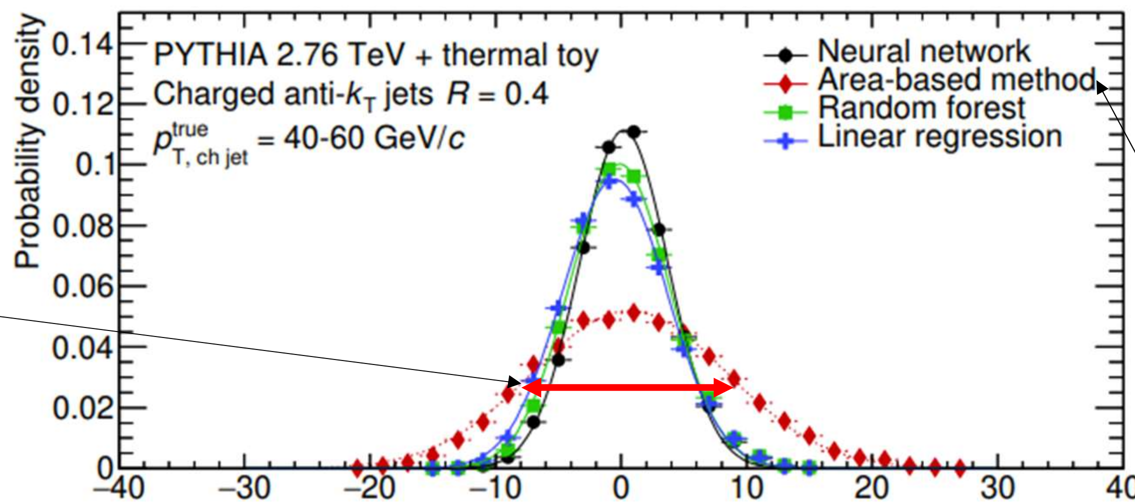
Interpretable ML for jet background subtraction

Tanner Mengel

Motivations

- Applications of ML to jet background subtraction achieve improved p_T resolution at LHC energies, particularly at low jet momentum.
- Can a neural network outperform traditional background subtraction methods at RHIC energies?

Width of δp_T determines momentum resolution of method.



$\delta p_T = \left(p_{T,jet}^{pred.} - p_{T,jet}^{truth}\right)$

$p_T^{Corr.} = p_T^{raw.} - \rho A_{jet}$

Phys. Rev. C 99, 064904 (2019)

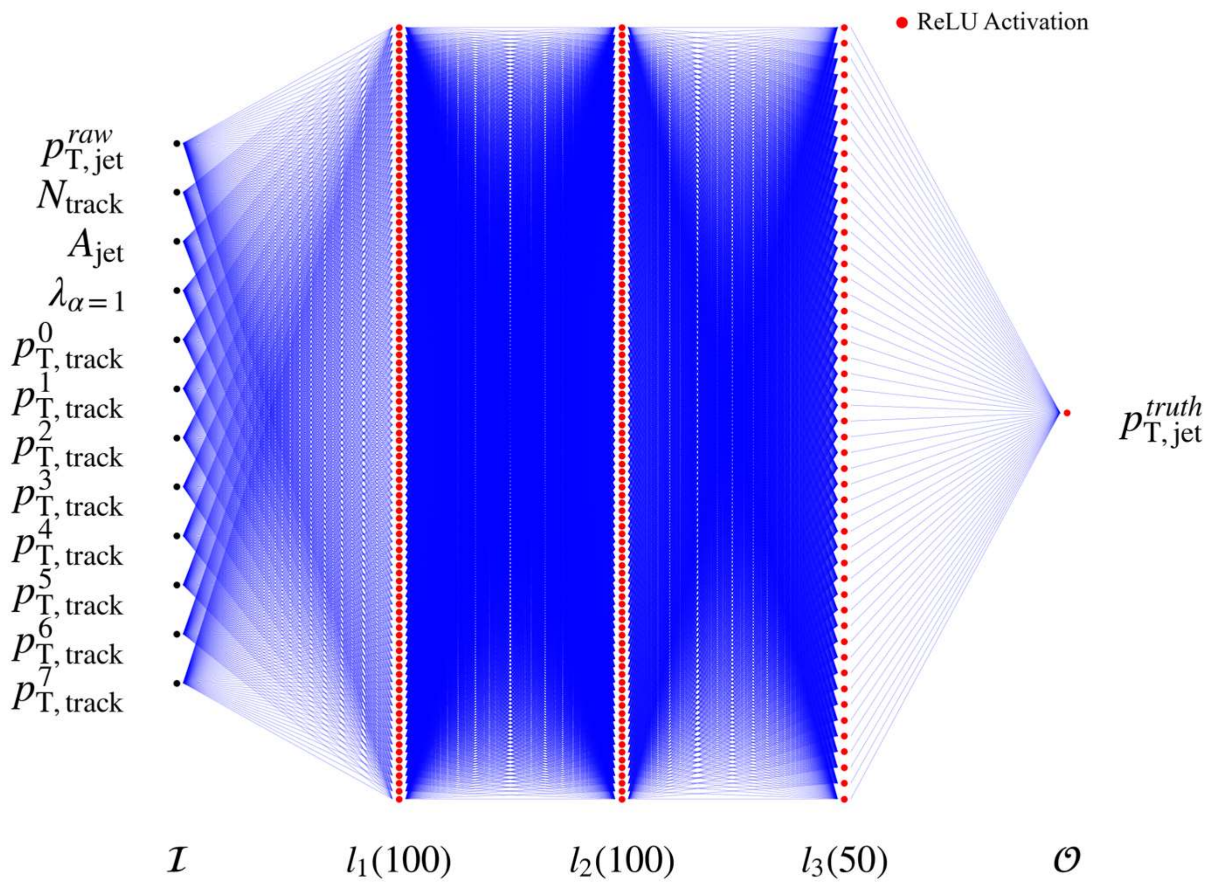
Dataset

- Merge PYTHIA pp collisions into TennGen heavy ion background
- Find charged anti- k_T jets in merged event and geometrically match them back to PYTHIA jets
- Use matched PYTHIA jet momentum as ground truth $p_{T,jet}^{true} \equiv p_{T,jet}^{pythia}$

$+$ TennGen

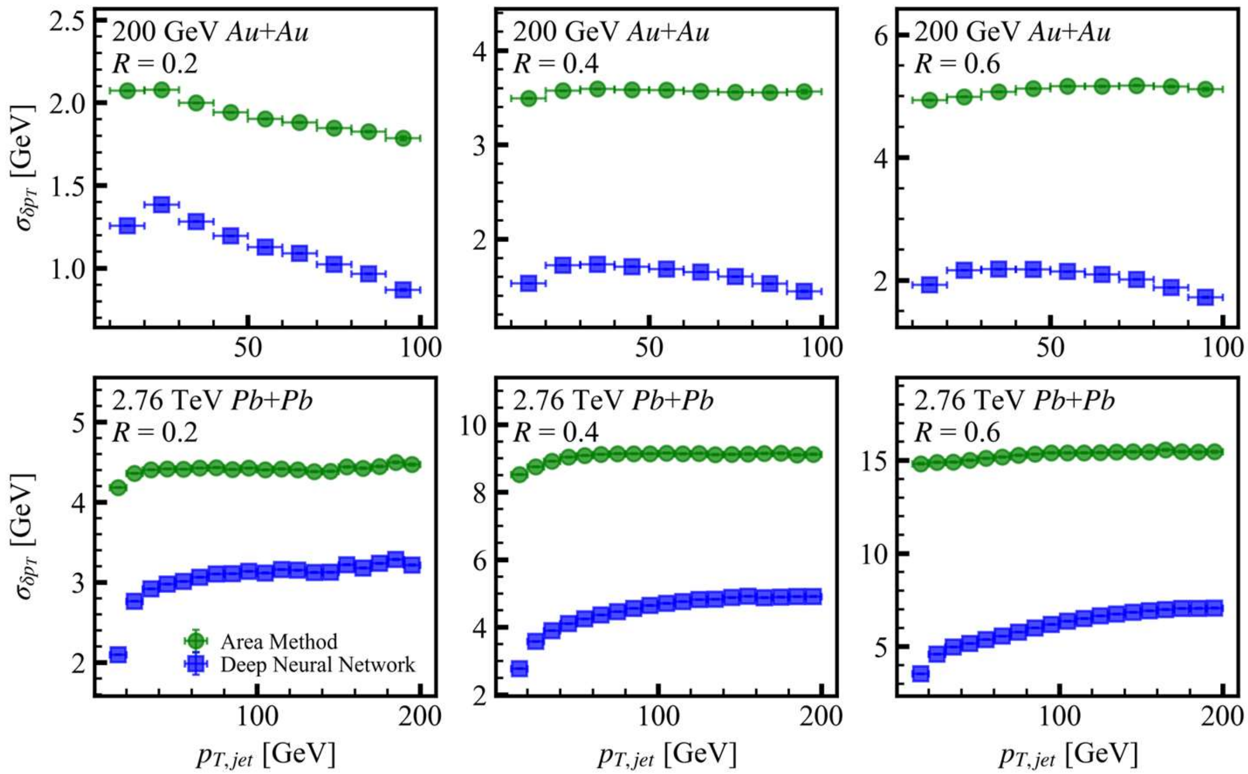
Architecture

- Sequential dense network with 3 hidden layers
- Mean squared error loss
- ADAM optimizator



