Hashing Classification for charged particle tracking

Luiza Adelina Ciucu (ATLAS)

17 July 2020

UNIVERSITÉ
DE GENÈVE

FACULTY OF SCIENCE
Physics Section

Introduction

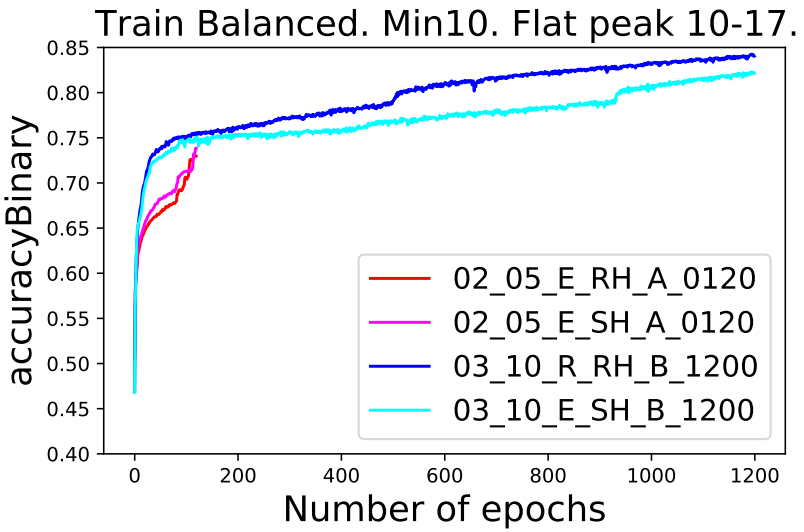
- 100 events. For each group of 10: 7 train, 3 test.
- If nbPositiveHit<10, set nbPositiveHit=0 and output made only of -1.
- For Train used Balanced (130k), for Test use Unbalanced (3.2M).
- Balancing Train in two steps, as shown last time.
  - Make peak flat between 10-17 (with value of 17).
  - Reduce nbPositiveHit=0 until 50% Pos, 50% Neg.
- Train in Balanced, evaluate in Test in Unbalanced.
- Last time tuned the hyperparameters for the best model (dark blue in our plots): `02_05_E_SH_A_0120` $\rightarrow$ `03_10_R_RH_B_1200`.
  - 02 $\rightarrow$ 03 hidden layers
  - 05 x 20 $\rightarrow$ 10 x 20 nodes on each hidden layer.
    Input layer has 3 x 20 nodes, output layer has 1 x 20 nodes.
  - Elu $\rightarrow$ Relu activation function for all nodes on all the hidden layers.
  - Squared Hinge $\rightarrow$ Regular Hinge as loss function.
  - Added one dropout layer (0.2) at the end of the hidden layers.
  - 120 $\rightarrow$ 1200 epochs, each with batch size of 50000.
  - Kept Tanh activation function on the output layer.
- New studies today: calculate efficiency of particle reconstruction.

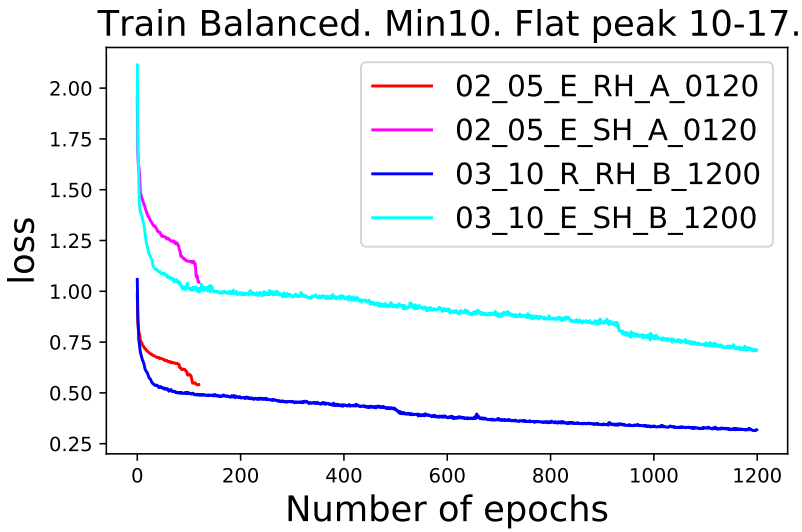
Calculate efficiency of particle reconstruction

