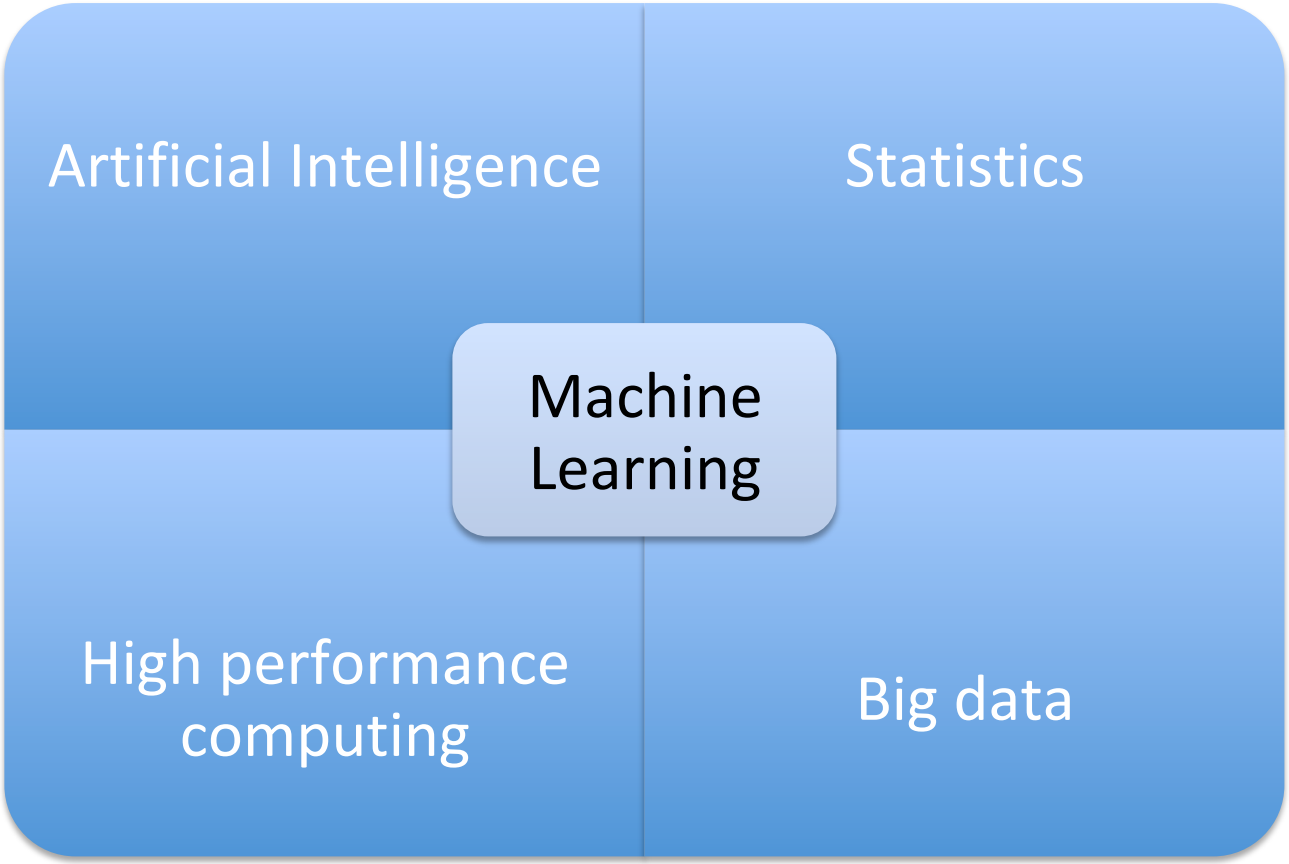
Machine learning for the Higgs boson physics at the LHC

Dr. Adrian Buzatu

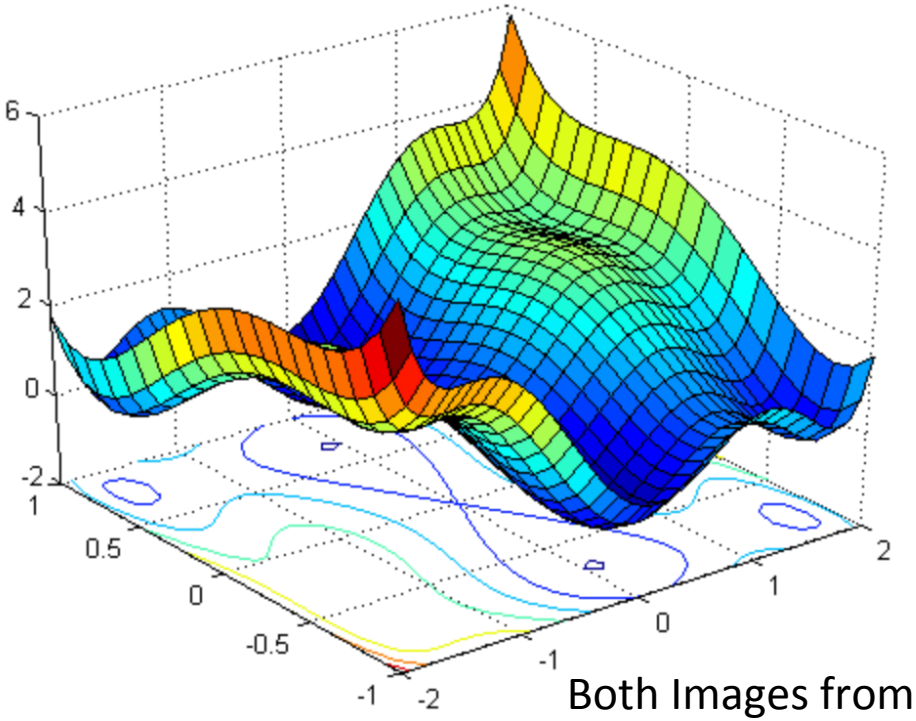
Higgs Couplings 2018

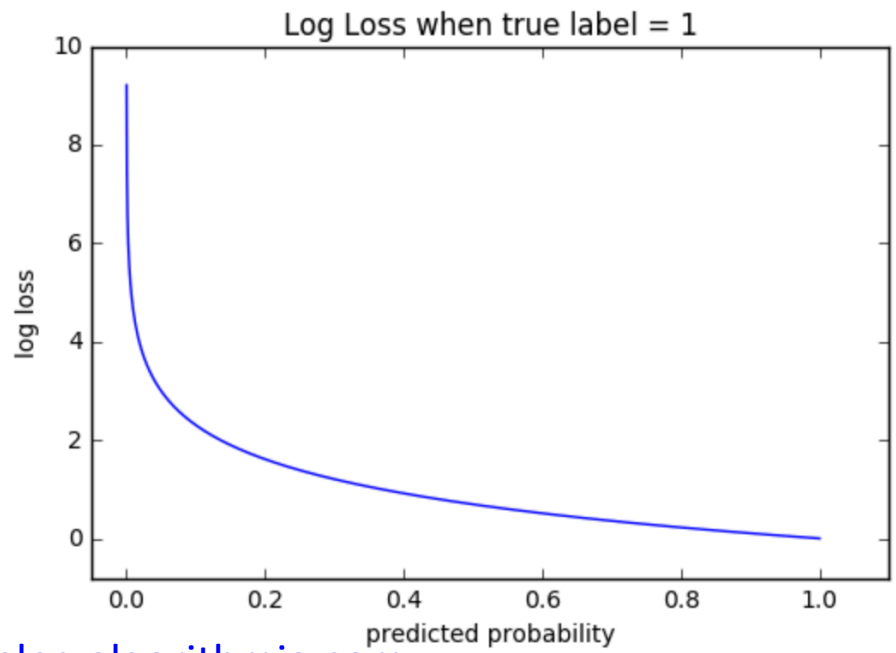
27 Nov 2018, Tokyo, Japan

Machine learning (ML) gives a computer system the ability to learn from real or simulated training data, in a supervised or an unsupervised way, without being explicitly programmed.



Supervised machine learning minimises a loss function that reflects how close are for all the data points the predicted outcome from the actual labelled data.





Both Images from blog.algorithmia.com

Finding the parameters that minimize the loss function via gradient descent.
The learned parameters define a learned model that can then make predictions.

1 epoch = going once through all the data points.

The more epochs, the more we minimize the loss function.

But if too many, risk to over-train.

Split dataset in two: train in one, test in the other.

Machine learning (ML) has been continuously used and improved upon since the Higgs boson discovery.

REVIEW Nature

https://doi.org/10.1038/s41586-018-0361-2

Machine learning at the energy and intensity frontiers of particle physics

Alexander Radovic[1]*, Mike Williams[2]*, David Rousseau[3], Michael Kagan[4], Daniele Bonacorsi[5,6], Alexander Himmel[7], Adam Aurisano[8], Kazuhiro Terao[4] & Taritree Wongjirad[9]

Our knowledge of the fundamental particles of nature and their interactions is summarized by the standard model of particle physics. Advancing our understanding in this field has required experiments that operate at ever higher energies and intensities, which produce extremely large and information-rich data samples. The use of machine-learning techniques is revolutionizing how we interpret these data samples, greatly increasing the discovery potential of present and future experiments. Here we summarize the challenges and opportunities that come with the use of machine learning at the frontiers of particle physics.

The standard model of particle physics is supported by an abundance of experimental evidence, yet we know that it cannot be a complete theory of nature because, for example, it cannot incorporate gravity or explain dark matter. Furthermore, many properties of known particles, including neutrinos and the Higgs boson, have not yet been determined experimentally, and the way in which the emergent properties of complex systems of fundamental particles arise from the underlying standard-model theory remains unknown.

