

## Identification of Quenched Jets

In this poster I show how machine learning techniques can help to evaluate how quenched a jet is. To achieve this, a classifier was trained using the Long Short-term Memory (LSTM) model. Simulations were done with Monte Carlo event generators such as Pythia8 [1] and JEWEL [2] which simulates jets in vacuum and in medium respectively.

## Feature Selection

Jets are recursively de-clustered into two subsets with Cambridge-Aachen (C/A) algorithm. The primary path, marked in red in Fig.2, follows the harder branch at each branching. Each node on the primary path contains a set of substructure variables which are used to train a classifier based on LSTM model.

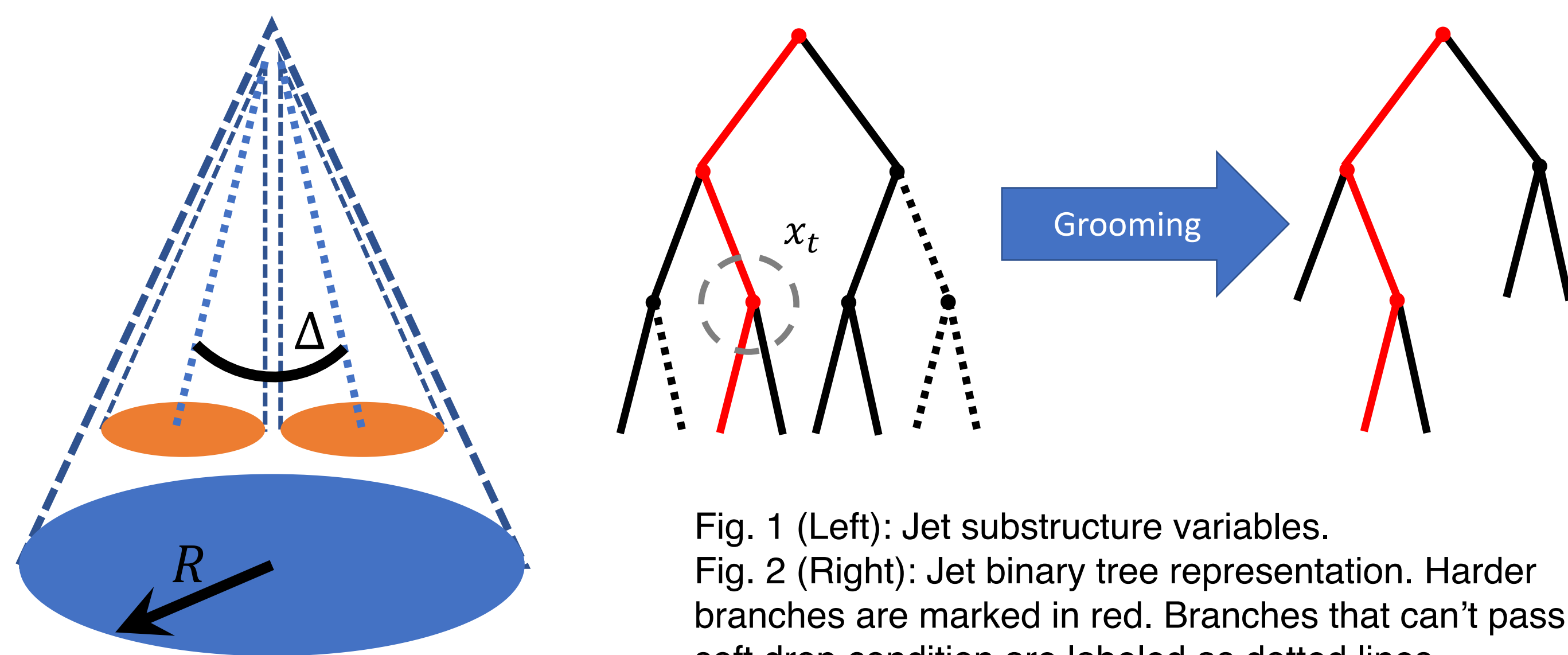


Fig. 1 (Left): Jet substructure variables.  
Fig. 2 (Right): Jet binary tree representation. Harder branches are marked in red. Branches that can't pass soft drop condition are labeled as dotted lines.

