TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

BACHELORARBEIT

Discrimination between $t\bar{t}\gamma$ and $t\bar{t}$-background in the signal region

Ausgeführt am Institut für Hochenergiephysik der Österreichischen Akademie der Wissenschaften

unter der Anleitung von
Dipl.-Ing. Dr.techn. Robert Schöfbeck

durch
Meikee PAGSINOHIN

Matrikelnummer: 01327477

Contents

1. Introduction **3**
- 1.1. Motivation . . . 3
- 1.2. Approach . . . 3

2. Background information **3**
- 2.1. CMS detector and data . . . 3
- 2.2. Proton-Proton-collision and $t\bar{t}\gamma$ . . . 4

3. Challenge in categorisation **5**

4. Analysis process **6**
- 4.1. Input parameters . . . 6
- 4.2. Analysis of variables . . . 7

5. MVA **12**
- 5.1. BDT - Boosted Decision Trees . . . 12
- 5.2. MLP - Multilayer Perceptron . . . 13
- 5.3. Training of MVA . . . 14
- 5.4. Results . . . 16

6. Conclusion **25**

A. Appendix **26**
- A.1. Full list of variables with almost identical distributions . . . 26
- A.2. Configuration options for MVA method: BDT . . . 31
- A.3. Configuration options for MVA method: MLP . . . 33

B. References **35**

C. List of Figures **36**

D. List of Tables **38**

1. Introduction

1.1. Motivation

The 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 performed through a multivariate analysis by developing and training models. At the beginning, different variables were analysed and compared, and the most promising were then used for training. The type of models, denoted MVA in the following, used for this project are boosted decision trees (BDT) and multilayer perceptrons (MLP). The results of the MVAs for three different variable categories were created, then analysed and presented in the course of this project.

2. Background information

2.1. CMS detector and data

The CMS detector (compact muon solenoid) is part of the Large Hadron Collider, a particle accelerator, which was built by the research organisation CERN in Geneva, Switzerland. LHC has a circumference of 27 km and is up to 175 m deep. Besides CMS, there are other experiments as well, such as ATLAS, ALICE and LHCb. CMS has multiple sub-detectors, which all surround the location, where the particles are accelerated and collided against each other.

In Figure 1, a sketch of the detector system can be seen. The inner tracking system is based on silicon detector technology (SI microstrip and pixel) and is used to track charged particles. It is enclosed by electromagnetic and hadron calorimeters (ECAL, HCAL) utilized to prompt showers. They have a barrel and an endcap part and additionally a forward calorimeter was installed to further extend the detector's coverage. The ECAL is built out of lead tungstate crystals and the HCAL is made out of brass and scintillator. Special for the CMS detector is the superconducting solenoid, which surrounds the inner detectors and calorimeters. It has a 6 m diameter with a magnetic field of 3.8 T used to divert the particle trajectory and to measure their momentum. The muon detectors are positioned in the outermost layer since muons may pass all inner layers [1, 2, 3].

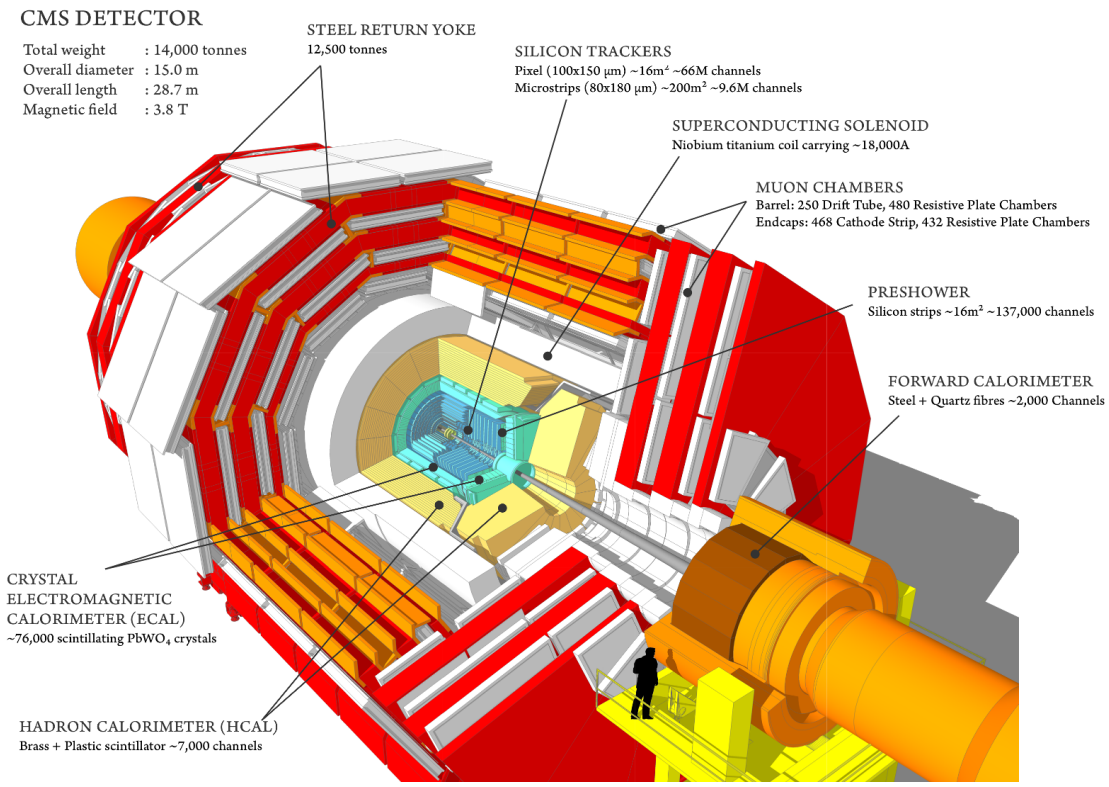


Figure 1: Detector system of CMS, sketch taken from [4]

The data is filtered by a two-step trigger system. The hardware-based system (L1) is triggered when muons, electrons, photons and jets with specific characteristics are detected. This reduced the data amount from 40 MHz to 100 kHz. The software-based high-level trigger selects events based on a rudimentary event reconstruction and reduces the rate from 100 kHz to 1 kHz. The data used during this project was recorded in 2016 at a centre-of-mass energy of 13 TeV. The signal and background processes were modelled using multiple Monte Carlo event generators [2].

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 at four interaction points. During the pp collision a top quark pair is produced. The top quark then decays to 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.

The $t\bar{t}\gamma$ events were selected for this project - 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 their 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 [1]. By analysing the top quark, the results can be used to test the Standard Model and its possible extensions [5].

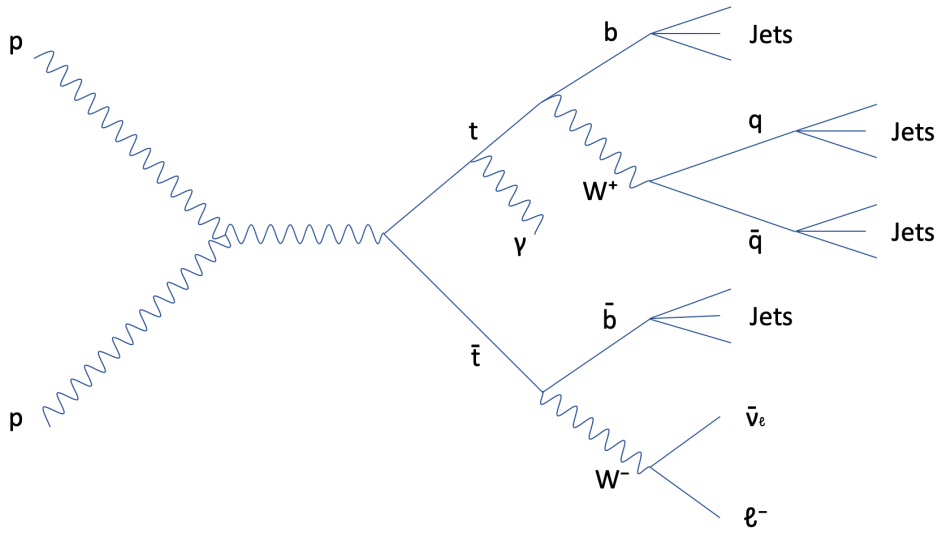


Figure 2: Proton Proton collision process

3. Challenge in categorisation

Events with the following conditions were selected:

- lepton has 30/35 GeV
- at least 3 jets were detected
- at least one of the jets have a b-tag
- exactly one photon

In Figure 3, the transverse momentum of the photon is visible. The background makes up nearly 50% of the signal region. The possible sources are [1, 5]:

- hadron or electron misidentified as photon
- incorrectly reconstructed photon
- photon radiated from different process
- additional particles not detected due to limited detector's acceptance
- hadron incorrectly reconstructed as electron

A neural network was developed to differentiate between the correctly reconstructed and the misidentified photons. What still needs to be 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.

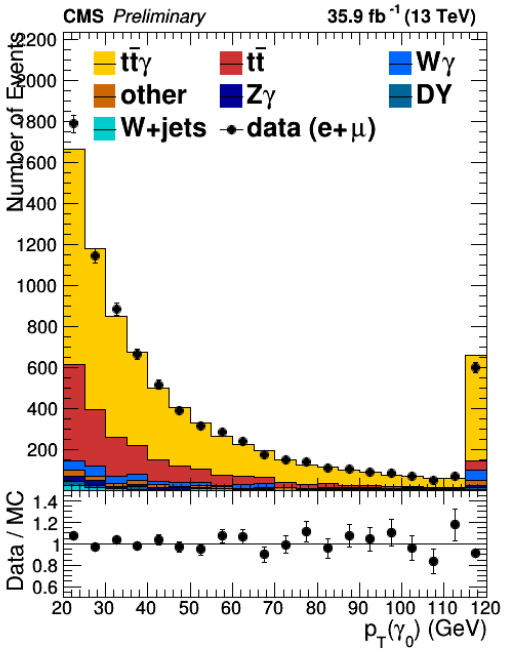


Figure 3: Distribution of p_T photon

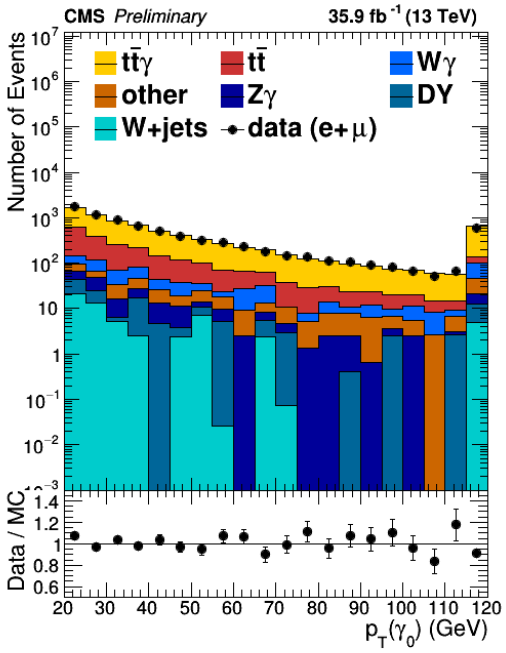


Figure 4: Distribution of p_T photon (log)