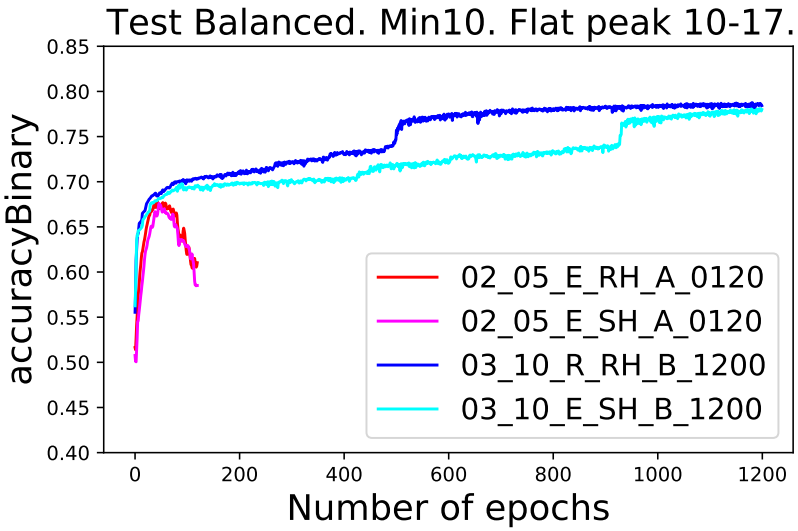
- numpy arrays: output and output predicted.
  - each row is a bucket.
  - 20 columns (20 hits)
- Loop over buckets. For each bucket first loop over hit indices:
  - for each hit index have the output and output predicted value (-1 or 1).
  - nbHitPositive = count nb of hits that are positive
  - nbHitTruePositive = count nb of hits that are both positive and predicted positive
- For the current bucket if nbHitPositive ≥ 10 consider the bucket has a truth particle in it.
- If in addition the bucket also has nbHitTruePositive/nbHitPositive $>$ 80%, consider the bucket has also reconstructed that particle.
- Increase counters for both. At the end of the loop over buckets, calculate efficiency as nbParticleReco / nbParticleTruth.
- Train or Test, eff, nb bucket, nb particle truth, nb particle recon.
- Train Balanced, particle efficiency=84.2%, 130k, 94k, 79k.
- Test Balanced, particle efficiency=74.9%, 62k, 45k, 34k.
- Test Unbalanced, particle efficiency=71.3%, 3219k, 1178k, 840k.

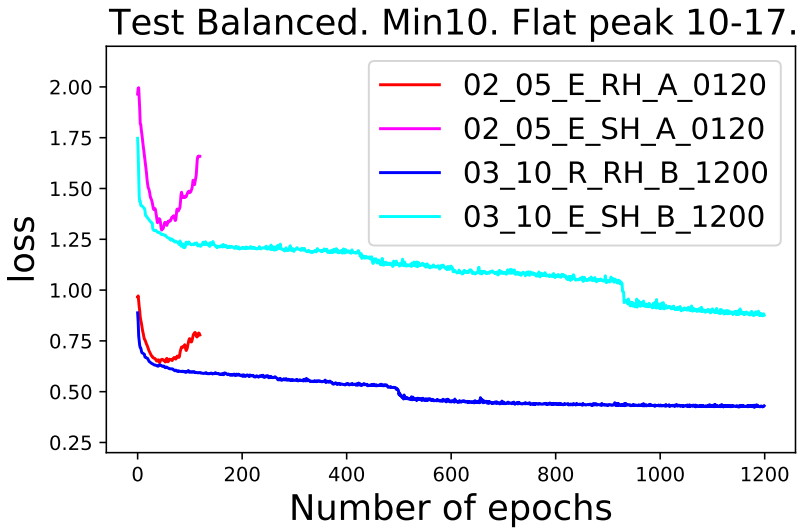
Accuracy and Loss from training

- For 120 epochs, Test shows unbalanced, for 1200 Test balanced.



