Using Machine Learning for Jet Measurements

Hannah Bossi (Yale University)

RHIC/AGS User's Meeting

June 9th, 2021 (Remote)

Wright Laboratory

Yale

Outline

What is ML and why is it useful for jets?

How is ML currently being used for jet measurements?

Ongoing challenges

Highlighting the future along the way!

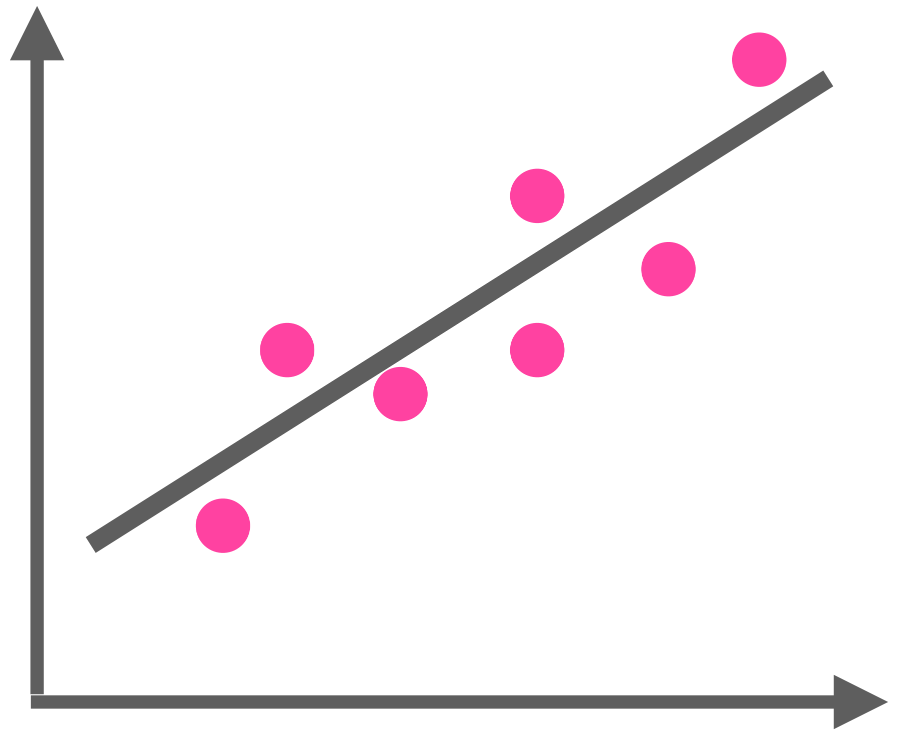
This is a brief and biased overview! Many great conferences and workshops dedicated to this topic! For example, check out ML4jets!

What is Machine Learning?

ML is a subset of artificial intelligence (AI) in which algorithms can be used to imitate human learning, i.e. gradually improving accuracy over time.

Supervised Learning

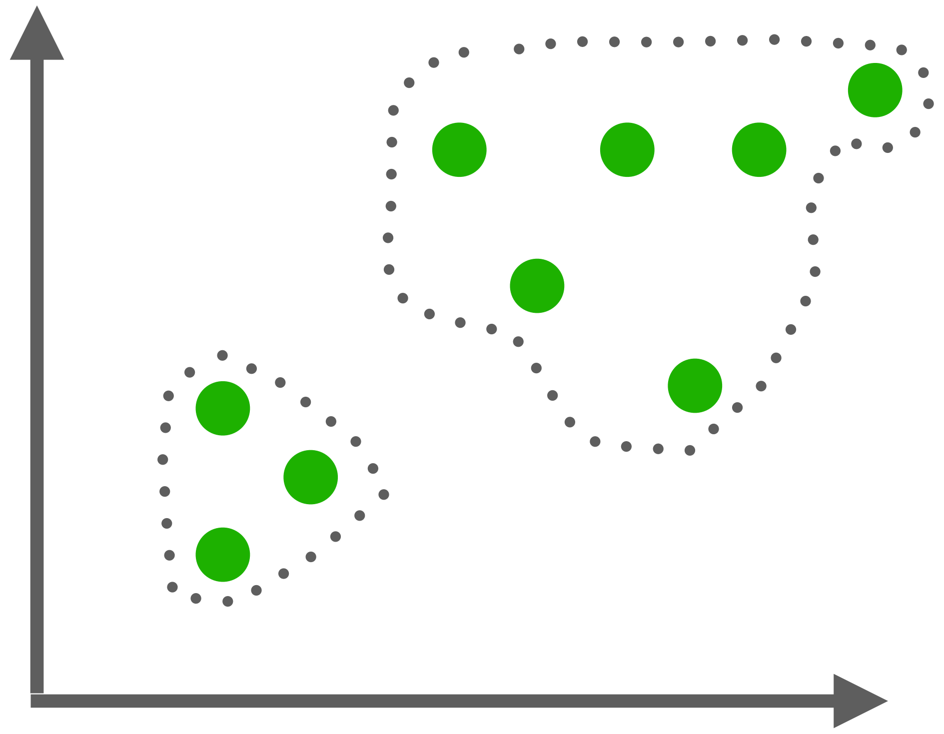
Algorithm learns from a labeled set of “true values”.



Driven by the Task

Unsupervised Learning

Algorithm finds structure in the data without knowing the desired outcome.



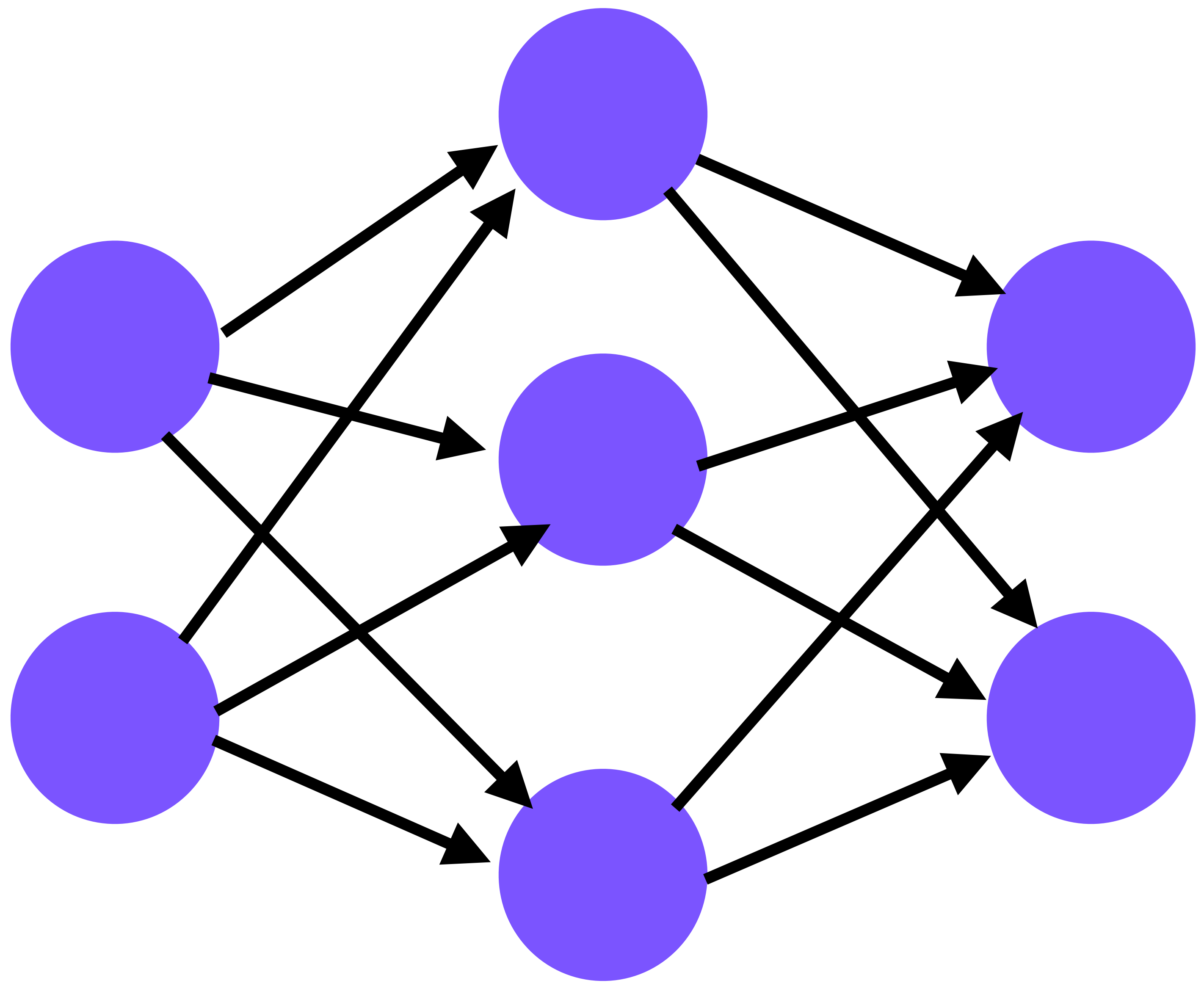
Driven by the Data

Reinforcement Learning

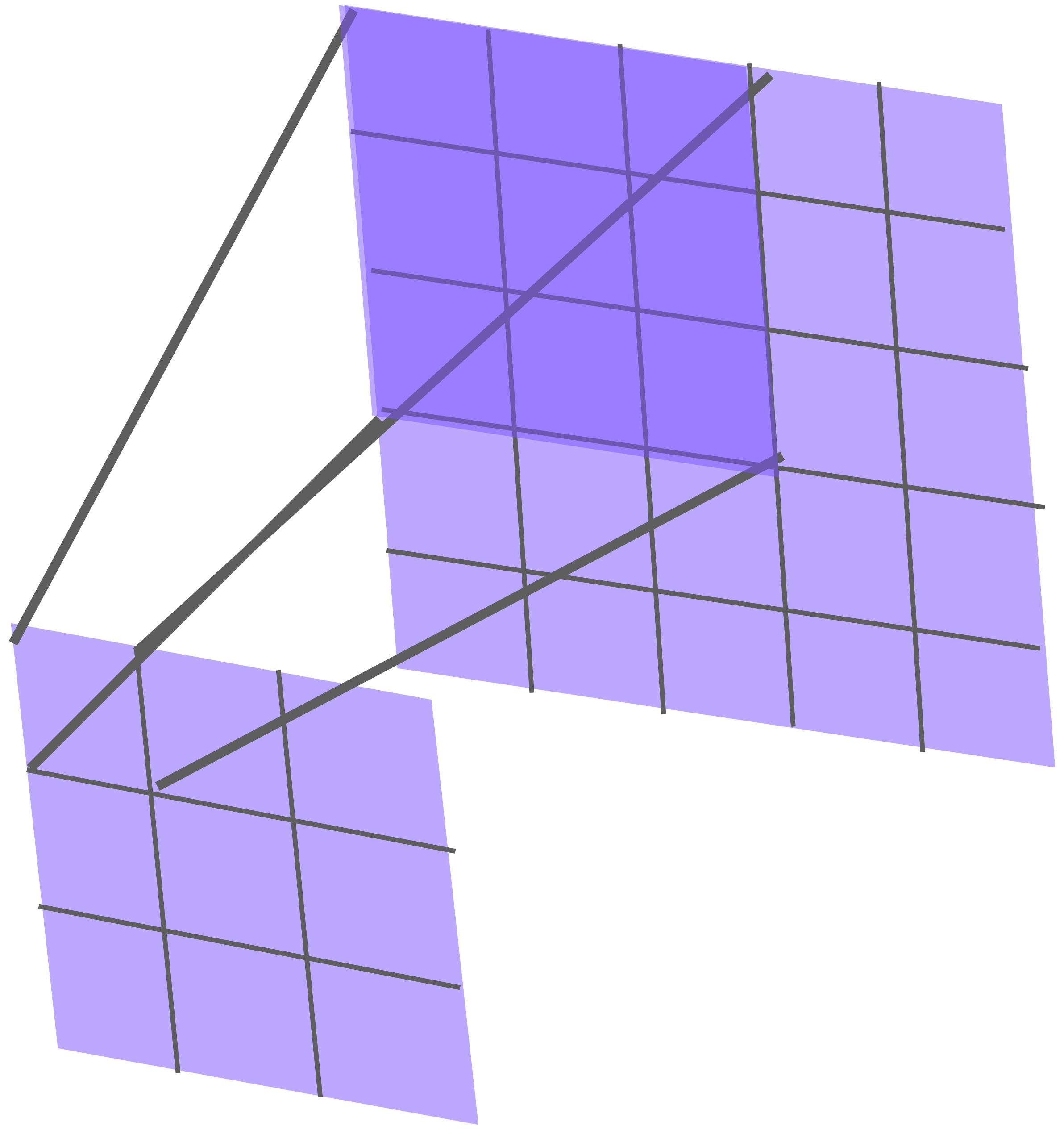
Algorithm learns in a reward based system to determine a series of actions.

Driven by the Reward

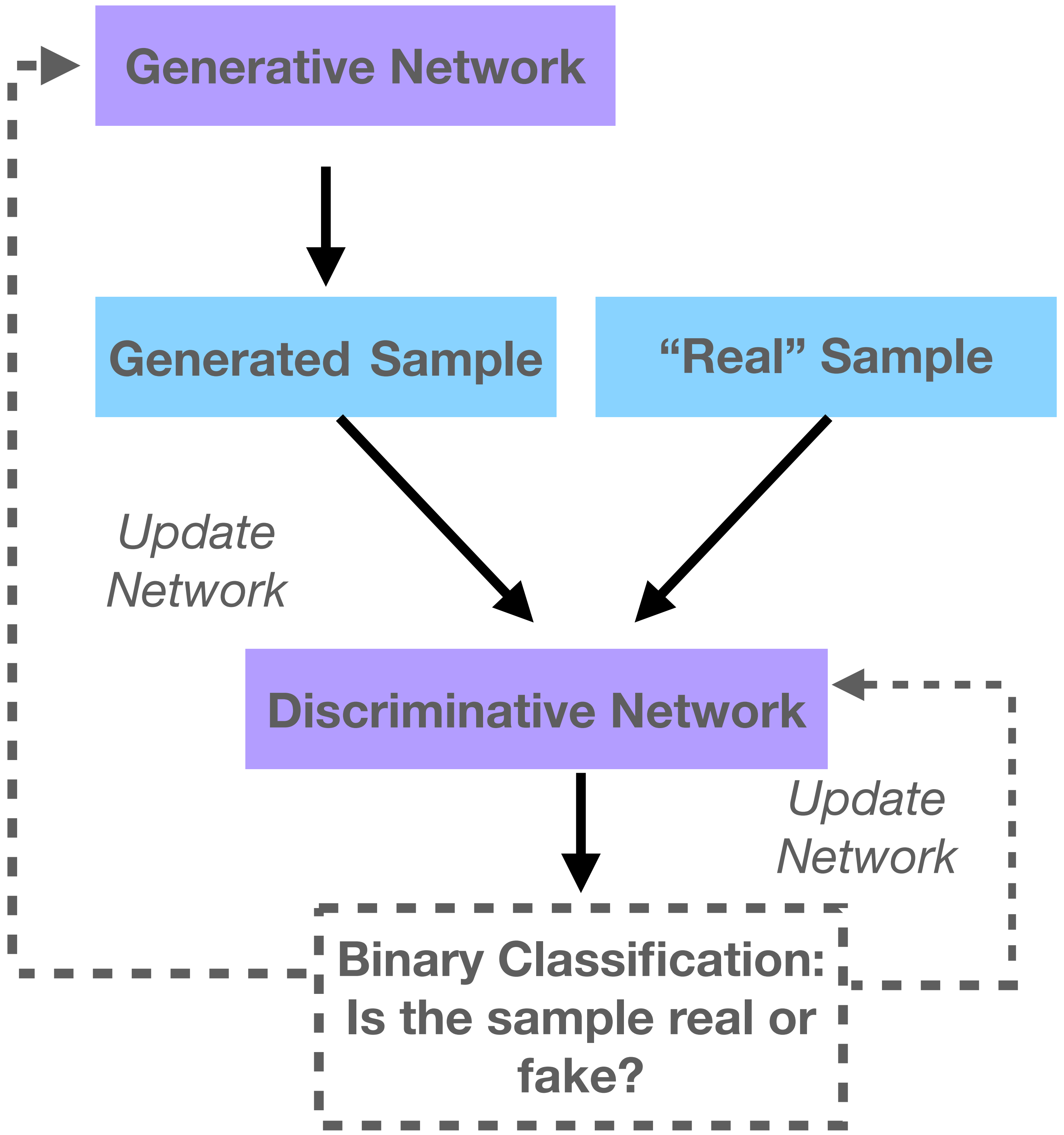
Different algorithms for different problems!



(Shallow or Deep) Neural Networks



Convolutional Neural Networks (CNNs)



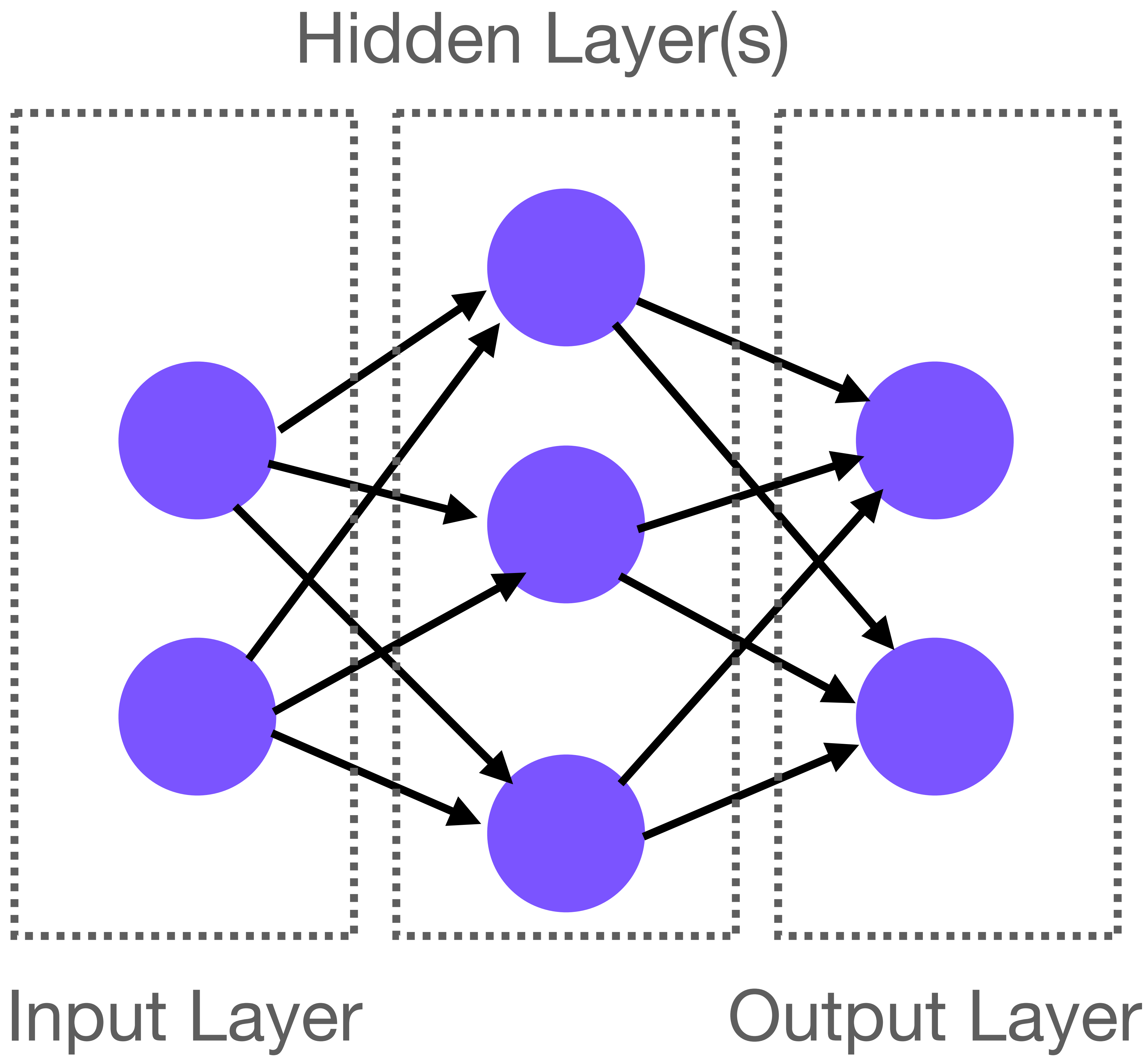
Generative Adversarial Networks (GANs)

Neural Networks

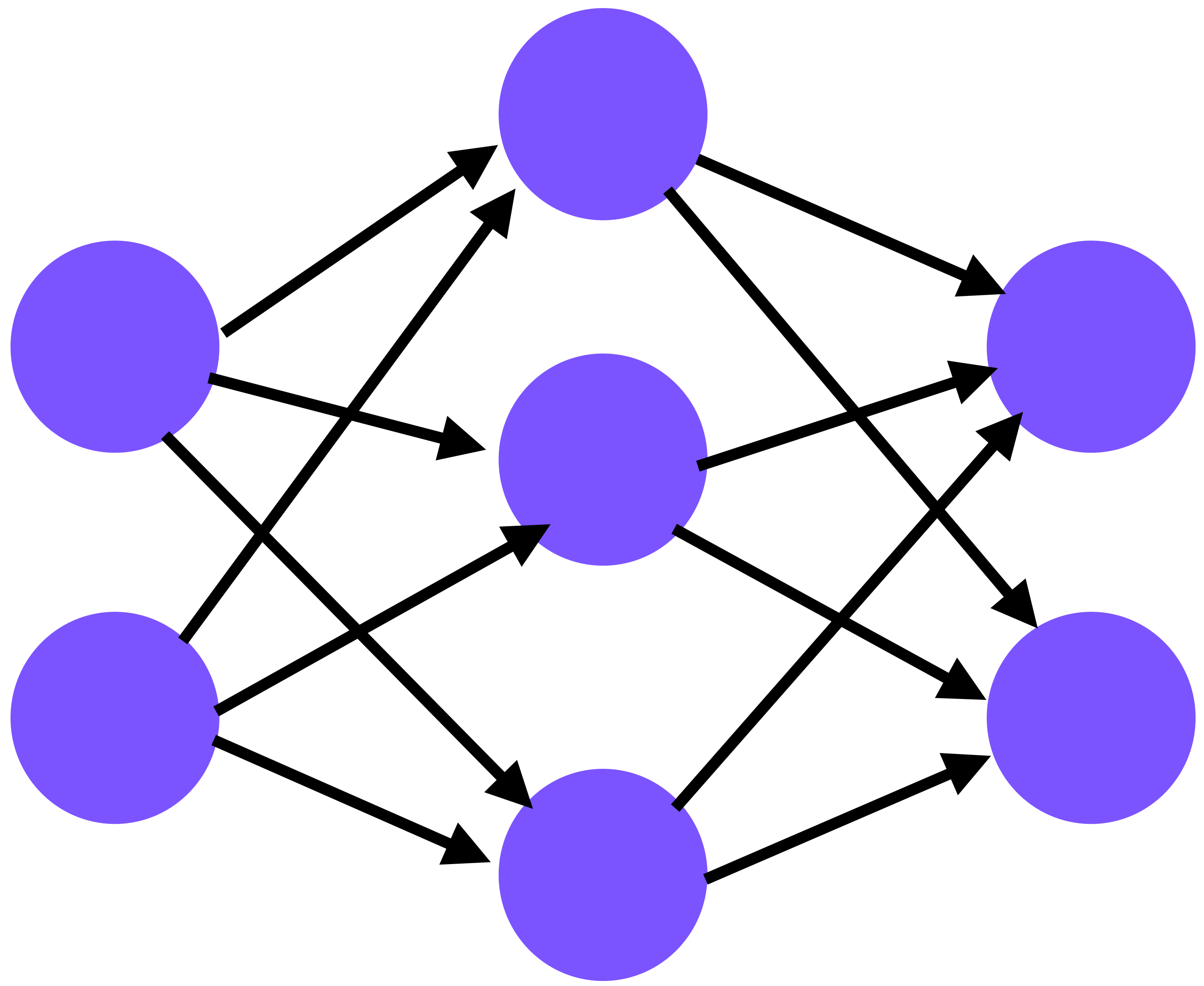
Flow of information happens between nodes.

Connections between nodes have weights which reflect the relative importance of different features.

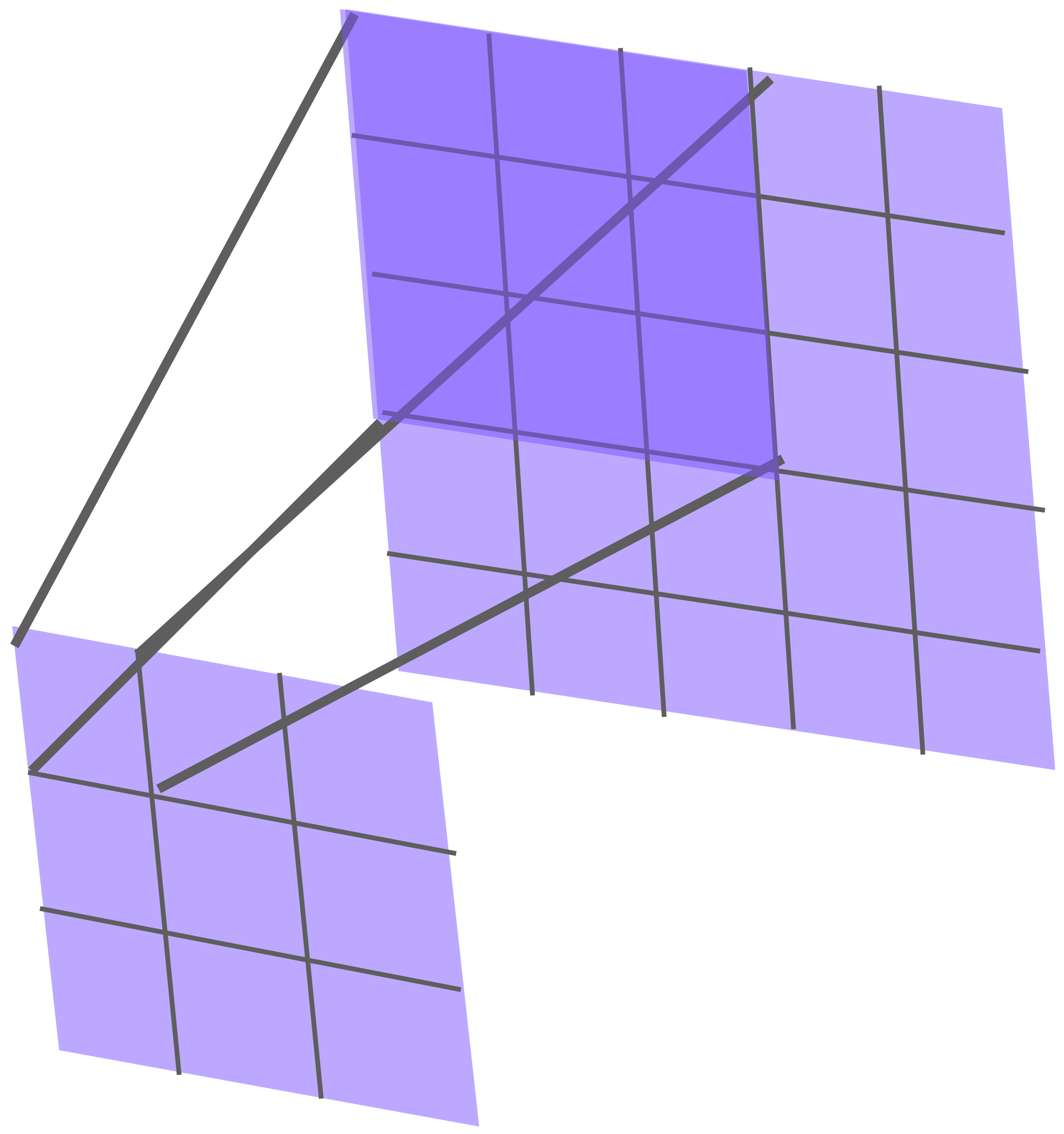
In training we seek to learn the set of weights which minimize the total error of the network.



