Universidade do Minho
Escola de Ciências

LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

Big Data HEP

[New physics searches @LHC]

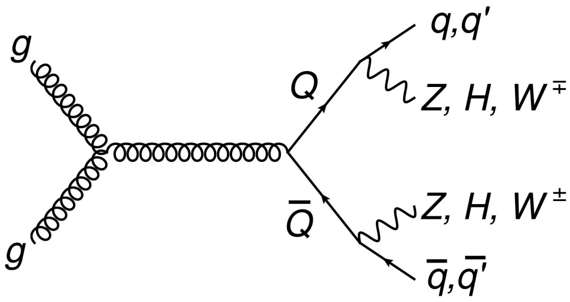
Nuno Castro, Tiago Vale

1st Big Data meeting
11th January 2019
Coimbra, Portugal

New physics searches

- New physics searches heavily rely on good **discrimination**
- **Machine learning** has already been proven to be a key tool for this task
- In this talk I will present a specific task I am involved
  - **Neural network** for classification
  - The methodology should be **transferable** to other physics contexts

Physics context

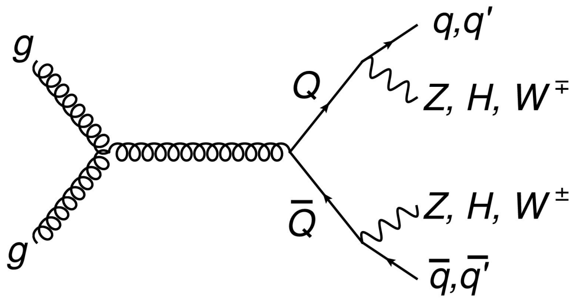


- Searches for vector-like quarks (VLQ) are a big area of interest in the LHC program of searches for new physics phenomena

- They arise from many non-SUSY models that tackle the hierarchy problem

- New **massive** objects (~1 TeV) that decay into Standard Model particles

Physics context



BB ➞
Z(ll)b Z(ll)b
W(lv)t Z(ll)b
W(lv)t W(lv)t
Z(ll)b W(qq)t
Z(ll)b Z(qq)b
Z(ll)b H(bb)b
W(lv)t W(qq)t
W(lv)t Z(qq)b
W(lv)t H(bb)b
W(qq)t W(qq)t
W(qq)t Z(qq)b
Z(qq)b W(qq)t
Z(qq)b Z(qq)b
W(qq)t H(bb)b
Z(qq)b H(bb)b
H(bb)b H(bb)b

TT ➞
Z(ll)t Z(ll)t
W(lv)b Z(ll)t
W(lv)b W(lv)b
Z(ll)t W(qq)b
Z(ll)t Z(qq)t
Z(ll)t H(bb)t
W(lv)b W(qq)b
W(lv)b Z(qq)t
W(lv)b H(bb)t
W(qq)b W(qq)b
W(qq)b Z(qq)t
Z(qq)t W(qq)b
Z(qq)t Z(qq)t
W(qq)b H(bb)t
Z(qq)t H(bb)t
H(bb)t H(bb)t

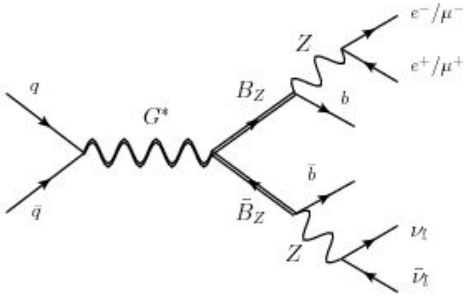
Single VLQ
(B,T,X,Y,B',q*)
➞
Z(ll)b
Z(ll)t
W(lv)b
W(lv)t
V(qq)b
V(qq)t
H(bb)b
H(bb)t

