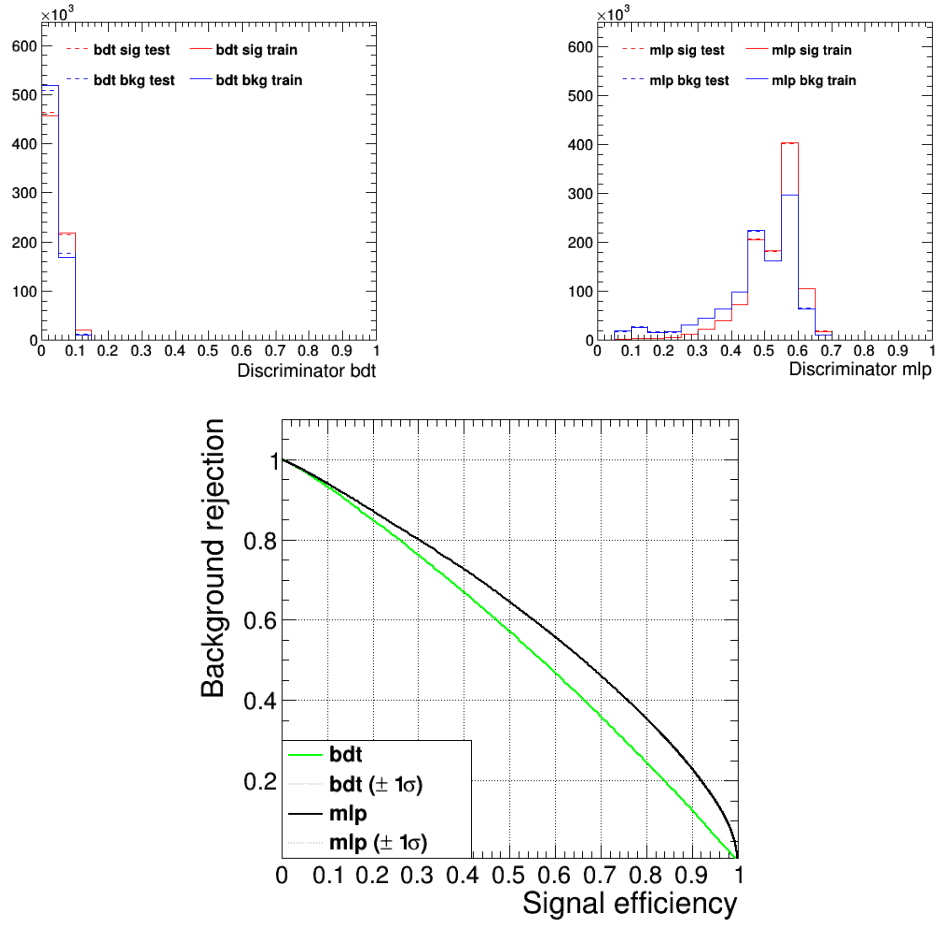
Figure 41: BDT-, MLP-distribution and ROC curve for variables from selection 2 with configuration 3
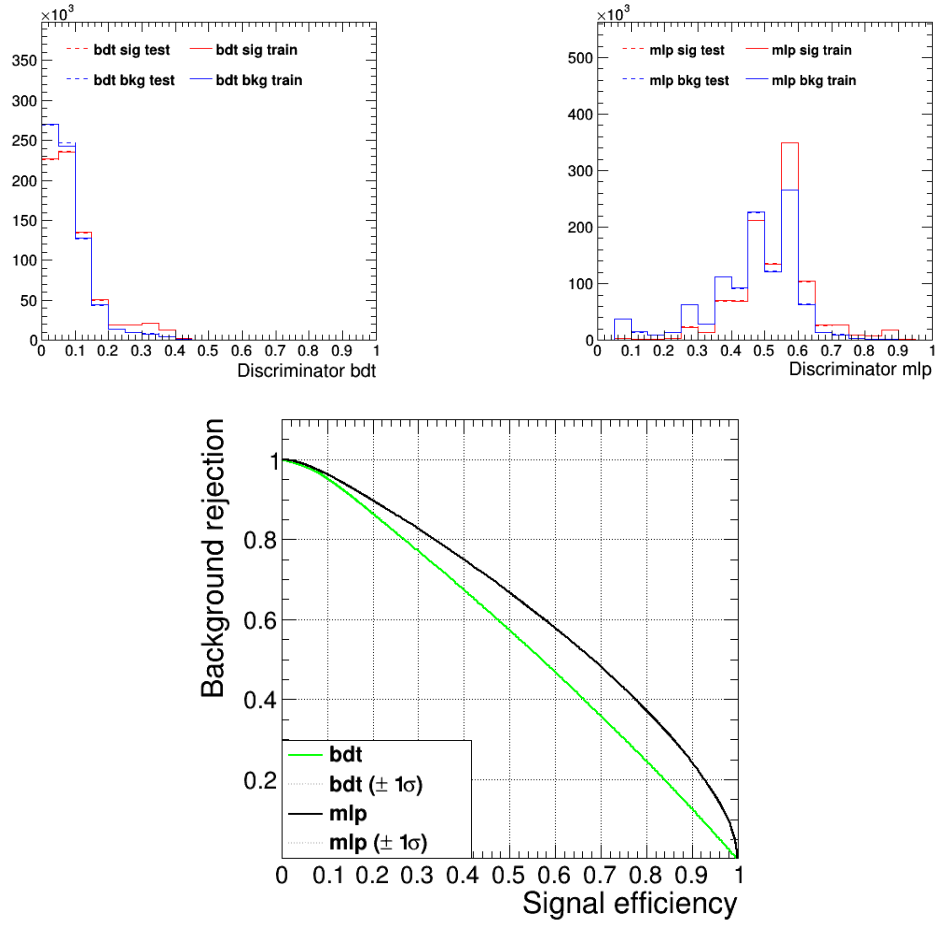
