Direct optimization of discovery significance

A. Mohamed, D. Krüecker, I. Melzer-Pellmann, I. Pidhurskyi, O. Turkot

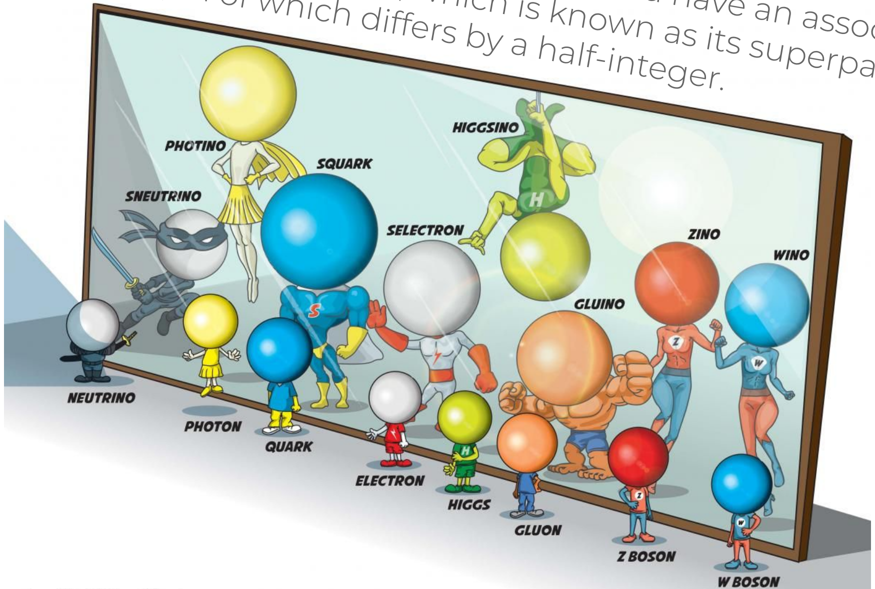
- SUSY
- DNN
- Discovery significance
- Binary classification
- Multiclass
- Summary

SUSY

SUSY is a principle that proposes a relationship between two basic classes of elementary particles: bosons and fermions.

Bosons (integer-valued spin) ⟷ SUSY ⟷ Fermions (half-integer spin)

Each particle from one group would have an associated particle in the other, which is known as its superpartner, the spin of which differs by a half-integer.