×2 for decays to light quarks

Also:
tttt ➞ 0, 1, 2, 3, 4 leptons + 4b + jets

➡ Many final states lead to inclusive searches like Zt + X

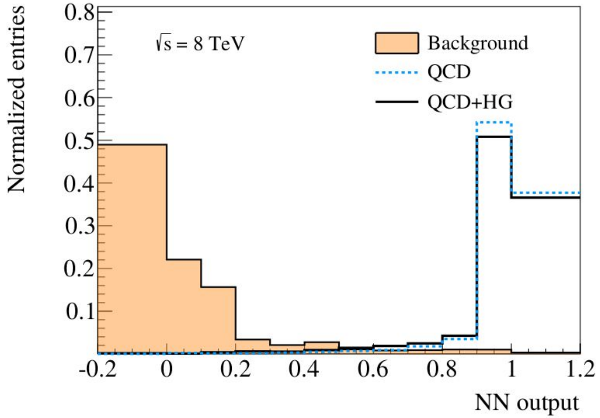
Physics context

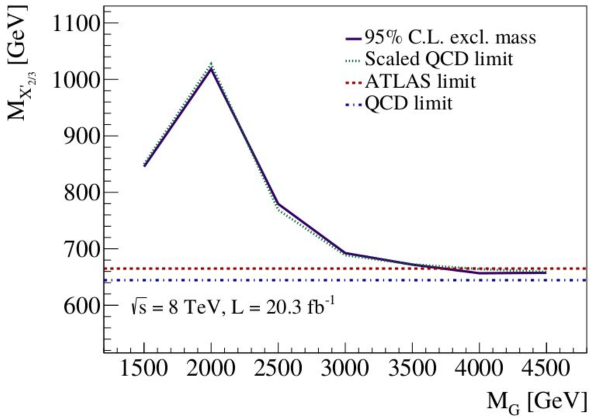


- However some of these BSM models introduce other particles that interact with VLQ
- In some models VLQ can be produced via SM gluon fusion or from the decay of a **heavier gluon** (~3 TeV)
- This should affect the interpretation of our **current** VLQ search results

Current studies

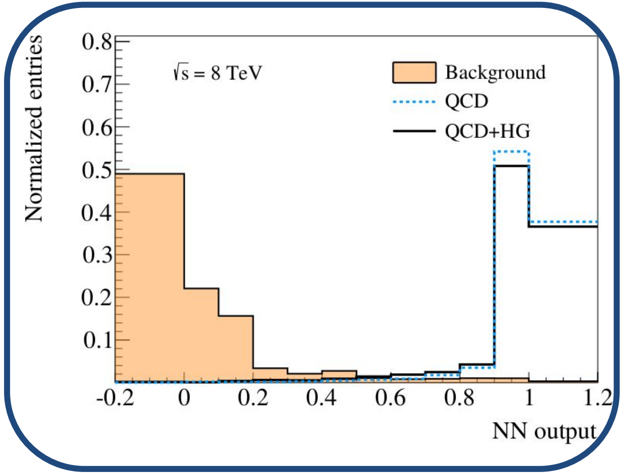
- A recast was previously done http://dx.doi.org/10.1007/JHEP11(2015)120
- Inclusive search targeting leptonic Z decays
- Good discrimination but very similar mass limits
- A neural network could be able to use the whole information for discrimination
  - Our approach:
    - Be **all inclusive** and profit from the whole phase space

