Different Machine Learning Approaches and Techniques in the Scope of a $t\bar{t}\gamma$ Cross Section Measurement at $\sqrt{s} = 13\,\text{TeV}$ in ATLAS

Steffen Korn
Supervised by
Arnulf Quadt

II. Physikalisches Institut
Georg-August-Universität Göttingen

13th Annual Helmholtz Alliance Workshop on "Physics at the Terascale" - DESY Hamburg
Computing and Machine Learning
26.11.2019

BMBF-Forschungsschwerpunkt
ATLAS-EXPERIMENT
Physik bei höchsten Energien mit dem ATLAS-Experiment am LHC
FSP 103
ATLAS

Gefördert durch
Bundesministerium für Bildung und Forschung

Content

- Motivation
- Event selection
- Backgrounds
- Tagging prompt photons
- Event level discrimination (ELD)

Previous ATLAS analysis with 2015/2016 data

Eur.Phys.J. C79 (2019) no.5, 382 (2019-05-03)

- Multi-Class approaches
- Difficulties and challenges
- Current status of the analysis using Run II d

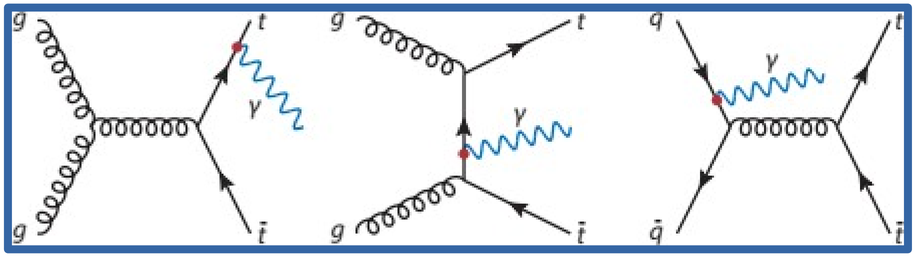
Full Run II analysis with 2015-2018 data (In progress)

- Conclusion

Motivation

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - q\bar{\psi}\gamma^{\mu}A_{\mu}\psi$$

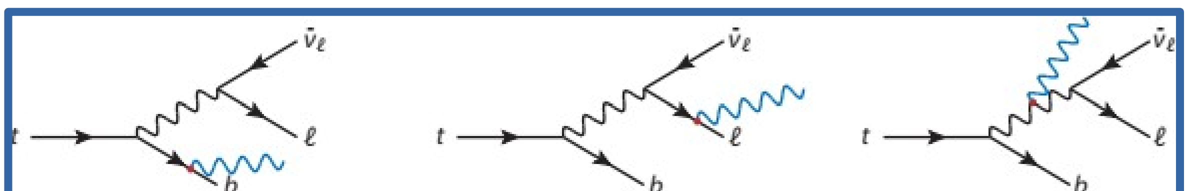
Radiative production ISR



Photon can be radiated from any charged particle in the initial and final state MVA studies on its origin:

II.Physik-UniGö-MSc-2018/04

FSR from b FSR from lepton FSR from W



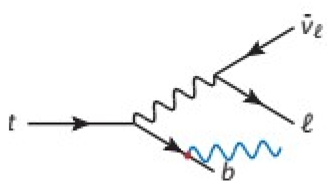
All these processes are treated as signal

- Measurement is probing the coupling between top and photon
- Test of the QED vector structure
- Tensor contributions?

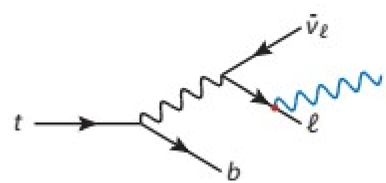
Signal Definition

	e+jets	μ+jets
Jets	≥ 4	
b-Jets	$\geq$ 1 tagged with MV2C10 85 % WP	
Leptons	one e	one μ
E_t^{miss}	-	-
$\Delta R(l, \gamma)$	≥ 1	
$m(l, \gamma)$	$\notin$ [85,95] GeV	

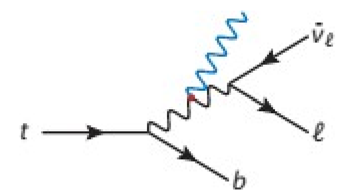
FSR from b



FSR from lepton

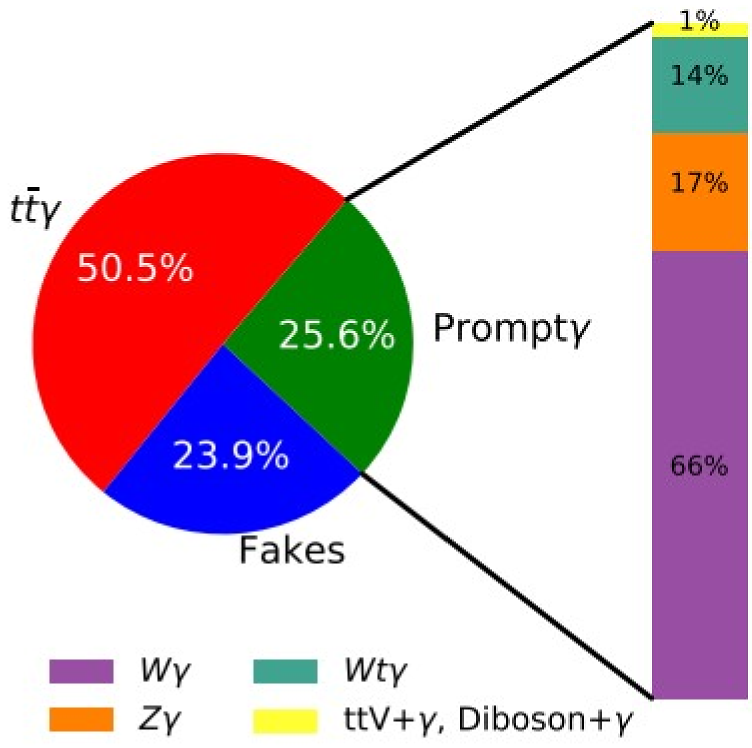


FSR from W



Prompt Photon Background

Prompt photon background composition:

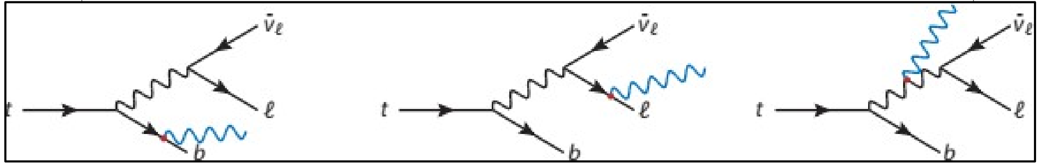


Backgrounds can be separated by using b-tagging information

Standard $t\bar{t}$ Background plus additional γ

- W+photon
- Z+photon
- Single top+photon
- Diboson processes and $t\bar{t}$V processes

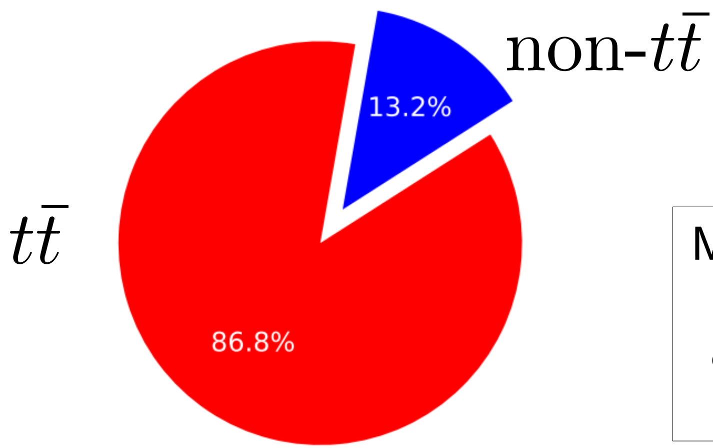
Using b-tagged jets:
- Invariant masses (e.g. between lepton and b-tagged jet)
- B-tagging scores
- → Powerfull discriminants



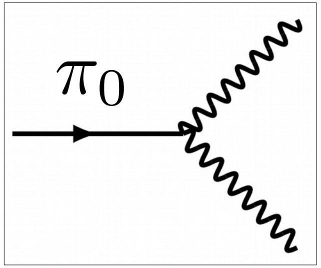
Fake Photon Background

Electronic fake photons (e-fakes)

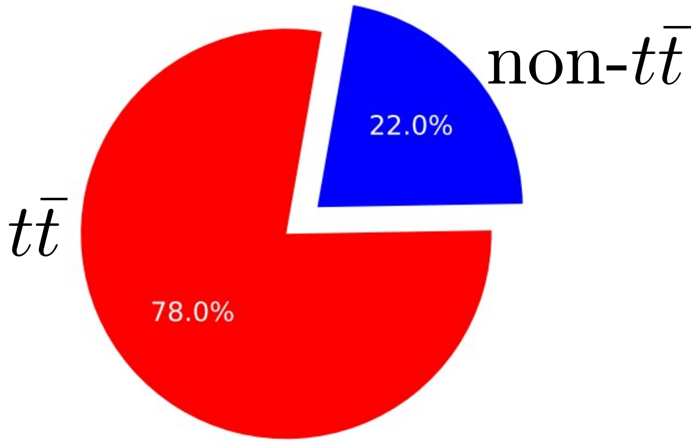
- Electrons misidentified as photons
- Originating from dileptonic channels
- Almost exlusively in the e+jets channel