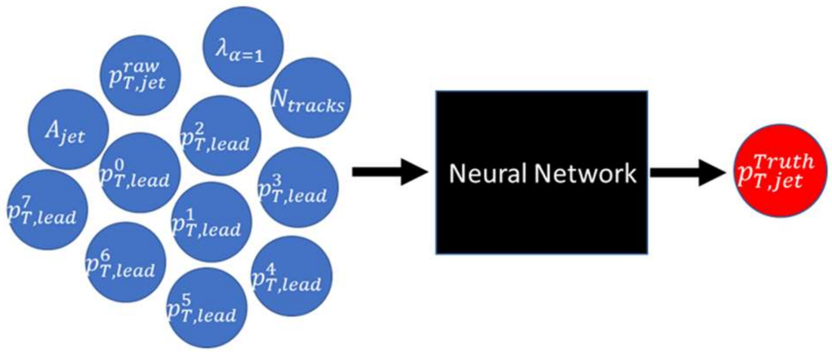
Initial Results

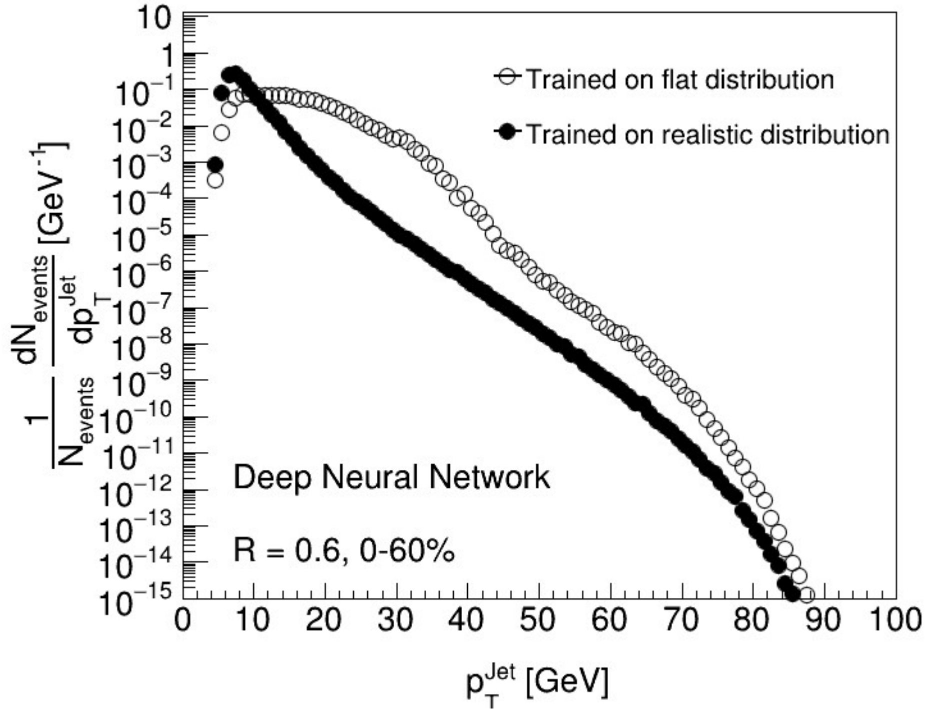
- Width of δ p_T from neural network is 2-3 times smaller for all jet p_T.



The Elephant

- Predictions biased by training data.
- Predictions only reliable within training phase space.
- Offers little/no explanation for underlying physics.





Revised Question

Why does

~~Can~~ a neural network outperform traditional background subtraction methods?

Interpretable ML

1. Method must be equivalently applicable to data and simulation.

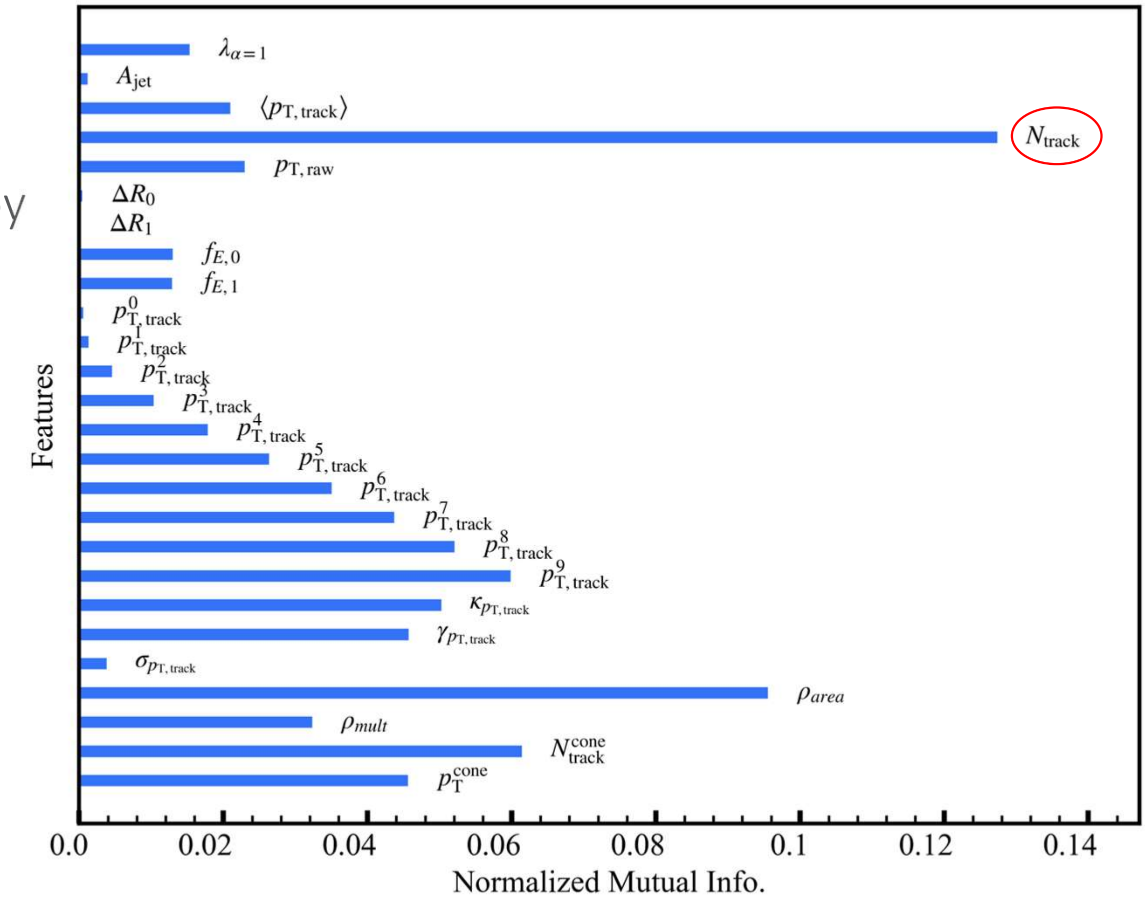
2. Predictions must be understood outside the range of training set.

