Sphalerons vs black holes

Results and progress update

Aurora Grefsrud

Høgskulen på Vestlandet

Universitetet i Bergen

Uniwersytet Warszawski

People involved

Kazuki Sakurai, UW, theory

Aurora Grefsrud, HVL, machine learning method and results

Anna Lipniacka, UiB

Rafal Maselek, UW

Fotis Koutroulis, UW

Andreas Papaefstathiou KENNESAW STATE UNIVERSITY

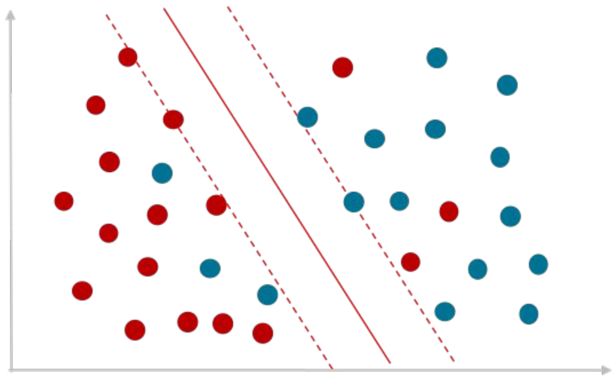
Trygve Buanes, Therese Sjursen, Steffen Mæland, Igor Slazyk (HVL)

Goal: Investigate new machine learning methods to separate **sphaleron** and **black hole** events

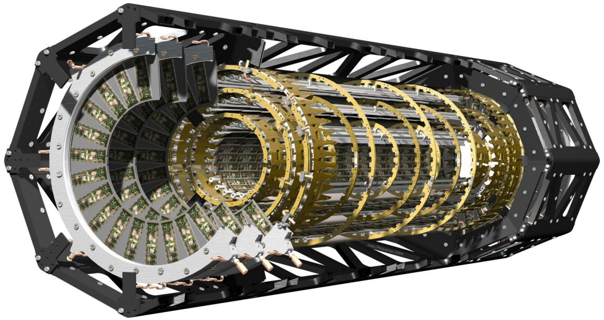
Datasets produced using BlackMax/Herwig7/Delphes:

1. Sphalerons, 9TeV sphaleron energy
2. Black holes, 10 TeV minimum mass
    a. 2, 4, 6 extra dimensions

Separate **training** and **testing** data sets.