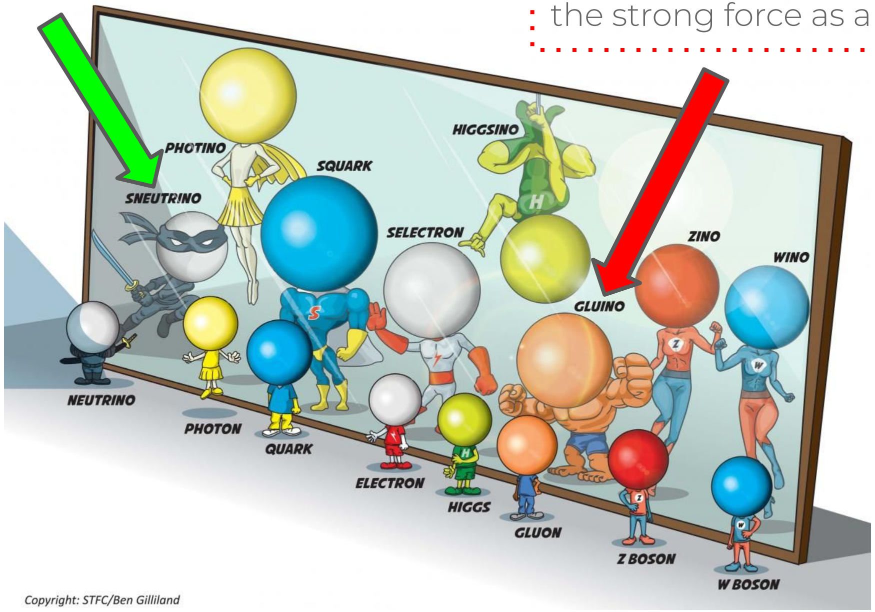


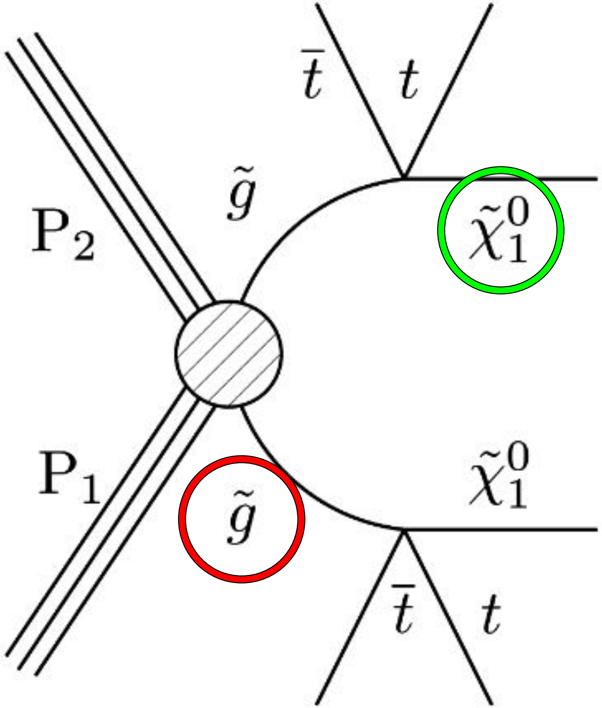
- Solves
  - hierarchy problems
  - Higgs mass
  - cosmological constant problem
  - …
- DM candidates
- Grand unification
- A chance for string theory
- …

SUSY

Lightest Supersymmetric Particle (LSP). One of the main DM candidates.

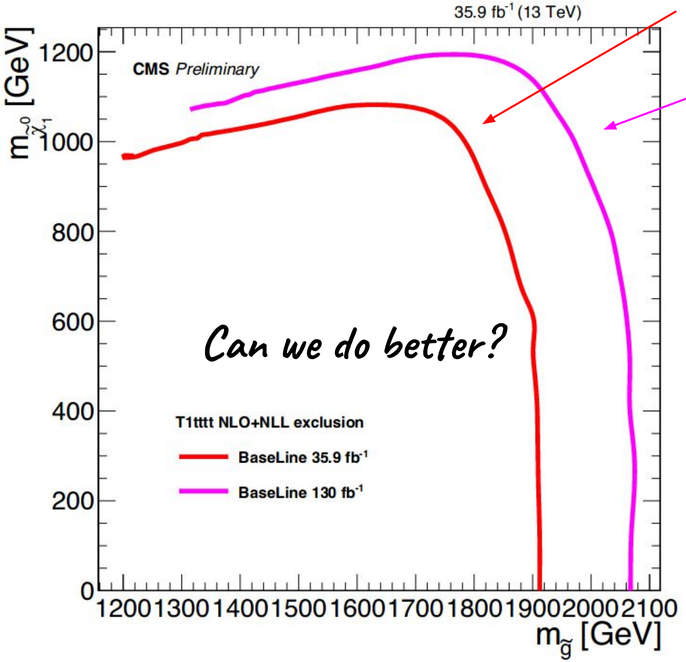
Majorana fermions and interact via the strong force as a color octet.