Hadronic fake photons (h-fakes)



- Jets faking photons and photons originating from jets
- Originating from dileptonic channels
- Contribution roughly equal in both lepton+jets channels

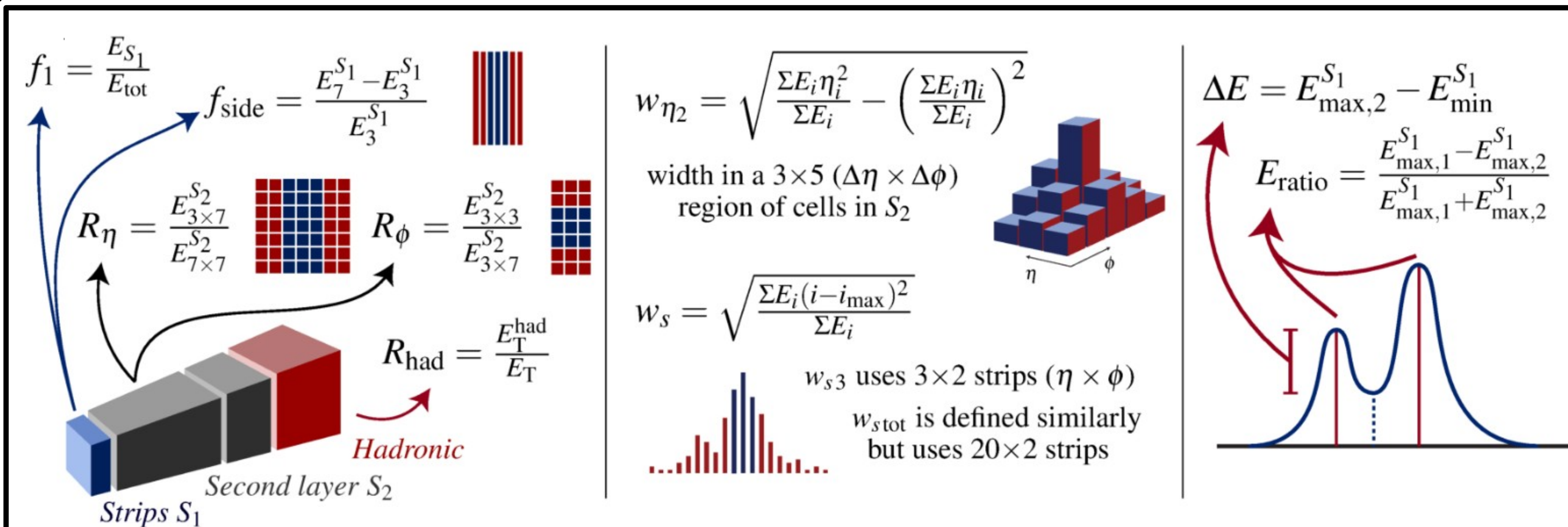


Most fake events (about 80%) are very similar to signal events

Prompt Photon Tagger

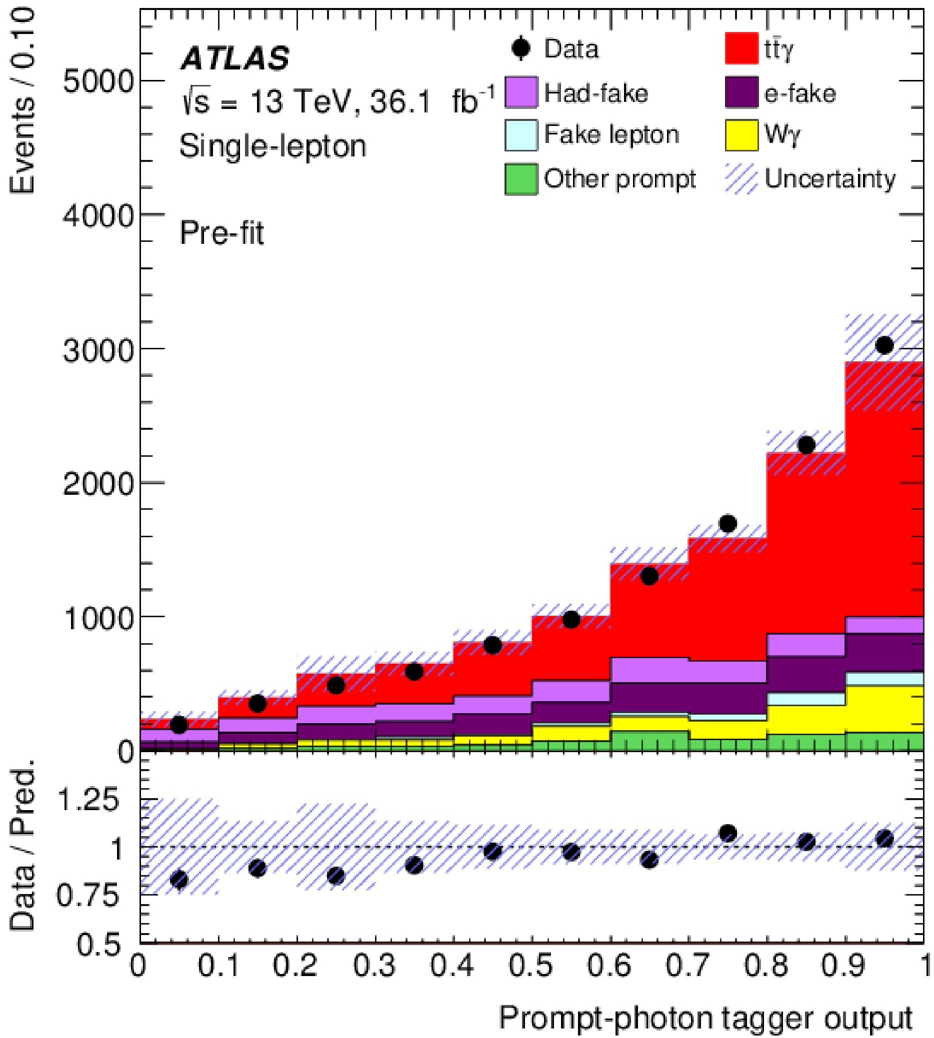
- Significant contribution of hadronic fakes in SR
- Use information from EM calorimeter to distinguish between fakes and prompt photons

Use Variables used for photon identification as inputs for a binary NN

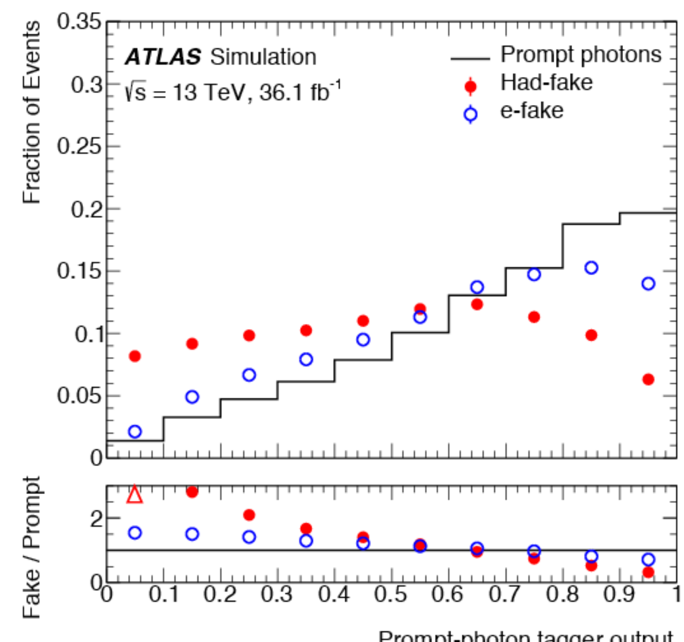


Energy Ratios **Widths** **Shower shapes**

Prompt Photon Tagger



- Binary NN to separate fake photons and prompt photons
- Using shower shape information only
- Analysis independent tool



Eur.Phys.J. C79 (2019) no. 5, 382(2019-05-03)