Big data at the LHC

The sensor arrays of the LHC experiments produce data at a rate of about one petabyte per second. Even after drastic data reduction by the custom-built electronics used to readout the sensor arrays, which involves zero suppression of the sparse data streams and the use of various custom compression algorithms, the data rates are still too large to store the data indefinitely—as much as 50 terabytes per second, resulting in as much data every hour as Facebook collects globally in a

HEP ML White Paper 1807.02876

Machine Learning in High Energy Physics Community White Paper

July 10, 2018

Abstract: Machine learning is an important applied research area in particle physics, beginning with applications to high-level physics analysis in the 1990s and 2000s, followed by an explosion of applications in particle and event identification and reconstruction in the 2010s. In this document we discuss promising future research and development areas in machine learning in particle physics with a roadmap for their implementation, software and hardware resource requirements, collaborative initiatives with the data science community, academia and industry, and training the particle physics community in data science. The main objective of the document is to connect and motivate these areas of research and development with the physics drivers of the High-Luminosity Large Hadron Collider and future neutrino experiments and identify the resource needs for their implementation. Additionally we identify areas where collaboration with external communities will be of great benefit.

Editors: Sergei Gleyzer[26], Paul Seyfert[11], Steven Schramm[28]

Combined performance

- Object identification
- Object E/p calibration

Physics analysis

- Separate Signals and several backgrounds
- Reduction of systematic uncertainties

Machine learning is used in other general aspects that indirectly help Higgs boson analyses.

Pattern recognition	Cluster hits
Track classification	Duplicate removal,
	Quality selection
Data quality monitoring	Outlier rejection
Data placement	Predict which samples are mostly needed

Read more at [CMS talk @ LHC 2015]

Machine learning applications have two main steps: learning (free choice) and inference (C++).

Historically: TMVA in ROOT has been the first (main) tool for BDT and NN.
Learning (left): now several modern tools (e.g. Python).
Inference (right): these tools need to adapt in the short term to our ROOT/C++ workflow In order to make inferences/predictions.

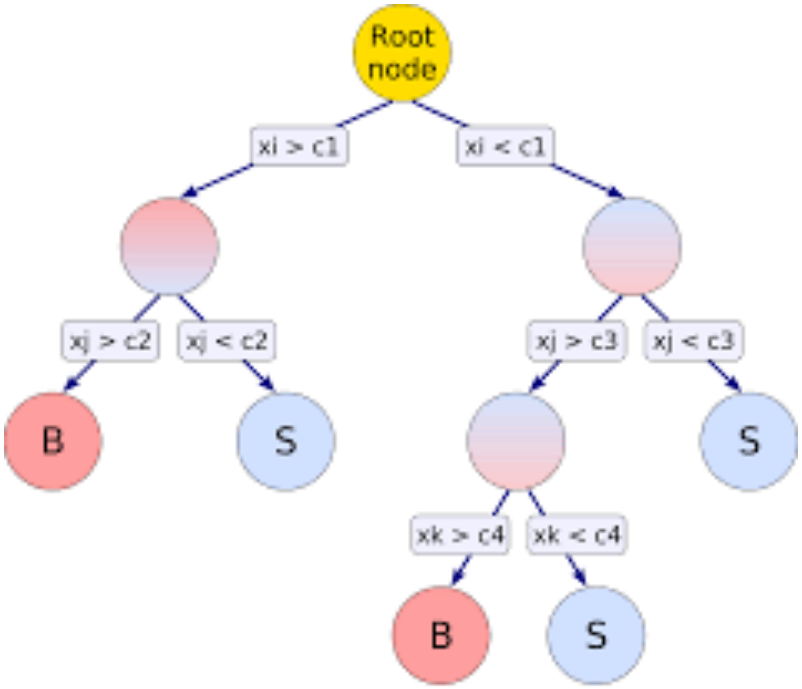
Learning (Training)

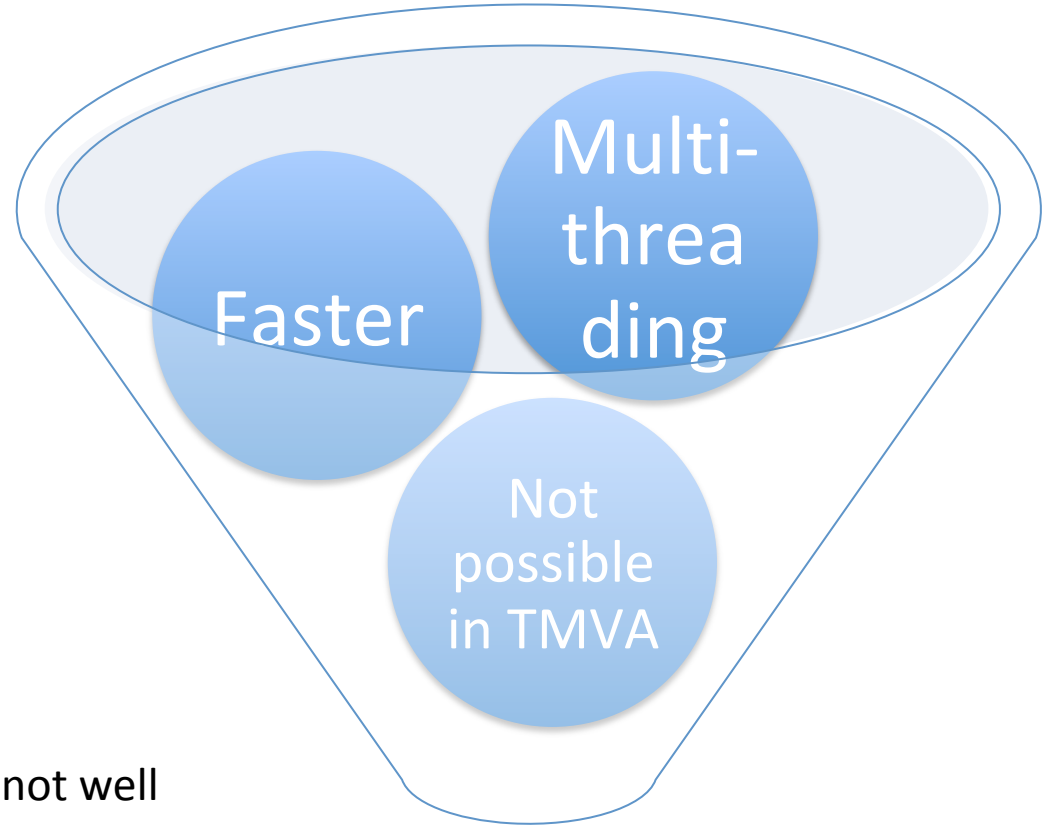
- R&D environment
- TMVA in ROOT
- Modern tools also available
  - BDT -> XGBoost and LightGBM
  - NN -> Keras with Theano or TensorFlow
- scikit-learn

Inference (prediction)

- Production environment
- BDT training -> .xml, ready by TMVA in ROOT, or in Athena with custom code
- XGBoost training evaluated directly in XGBoost with a C++ api
- NN training evaluated with lwtnn

Boosted decision tree (BDT) in TMVA vs Extreme Gradient Boosting in XGBoost





Boosting = creating iteratively new trees where larger weights are given to events not well "learned" by the previous tree.

Final result = weighted average of all trees.

XGBoost can be parallelized with multithreading, So can be faster and open R&D possibilities.

XGBoost

BDT for electron at CMS ID (classification), calibrate energy & predict uncertainty (regression), and ATLAS calibration.

- Select tracks and other constituents of electrons

CMS Identification (classification)

- Instead of parameterization of eta, ET, category
- Improves 20% in resolution
  35% in forward region
  Improves peak position

CMS Energy calibration (regression)

ATLAS Calibration

- For energy in front of calorimeter
- And outside of cluster

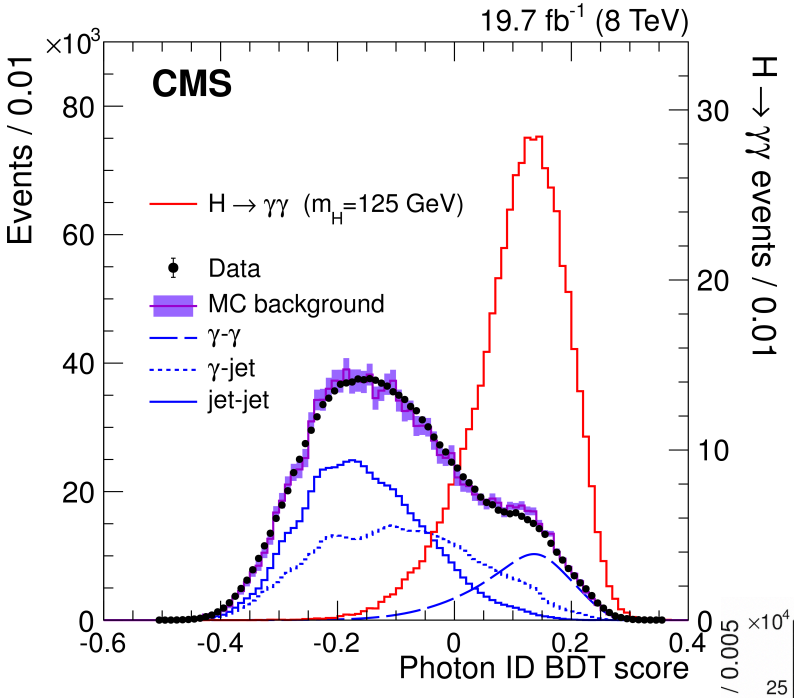
CMS Energy Uncertainty (regression)

Estimate the uncertainty on the energy measurement

ATL_PHYS_PUB_2016_015

CMS-EGM-13-001

BDT used for photon at CMS for ID (classification) and energy reconstruction (semi-parametric regression).