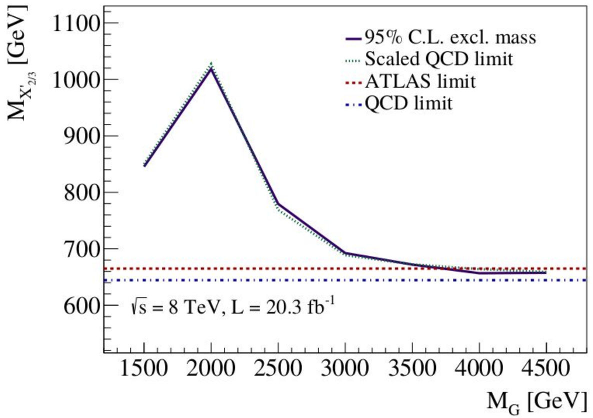




Current studies

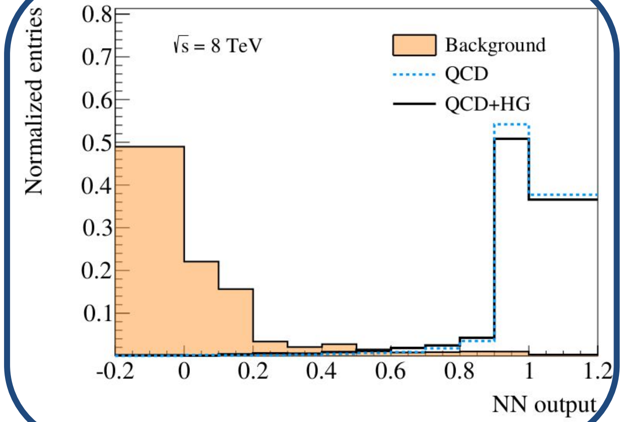
- A recast was previously done http://dx.doi.org/10.1007/JHEP11(2015)120
- Inclusive search targeting leptonic Z decays
- Good discrimination but very similar mass limits
- A neural network could be able to use the whole information for discrimination
  - Our approach:
    - Be **all inclusive** and profit from the whole phase space

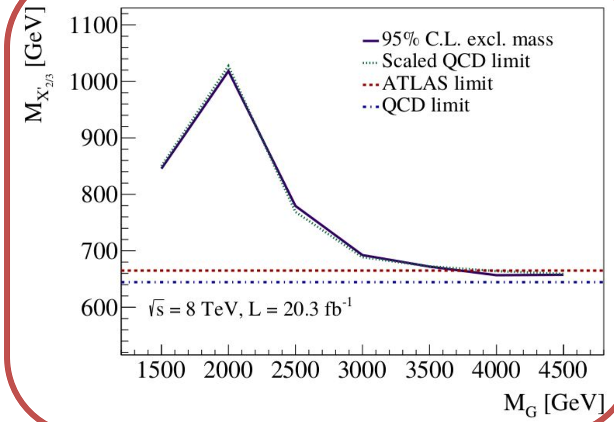




Current studies

- A recast was previously done http://dx.doi.org/10.1007/JHEP11(2015)120
- Inclusive search targeting leptonic Z decays
- Good discrimination but very similar mass limits
- A neural network could be able to use the whole information for discrimination
  - Our approach:
    - Be **all inclusive** and profit from the whole phase space





Generation parameters

- Composite Higgs model

- Heavy gluon (HG) mass of 3 TeV, VLT of 1 TeV

- **Signal** processes:
  - gg -> TT
  - gg -> G -> TT
- **Backgrounds**:
  - ttbar
  - Z + bb
  - Z + cc

This is just a first approach to get everything going, more mass points and backgrounds will have to be added

Neural network

- **Keras** with pandas and scikit-learn
  - Tensorflow as the backend
- Inputs are normalized, standardized and ran through PCA to decorrelate
- Adamax with binary cross-entropy
- **First** architecture approach:
  - 3 layers of 100 nodes
  - selu as activation layer
  - Batch normalization in between each dense layer and its activation layer
  - Sigmoid in the output layer
  - Bayesian optimization machinery in place

Neural network

- Feature selection still not fully studied (machinery is in place, though, with feature ranking)

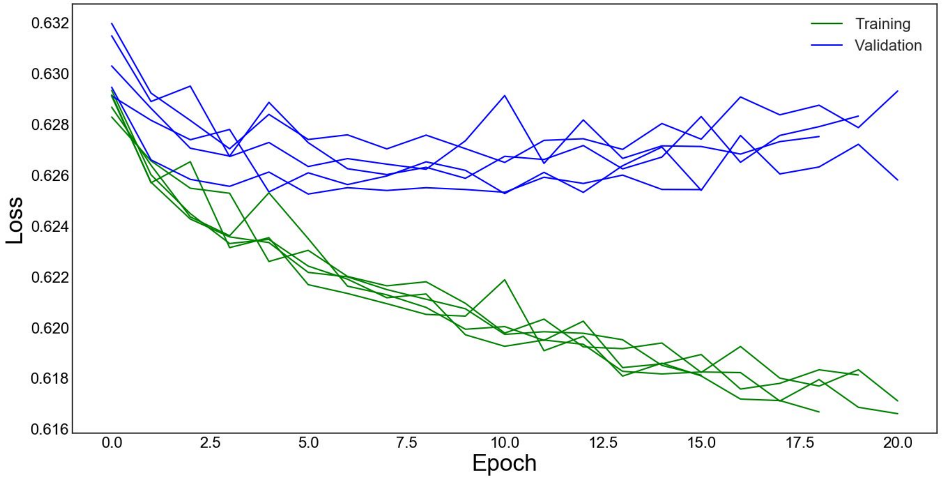
- **Jets** (R = 0.4):
  - pT, mass, eta, phi, btag
  - 3 most energetic
- **Large-R** (1.0) jets:
  - pT, mass eta, phi, tau (1-5)
  - 3 most energetic
- **Leptons** (electrons and muons):
  - pT, eta, phi
  - 2 most energetic
- **MET**