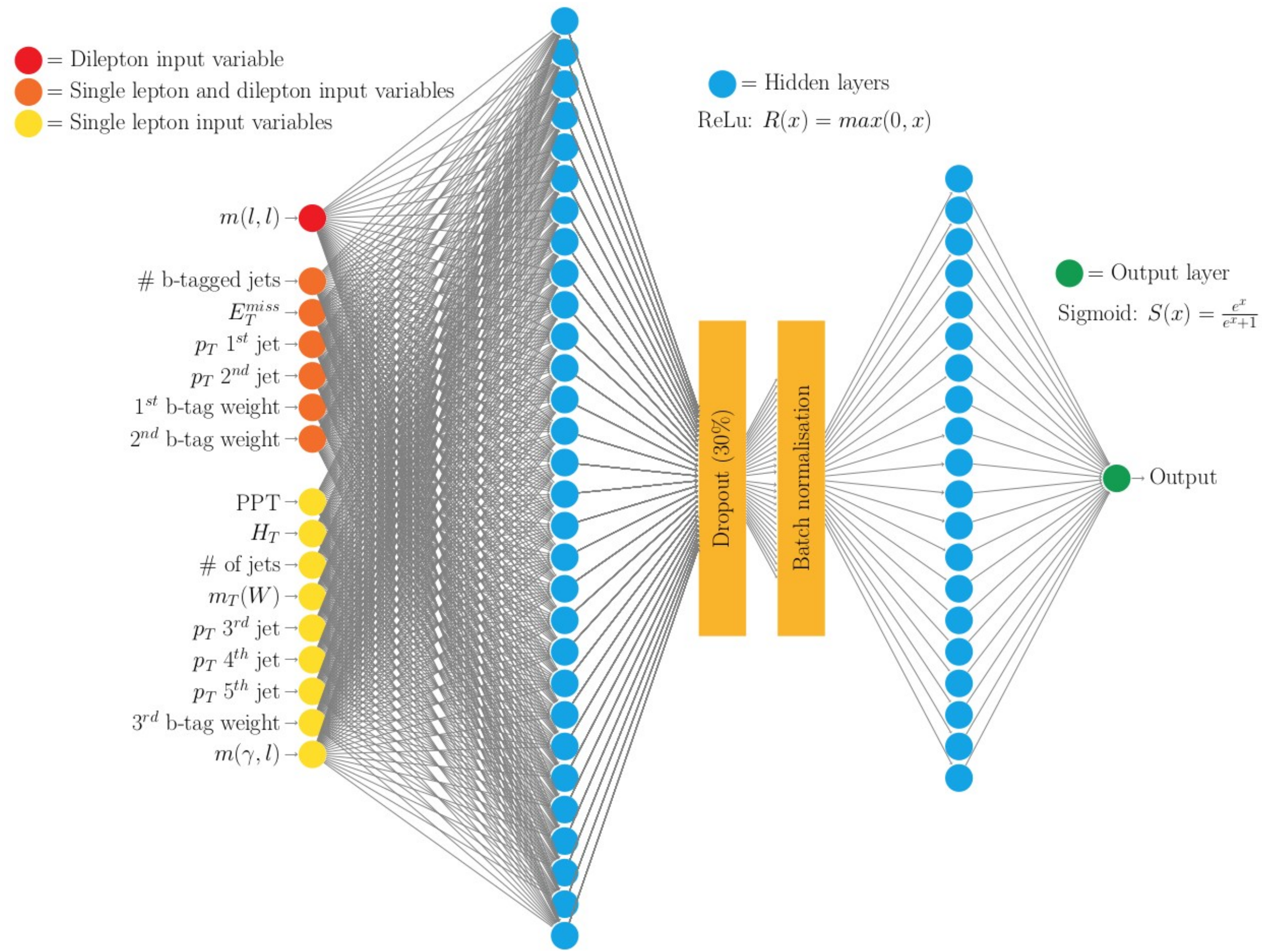
- Event level discriminator (**ELD**)

  II.Physik-UniGö-Diss-2018/01
- Signal vs. total background
- Binary NN
- Use **PPT** as input

  II.Physik-UniGö-MSc-2017/07

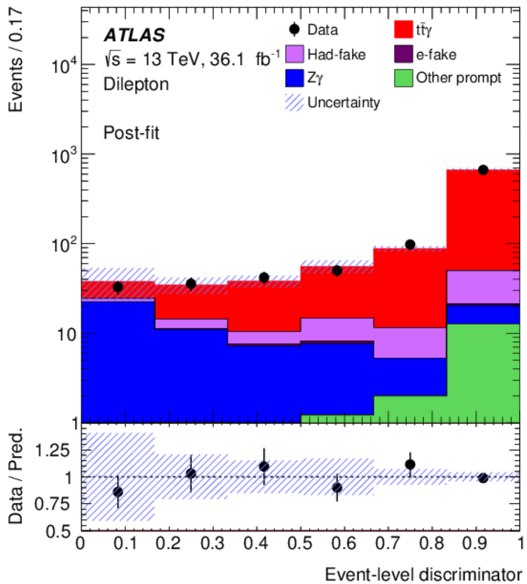
  This is used in:

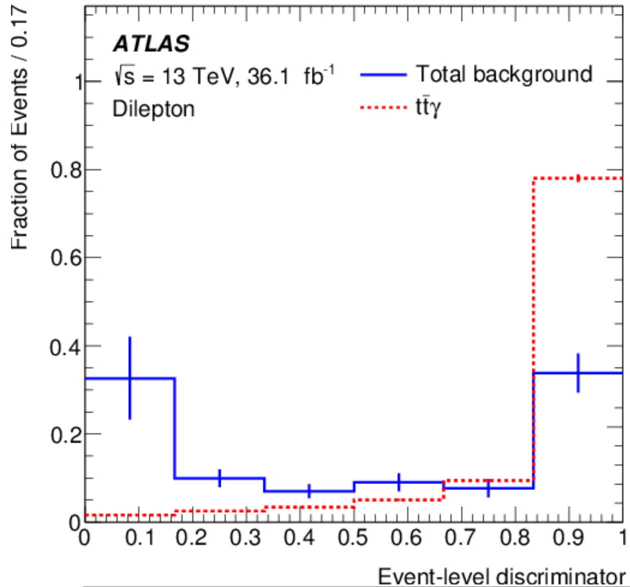
  Eur.Phys.J. C79 (2019) no. 5, 382(2019-05-03)

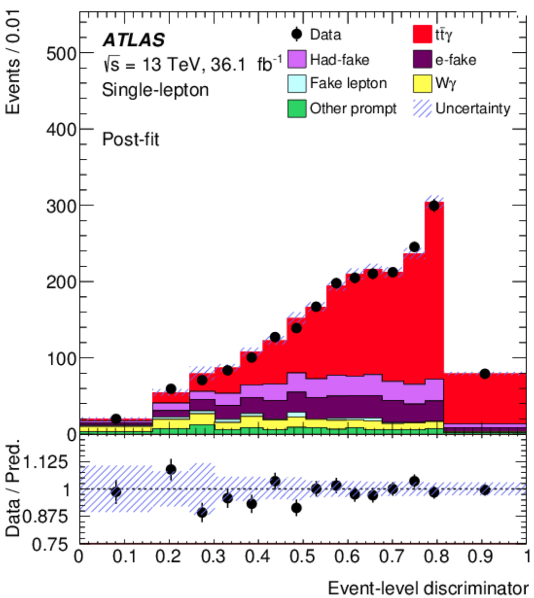


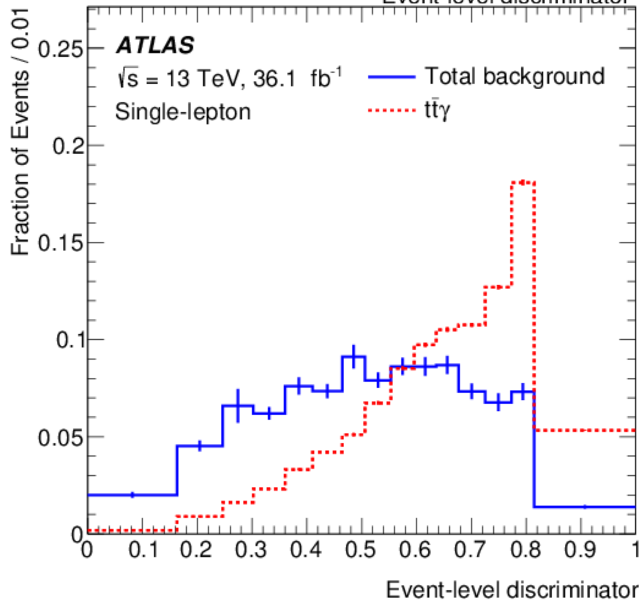
ELD 2

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

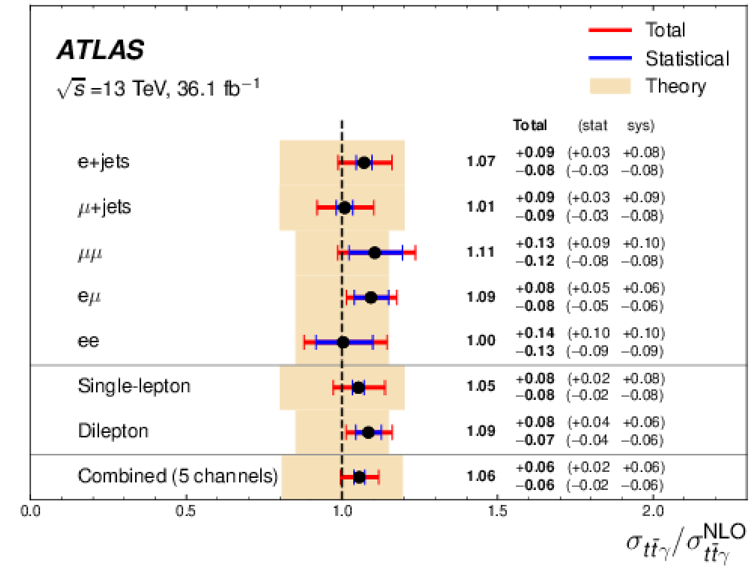








- Fiducial cross section measurement
- Using single-lepton and dilepton channels
- 2 NN (***PPT*** and ***ELD***)



$$\text{Single-lepton: } 521 \pm 9(\text{stat.}) \pm 41(\text{sys.})\text{fb}^{-1}$$
$$\text{Dilepton: } 69 \pm 3(\text{stat.}) \pm 4(\text{sys.})\text{fb}^{-1}$$

Eur.Phys.J. C79 (2019) no. 5, 382(2019-05-03)