3. Systematic uncertainties can be assessed for predictions.

4. Learned relationships can be directly observed.

Domain Knowledge

- Jet multiplicity has largest mutual information to truth momentum
- Background fluctuations are driven by multiplicity

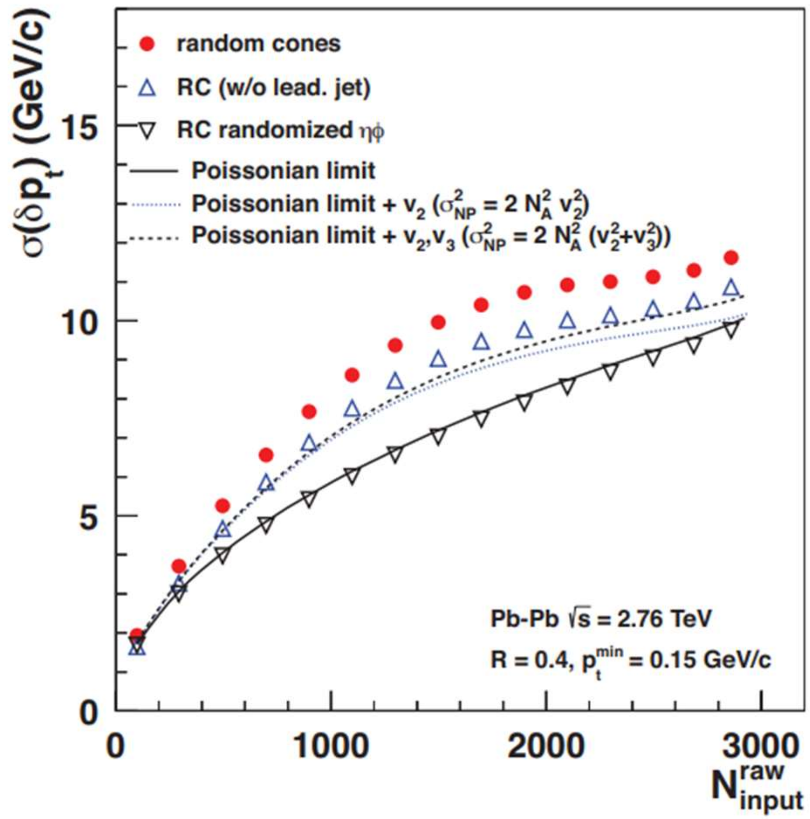


Multiplicity Method

- Form eliminates contributions from Poissonian fluctuations and hydrodynamical flow

$$\sigma(\delta p_T) = \sqrt{N \cdot \sigma_{p_T}^2 + N\langle p_T^{bkgd}\rangle^2 + 2N^2\langle p_T^{bkgd}\rangle^2 \sum_{n=1}^{\infty} v_n^2}$$

$$p_T^{corr.} = p_T^{raw} - \langle p_T^{bkgd}\rangle \cdot N_{bkgd}$$

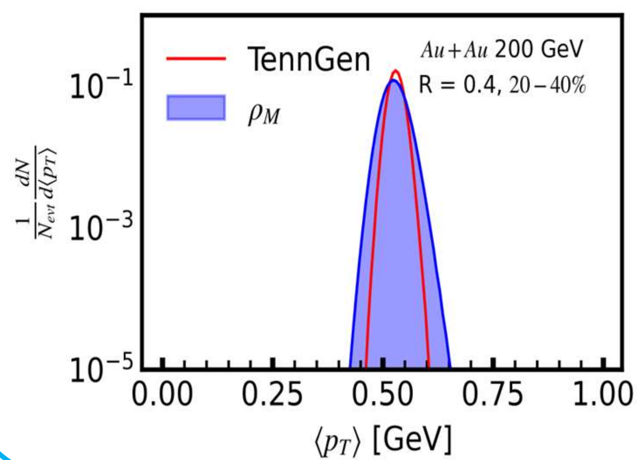


JHEP 03 (2012) 053

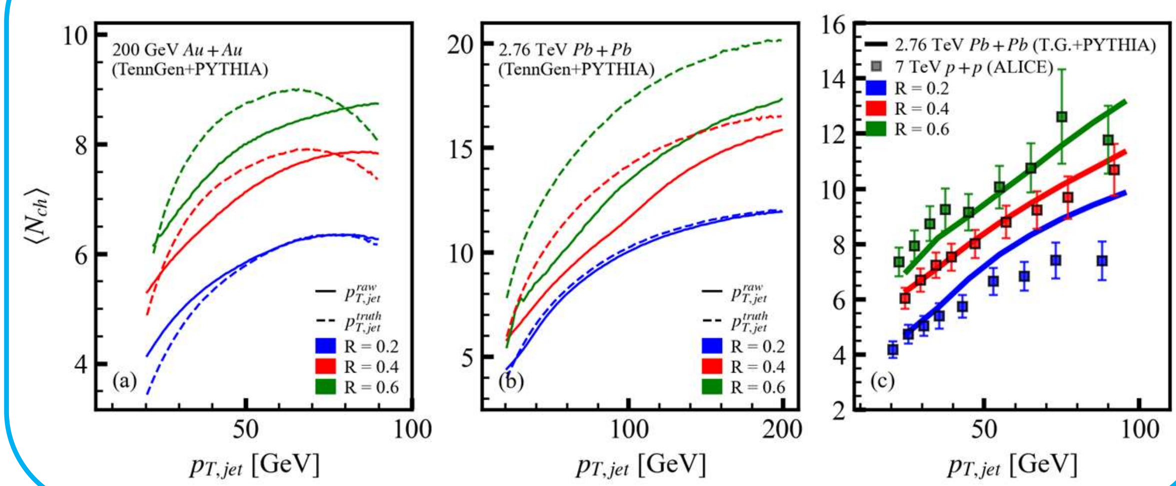
Multiplicity Parameters

$$p_T^{corr.} = p_T^{raw} - \rho_M \cdot \left(N - \langle N_{pythia} \rangle\right)$$

$$\rho_M \equiv median\left\{\langle p_{T,track}\rangle^{k_T\ jets}\right\} \approx \langle p_T^{bkdg}\rangle$$

