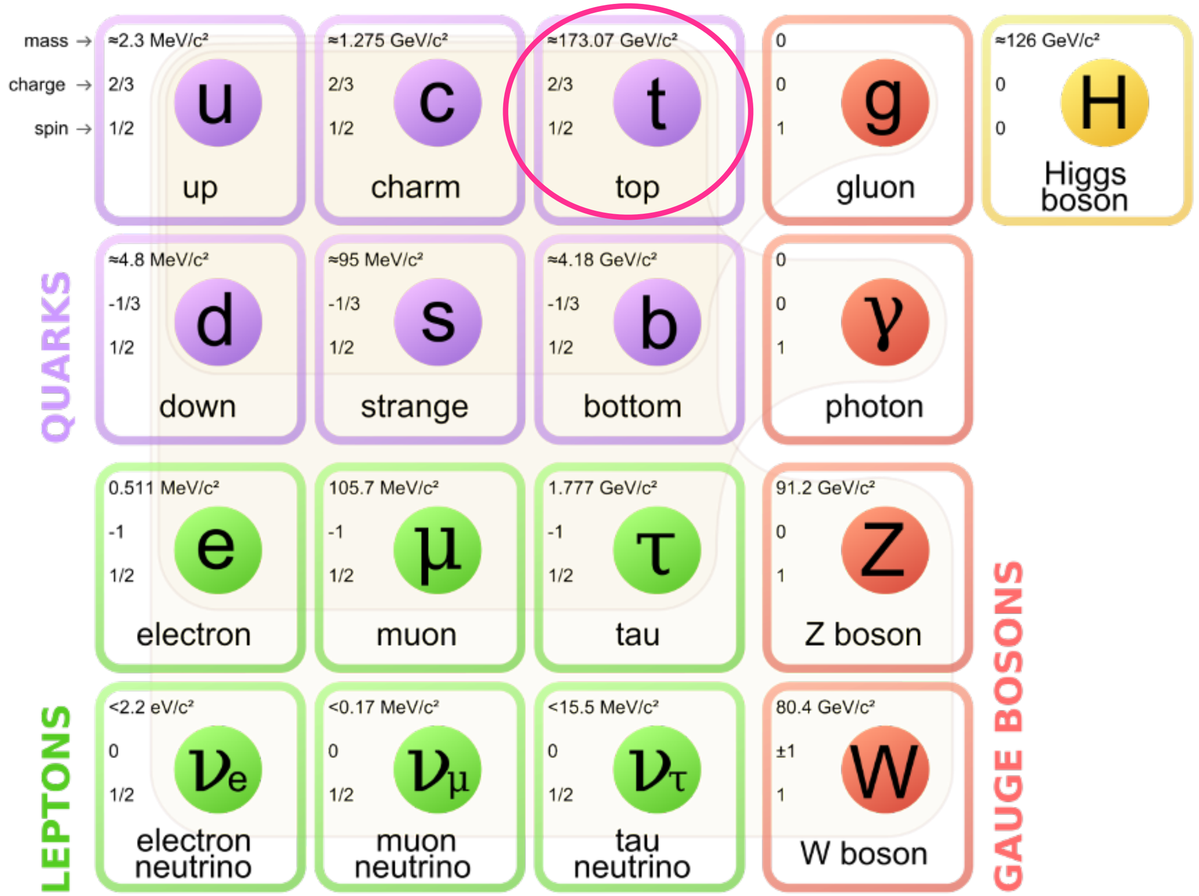
Separating the production of single top quarks in association with a Z boson from background events with Machine Learning techniques at the ATLAS experiment

BSc candidate:
Niccolò Laurora

Supervisor: Lidia Dell'Asta

The Standard Model and the top quark.



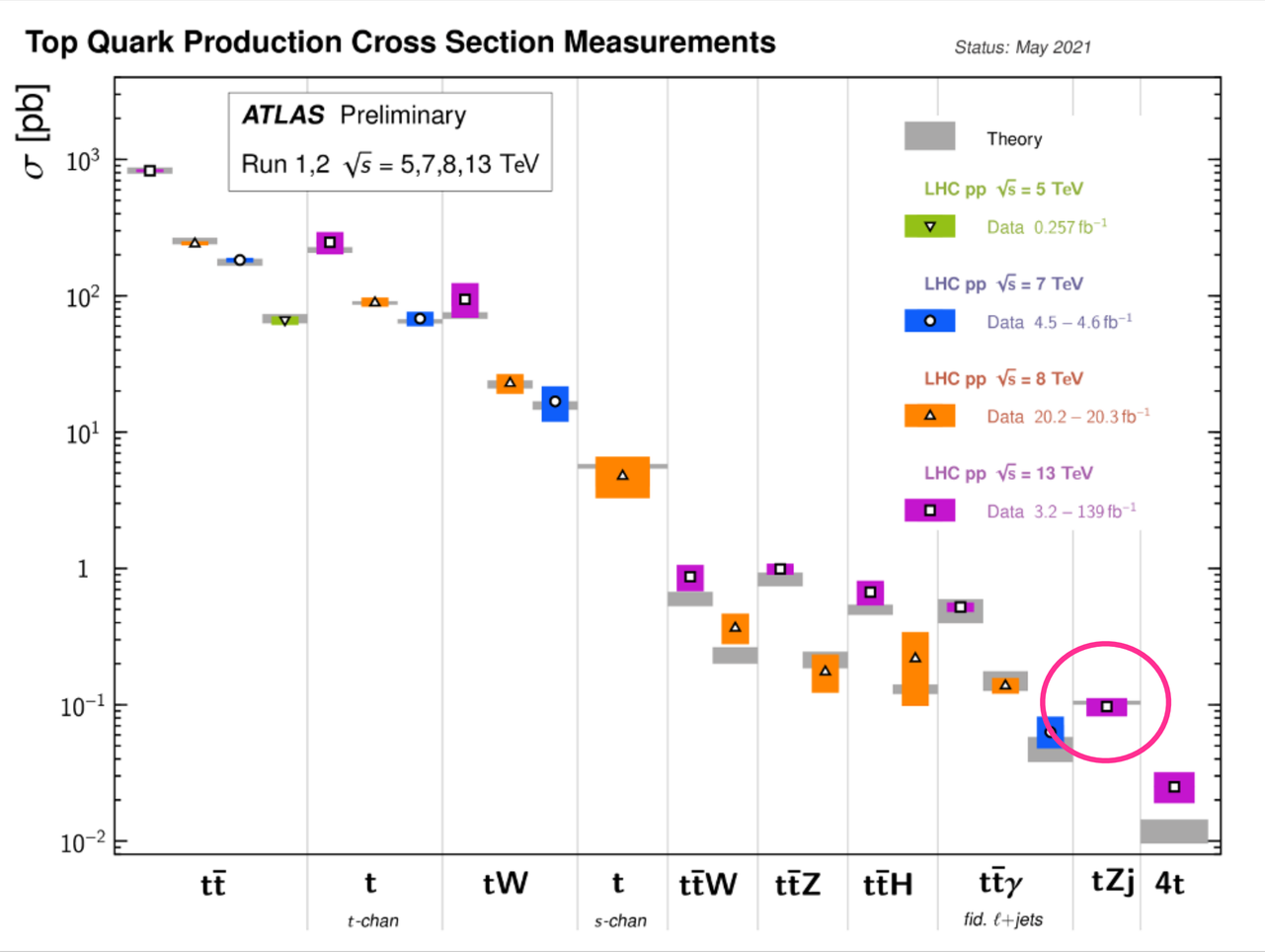
Standard Model:

- Fermions: quarks and leptons;
- Bosons: four vector, one scalar;

Top quark:

- Heaviest particle in the SM (172.69 ± 0.30 GeV);
- Unitary coupling to the Higgs boson;
- No bound states: its lifetime (5×10^{-25} s) is shorter than time needed to form hadrons;
- Decay: $t \rightarrow Wb$ (100 %);

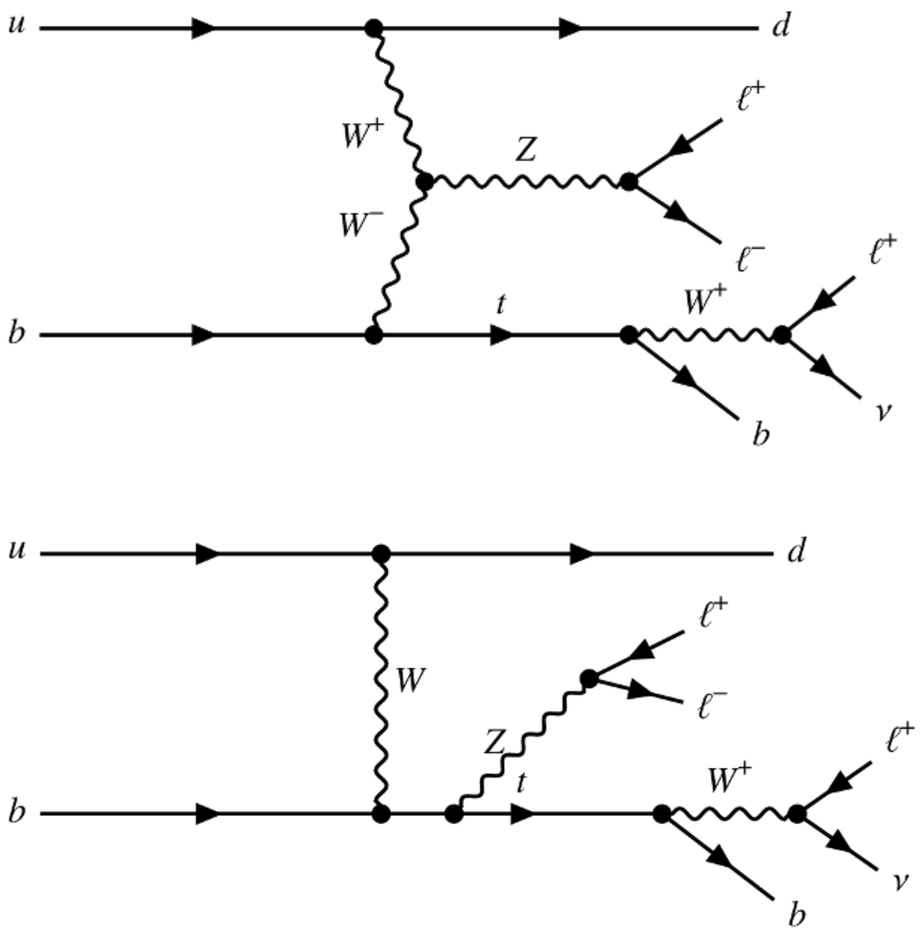
Top quark production at LHC.



Production channels:

- Strong pair production (dominant): σ ≈ 830 pb;
- Electroweak single top production (smaller σ);
- Associated Z boson production (rare): tt̄Z, **tZq**, tWZ;

tZq: single top production in association with a Z boson.



- Electroweak process → Small cross-section (102 fb at √s = 13 TeV, NLO);
- Observed by ATLAS and CMS (97 ± 13 (stat.) ± 7 (syst.) fb with 139 fb^{-1});
- Sensitive to top-Z coupling, background to tHq;
- **Trilepton** final state (BR=2.17 %):
  - 3 leptons;
  - one neutrino;
  - one light jet;
  - one jet from *b*-quark (*b*-jet);

Purpose of my work.

Explore a more refined multivariate analysis based on a multi-class classifier.

- The **classifier** problem: which classifier should the analysis use? Binary or Multi-Class?
- The **selection** problem: which selection should the analysis use? Default or Loose?

Event Selection and Signal Region.