From Binary to Multi-Class

Previous Analysis

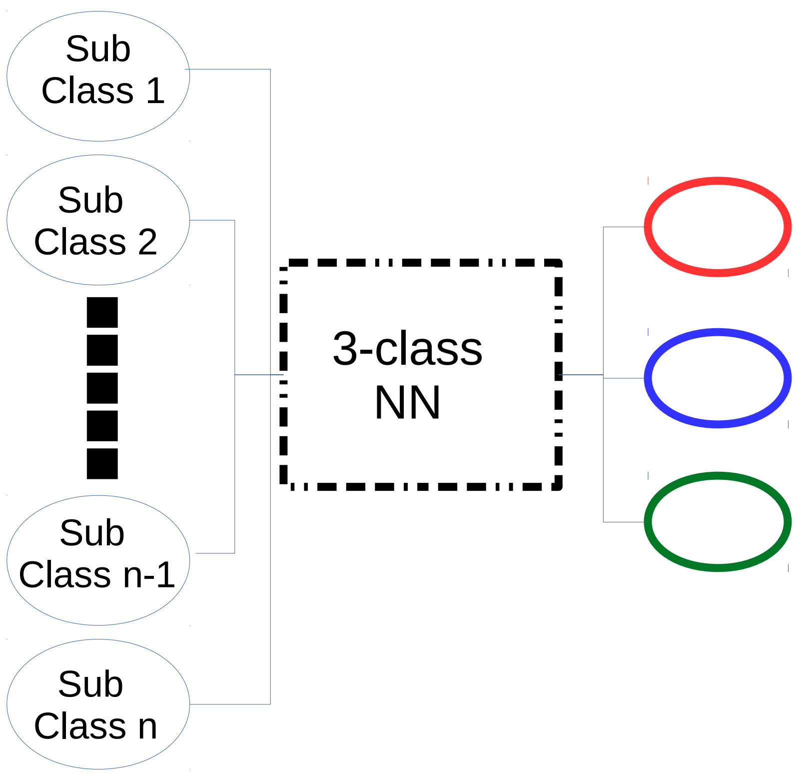
- Binary ELD and PPT
- Data taken in 2015/2016

Now

- Investigate
  Multi-class approaches
- Full Run II dataset

One-vs-All and One-vs-One

General Structure is based on feed-forward NN built with Keras and Tensorflow



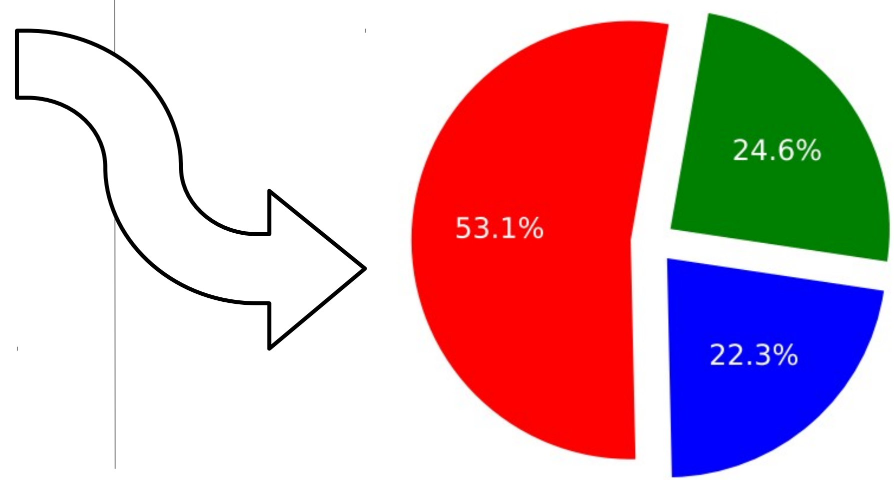
Binary-Sub-Classifier

One-vs-All

- Separating **one** class from **all** remaining classes

One-vs-One

- Separating **one** class from **signal** class

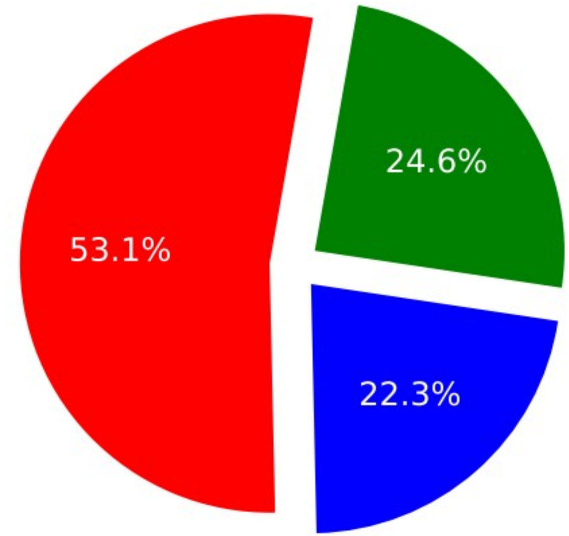


Handling Imbalanced Classes

Taking expected numbers of events into account during training

$$t\bar{t}\gamma \,:\, \text{Fakes} \,:\, \text{Prompt}\gamma \neq 1:1:1$$

Leads to



Rescaling of weights for training is necessary

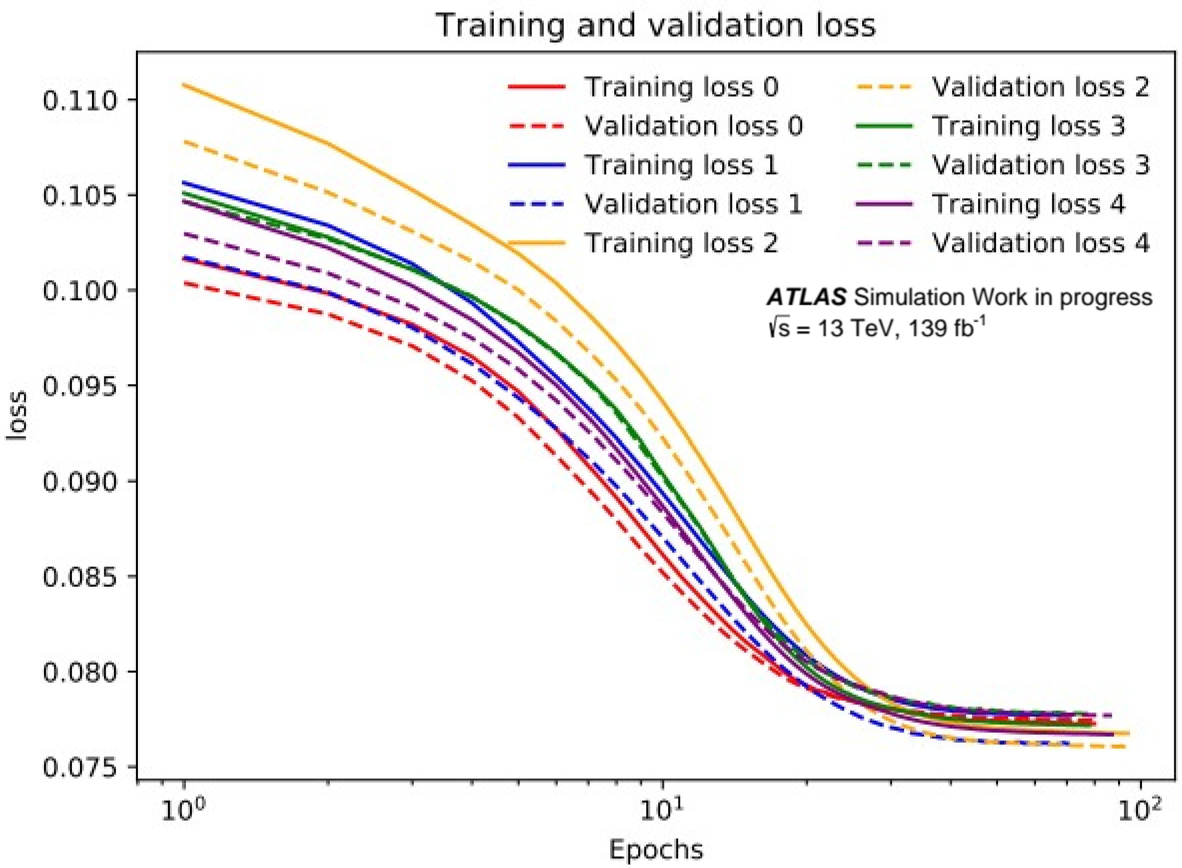
Binary sub-classifier:

$$\sum W_{\text{train,Signal}} : \sum W_{\text{train,Background}} = 1:1$$

Multi-class:

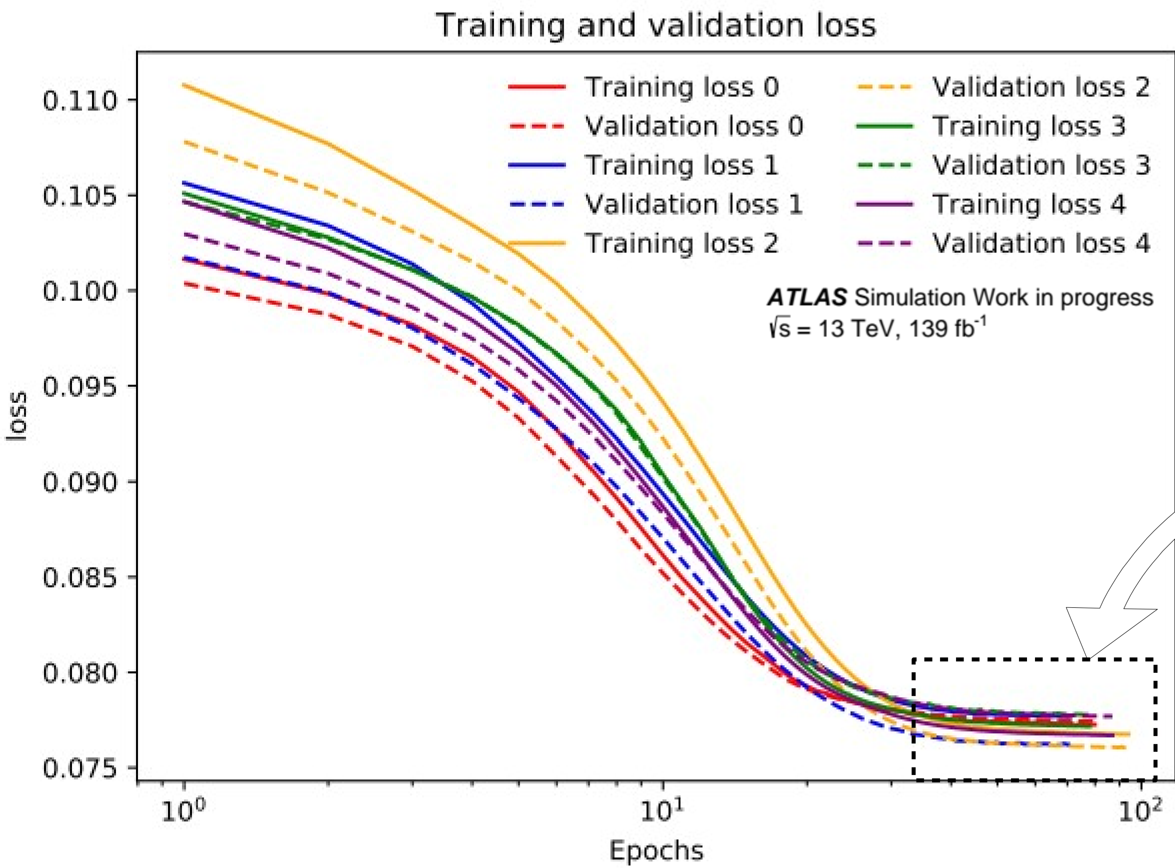
$$\sum W_{\text{train},1} : \sum W_{\text{train},2} : \sum W_{\text{train},3} = 1:1:1$$

Training Stacked Models



- Nadam
  (Adam Optimizer with
  Nesterov momentum)
- 5-fold cross validation
- Binary/Categorical
  cross entropy
  as loss function
- Using early stopping procedure to prevent overfitting
- Using dropout to reduce overfitting further

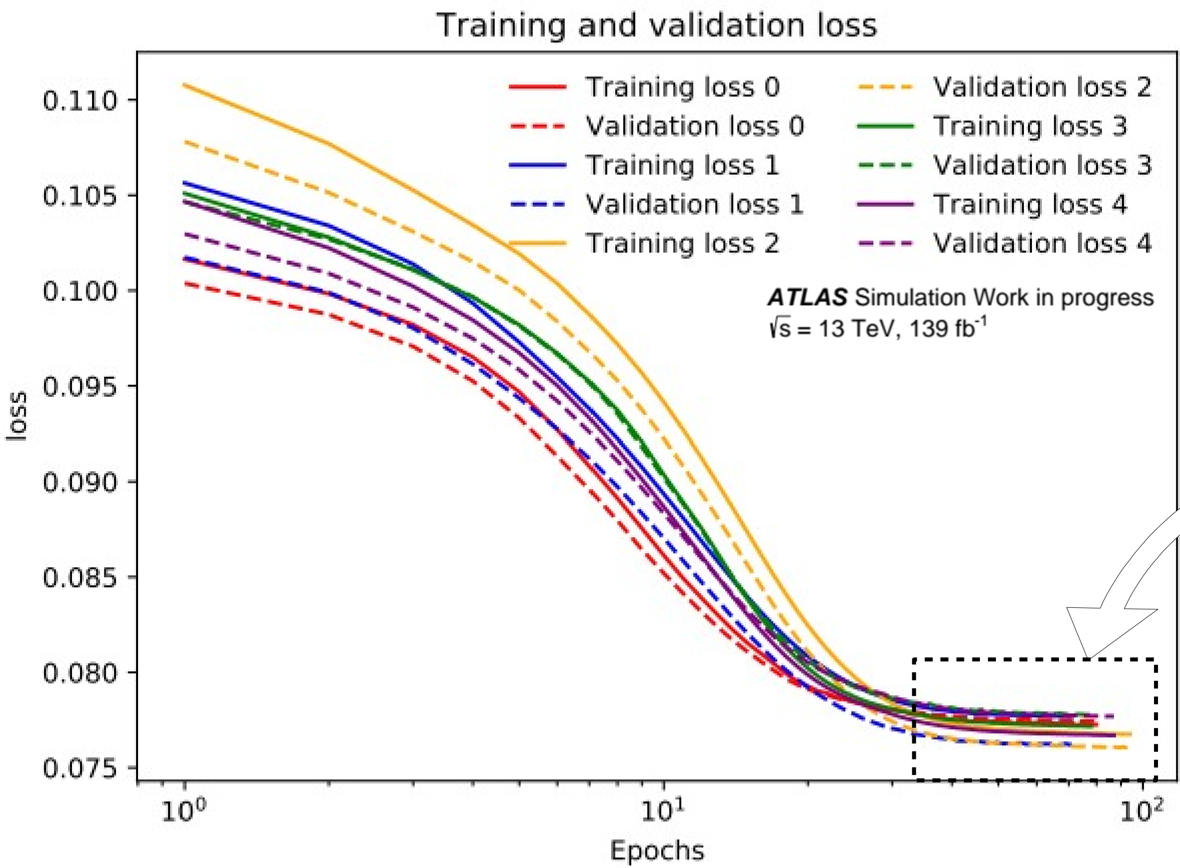
Training Stacked Models



- NAdam
- 5-fold cross validation
- Binary/Categorical cross entropy
- Early stopping
- Dropout to reduce overfitting

Validation loss < training loss
Since droped nodes are not used when the training loss is calculated

Training Stacked Models



- NAdam
- 5-fold cross validation
- Binary/Categorical cross entropy
- Early stopping
- Dropout to reduce overfitting

Validation loss < training loss
Since droped nodes are not used when the training loss is calculated

Batch size = 100.000 events
→ better gradient estimation

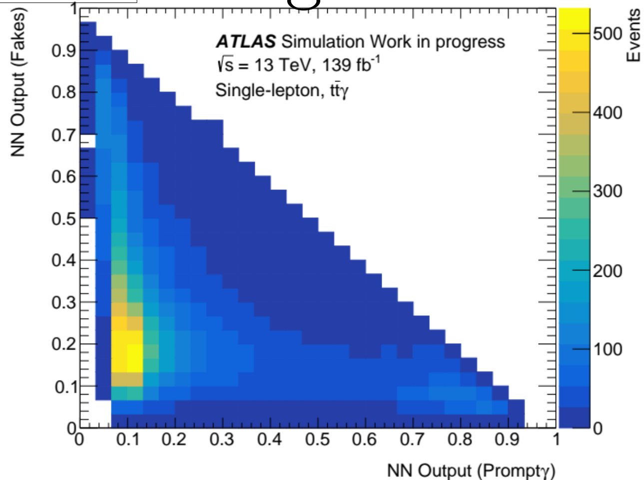
Smooth graphs

Utilising Multidimensionality

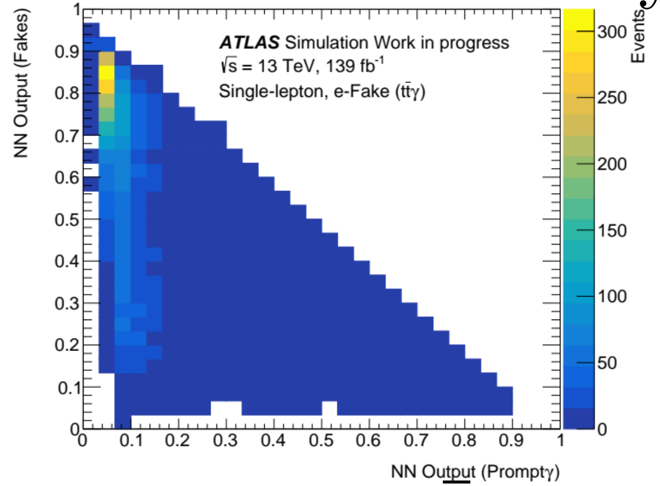
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