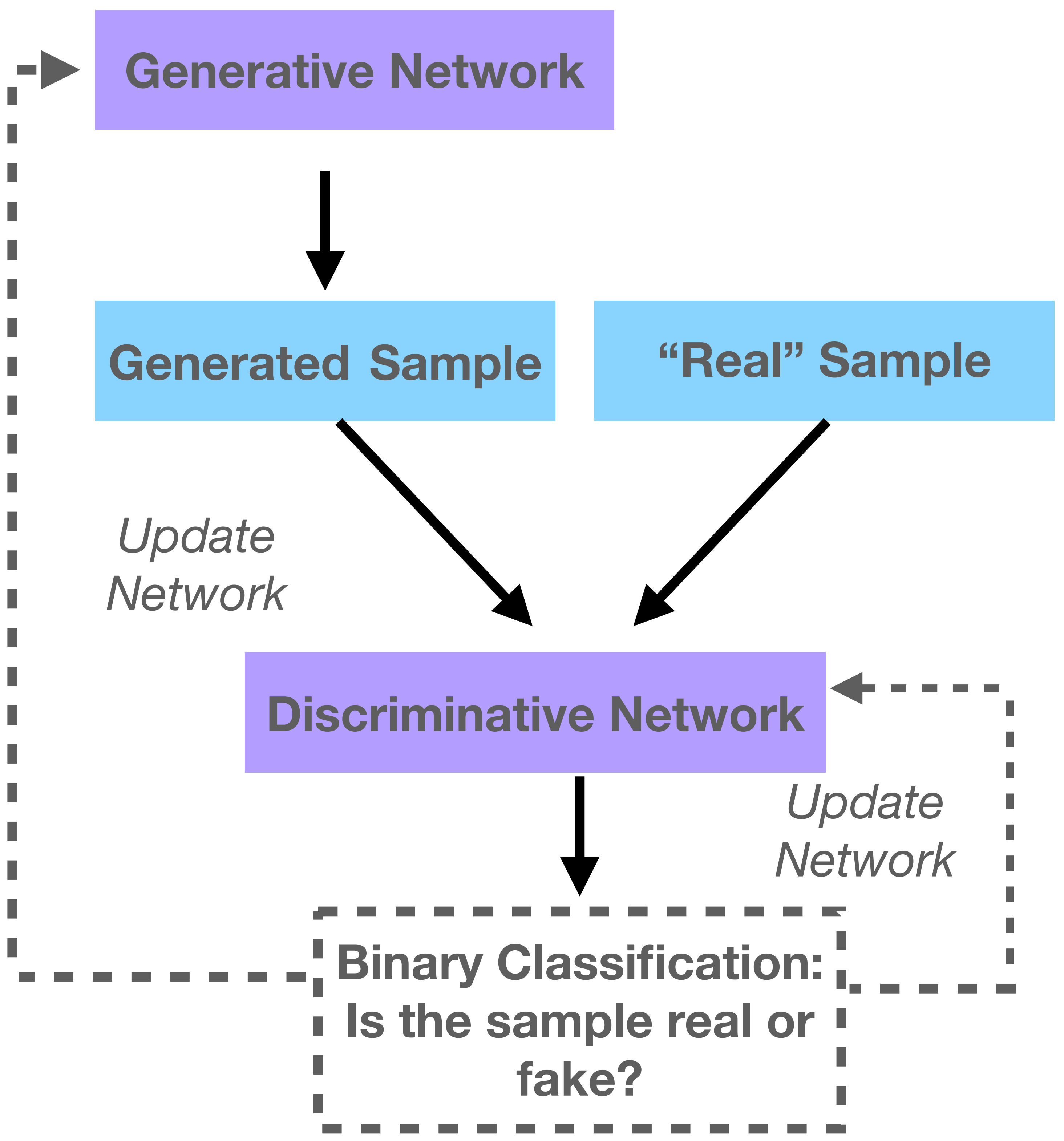
Different algorithms for different problems!



(Shallow or Deep) Neural Networks → *Great for making predictions!*

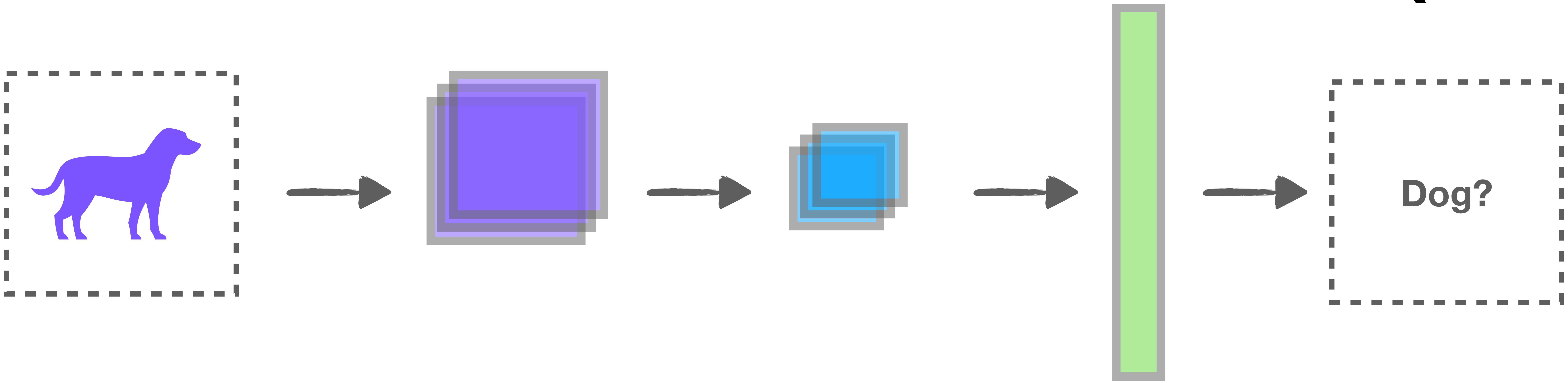


Convolutional Neural Networks (CNNs)



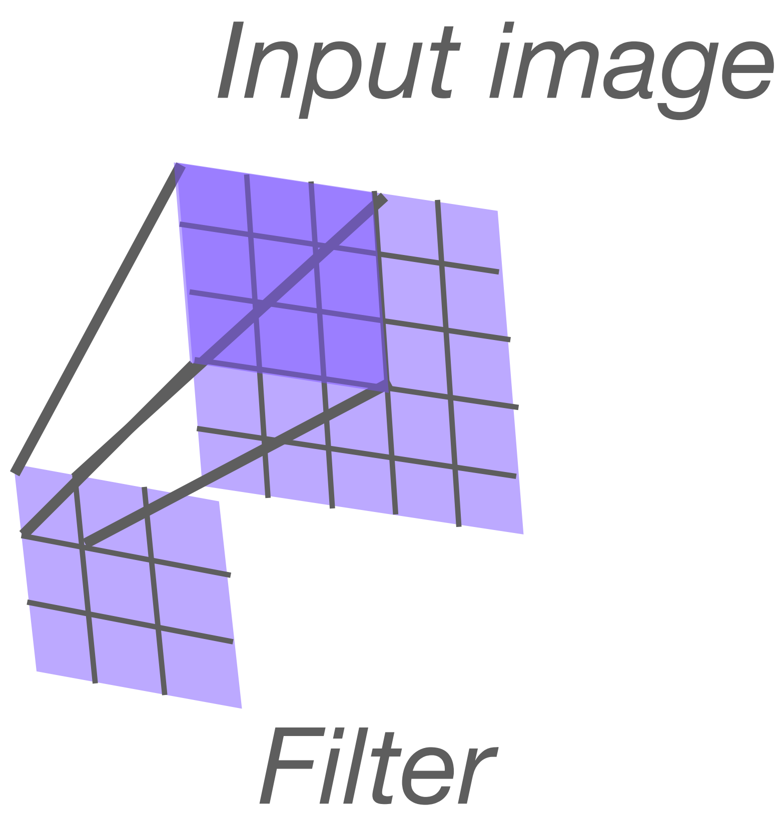
Generative Adversarial Networks (GANs)

Convolutional Neural Networks (CNNs)

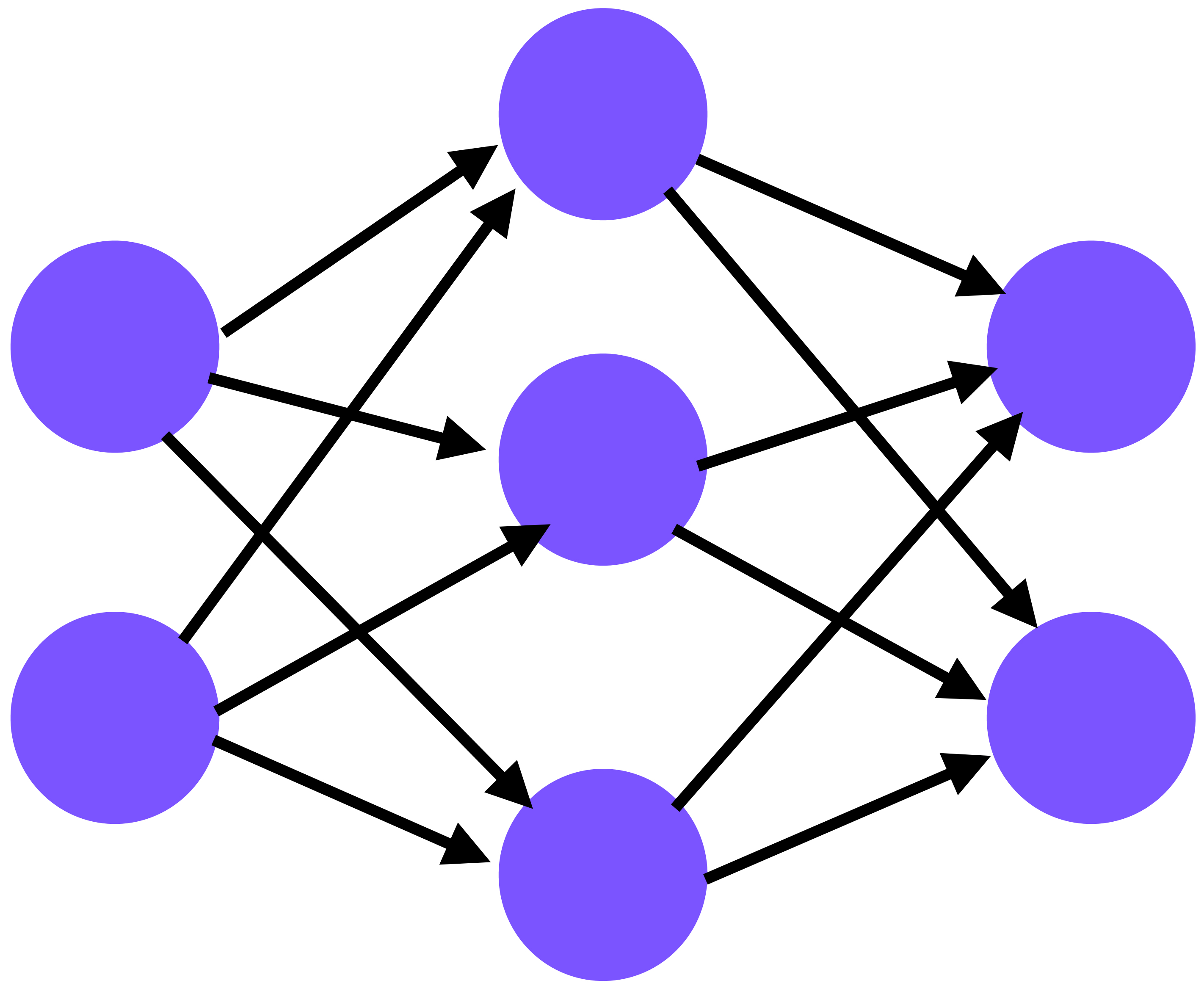


Input Layer | Convolution Layer | Pooling Layer | Fully Connected Layer | Output Layer

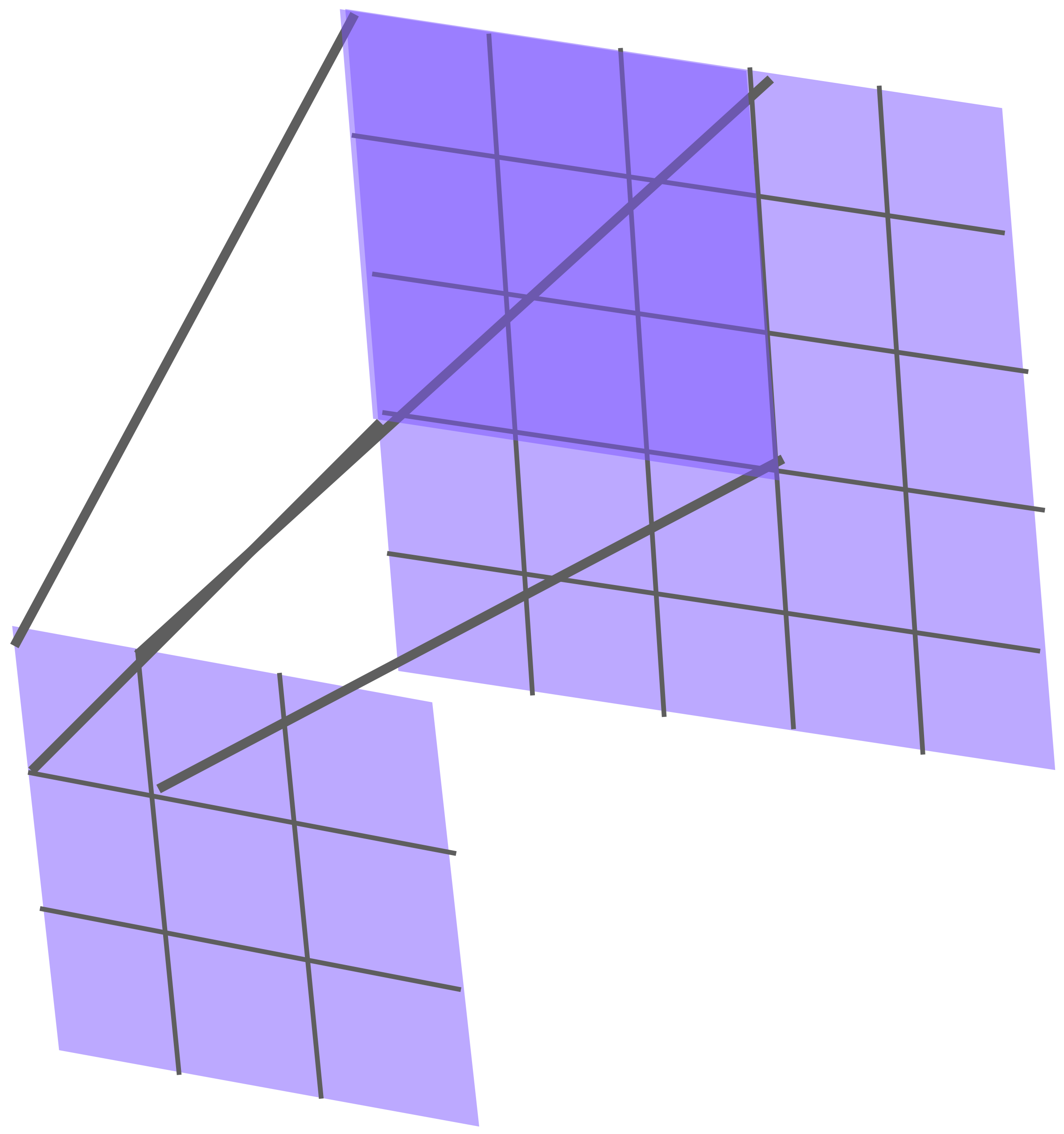
→ Key component of a CNN is the **convolution layer**, which (with the help of a filter) will determine if a feature/pattern is present.



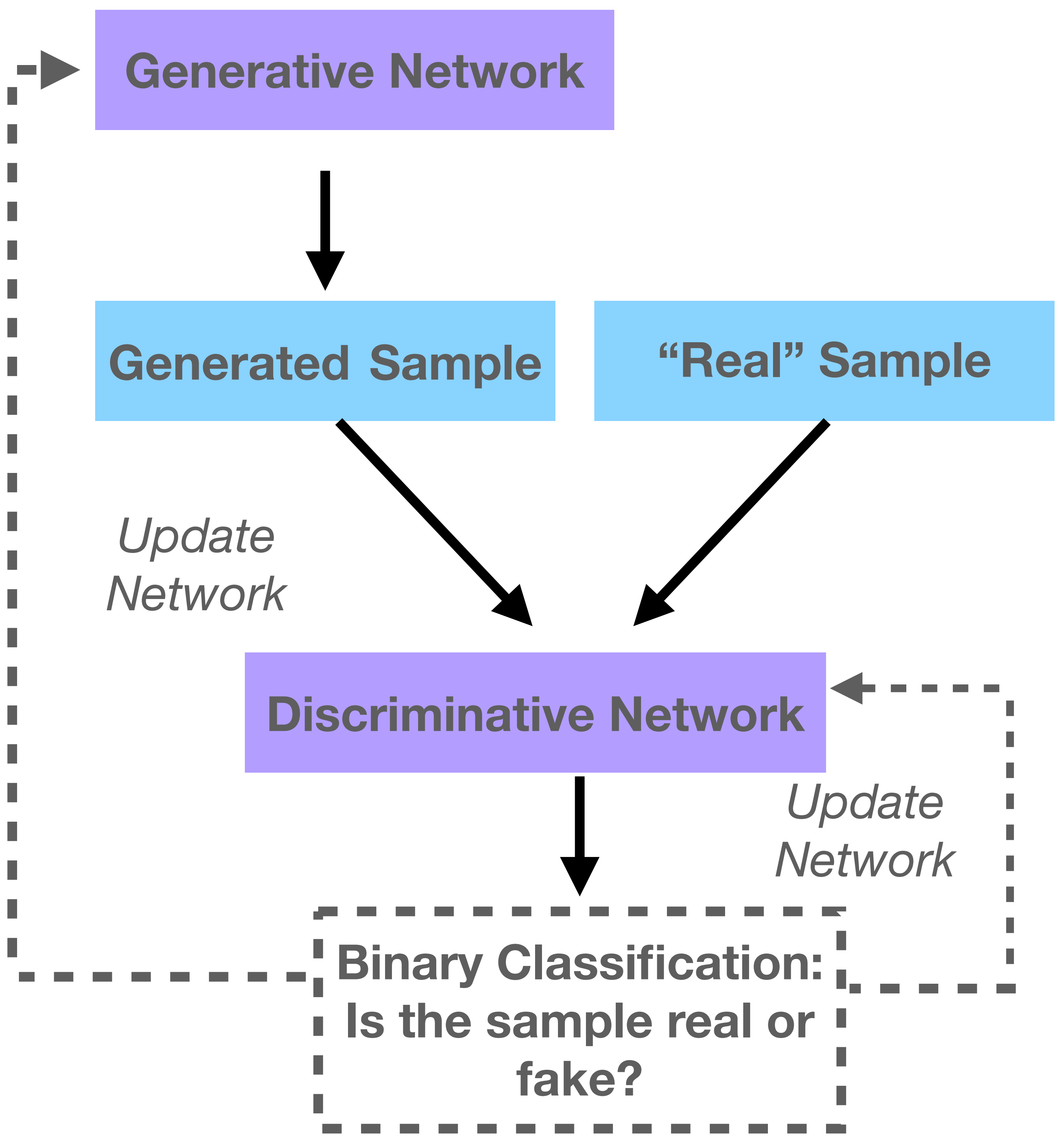
Different algorithms for different problems!



(Shallow or Deep) Neural Networks → *Great for making predictions!*



Convolutional Neural Networks (CNNs) → *Great for image processing!*



Generative Adversarial Networks (GANs)

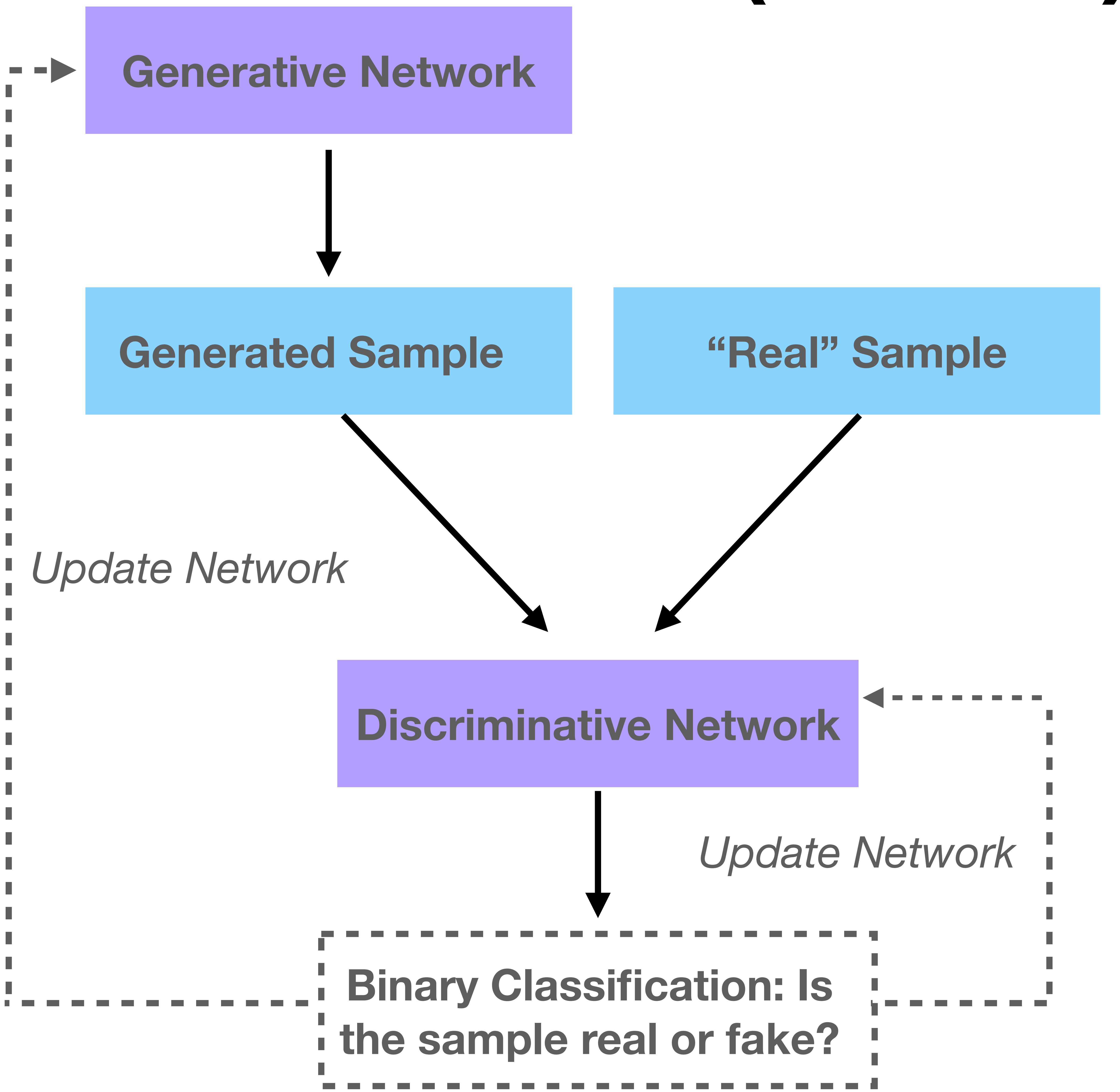
Generative Adversarial Networks (GANs)

Two networks compete with one another in a game.

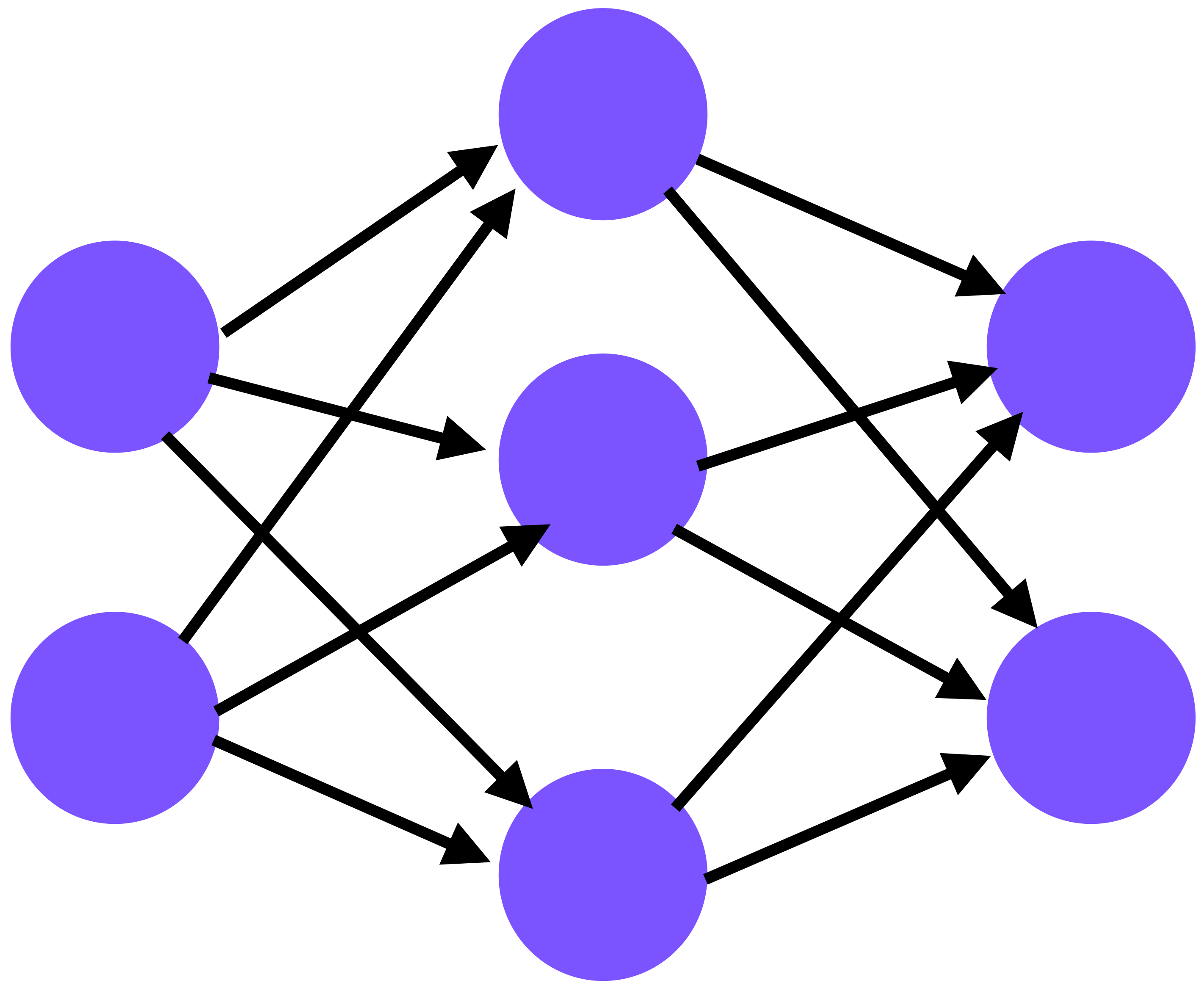
The generative network seeks to fool the discriminative network.