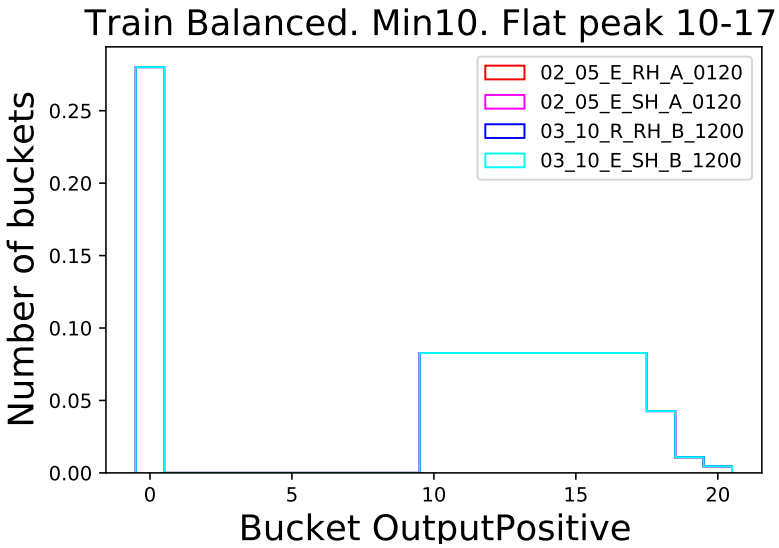


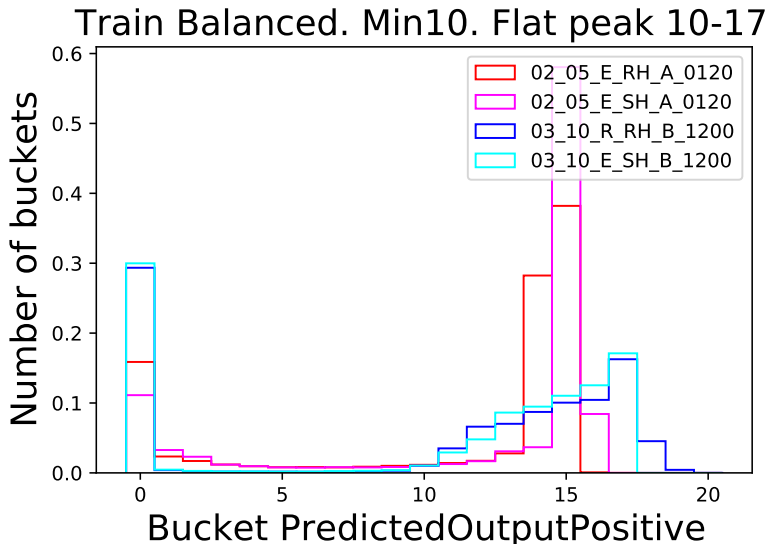


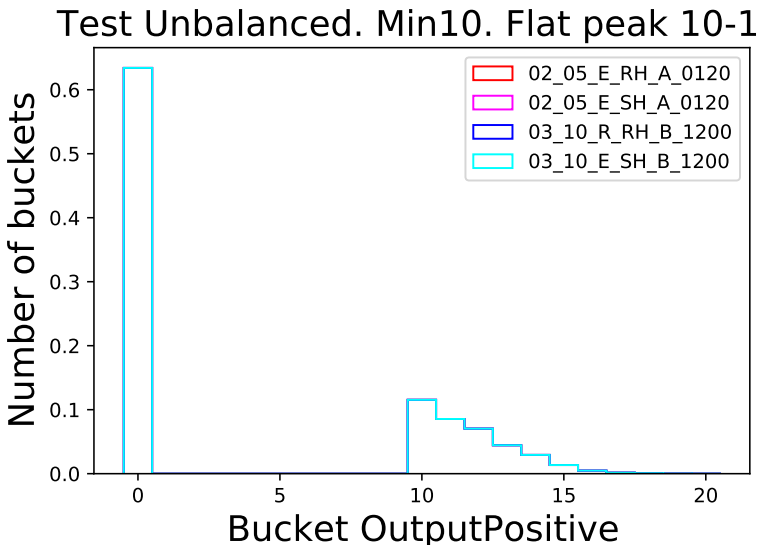