SUSY

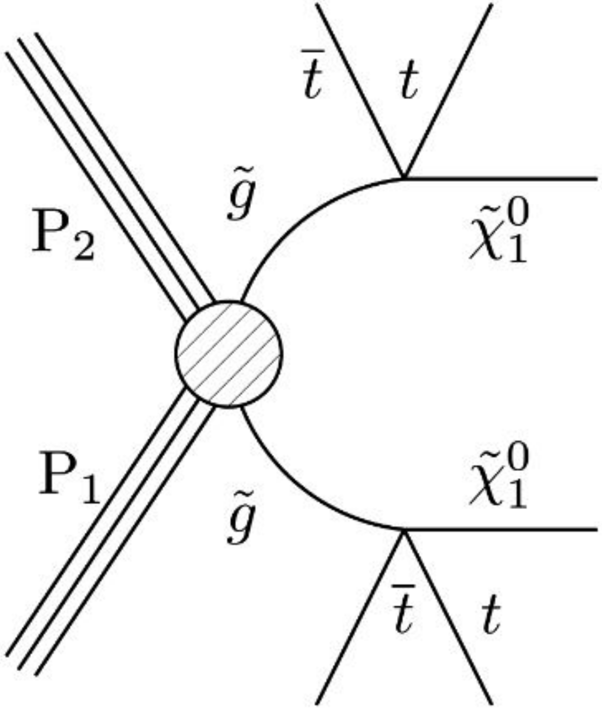
95% C.L. limits



CMS data for 2016

for 2016 + 2017 + 2018

~2 σ exclusion ranges we are able to set, considering there is no SUSY in real world.



SUSY

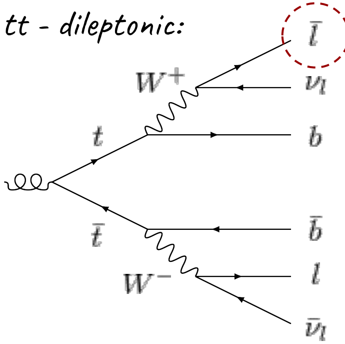
$t \to W + \underline{b\text{-jet}}$

$W \to ll \,/\, qq$

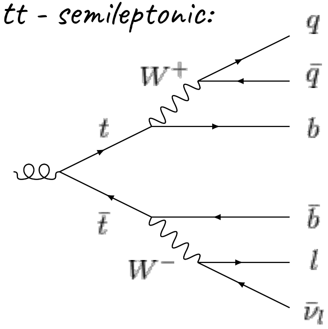
Main backgrounds:

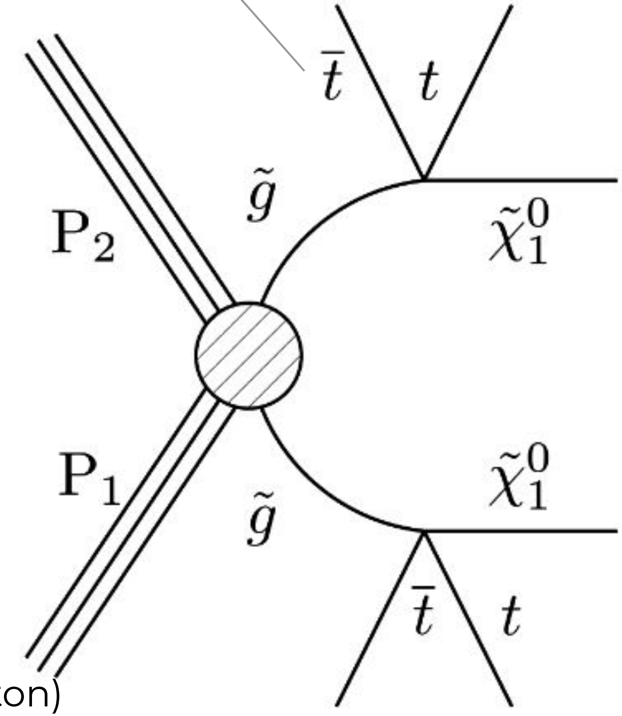
can get undetected

tt - dileptonic:



tt - semileptonic:



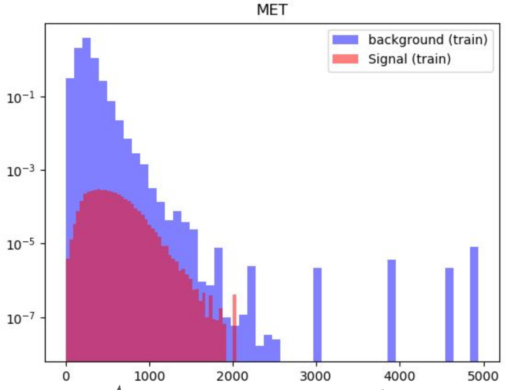


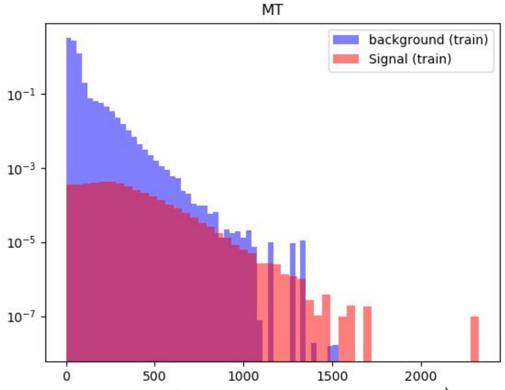
Signs:

- b-jets (and may also have jets from hadronic W-decays)
- leptons from W-decays (we constrain events to have 1 lepton)
- Missing Transverse Energy (**MET**) due to neutralinos and neutrinos
- and kinematics …

SUSY

- can discriminate signal and background by kinematic features
- but it may still be a difficult task due to combinatorial background



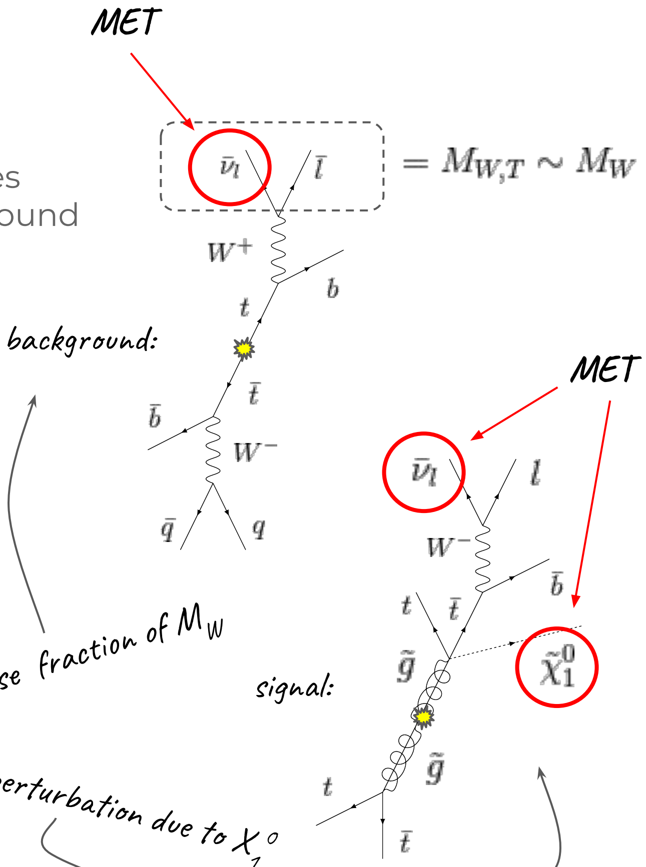


these are some of the best available sources for discrimination

$$\mathrm{MT} = \sqrt{P_{T,miss} P_{T,lep}(1-\cos(\Delta\phi))}$$

= transverse fraction of M_W

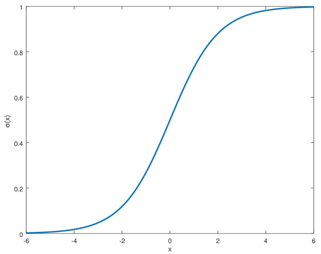
= $M_{W,T}$ + perturbation due to χ_1^0

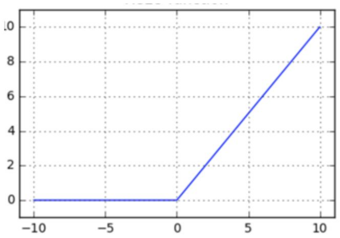


- can combine various features for better discrimination

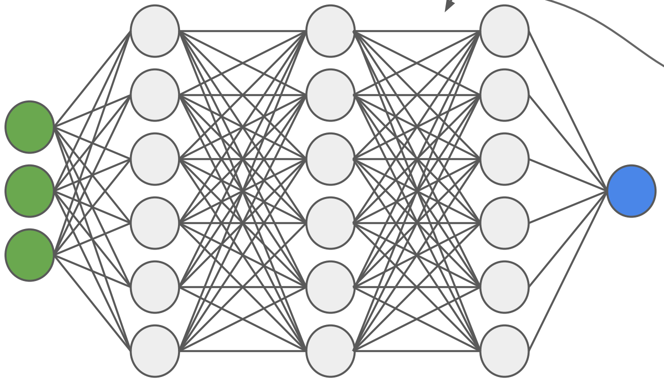
=> looks like a job for machine learning!