The discriminative network seeks to find the real sample from the generated samples.

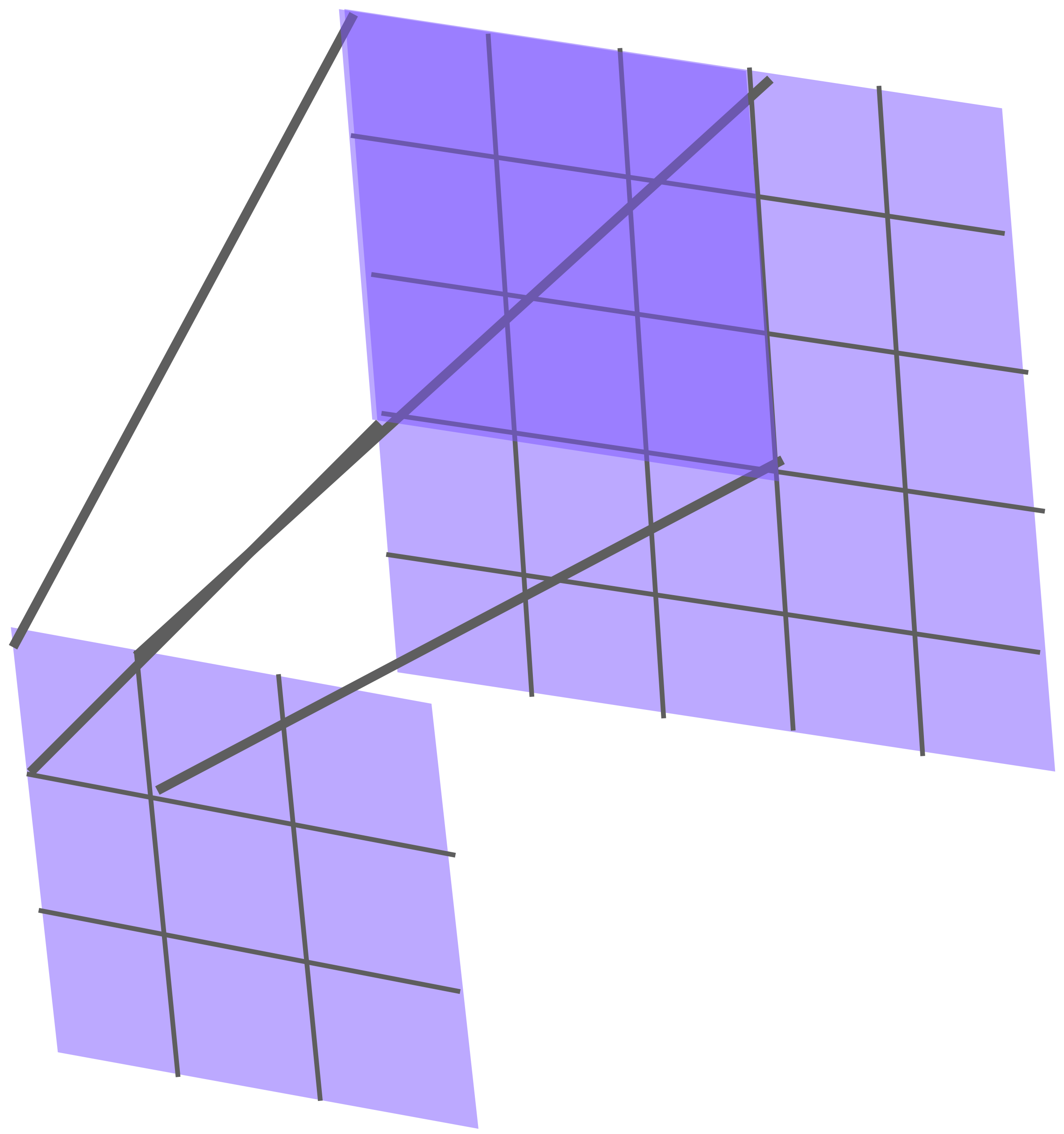
Indirect training → generative network never sees the true distribution!



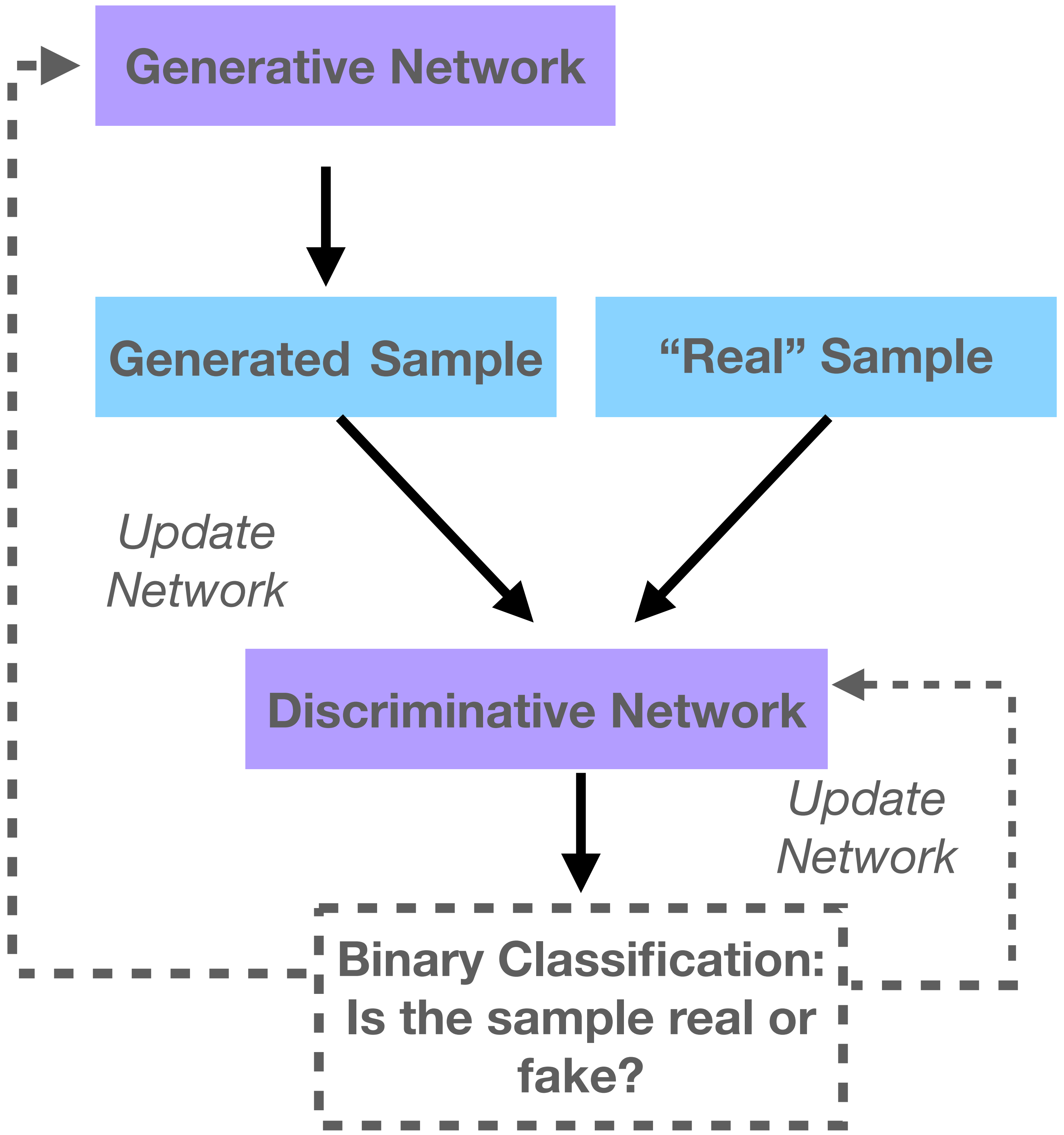
Different algorithms for different problems!



(Shallow or Deep) Neural Networks → *Great for making predictions!*



Convolutional Neural Networks (CNNs) → *Great for image processing!*

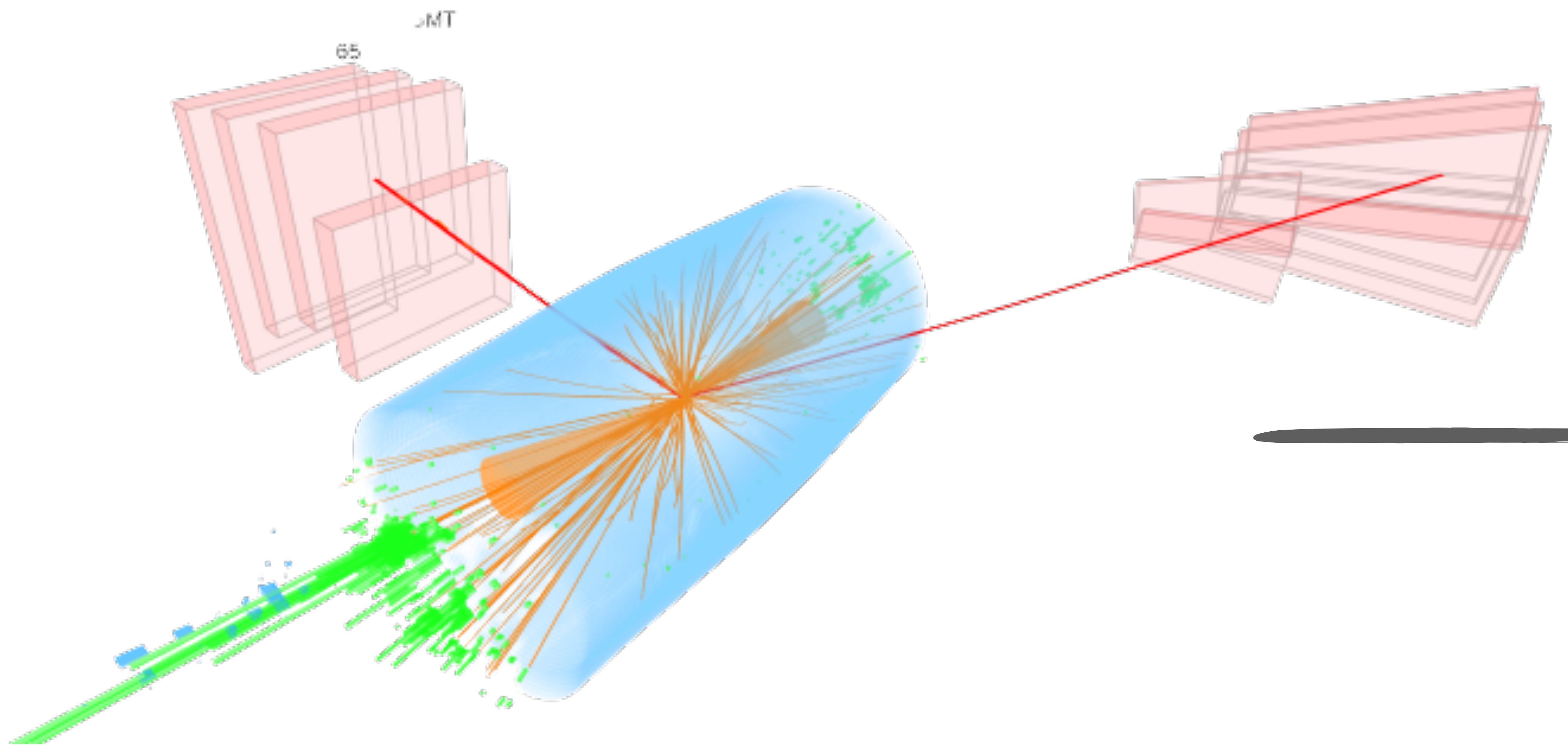


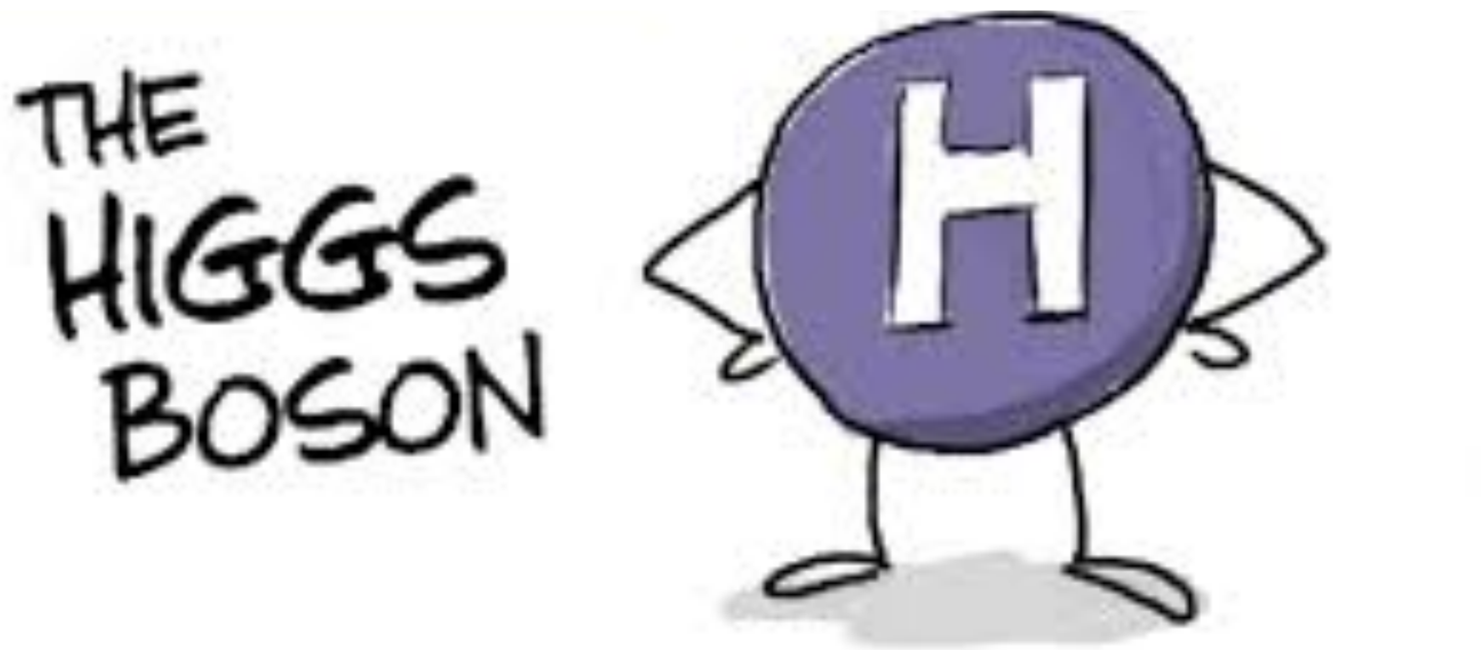
Generative Adversarial Networks (GANs) → *powerful tool for generating samples*!

ML and HEP

Goal of HEP measurements: *To extract relevant physics information from available data!*

Conventional approach: Use a series of boolean decisions motivated by physics or experimental constraints to make a selection then perform a statistical analysis on selected data.





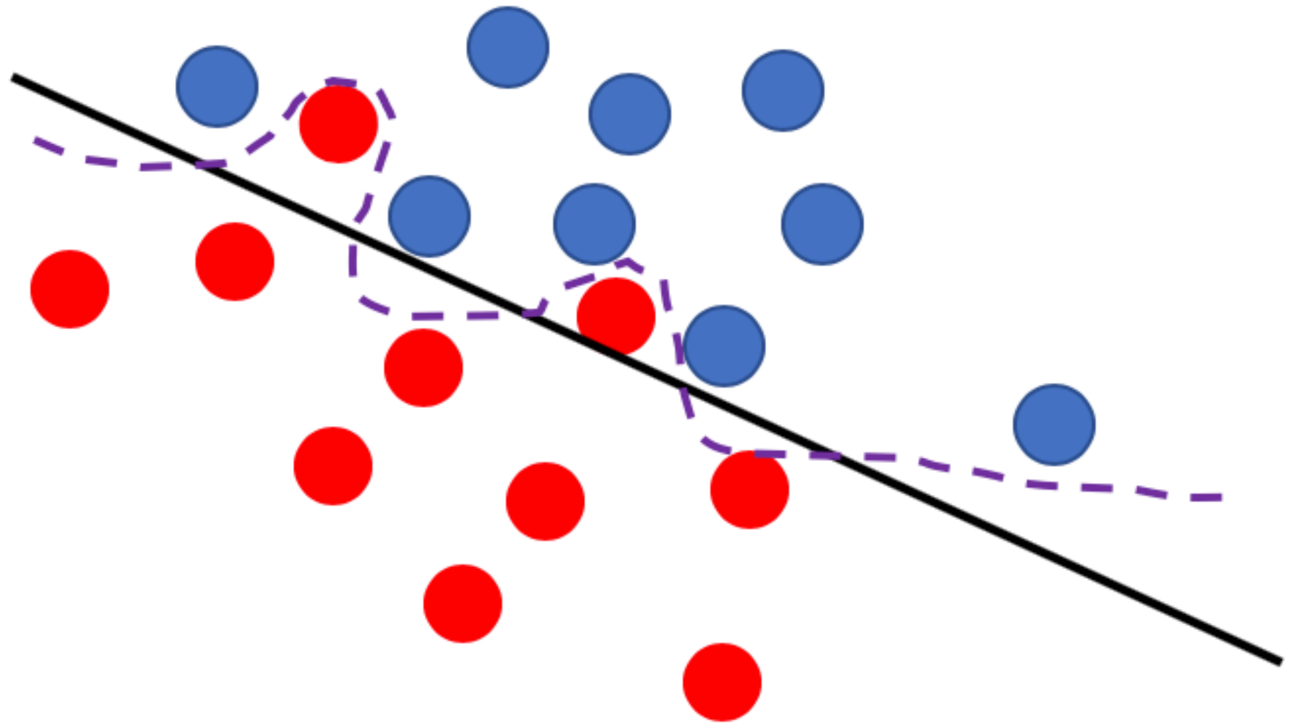
Optimal decision difficult to derive from expert knowledge alone! Employ algorithms that utilize multiple variables simultaneously → inspired countless ML applications! (Living Review)

Limitations of ML

ML is not a magic fix!

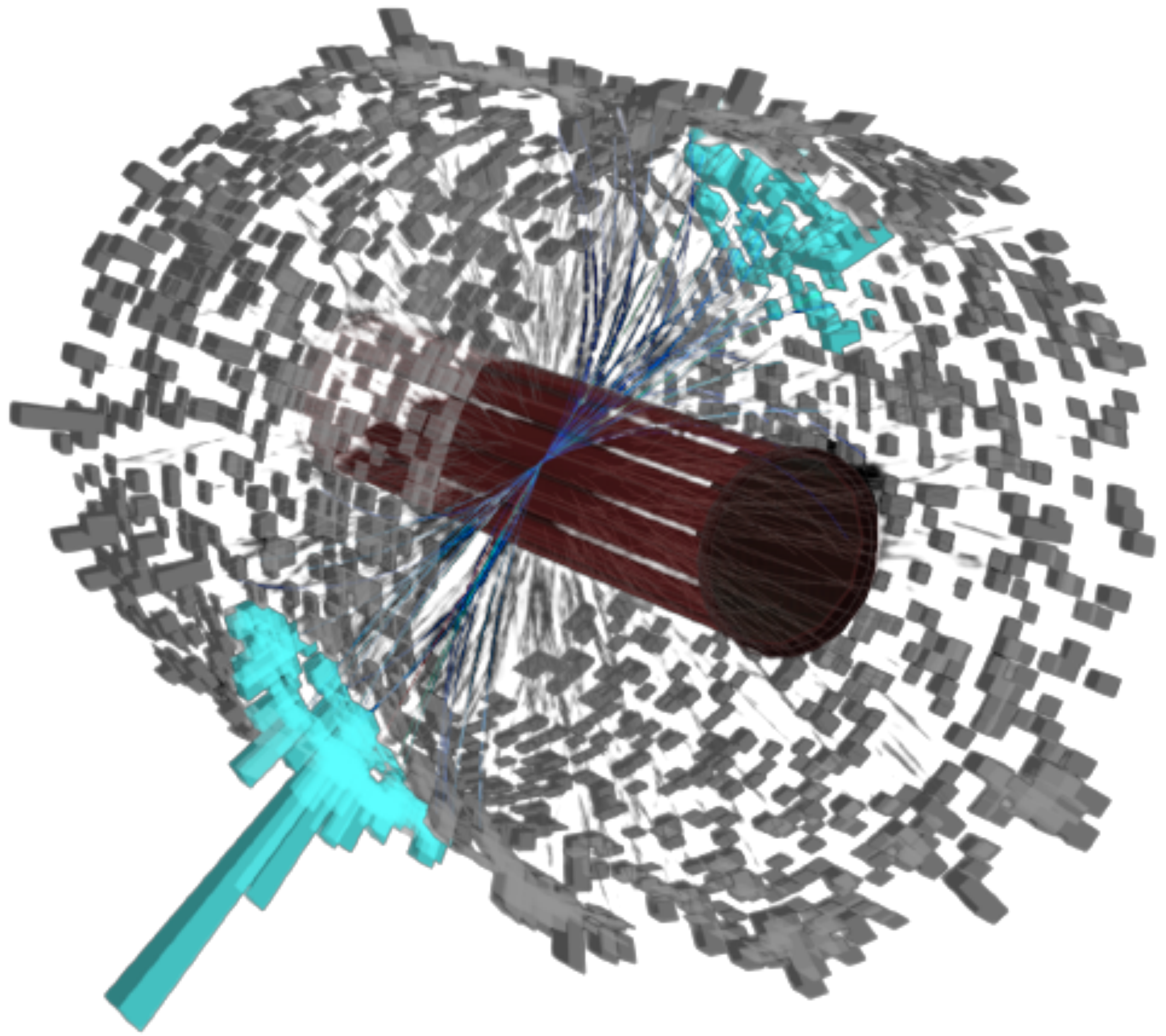
- *Put garbage in, get garbage out!*
- ML cannot replace domain knowledge.
- ML is not a causation tool.
- Model should be generalizable (i.e. should perform well on unseen data).

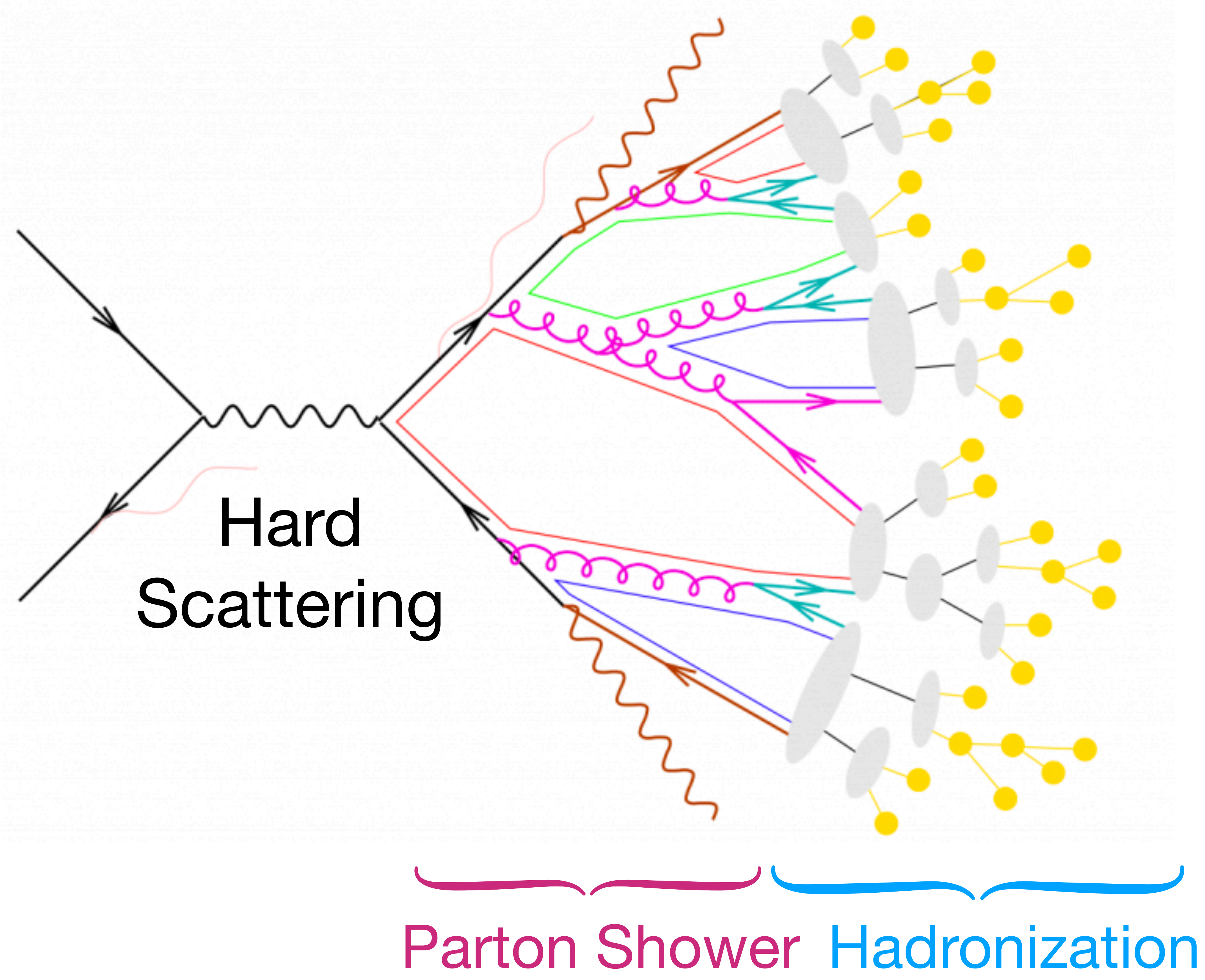
Don't want to be finding cloudy days when you should be finding tanks!



ML and Jets

Jets are experimentally and theoretically complex multi-dimensional objects sensitive to many physics scales! → *Talk today will be from the perspective of jets in heavy-ions (HIs) where that is especially true!*





→ Good candidate for ML!

Outline

What is ML and why is it useful for jets?

How is ML currently being used for jet measurements?