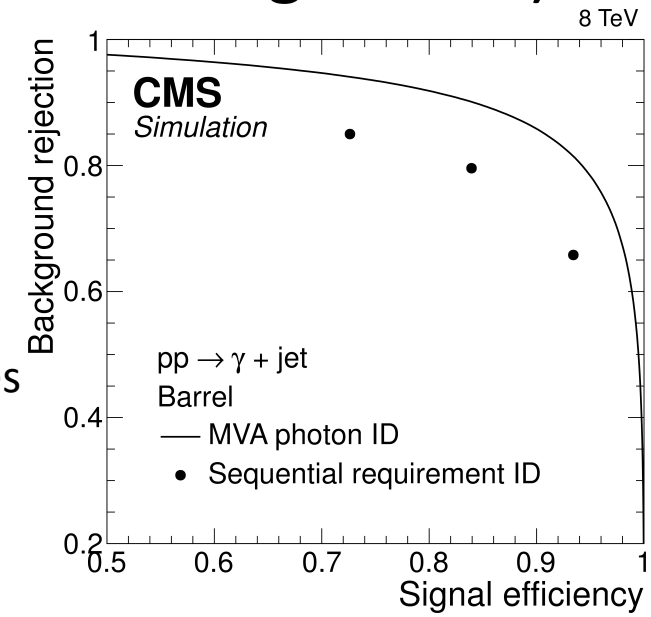
CMS Trained BDT in γ+jet MC

S=prompt γ

B=non-prompt γ

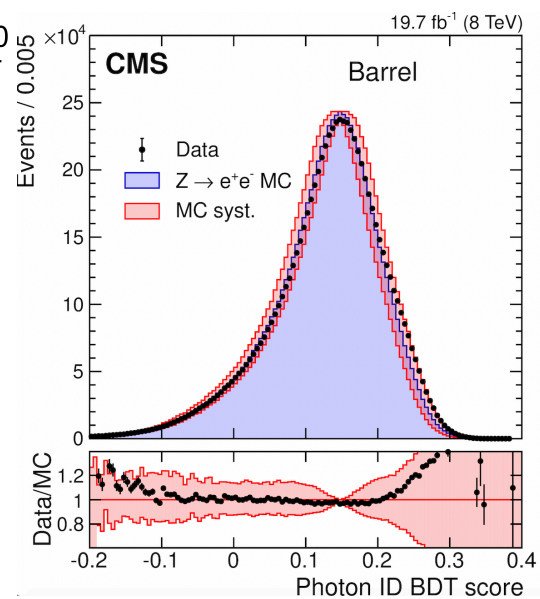
Inputs:

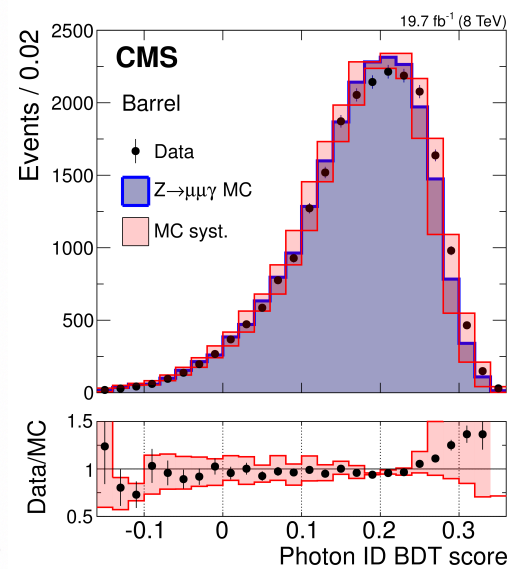
- all shower variables
- Isolation variables
- Eta, Et, etc



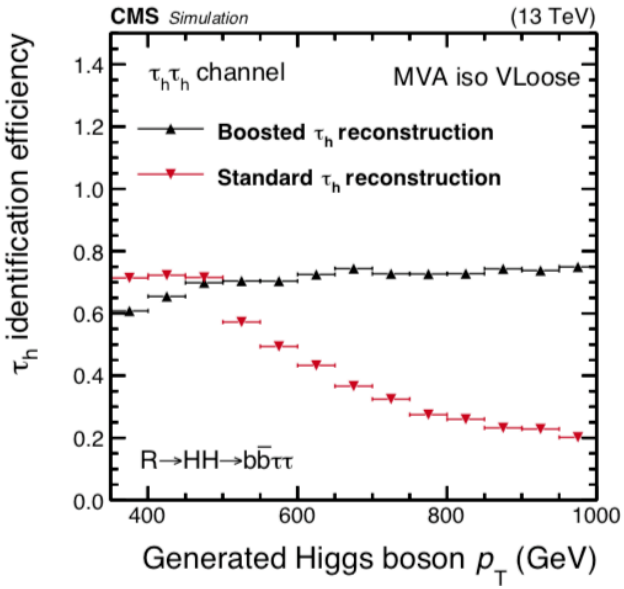
CMS also calibrates energy With a semi-parametric regression. [CMS-EGM-14-001](CMS-EGM-14-001)

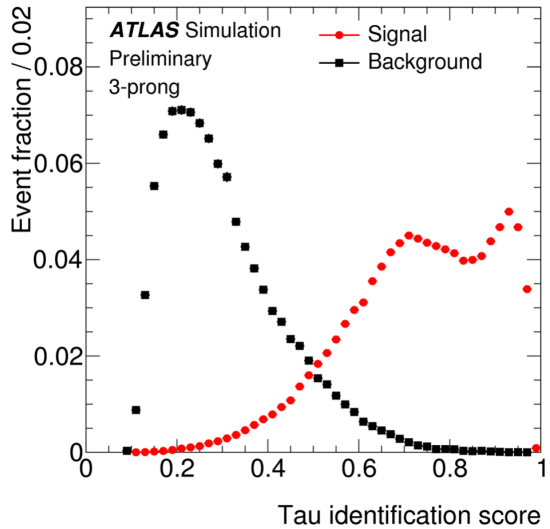
ATLAS uses BDT for photon calibration, same approach as for electron. PERF-2013-05. ATL-PHYS-PUB-2017-022.

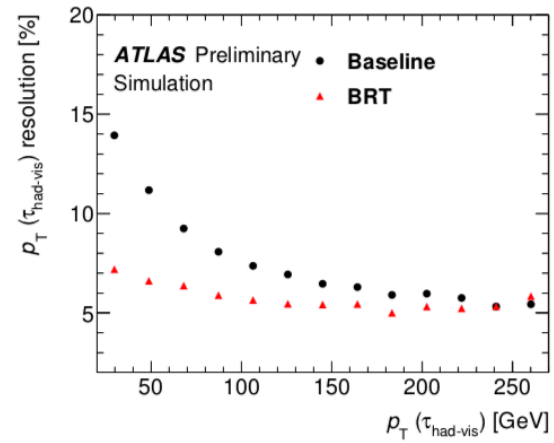




BDT for hadronic tau at CMS for ID (classification), at ATLAS for ID & energy calibration (regression).



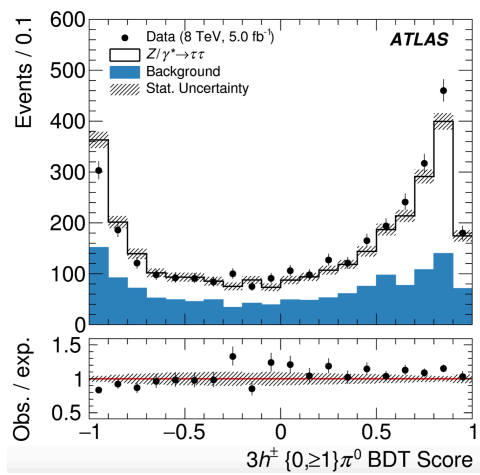




ATLAS BDT (BRT) regression improves resolution.

CMS & ATLAS each two BDTs for ID:

- tau (had) vs jet (q, gluon)
- tau (electron) vs electron
- Also boosted di-tau reco.
- [CMS-TAU-16-003]();
- [ATL-PHYS-PUB-2015-045]()