End-to-end classification using computer vision inspired techniques



Images

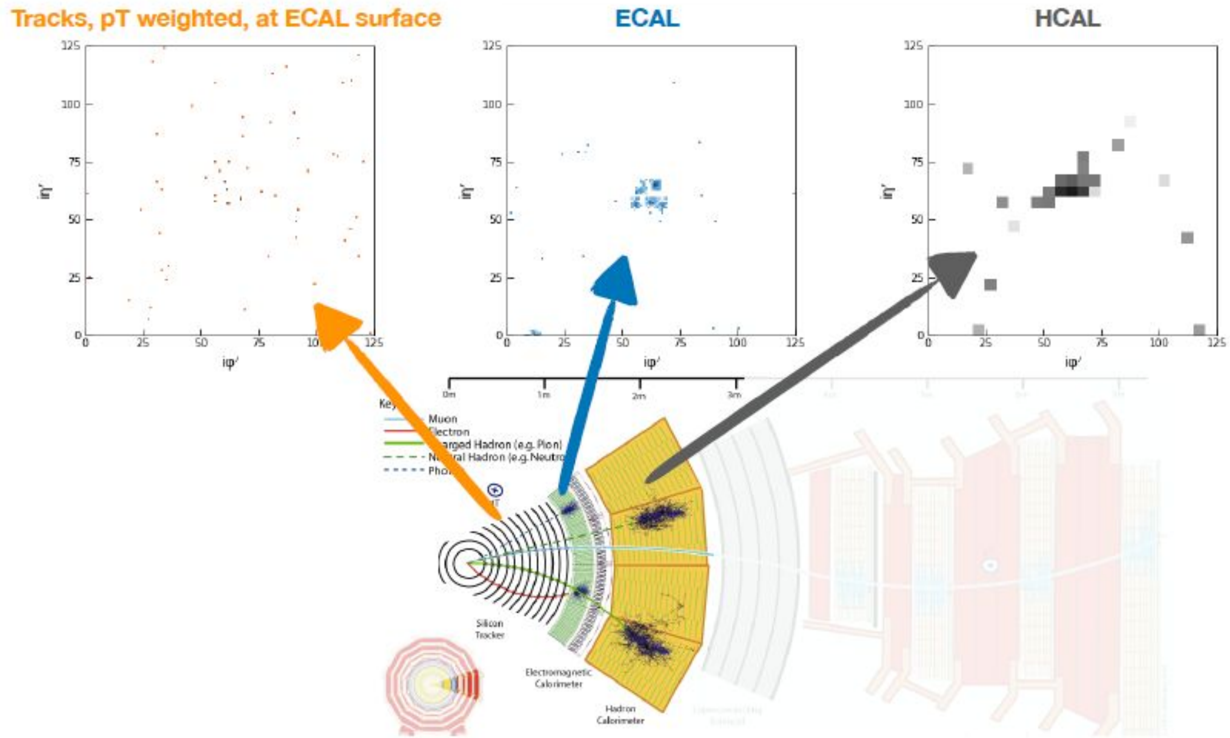
Resolution: 50x50

(R, G, B) = (EMCal, HCal, tracks)

Intensity ∝ Energy deposit

Process based on this paper:

https://arxiv.org/abs/1807.11916



Convolutional neural network

ResNet18.

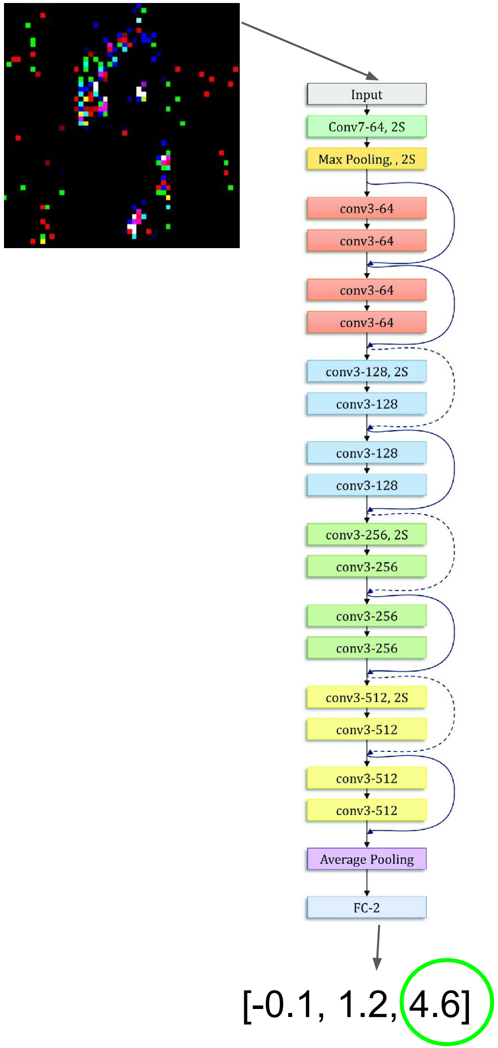
Added **circular convolution** for panoramic images.

Takes in 50x50 = **2500 features** (the image).

Outputs a tensor of values representing the classes, and the maximum value gives the **predicted class**.

Metric: In the end we can calculate the % correctly classified in each class, the **accuracy**.