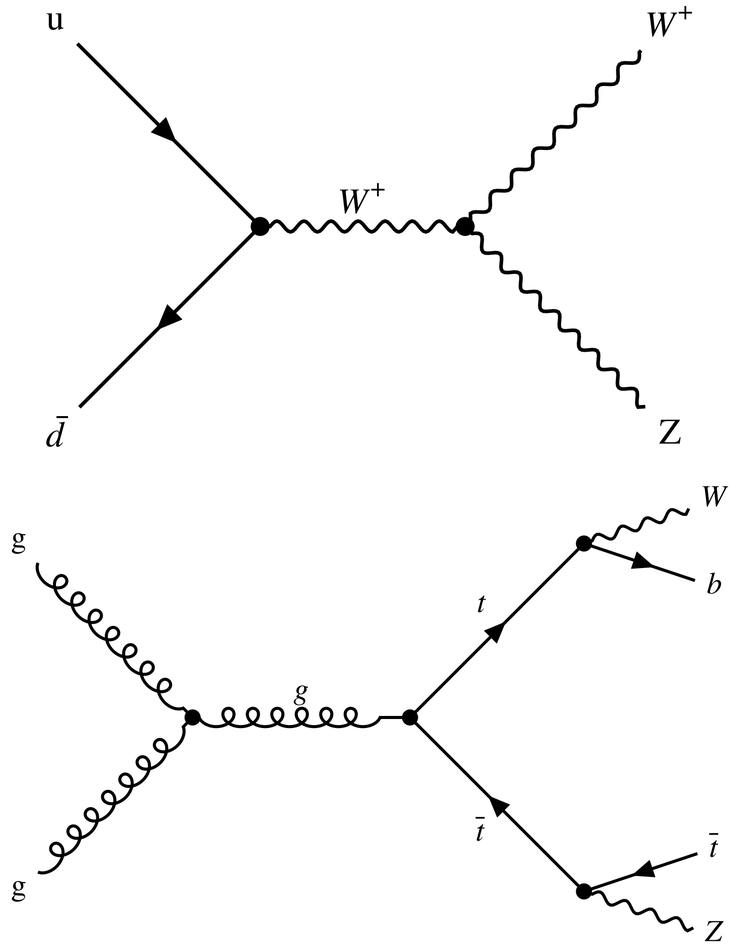
Three signal regions: **2j1b**, **3j1b**, **nj1b**.

2j1b	3j1b	nj1b
1 OSSF pair $\lvert m_{\ell\ell} - m_Z \rvert < 10$ GeV 2 jets, $\lvert\eta\rvert < 4.5$ 1 b-jet, $\lvert\eta\rvert < 2.5$	1 OSSF pair $\lvert m_{\ell\ell} - m_Z \rvert < 10$ GeV 3 jets, $\lvert\eta\rvert < 4.5$ 1 b-jet, $\lvert\eta\rvert < 2.5$	1 OSSF pair $\lvert m_{\ell\ell} - m_Z \rvert < 10$ GeV 2 or 3 jets, $\lvert\eta\rvert < 4.5$ 1 b-jet, $\lvert\eta\rvert < 2.5$

Two selections: **Default**, **Loose**.

Selection	Jet p_T	b-tag working point	Leptons p_T
Default	35 GeV	70%	27,20,15 GeV
Loose	20 GeV	85%	27,15,10 GeV

Dominant background.

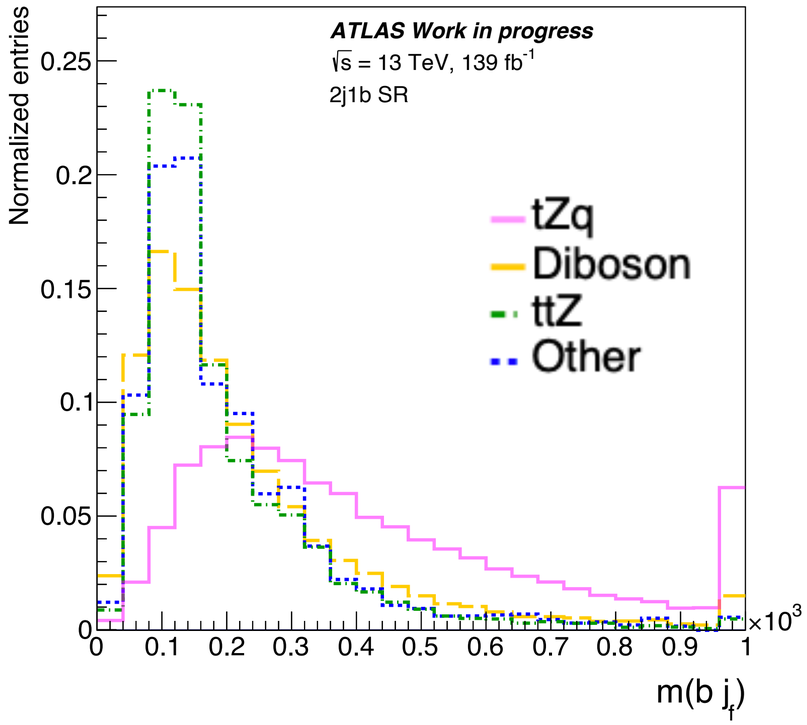


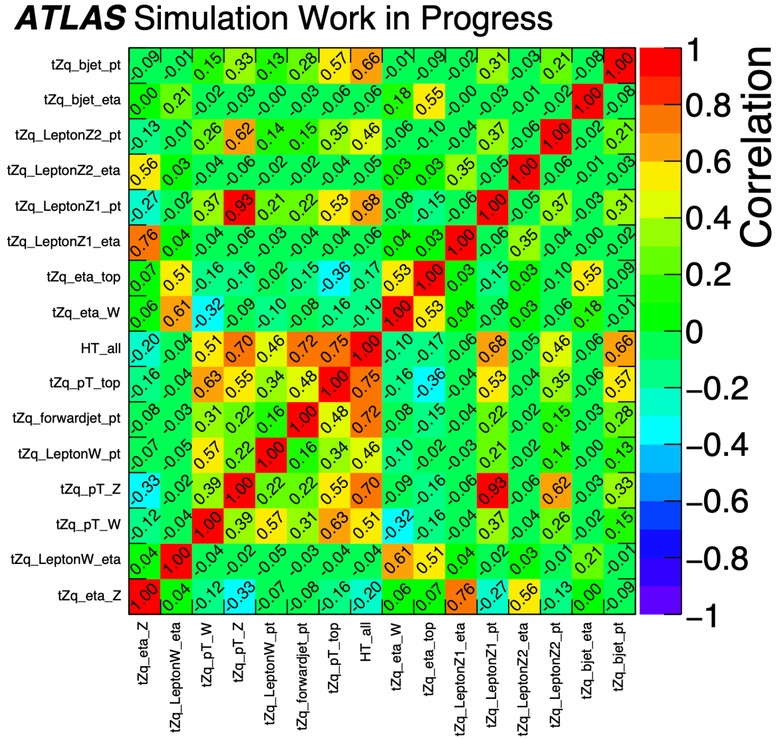
- **Diboson**: mostly WZ events. Dominant background in 2j1b and 3j1b SRs;
- **$t\bar{t}Z$**: strong process of a pair production in association with a Z boson.
  Dominant background in 3j1b SR;

.