Input vectors to the LSTM network  $[x_0, x_1, \dots, x_N]$  are taken from jet binary tree which come from recursive de-clustering. At step  $t$ , the feature vector  $x_t$  is a combination of substructure variables.

$$z = \frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}}, \quad \Delta = \sqrt{(\varphi_1 - \varphi_2)^2 + (\eta_1 - \eta_2)^2}$$

$$k_{\perp} = p_{T,2} * \Delta, \quad m = \text{inv\_mass}(j_1, j_2)$$

## Simulation & Training

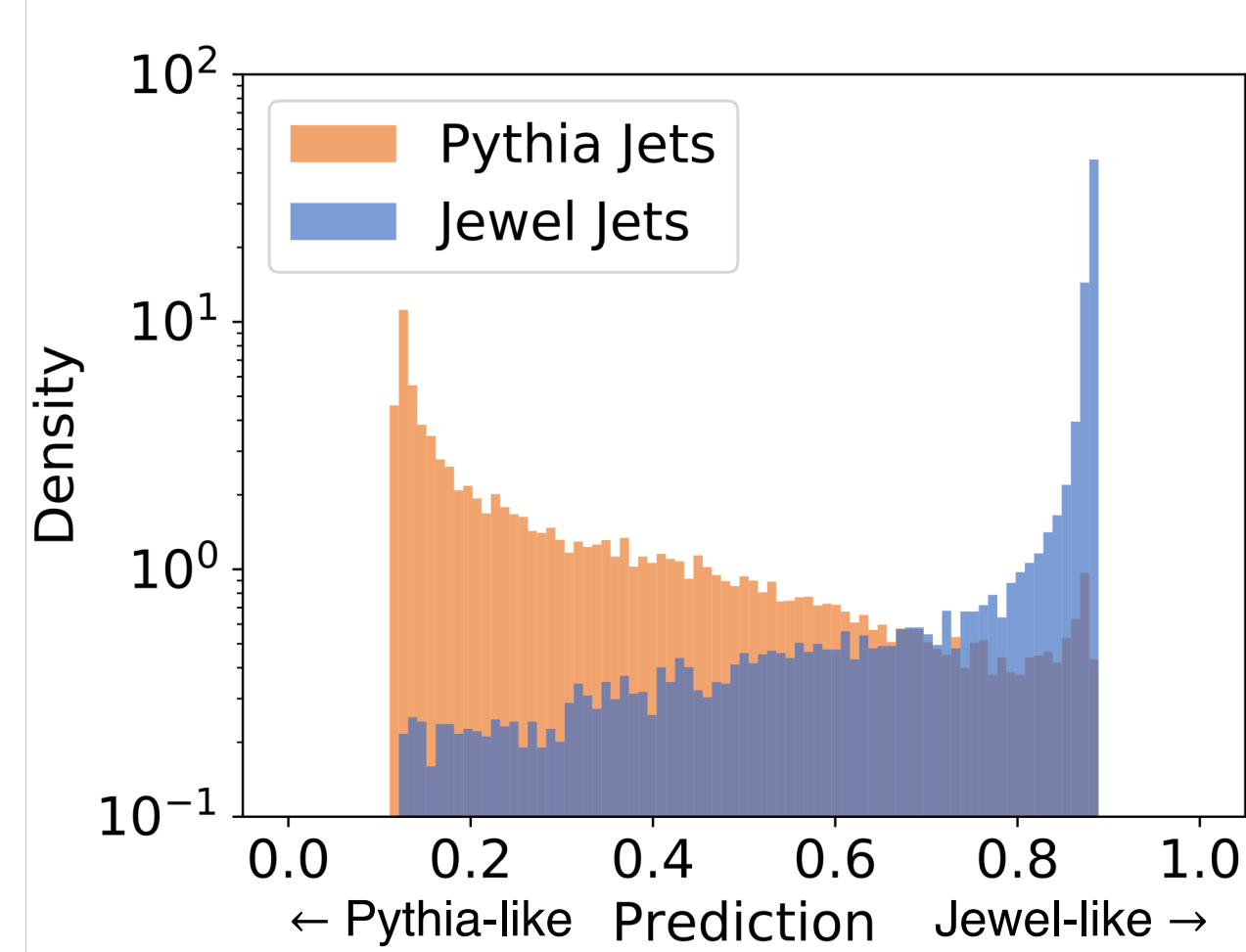


Fig. 3: Distributions of predicted values that come from a classifier trained with 4 features  $[lnz, ln\Delta, lnk_{\perp}, lnm]$ .