Neural network

- Training until 10 epochs **without** improvement
- Using weights corresponding to **lowest validation loss**
- K-fold cross validation to reduce variance
- **First** approach:
  - Test gg-> TT against gg->G->TT
  - Use this to develop code, study outputs
  - Next step is to train each of the two scenarios against the backgrounds

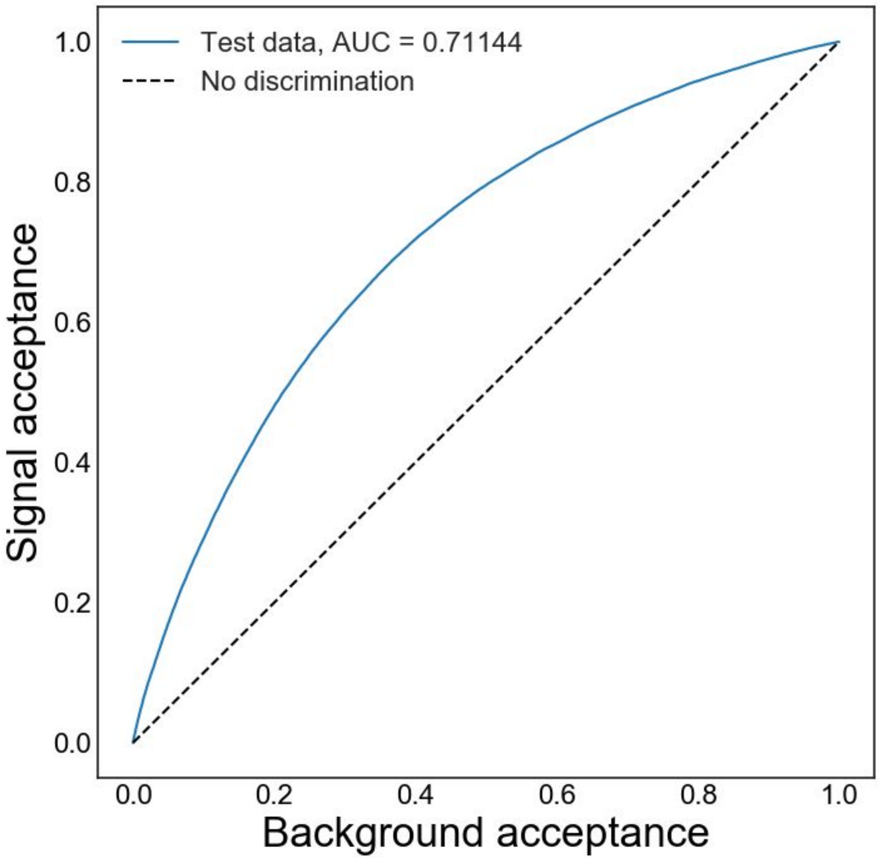
HG vs
SM pair-production

- Stable training
- Best validation loss will be chosen