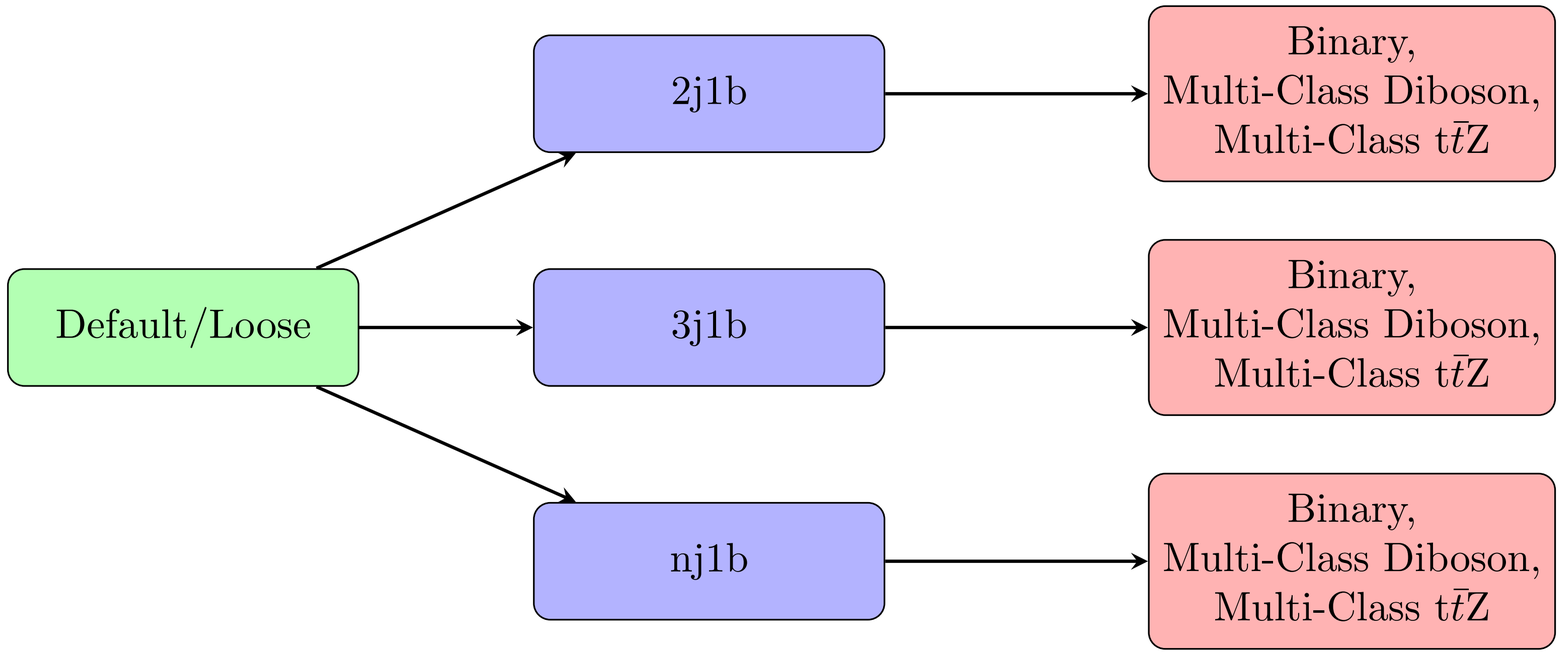
Feature selection.

41 Variables → Separation Strength → Pearson's Correlation → 23 Variables





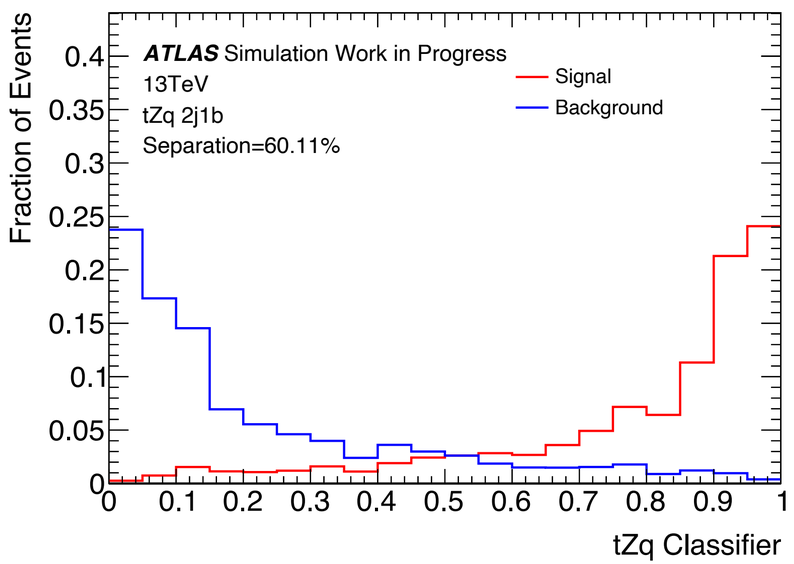
Deep Neural Networks.



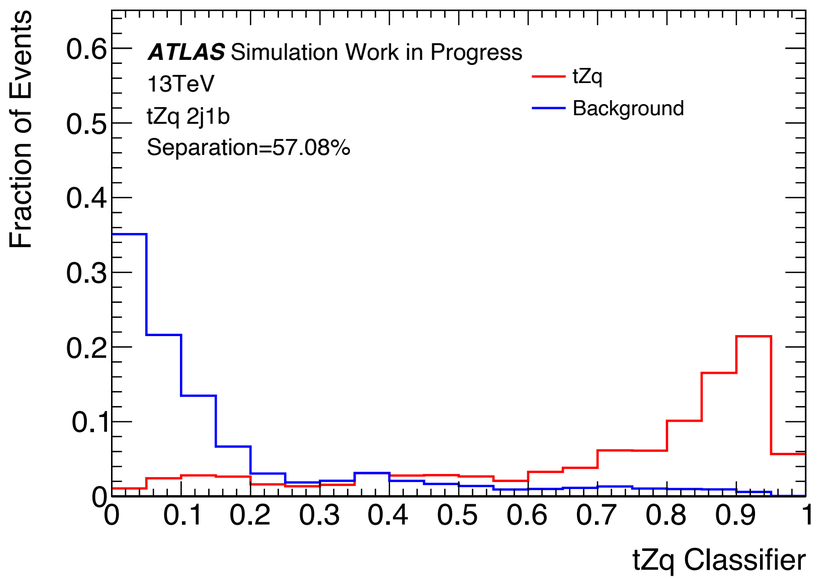
Classifiers:

- **Binary**:
  signal vs background;
- **Multi-class**:
  three classes.
  - Signal (tZq);
  - Dominant background (diboson/t$\bar{t}$Z);
  - All other processes;

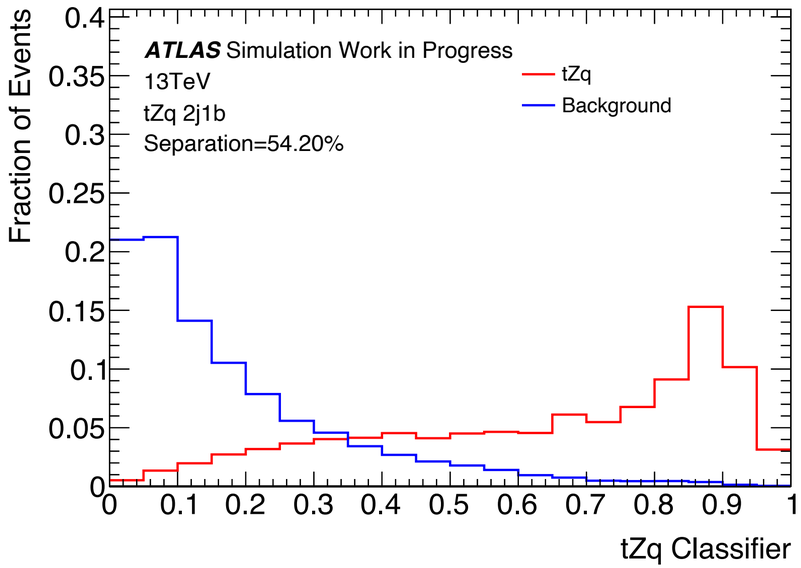
The classifier problem.



Binary



Multi-class Diboson



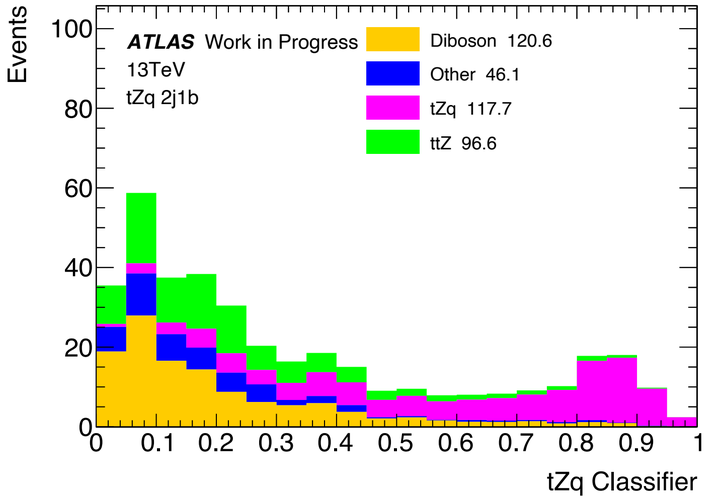
Multi-class $t\bar{t}Z$

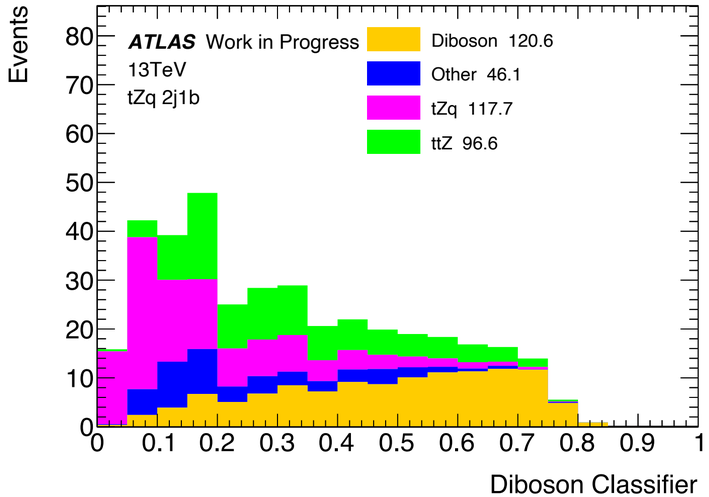
The classifier problem.

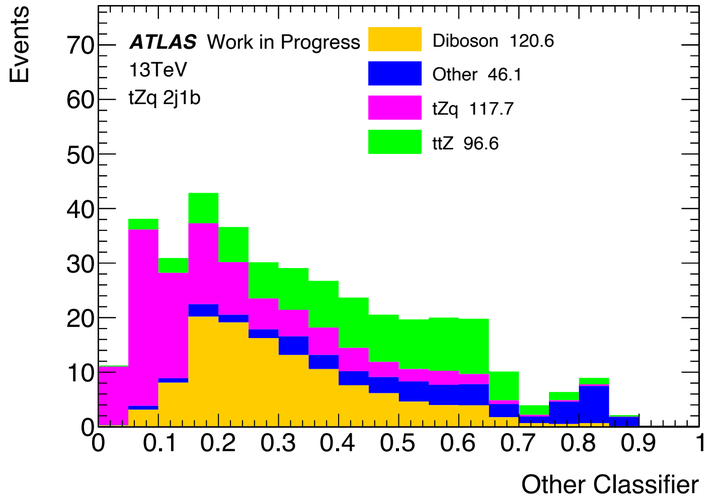
SR	Classifier	Separation Binary	Separation Diboson	Separation $t\bar{t}Z$
2j1b	Default	43.77 %	44.47 %	44.08 %
	Loose	60.11 %	57.08 %	54.20 %
3j1b	Default	39.15 %	37.11 %	38.42 %
	Loose	53.03 %	51.43 %	49.05 %
nj1b	Default	40.60 %	38.72 %	40.48 %
	Loose	57.17 %	56.22 %	52.84 %

Binary classifiers $\geq$ multi-class classifiers

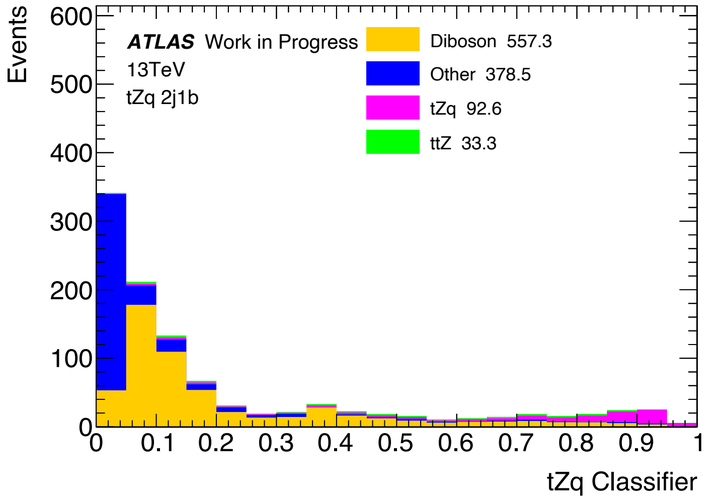
The selection problem.