Data augmentation: Random flips across η = 0 and random rotations in φ-direction, completely necessary for training



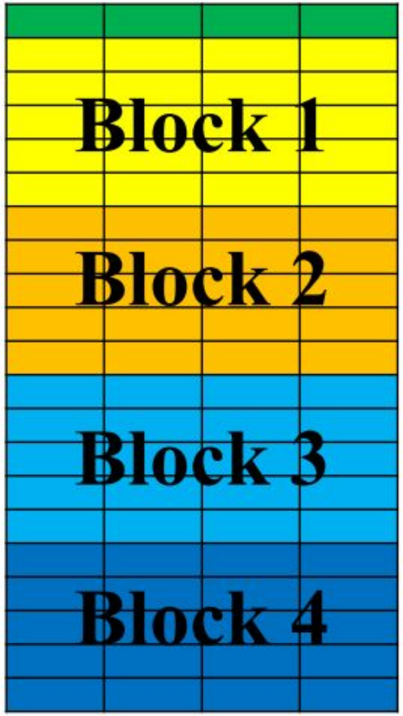
XGBoost

XGBoost – external open-source library (framework) based on the Gradient Boosting. In comparison to the regular Gradient Boosting algorithm, the XGBoost increases speed and performance significantly.

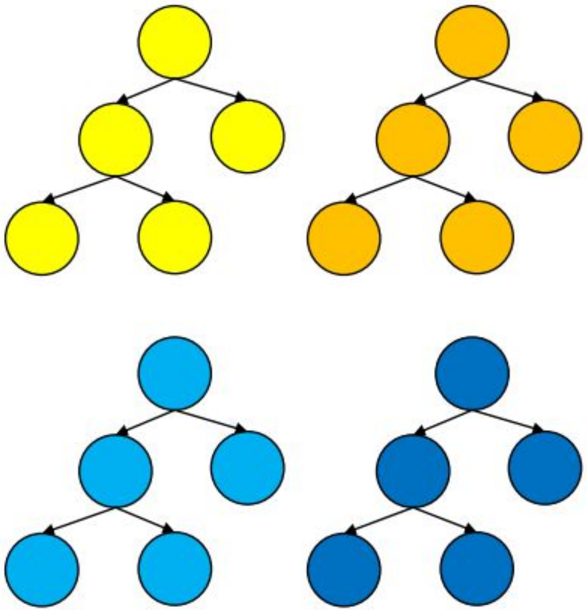
dmlc XGBoost

Major improvements:

- Parallelized tree building
- Tree pruning
- Efficient handling of missing data
- Regularization to prevent overfitting
- In-built cross-validation capability
- Hardware optimization