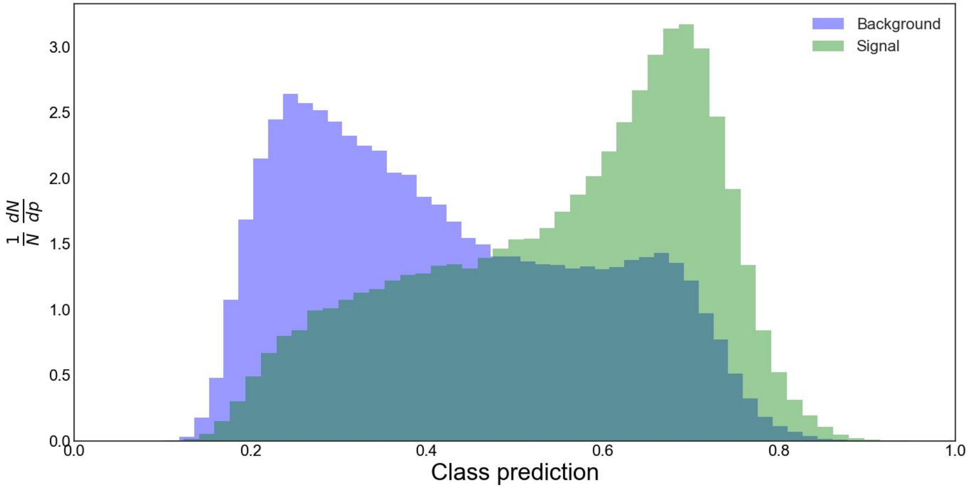
HG vs
SM pair-production

- Decent performance
- Still room for improvement



HG vs
SM pair-production

- Decent discrimination
- Background and signal are HG and SM pair-production

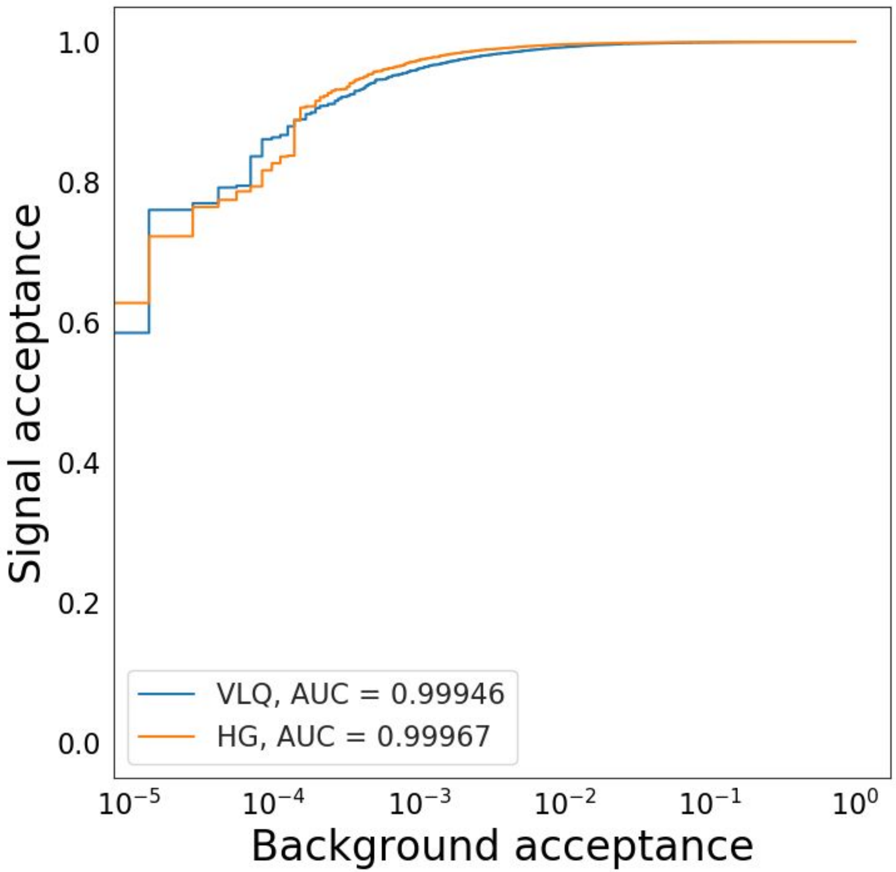


Signals vs Backgrounds

- This was a first approach to get the machinery set
- A recast is done training both signals against the backgrounds
- Afterwards test the impact on search results and reinterpret them