Figure 42: BDT-, MLP-distribution and ROC curve for variables from selection 3 with configuration 1
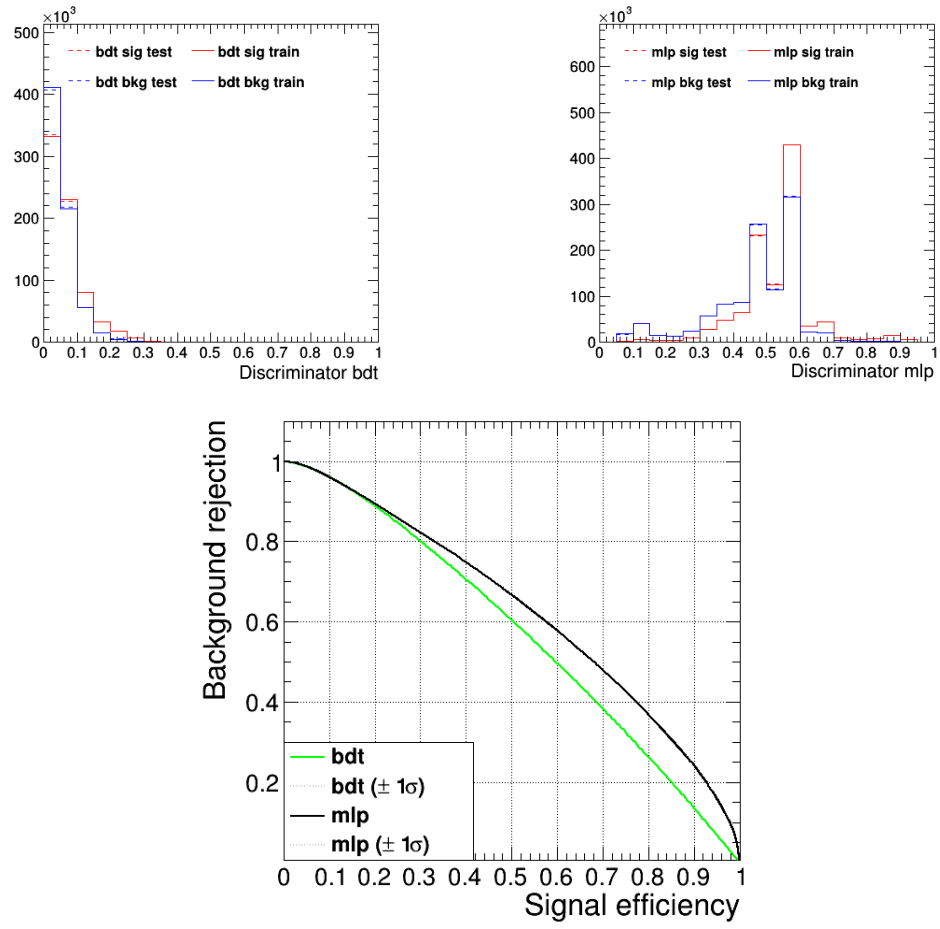
Figure 43: BDT-, MLP-distribution and ROC curve for