Ongoing challenges

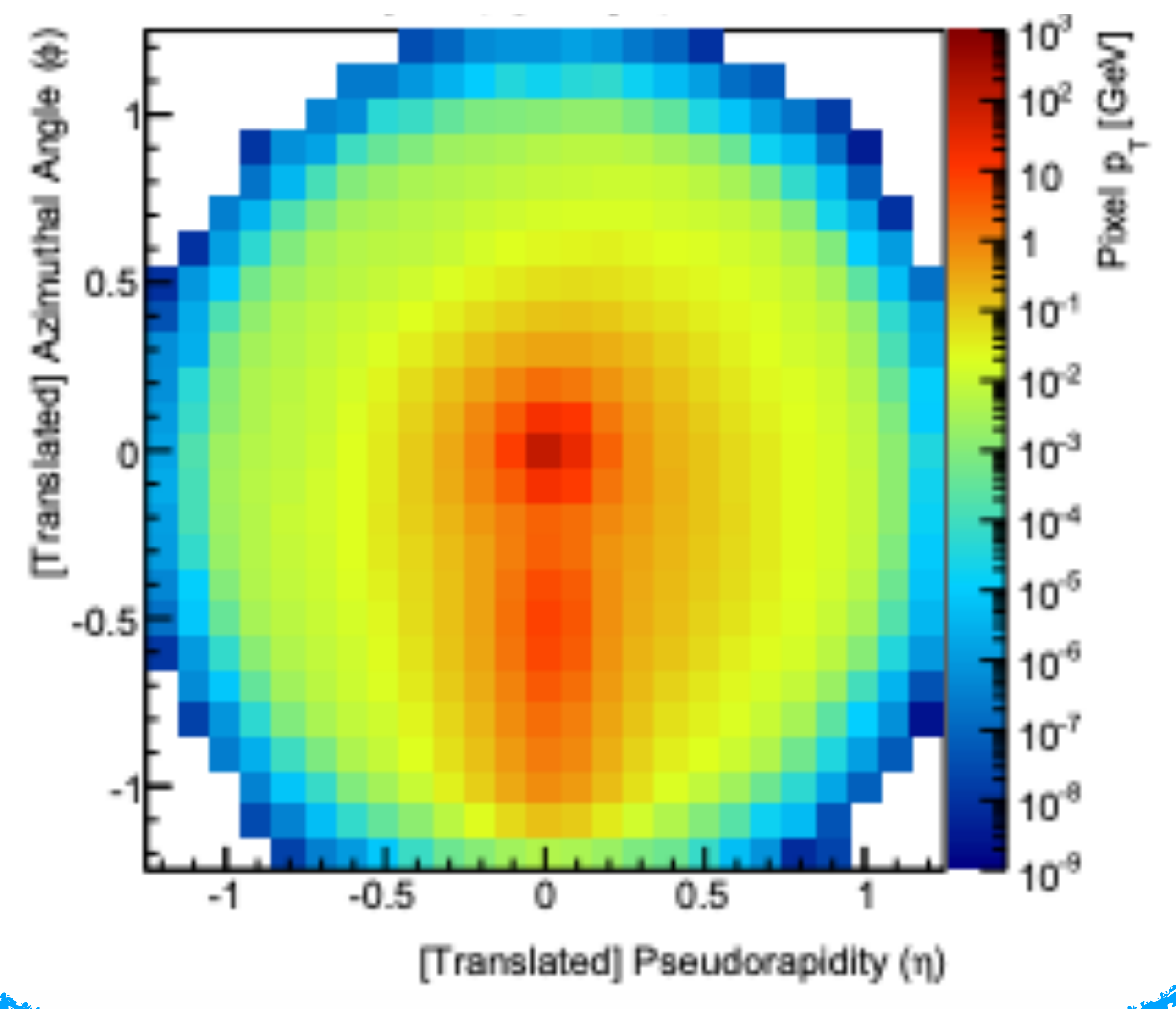
Different representations of jets

Optimal choice depends on the problem! (May be a combination of these!)

Jets as Images

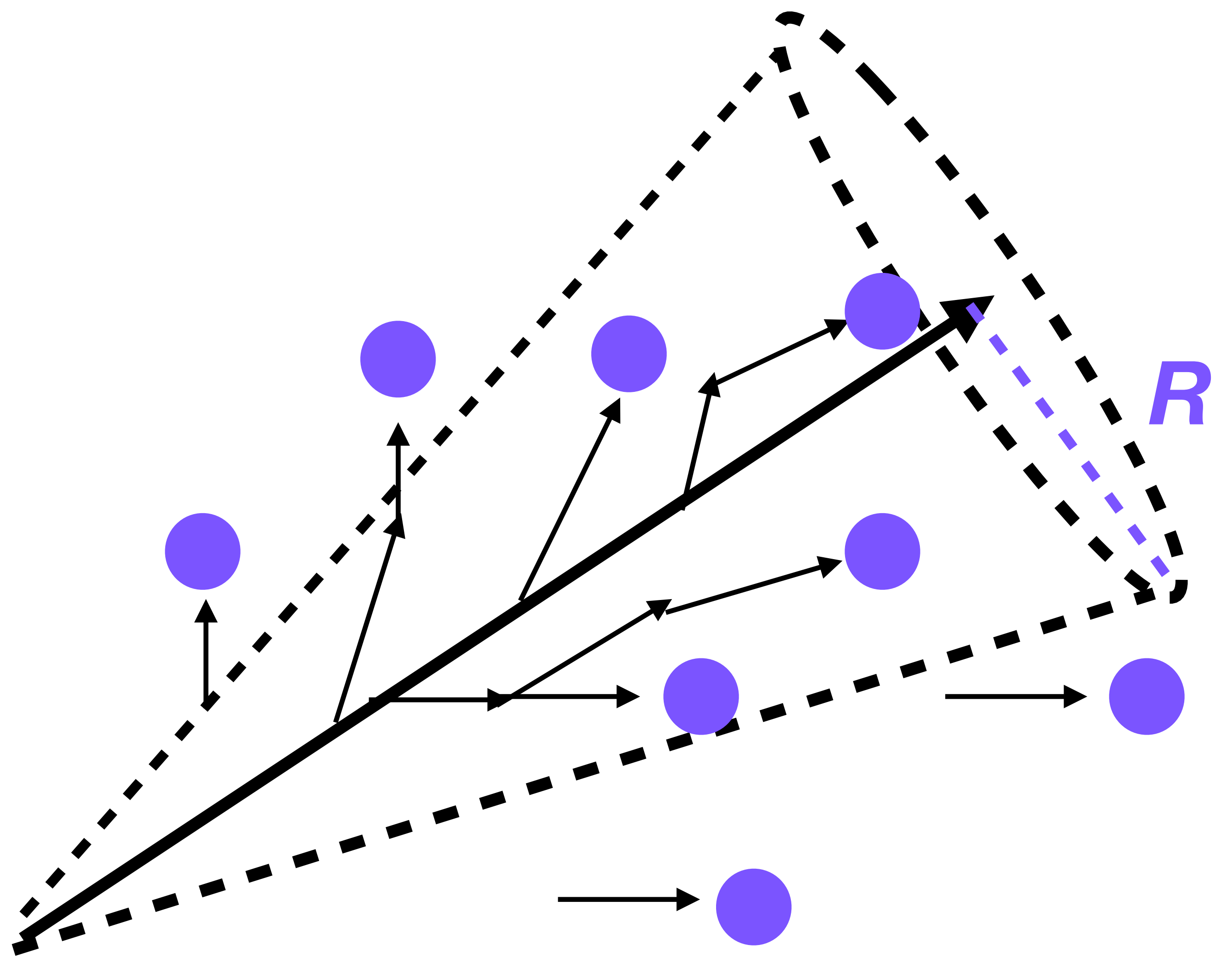
Advantage: powerful tools available for image classification with ML!

arXiv: 1150.05190



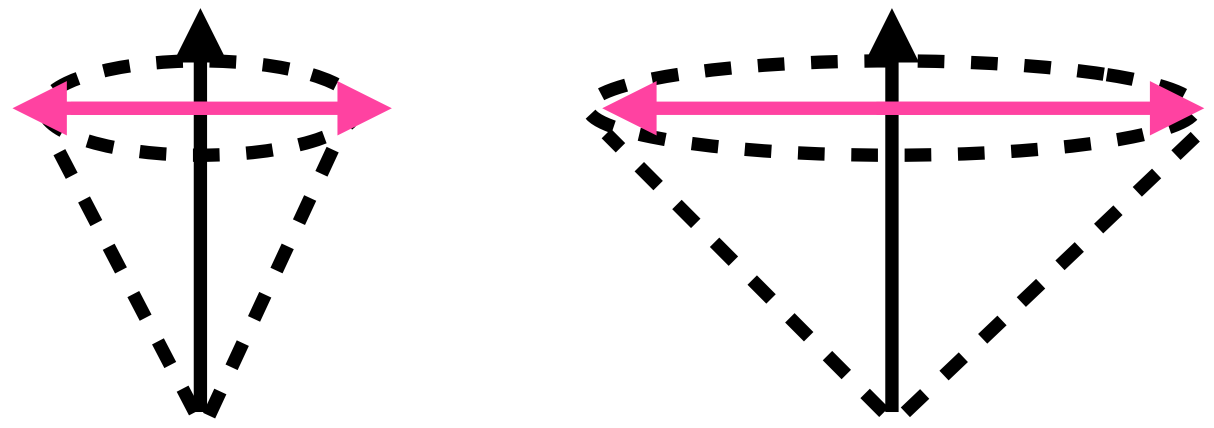
Jets as a collection

Ex: declustering history, ordering of constituent p_{T}

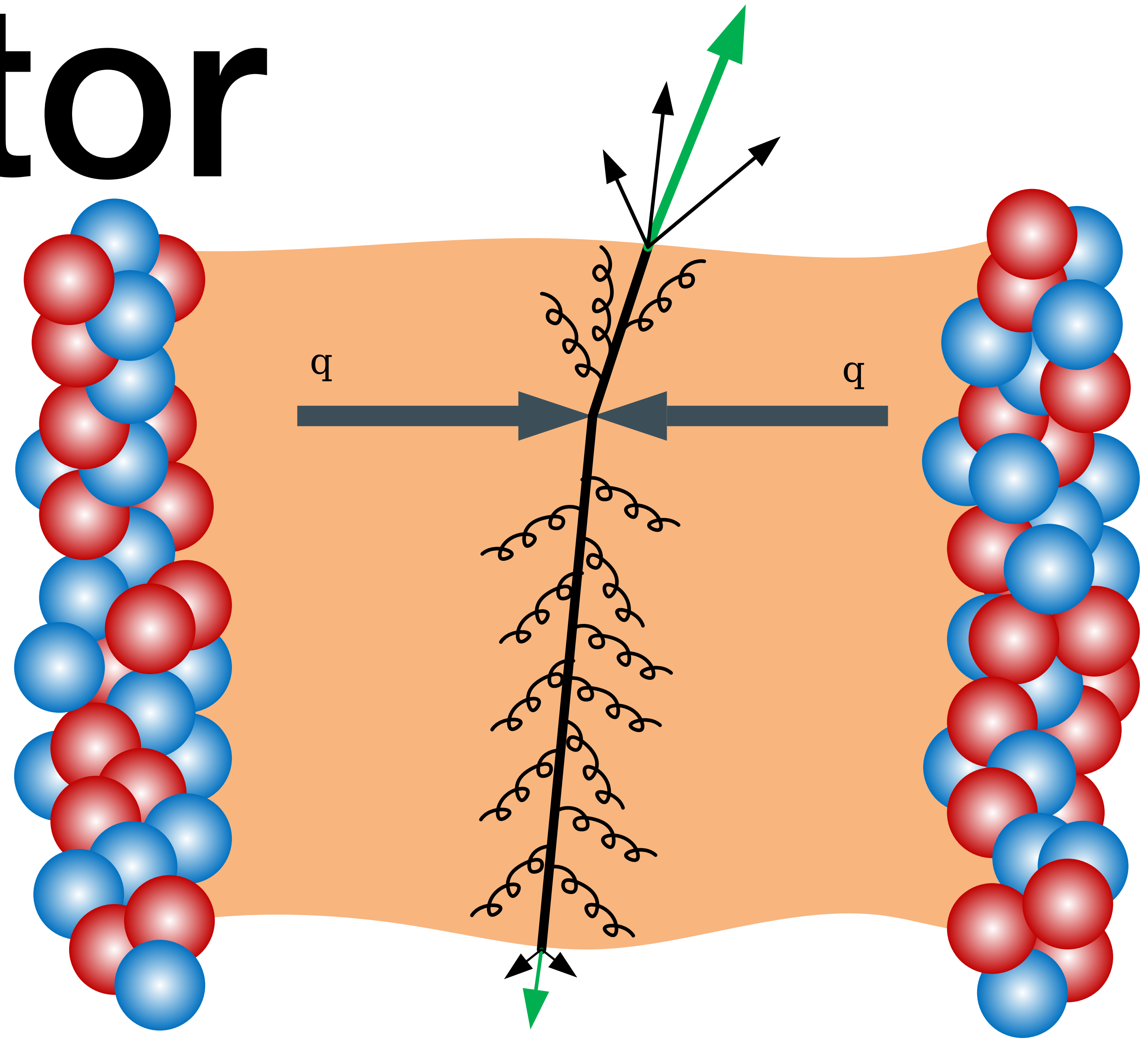


Jets as a single object

Ex: Jet mass, radial moment, other jet shapes…



ML background estimator



Use machine learning (ML) to correct the jet for the large uncorrelated background in heavy-ion collisions!

Conventional approach: Apply a minimum p_{T} requirement on the leading track of the jet, correct the jet for the background with a pedestal subtraction.

ML approach: Use ML to construct the mapping between measured and corrected jet without a leading track bias.

Jet Properties

(Including constituent properties)

ML →

Corrected Jet p_{T}

→

Unfold for fluctuations and detector effects

R.Haake, C. Loizides Phys. Rev. C 99, 064904 (2019)

Process

Training (PYTHIA fragmentation)

Train on “hybrid event” created by embedding PYTHIA jets into Pb-Pb Background

Shallow neural network

Testing

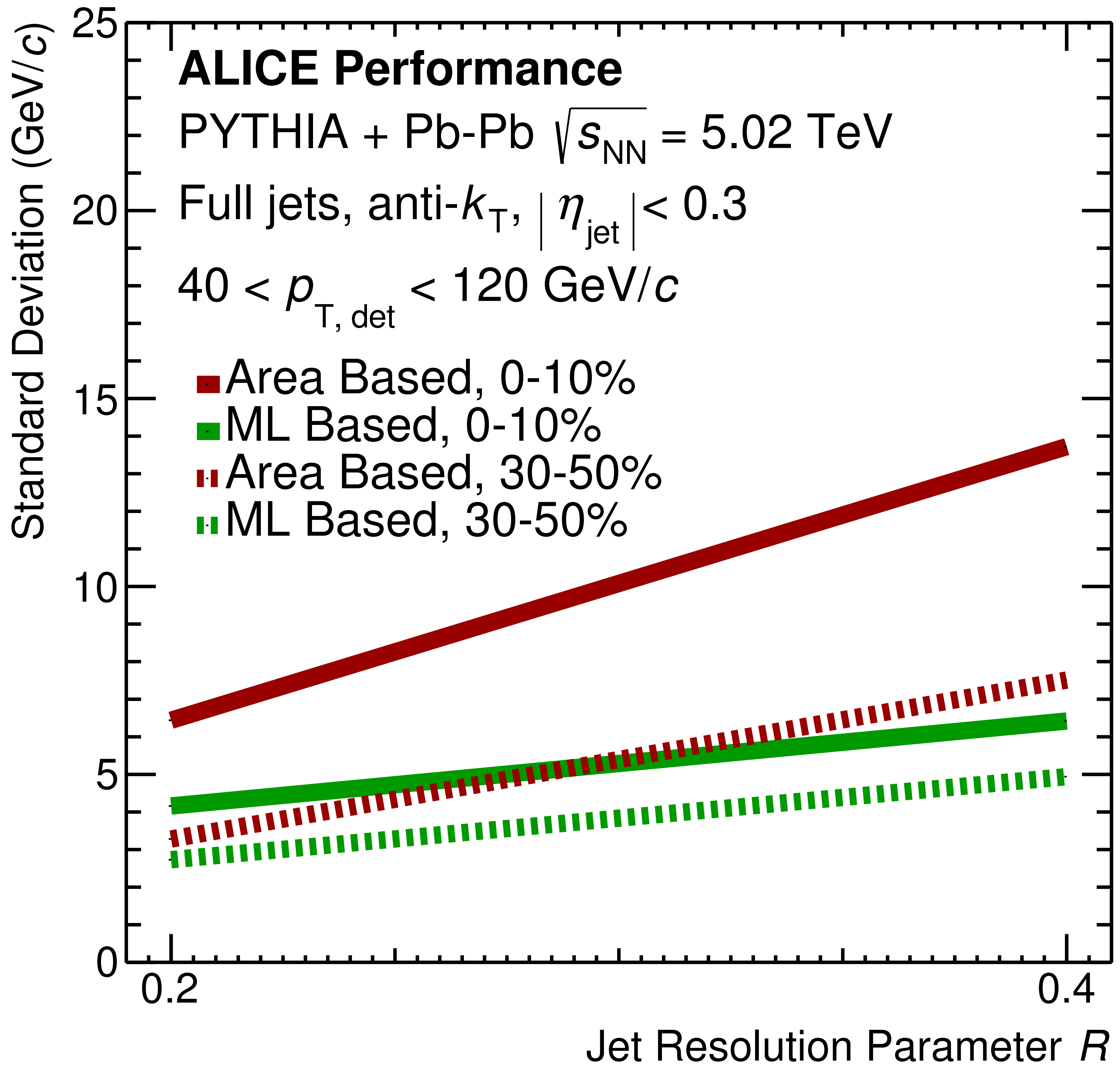
Apply ML estimator to hybrid events not used in training.

Key is that this background is *realistic.*

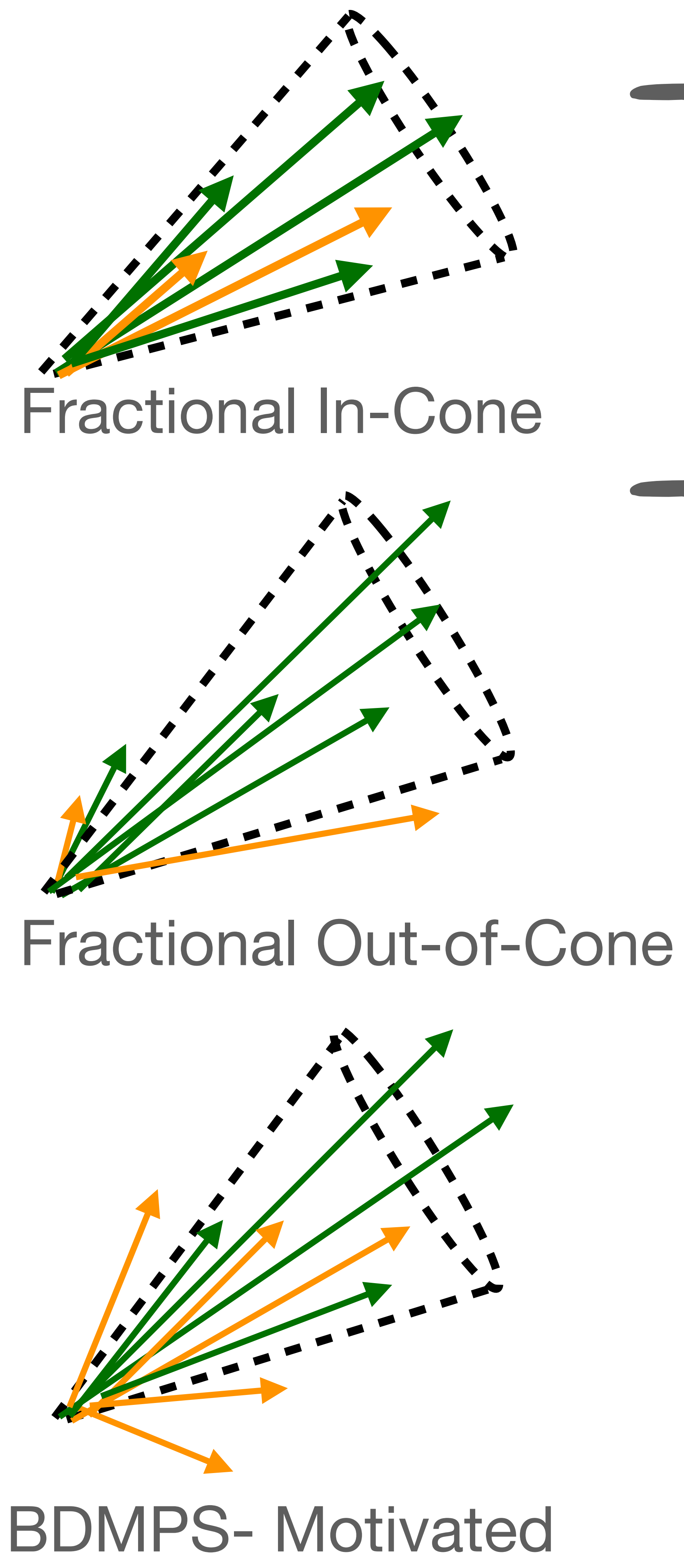
Do we get back the signal we put in?

Testing performance and potential bias

Are we getting closer to the "truth"?

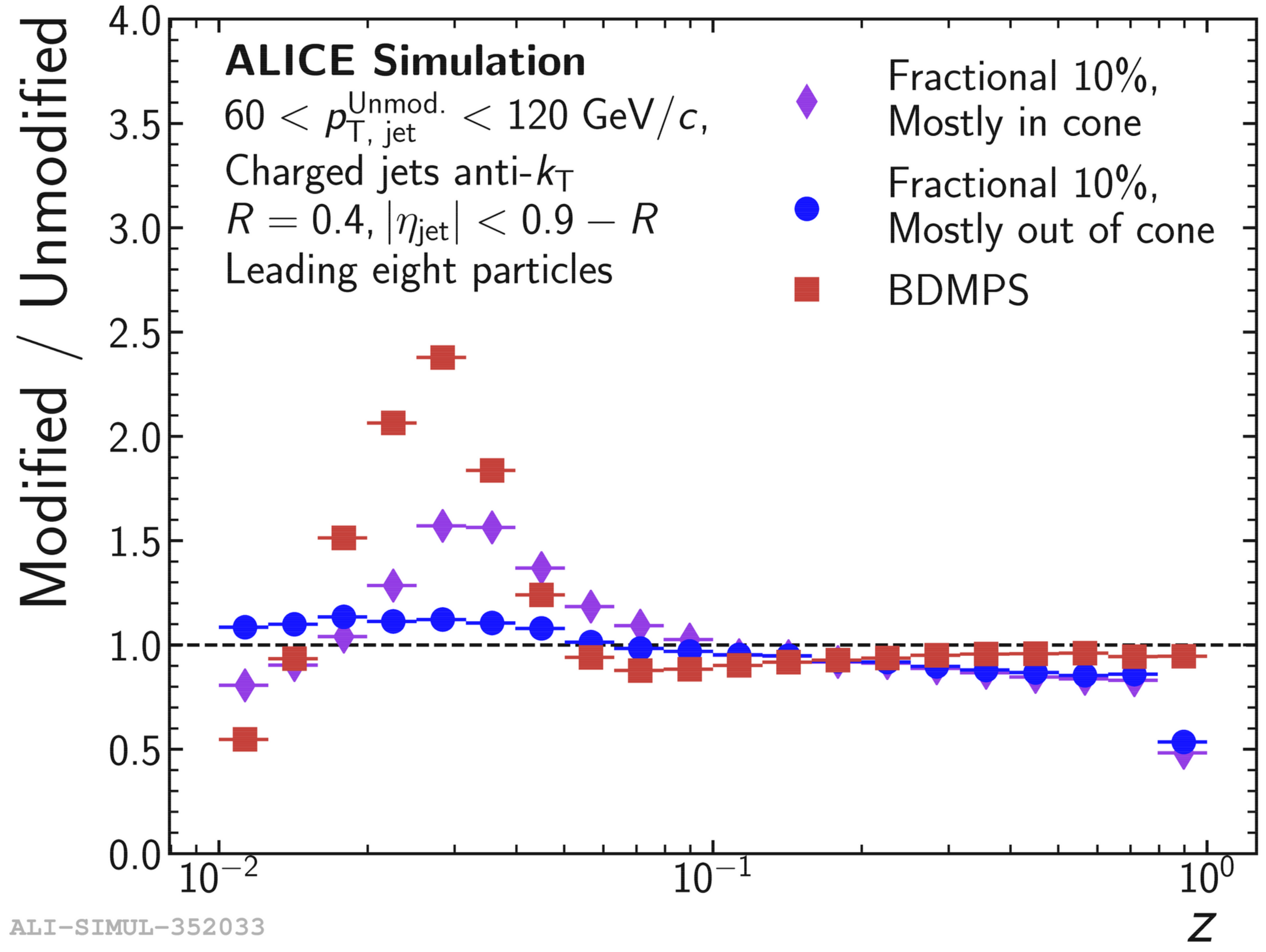


Residual fluctuations significantly reduced!



Learning on constituents introduces a bias towards PYTHIA fragmentation!

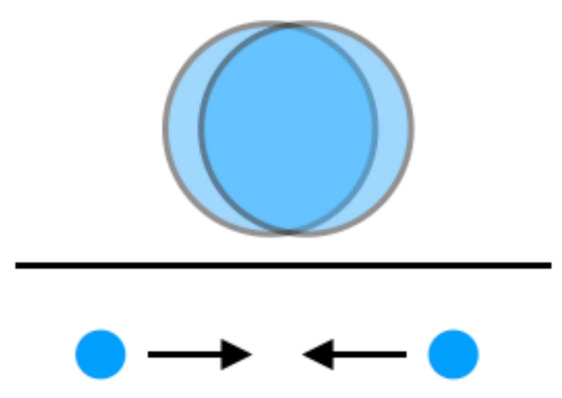
Modify PYTHIA jets to change the fragmentation.

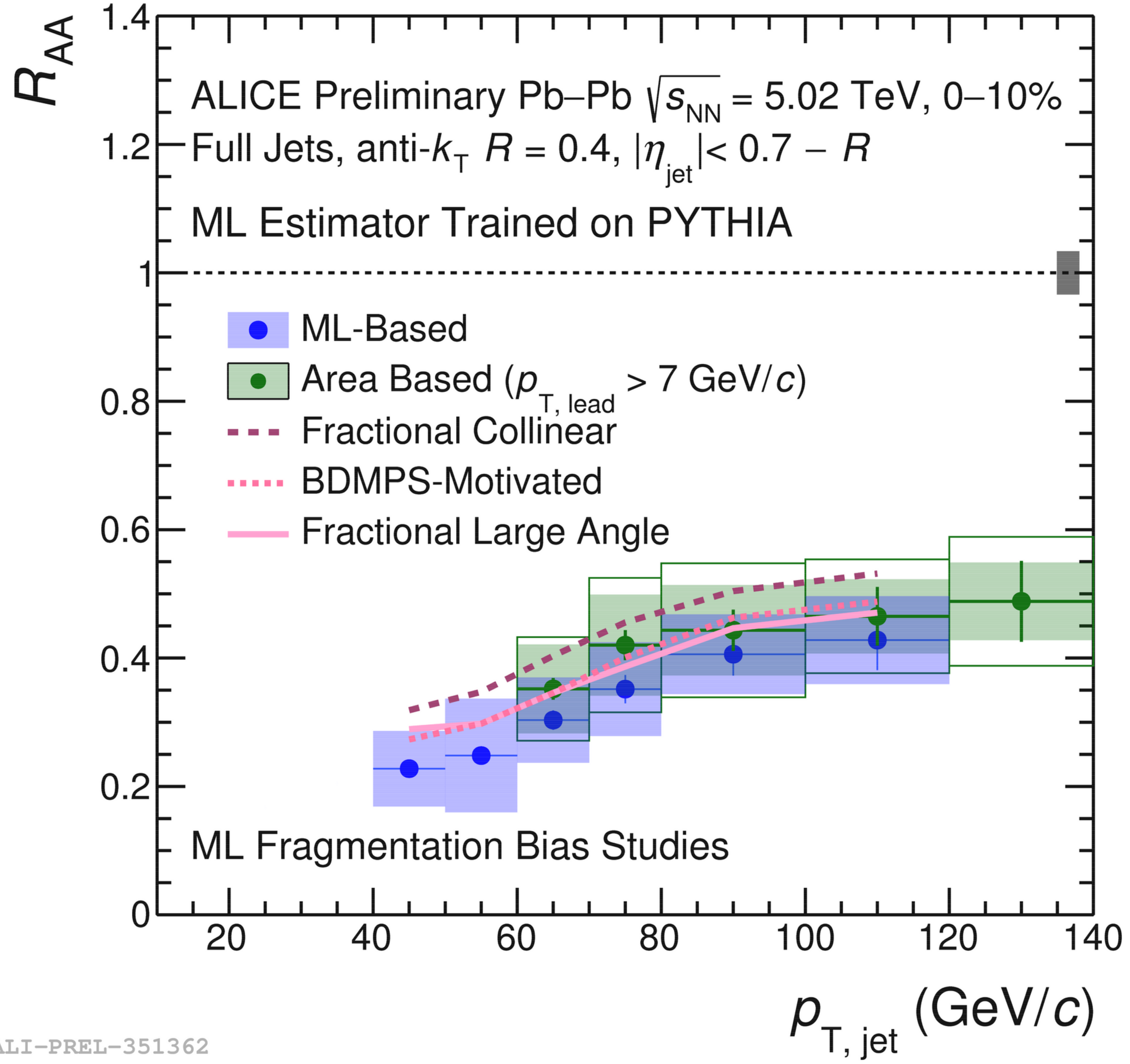


Train on the modified toy model and apply to data; measure bias.