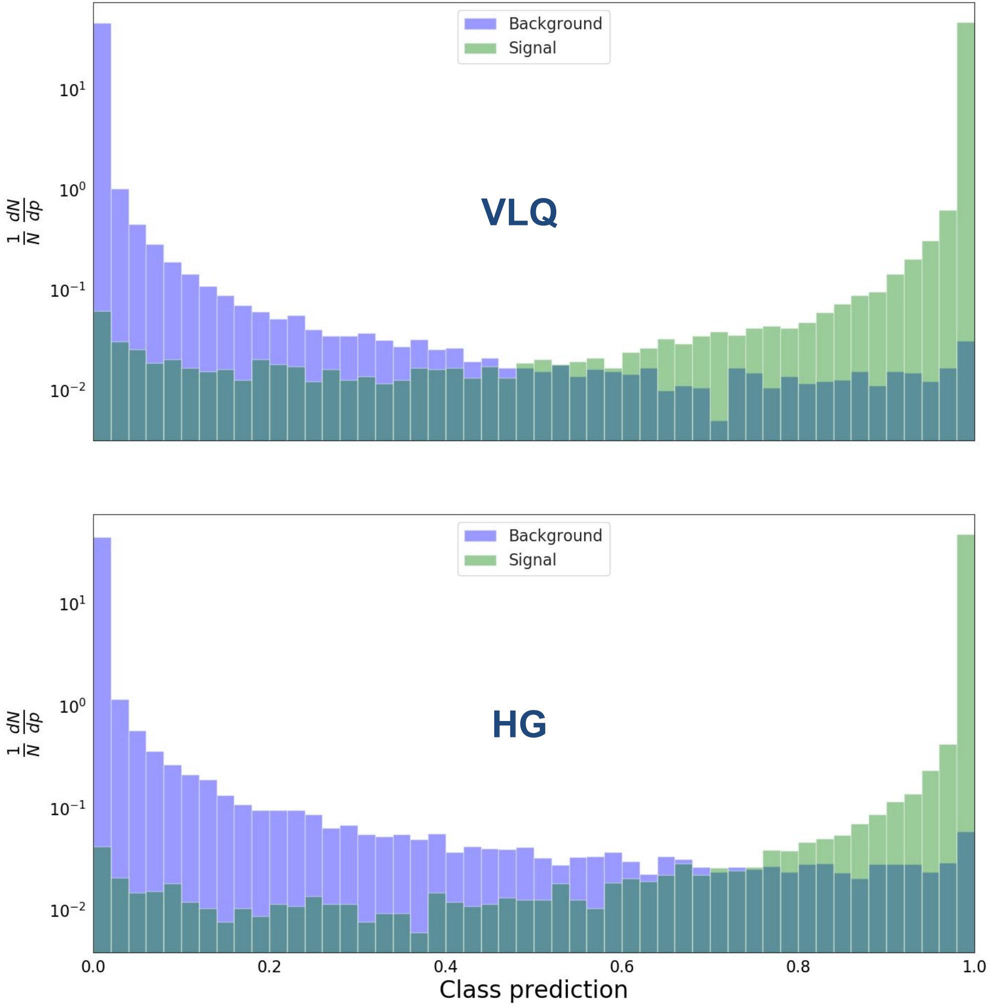
Signals vs Backgrounds

- Good performance
- **Very fresh** so still to be fully validated



Signals vs **Backgrounds**

- Good discrimination
- Should be an excellent fitting variable
- Still be seen if it leads to big effects in the mass limits



Concluding Remarks

- Machinery set for a heavy gluon recast
- Neural network approach seems to be very powerful
- Results are **very fresh** so some digestion is still to be done
  - Tweak training to assure no mistakes were made and everything is robust
- Machinery set for the next steps:
  - Feature engineering and selection
  - Bayesian optimization to select architecture
  - Generation with MadGraph
    - More backgrounds
    - More mass points
- All should be **applicable** to other physics scenarios

Thanks

tiago.vale@cern.ch

Signals vs Backgrounds

- Good discrimination
- Should be an excellent fitting variable