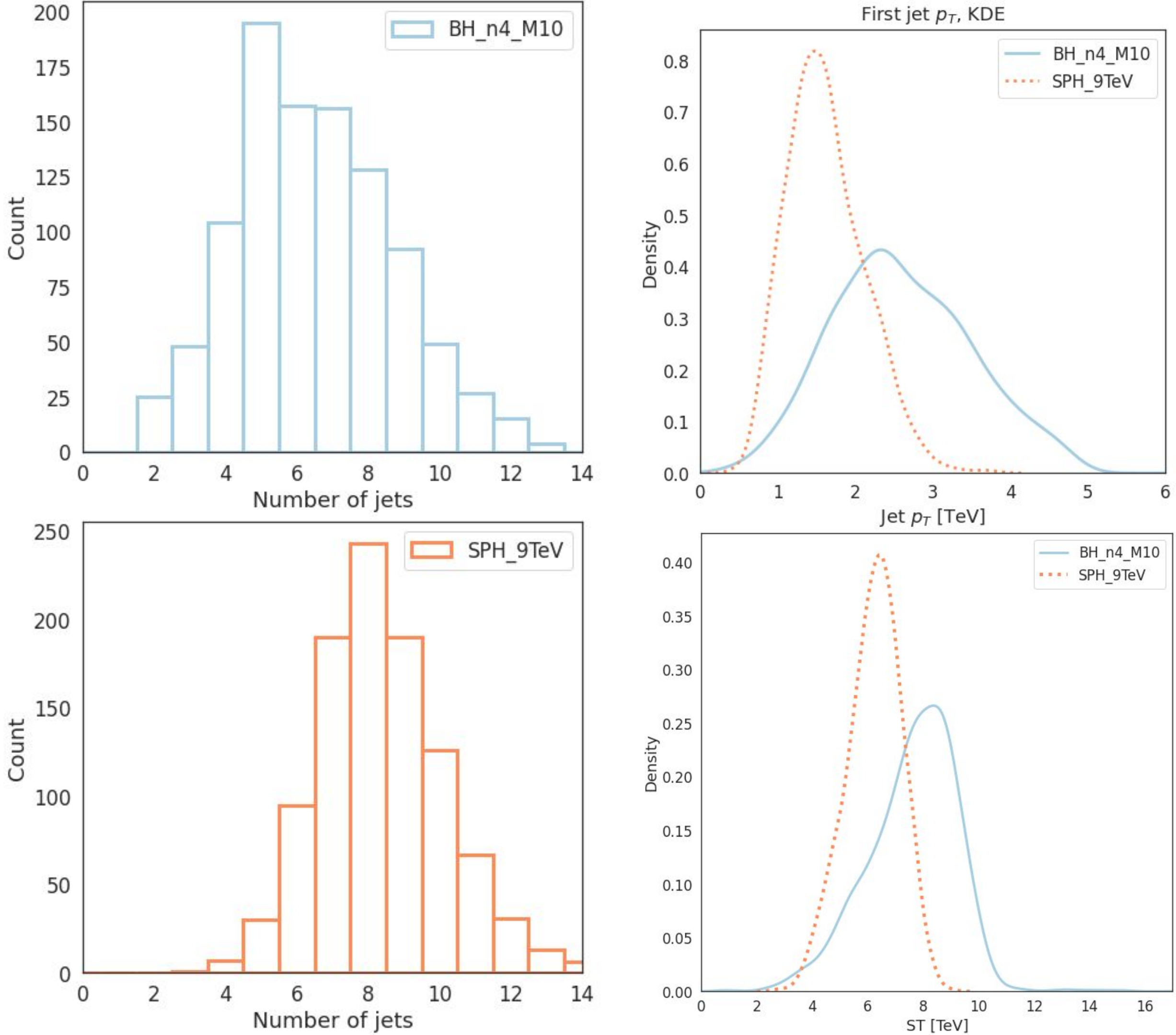
Parallelism



Binary classification

SPH_9TeV

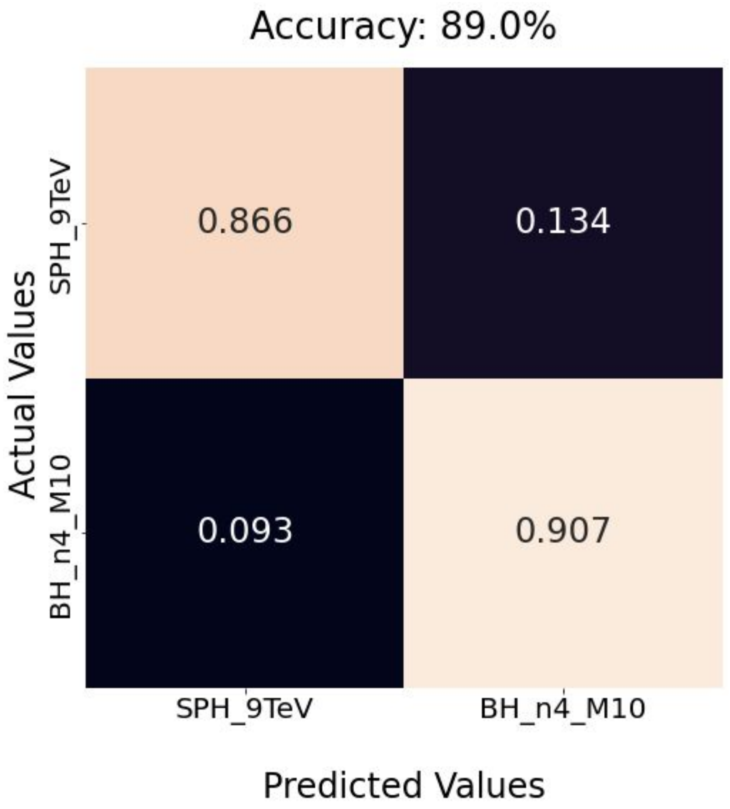
BH_n4_M10



ResNet18 results

After training 5 models for 30 epochs:

- Mean accuracy: **90.7%**
- Standard deviation: 0.7%



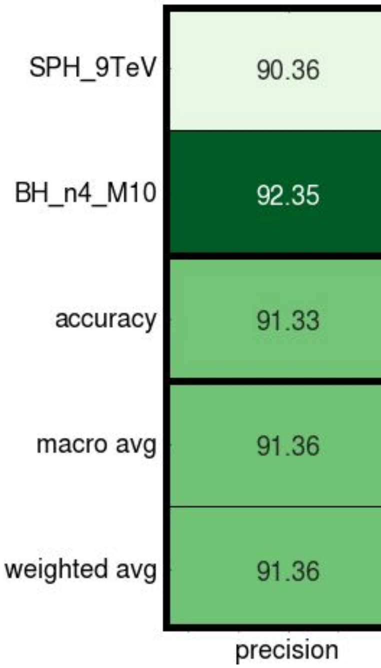
XGBoost results

Input features:

- Five most energetic hits in
  - ECal
  - HCal
  - Tracks

After running the experiment 5 times:

- Mean accuracy: **86.1%**
- Standard deviation: 0.1%



Input features:

- First eight jets
- First two leptons
- MET

After running the experiment 5 times:

- Mean accuracy: **91.4%**
- Standard deviation: 0.2%

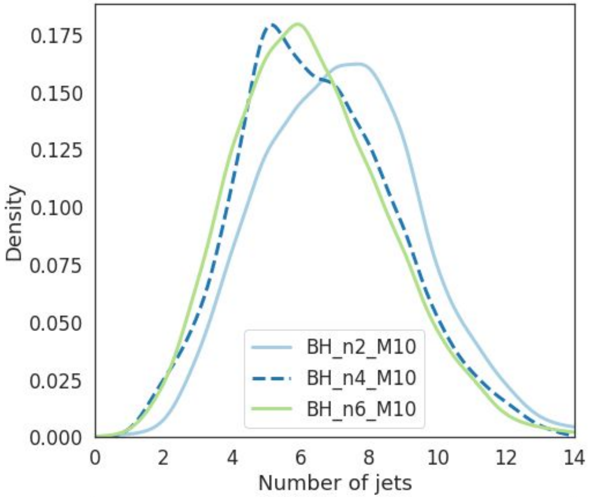
Multi classification

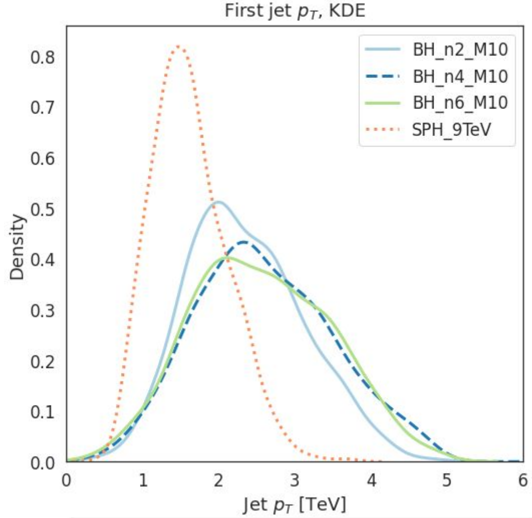
SPH_9TeV

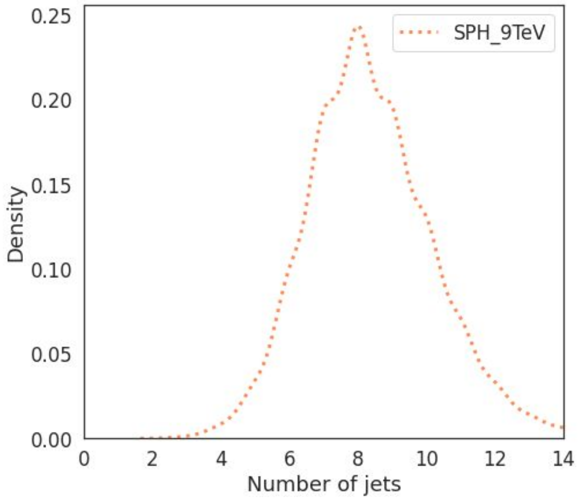
BH_n2_M10

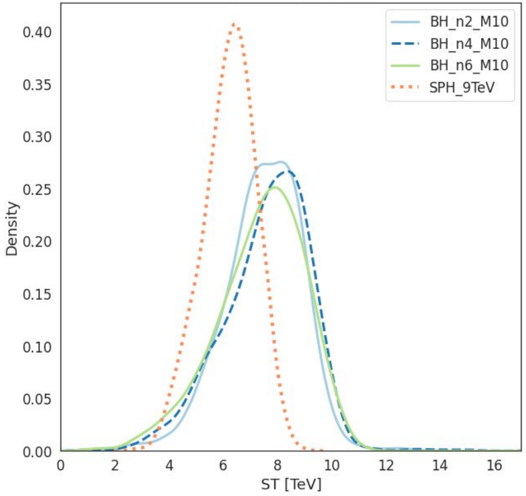
BH_n4_M10

BH_n6_M10







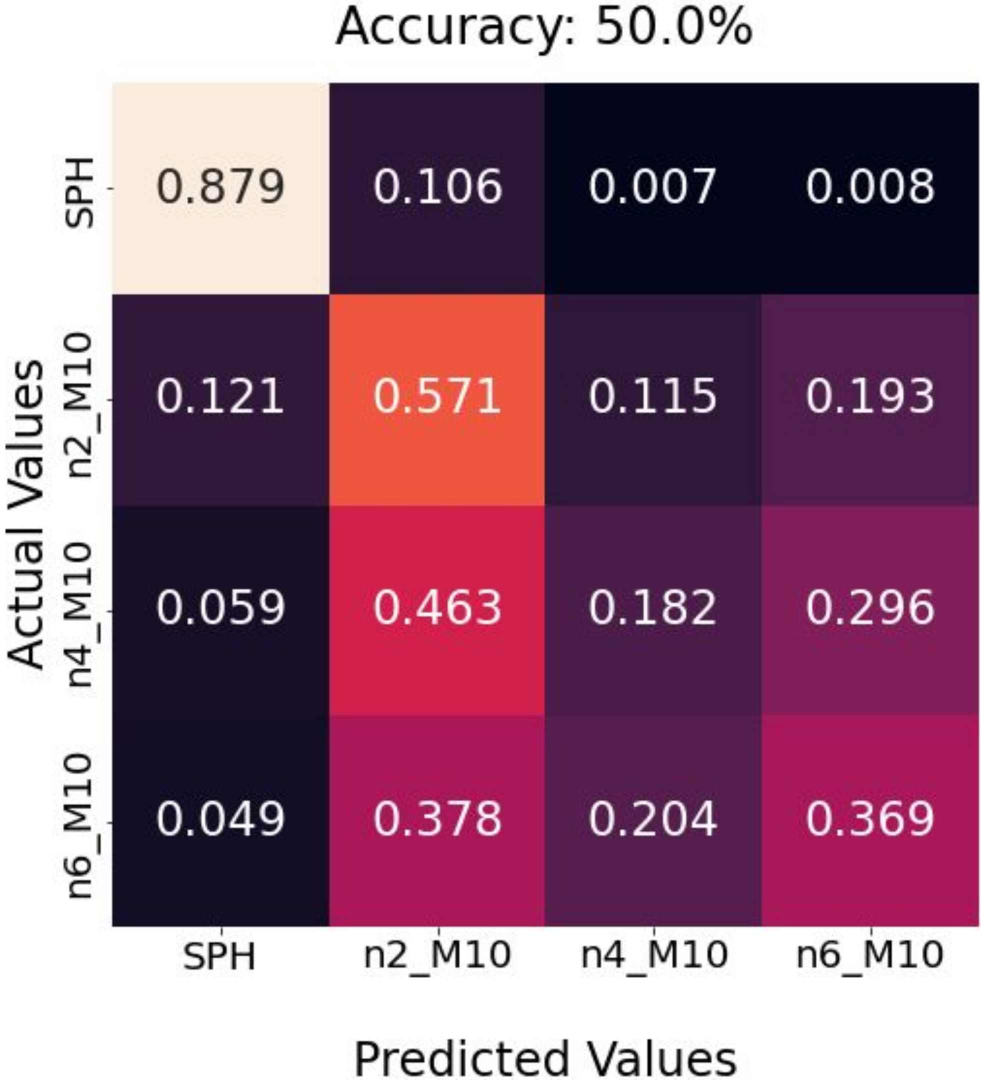


Resnet18 results

After training 5 models for 30 epochs:

- Mean accuracy: **49.9%**
- Standard deviation: 0.6%

Struggling to separate the three black hole types.



XGBoost results

Input features:

- Five most energetic hits in
  - ECal
  - HCal
  - Tracks

After running the experiment 5 times:

- Mean accuracy: **46.5%**
- Standard deviation: 0.2%

	precision
SPH_9TeV	75.18
BH_n2_M10	40.56
BH_n4_M10	37.17
BH_n6_M10	41.68
accuracy	50.61
macro avg	48.65
weighted avg	48.65

Input features:

- First eight jets
- First two leptons
- MET

After running the experiment 5 times:

- Mean accuracy: **50.9%**
- Standard deviation: 0.2%

Summary results

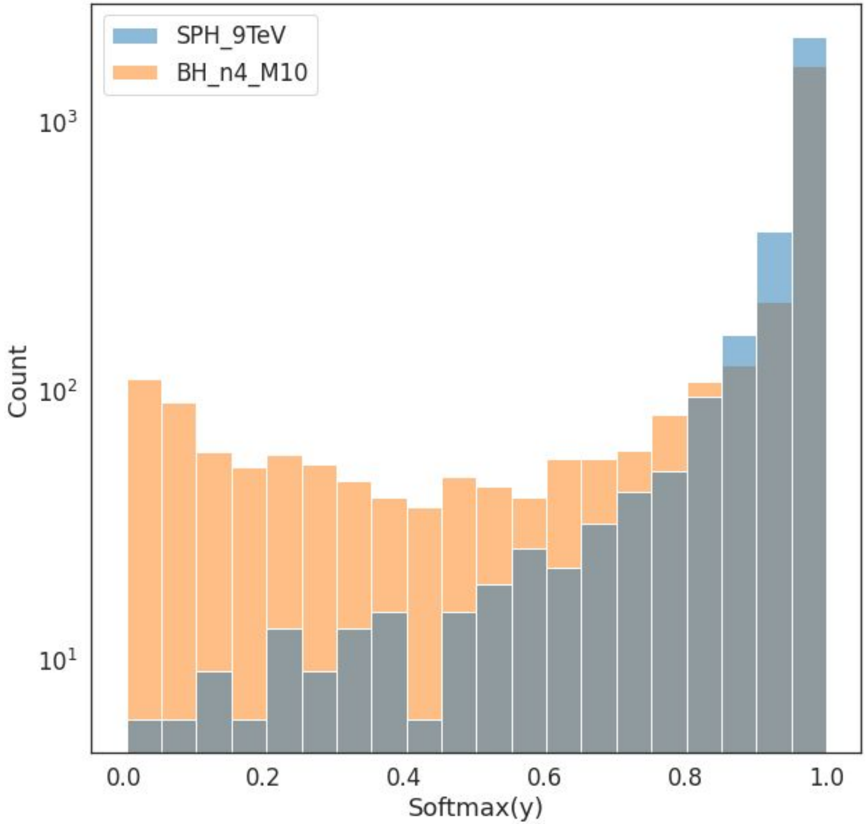
	Resnet18 low level	XGBoost low level	XGBoost high level
Binary classification	0.907 +- 0.007	0.861 +- 0.001	0.914 +- 0.002
Multi classification	0.499 +- 0.006	0.465 +- 0.002	0.509 +- 0.002

Can we trust the CNN?

- Softmax(y) transforms the output vector $y = [y_1, y_2, ...]$ to a new vector with values such that $sum(y_i) = 1$ and $0 < y_i < 1$.