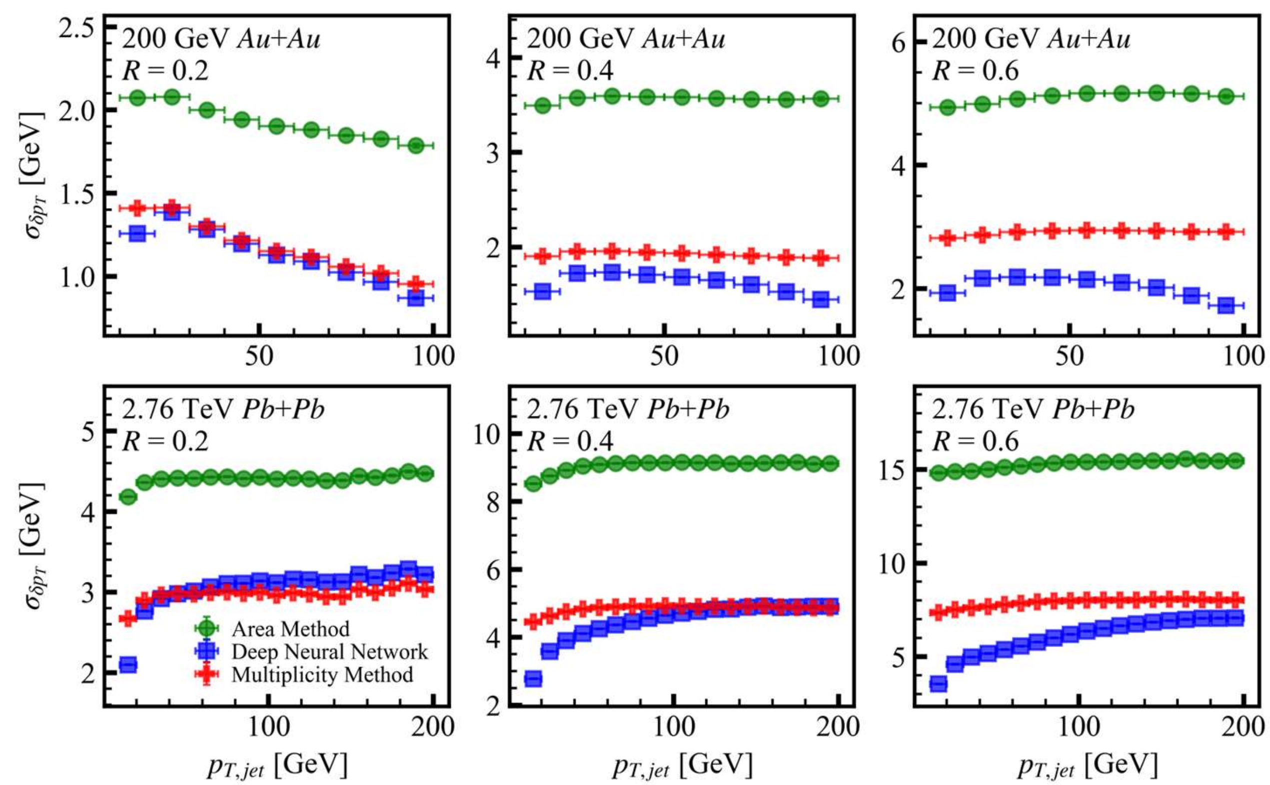


$$N_{bkgd} = N - N_{signal} \approx N - \langle N_{pthyia} \rangle$$



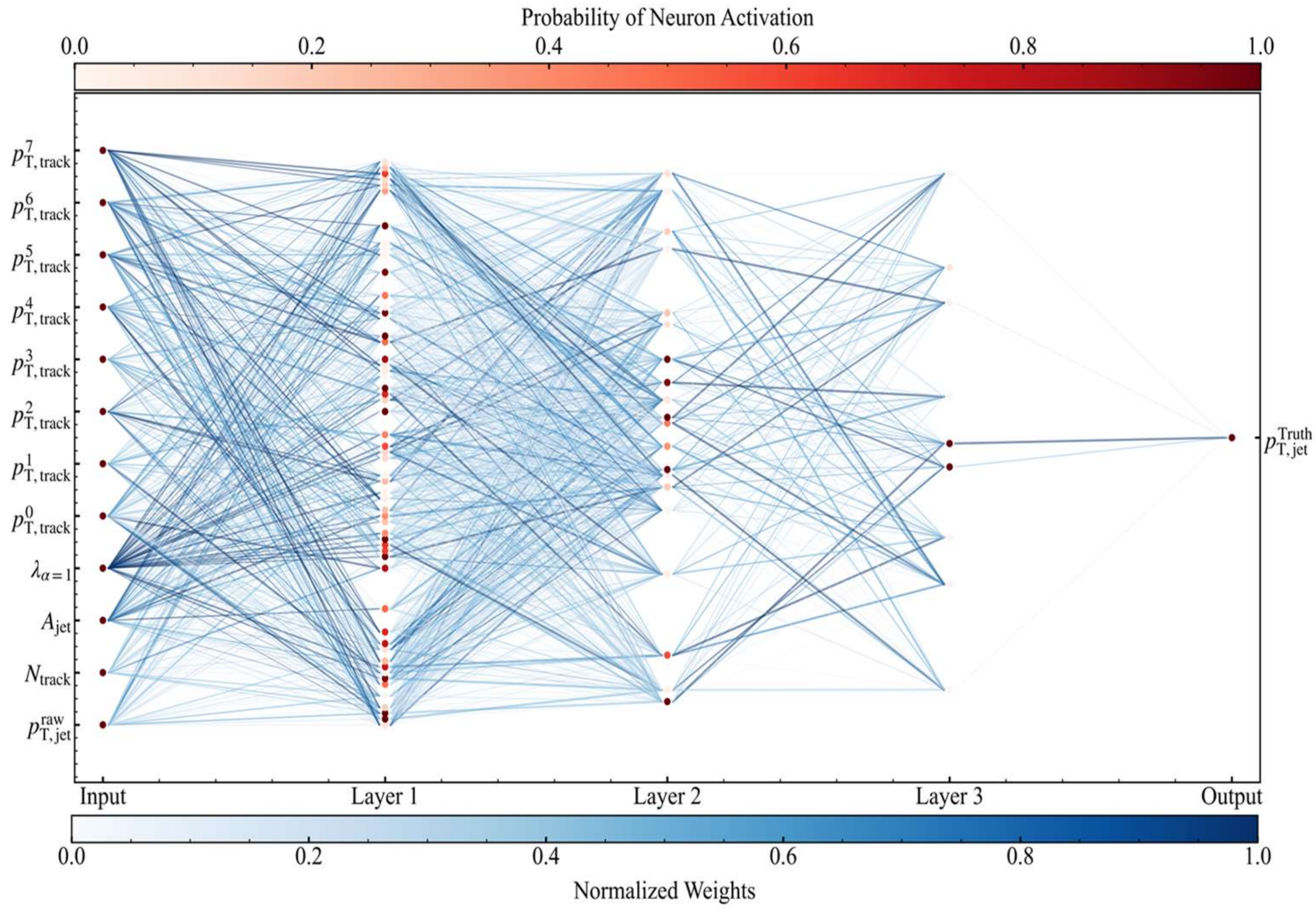
Momentum Resolution

- Multiplicity method reproduces much of improvement achieved by neural network.



Complexity vs. Accuracy

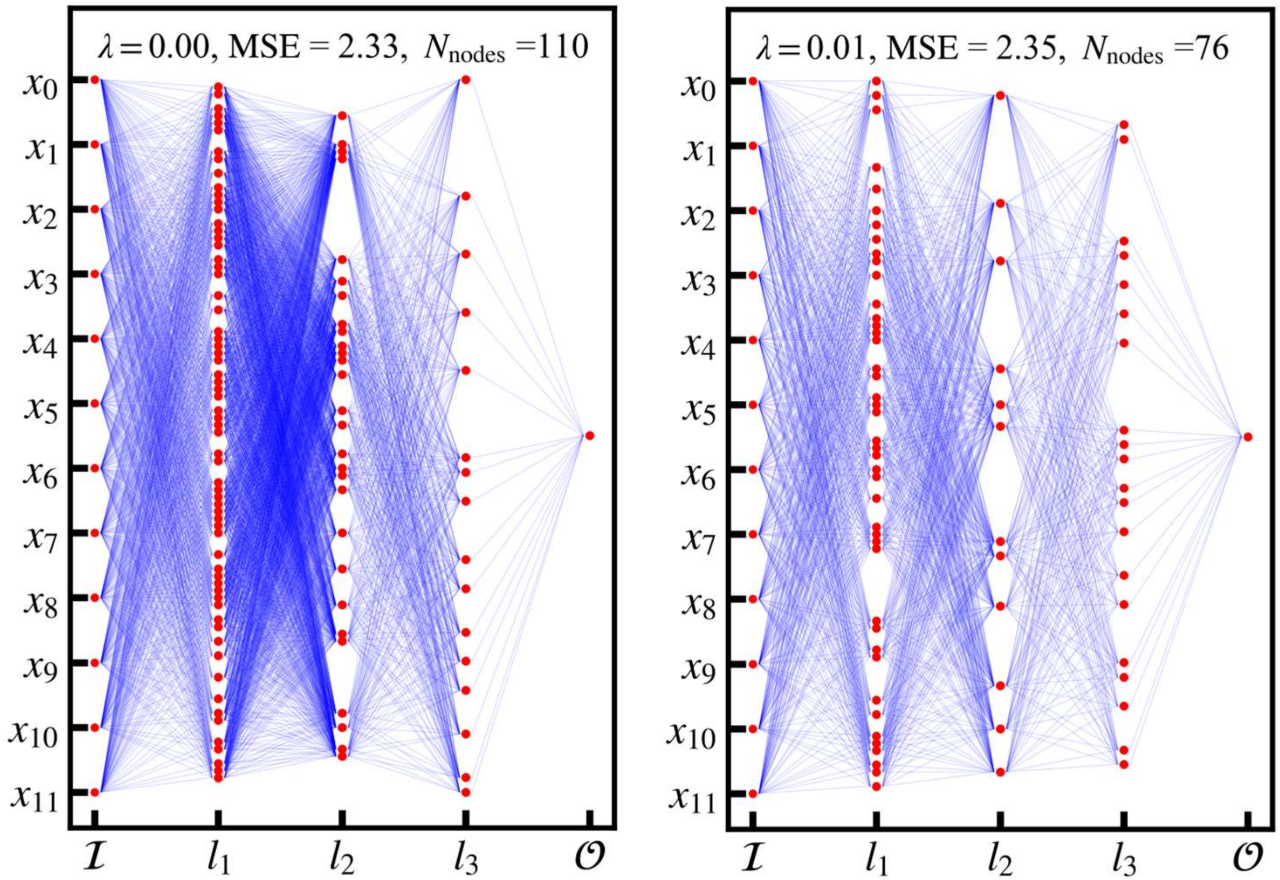
- Multiplicity method has 2 parameters
- Neural network has over 16,000 learned weights