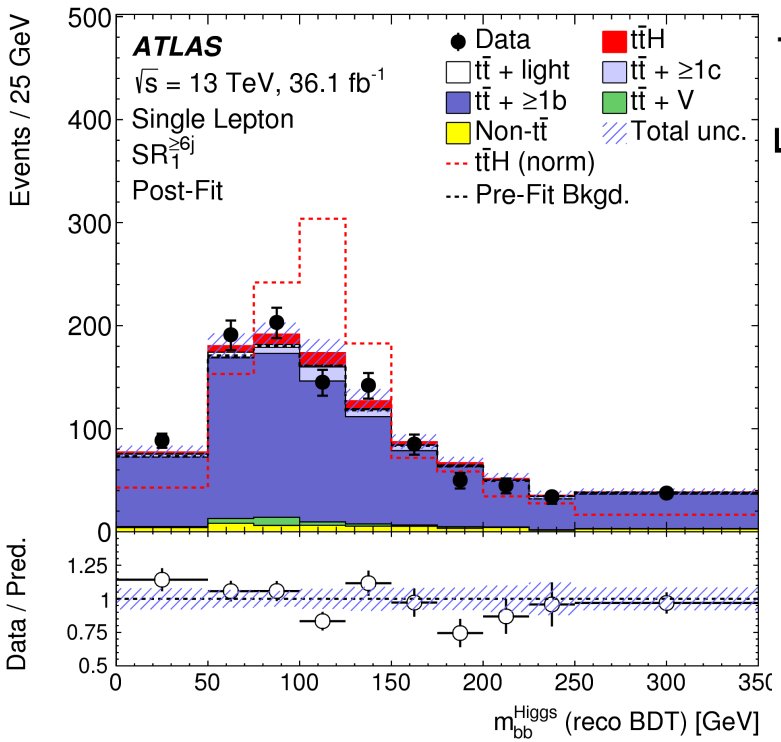


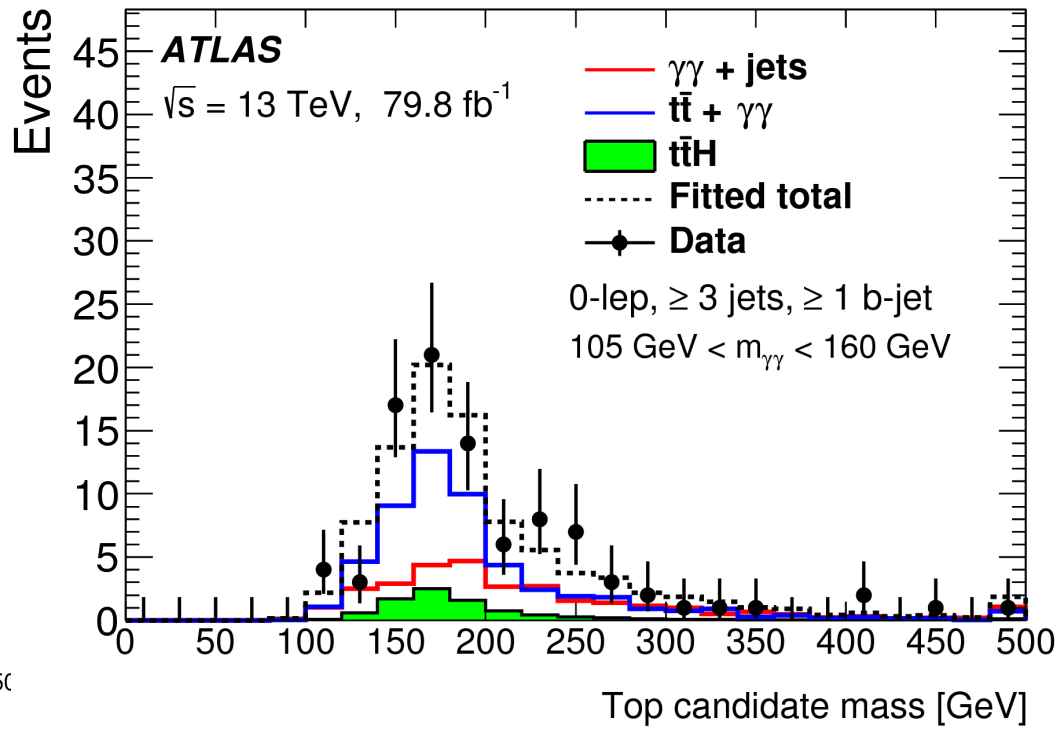
Inputs from baseline method, plus tau particle flow (using tracks for low pT), plus other calorimeter and tracking variables. [ATLAS-CONF-2017-029]()

- ATLAS differentiate different decay modes of already identified tau by counting pi^0 [PERF-2014-06]().

Tau group was first in ATLAS to introduce a BDT ID at trigger level.

Example of BDT classification to choose which jets to group to reconstruct Higgs or top quark candidate mass.





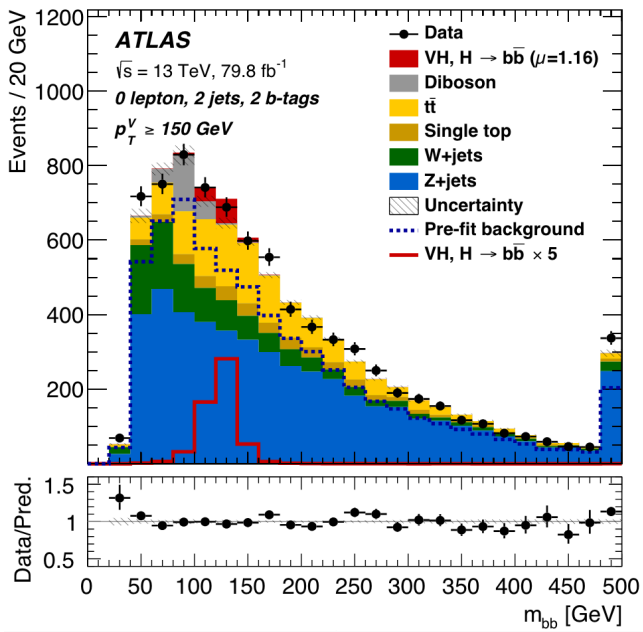
ATLAS ttH(bb) **H(bb) reconstruction**
Classification BDT choosing the right pairings of jets to form the H(bb) candidate in the resolved categories.

Right matching in about 30-50% of Higgs signal events. arXiv:1712.08895

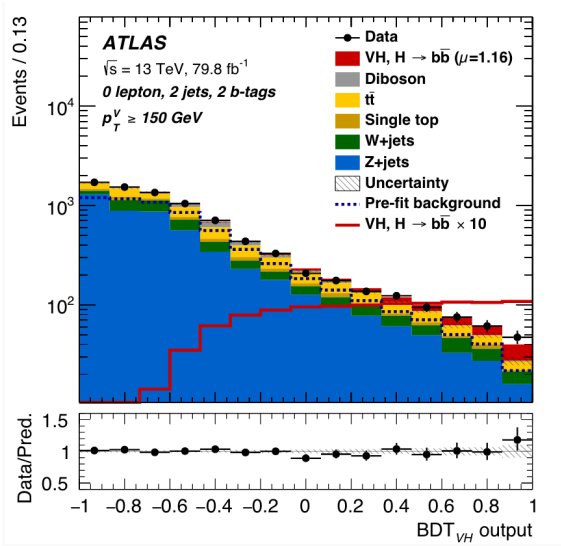
ATLAS ttH(gamma gamma**)**
top quark mass reconstruction.
Classification BDT to choose which jets more likely to form the top quark candidate.

arXiv:1806.00425

BDT classification for S/B separation adds ~25% in sensitivity on top of m(bb) alone in ATLAS VH(bb).



Variable	0-lepton	1-lepton	2-lepton
p_T^V	$\equiv E_T^{miss}$	×	×
E_T^{miss}	×	×	
$p_T^{b_1}$	×	×	×
$p_T^{b_2}$	×	×	×
m_{bb}	×	×	×
$\Delta R(b_1, b_2)$	×	×	×
$\lvert\Delta\eta(b_1, b_2)\rvert$	×		
$\Delta\phi(V, bb)$	×	×	×
$\lvert\Delta\eta(V, bb)\rvert$			×
m_{eff}	×		
$\min[\Delta\phi(\ell, b)]$		×	
m_T^W		×	
$m_{\ell\ell}$			×
$E_T^{miss}/\sqrt{S_T}$			×
m_{top}		×	
$\lvert\Delta Y(V, bb)\rvert$		×	
	Only in 3-jet events		
$p_T^{jet_3}$	×	×	×
m_{bbj}	×	×	×



BDT using high level variables in ATLAS H->bb and VH observations.
[Phys.Lett. B786 (2018) 59-86]

In CMS a BDT is used to classify H->mumu from bkg, split in analyses regions based on this, then set limit on m(μμ).

HIG_17_019

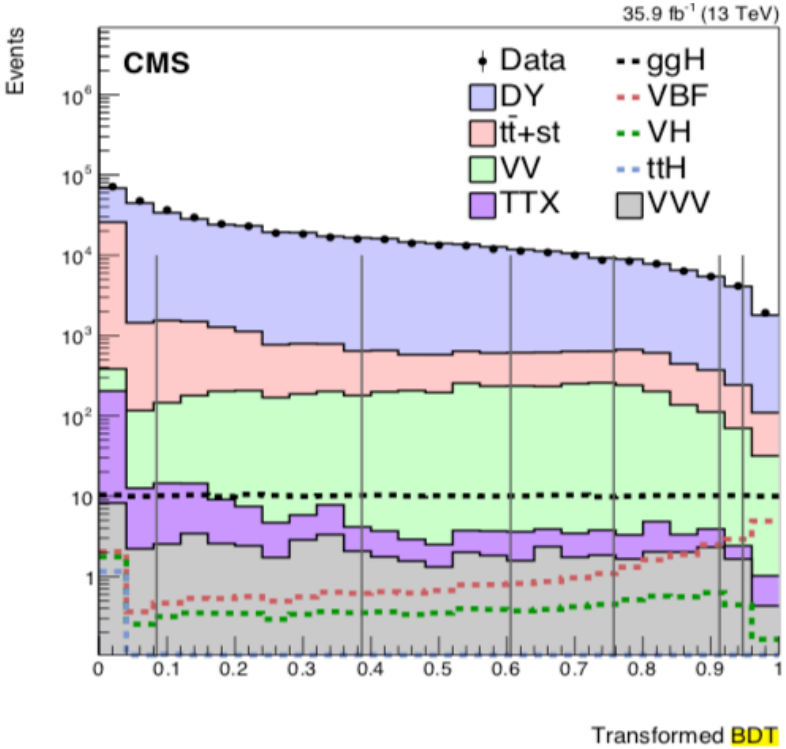
A greedy category optimisation based on a single tree.