4. Analysis process

4.1. Input parameters

The amount of detected particles and the pseudorapidity (η = -ln(tan($\frac{\vartheta}{2}$)), whereas ϑ is the polar angle) of leptons, photons, first and second jets and b-jets were analysed. B-jets are jets with a b-tag and the first and second jet refer to the jet with the highest and second highest p_T respectively. The azimuthal opening angle around the z-axis ($\Delta\phi$) and transverse momentum (p_T) of these particles were looked at as well. The distance ($\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$) between lepton and jet, lepton and photon and photon and jet were also considered. Additional variables are:

- total transverse momentum of all jets; higher amount means either a large number of jets or jets with high energies were detected
- maximum mass of three jets; during the decay process of a top quark to a b-quark and W-boson, which further decays to two quarks (see chapter 2.2), the variable would then have a peak at the top quark mass

- transverse mass of an object depending on input variables; if a lepton and neutrino were used in the following formula, then the transverse mass of a W-boson is calculated:

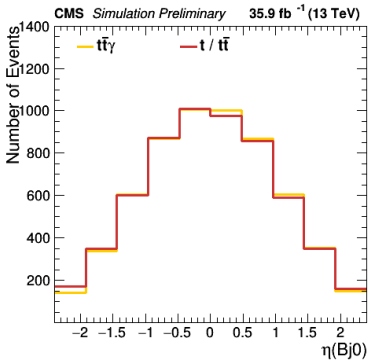
$$m_T(W) = \sqrt{2p_T(\ell)p_T^{miss}[1 - \cos((\Delta\phi(\ell, \vec{p}_T^{miss}))]}$$

- currently used variable to differentiate between signal and background events
- proportion of energy coming from neutral particles detected in ECAL (e.g. photon)
- proportion of energy coming from charged particles detected in ECAL (e.g. electron)
- proportion of energy coming from neutral particles detected in HCAL (e.g. neutral meson, neutron)
- proportion of energy coming from charged particles detected in HCAL (e.g. charged meson, proton)
- relative isolation of the lepton to all particles in a radius of ΔR = 0.3
- relative isolation of the lepton to charged particles in a radius of ΔR = 0.3
- relative isolation of the lepton to neutral particles in a radius of ΔR = 0.3
- result of neural network to identify first and second jets coming from b-quark (b-tag)
- $\Delta\phi$ between lepton and photon
- $\Delta\phi$ of missing transverse momentum (MET)
- transverse momentum of MET

4.2. Analysis of variables

The variables were analysed by comparing the distribution of events with a uniform scaling. A uniform scaling was needed to have a better view of relative differences between signal ($t\bar{t}\gamma$) and background ($t\bar{t}$). If the variable shows a different behaviour for each type of event, it could be an indication of discriminatory power.

Figure 5 and 6 depict some examples of variables, where the distributions are almost identical and were therefore not analysed any further. A full list of these variables are listed in the Annex, chapter A.1.



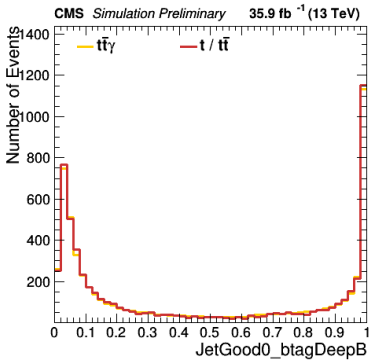
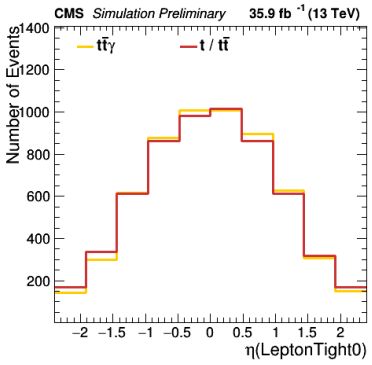


Figure 5: pseudorapidity of first b-jet and result of neural network to identify b-jets suggest no discriminatory power



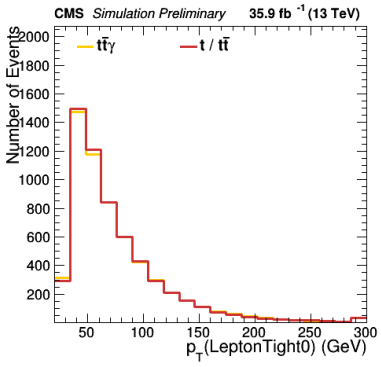


Figure 6: pseudorapidity and transverse momentum of lepton suggest no discriminatory power

Variables, where some differences were visible in the signal region, are seen in Figure 7 to 12. These were candidates for the first set of selected variables.

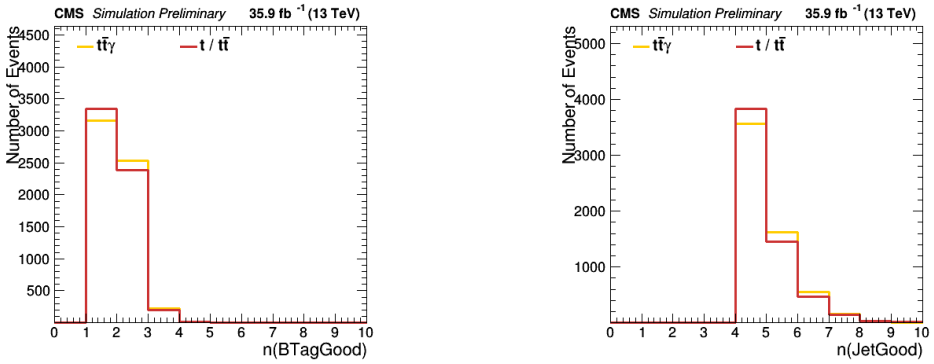


Figure 7: number of jets and b-jets show slightly different behaviour between signal and background

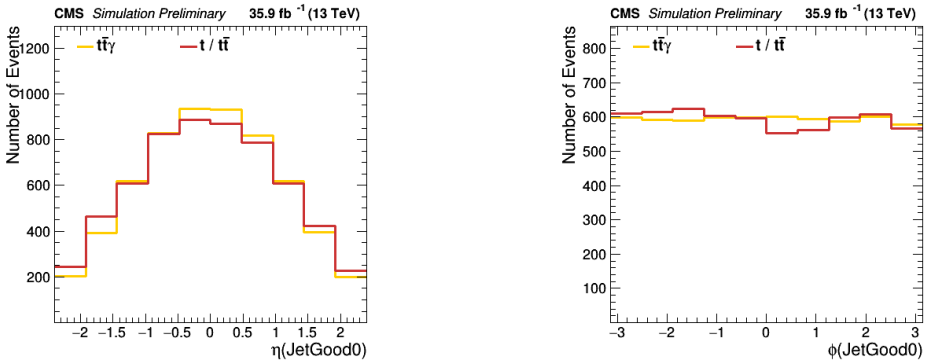


Figure 8: pseudorapidity and $\Delta\phi$ of the first jet show slightly different behaviour between signal and background

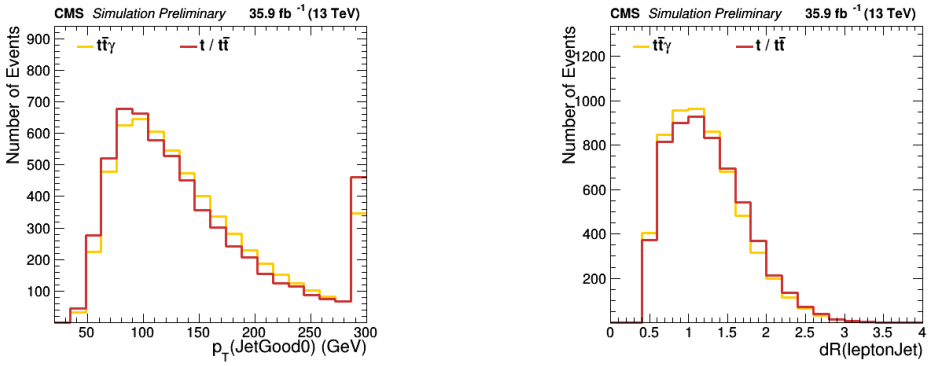


Figure 9: transverse momentum of first jet and distance between lepton and jet show slightly different behaviour between signal and background

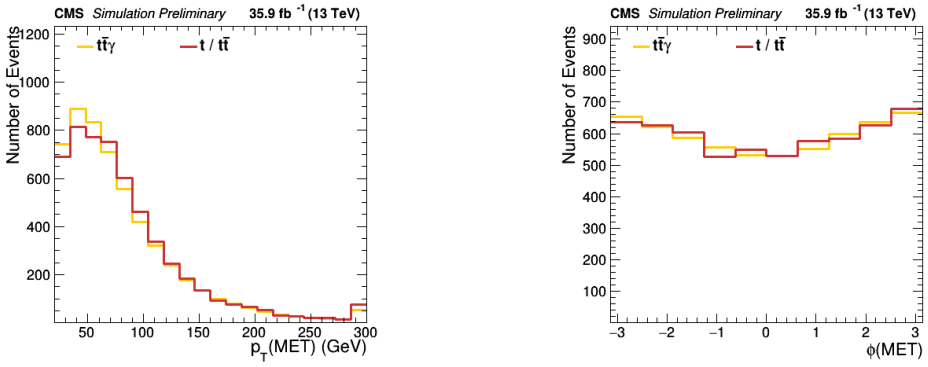


Figure 10: transverse momentum and $\Delta\phi$ of missing transverse momentum show slightly different behaviour between signal and background

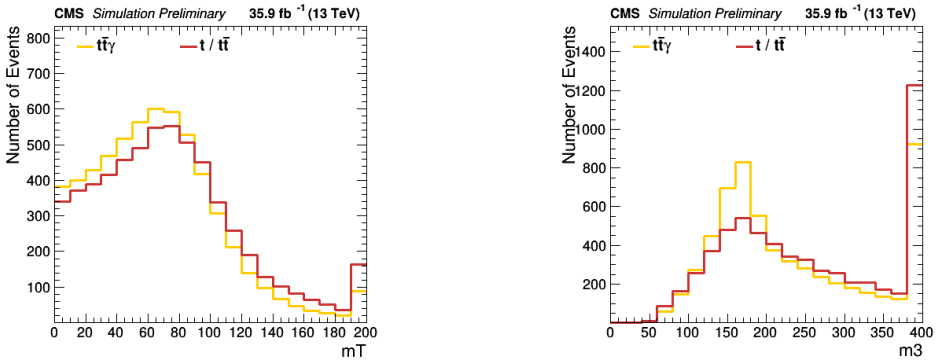


Figure 11: transverse momentum of W-boson and maximum mass of three jets show slightly different behaviour between signal and background

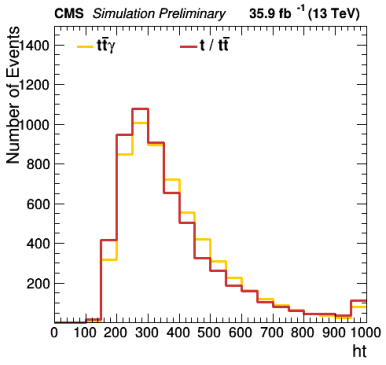


Figure 12: transverse momentum of all jets show slightly different behaviour between signal and background

The currently used variable to differentiate between $t\bar{t}\gamma$ and $t\bar{t}$, seen in Figure 13, shows quite a good distinction and is therefore added to the list (second selection).