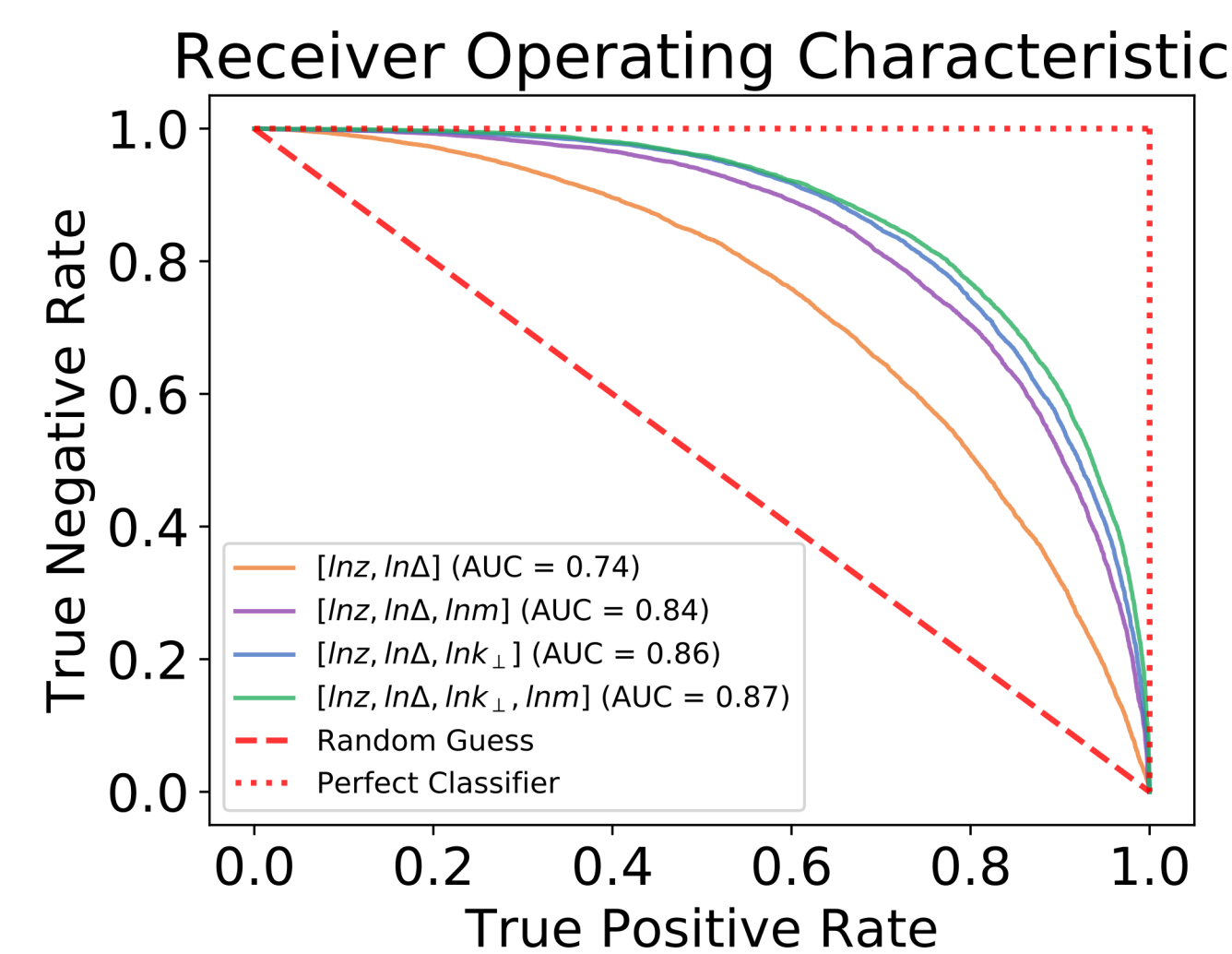


Fig. 4: The distinguishing abilities of different classifiers compared with the help of ROC (Receiver Operating Characteristic) curves.

MC	$\hat{p}_t$ (GeV)	Jet Radius	Jet Clustering	Jet De-clustering	Label
Pythia8	120	0.4	Anti-kt	C/A	0
Jewel <sup>2</sup>	120	0.4	Anti-kt	C/A	1

Training	Learning Rate	LR decay (rate)	No. of Epochs	Loss Func	Batch Size
	0.005	Exponential (0.8)	10	BCE	100

Others	ML Framework	No. of Jets	Hyper-tuning	Training Time <sup>2</sup>	GPU
	Pytorch [3]	15k + 12k	No	< 30 Sec/Epoch	No

Classifier	Input Features	Input Size	Hidden Size	No. Hidden Layers	No. of Parms <sup>3</sup>
1	$[lnz, ln\Delta]$	2	2	5	240
2	$[lnz, ln\Delta, lnm]$	3	2	5	248
3	$[lnz, ln\Delta, lnk_{\perp}]$	3	2	5	248
4	$[lnz, ln\Delta, lnk_{\perp}, m]$	4	2	5	256

Table: Simulation and training details.