variables from selection 3 with configuration 2
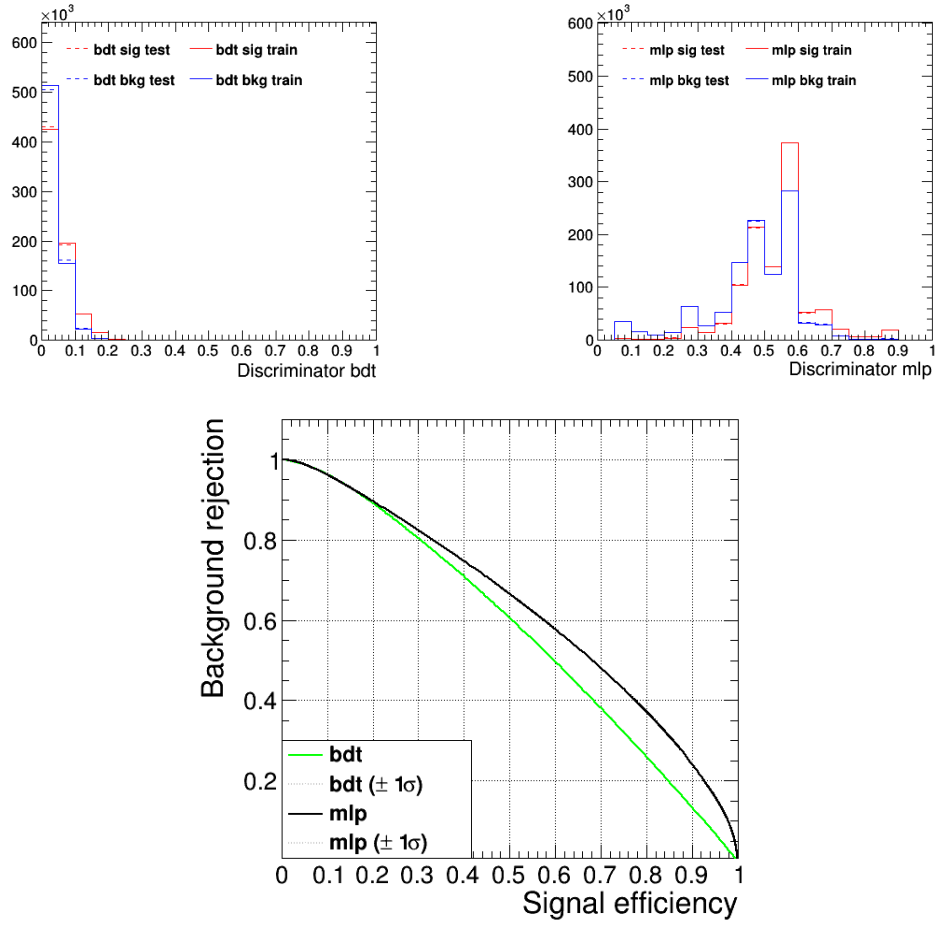
Figure 44: BDT-, MLP-distribution and ROC curve for variables from selection 3 with configuration 3