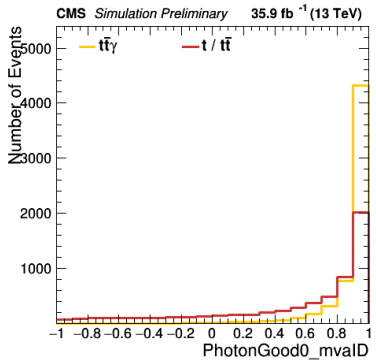
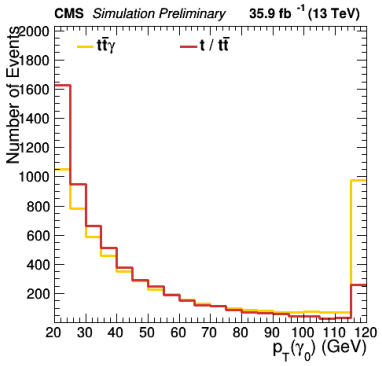


Figure 13: currently used variable to differentiate between signal and background events suggests good distinction

Figure 14 to 16 are the photon variables. These demonstrate the most promising behaviour for discriminating between $t\bar{t}\gamma$ and $t\bar{t}$ events and were additionally put into the third selection category.



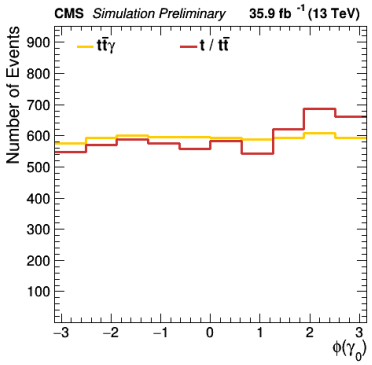


Figure 14: transverse momentum and $\Delta\phi$ of photon indicate good discriminatory power

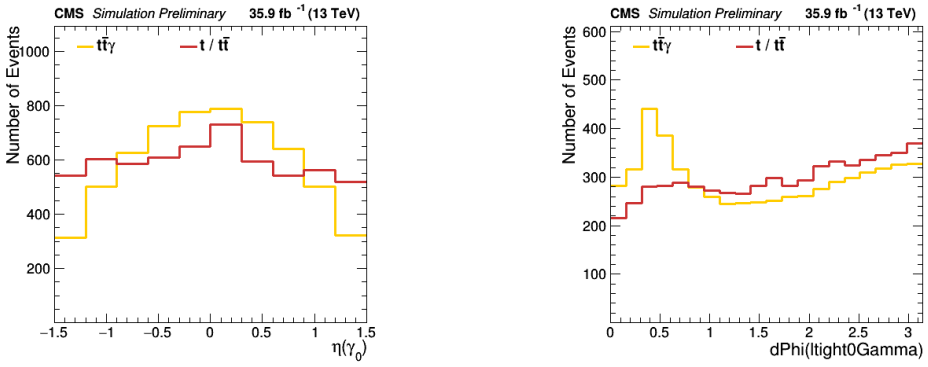


Figure 15: pseudorapidity of photon and $\Delta\phi$ between lepton and photon indicate good discriminatory power

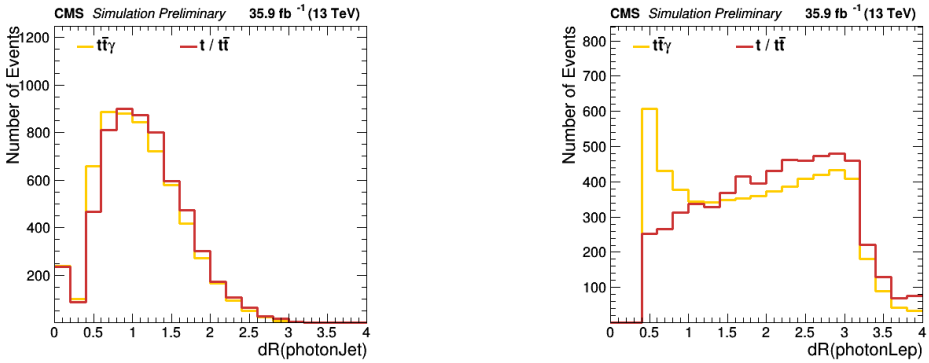


Figure 16: distance between photon and jet and between photon and lepton indicate good discriminatory power

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, multiple 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 that separates 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 criteria 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 classifiers. 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 [6].

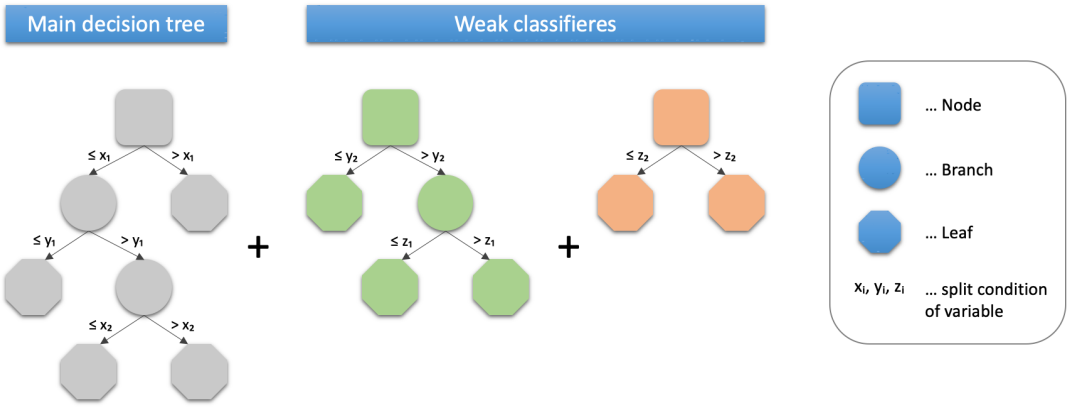
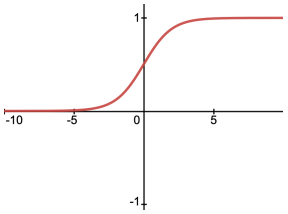


Figure 17: Outline of Boosted Decision Trees

5.2. MLP - Multilayer Perceptron

The model is built out of multiple perceptrons, which are arranged in layers shown in Figure 19. The value of one perceptron is the sum of all weighted inputs. This value is then transformed through an activation function and a threshold. The activation function is a linear function for the input- and outputlayer, while it is usually the log-sigmoid transfer function for the hiddenlayers, as seen in Figure 18. All layers have a threshold, except the inputlayer.



$$f(x) = \frac{1}{1 + e^{-x}}$$

Figure 18: 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, a so called weight coefficient. All coefficients of one layer can be written as a weight matrix. The first and last layer are called input- and outputlayer respectively, while the intermediate layers are the hid-

denlayers. A MLP has a minimum of one hiddenlayer, therefore contains at least three layers [7].

If the signal is only transmitted in one direction, therefore the MLP doesn't have any 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 expected 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 they 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 [8].

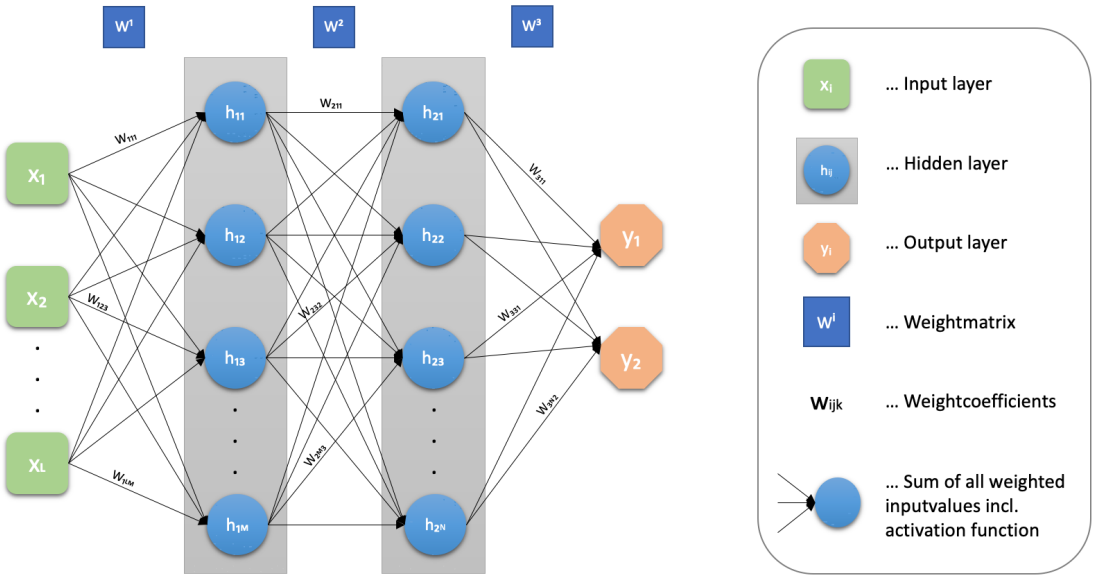


Figure 19: Outline of Multilayer Perceptron

5.3. Training of MVA

The default values of the following configuration options were used for the models:

- BDT
  - Minimum percentage required in a leaf node (MinNodeSize): 5%
  - Separation criterion for node splitting (SeparationType): GiniIndex
- MLP
  - Number of training cycles (NCycles): 500
  - ANN learning rate parameter (LearningRate): 0.02

  - Decay rate for learning parameter (DecayRate): 0.01
  - Neuron input function type (NeuronInputType): sum
  - Training Type (TrainingMethod): BP (Back-Propagation)

A full list of all parameters, configuration options as well as their default values are given in the Annex, chapter A.2 and A.3 [9].

These parameters were adapted to potentially improve the result:

- BDT
  - Maximum depth of decision trees allowed (MaxDepth)
  - Number of trees in the forest (NTrees)
  - Number of grid points used to find optimal cut (nCuts)
- MLP
  - Number of hidden layers (HiddenLayers)

Their default values were used as a basis, listed in the second row in Table 1. The MVAs were additionally trained with a lower and a higher configuration (row 2 and 3):

	BDT			MLP
Configuration	MaxDepth	NTrees	nCuts	HiddenLayers
1	1	400	10	5
2	3	800	25	7
3	5	1000	50	10

Table 1: Configuration combinations used in MVAs

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 (receiver operating characteristic curve) show the background rejection versus signal efficiency. The AUC (area under the ROC-curve) is used as a parameter to demonstrate how well a model performs. 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 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, called false positive

Examples for different types of ROC-curves are displayed in Figure 20. The green curve indicate a good discriminatory power. The orange curve shows that the model assigns the events the wrong way around. If the ROC-curve follows the identity line, then the model doesn't have any discriminatory power.

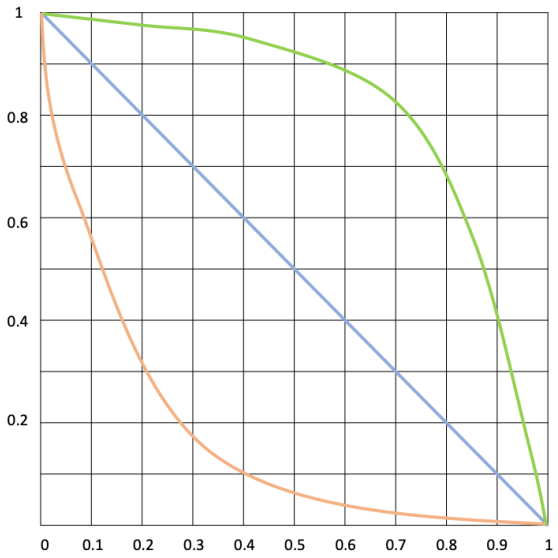


Figure 20: ROC-curve examples

5.4. Results

The first set of selected variables (non-photon) were put into the models and are seen in Figure 21 - 23. The ROC-curves indicate a slight improvement in the split between $t\bar{t}\gamma$- and $t\bar{t}$-events (blue line represents the identity line). At a signal efficiency of 70%, a background rejection of about 40% for MLP and 35% for BDT can be achieved.