## LSTM Model

Long short-term memory is an artificial recurrent neural network architecture and is capable of processing sequential data. This design makes it well-suited for making predictions on jets considering that the way of calculating substructure variables implies sequential information about how branching occurs. Recent study has shown successful implementation of a LSTM-based jet groomer [4].

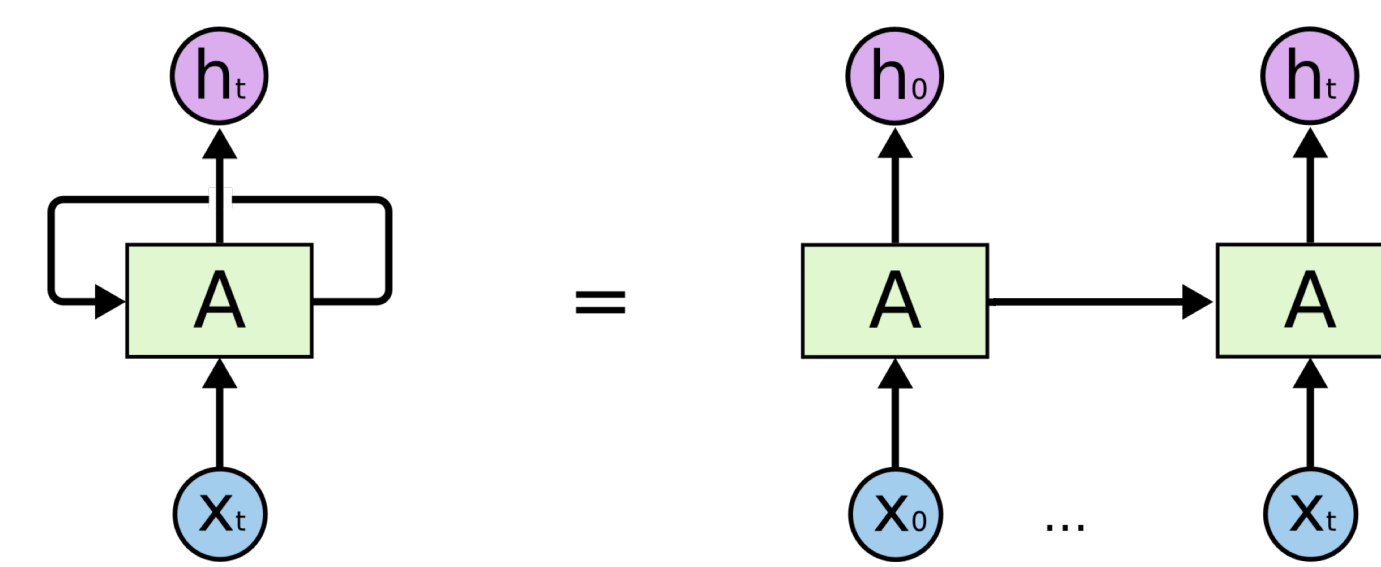


Fig. 5: Recursive neural network (RNN) [5].