6 Conclusion

It is very difficult to differentiate between $t\bar{t}\gamma$ and $t\bar{t}$ events because they show very similiar behaviour in the signal region. Nevertheless, a slight improvement was possible by training a MVA with relevant variables. Especially photon variables provide a bigger improvement. MLPs show a better performance than BDTs, which suggests the selection of this type of model.

7 Bibliography

8 List of figures

9 Appendix

9.1 Full list of variables with almost identical distributions
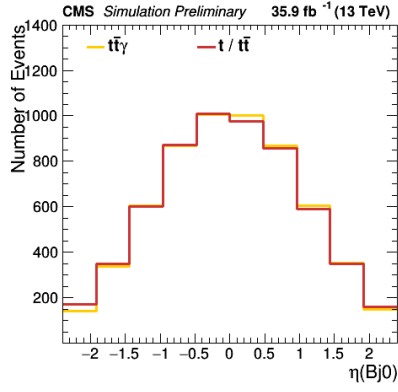
Figure 45: Bj0_eta
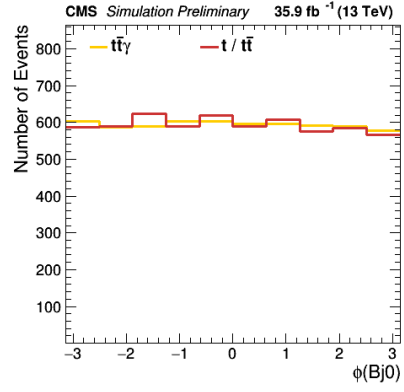
Figure 46: Bj0_phi
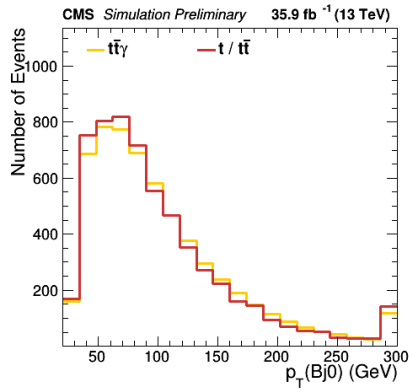
Figure 47: Bj0_pt
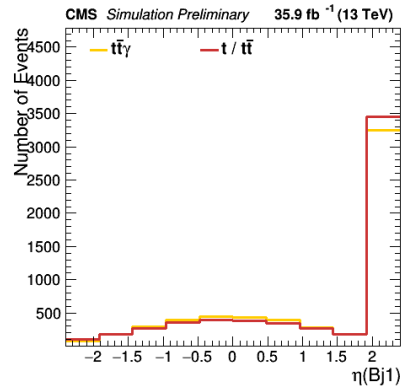