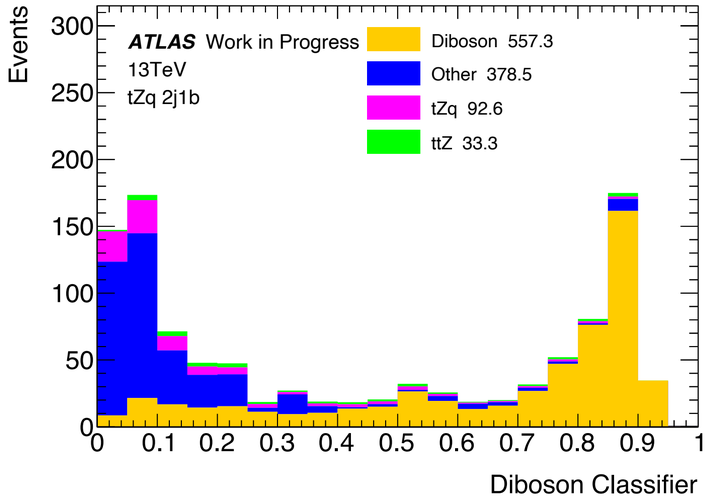


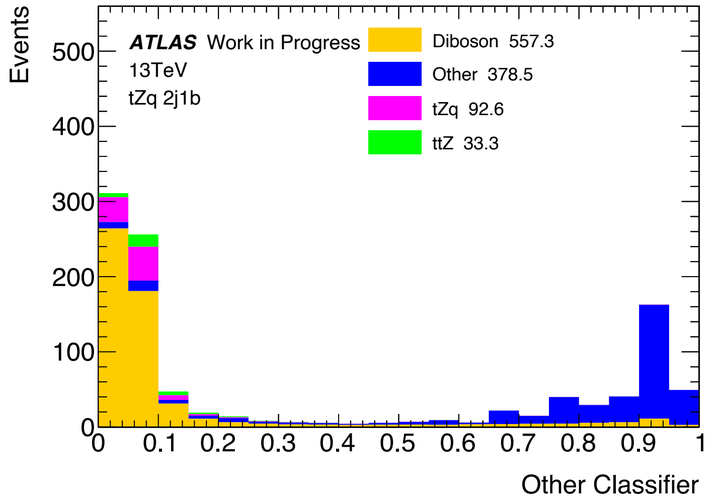




Default







Loose

The selection problem.

SR	Classifier	Separation Binary	Separation Diboson tZq	Diboson	Other
2j1b	Default	43.77 %	44.47 %	28.32 %	27.04 %
	Loose	60.11 %	57.08 %	52.94 %	59.91 %
3j1b	Default	39.15 %	37.11 %	21.28 %	21.60
	Loose	53.03 %	51.43 %	45.48 %	46.32 %
nj1b	Default	40.60 %	38.72 %	24.26 %	20.46
	Loose	57.17 %	56.22 %	52.39 %	55.87 %

SR	Cut on tZq node	$S/\sqrt{S+B}$ Default	Loose
2j1b	Binary	7.82	6.09
	Multi-class diboson	7.86	5.86
3j1b	Binary	5.10	6.09
	Multi-class diboson	5.03	6.01
nj1b	Binary	9.18	8.67
	Multi-class diboson	9.08	8.62

Separation Default $\leq$ Separation Loose

Significance Default $\geq$ Significance Loose

Conclusions and future prospects.

- The **classifier** problem: binary separation $\geq$ multi-class separation.
  Is the separation the best metric to choose the classifier?
- The **selection** problem:

Loose vs Default

	Selection	
	Default	Loose
Separation	Worse	Better
Significance	Better	Worse

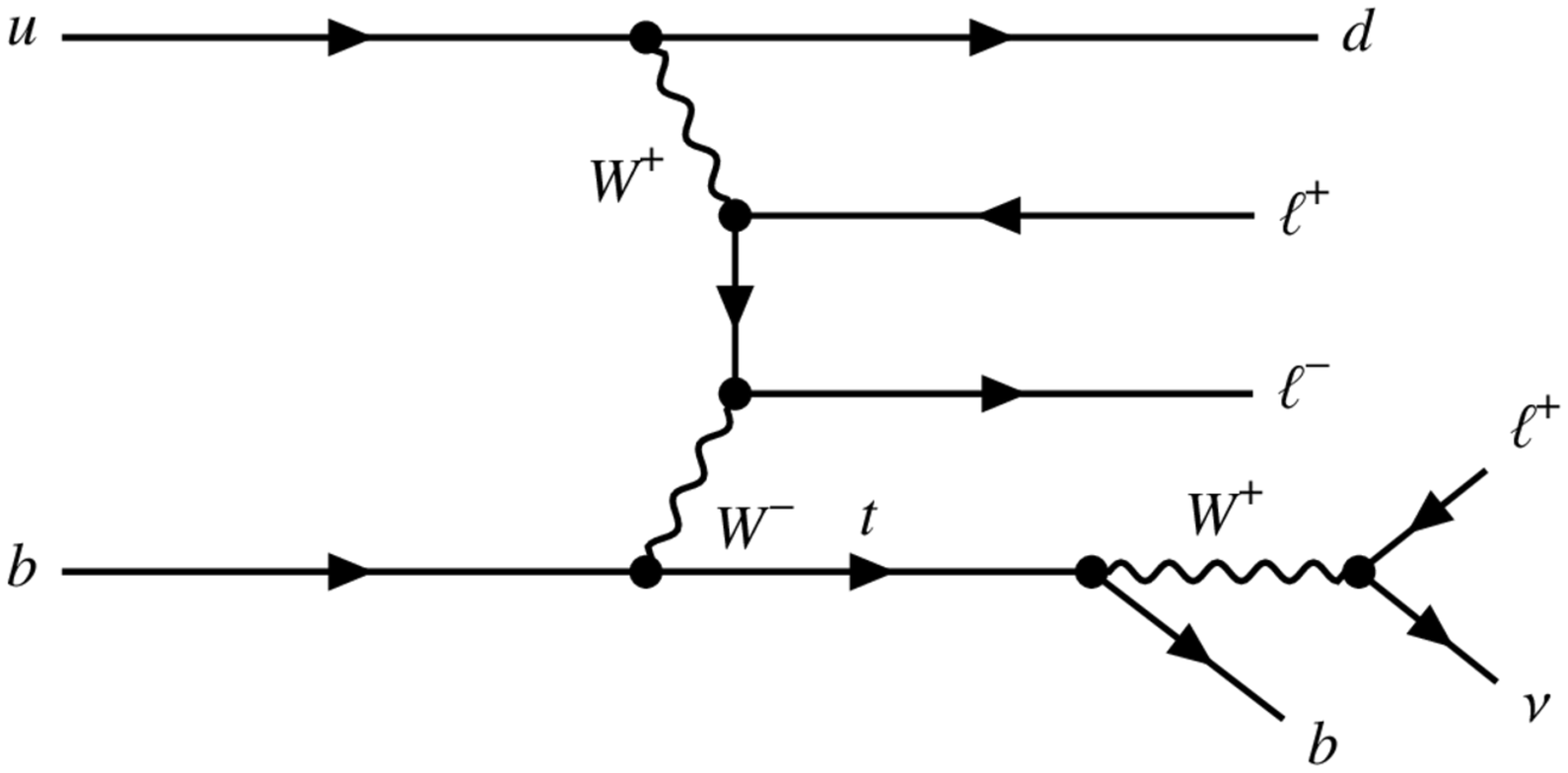
Transfer Learning:

- Training on Loose selection;
- Work on Default selection;

Thank you for your attention.

Niccolò Laurora

Non-resonant tZq Feynman diagram.



Separation.

$$\langle S^2 \rangle = \frac{1}{2} \sum_{i=1}^{bin} \frac{(S_i - B_i)^2}{S_i + B_i}$$

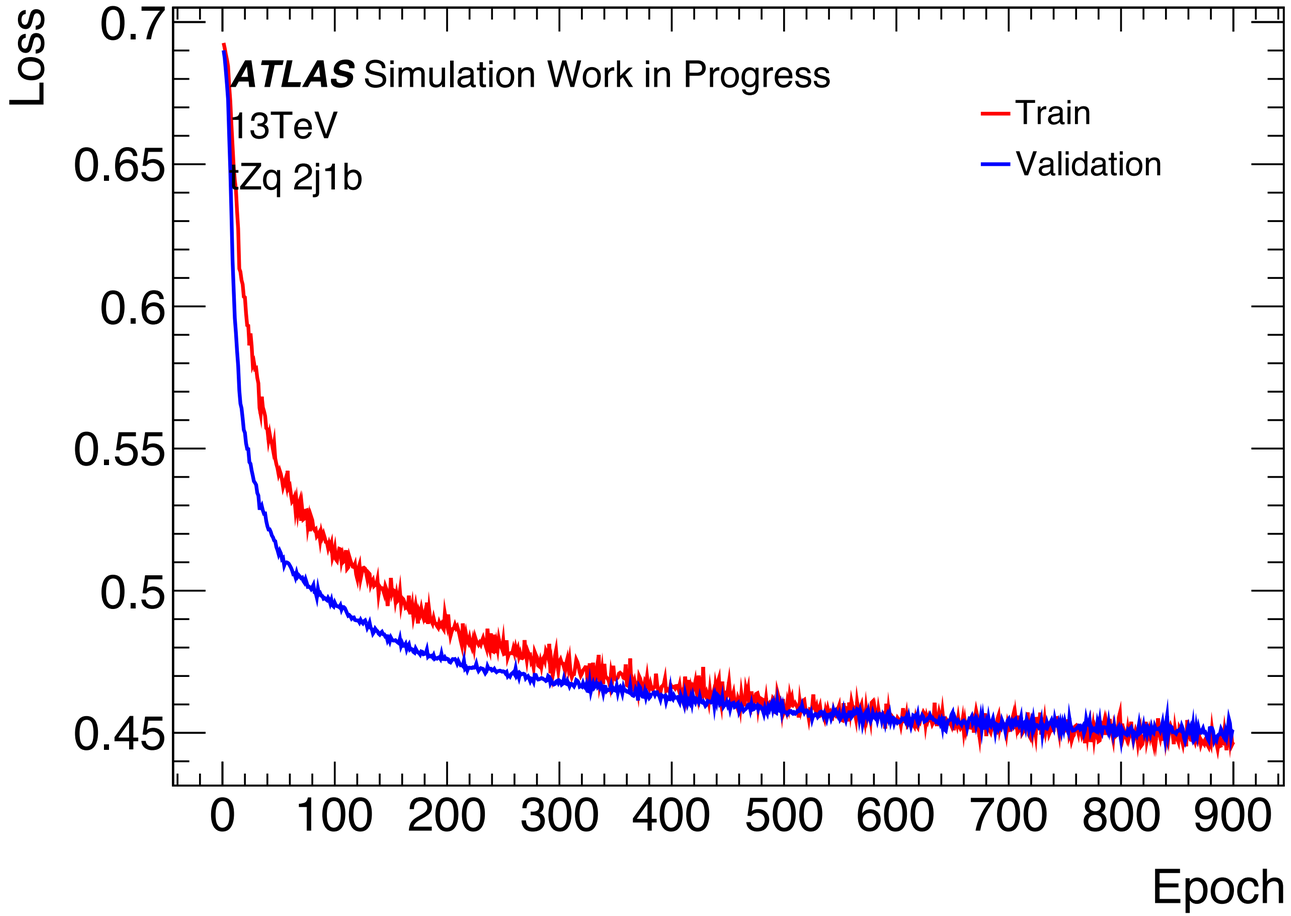
Variables.