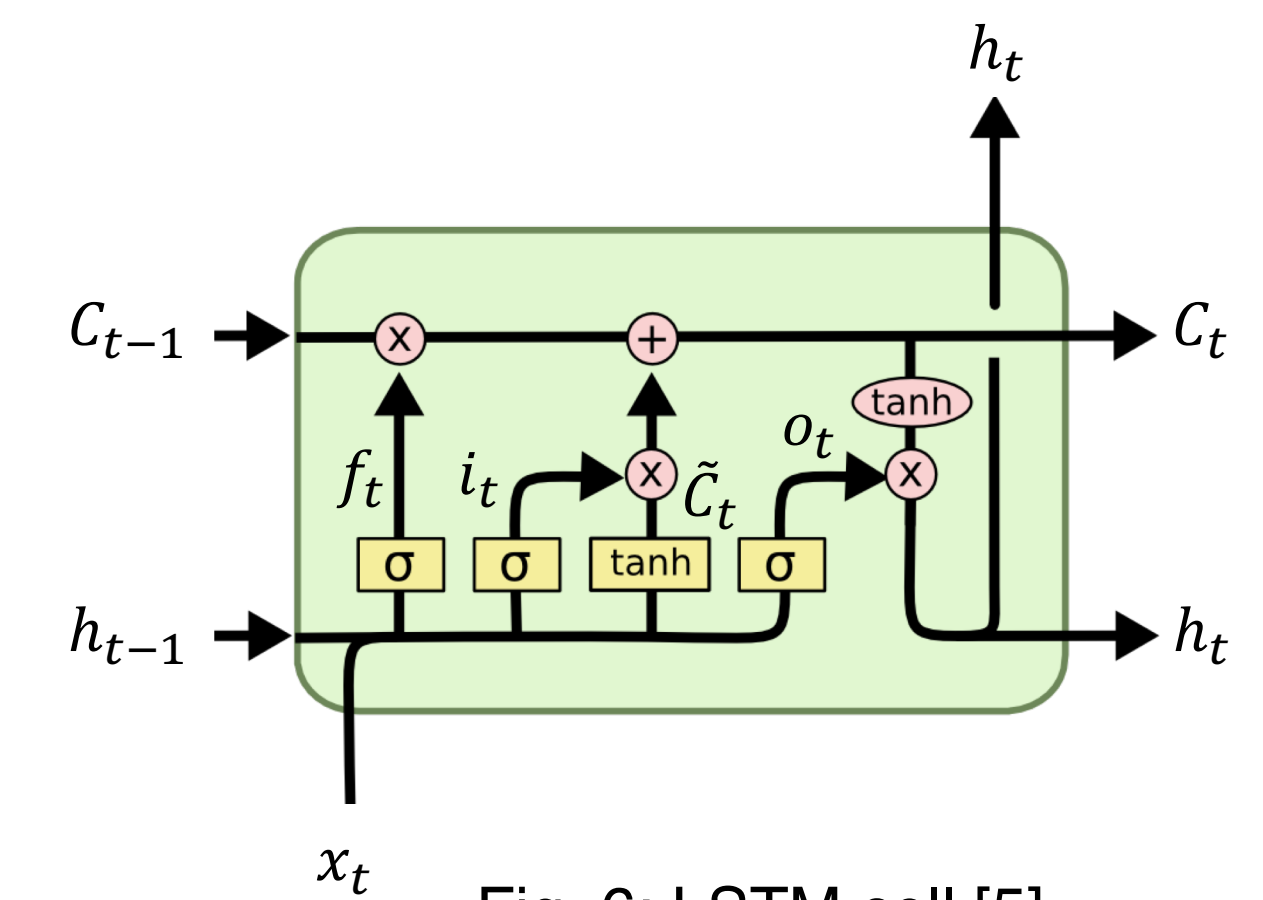


Fig. 6: LSTM cell [5].

Calculations that happen in a LSTM cell can be expressed as follows:

- Forget gate:  $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
  - Input gate:  $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$   
 $\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$
  - Update gate:  $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$
  - Output gate:  $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$   
 $h_t = o_t * \tanh(C_t)$
- ( $W, b$ : learnable weights and biases.)

## Substructure Variables

Jets with different predicted values populate the Lund radiation plane in different ways (Fig. 7). Substructure variables such as  $z$ ,  $\Delta$  and jet masses are compared (Fig. 8). Classifier no. 4 is used in making predictions.