Reducing Redundancy

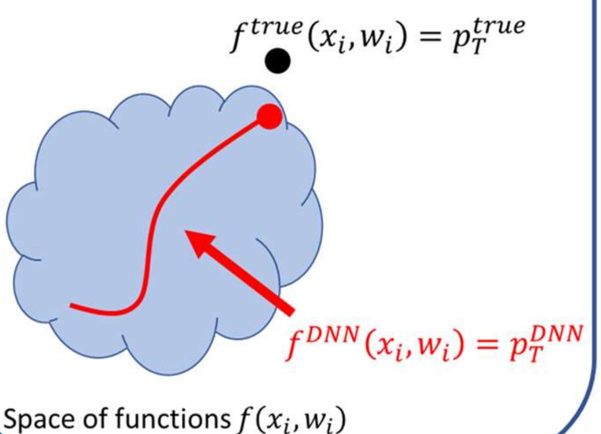
- Introduce kernel regularization term to loss function.

$$\mathcal{L} = \frac{1}{n}\sum_{i}^{n} |p_{T,jet}^{pred.} - p_{T,jet}^{truth}|^2 + \boxed{\lambda \|\boldsymbol{W}\|^2}$$

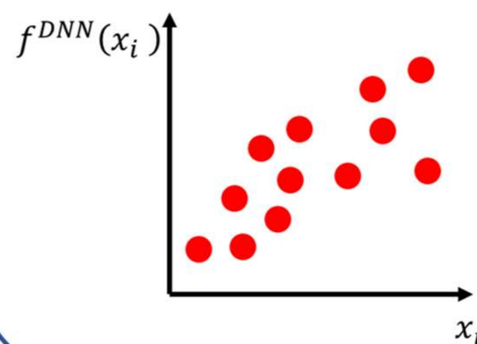


Neural Network Mapping

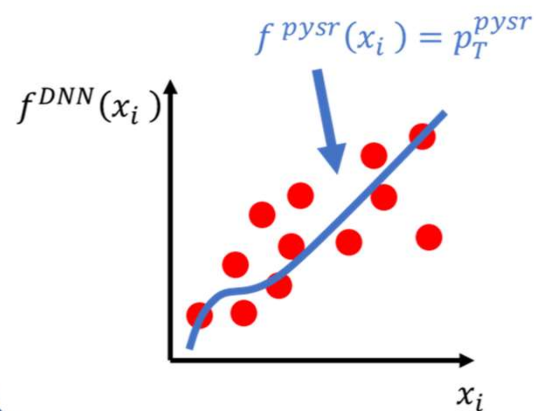
Use a deep neural network to map input jet features to the truth momentum.



Sample output of neural network across full range of input phase space.

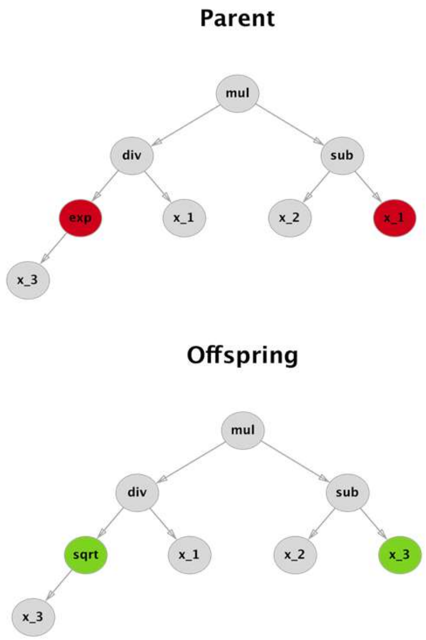


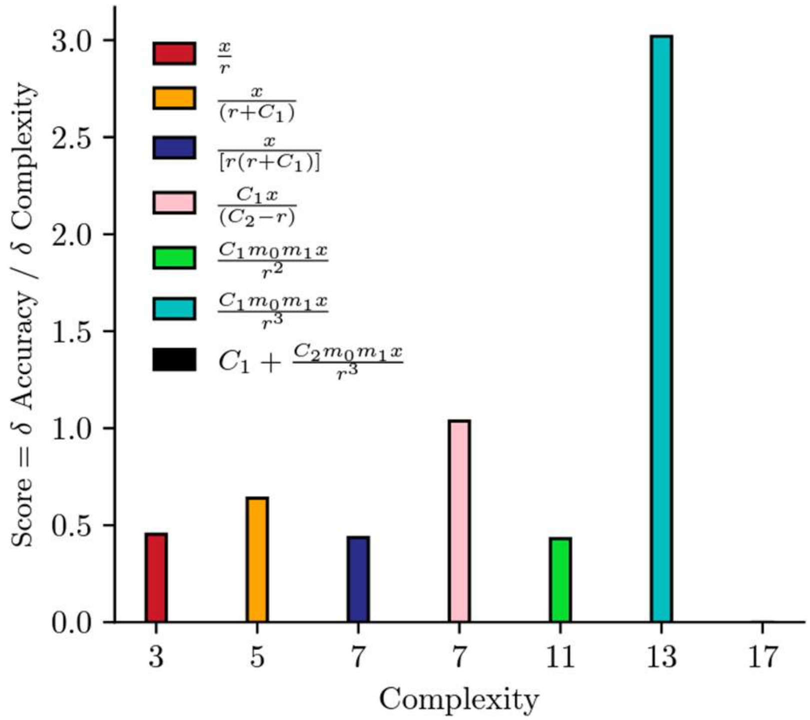
Fit jet features to neural network momentum prediction with symbolic regression.



PySR: Symbolic Regression

- PySR searches space of analytic expressions via multi-population evolutionary algorithm