$m(b, j_f)$	$m(top)$	$q(\ell_W)$
$m_T(W)$	$m(Z)$	$j_b(b-tagging)$
$p_T(j_b)$	$\lvert\eta(j_b)\rvert$	$j_f(b-tagging)$
$\lvert\eta(j_f)\rvert$	$\lvert\eta(Z)\rvert$	$mindiffmass$
$\lvert\eta(\ell_W)\rvert$	$p_T(\ell_W)$	$\Delta\phi(W, j_{closest})$
$p_T(j_f)$	$p_T(top)$	$\Delta R(top, Z)$
$\Delta R(j_f, Z)$	$\Delta\phi(top, Z)$	$\Delta R(\ell_W, j_{closest})$
$p_T(Z)$	$p_T(W)$	

DNNs' specifics.

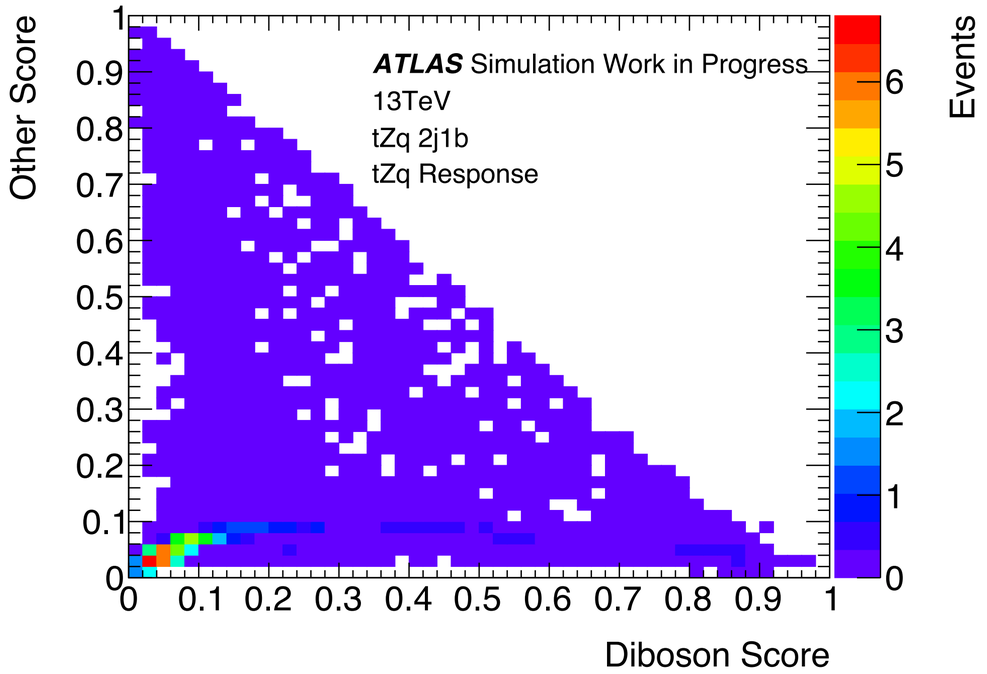
- 5 hidden layers: 20, 30, 30, 30, 20;
- Activation function: ReLU;
- 2000 epochs;
- Batch Size equal to 1000;
- 4 Dropout layers;
- 3 folds;
- MinMax feature scaler;
- Early Stopping: 100 epochs, $\Delta = 0.0005$;
- Validation size 25 %;
- Learning rate differs from one DNN to another (~ 0.0001);
- No BatchNormalisation layers;

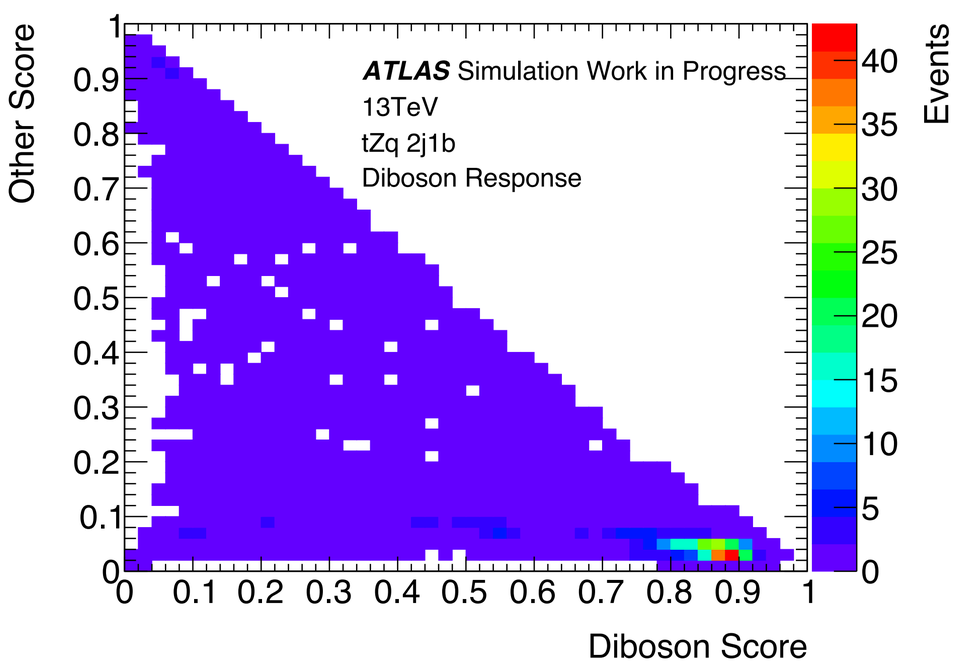
Loss Functions.



Every Loss function showed a well converging model.

Future prospects.





The separation might not be the best metric to choose the classifier.

2D Response Plots:

- SR of scores;
- CR of scores;

Selections' Yields.

Event	Default		Loose	
	SR 2j1b	SR 3j1b	SR 2j1b	SR 3j1b
tZq	118	66	93	99
Diboson	235	69	557	450
$t\bar{t}Z$ + tWZ	97	148	33	79
Z + jets	10	10	368	205
Others	9	6	6	6
Total	469	299	1057	839

b-tagging.

b-jets: hadronic jets from a b-quark.

- b-hadron longer lifetime;
- Secondary vertex;

b-tagging:

- Reconstructed in the ID ($|\eta| < 2.5$);
- Many b-jets identification algorithms (based on vertex and tracks informations);