- Softmax(y) value interpretation:
  - Close to 1 - very confident and right
  - Close to 0.5 - very uncertain
  - Close to 0 - very confident and wrong

- Majority are confidently classified right
- BH are much more likely than SPH to be confidently classified wrong



Discussion points

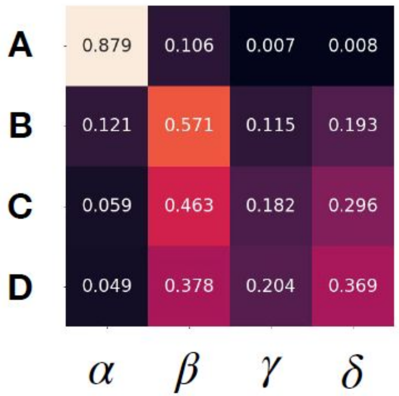
- How do we understand the predictions from the network.
    - Not probabilistic values.
        - Relation between softmax and how certain a prediction is?
    - Would we expect the same accuracy for ‘real life scenario’?
        - Proposed statistical method using the softmax function to make “probabilities” from the output vector
            - Rafal and Kazuki have the details
        - We can make some experiments to simulate the effect of having just a few events available. How many do we need to make conclusions?

Paper progress

- Paper draft has been started
- Results are in
- Just write it :)

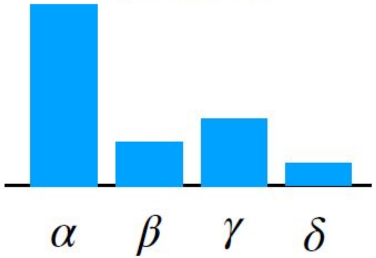
For a given size of signal events, observed at the LHC with a given integrated lumi, with what accuracy can we say Model-X is realised in nature?

- ML gives a "label" $(\alpha, \beta, \gamma, \cdots)$ to each signal event. We can assign some number ("probability") to a possible Model (A, B, C, ...) depending on the label.

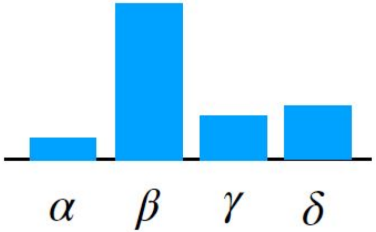


- Using MC simulation, we can create a (normalised) "template" histograms for each model.

Model A

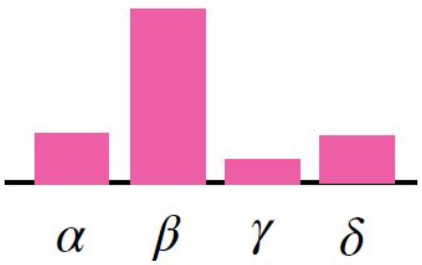


Model B

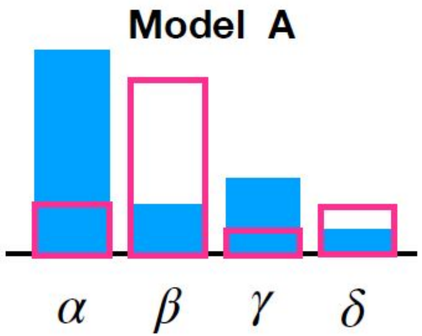


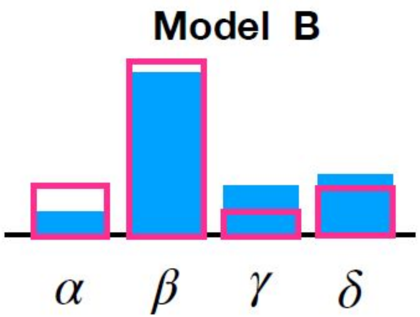
...

- For a given size of signal events, observed at the LHC with a given integrated lumi, we can create the same histogram.



- We compare the observed histogram with the template and calculate χ^2 ==> p-value





...

- **Those p-values give us the likelihood that those models are realised in nature. The likelihood is improved (gets smaller or larger) as the integrated luminosity increases.**