One-vs-All

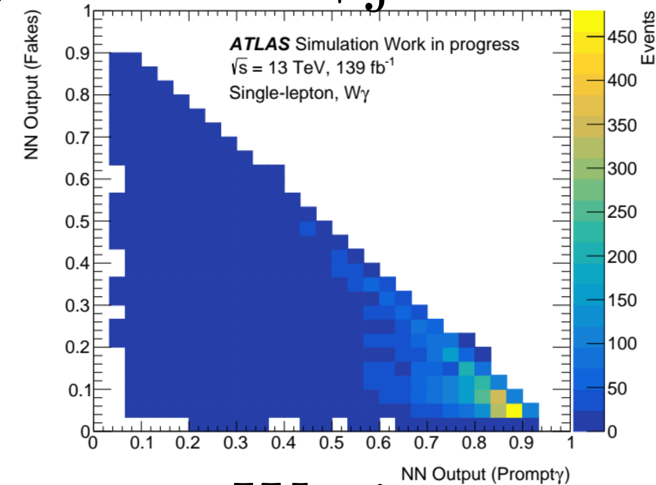
Signal



e-fakes from $t\bar{t}$-decay

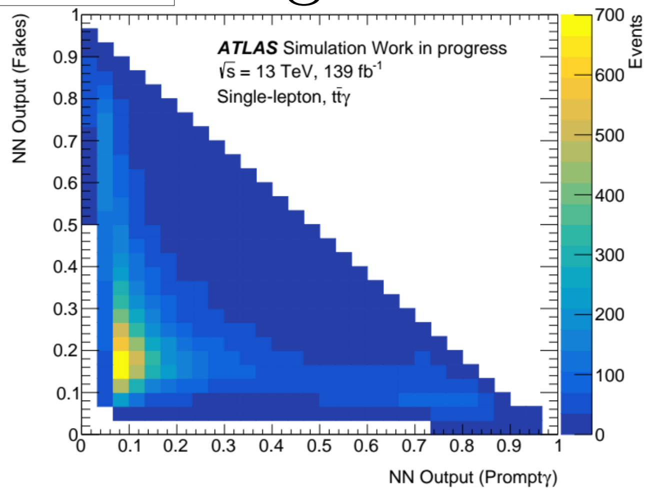


W+jets

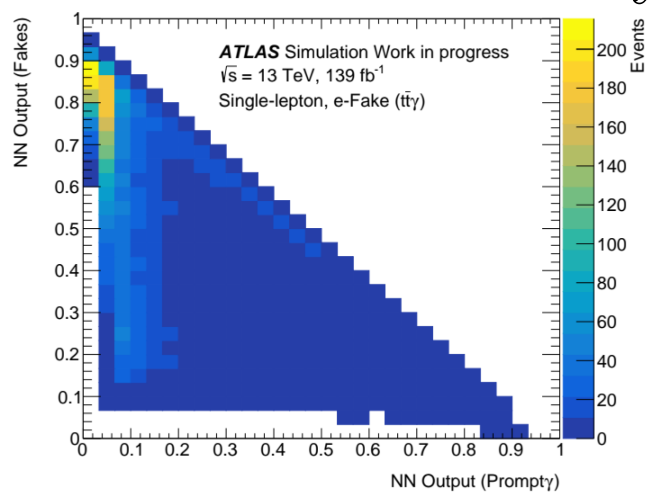


One-vs-One

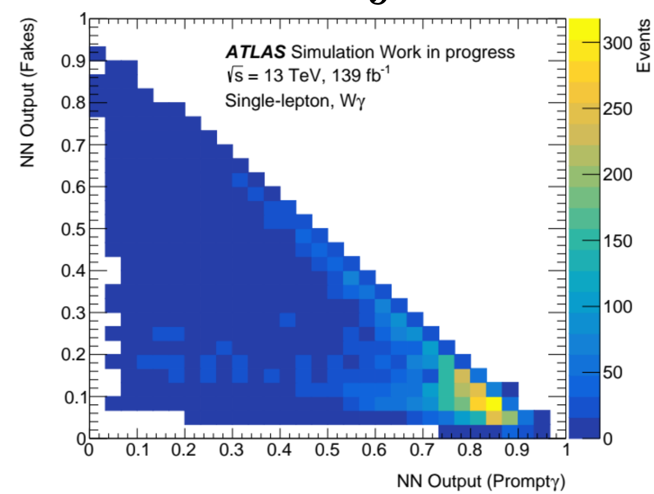
Signal



e-fakes from $t\bar{t}$-decay



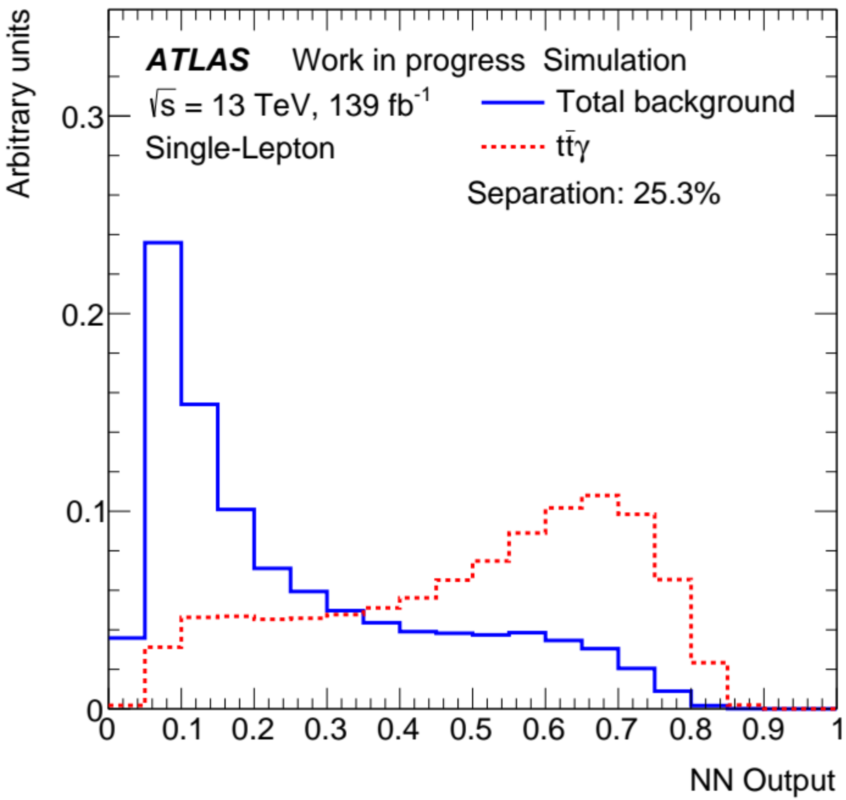
W+jets



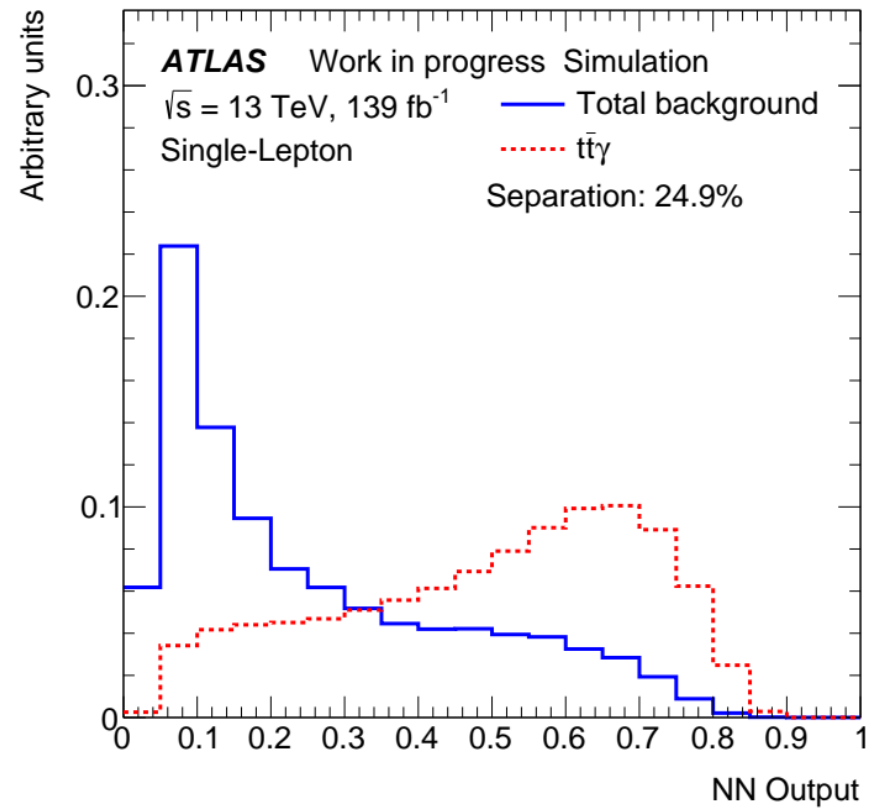
Separating Signal from Background

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

One-vs-One



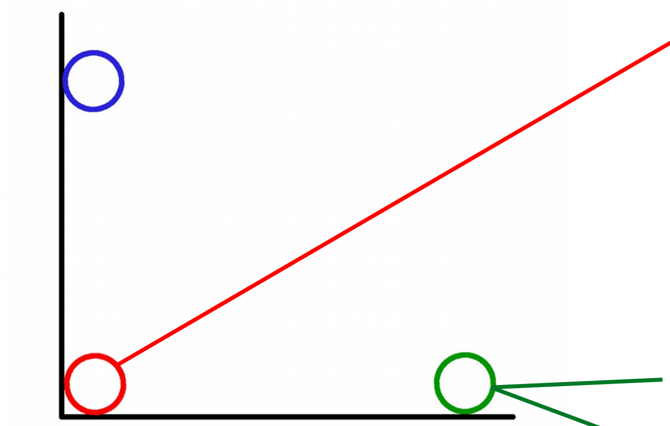
One-vs-All

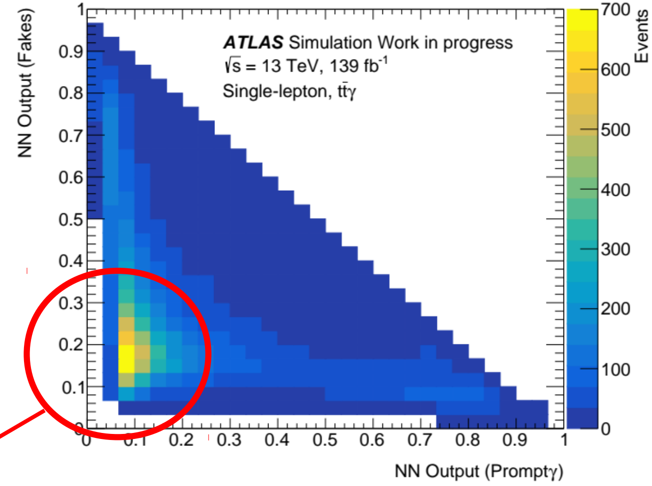


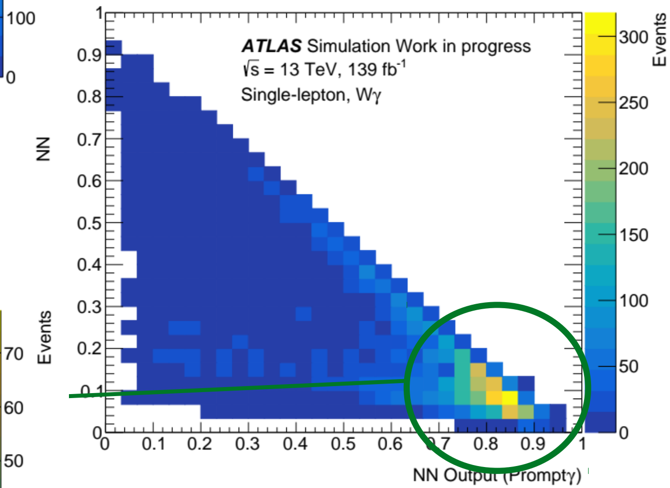
Both approaches yield similar results.