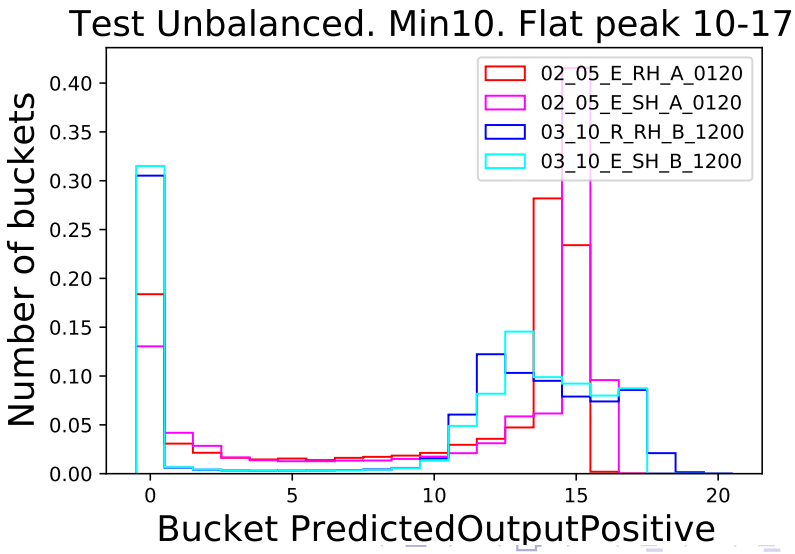
Output and output predicted 1D

- Older ones have a sharp peak, the new ones are more flat.