DNN

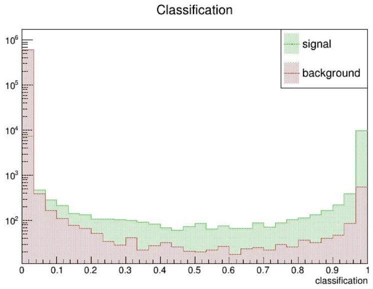


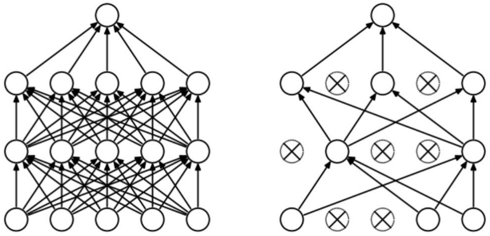


- input layer (20 features) + 2 hidden layers
  - fully connected
  - 256 nodes each
  - with *ReLU* activation
  - *10% dropout* to fight overtraining



- single output
  - *sigmoid* activation
  - -> signal/background classification
- optimization with *Adadelta* (adaptive learning rate).

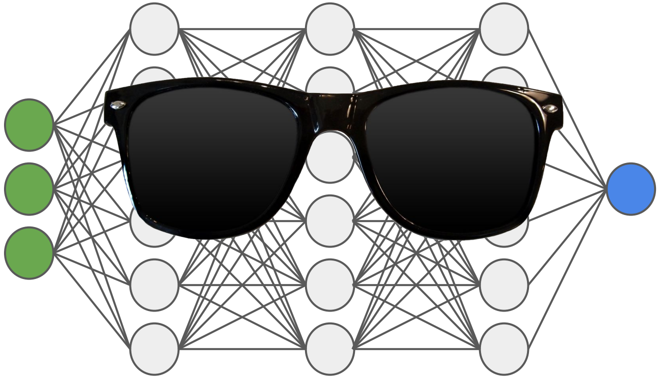




(a) Standard Neural Net (b) After applying dropout.

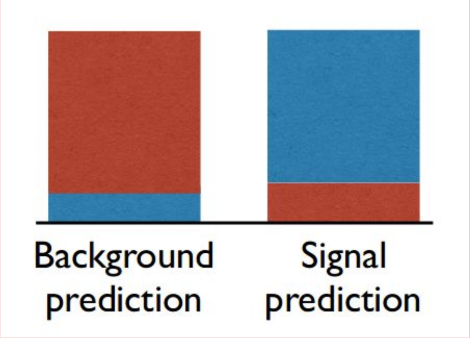
Discovery significance

arXiv: 1806.00322



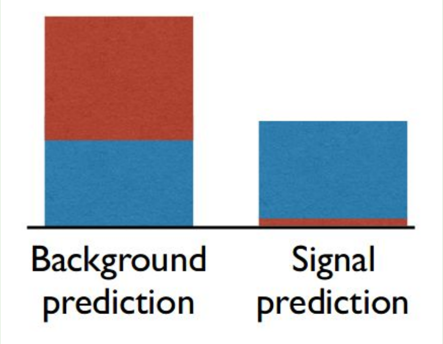
Classic approach:

- optimize accuracy
- loss e.g. crossentropy



HEP approach:

- optimize significance i.e. less signal but pure
- approximated with Asimov estimate, Z_A
- loss = $1 / Z_A$