Applying ML to data

$$R_{AA} = \frac{\frac{1}{N_{\rm event}} \frac{d^2 N_{\rm jet}^{\rm PbPb}}{dp_{\rm T} dy}\Big|_{\rm cent}}{\langle T_{\rm AA} \rangle \frac{d^2\sigma_{\rm jet}^{\rm pp}}{dp_{\rm T} dy}}$$





ML extends measurements lower in p_T with reduced systematic uncertainties!

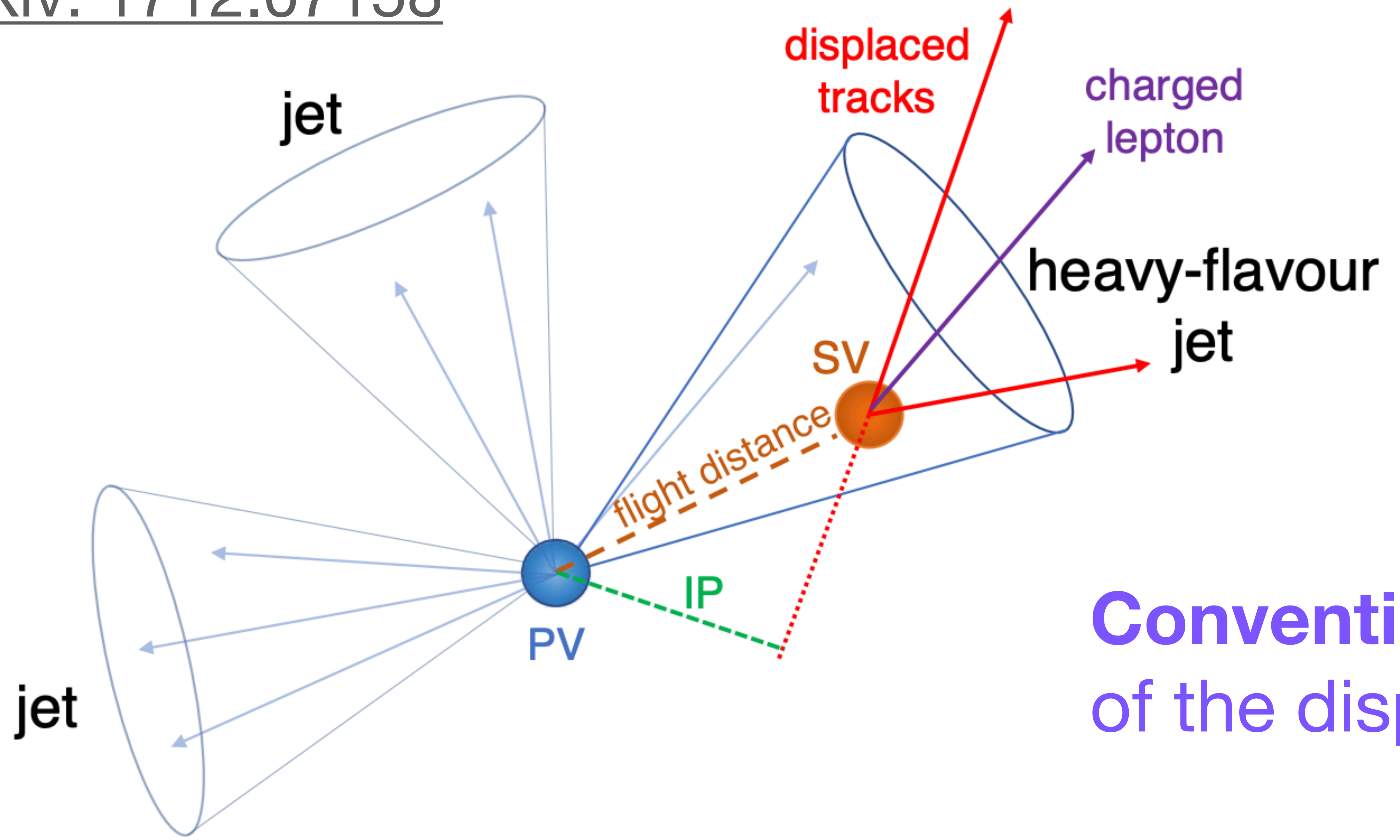
Useful to go to low p_T for kinematic overlap with RHIC!

Method is relatively robust to the explored biases!

Future: Apply similar techniques to other variables such as jet substructure!

Heavy Flavor Tagging

arXiv: 1712.07158



HF Jets have….

- large impact parameter of tracks
- displaced secondary vertices

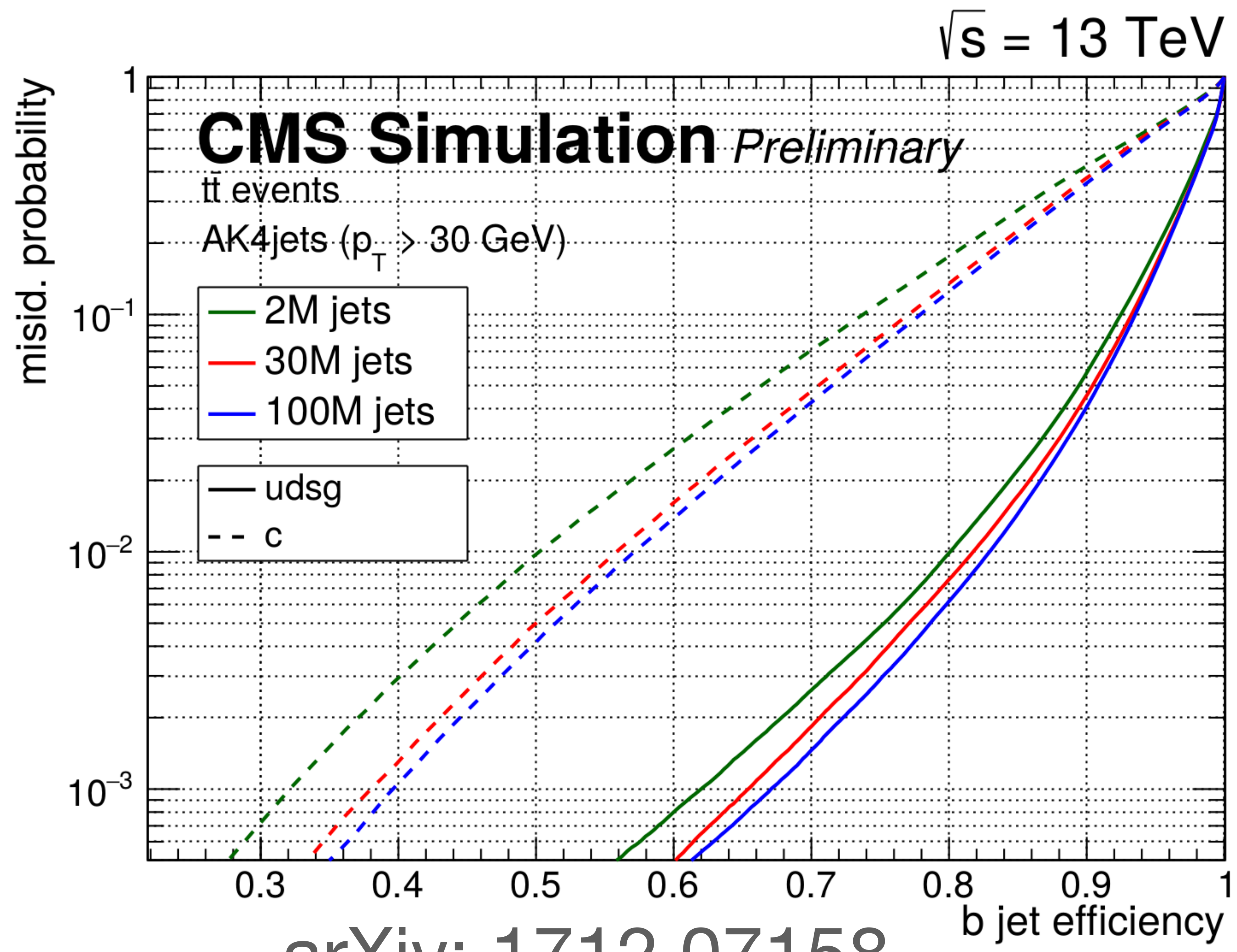
Conventional approach: Apply cuts on the properties of the displaced vertices.

ML approach: Use low-level jet parameters such as constituents, secondary vertices, track impact parameters etc. Learn from simulation in a supervised approach.

Heavy Flavor Tagging

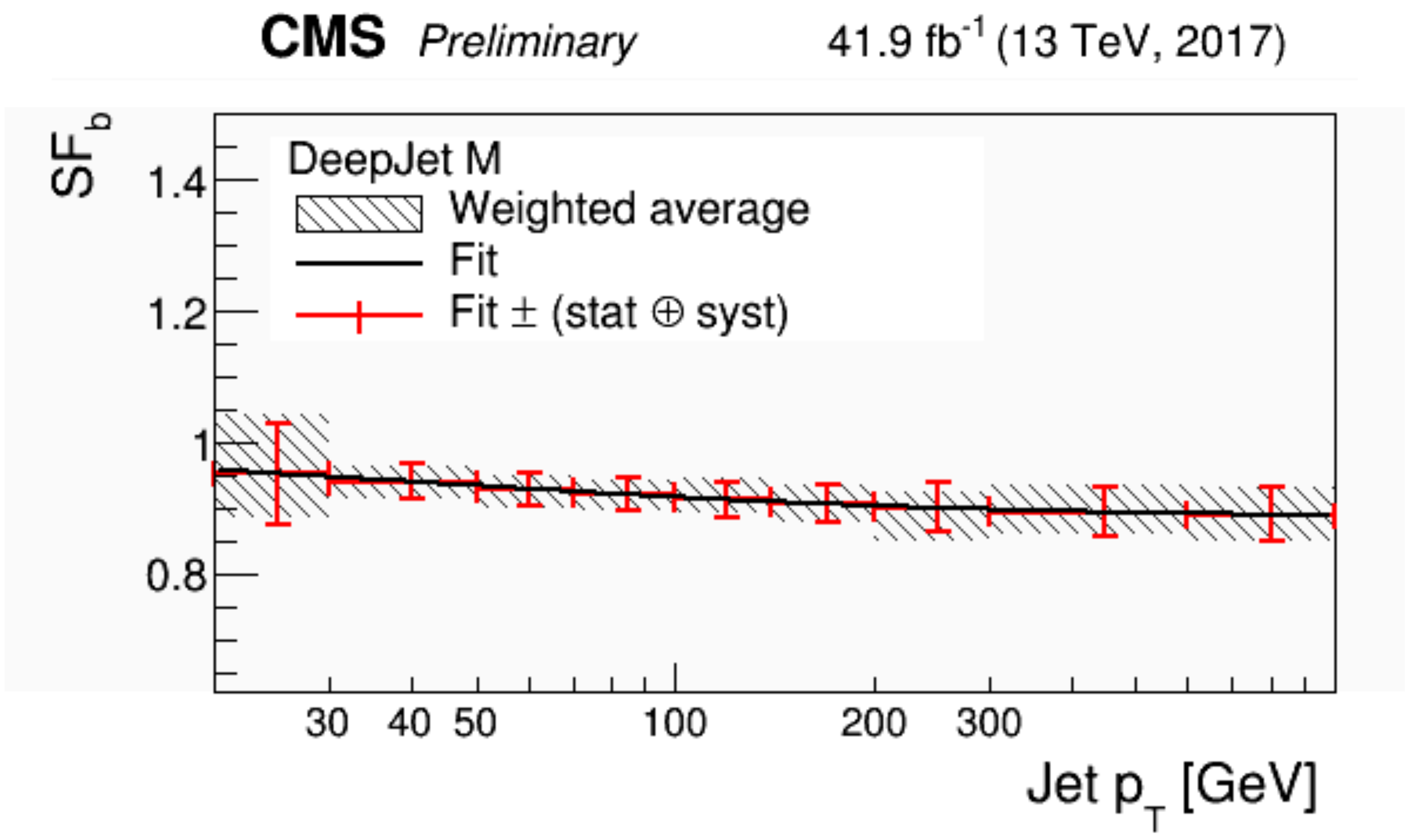
Many ongoing experimental efforts showing nice improvement, focus on one CMS application to data using a deep neural network!

- Lower the curve, the better the performance
- Performance tends to improve with the size of training data



arXiv: 1712.07158

Applied to data!

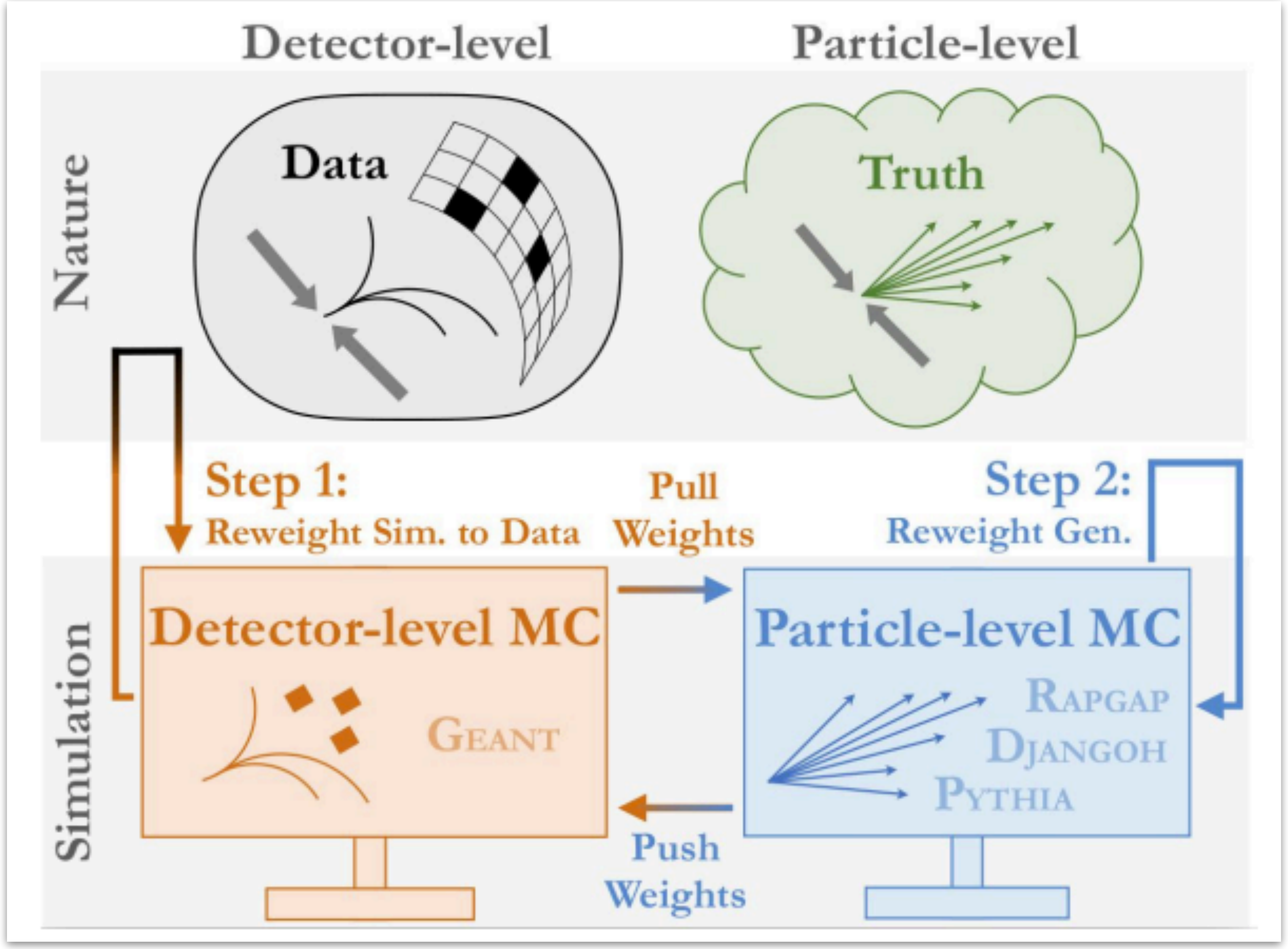


Future: Extend to p-Pb and Pb—Pb data!

OmniFold: Simultaneous Unfolding of All Observables

Purpose of Unfolding: Correct for detector effects and background fluctuations.

Conventional Approach: Use an unfolding procedure on a binned distribution and repeat for each desired observable!



ML Approach: Use ML to calculate weighting factors and unfold the phase space all at once, before the choice of binning or observable!

Really beneficial to…

→ have an unbinned result!

→ unfold with more dimensions which can correct for more complex effects.

(Ex: Two jets with different substructures acquire same mass.)

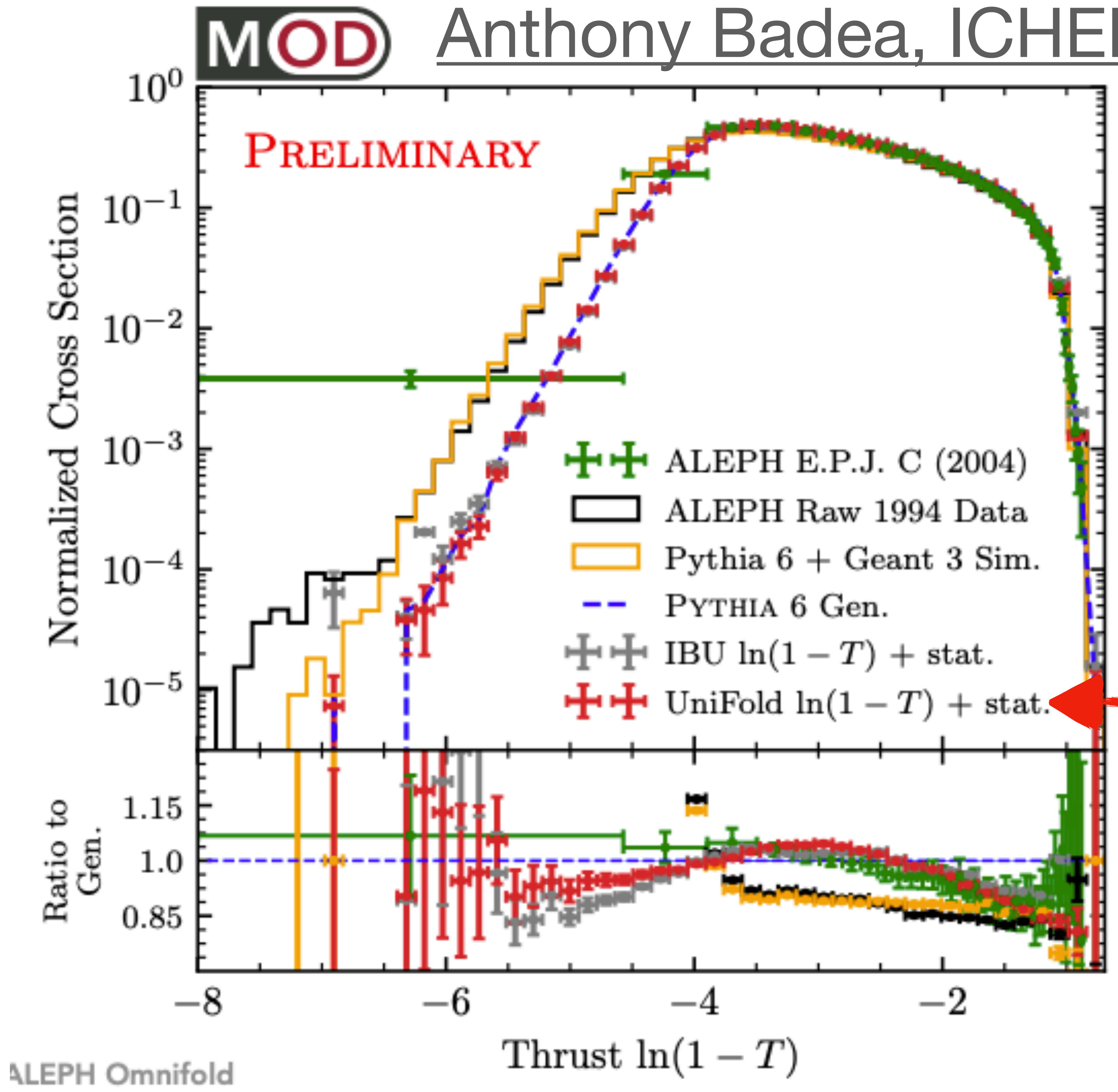
OmniFold

arXiv: 1911.09107

Applying OmniFold to Data

Two applications to data!

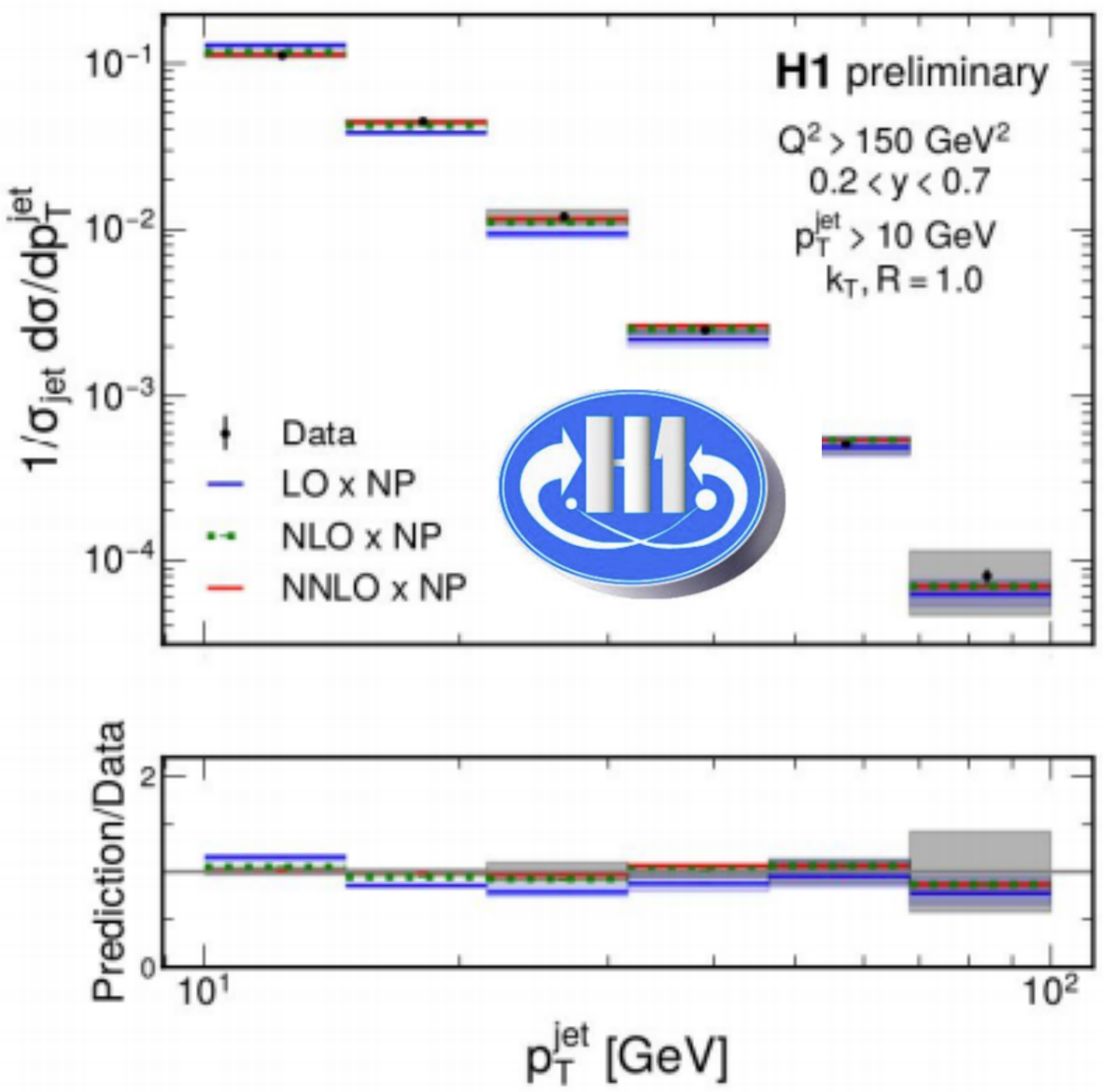
Anthony Badea, ICHEP



unbinned IBU

ALEPH Archived Data (Thrust)

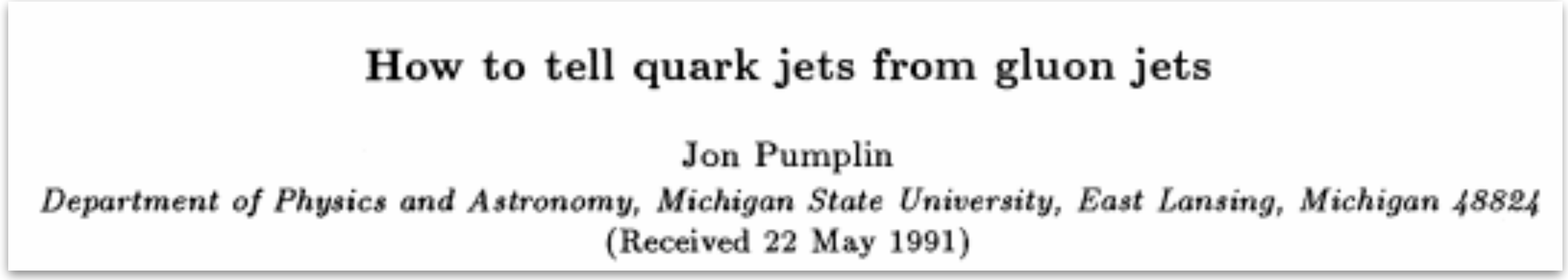
Miguel Arratia, DIS



H1 Archived Data (Jet p_{T})

Future: This is just the beginning! Hopefully many more applications to data!