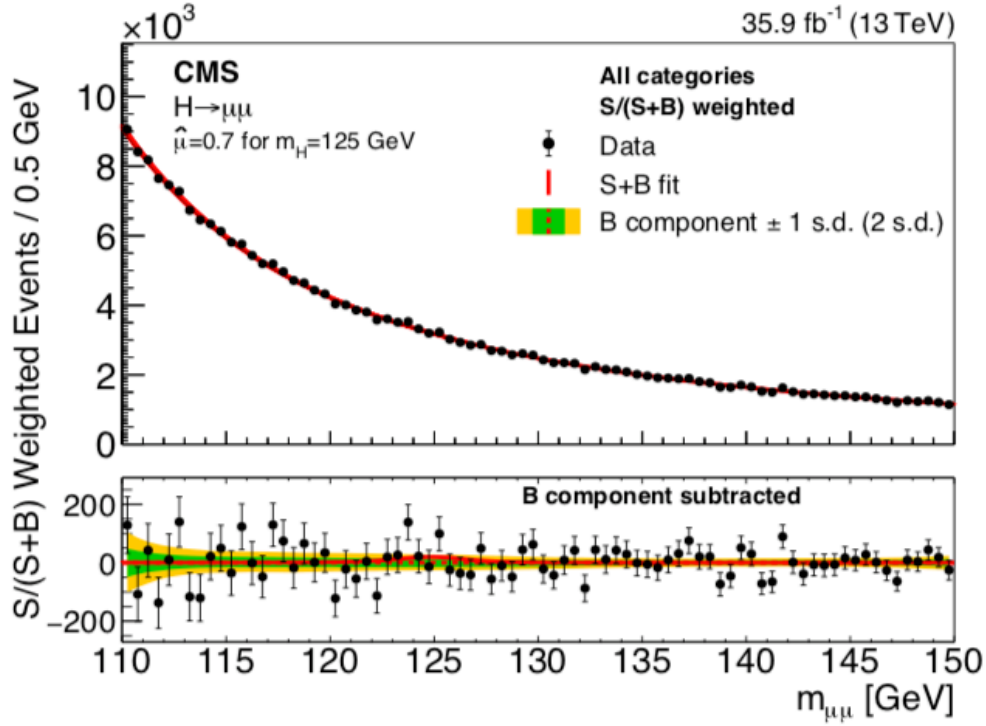
Using input variables largely uncorrelated to m(μμ) to later optimise and fit on m(μμ).

Choose variables to discriminate from various backgrounds.

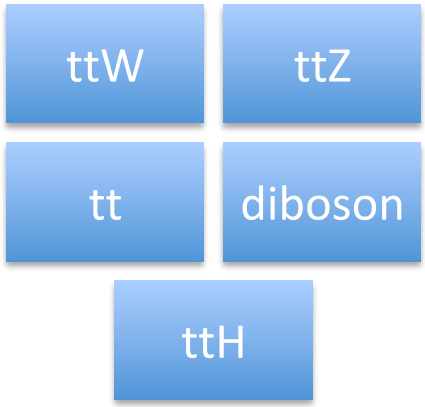
Vertical lines on left represent the various categories as a function of BDT.

Right is all categories weighted by the expected signal + background.

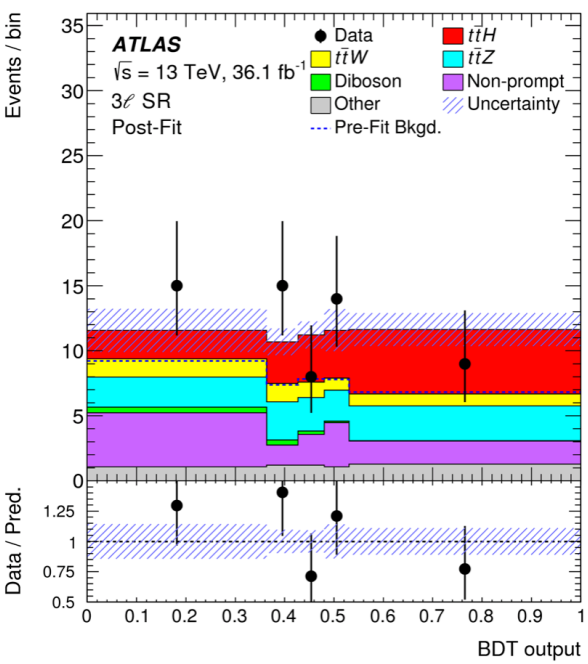


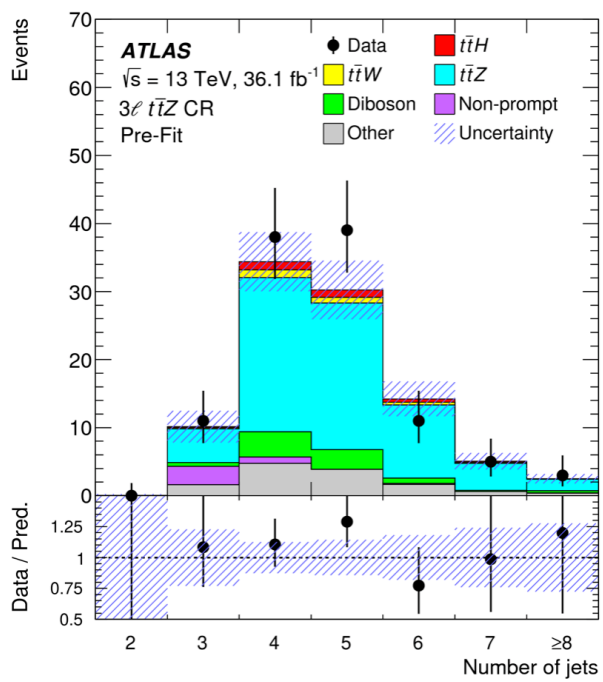


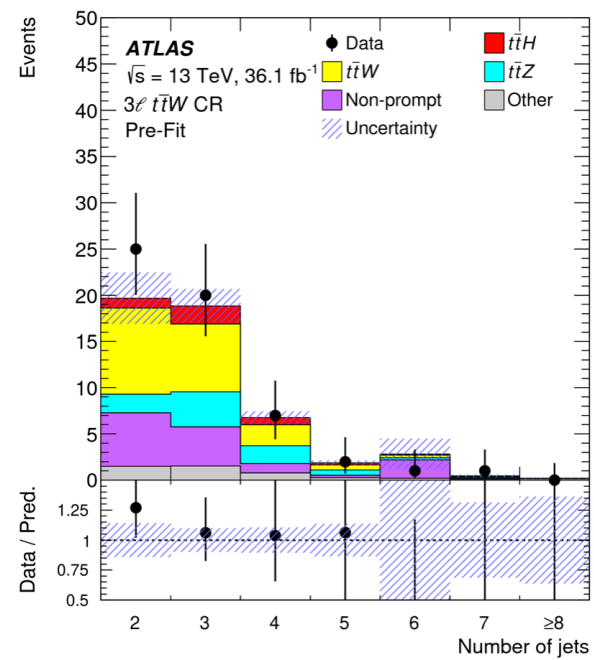
A multinomial BDT in XGBoost to separate signal from 4 different background processes in ttH(3-lep).



About 25 Inputs:
Lepton properties; Jet properties; Angular distances; MET.
arXiv: 1712.08891

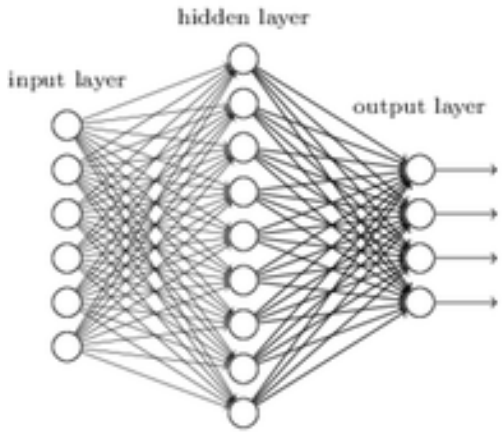




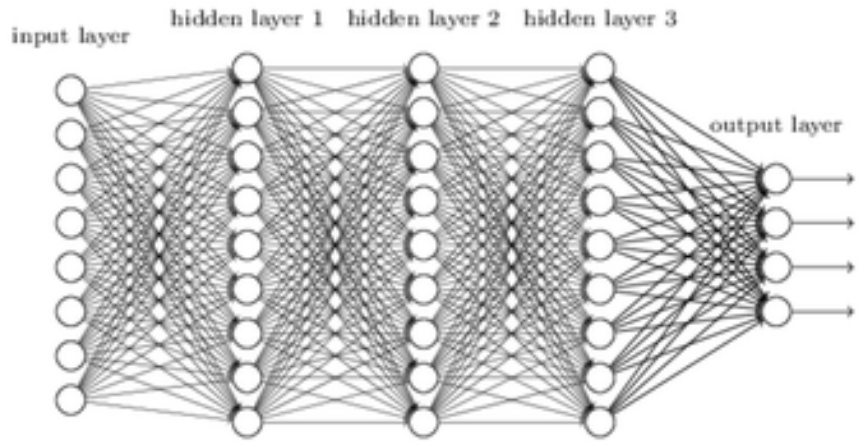


Neural networks from shallow simple structure in TMVA to complex structures in Keras.