Discovery significance

$$Z_A = \left[2\left((s+b)\ln\left[\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right] - \frac{b^2}{\sigma_b^2}\ln\left[1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right]\right)\right]^{1/2}$$

systematic uncertainty on background

signal counts:

$$s = W_s \sum_{i\in sig_{batch}} w_i y_i^{pred}$$

$$W_s = \frac{L\sum_{i\in sig_{data}} w_i}{\sum_{i\in sig_{batch}} w_i}$$

background counts:

$$b = W_b \sum_{i\in bg_{batch}} w_i y_i^{pred}$$

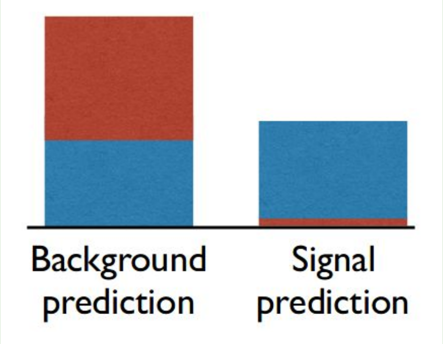
$$W_b = \frac{L\sum_{i\in bg_{data}} w_i}{\sum_{i\in bg_{batch}} w_i}$$

$$w_i = \frac{\sigma_i \epsilon_i}{N}$$

Differentiable => can do gradient descent (training)

HEP approach:

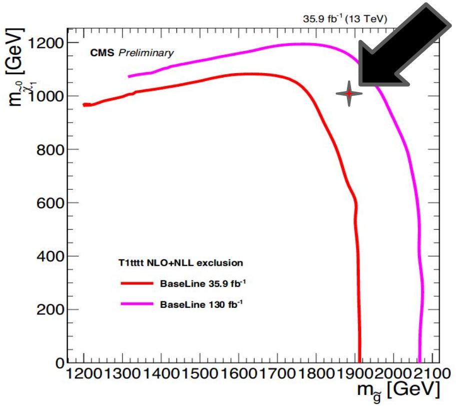
- optimize significance i.e. less signal but pure
- approximated with Asimov estimate, Z_A
- loss = 1 / Z_A



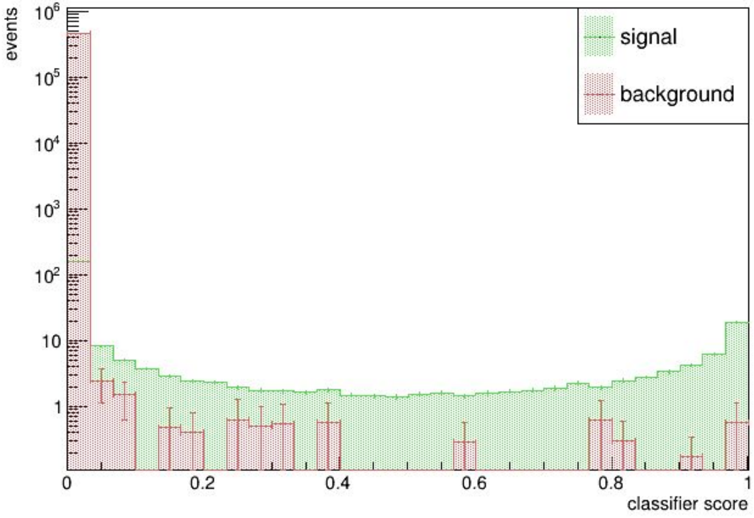
Binary classification

Note: using data for 2016 (35.9 fb^{-1})

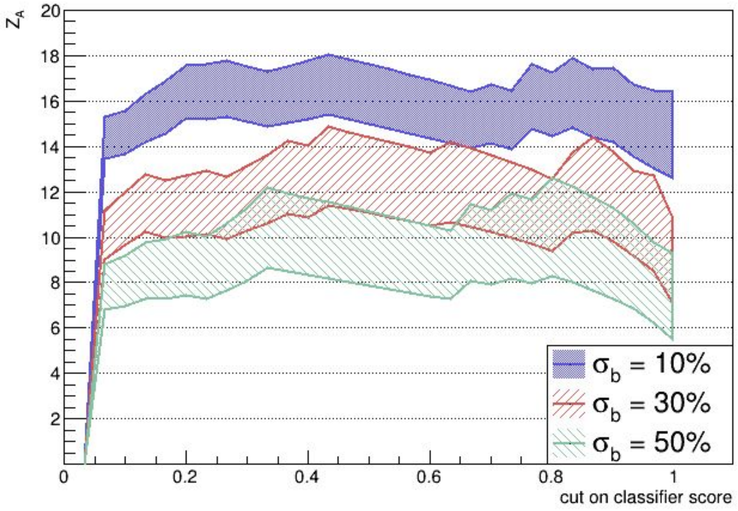
M_{LSP} = 1000 GeV
M_{gluino} = 1900 GeV