Quark vs. Gluon Jets with ML


How to tell quark jets from gluon jets

Jon Pumplin

Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

(Received 22 May 1991)

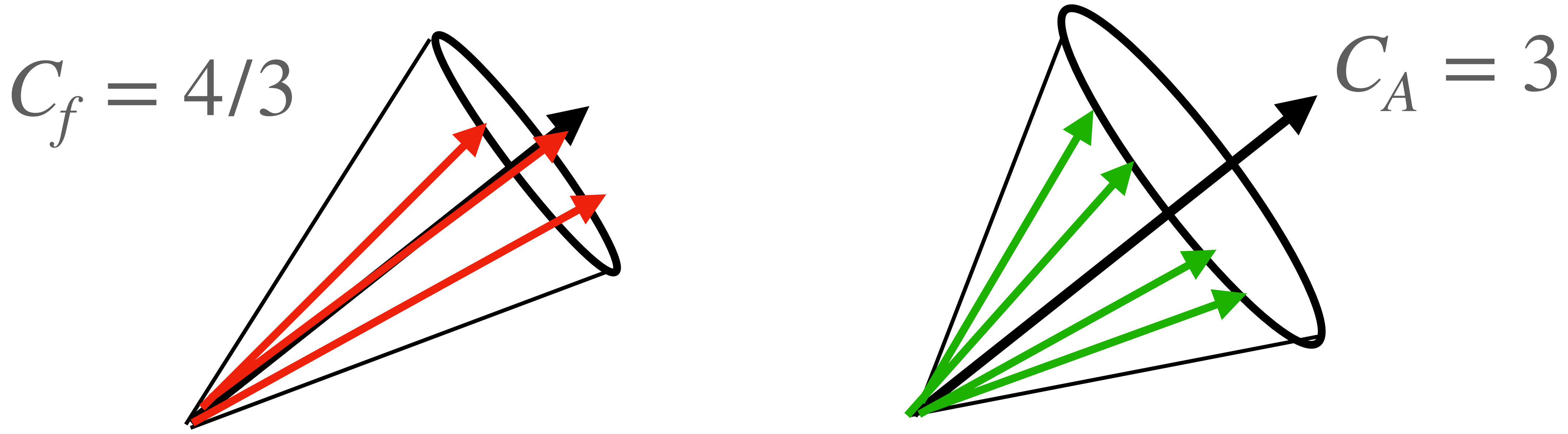
Long-standing effort starting around 1991!

Quark and gluon jets have different color factors and substructure!

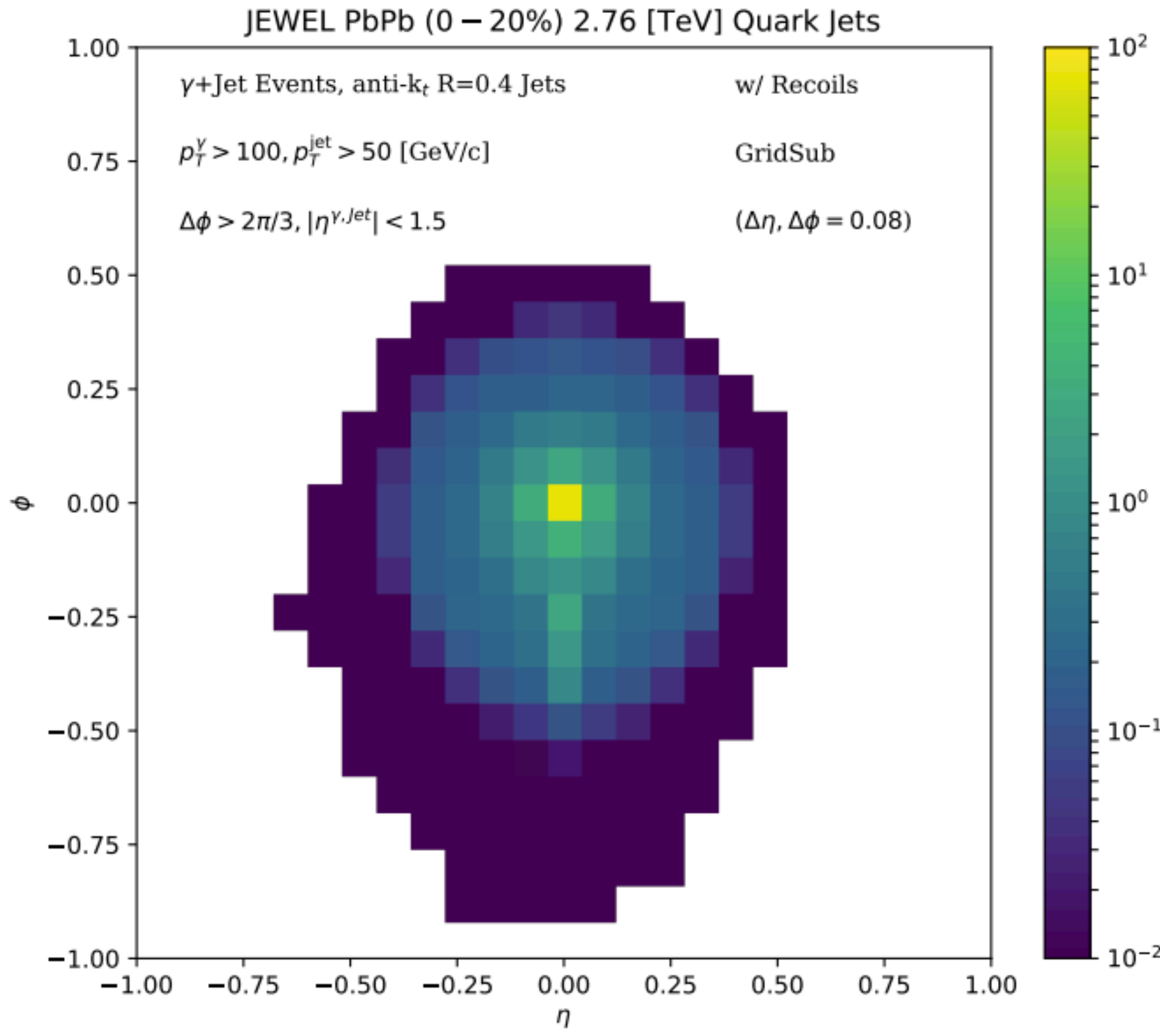
We will focus on an effort in HI collisions!

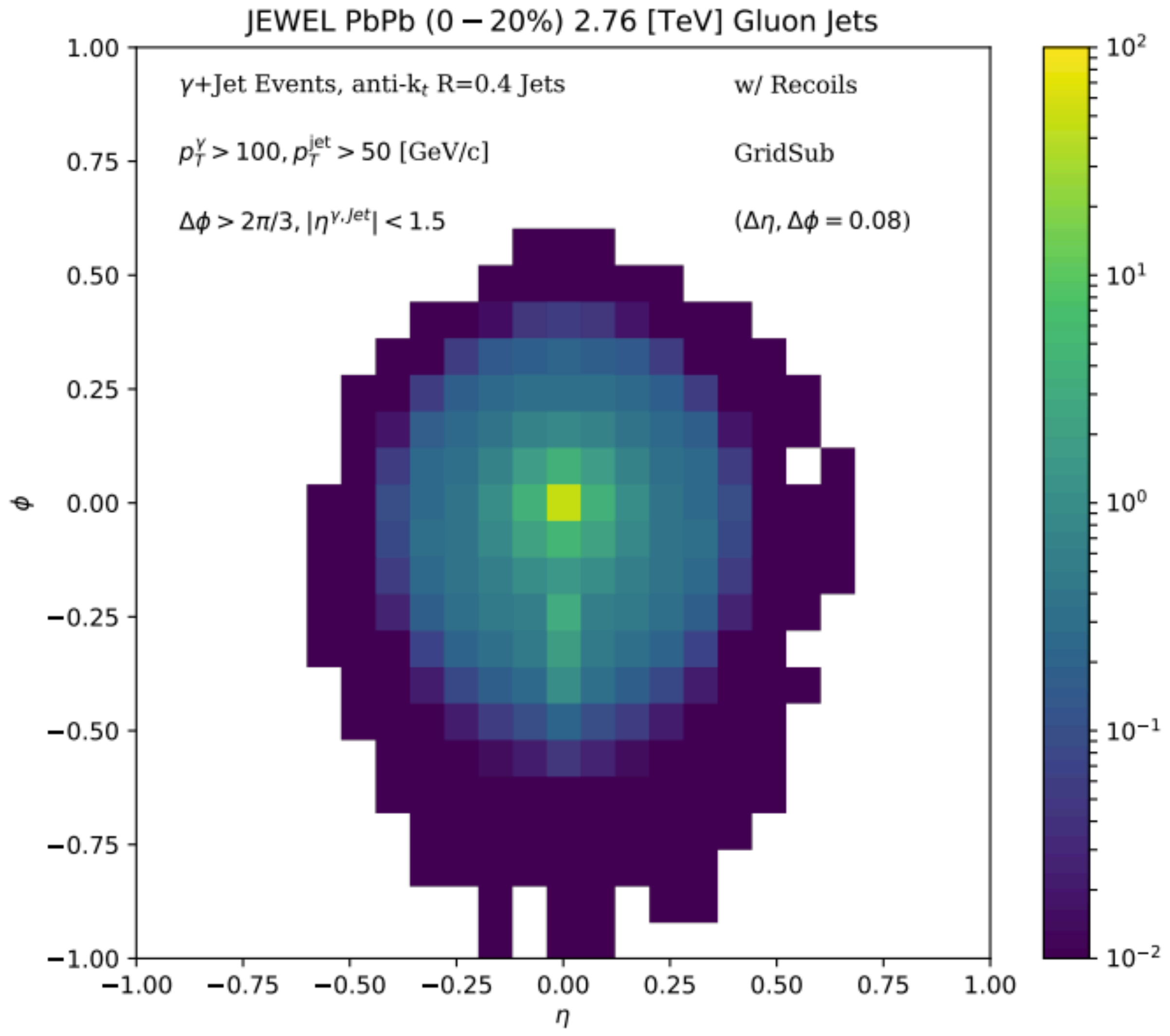
Quark Jets

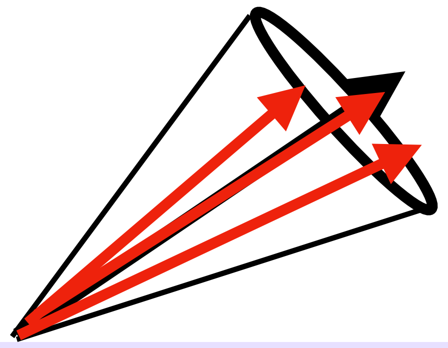
Gluon Jets

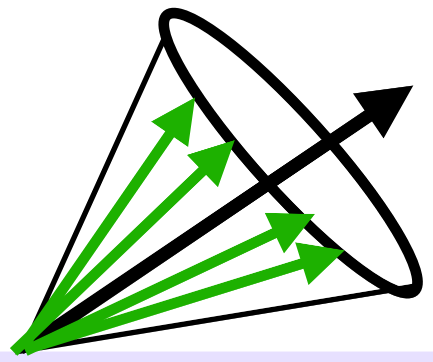


Y. Chien, R. Elayavalli: 1803.3589







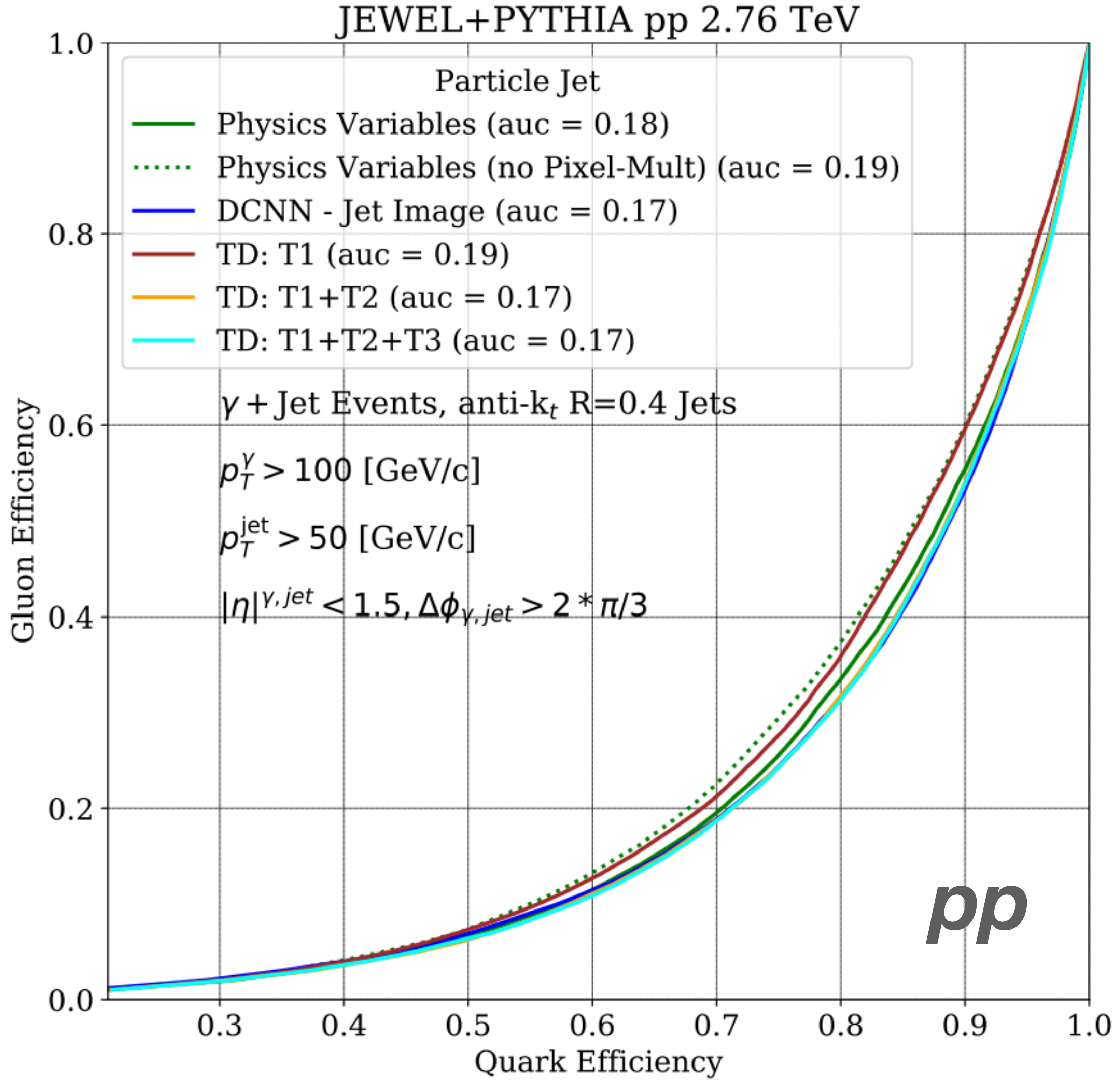


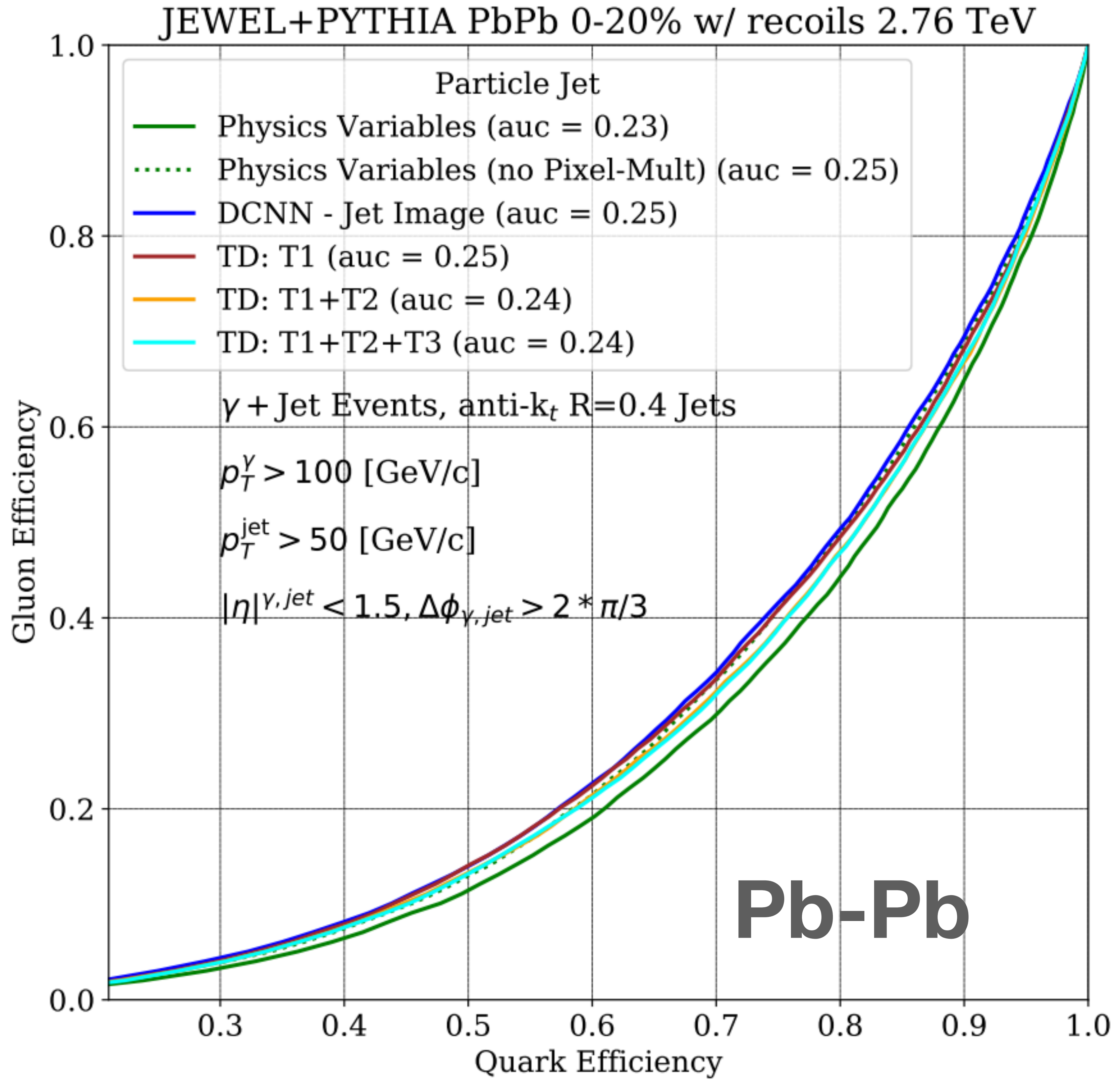
Use jet images with deep CNNs (DCNNs) to discriminate q/g.

Train using jet images in JEWEL! (Supervised Learning)

Quark vs. Gluon Jets with ML

Y. Chien, R. Elayavalli: 1803.3589





- *Lower the curve, the better the performance.*
- All methods explored tend to perform consistently, indicating that they may be picking up on similar features.

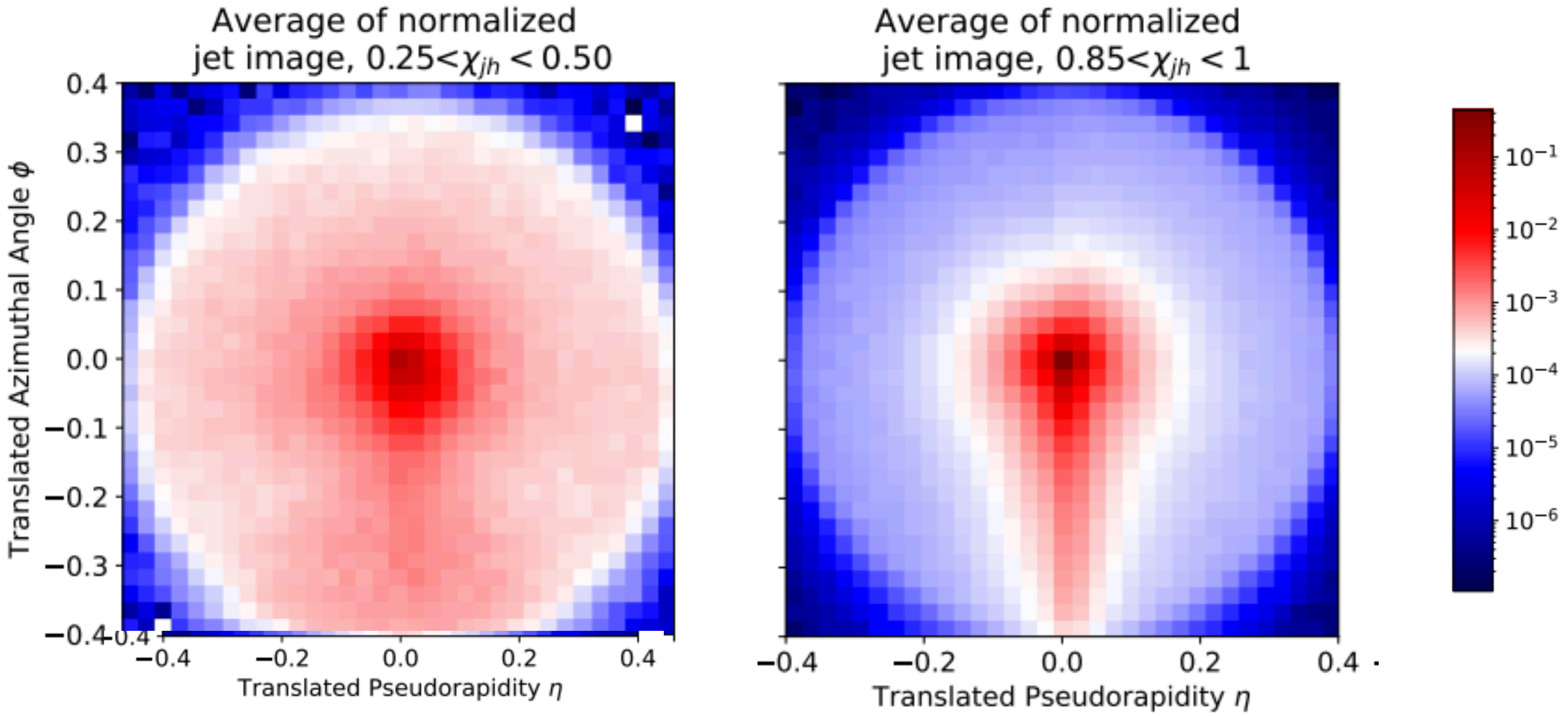
- The performance worsens for Pb—Pb, due to the large UE.

Quark and gluon discrimination is a difficult and ongoing effort in HIs!

Future: Apply these methods to data in pp and Pb—Pb!

Deep Learning Jet Modifications

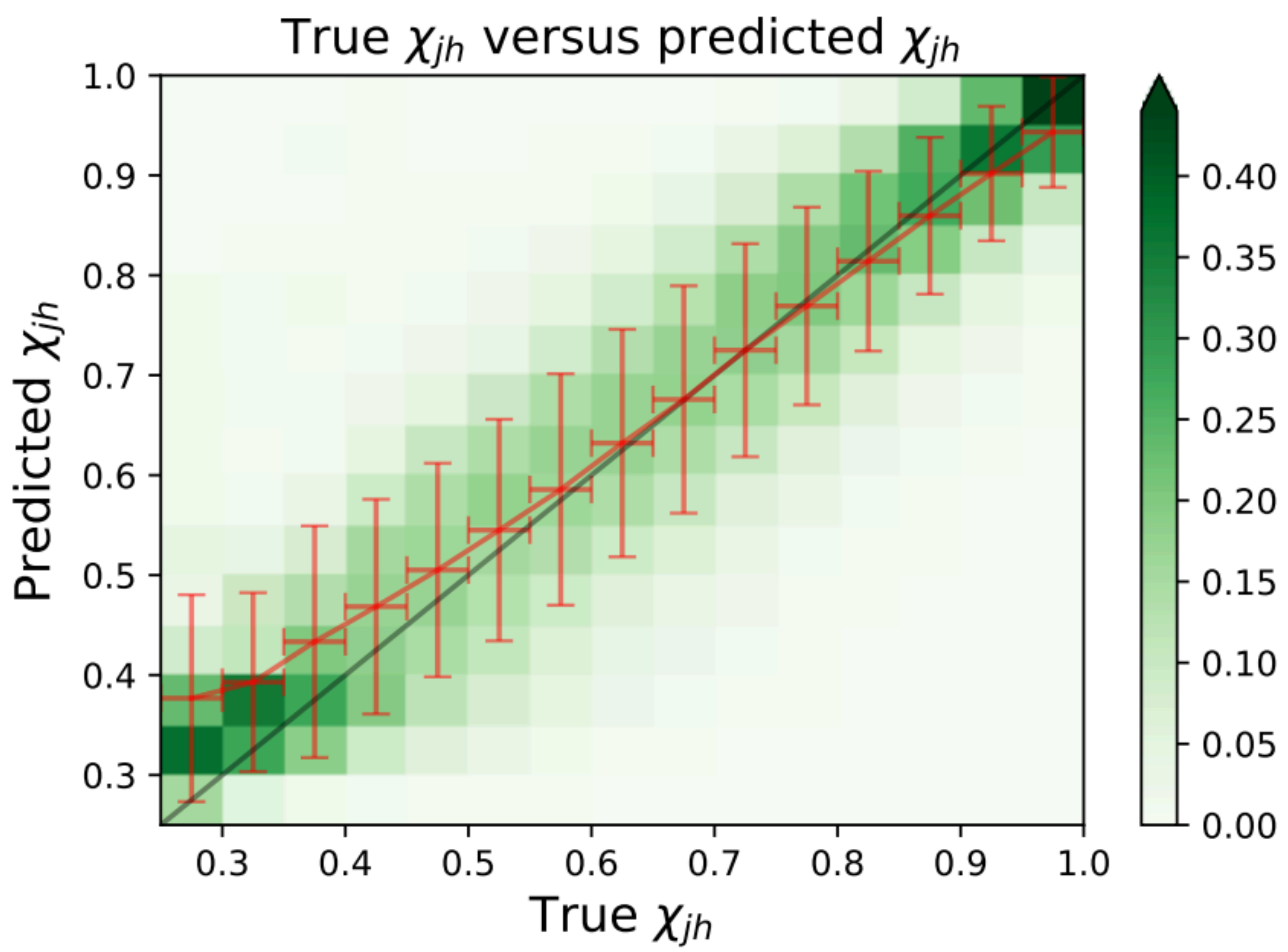
Use supervised learning on jet images with a CNN to perform the regression task of predicting the energy loss ratio in HI collisions (hybrid model).



quenched **unquenched**

$$\chi_{jh} = \frac{E_f^h}{E_i^h}$$

Y. Du, D. Pablos, K. Tywoniuk: 2101.07797



Shows good performance!