"Non-deep" feedforward neural network



Deep neural network

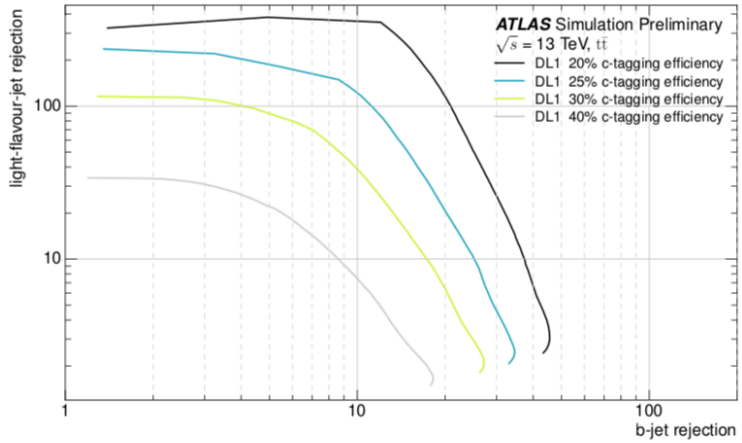


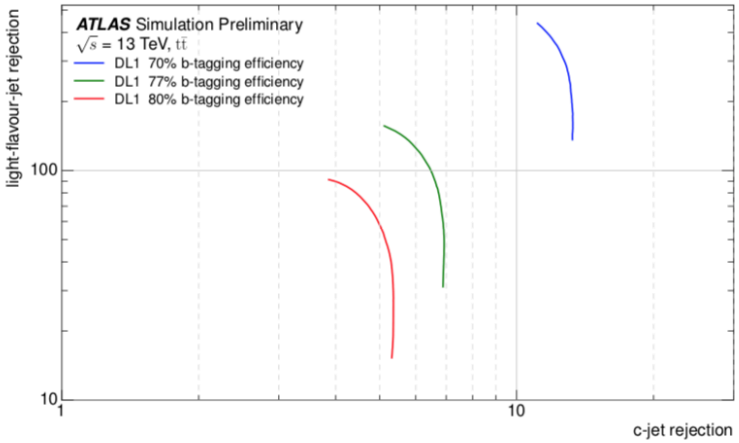
High level engineered variables

Low level reconstruction variables

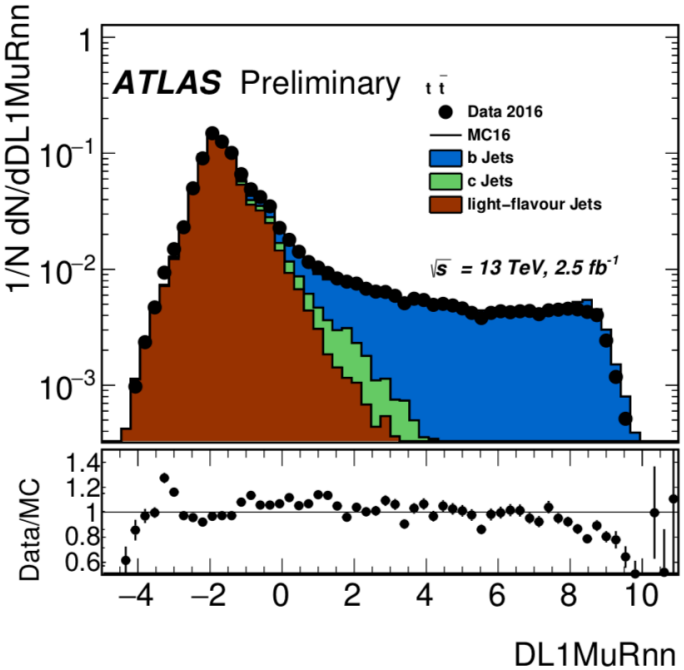
Multi-layered networks already used in Run-1 in TMVA and JETNET.
Deep networks allow to learn more complex functions -> typically better performance.
Requires more (simulated) data. Requires more computing power.

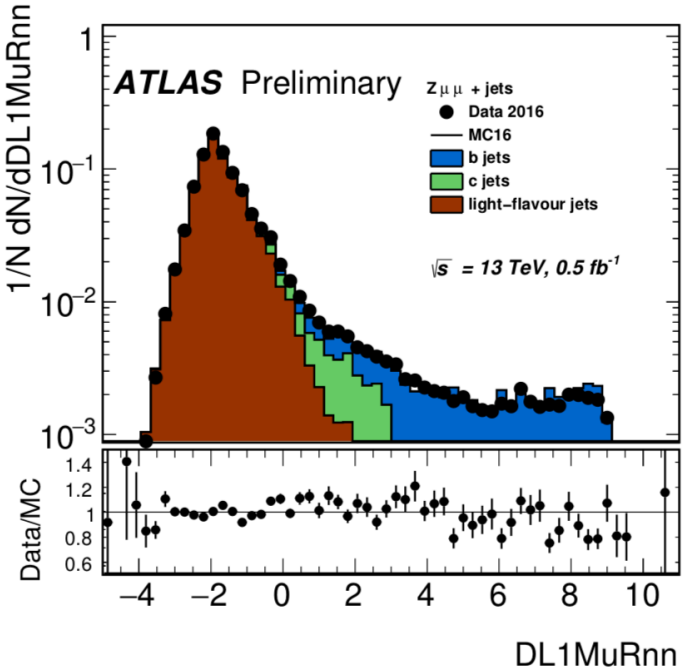
Deep neural network (DNN) vs BDT for b-tagging. For b-tagging similar performance, opens R&D.





ATL_PHYS_PUB_2017_013



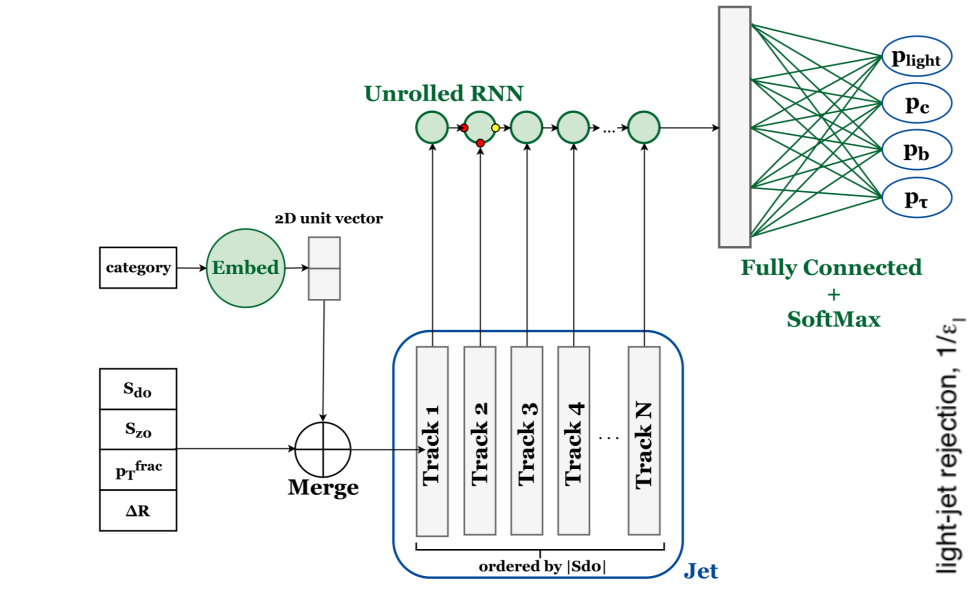


Same inputs, but at lower level.

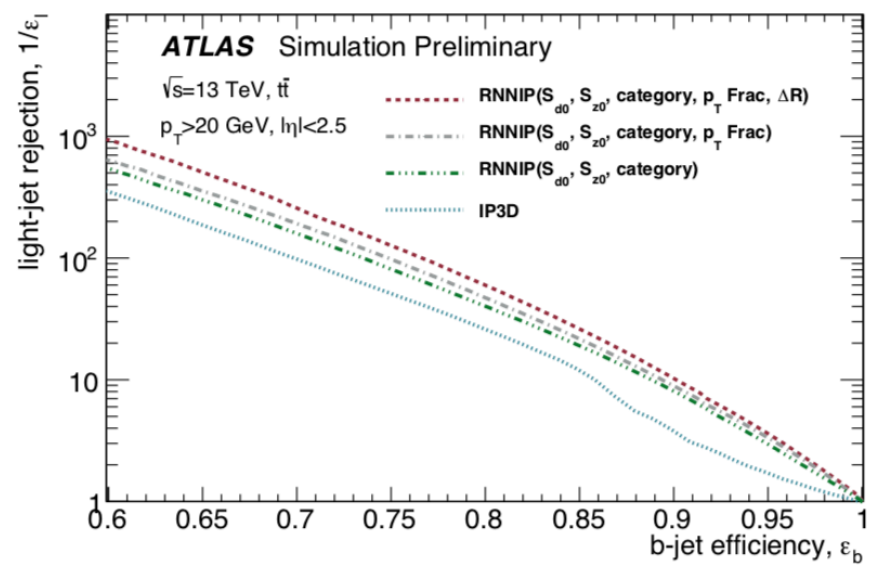
Three outputs as probabilities of b, c, l (no tau).

Can be used for c-tagging too.

Recurrent neural networks (RNN) used in b-tagging jet identification (classification) using track correlations.