Figure 48: Bj1_eta
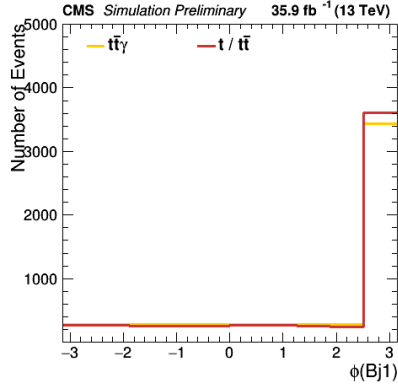
Figure 49: Bj1_phi
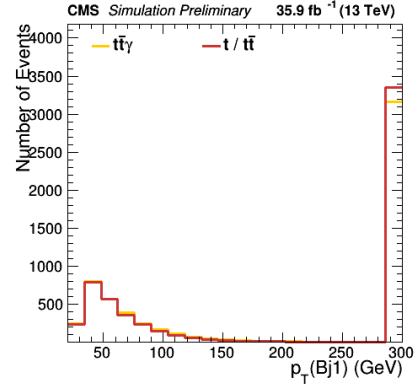
Figure 50: Bj1_pt
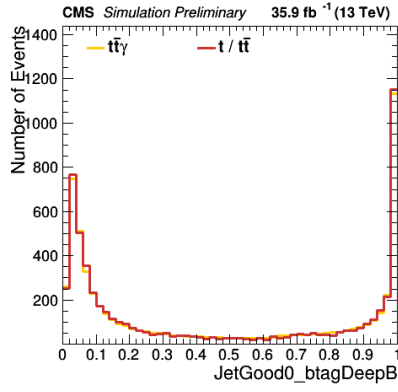
Figure 51: JetGood0_btagDeepB
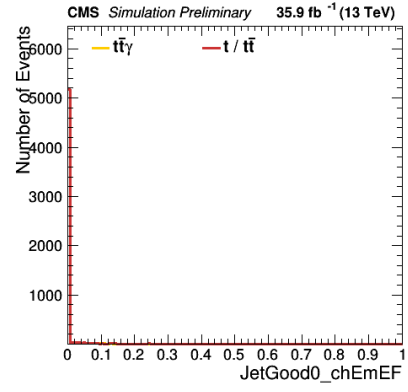
Figure 52: JetGood0_chEmEF
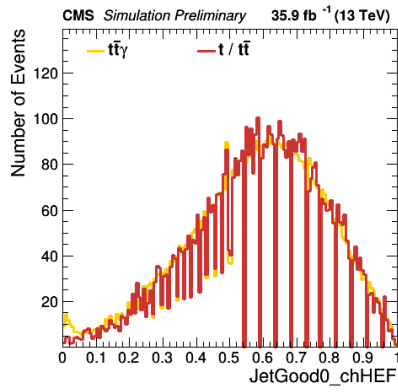
Figure 53: JetGood0_chHEF
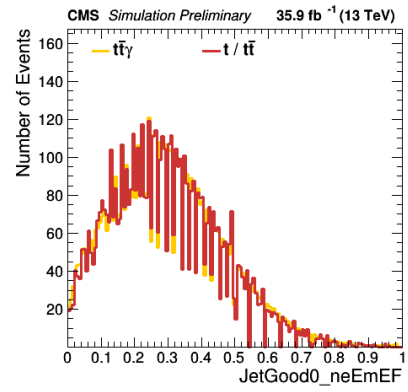
Figure 54: JetGood0_neEmEF
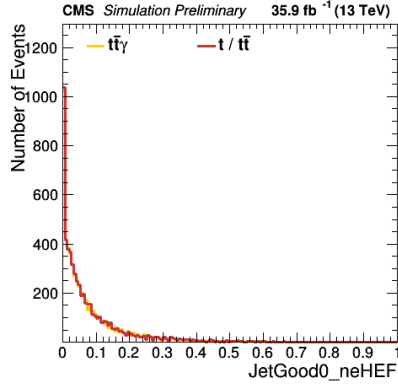
Figure 55: JetGood0_neHEF
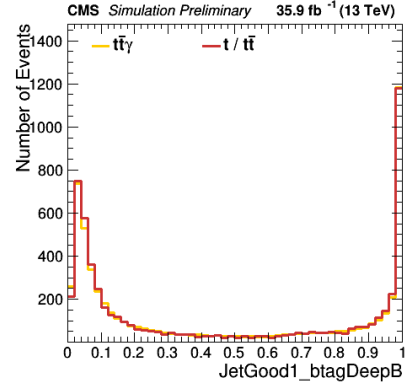
Figure 56: JetGood1_btagDeepB
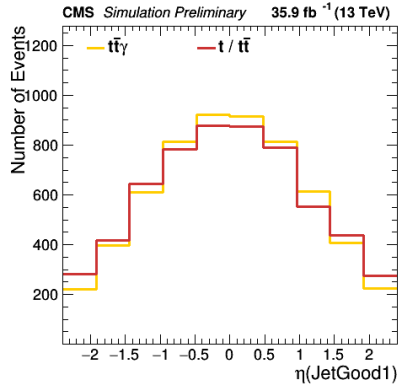
Figure 57: JetGood1_eta
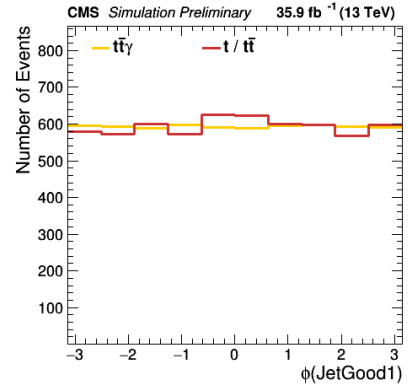
Figure 58: JetGood1_phi
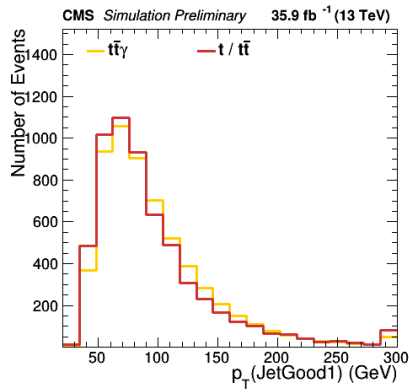