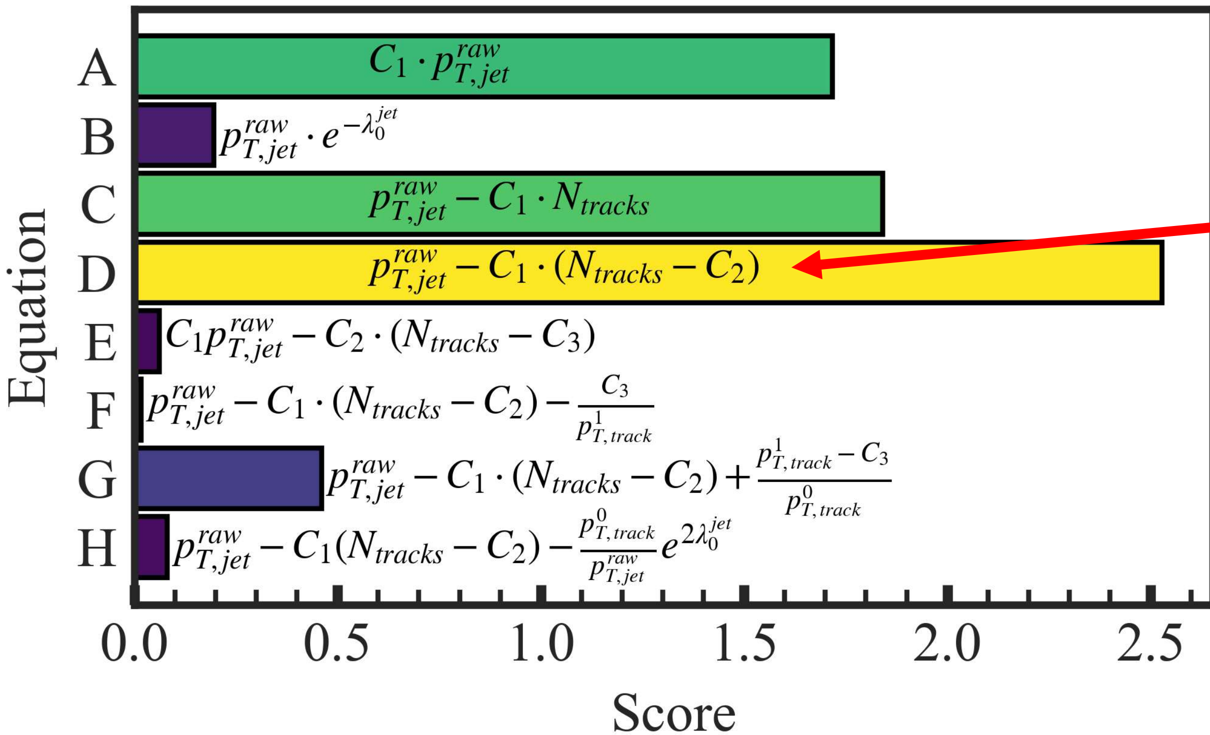


https://arxiv.org/abs/2305.01582

https://arxiv.org/abs/2202.02306

Mapping Results

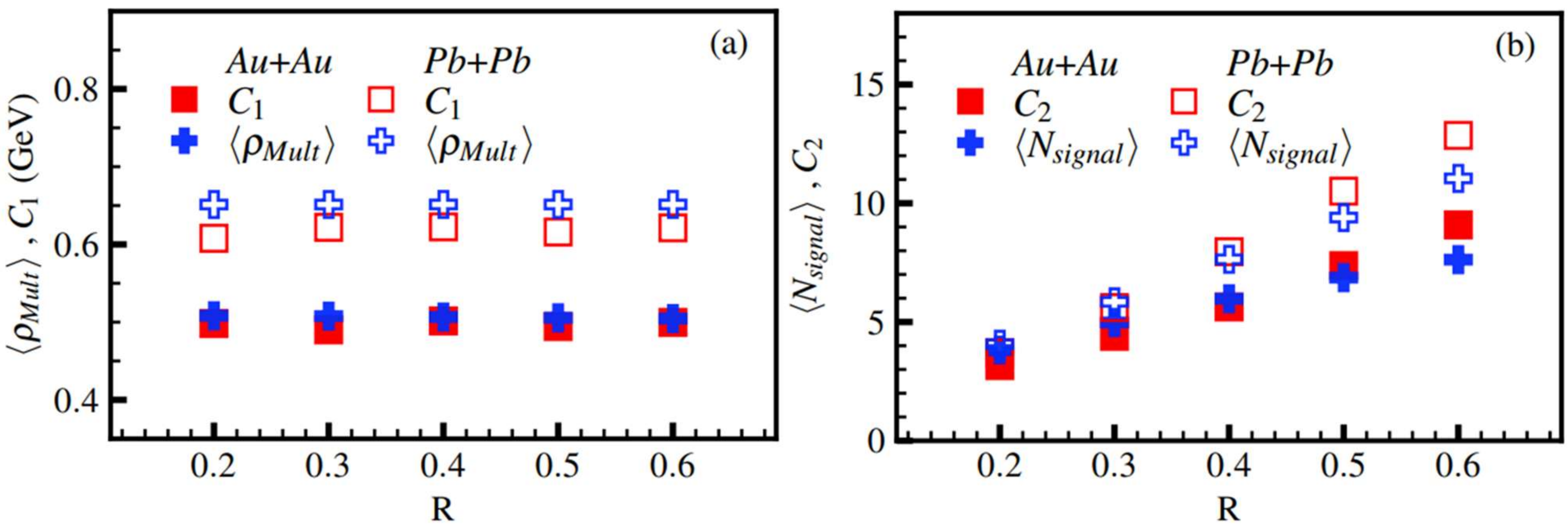
- Trained on neural network input features to predict neural network output.



Learned parameters

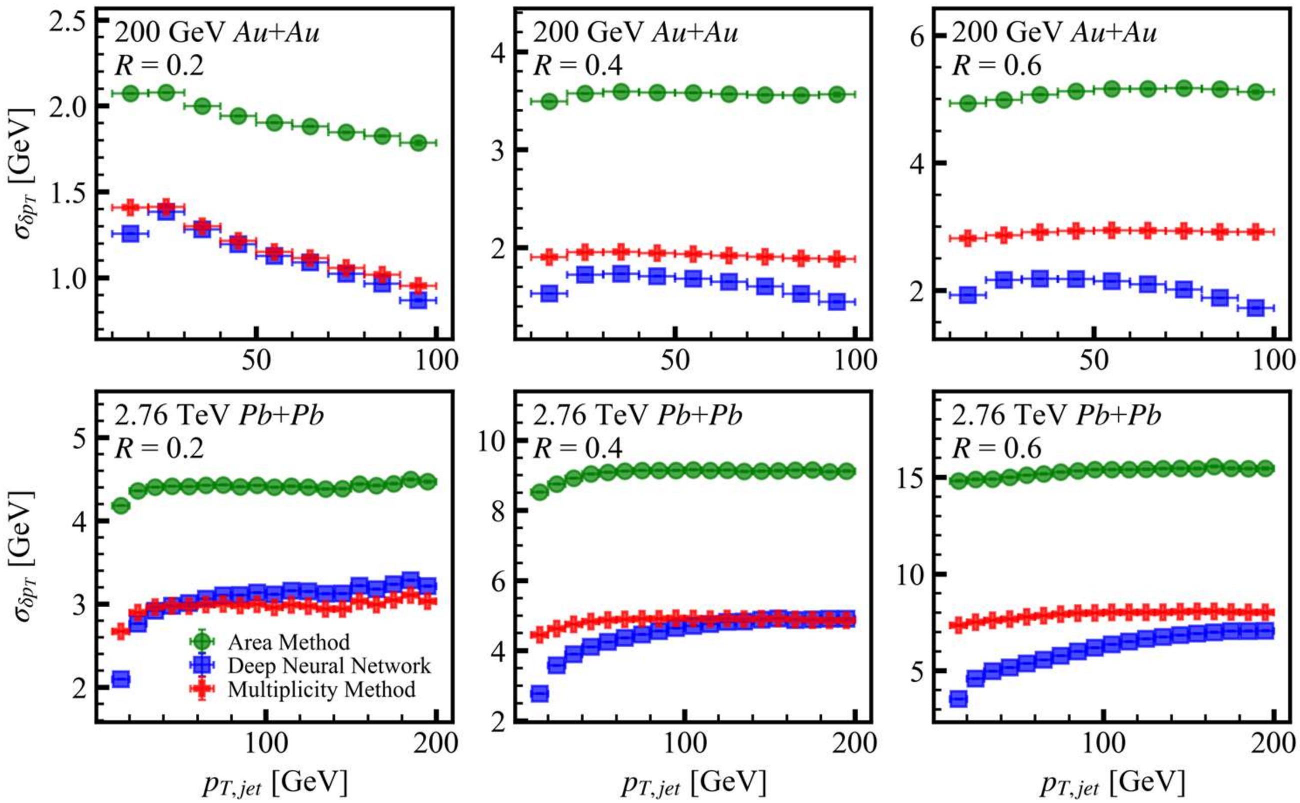
- Constants learned by PySR are approximately the terms used in multiplicity background subtraction method.

$$p_T^{mult} = p_T^{raw} - \rho_M \cdot (N - \langle N_{pythia} \rangle)$$
$$p_T^{pysr} = p_T^{raw} - C_1 \cdot (N - C_2)$$