ATLAS b-tagging with RNN
conf proceeding



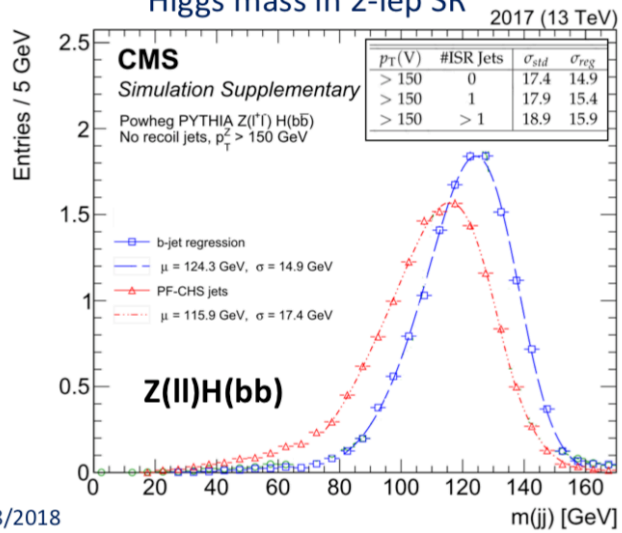
Output of RNN is one of inputs of DNN b-tagging.

Exploit the track-to-track correlation present in b-jets, but not in light jets.
Tracks ranked by significance are introduced sequentially to the RNN.
Multi-nominal, with four outputs: probabilities of b, c, l and tau..

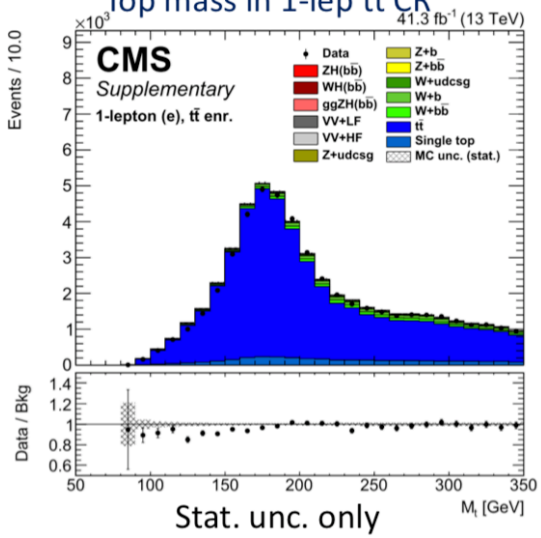
CMS uses two DNNs in VH(bb) observation. b-jet energy calibration (regression) and S/B separation (classification).

Using same high level variables from BDT and adding more low level variables (paper, talk).

