Neural Network KLM trigger

by Anthony Little

Belle II
University of Sydney

TRG Meeting 4/10/24

Motivation

- More sophisticated trigger over current one
- Improved exclusion of Cosmic Muon Background
- Improved Muon and Hadron identification

Data and Clustering

Data

- Use HLT calibrated data from experiment 30 run 3508 for training and run 3505 for testing for muon and hadron data
- Use experiment 30 run 3461 for training and run 3460 for testing for cosmic
- Added stipulation that Muon events require at least MuonID reconstruction ≥ 0.9 and Hadron events require a KLID reconstruction ≥ 0.3

Clustering

- Simple clustering algorithm
- If a single sector has 7 or more hits cluster all hits in that a sector

Neural Network based KLM TRG

General Framework

- Deep neural network using trigger level info
- Output Identifier of particle type (Muon, Hadron, Cosmic)
- Output is a 3 length softmax likelihood function
- L(Muon), L(Hadron), L(Cosmic)
- Separate NNs for EKLM and BKLM

THE UNIVERSITY OF SYDNEY

Neural Network overview

Software

- Developed in Tensorflow and using a Keras based Deep Neural Network

Input Features

- First Layer
- Total Number of Unique Layers
- Simple average: $0.5 * (max(strips) + min(strips))$
- Simple average of ϕ and z strips in First Layer of cluster
- Simple average of ϕ and z strips in Last Layer of cluster
- Simple average of ϕ and z strips in First Layer outside the cluster

Output Features

- 3 length softmax output of likelihood of particle type
- $(L_{Muon}, L_{Hadron}, L_{Cosmic})$

Neural Network Development: MC vs Raw

- Original model worked off MC data, generated version v08-01-00
- Muons and Hadron were ParticleGun simulations and CRYInput used for Cosmic
- When tested on raw data, model performed very poorly if trained on MC
- Main issue of concern was the clustering algorithm is rather simple, and affected by background
- CRYInput wasn't able to produce cosmic similar to those in the raw data

Neural Network Development: Positional Information

- Model was supposed to output positional information of ϕ, θ and dz of cluster
- dz and ϕ predictions had very large uncertainties even when trained on MC
- Switch to training on raw data made predicting positional info much more resource intensive so is removed

Neural Network Development: Input Features

- Model used Sector and Section info for positional info, useless with purely classification based model
- Simple average replaced difference of ϕ and z as it performed better for raw data
- Total number of hits of cluster had very good separation for MC data, but due to simple clustering, no improvement on results for raw data

THE UNIVERSITY OF SYDNEY

Configuration of Neural Network overview

- Total trainable parameters of 1667
- 65% sparsity pruned

Layer	Nodes	Activation
Trivial Input	9 (non-trainable)	N/A
1st Hidden	64	tanh
2nd Hidden	16	tanh
ID output	3	softmax

THE UNIVERSITY OF SYDNEY

Training of Models

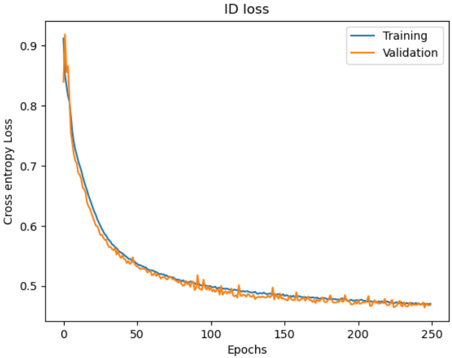
Event distribution

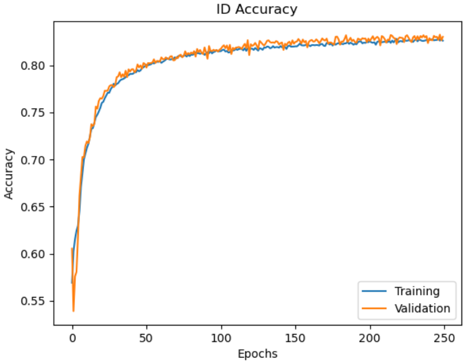
- Required at least one Primary Sectors (minimum 7 hits)
- 8000 Muon events, 8000 Hadron events, 8000 Cosmic events
- 80% events used for training
- 20% events for validation
- 4000 Muon, Hadron and Cosmic events respectively for testing (different run)