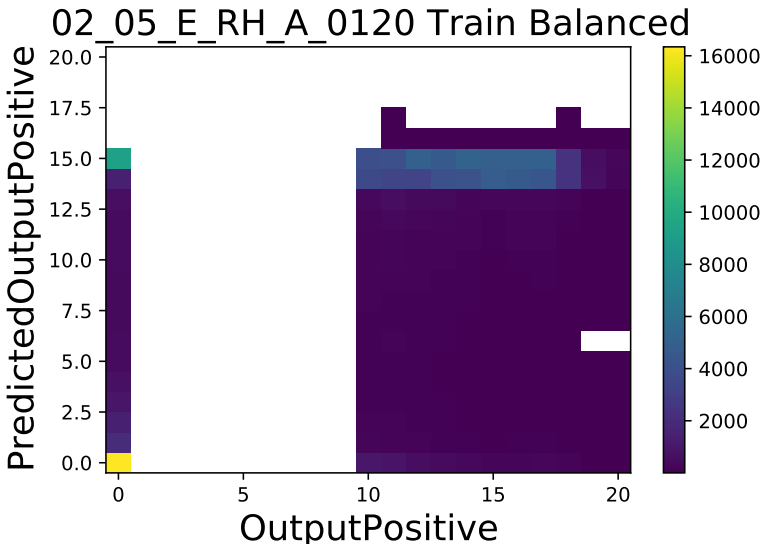


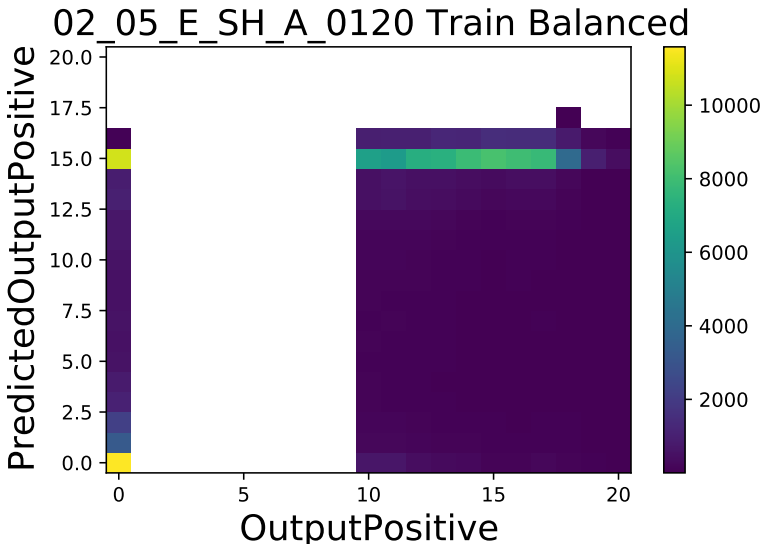


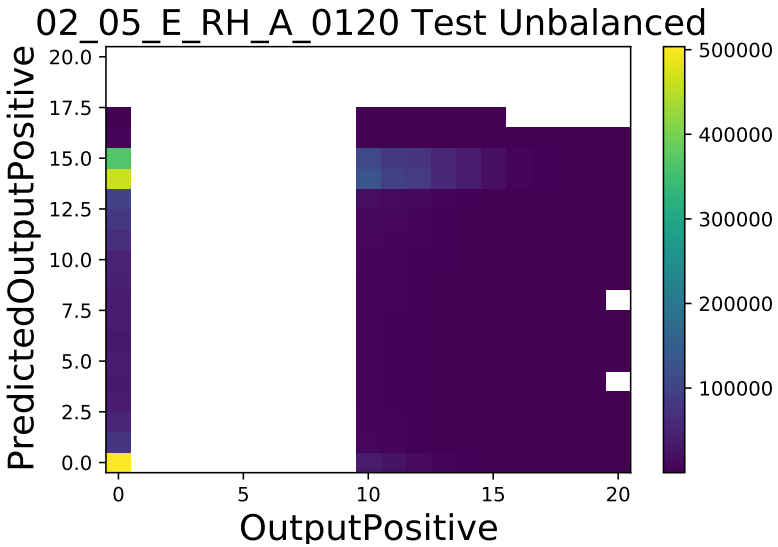


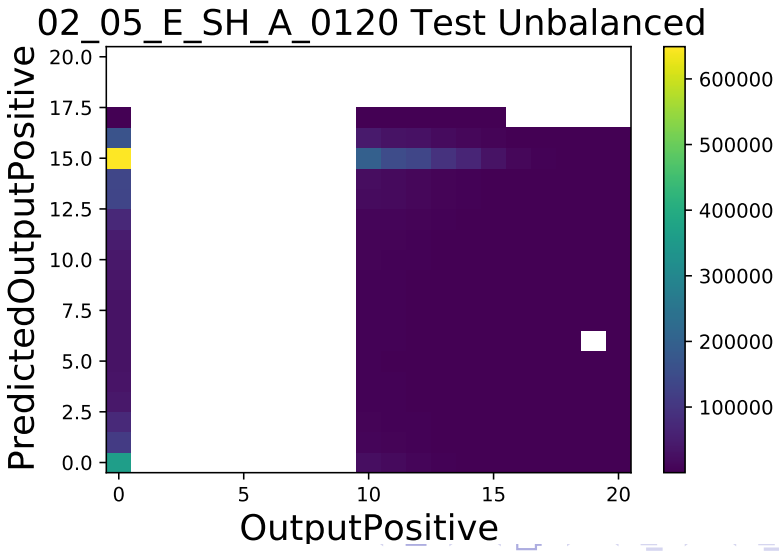
Output and output predicted 2D 1/2 - last week

- No diagonal in either Train or Test.