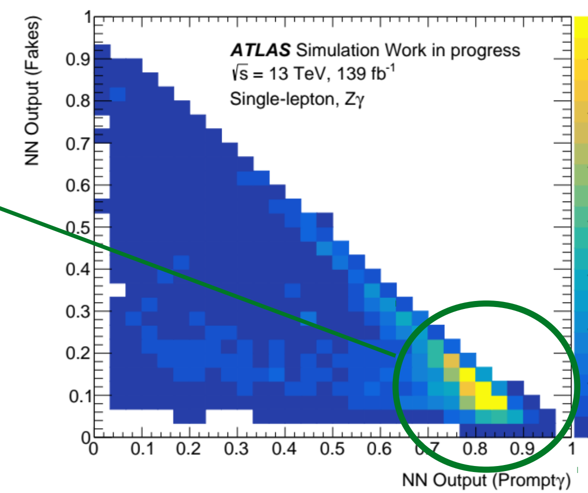
Analysis Impact

Defining control regions by applying cuts on NN output









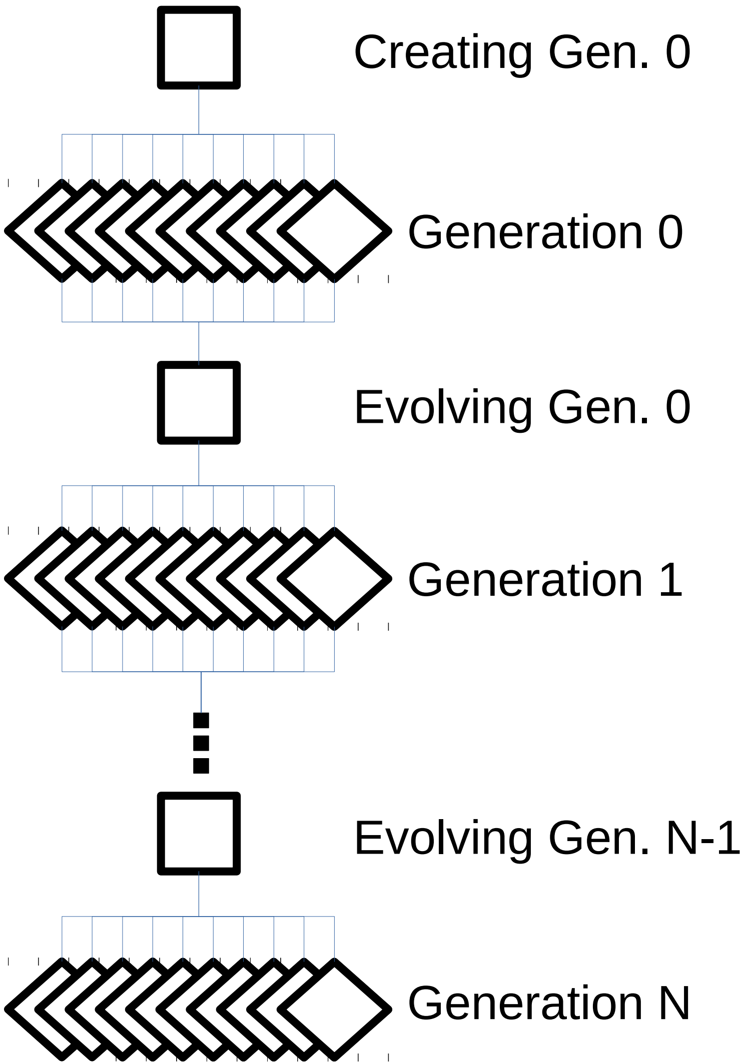
Conclusion

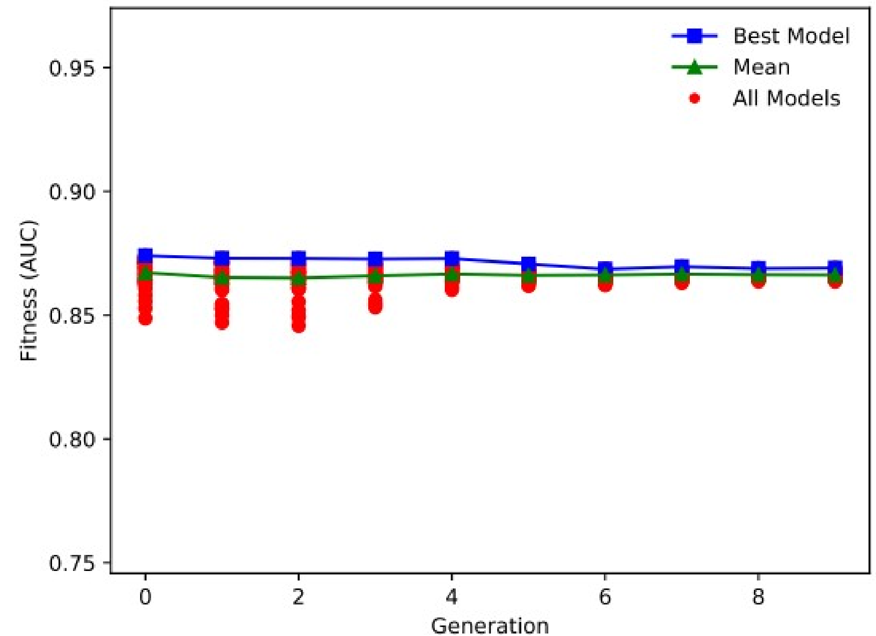
- Different machine learning techniques were presented
- Previous analysis focused on binary approaches
- We are now moving to multi-class approaches
- New multi-class approaches provide the opportunity to define new dedicated control regions and as well as refined signal regions
- These control regions can be used to constrain the prompt photon and fake photon backgrounds

Thank you for your attention!