Higgs mass in 2-lep SR

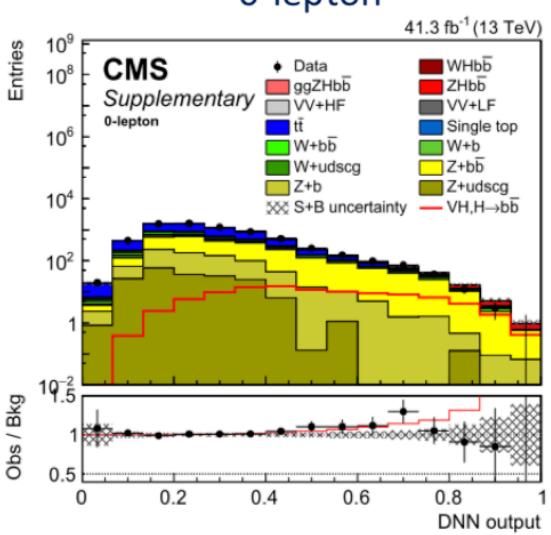


Top mass in 1-lep tt CR

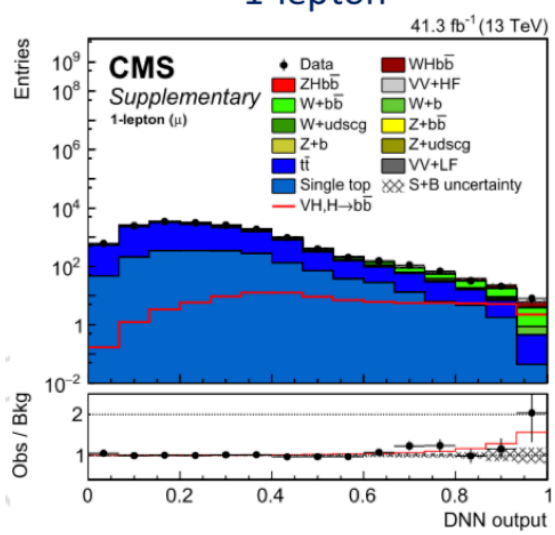


28/08/2018

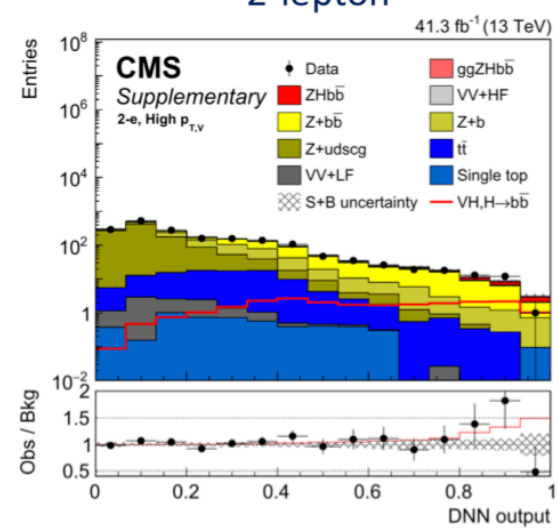
0-lepton



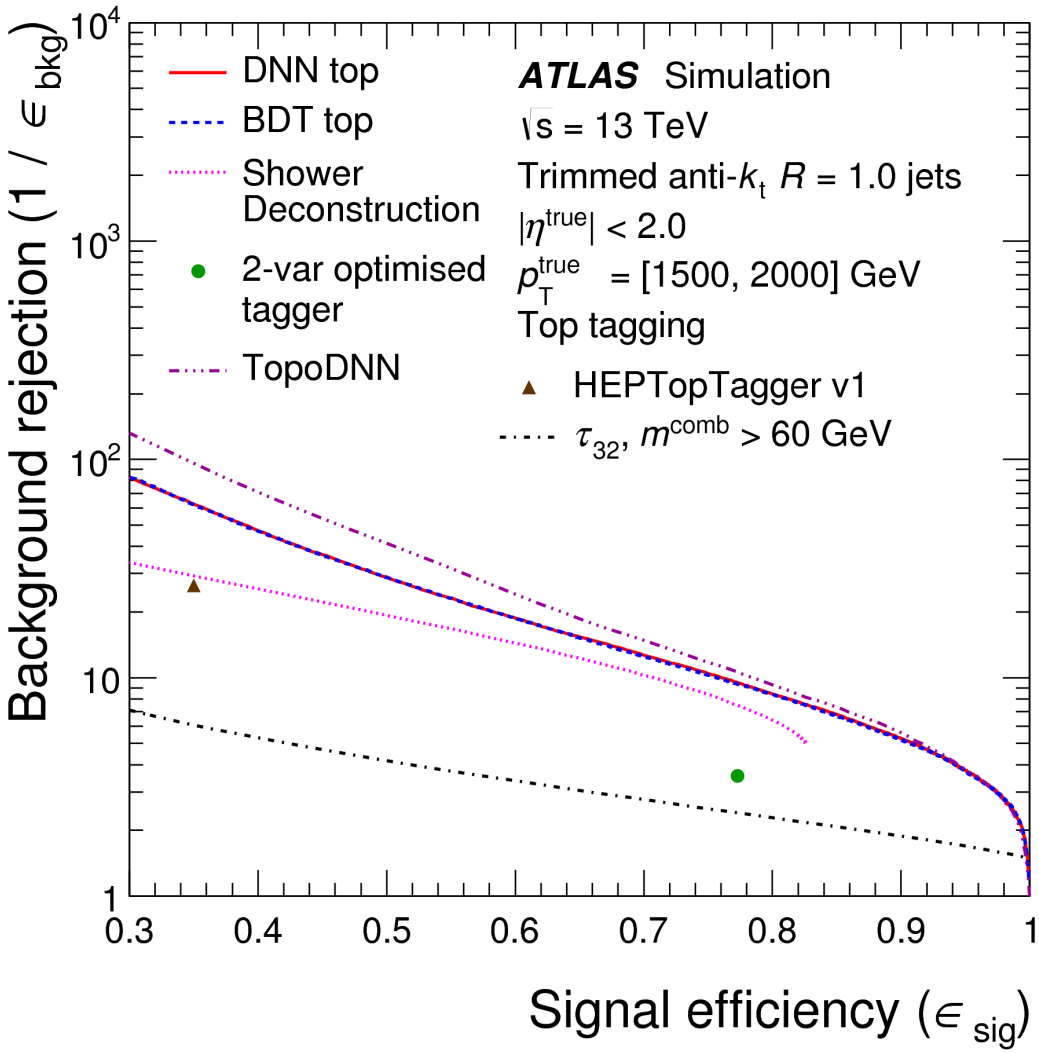
1-lepton



2-lepton

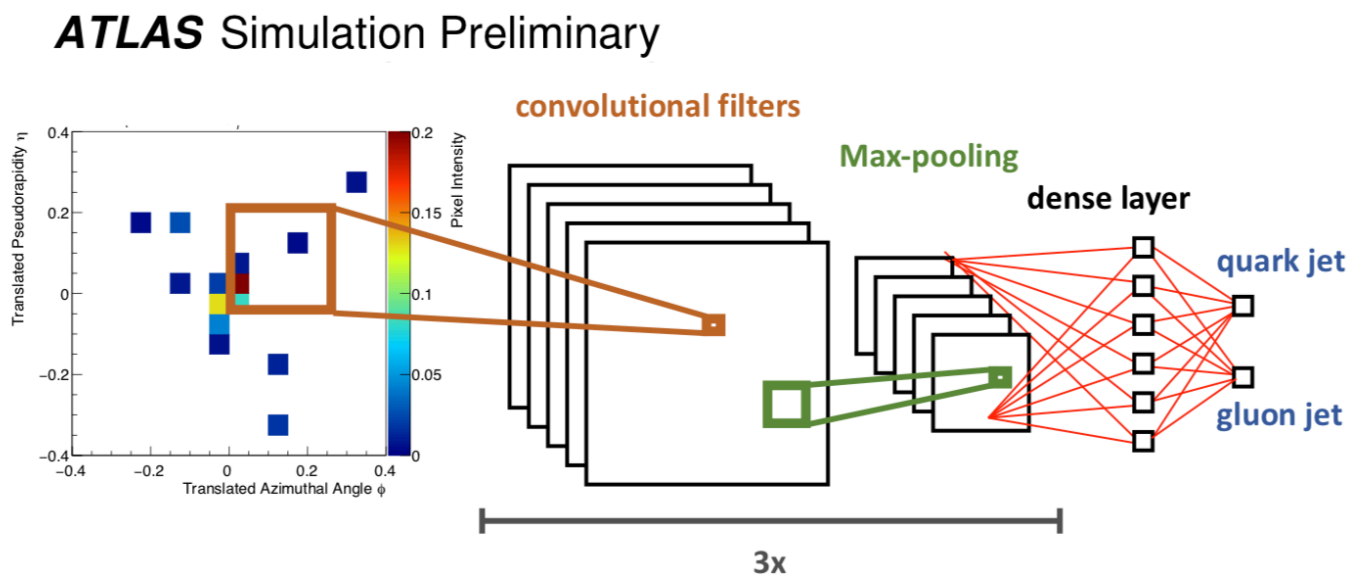


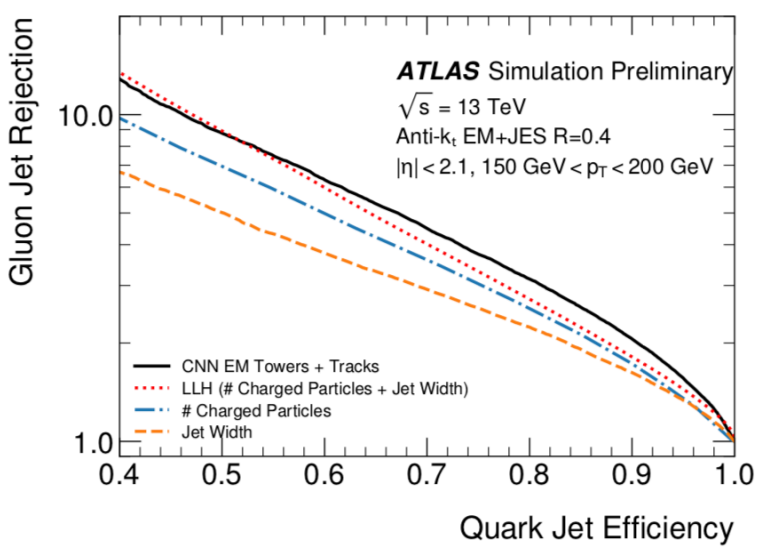
Using only high level variables, no performance difference between BDT and DNN. But adding also lower-level information improves performance (TopoNN).

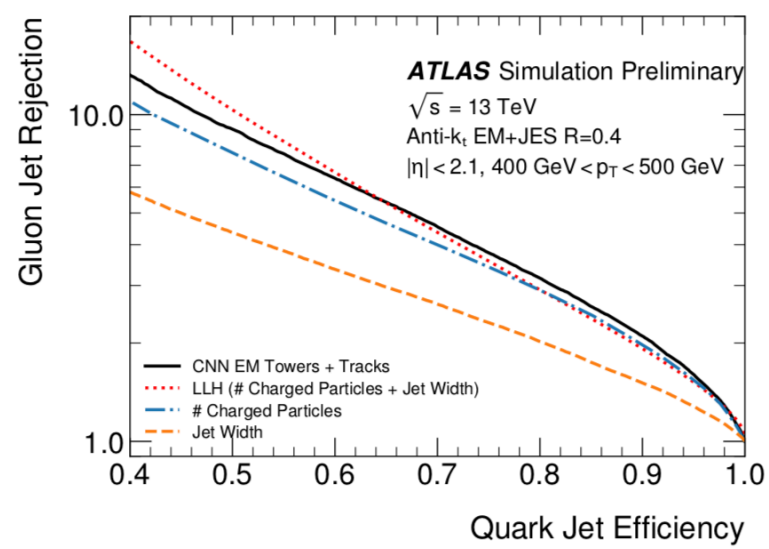


Example from
fat jet tagging
for W boson
or top quark
ATLAS arXiv 1808.07858

Convolutional neural networks (CNN) classify jet images, like in the quark/gluon tagger (ATL-PHYS-PUB-2017-017).







Convolutional neural networks assume (translational) invariances as found in images. Images are scanned with (learned) filter matrices.

CMS uses boosted decision trees (TMVA, XGBoost) and (deep) neural networks (TMVA, Keras).

Object	Identification	Calibration
Electrons	BDT	BDT
Photons	BDT	BDT
Tau leptons	BDT	non-MVA
b-jets	DNN	DNN

Higgs analysis	Sig vs Bkg Separation
H -> gamma gamma	(mass factorised) BDT
VH(bb)	DNN
ttH(bb) 1-lepton	DNN (with MEM)
ttH(bb) 2-lepton	BDT
ttH(WW)	BDT
H -> mumu	BDT
HH ->bbgg	BDT
HH -> bbbb	BDT

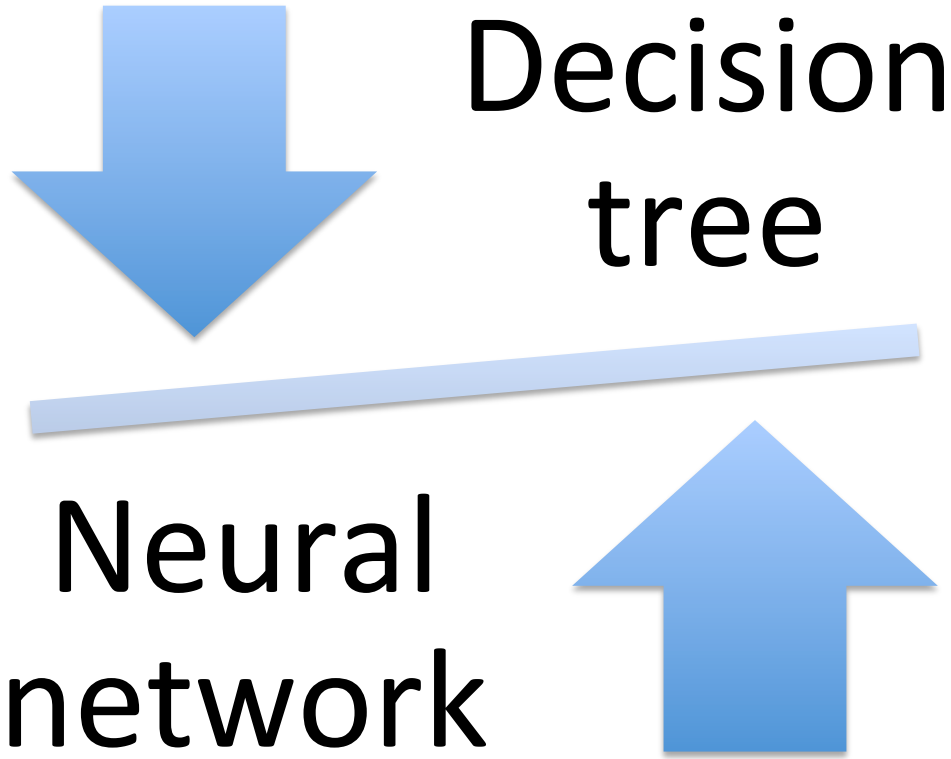
ATLAS uses boosted decision trees (TMVA, XGBoost) and (deep) neural networks (TMVA, Keras).

Object	Identification	Calibration
Electrons	non-MVA	BDT
Photons	non-MVA	BDT
Tau leptons	BDT	BDT
b-jets	MVA, DNN, RNN	non-MVA

Higgs analysis	Sig vs Bkg Separation
H->WW	BDT
VH(bb)	BDT
ttH(bb)	BDT for reco, BDT for S vs B
ttH(multilepton)	BDT for S vs B in most channels
ttH(gamma gamma)	BDT
BSM H+ -> tb	BDT
HH->bbtautau	BDT
H->mumu	BDT

Page ATLAS Public results

Decision tree or neural network?
Often similar performance, choice from ease of use.



Machine learning has many other methods that we only start to exploit at CMS/ATLAS:

- Generative models (fast MC simulation for calorimeter energy)
- Auto-encoders (anomaly detection)
- Adversarial training (a S/B classifier independent of particle mass)

Conclusion: Machine learning used and developed further for many aspects of Higgs boson analyses.

Combined performance

- Object identification
- Object E/p calibration

Physics analysis

- Separate Signals and several backgrounds
- Reduction of systematic uncertainties

TMVA continues to be used, but modern tools become popular.
BDT -> XGBoost (faster; classify several classes at the same time).
NN -> DNN (jet tagging; RNN for all tracks in a jet; CNN for jet images).
Modern (Python) tools need to be adapted to our (C++) frameworks.

Stay tuned for end of Run-2 papers and new ML techniques!