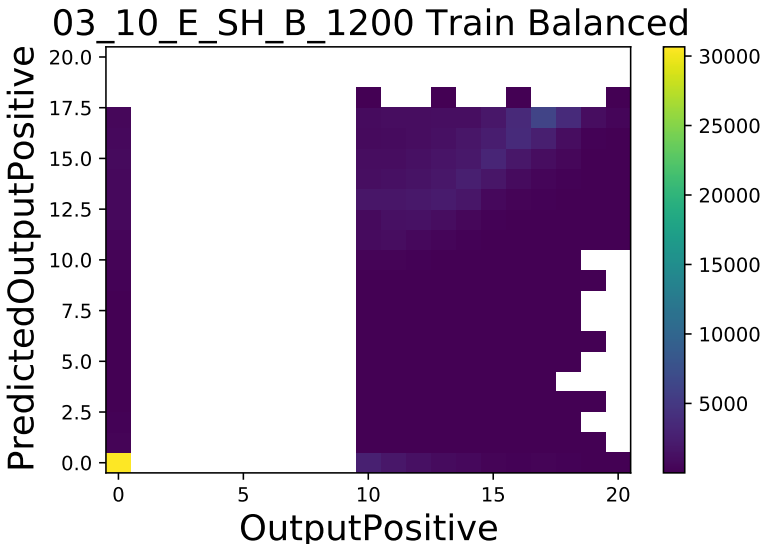


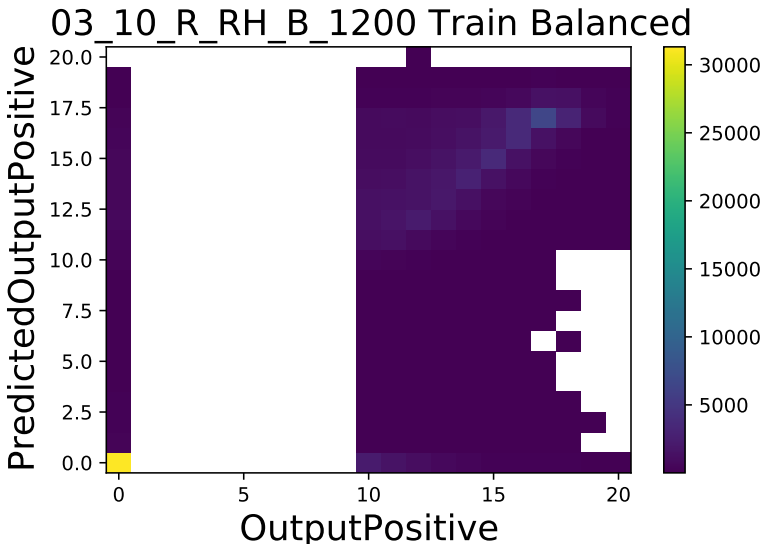


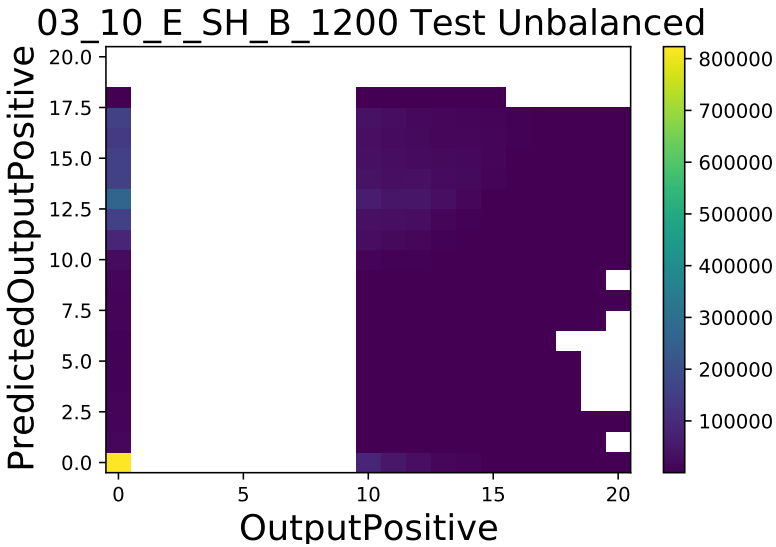


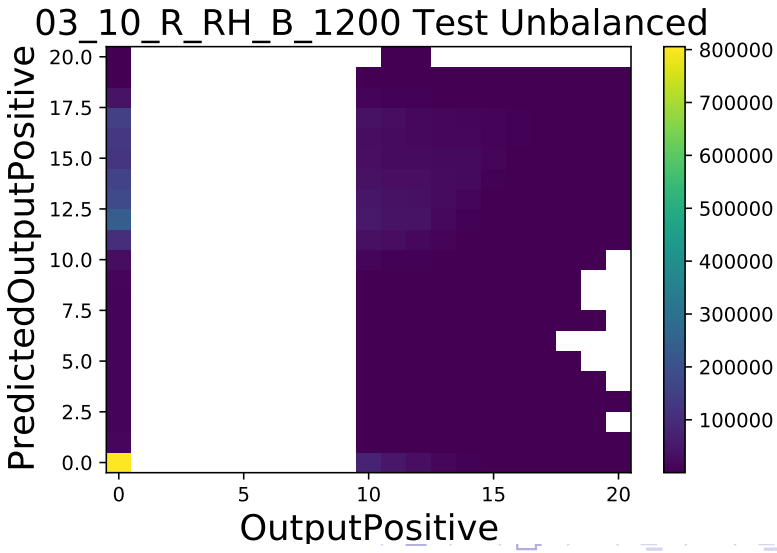
Output and output predicted 2D 2/2 - new models