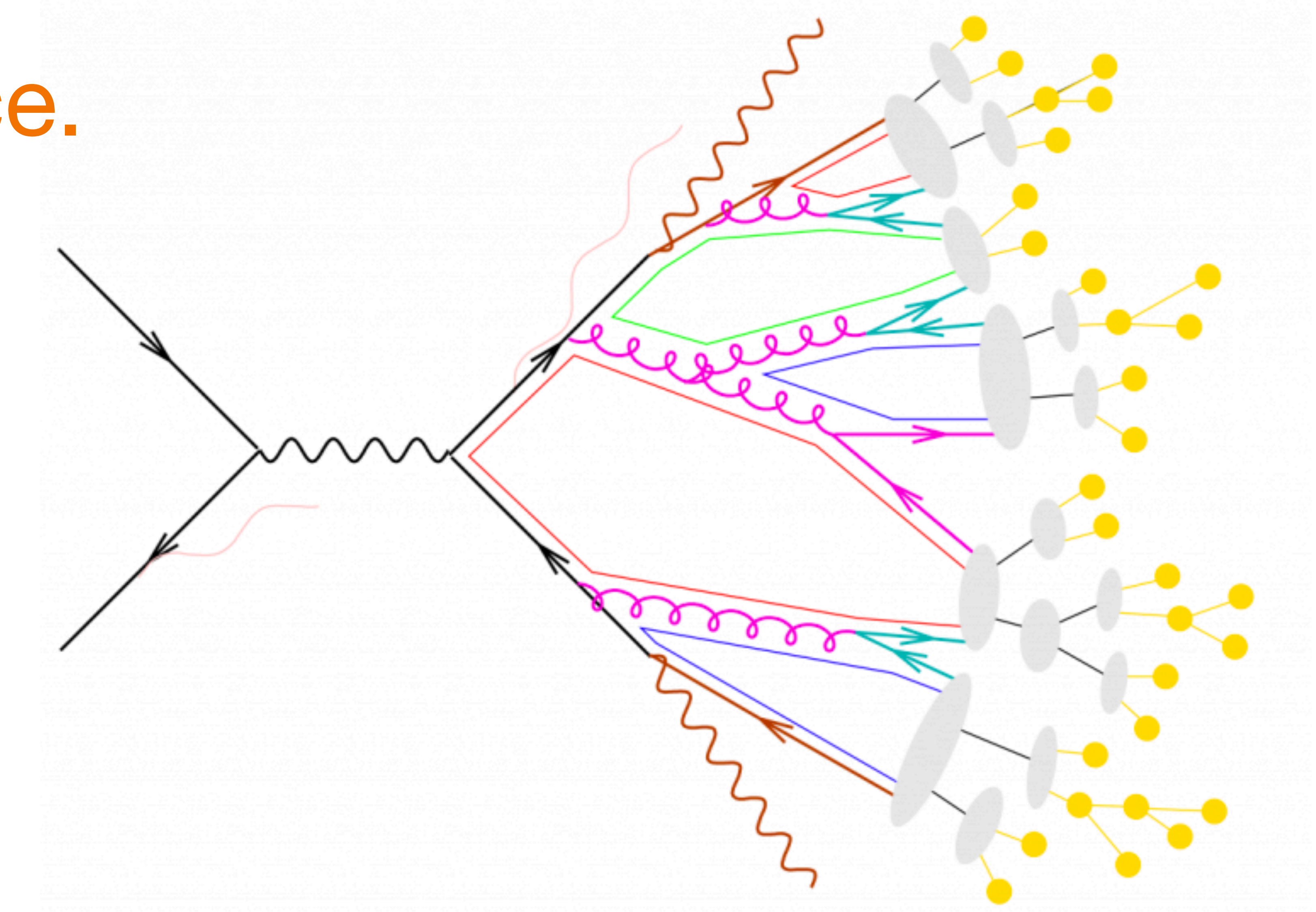
Very useful to separate and study quenched vs. unquenched jets as well as extracting the initial energy of the jet.

Future: Apply these methods to different models & variables, improve performance.

ML for Underlying Physics

Could we use ML to directly access some of these underlying physics mechanisms?

"Data"-based learning complements simulation-based inference.



~ Given an answer
~ "White Box" ML
~ Underlying physics

~ Domain knowledge
~ "Black Box" ML
~ Answer

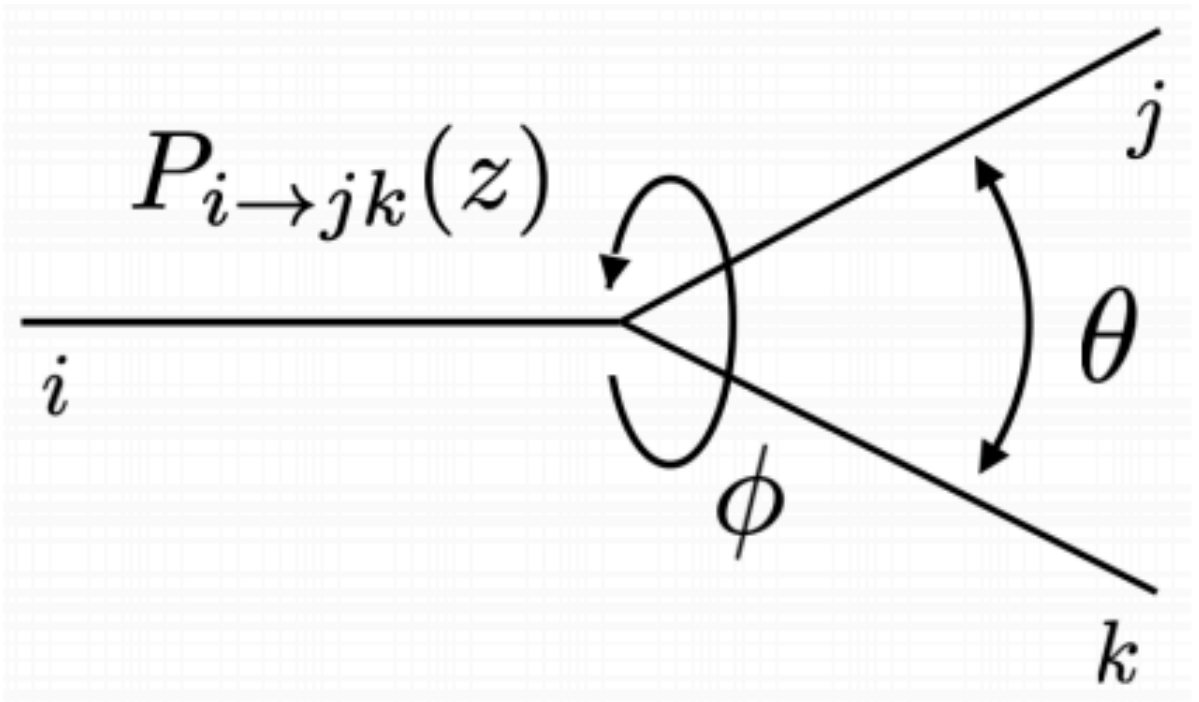
This is a long term effort!

→ Learning from data is difficult due to systematic experimental biases.

→ Helpful in understanding uncertainties or shortcomings of models!

Proof of Concept

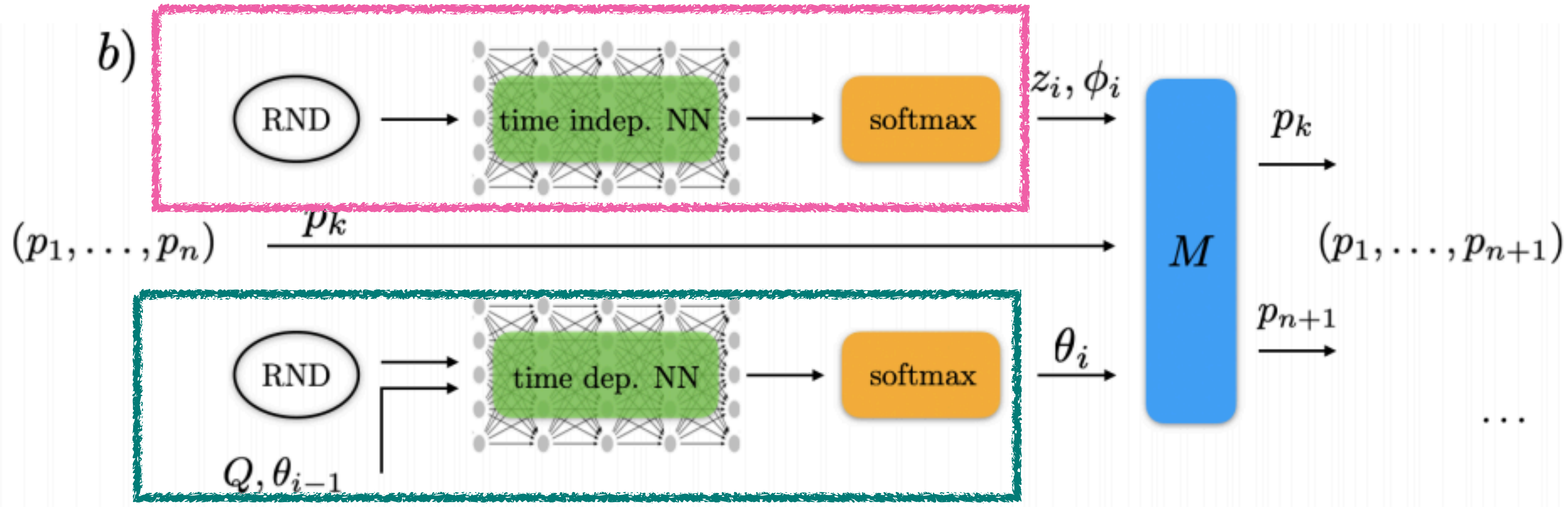
Y. Lai, D. Neill, M. Ploskon, F. Ringer: 2012.06582

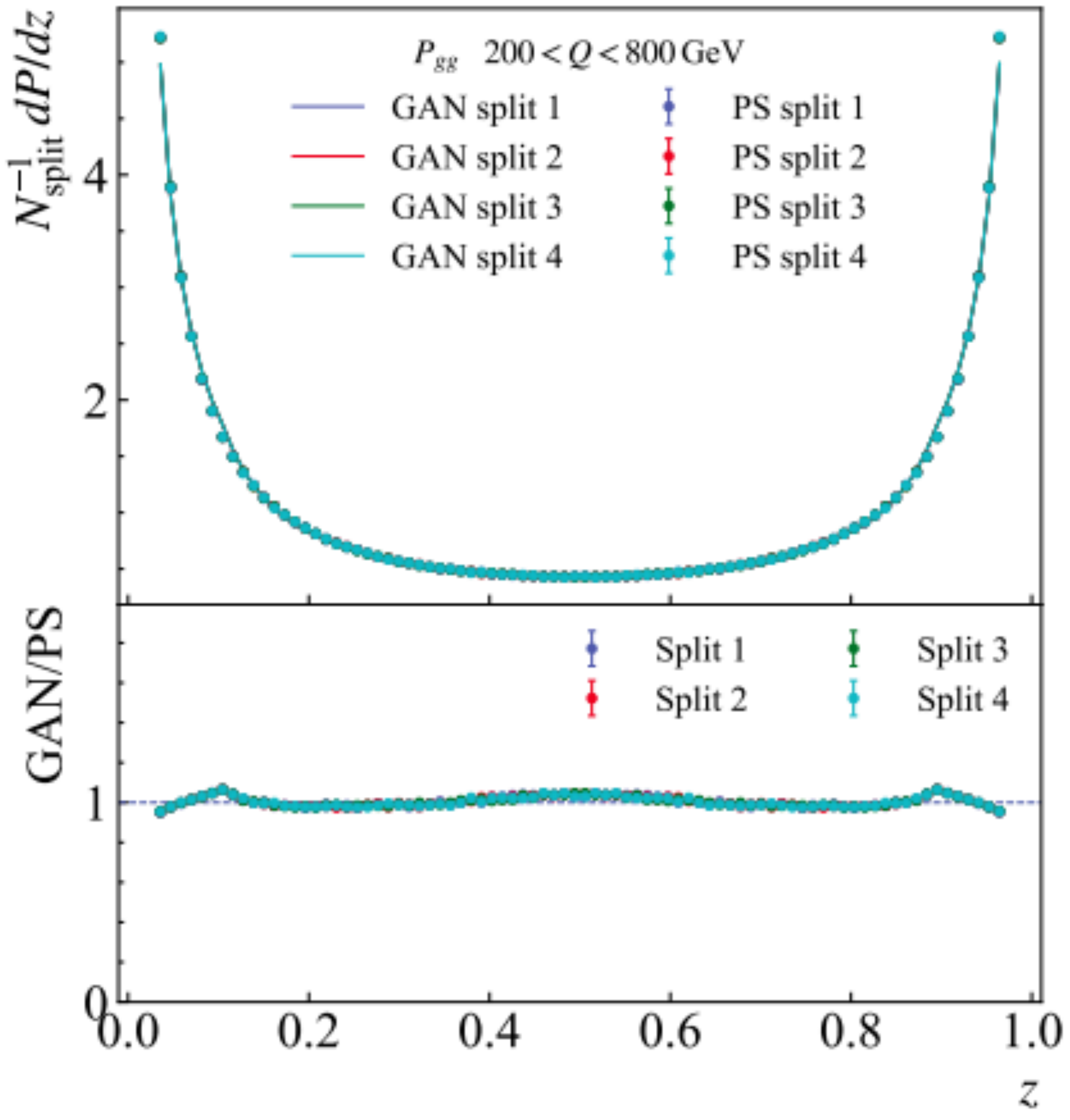


Extract info from the network in white-box ML.

This is done by splitting the GAN into two components.

1) Time independent learns the z, ϕ





2) Time dependent learns the θ

Was able to reproduce AP splitting function.

Future: Can we use ML to learn physics from data?

Outline

What is ML and why is it useful for jets?

How is ML currently being used for jet measurements?

Ongoing challenges

Ongoing Challenges

How do we quantify uncertainty?

How can we construct more interpretable models?

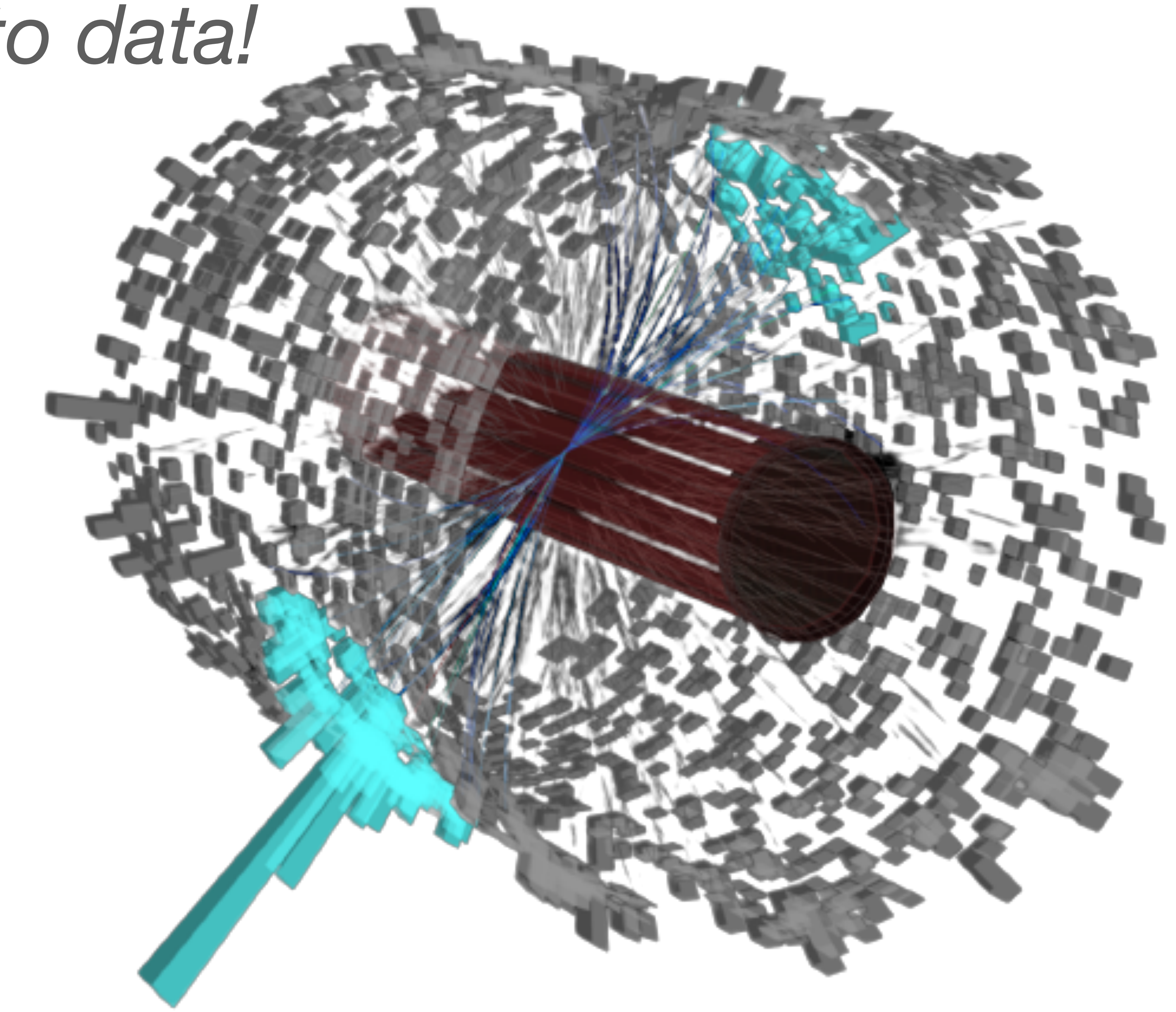
Do we need to standardize ML applications across experiments?

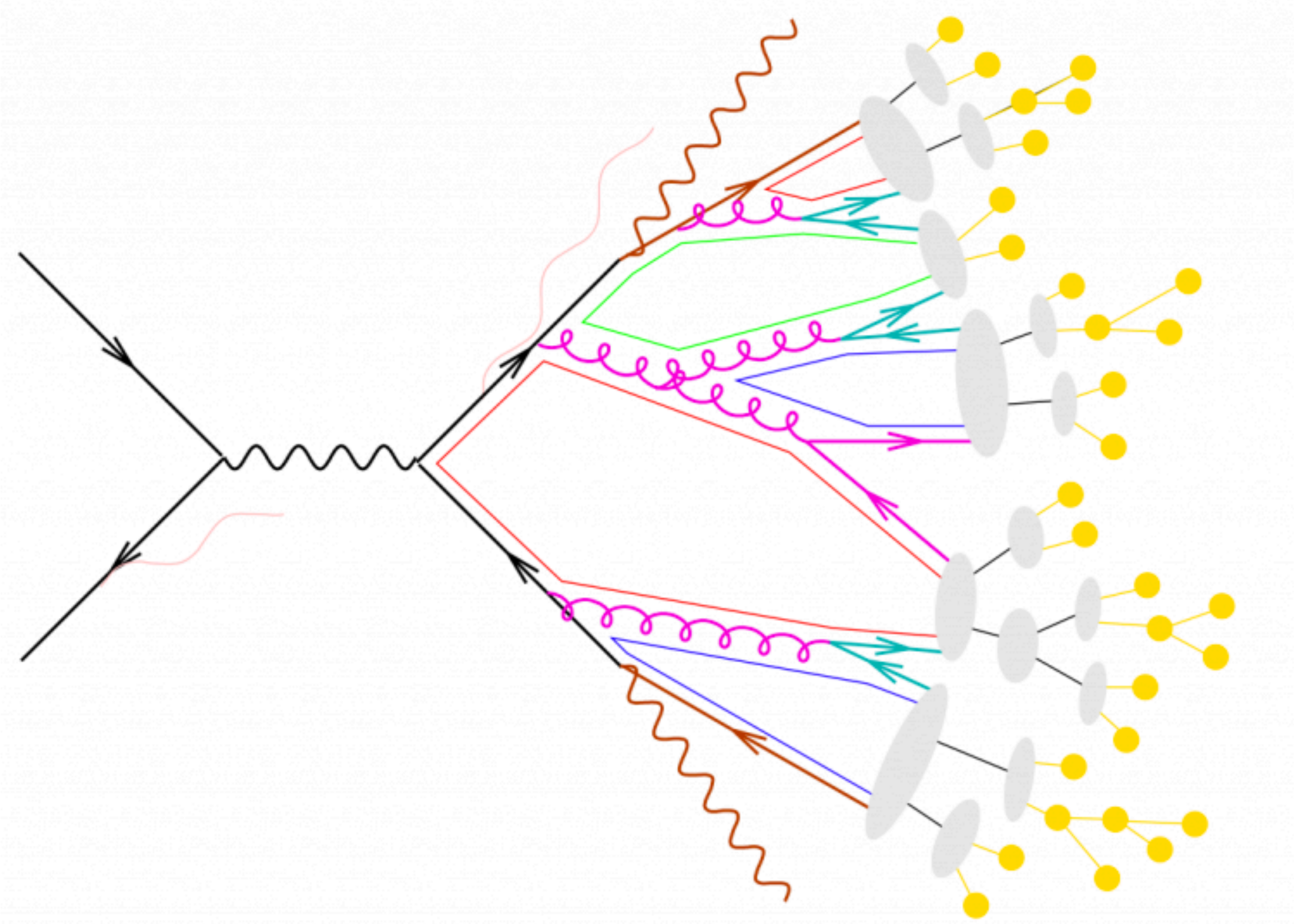
Many future and ongoing efforts to address these challenges!

Summary and Conclusions

Machine learning is a great tool for both understanding jet physics and making jet measurements, especially in heavy-ions!

Despite many great developments, still ongoing challenges → especially for applying ML to data!





Coming years are an exciting time for ML + jets → lots of uncharted phase space and opportunities to get involved!

Interested in ML? Here's where to start!

Resources Hub For Machine Learning

Here is a collection of resources that I've found helpful in order to learn about machine learning and explore its applications for jet measurements! Feel free to distribute this guide to anyone who may be able to benefit from it! If there are any resources that you would like to see included in this list, email me at **hannah.bossi@yale.edu** and I will happily add them! Enjoy! ~ Hannah

Resource guide available HERE!

Includes resources useful in learning about ML in general and also about ML for jet physics.

Thanks!!

Backup

Technical Details of the ML

Regression task where the regression target is the detector level jet p_{T}.

Here we are prioritizing a simple model!

Training sample 10%, testing sample 90%.

Implemented in *scikit-learn*. Default parameters used unless otherwise specified.

Shallow Neural Network

Shallow, 3 layers with [100, 100, 50] nodes

ADAM optimizer, stochastic gradient descent algorithm.

Nodes/neurons activated by a ReLU activation function.

Linear Regression

Normalization set to true by default.

Random Forest

Ensemble of 30 decision trees.

Maximum number of features set to 15.

Features for training

Ask ourselves two questions

How important is the feature to the model? → Feature Scores

Higher the feature score, more often variable is used in training.

How correlated is the feature with other features?

Feature	Score	Feature	Score
Jet p_{T} (no corr.)	**0.1355**	$p^{1}_{\mathrm{T,const}}$	0.0012
Jet mass	0.0007	$p^{2}_{\mathrm{T,const}}$	**0.0039**
Jet Area	0.0005	$p^{3}_{\mathrm{T,const}}$	0.0015
Jet p_{T} (area based corr.)	**0.7876**	$p^{4}_{\mathrm{T,const}}$	0.0011
LeSub	0.0004	$p^{5}_{\mathrm{T,const}}$	0.0009
Radial moment	0.0005	$p^{6}_{\mathrm{T,const}}$	0.0009
Momentum dispersion	0.0007	$p^{7}_{\mathrm{T,const}}$	0.0008
Number of constituents	0.0008	$p^{8}_{\mathrm{T,const}}$	0.0007
Mean of constituent p_{T}s	**0.0585**	$p^{9}_{\mathrm{T,const}}$	0.0006
Median of Constituent p_{T}s	0.0023	$p^{10}_{\mathrm{T,const}}$	0.0007

Iteratively remove unimportant or highly correlated features!

R.Haake, C. Loizides Phys. Rev. C 99, 064904 (2019)

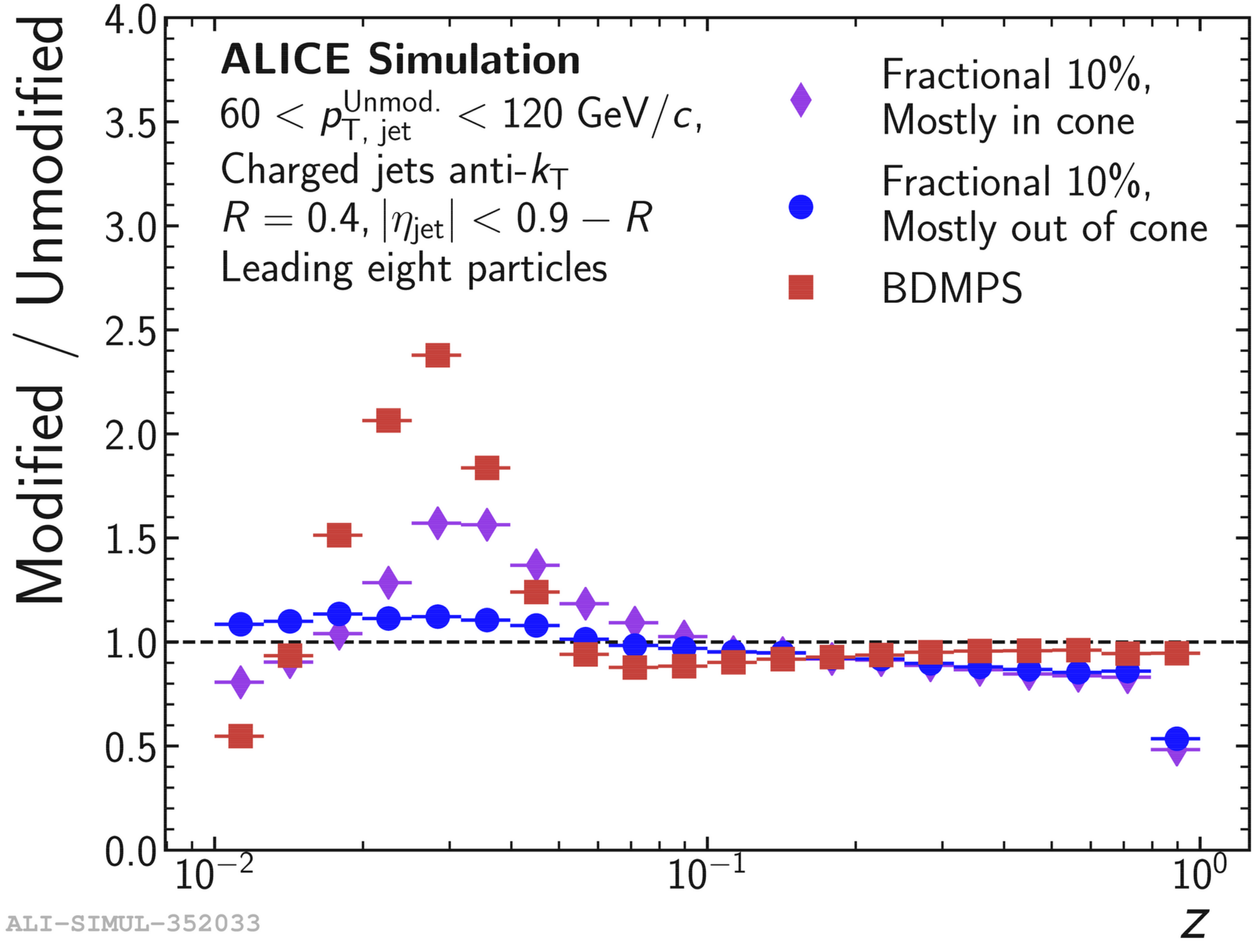
Toy model studies

We study a toy model with three different ways to alter constituents of the jet, changing the fragmentation.