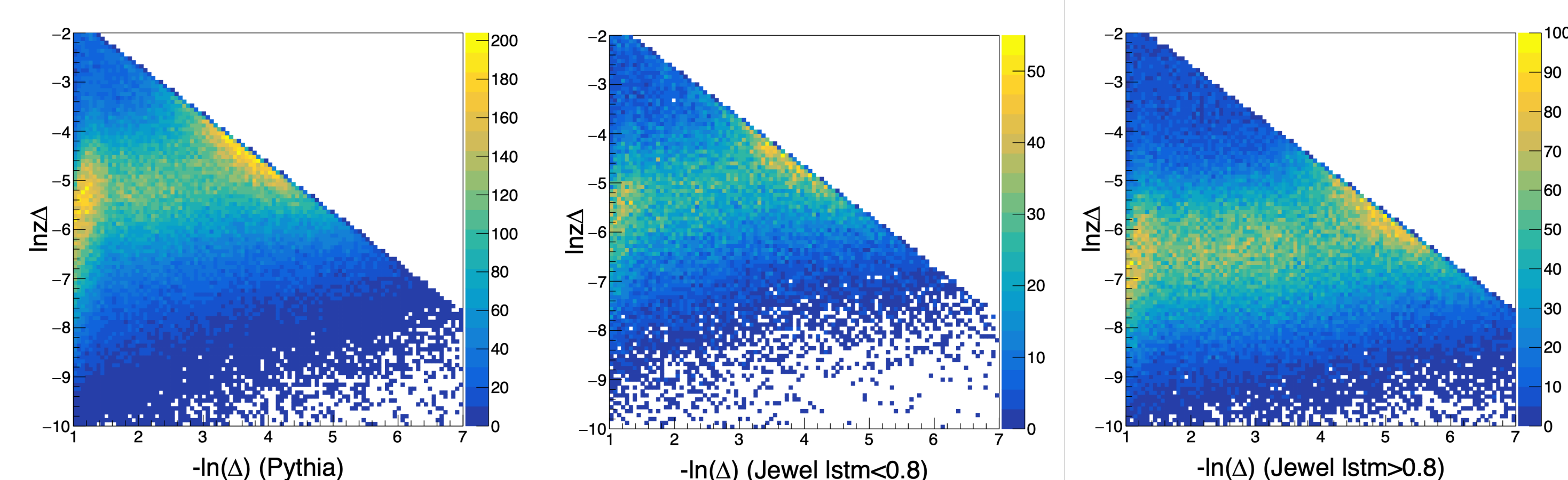


Fig. 7: Lund radiation planes.

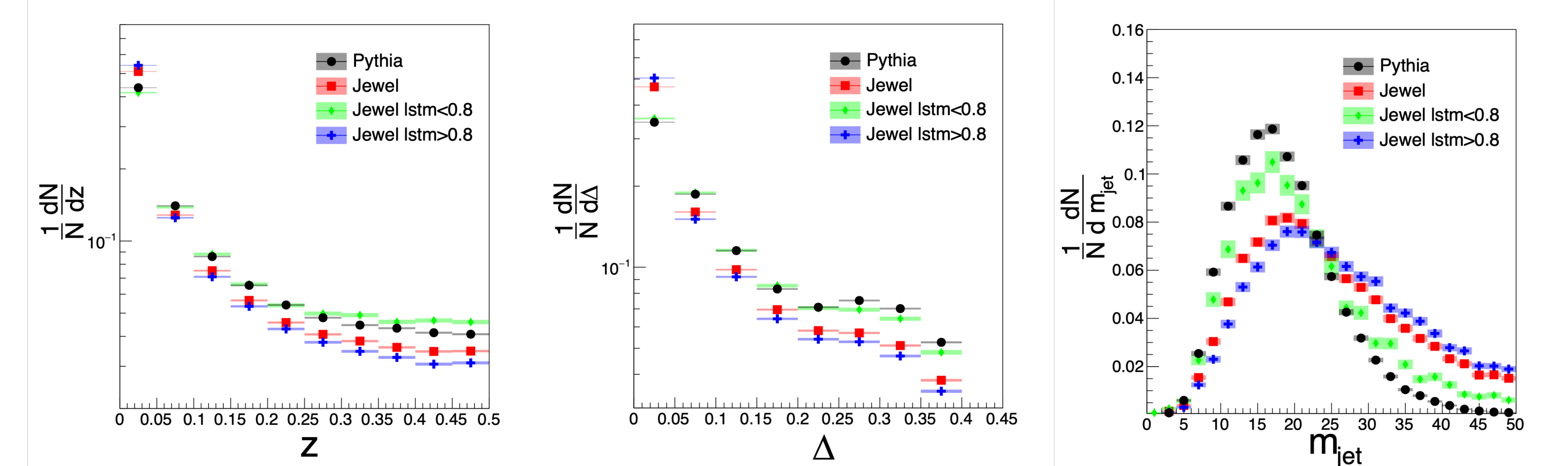


Fig. 8: Distributions of  $z$ ,  $\Delta$  and the jet masses.

Jewel jets with smaller predicted values (Istm<0.8) show similarities to Pythia jets which can be observed from distributions of substructure variables (see Fig. 8).

## Conclusion

This study has shown that the LSTM network can take jet substructure variables as input and develop its identification ability thus making it a practical approach to identify quenched jets.

It also shows that different feature selections may result in classifiers with different distinguishing abilities, as shown in Fig. 4.