- In Train balanced, a diagonal is seen. In Test unbalanced it's harder.





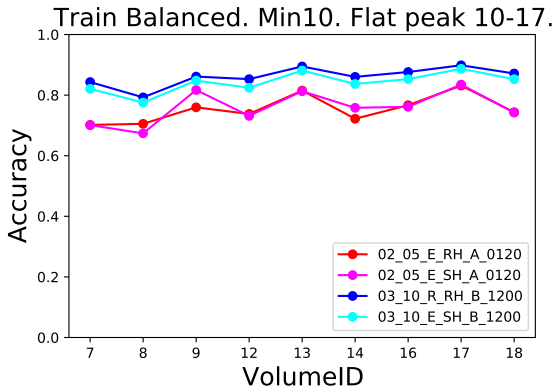


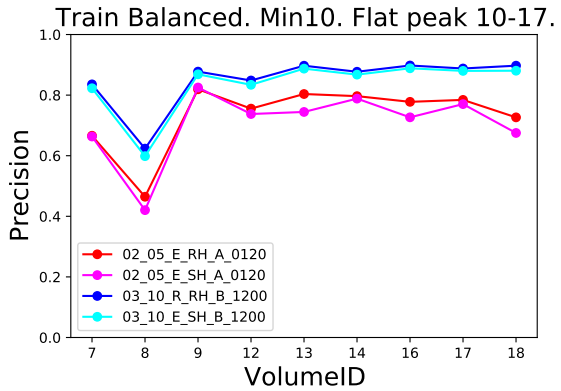


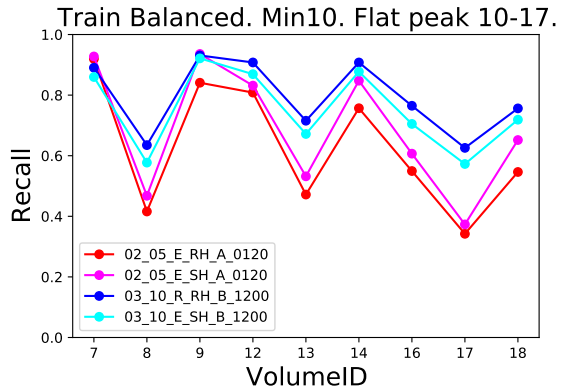
Metrics for each VolumeID.