Hyper-parameters

- Learning rate: 0.001
- Batch size: 64
- Epochs: 250
- Loss: Cross Entropy

Particle ID likelihoods Model Training Results





hls4ml Model Synthesis/Performance

hls4ml

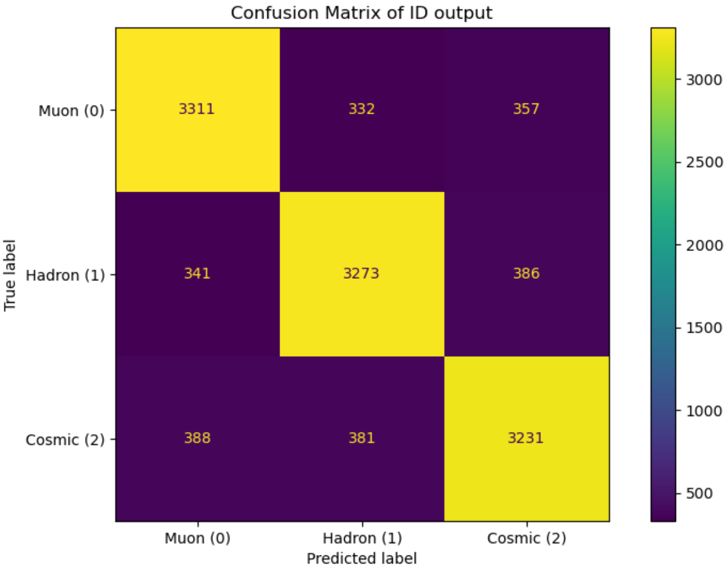
- Program to convert Keras based model in python, to work in Verilog/Vivado and on FPGAs used in trigger
- FPGA part used: XCVU080-FFVB2104-2-E (UT4)
- More details on backup slides

Total Latency	Cycles	BRAM	DSP	FF	LUT
115 ns	16	1%	45%	$< 1\%$	5%

Result	Keras Model	hls4ml Model
Accuracy	83.0%	81.9%

THE UNIVERSITY OF SYDNEY

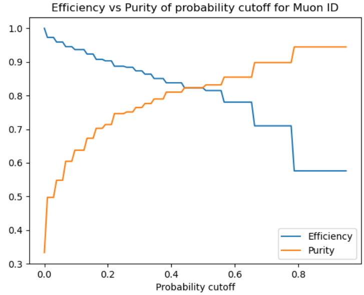
Confusion Matrix for hls4ml model

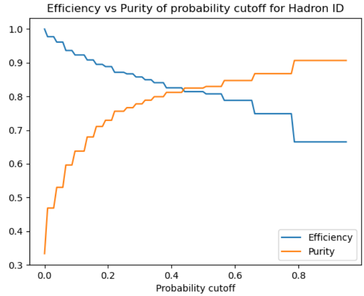


VERSITY OF
NEY

Model Output into trigger conditions

- Need to use these likelihoods to use generate actual trigger conditions
- Cut along likelihood to determine if this event is worth keeping
- Independent cuts along Muon and Hadron likelihood didn't show very good results
- Keeping 90 – 95% efficiency results in purities between 60 – 70%





RSITY OF
NEY

Binary likelihoods: Idea

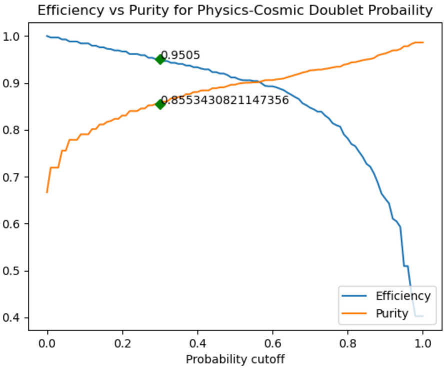
- Main concern is cosmic misidentification
- Muons being identified as hadrons and vice versa aren't as bad
- Make cuts along binary likelihoods using model output
- e.g. Muon: $\frac{L_{Muon}}{L_{Muon}+L_{Cosmic}}$
- Looks at the models correlation between likelihood predictions
- Direct calculation of trigger rate and cosmic misidentification

THE UNIVERSITY OF SYDNEY

Binary likelihoods: Use

- Event has 2 triggers: Muon: 0 or 1 Hadron: 0 or 1
- 2 stage trigger
- Cut along entire physics (non-cosmic) likelihood: $\frac{L_{Muon}+L_{Hadron}}{L_{Muon}+L_{Hadron}+L_{Cosmic}}$
- If passed, then cut along Muon-Cosmic and Hadron-Cosmic likelihoods, if pass then muon and/or hadron trigger
- Additionally: If model predict 1 likelihood for physics event, skips cuts and is passed
- If they events do not make initial cut or don't trigger a muon or hadron, they are labelled as cosmics