In the second selection category, the currently used variable to differentiate between signal and background events was added (Figure 24 - 26). The result has improved, the AUC increased significantly. At a signal efficiency of 70%, a background rejection of about 46% for MLP and 37% for BDT can be achieved.

Now adding the photon variables for the third selection shows a better success in the distinction process. They can be seen in Figure 27 - 29. At a signal efficiency of 70%, a background rejection of about 48% for MLP and 38% for BDT can be achieved.

Between the three configuration combinations, there is only a slight difference for the MLP model, but for the BDT model there is an increase in the AUC for configuration 2 visible. Training- and test-results are nearly identical, so there are no signs of overfitting. The MLP model performed better than the BDT model for all selection categories. The third set of variables shows the best result.

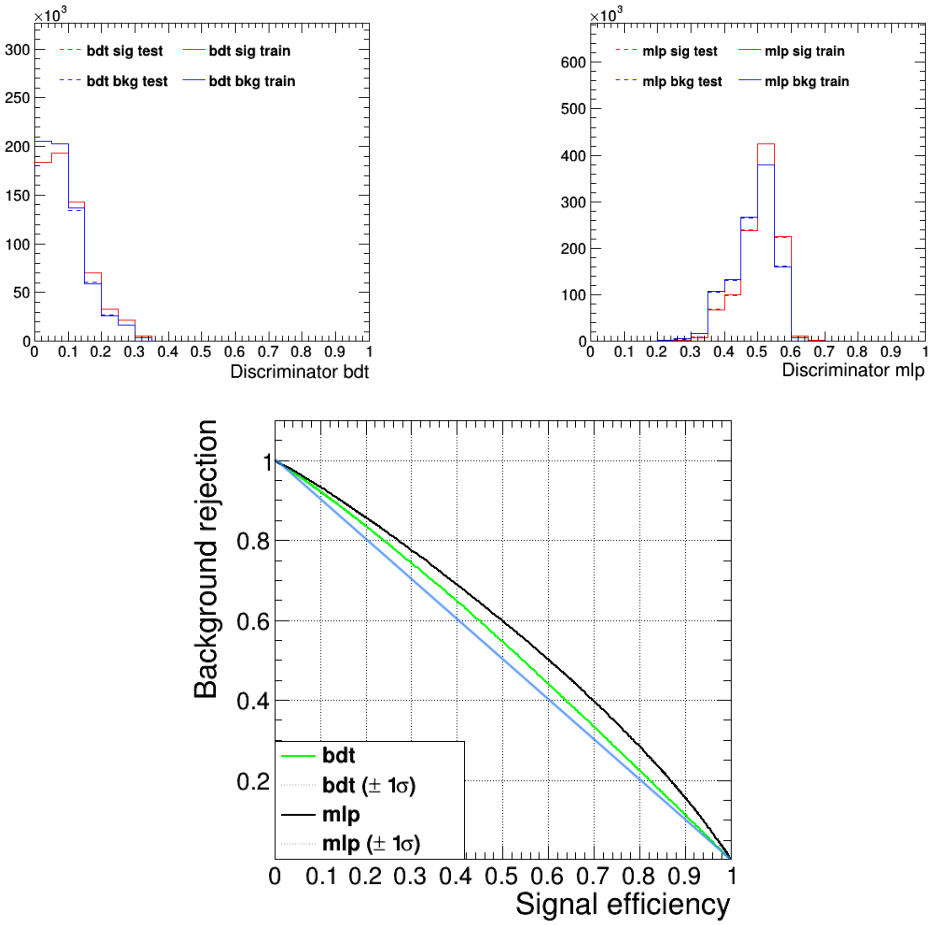


Figure 21: BDT-, MLP-distribution and ROC curve for variables from selection 1 with configuration 1

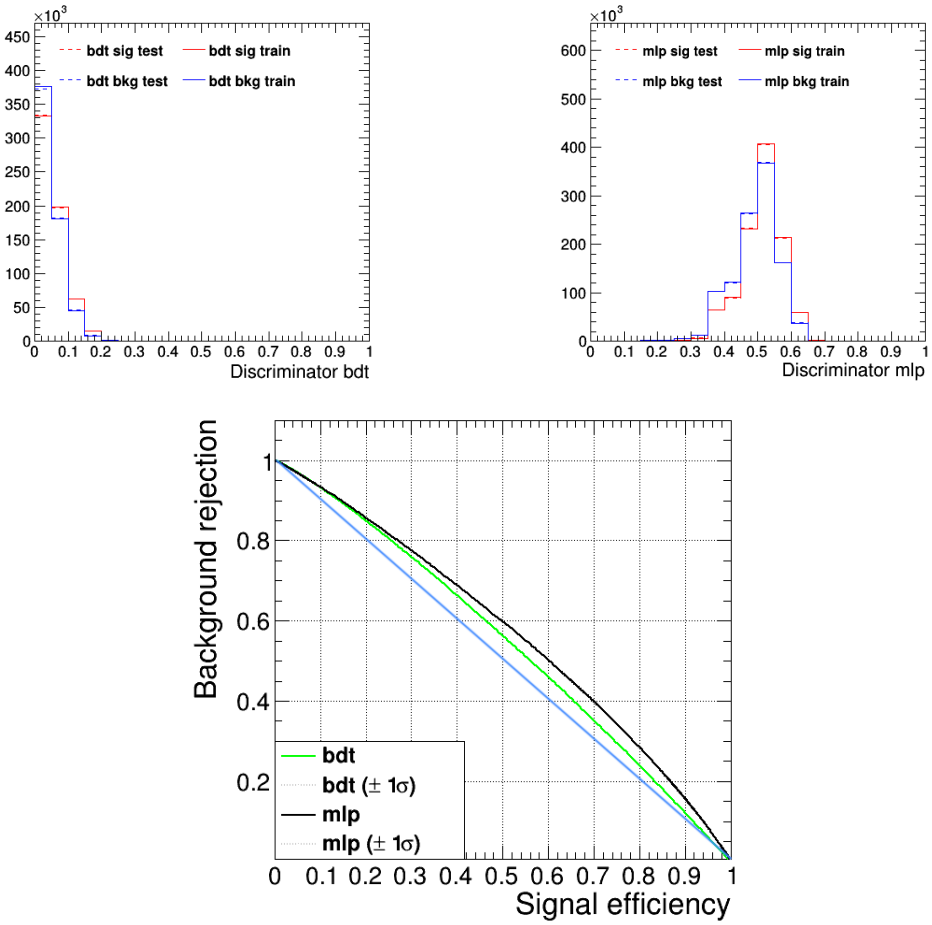


Figure 22: BDT-, MLP-distribution and ROC curve for variables from selection 1 with configuration 2

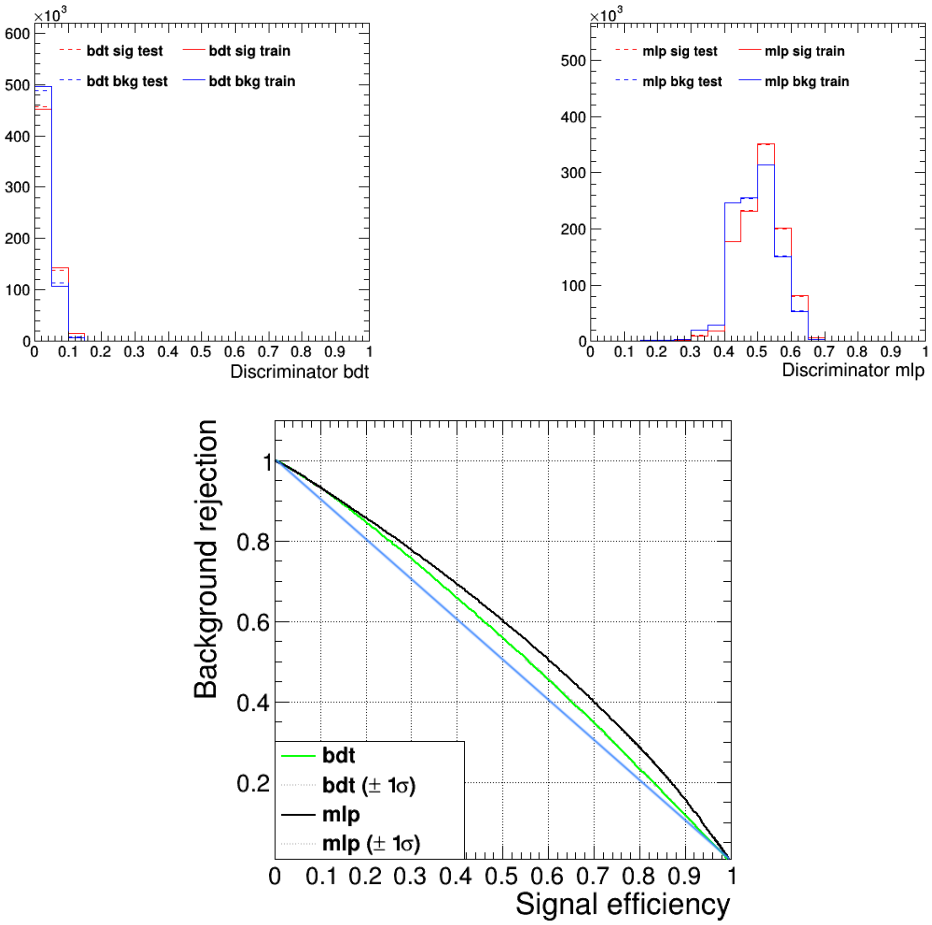


Figure 23: BDT-, MLP-distribution and ROC curve for variables from selection 1 with configuration 3