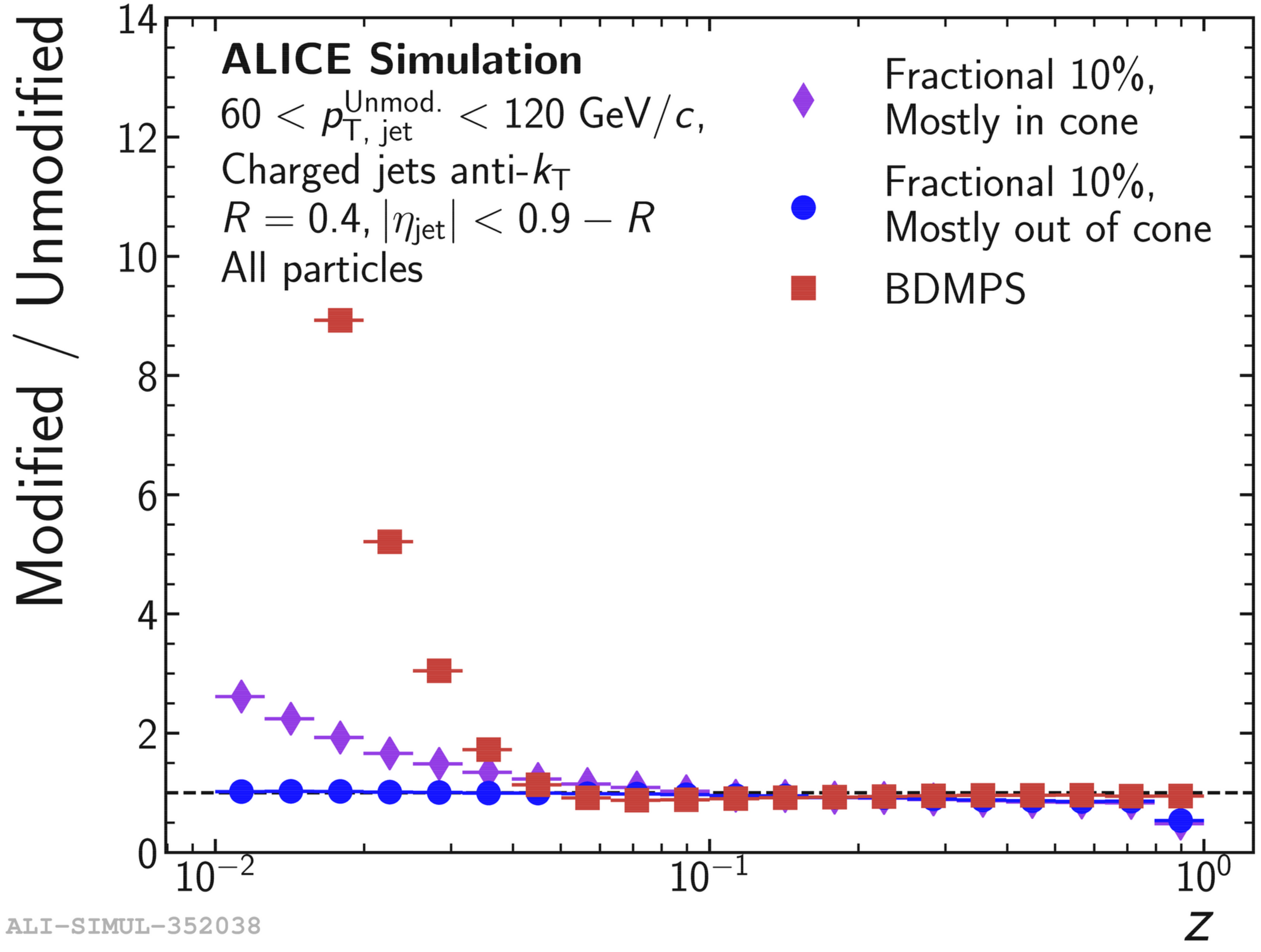
Use prior knowledge of behavior at intermediate p_{T} to create a variation in the fragmentation!

Modification to the fragmentation function

Leading 8 particles



Inclusive particles



Toy model modifications indeed modify the fragmentation, some modifications are more extreme than others.

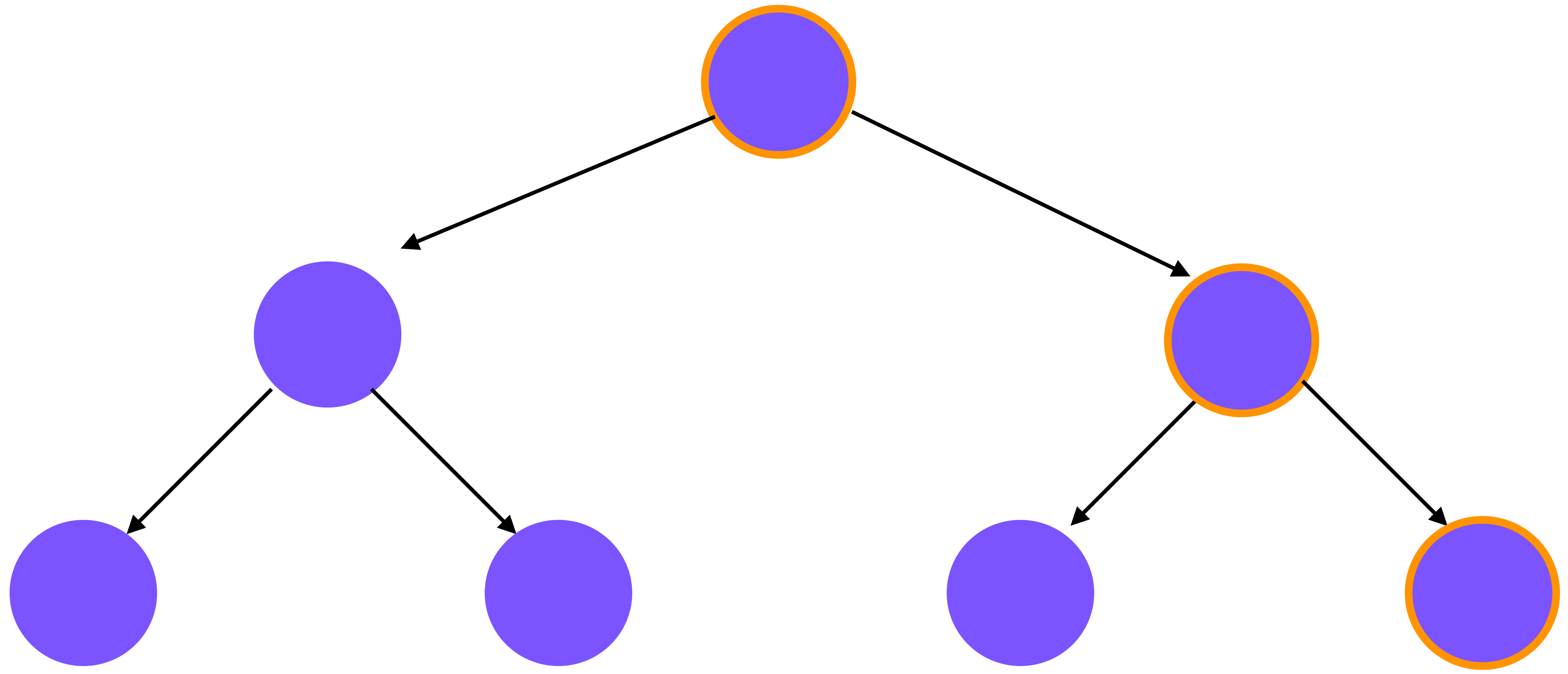
8 leading particles are what we chose to train on.

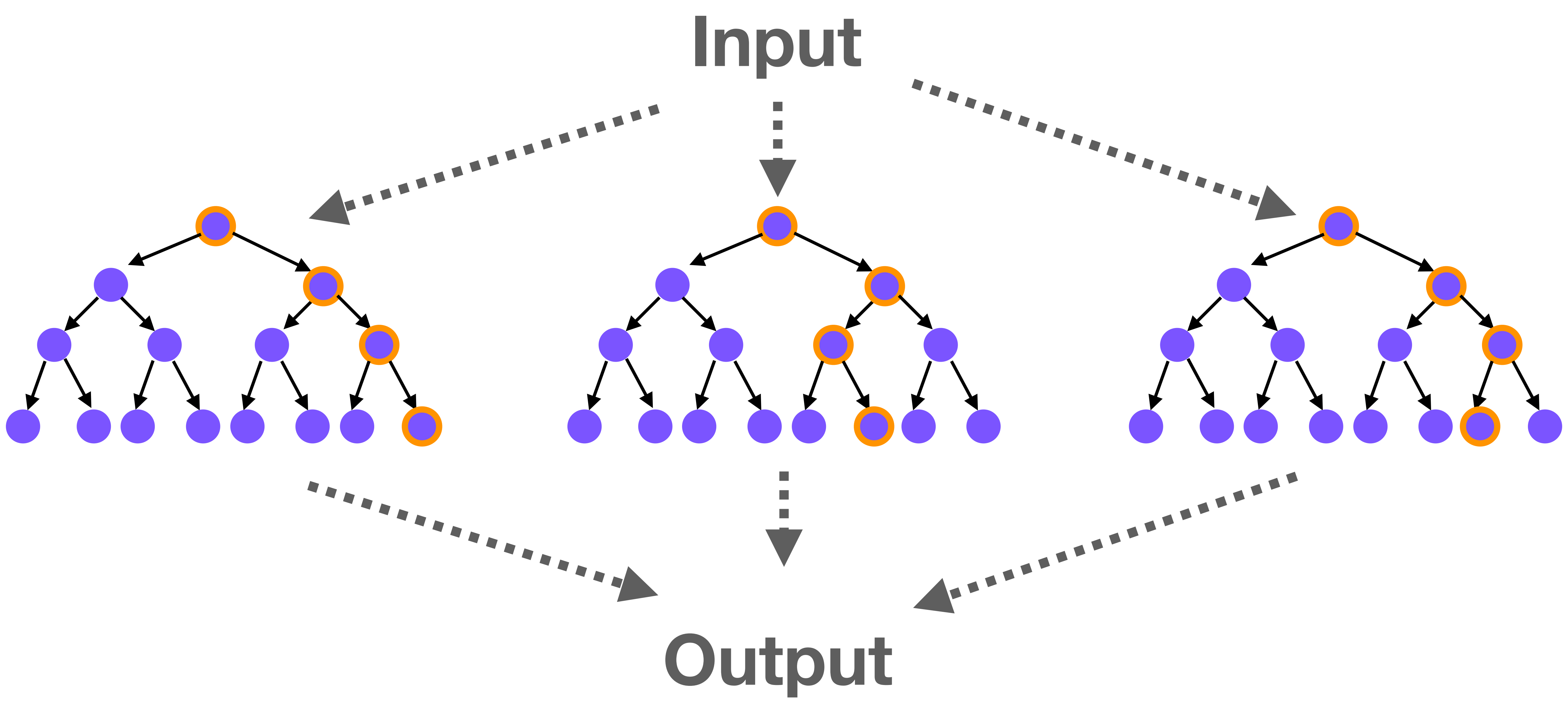
Intro to Random Forest

Random forests are composed of decision trees.

Decision trees are a set of rules organized in a tree structure.

Each node is a rule which subdivides the dataset into two or more parts (think 20 questions).





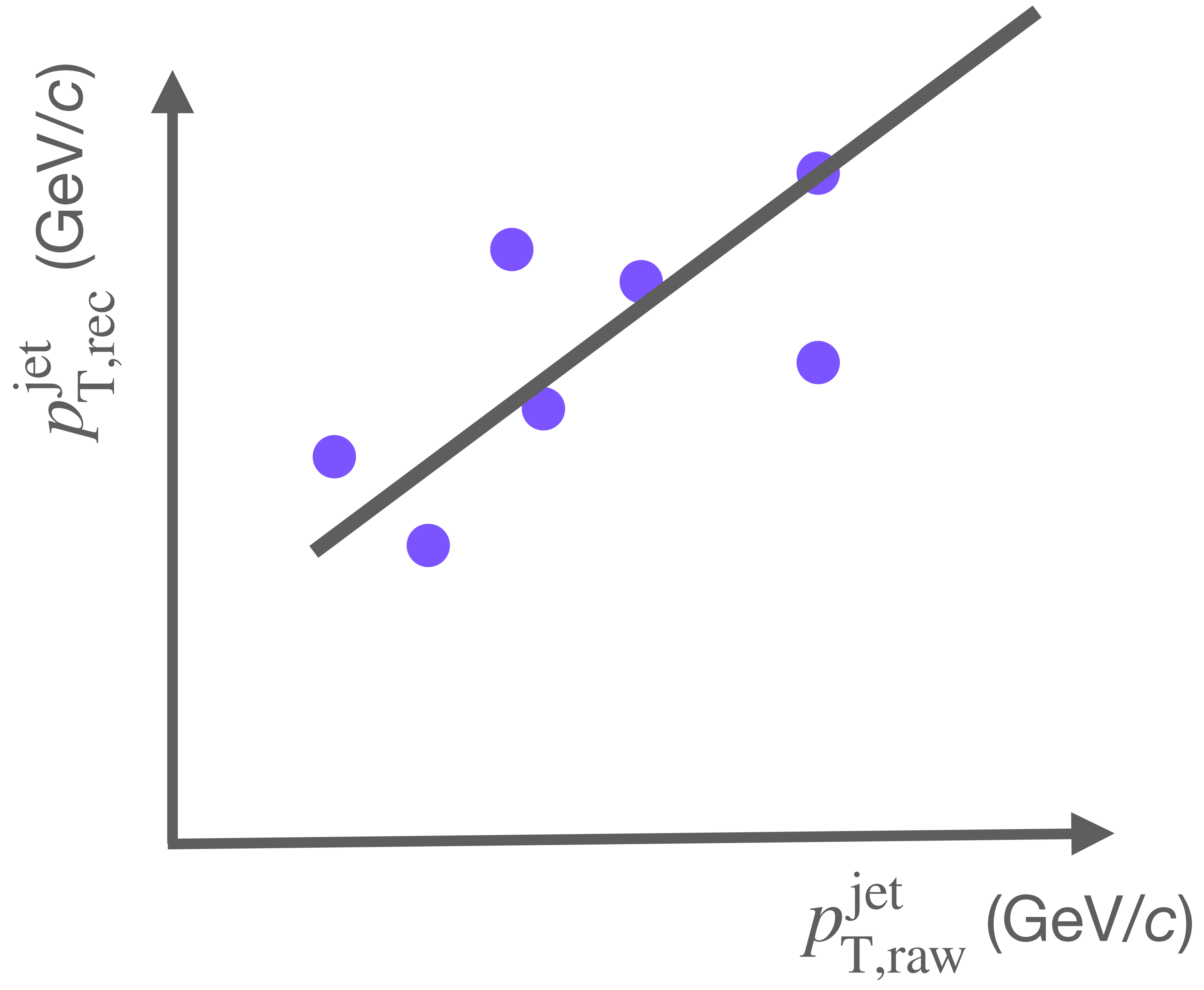
Output of the random forest is a combination of the output of each of the decision trees.

In training, the algorithm sets up the rules of each decision tree.

Intro to Linear Regression

Linear regression predicts the value of a **dependent variable** based on a given **independent variable** (feature x1 with a given weight w1).

$$y = b + w_1 x_1$$



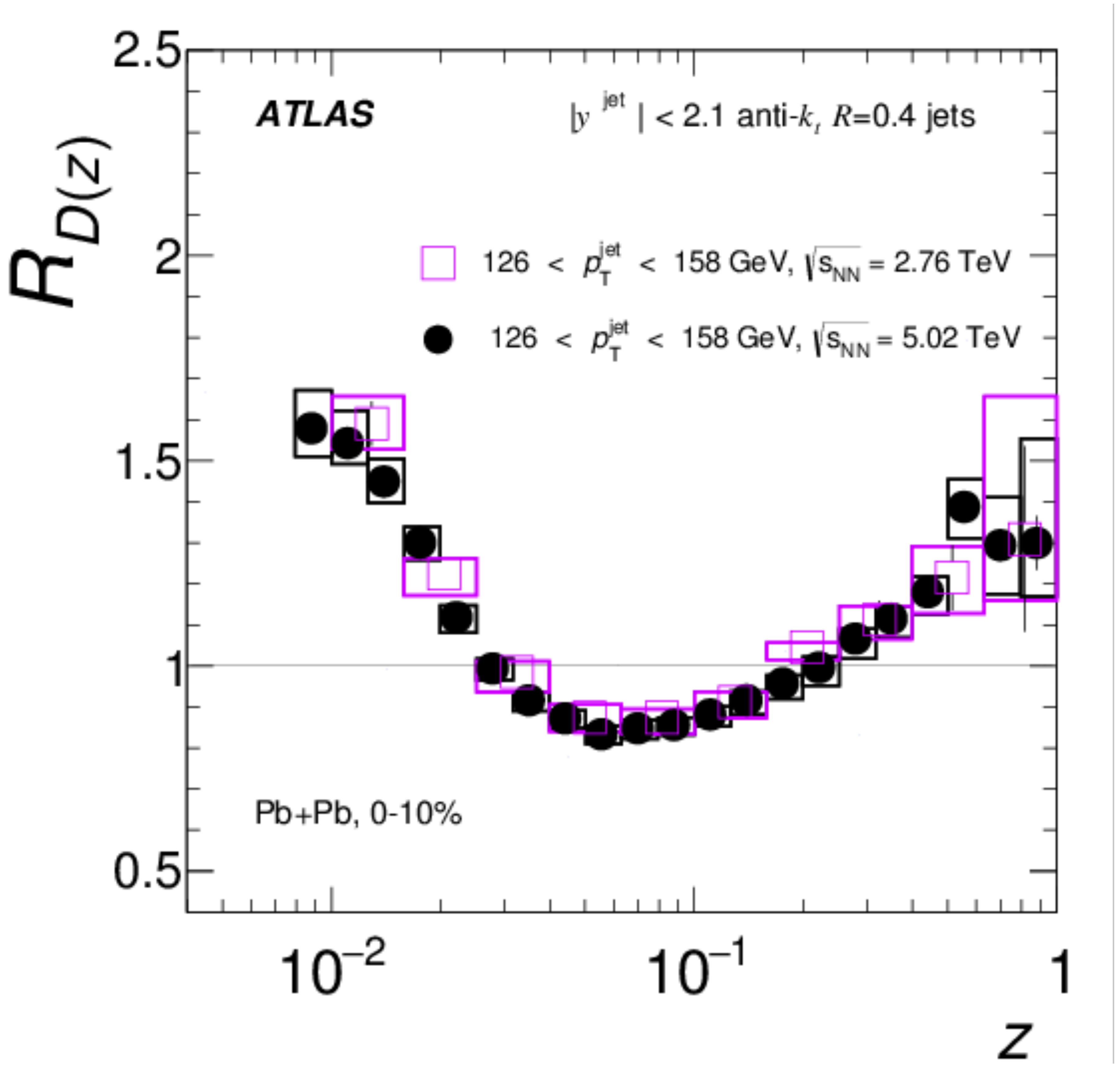
The example at the right is a simplified view in reality we have multiple features each having a separate weight.

$$y = b + w_1 x_1 + w_2 x_2 + w_3 x_3 \ldots\ldots$$

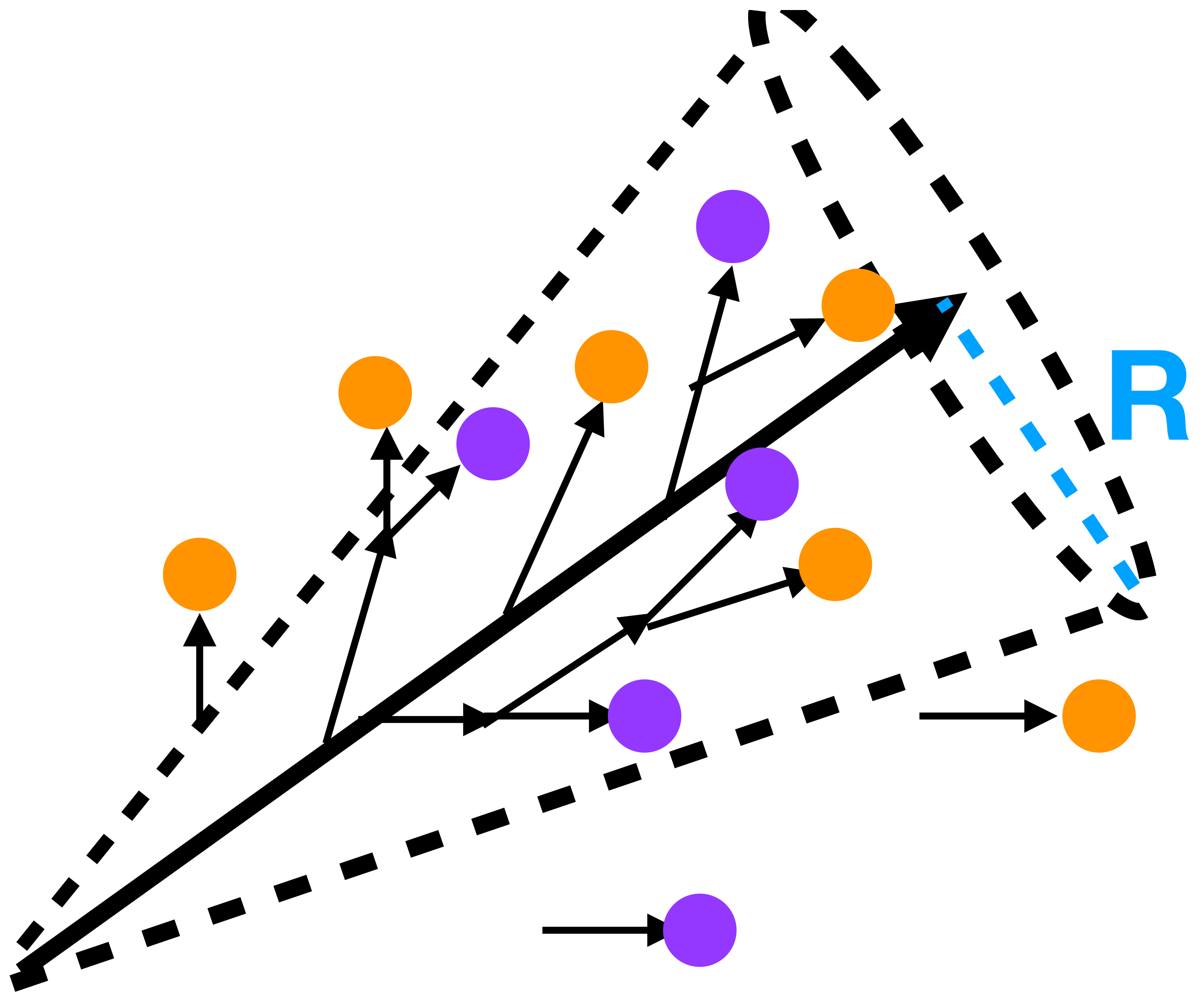
Training determines the optimal weight for each feature.

Fragmentation bias

Learning on constituents introduces a fragmentation bias.



[Phys. Rev. C 98, 024908 (2018)]



We learn on a PYTHIA fragmentation.

We know that fragmentation in heavy-ion collisions is modified by the presence of the medium.

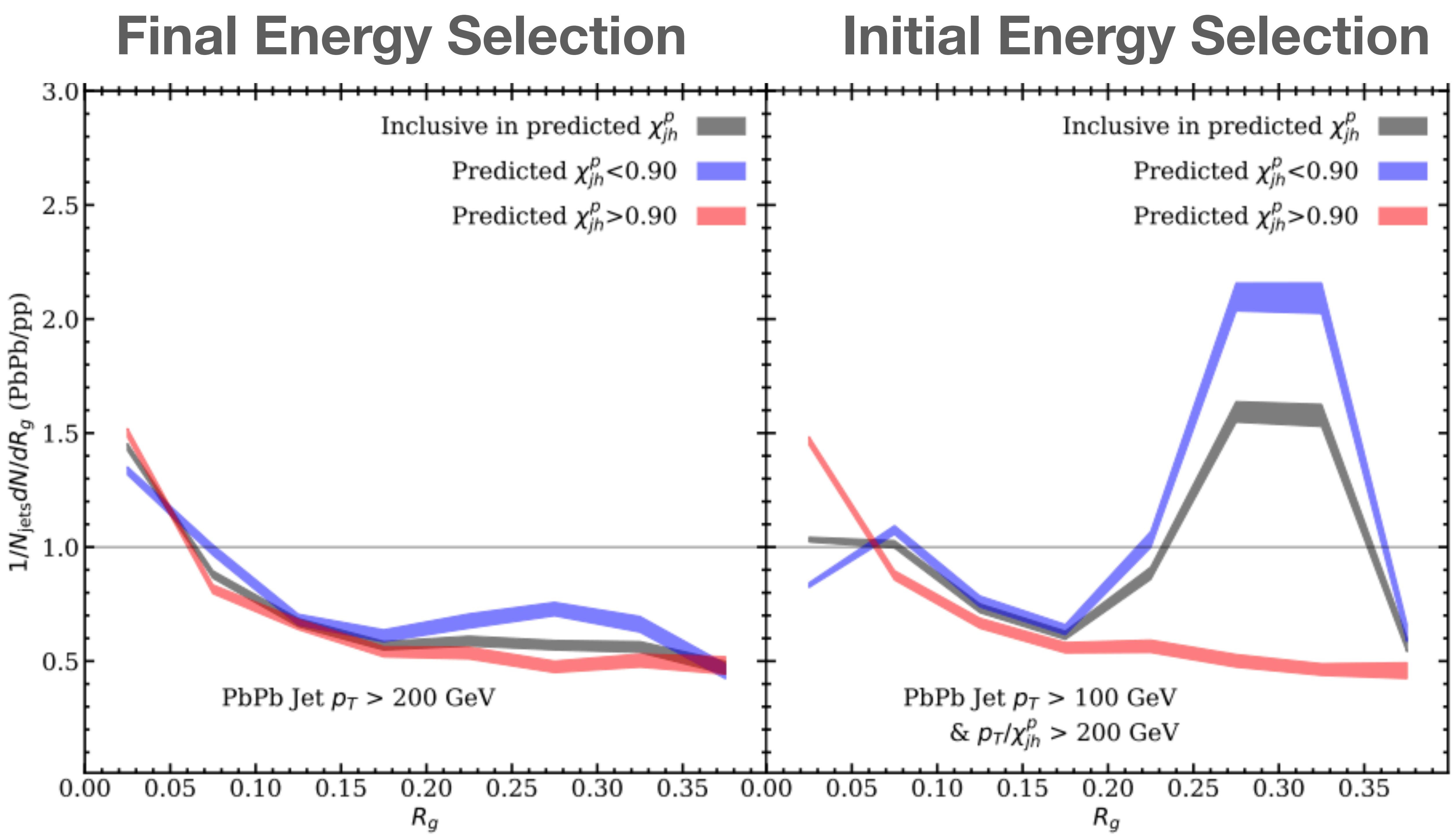
We want to investigate how this impacts the final result we get with ML!

Deep Learning Jet Modifications

Ex: Groomed jet radius

arXiv:2101.07797

Small R_g (Collimated)   Large R_g (wide)



Unquenched jets != jets in vacuum
→ selection bias!

Jets that fall into the “unquenched class” tend to be narrower than average jet population in vacuum.

ML is a cool tool to begin to think about selection biases and its impact on how we see quenching!

Future: Apply these methods to different models & variables, improve performance.