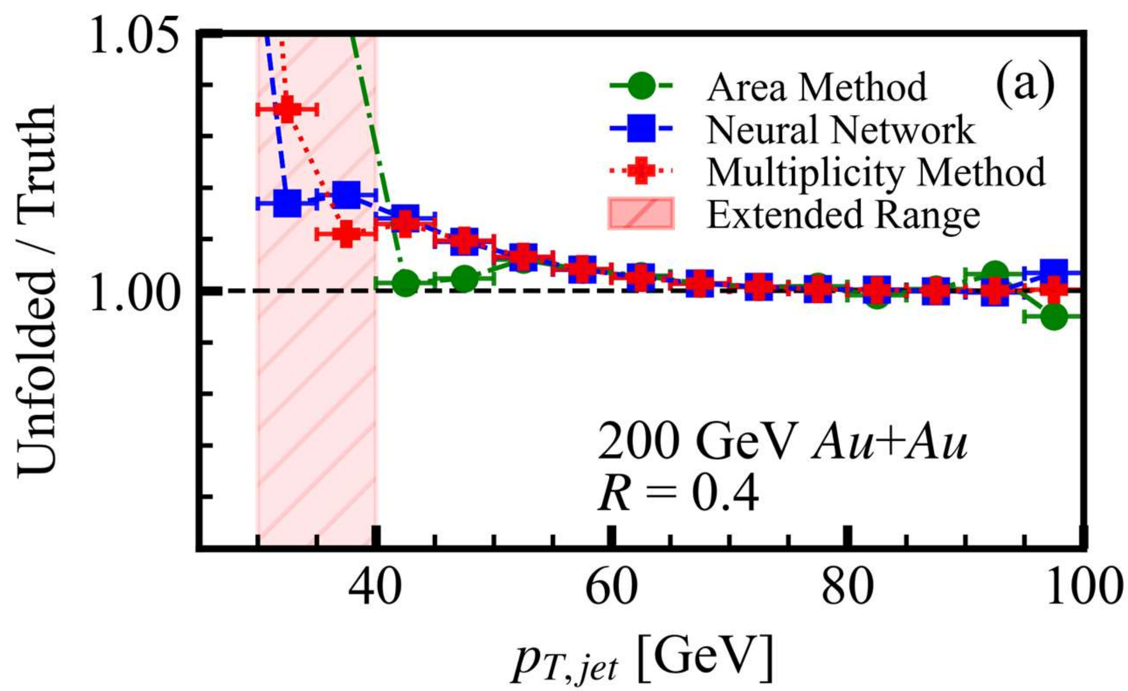
Learning from ML

- Neural network picks up on multiplicity relationship
- Multiplicity dependence is clear from background fluctuations



Increased Kinematic Reach

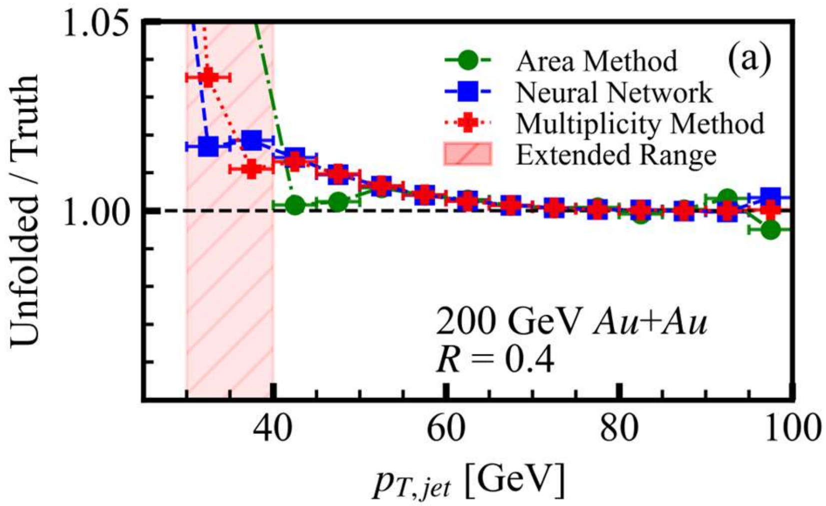
- Multiplicity and neural network methods are both stable to at least 10 GeV lower in momentum for all datasets.



Conclusions

- Multiplicity method offers almost all the improvement of neural network without the draw backs of traditional ML.
- Use of machine learning in science needs to be interpretable https://arxiv.org/abs/2303.08275
- PySR: https://github.com/MilesCranmer/PySR

$$p_T^{mult} = p_T^{raw} - \rho_M \cdot \left(N - \langle N_{pythia} \rangle\right)$$

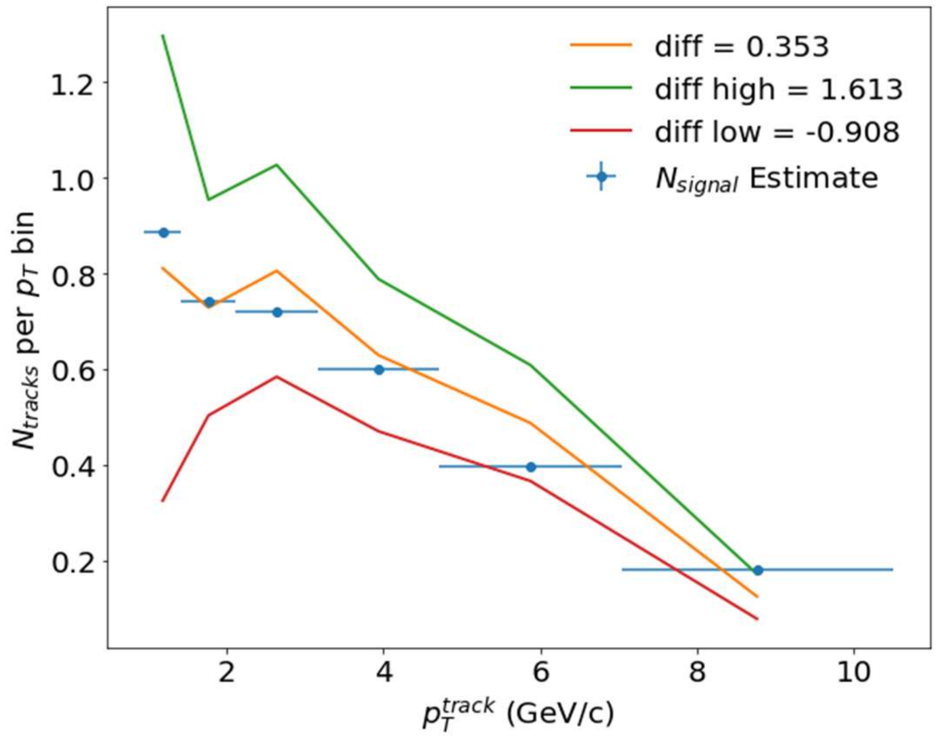


Special Thanks to Charles Hughes, Antonio Silva and Patrick Steffanic for their contributions to this study.

Backup

Uncertainty from $\langle N_{pthyia} \rangle$

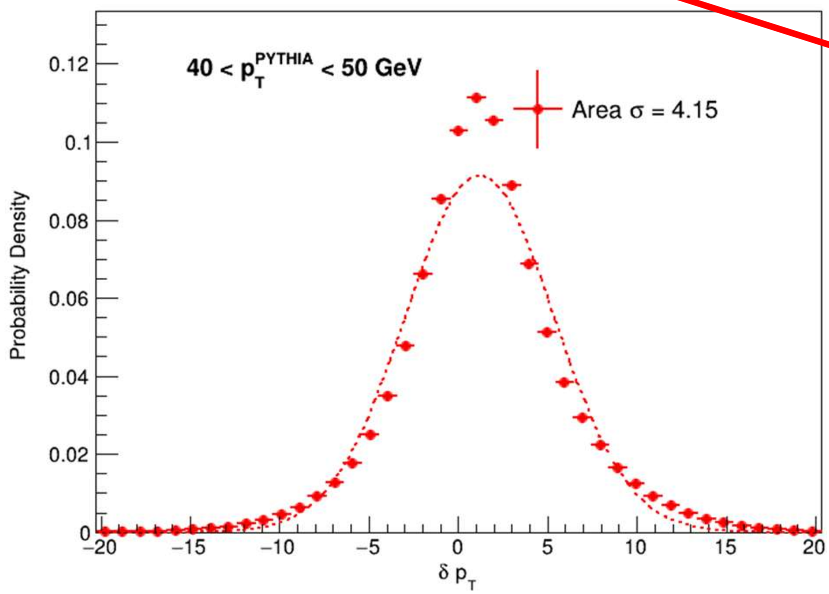
- Enhancement in jet multiplicity can be estimated via measured jet fragmentation functions
- At most 1 particle difference at low jet momentum