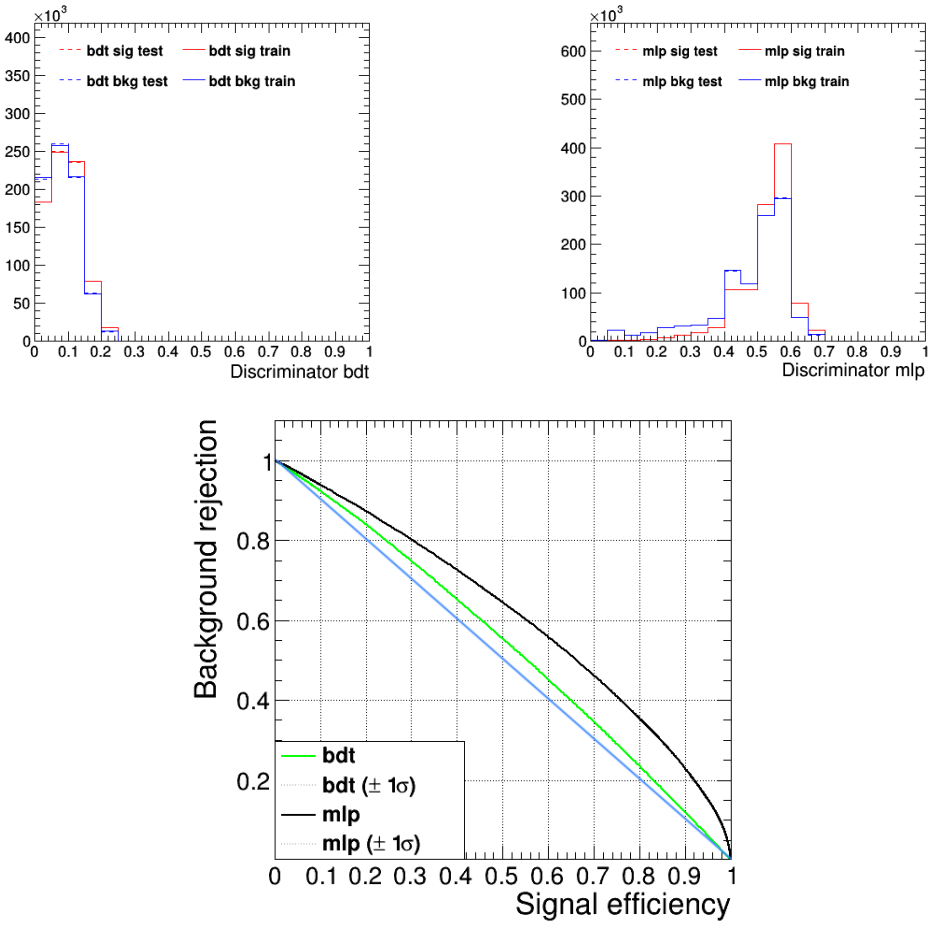


Figure 24: BDT-, MLP-distribution and ROC curve for variables from selection 2 with configuration 1

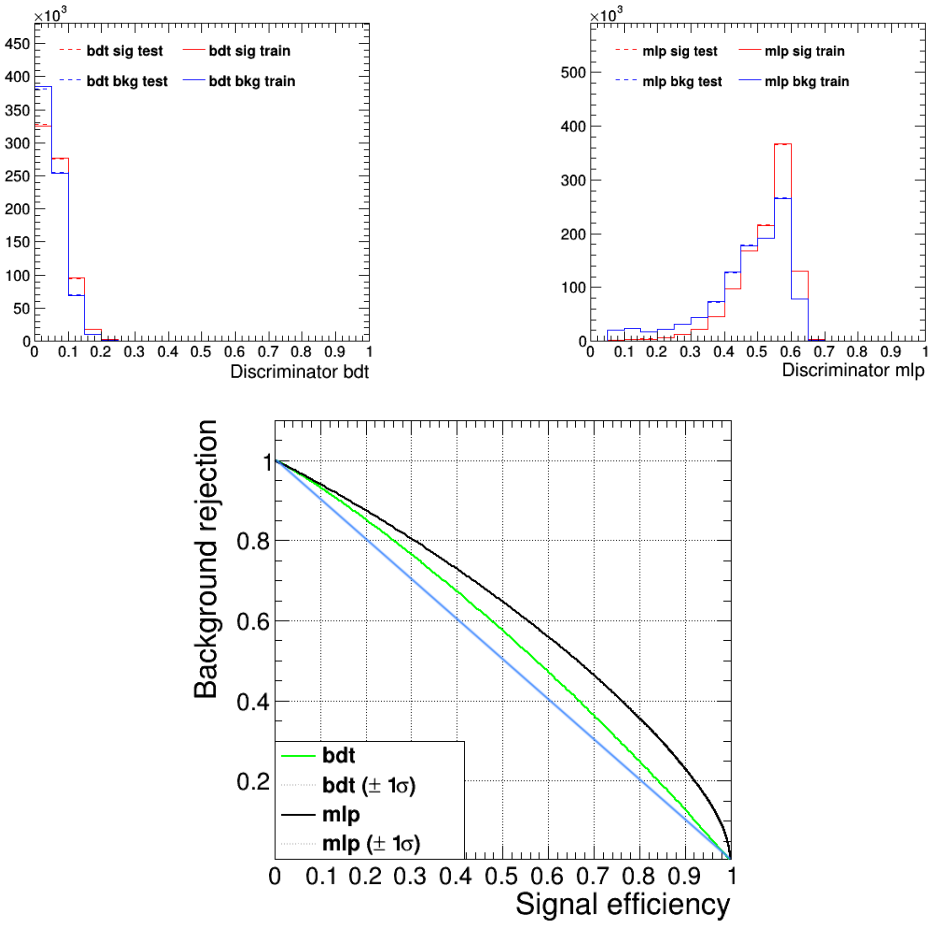


Figure 25: BDT-, MLP-distribution and ROC curve for variables from selection 2 with configuration 2

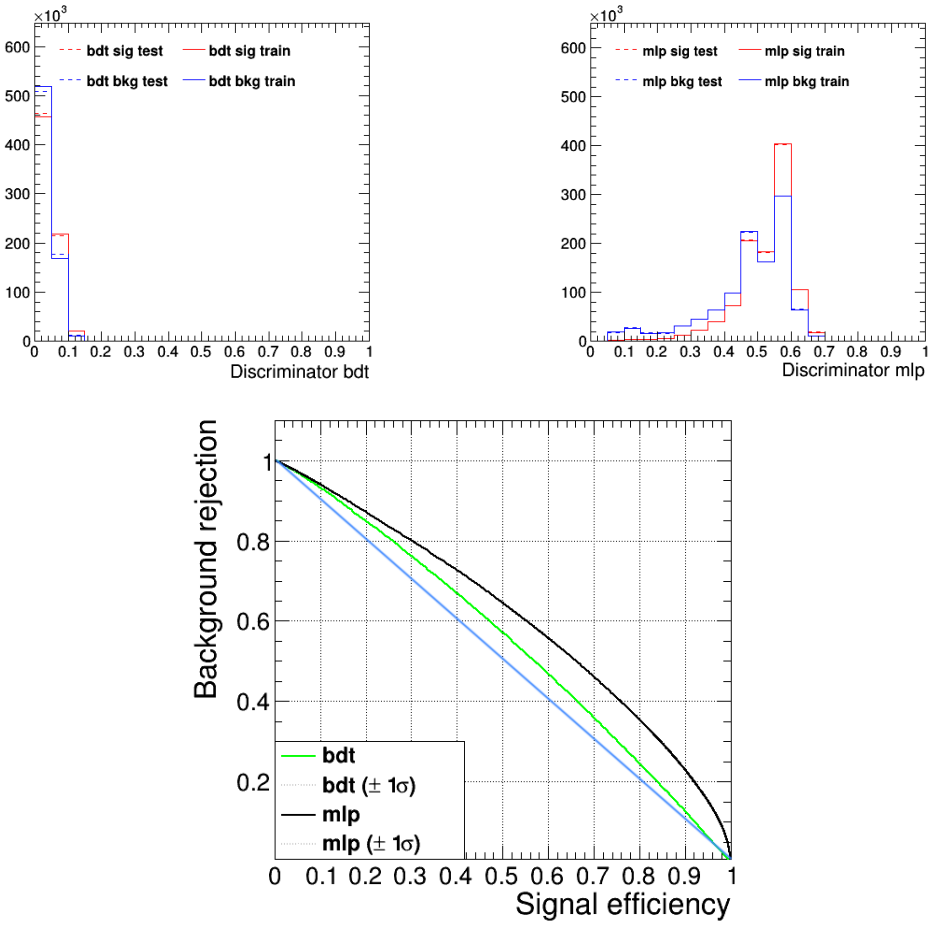


Figure 26: BDT-, MLP-distribution and ROC curve for variables from selection 2 with configuration 3

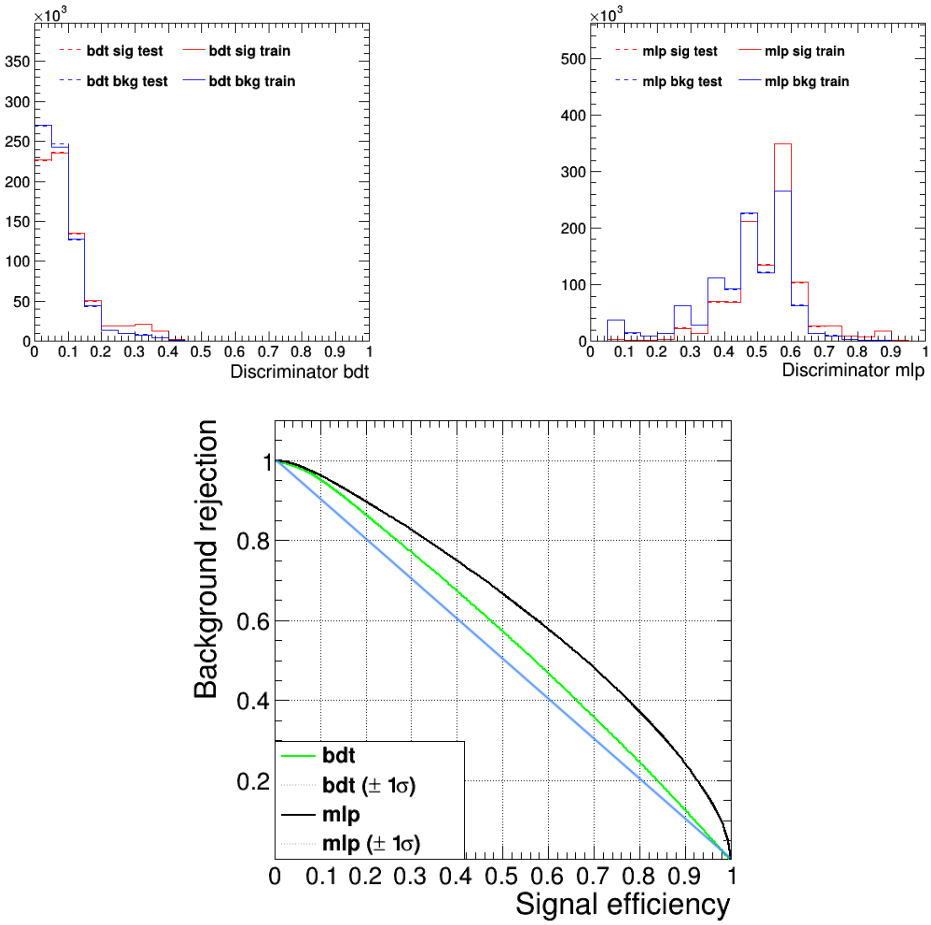


Figure 27: BDT-, MLP-distribution and ROC curve for variables from selection 3 with configuration 1