- Pure signal sample separated
- **High** significance is reached
- ...even with 50% background uncertainty



Classification after training for σ_b = 30%

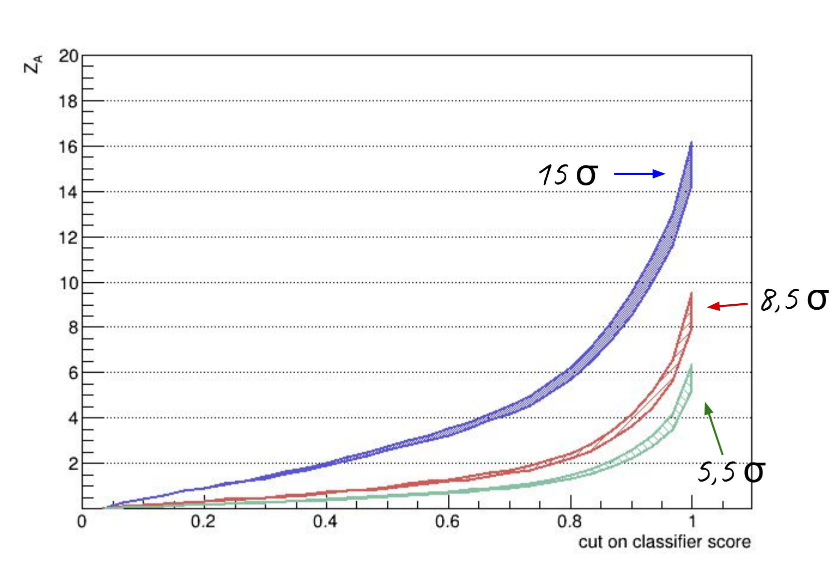


significance > 8σ even at σ_b = 50%

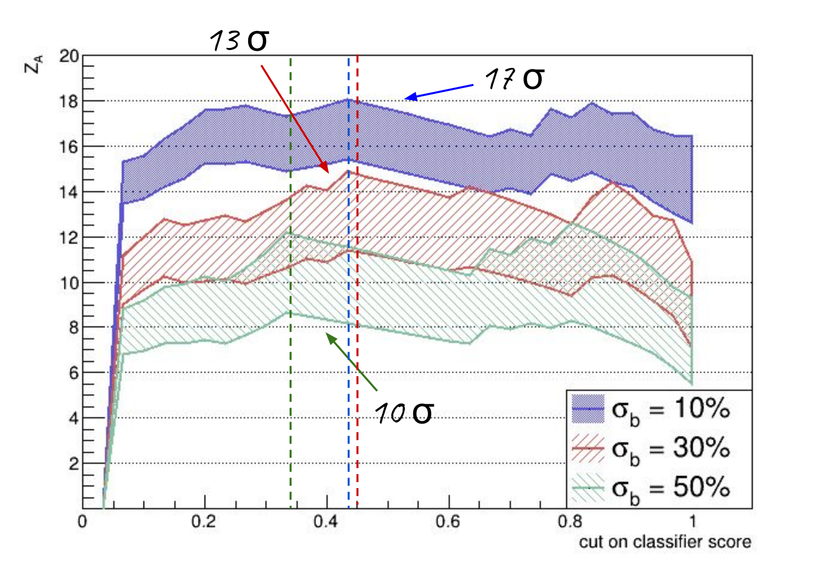
Binary classification

M_{LSP} = 1000 GeV
M_{gluino} = 1900 GeV

Binary crossentropy

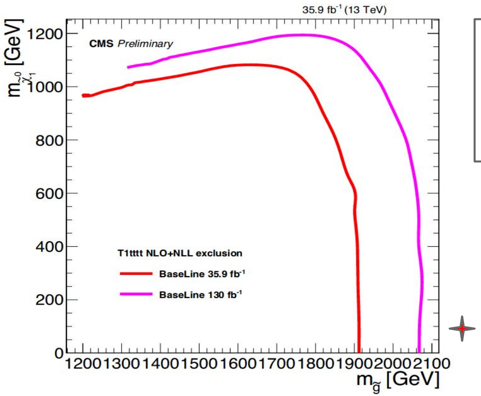


Asimov loss



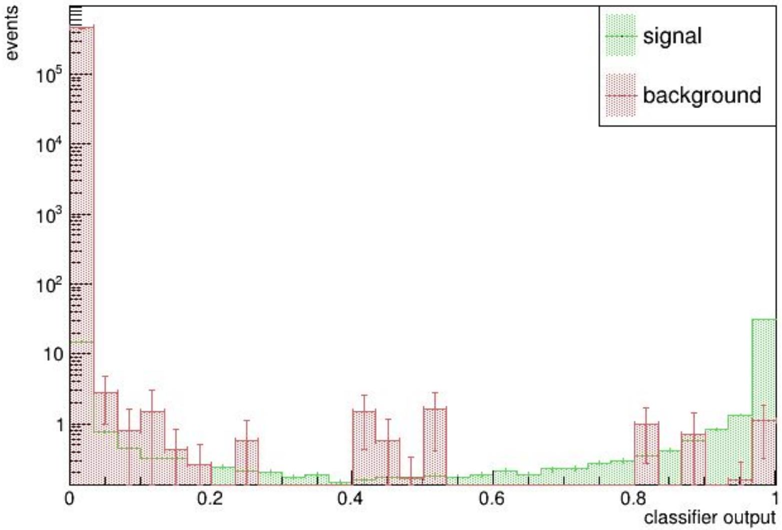
Binary classification

Note: using data for 2016 (35.9 fb^{-1})

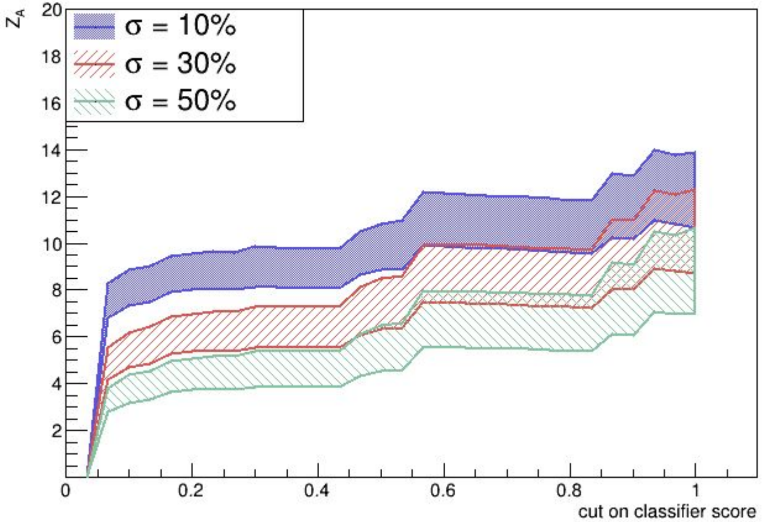


M_{LSP} = 100 GeV
M_{gluino} = 2200 GeV

- Mass point beyond the limit obtained with 130 fb^{-1} of data => still high significance.



Classification after training for σ_b = 10%



significance up to 7σ at σ_b = 50%

Summary

Obtained results:

- Implemented Asimov estimate as loss for DNN training.
- New approach outperforms classic crossentropy.
- Applied to mass points not excluded by the latest limits => better limits can be obtained.

Also studied:

- Multiclass with Asimov loss combined with crossentropy does not improve results compared to binary classification.