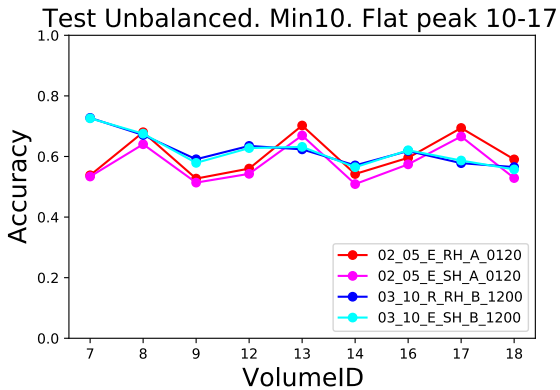
- The new ones (bluish) have a clearly better Train Balance and only a slightly better Test Unbalanced than the old ones (reddish).

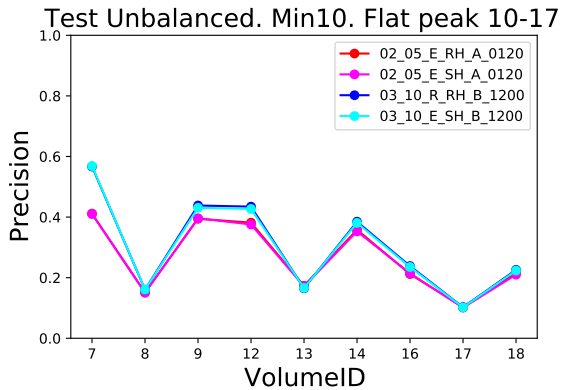
Accuracy	Precision	Recall
$\frac{TP+TN}{TP+FP+FN+TN}$	$\frac{TP}{TP+FP}$	$\frac{TP}{TP+FN}$

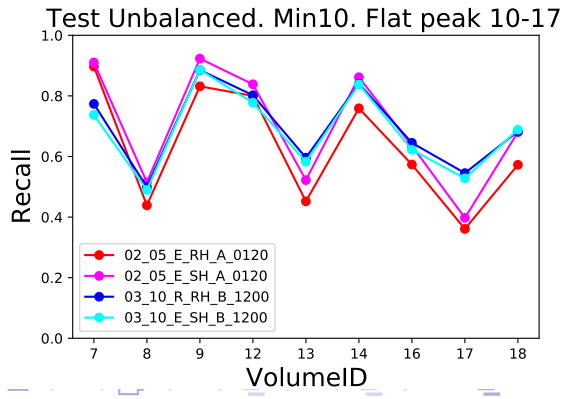












Conclusion 1/2

- Improved model from last week.
- In balanced dataset (Train and Test) a diagonal can be nicely seen.
- In Test unbalanced it is harder to see,
  but still values look relatively flat in 1D.
- Changes to the model for the best choice:
  `02_05_E_SH_A_0120` $\rightarrow$ `03_10_R_RH_B_1200`.
  - 02 $\rightarrow$ 03 hidden layers
  - 05 x 20 $\rightarrow$ 10 x 20 nodes on each hidden layer.
    Input layer has 3 x 20 nodes, output layer has 1 x 20 nodes.
  - Elu $\rightarrow$ Relu activation function for all nodes on all the hidden layers.
  - Squared Hinge $\rightarrow$ Regular Hinge as loss function.
  - Added one dropout layer (0.2) at the end of the hidden layers.
  - 120 $\rightarrow$ 1200 epochs, each with batch size of 50000.
  - Kept Tanh activation function on the output layer.

Conclusion 2/2

- Evaluated for this best model the efficiency of particle reconstruction:
- Train or Test, eff, nb bucket, nb particle truth, nb particle recon.
- Train Balanced, particle efficiency=84.2%, 130k, 94k, 79k.
- Test Balanced, particle efficiency=74.9%, 62k, 45k, 34k.
- Test Unbalanced, particle efficiency=71.3%, 3219k, 1178k, 840k.
- This model seems good enough.
- Next step: add these results in the thesis.