Figure 59: JetGood1_pt
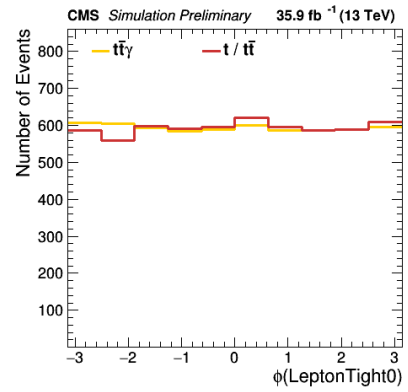
Figure 60: LeptonTight0_phi
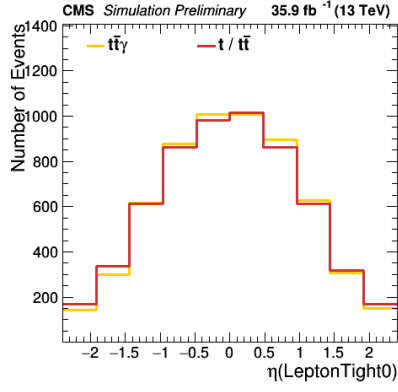
Figure 61: LeptonTight0_eta
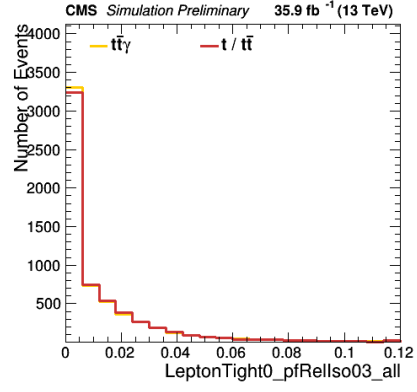
Figure 62: LeptonTight0_pfRelIso03_all
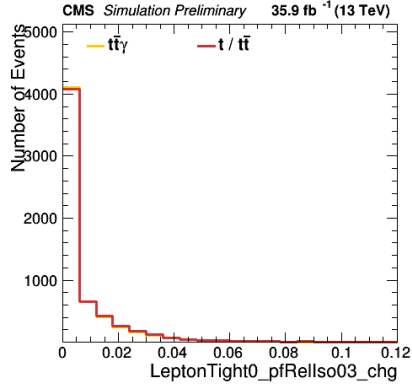
Figure 63: LeptonTight0_pfRelIso03_chg
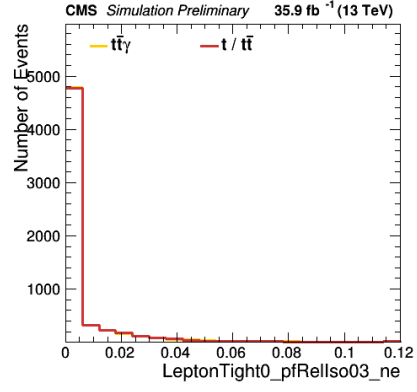
Figure 64: LeptonTight0_pfRelIso03_ne
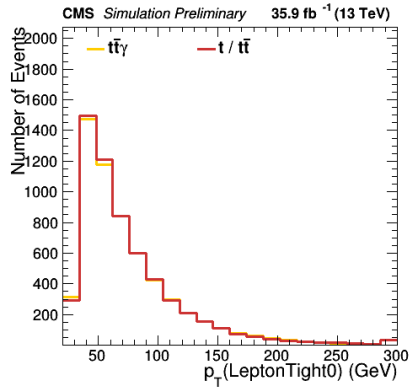
Figure 65: LeptonTight0_pt
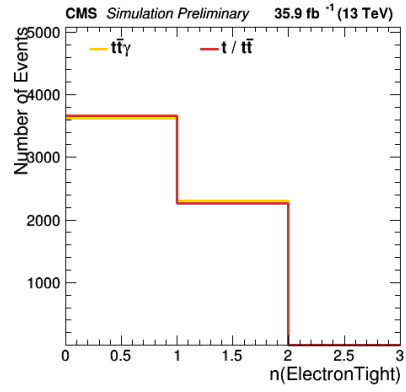
Figure 66: nElectronTight
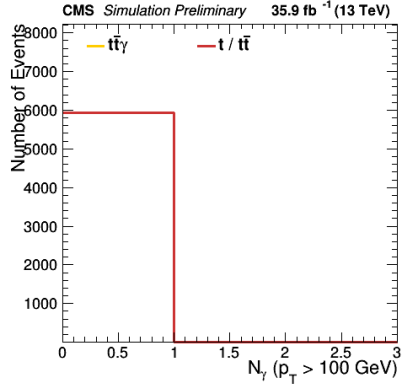
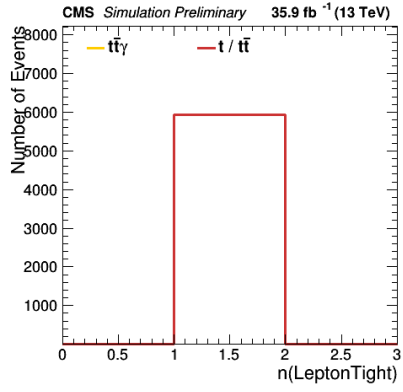Figure 67: nHighPTPhotons
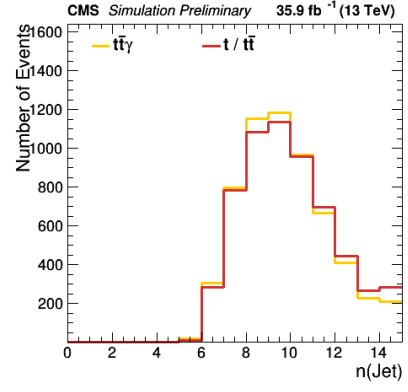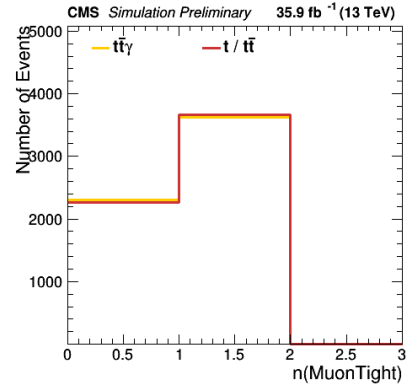