Backup

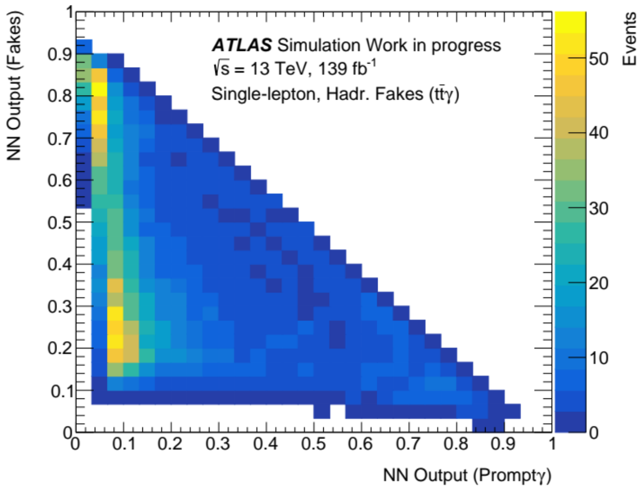
Optimizing Sub-Classifiers

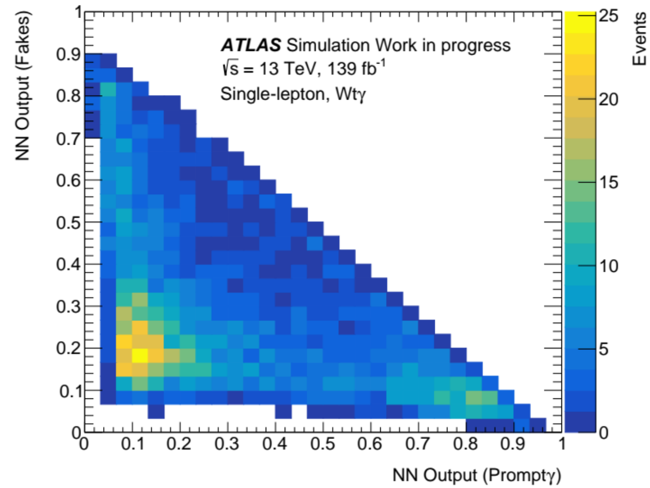


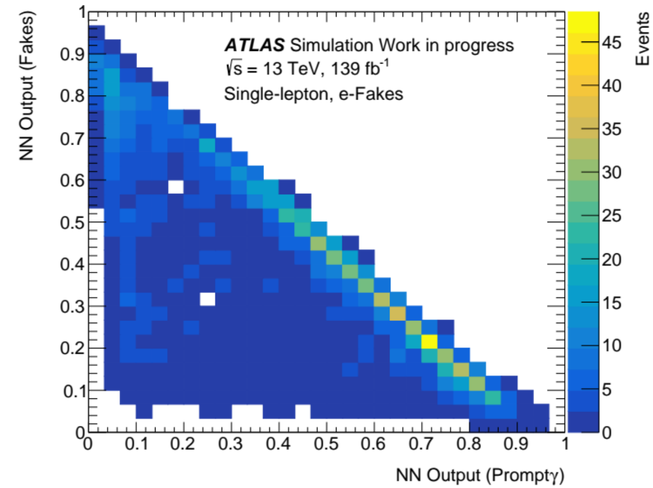


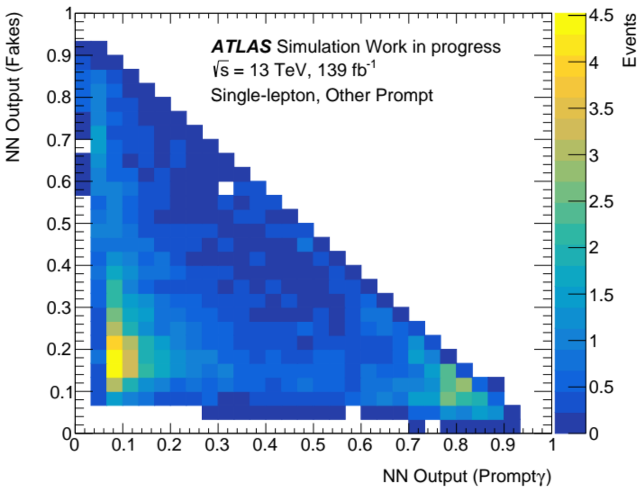
- New generations based on old generations
- Small changes for best model
- Network chosen based on AUC score in training and testing
- Optimizing all sub-classifiers to improve over-all performance
- Training of ~5000 configurations

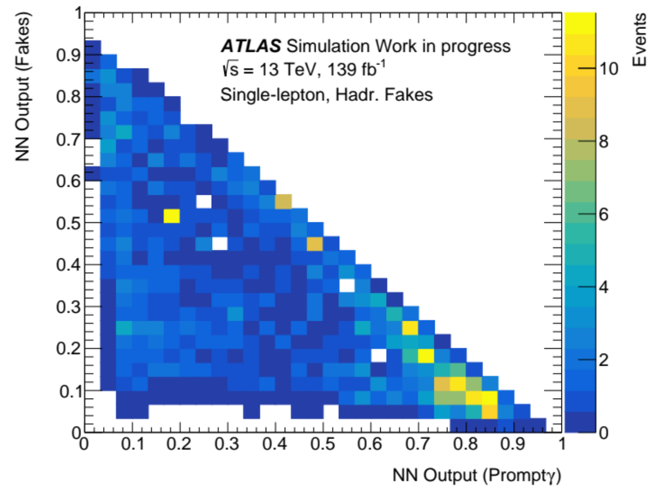
Other 2D Distributions





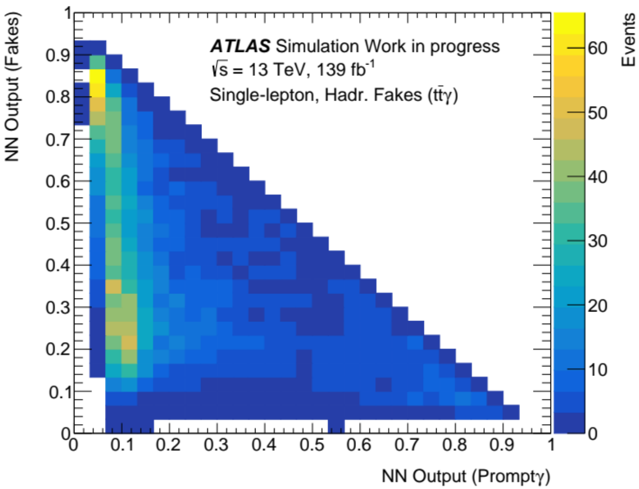


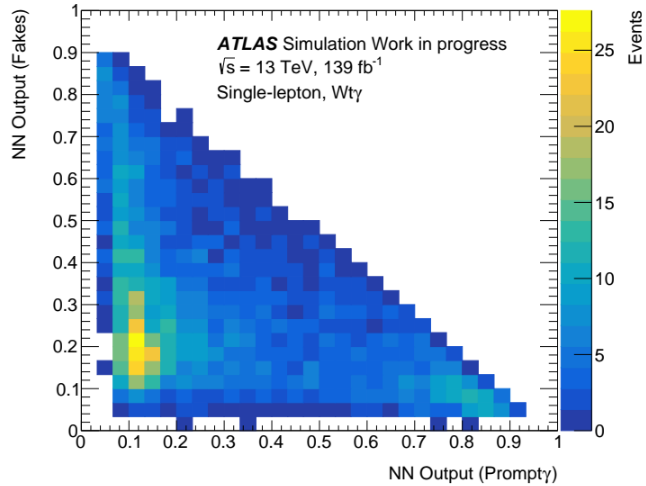


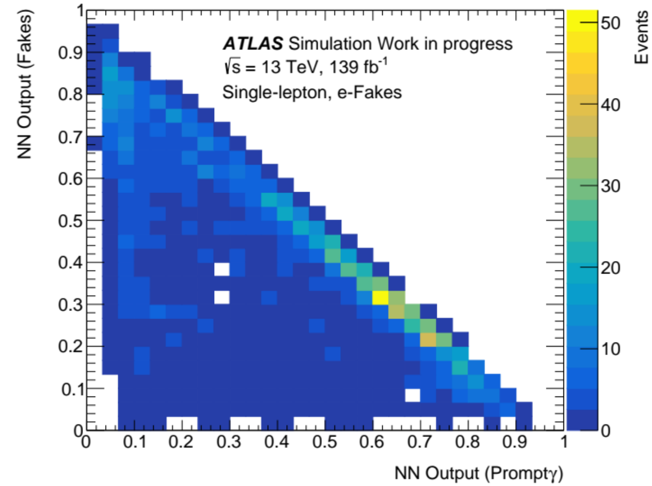


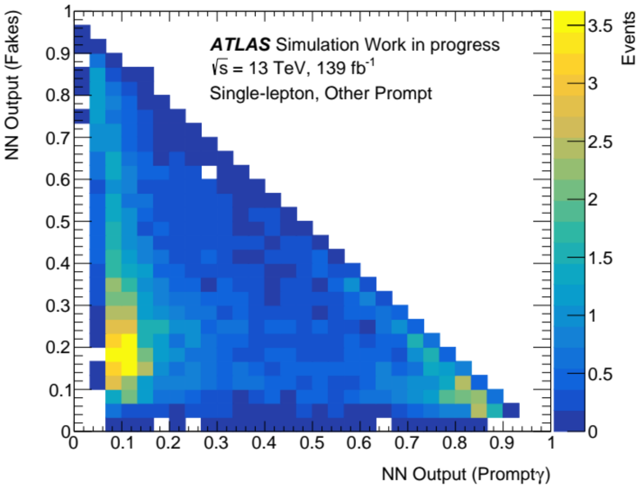
Other 2D distributions
For One-vs-One output

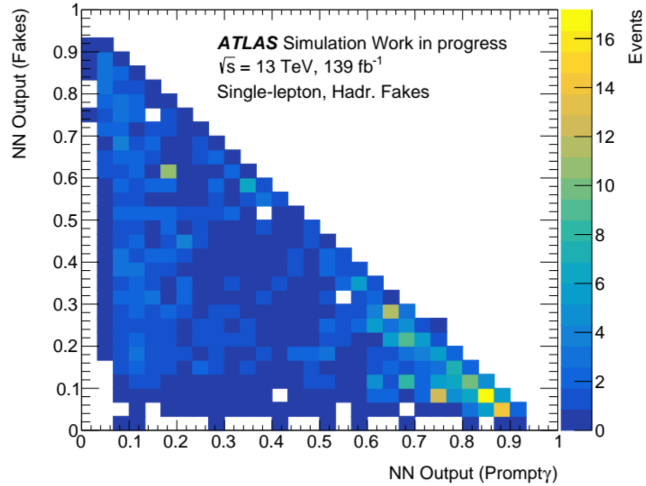
Other 2D Distributions











Other 2D distributions
For One-vs-All output