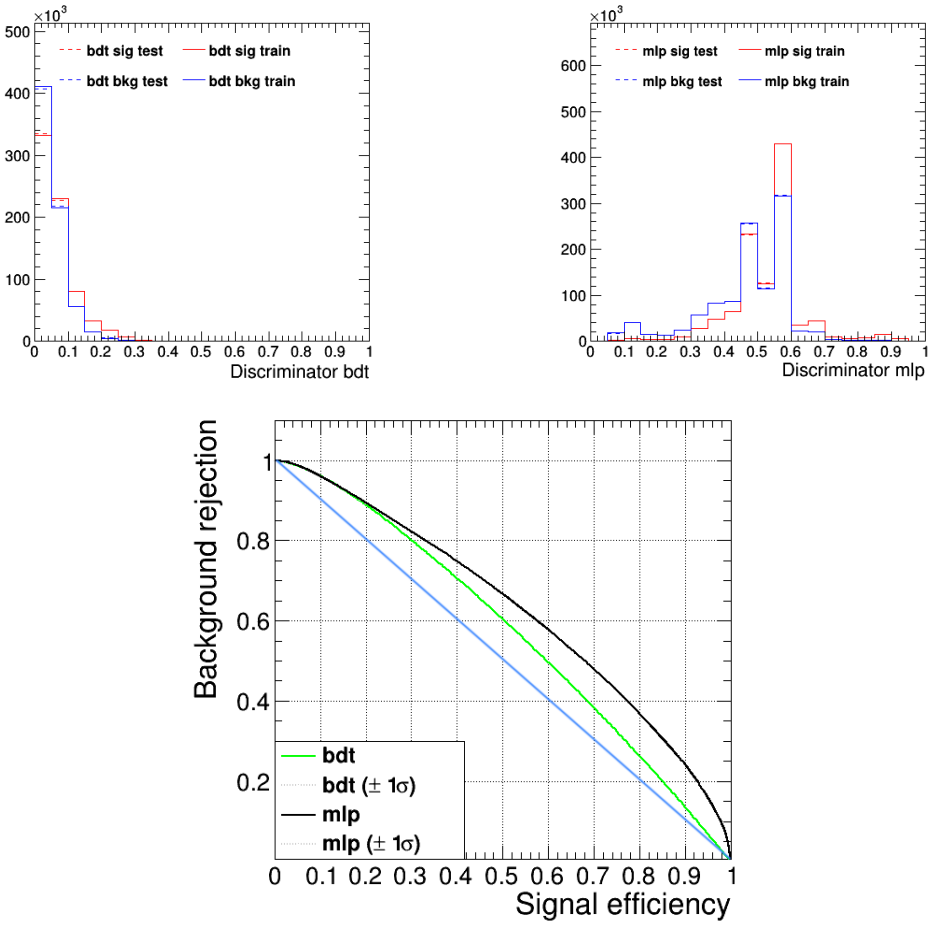


Figure 28: BDT-, MLP-distribution and ROC curve for variables from selection 3 with configuration 2

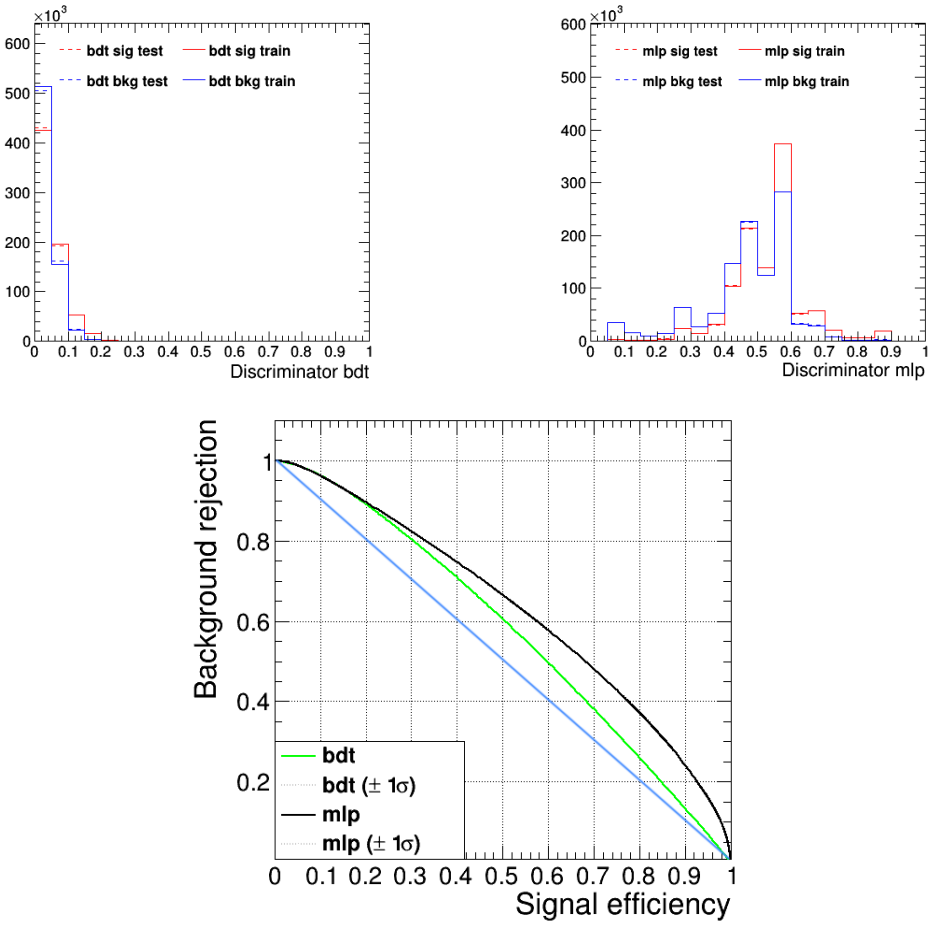


Figure 29: BDT-, MLP-distribution and ROC curve for variables from selection 3 with configuration 3

6. Conclusion

It is very difficult to differentiate between $t\bar{t}\gamma$ and $t\bar{t}$ events because they show very similar behaviour in the signal region. Nevertheless, a slight improvement was possible by training a MVA with relevant variables. At a signal efficiency of 70%, a background rejection of up to 48% for MLP and up to 38% for BDT can be achieved. Especially photon variables provide a bigger improvement. MLPs show a better performance than BDTs with a 5-10% difference observed in background rejection at 70% signal efficiency, which suggests the selection of this type of model.

A. Appendix

A.1. Full list of variables with almost identical distributions

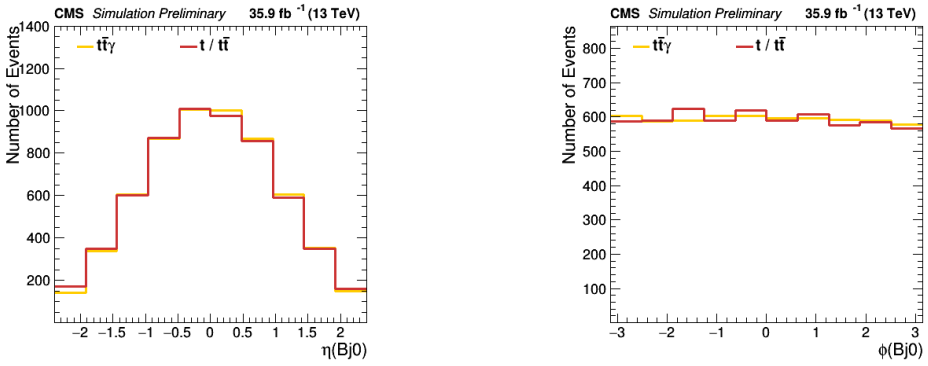


Figure 30: pseudorapidity and $\Delta\phi$ of first b-jet suggest no discriminatory power

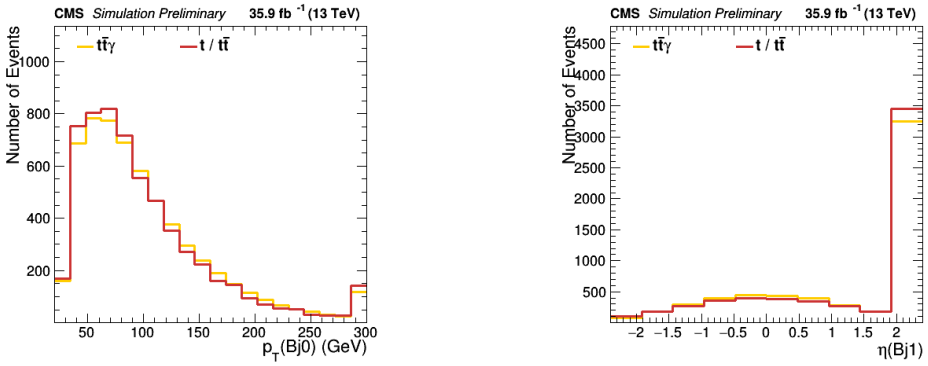


Figure 31: transverse momentum of first b-jet and pseudorapidity of second b-jet suggest no discriminatory power

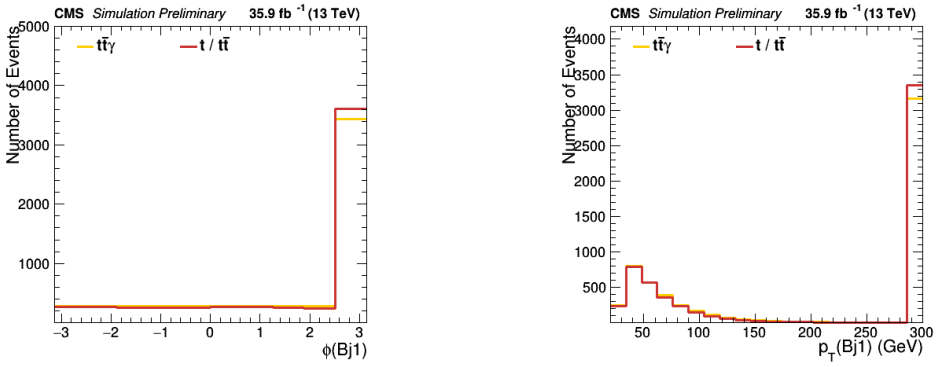


Figure 32: $\Delta\phi$ and transverse momentum of second b-jet suggest no discriminatory power

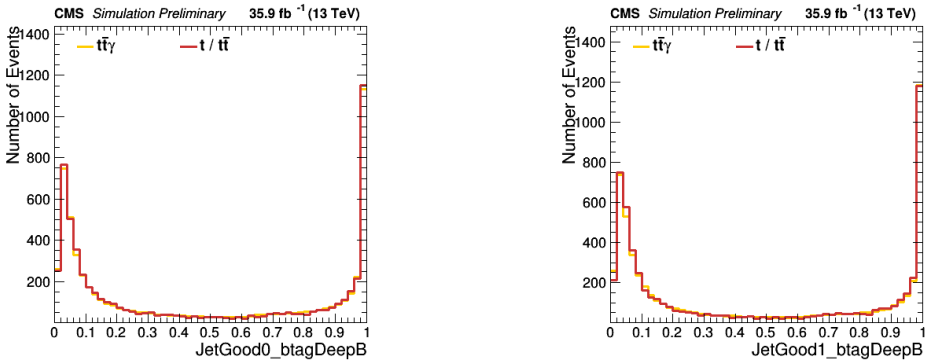


Figure 33: results of neural network to identify first and second b-jet suggest no discriminatory power

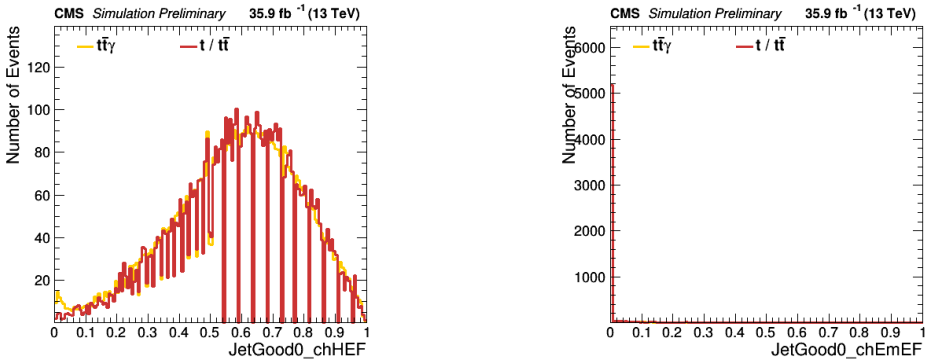


Figure 34: proportion of energy coming from charged particles detected in HCAL and detected in ECAL suggest no discriminatory power