Figure 68: nJet
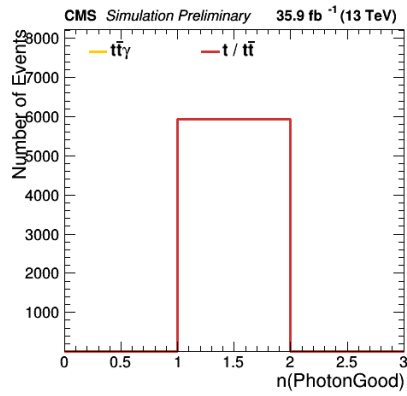

Figure 69: nLeptonTight

Figure 70: nMuonTight

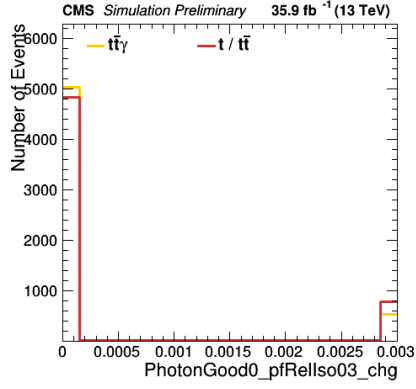Figure 71: nPhotonGood

9.2 Excluded variables
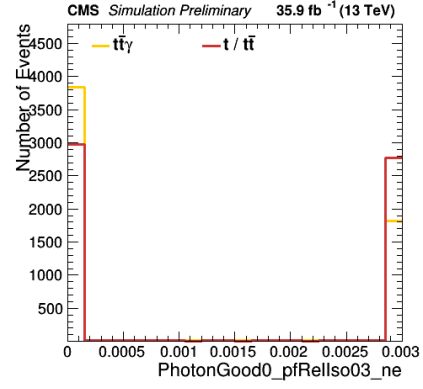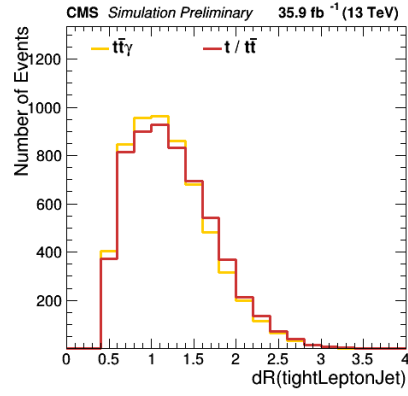

Figure 72: PhotonGood0_pfRelIso03_chg

Figure 73: PhotonGood0_pfRelIso03_ne

Figure 74: tightLeptonJetdR