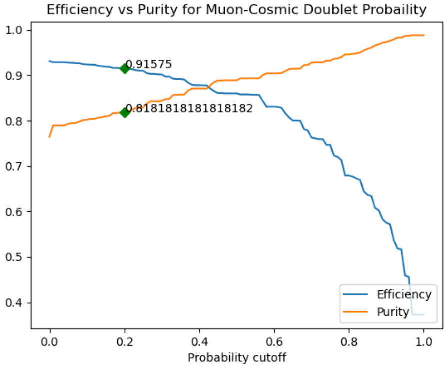
Results of Binary likelihood: Muon EvP



- Trigger cares about High efficiency: +90-95%
- Highlighted cut at 0.3
- 95.1% of Physics events are kept, and of the kept events 85.5% are Physics events
- Trigger rate needs real numbers: 7604 correctly identified as Physics events, 1284 Cosmics identified as Muons

THE UNIVERSITY OF SYDNEY

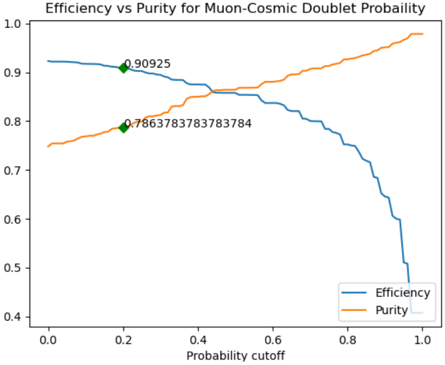
Results of Binary likelihood: Muon EvP



- Trigger cares about High efficiency: 90-95%
- Highlighted cut at 0.2
- 91.6% of Muon events are kept, and of the kept events 81.8% are Muons
- 3663 correctly identified Muons, 814 Cosmics identified as Muons
- Less important: 167 Hadrons identified as Muons

THE UNIVERSITY OF SYDNEY

Results of Binary likelihood: Hadron EvP



- Trigger cares about High efficiency: 90-95%
- Highlighted cut at 0.2
- 90.9% of Hadron events are kept, and of the kept events 78.6% are Hadrons
- 3637 correctly identified Hadrons, 988 Cosmics identified as Hadrons
- Less important: 137 Muons identified as Hadrons

THE UNIVERSITY OF SYDNEY

Multiple conditions

- Will get events will multiple conditions of both Muon and Hadron
- 1130 Muon events and 565 Hadron events and 516 Cosmic events are labelled as both Hadron and Muon triggers
- Roughly half of misidentified cosmics are misidentified both for Muon and Hadron triggers

Events lost before model

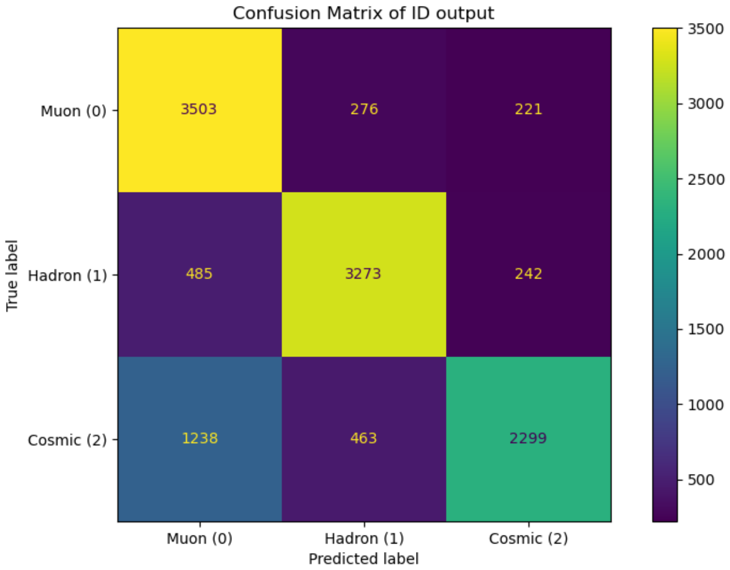
- How many events am I losing with the prepossessing steps?
- Clustering algorithm is simple, and isn't main issue
- Main issue is requiring hits in both ϕ and Z axis in the First Layer of the cluster
- 2 ways to solve: improve clustering algorithm or create backups
- In future will use Richard's Straight Line Fitter so seems unnecessary to improve clustering algorithm (and make this problem irrelevant)
- Instead created a Backup First and Last Layer, i.e. Second and Second Last Layer of the cluster