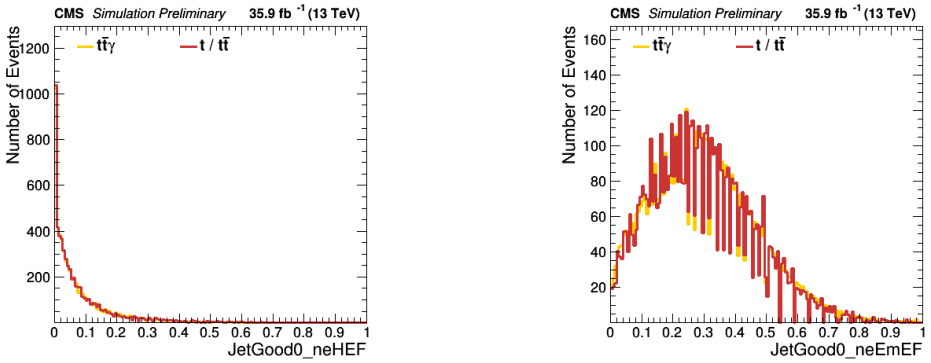


Figure 35: proportion of energy coming from neutral particles detected in HCAL and detected in ECAL suggest no discriminatory power

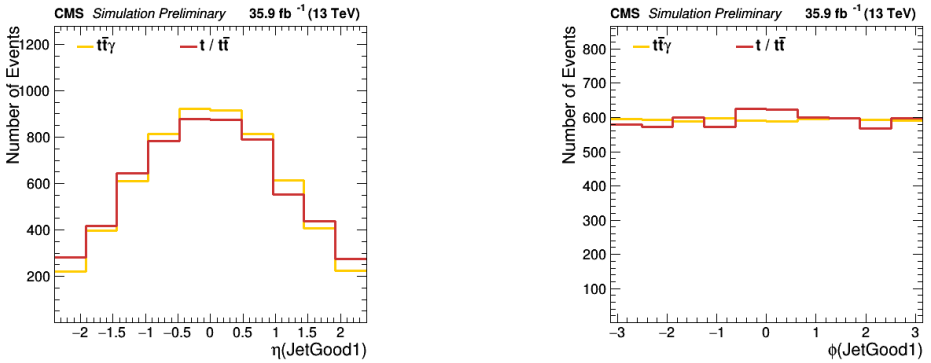


Figure 36: pseudorapidity and $\Delta\phi$ of second jet suggest no discriminatory power

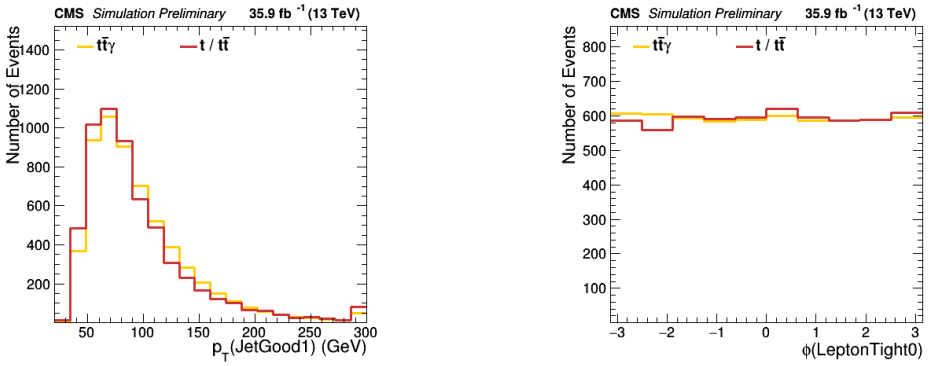


Figure 37: transverse momentum of second jet and $\Delta\phi$ of lepton suggest no discriminatory power

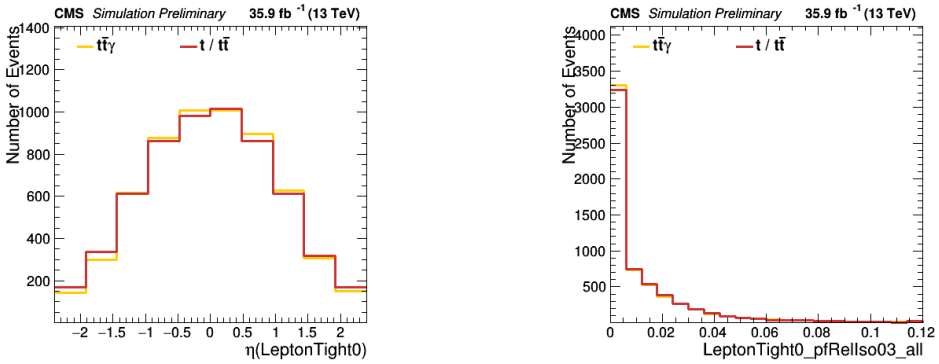


Figure 38: pseudorapidity of lepton and relative isolation of lepton to all particles suggest no discriminatory power

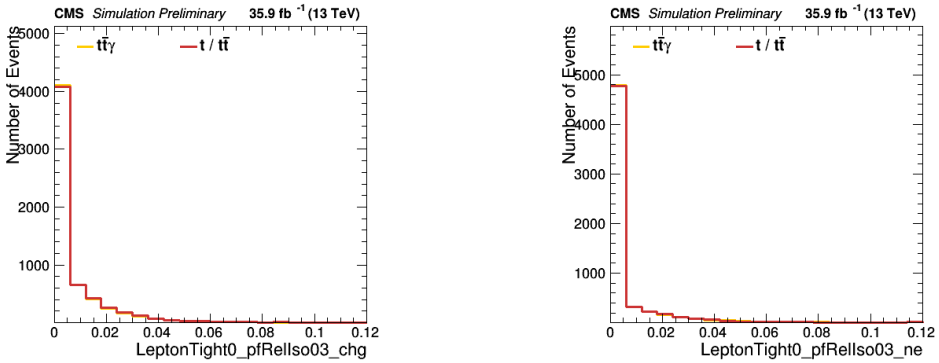


Figure 39: relative isolation of lepton to charged and neutral particles suggest no discriminatory power

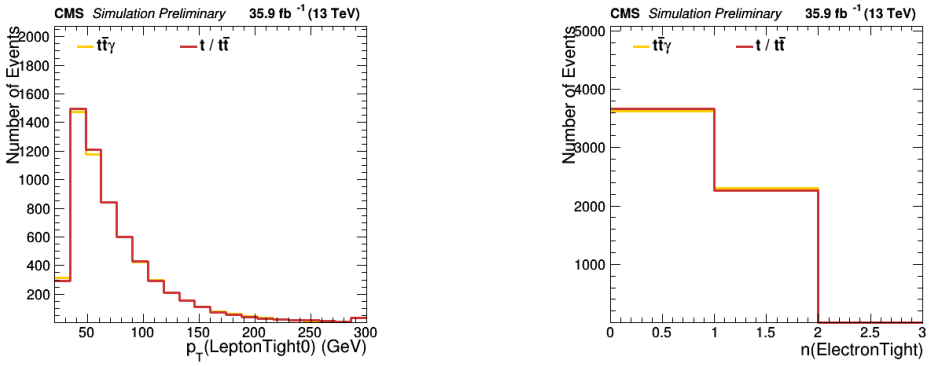
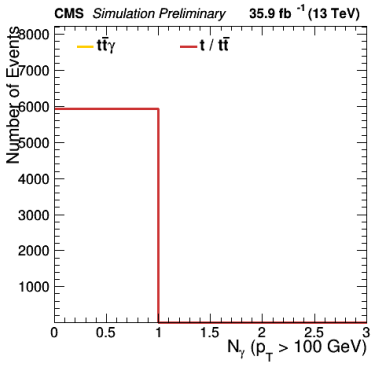


Figure 40: transverse momentum of lepton and number of electrons suggest no discriminatory power



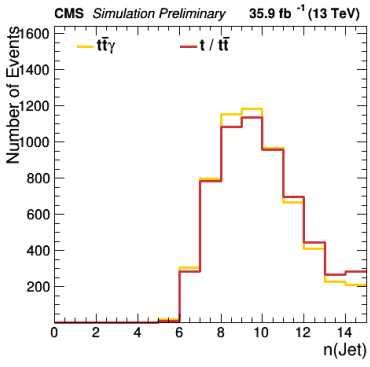
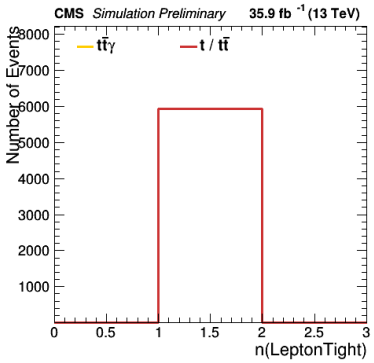


Figure 41: number of photons with high p_T and number of jets suggest no discriminatory power



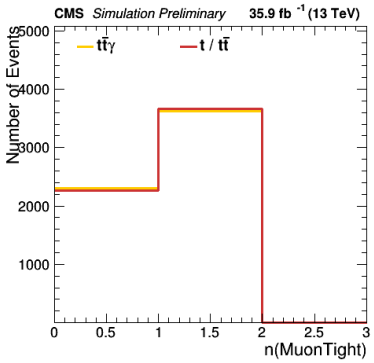


Figure 42: number of leptons and muons suggest no discriminatory power

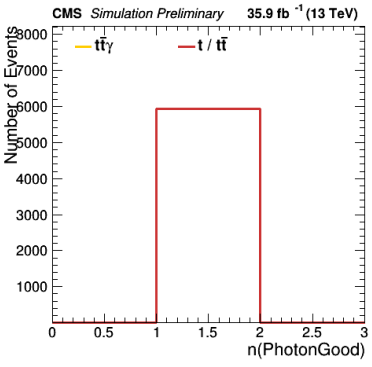


Figure 43: number of photons suggests no discriminatory power

A.2. Configuration options for MVA method: BDT

Note: Table taken from [9], p. 116-118.