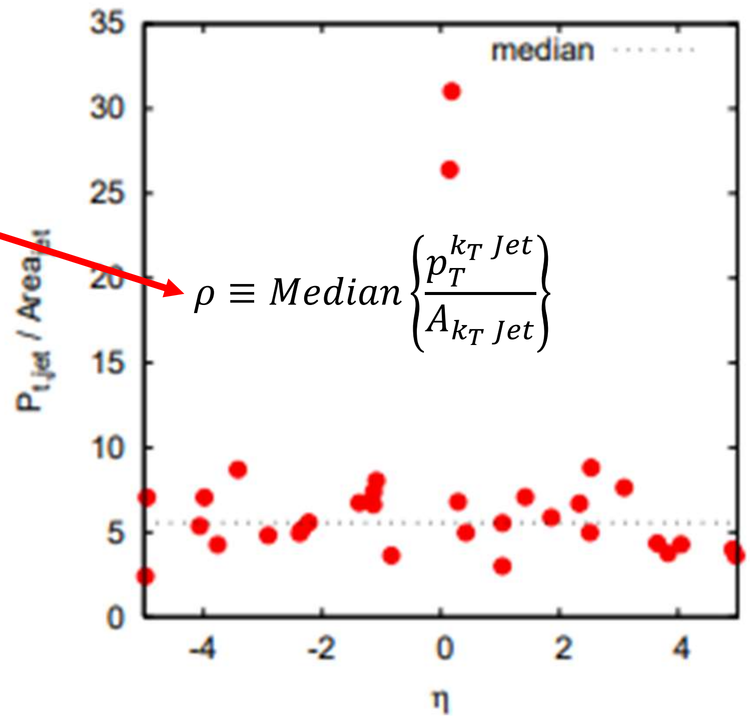
Area Based Method Recap

- Area based background subtraction:

$$p_T^{Corr.} = p_T^{raw.} - \rho A_{jet}$$



$$\rho \equiv Median\left\{\frac{p_T^{k_T\,Jet}}{A_{k_T\,Jet}}\right\}$$

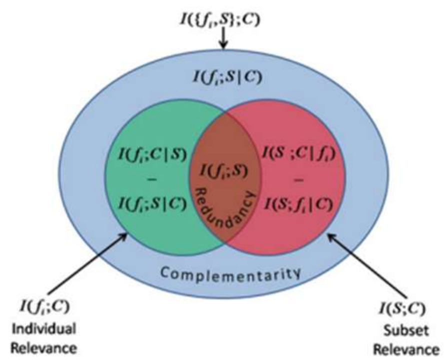


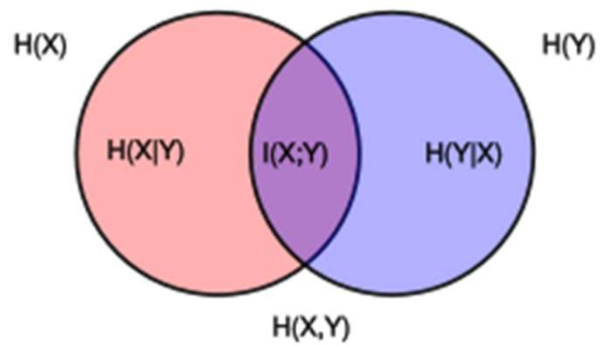
Phys.Lett.B 659 (2008) 119-126

Mutual Information

- ‘Distance’ between joint probability and individual probability between two random variables
- The mutual information provides a measure of the relevance an input feature has in predicting the target variable

$$I(X;Y) \equiv H(X) - H(X|Y)$$



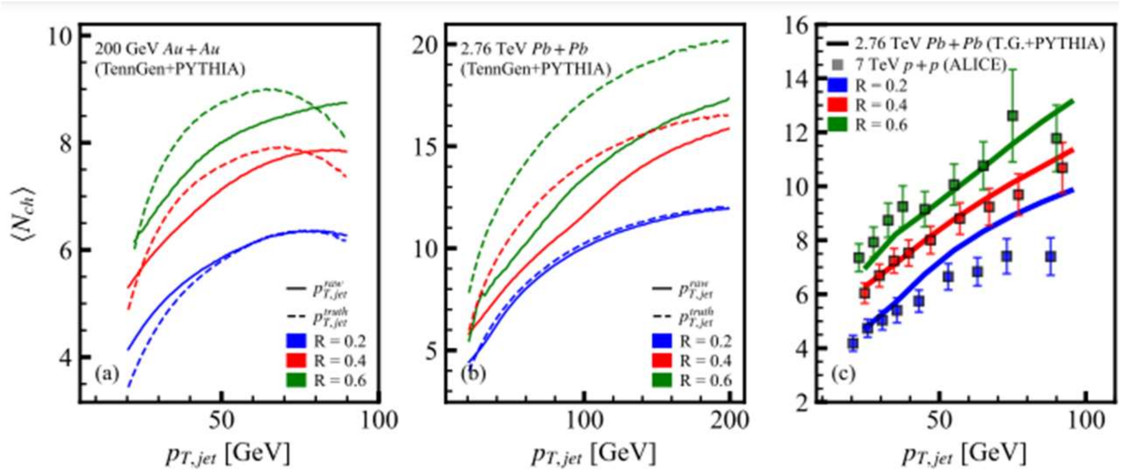


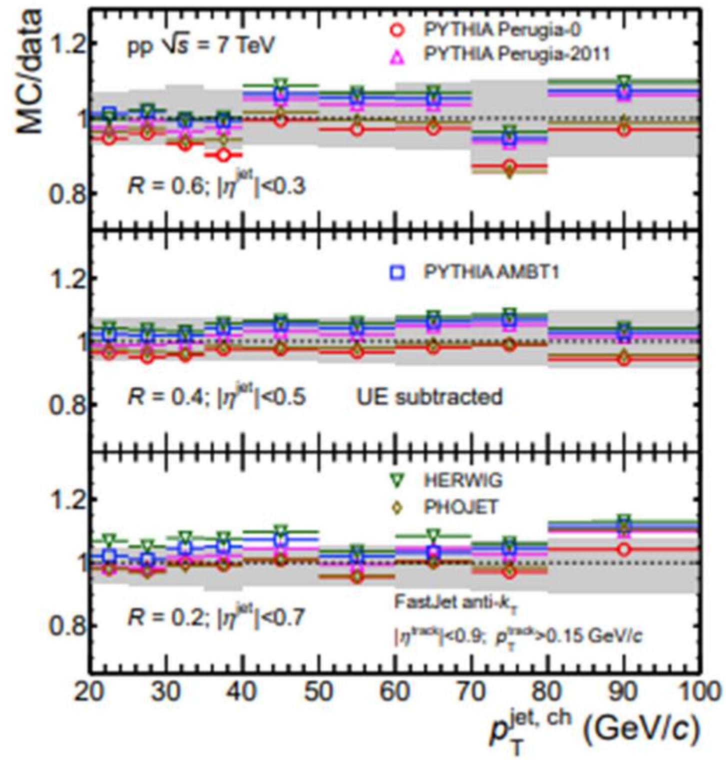
https://link.springer.com/article/10.1007/s00521-013-1368-0

Estimating $N_{tracks}^{B.G.}$

- Use average PYTHIA multiplicity $\left\langle N_{tracks}^{pp\ jet} \right\rangle_{p_{T,jet}}$ to approximate $N_{signal\ tracks}^{jet}$.

$$\boldsymbol{p_T^{Corr.} = p_{T,jet}^{raw} - \rho_M \cdot [N_{tracks}^{tot.} - \left\langle N_{tracks}^{pp,jet} \right\rangle_{p_{T,jet}}]}$$





Event Simulation

- PYTHIA8 (Signal):
  - 25 Million (1 million per p_T hard bin) p+p events at 200 GeV, Tune 14
- TennGen (Background): *C. Hughes et al Phys. Rev. C 106 (2022), 044915*
  - **Multiplicity**: Sampled from corrected N_{ch} distribution *STAR Phys.Rev.C 79 (2009), 034909*
  - $\boldsymbol{p_T}$: Identified particle p_T spectrum fit with Boltzmann-Gibbs Blast wave *PHENIX Phys.Rev.C 88 (2013) 2, 024906*
  - $\boldsymbol{\phi}$: Identified particle flow harmonics (v_2, v_3, v_4) *PHENIX Phys.Rev.C 93 (2016) 5, 051902*
  - $\boldsymbol{\eta}$: Uniform distribution |η| < 1.1
- Merge PYTHIA8 charged particles with TennGen Au+Au 200 GeV background
- Find anti-k_T jets
  - Only save jets with $p_{T,jet}^{Truth}$ > 10.0 GeV
  - |η| < 1.1 – R, 0 < $\phi \leq 2\pi$
  - ~30 Million jets per dataset
- Take matched jet $p_{T,jet}^{PYTHIA}$ to be truth value
  - Train-Test split: 50/50%