Option	Default	Description
NTrees	800	Number of trees in the forest
MaxDepth	3	Max depth of the decision tree allowed
MinNodeSize	5%	Minimum percentage of training events required in a leaf node (default: Classification: 5%, Regression: 0.2%)
nCuts	20	Number of grid points in variable range used in finding optimal cut in node splitting
BoostType	AdaBoost	Boosting type for the trees in the forest (note: AdaCost is still experimental) (AdaBoost, RealAdaBoost, Bagging, AdaBoostR2, Grad)
AdaBoostR2Loss	Quadratic	Type of Loss function in AdaBoostR2 (Linear, Quadratic, Exponential)
UseBaggedGrad	False	Use only a random subsample of all events for growing the trees in each iteration. (Only valid for GradBoost)
Shrinkage	1	Learning rate for GradBoost algorithm
AdaBoostBeta	0.5	Learning rate for AdaBoost algorithm
UseRandomisedTrees	False	Determine at each node splitting the cut variable only as the best out of a random subset of variables (like in RandomForests)
UseNvars	2	Size of the subset of variables used with RandomisedTree option
UsePoissonNvars	True	Interpret UseNvars not as fixed number but as mean of a Possion distribution in each split with RandomisedTree option
BaggedSampleFraction	0.6	Relative size of bagged event sample to original size of the data sample (used whenever bagging is used (i.e. UseBaggedGrad, Bagging, ..))
UseYesNoLeaf	True	Use Sig or Bkg categories, or the purity=S/(S+B) as classification of the leaf node ->Real-AdaBoost

NegWeightTreatment	InverseBoost-NegWeights	How to treat events with negative weights in the BDT training (particular the boosting) : IgnoreInTraining; Boost With inverse boostweight (InverseBoost-NegWeights); Pair events with negative and positive weights in traning sample and *annihilate* them (PairNegWeights-Global, experimental!), Pray
NodePurityLimit	0.5	In boosting/pruning, nodes with purity >NodePurityLimit are signal; background otherwise.
SeparationType	GiniIndex	Separation criterion for node splitting (CrossEntropy, GiniIndex, GiniIndexWithLaplace, MisClassificationError, SdivSqrtSPlusB, RegressionVariance)
DoBoostMonitor	False	Create control plot with ROC integral vs tree number
UseFisherCuts	False	Use multivariate splits using the Fisher criterion
MinLinCorrForFisher	0.8	The minimum linear correlation between two variables demanded for use in Fisher criterion in node splitting
UseExclusiveVars	False	Variables already used in fisher criterion are not anymore analysed individually for node splitting
DoPreselection	False	Apply automatic pre-selection for 100% efficient signal (bkg) cuts prior to training
RenormByClass	False	Individually re-normalize each event class to the original size after boosting
SigToBkgFraction	1	Sig to Bkg ratio used in Training (similar to NodePurityLimit, which cannot be used in real adaboost)
PruneMethod	NoPruning	Note: for BDTs use small trees (e.g.MaxDepth=3) and NoPruning; Pruning: Method used for pruning (removal) of statistically insignificant branches (NoPruning, ExpectedError, CostComplexity)
PruneStrength	0	Pruning strength
PruningValFraction	0.5	Fraction of events to use for optimizing automatic pruning.

nEventsMin	0	deprecated: Use MinNodeSize (in % of training events) instead
GradBaggingFraction	0.6	deprecated: Use *BaggedSampleFraction* instead: Defines the fraction of events to be used in each iteration, e.g. when UseBaggedGrad=kTRUE.
UseNTrainEvents	0	deprecated: Use *BaggedSampleFraction* instead: Number of randomly picked training events used in randomised (and bagged) trees
NNodesMax	0	deprecated: Use MaxDepth instead to limit the tree size

Table 2: Full list of configuration options and their default values (BDT)

A.3. Configuration options for MVA method: MLP

Note: Table taken from [9], p. 98-99.

Option	Default	Description
NCycles	500	Number of training cycles
HiddenLayers	N,N-1	Specification of hidden layer architecture
NeuronType	sigmoid	Neuron activation function type
RandomSeed	1	Random seed for initial synapse weights (0 means unique seed for each run; default value '1')
EstimatorType	MSE	MSE (Mean Square Estimator) for Gaussian Likelihood or CE (Cross- Entropy) for Bernoulli Likelihood, linear, sigmoid, tanh, radial
NeuronInputType	sum	Neuron input function type (sum, sqsum, abssum)
TrainingMethod	BP	Train with Back-Propagation (BP), BFGS Algorithm (BFGS), or Genetic Algorithm (GA - slower and worse)
LearningRate	0.02	ANN learning rate parameter
DecayRate	0.01	Decay rate for learning parameter
TestRate	10	Test for overtraining performed at each #th epochs
EpochMonitoring	False	Provide epoch-wise monitoring plots according to TestRate (caution: causes big ROOT output file!)

Sampling	1	Only 'Sampling' (randomly selected) events are trained each epoch
SamplingEpoch	1	Sampling is used for the first 'SamplingEpoch' epochs, afterwards, all events are taken for training
SamplingImportance	1	The sampling weights of events in epochs which successful (worse estimator than before) are multiplied with SamplingImportance, else they are divided.
SamplingTraining	True	The training sample is sampled
SamplingTesting	False	The testing sample is sampled
ResetStep	50	How often BFGS should reset history
Tau	3	LineSearch size step
BPMode	sequential	Back-propagation learning mode: sequential or batch
BatchSize	-1	Batch size: number of events/batch, only set if in Batch Mode, -1 for BatchSize=number_of_events
ConvergenceImprove	1e-30	Minimum improvement which counts as improvement (<0 means automatic convergence check is turned off)
ConvergenceTests	-1	Number of steps (without improvement) required for convergence (<0 means automatic convergence check is turned off)
UseRegulator	False	Use regulator to avoid over-training
UpdateLimit	10000	Maximum times of regulator update
CalculateErrors	False	Calculates inverse Hessian matrix at the end of the training to be able to calculate the uncertainties of an MVA value
WeightRange	1	Take the events for the estimator calculations from small deviations from the desired value to large deviations only over the weight range

Table 3: Full list of configuration options and their default values (MLP)

B. References

[1] J. Erdmann. *Measurement of the inclusive ttgamma cross section at sqrt(s) = 7 TeV with the ATLAS detector*, 2012, 1206.5696.

[2] CMS-Collaboration. *Measurement of the inclusive and differential $\mathrm{t\bar{t}}\gamma$ cross sections in the single-lepton channel and EFT interpretation at $\sqrt{s}$ = 13 TeV*, 2021, 2107.01508.

[3] CMS-Collaboration, S. Chatrchyan, G. Hmayakyan, V. Khachatryan, A. M. Sirunyan, W. Adam, T. Bauer, T. Bergauer, H. Bergauer, M. Dragicevic, and et al. *The CMS experiment at the CERN LHC*. Journal of Instrumentation, 3(08):S08004–S08004, 08 2008. doi:10.1088/1748-0221/3/08/s08004.

[4] *Detector | CMS Experiment*. `https://cms.cern/detector`. Accessed: 2021-10-01.

[5] M. Aaboud, G. Aad, B. Abbott, D. C. Abbott, O. Abdinov, B. Abeloos, D. K. Abhayasinghe, S. H. Abidi, O. S. AbouZeid, and et al. *Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}\gamma$ production in leptonic final states at $\sqrt{s}$ = 13 TeV in ATLAS*. The European Physical Journal C, 79(5), 05 2019. doi:10.1140/epjc/s10052-019-6849-6.

[6] Y. Coadou. *Boosted Decision Trees and Applications*. EPJ Web of Conferences, 55:02004–, 07 2013. doi:10.1051/epjconf/20135502004.

[7] J. Nazzal, I. El-Emary, and S. Najim. *Multilayer Perceptron Neural Network (MLPs) For Analyzing the Properties of Jordan Oil Shale*. World Applied Sciences Journal, 5, 01 2008.

[8] P. Marius, V. Balas, L. Perescu-Popescu, and N. Mastorakis. *Multilayer perceptron and neural networks*. WSEAS Transactions on Circuits and Systems, 8, 07 2009.

[9] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, and et al. *TMVA - Toolkit for Multivariate Data Analysis*, 2009, physics/0703039v5.

C. List of Figures

1. Detector system of CMS, sketch taken from [4] . . . 4
2. Proton Proton collision process . . . 5
3. Distribution of p_T photon . . . 6
4. Distribution of p_T photon (log) . . . 6
5. pseudorapidity of first b-jet and result of neural network to identify b-jets suggest no discriminatory power . . . 8
6. pseudorapidity and transverse momentum of lepton suggest no discriminatory power . . . 8
7. number of jets and b-jets show slightly different behaviour between signal and background . . . 9
8. pseudorapidity and $\Delta\phi$ of the first jet show slightly different behaviour between signal and background . . . 9
9. transverse momentum of first jet and distance between lepton and jet show slightly different behaviour between signal and background . . . 9
10. transverse momentum and $\Delta\phi$ of missing transverse momentum show slightly different behaviour between signal and background . . . 10
11. transverse momentum of W-boson and maximum mass of three jets show slightly different behaviour between signal and background . . . 10
12. transverse momentum of all jets show slightly different behaviour between signal and background . . . 10
13. currently used variable to differentiate between signal and background events suggests good distinction . . . 11
14. transverse momentum and $\Delta\phi$ of photon indicate good discriminatory power 11
15. pseudorapidity of photon and $\Delta\phi$ between lepton and photon indicate good discriminatory power . . . 12
16. distance between photon and jet and between photon and lepton indicate good discriminatory power . . . 12
17. Outline of Boosted Decision Trees . . . 13
18. log-sigmoid function . . . 13
19. Outline of Multilayer Perceptron . . . 14
20. ROC-curve examples . . . 16
21. BDT-, MLP-distribution and ROC curve for variables from selection 1 with configuration 1 . . . 17
22. BDT-, MLP-distribution and ROC curve for variables from selection 1 with configuration 2 . . . 18
23. BDT-, MLP-distribution and ROC curve for variables from selection 1 with configuration 3 . . . 19
24. BDT-, MLP-distribution and ROC curve for variables from selection 2 with configuration 1 . . . 20
25. BDT-, MLP-distribution and ROC curve for variables from selection 2 with configuration 2 . . . 21

26. BDT-, MLP-distribution and ROC curve for variables from selection 2 with configuration 3 . . . 22
27. BDT-, MLP-distribution and ROC curve for variables from selection 3 with configuration 1 . . . 23
28. BDT-, MLP-distribution and ROC curve for variables from selection 3 with configuration 2 . . . 24
29. BDT-, MLP-distribution and ROC curve for variables from selection 3 with configuration 3 . . . 25
30. pseudorapidity and $\Delta\phi$ of first b-jet suggest no discriminatory power . . . 26
31. transverse momentum of first b-jet and pseudorapidity of second b-jet suggest no discriminatory power . . . 26
32. $\Delta\phi$ and transverse momentum of second b-jet suggest no discriminatory power . . . 27
33. results of neural network to identify first and second b-jet suggest no discriminatory power . . . 27
34. proportion of energy coming from charged particles detected in HCAL and detected in ECAL suggest no discriminatory power . . . 27
35. proportion of energy coming from neutral particles detected in HCAL and detected in ECAL suggest no discriminatory power . . . 28
36. pseudorapidity and $\Delta\phi$ of second jet suggest no discriminatory power . . . 28
37. transverse momentum of second jet and $\Delta\phi$ of lepton suggest no discriminatory power . . . 28
38. pseudorapidity of lepton and relative isolation of lepton to all particles suggest no discriminatory power . . . 29
39. relative isolation of lepton to charged and neutral particles suggest no discriminatory power . . . 29
40. transverse momentum of lepton and number of electrons suggest no discriminatory power . . . 29
41. number of photons with high p_T and number of jets suggest no discriminatory power . . . 30
42. number of leptons and muons suggest no discriminatory power . . . 30
43. number of photons suggests no discriminatory power . . . 30

D. List of Tables

1. Configuration combinations used in MVAs 15
2. Full list of configuration options and their default values (BDT) 33
3. Full list of configuration options and their default values (MLP) 34