THE UNIVERSITY OF SYDNEY

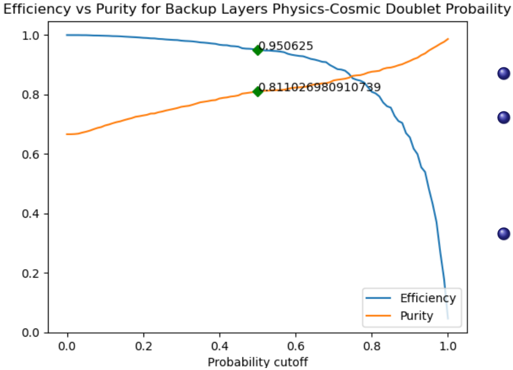
Backup layers

- Muon raw data has 20003 events, containing between 2+ Muons
- Number of particles reconstructed: 40249
- 75% of Muons interact with the barrel, so have $\approx$ 30000 potential clusters
- Current reprocessing only generates 16036 valid clusters
- By adding this backup First/Last Layers we have 21932 clusters
- However with these added events we see an issue of a drop in performance of the model, specifically is Muon-Cosmic separation

Confusion Matrix for Extra clusters

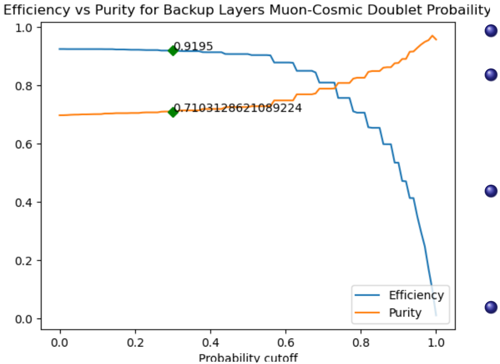


Results of Backup Layers Binary likelihood: Physics



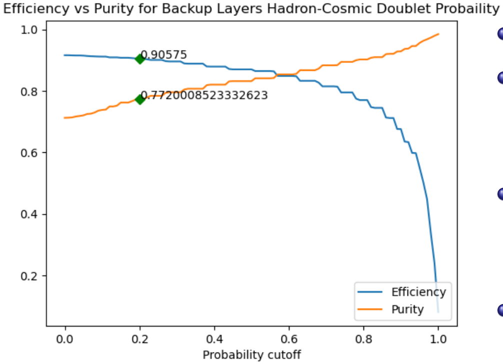
- Highlighted cut at 0.5
- 95.1% of Physics events are kept, and of the kept events 81.1% are Physics
- 7605 correctly identified Physics events, 2288 Cosmics identified as Physics

Results of Backup Layers Binary likelihood: Muon EvP



- Highlighted cut at 0.3
- 92.0% of Muon events are kept, and of the kept events 71.0% are Muons
- 3678 correctly identified Muons, 1500 Cosmics identified as Muons
- Less important: 105 Hadrons identified as Muons

Results of Binary likelihood: Hadron EvP



- Highlighted cut at 0.2
- 90.6% of Hadron events are kept, and of the kept events 77.2% are Hadrons
- 3623 correctly identified Hadrons, 1070 Cosmics identified as Hadrons
- Less important: 178 Muons identified as Hadrons

THE UNIVERSITY OF SYDNEY

Possible fix

- By reducing the sparsity reduction of the model was able to get better results, close to 81% accuracy during training
- However this sparsity reduction makes it resource usage too much
- Possible need some added input features/added complexity to overcome these new clusters

THE UNIVERSITY OF SYDNEY

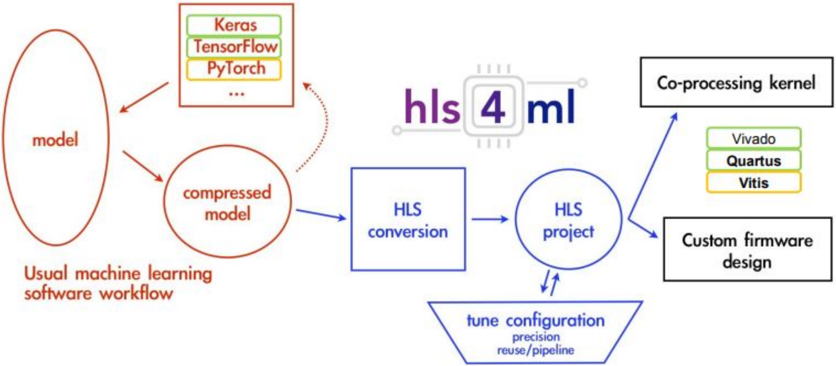
Conclusion

- Current model without backup layers: 95.1% effiency and 85.5% purity for total physics events
- Muon: 91.6% efficiency 81.8% purity
- Hadron: 90.9% efficiency 78.6% purity
- Want to investigate extra cluster issue to reach similar results
- Double check tests of Vivado test bench
- Work on this projects for 1-2months going forward then move on to my physics work
- Right up code/possible Belle2Note and still consult on future of project beyond that

THE UNIVERSITY OF SYDNEY

BACKUP SLIDES

hls4ml



FastML Team. hls4ml (Version v0.8.0) [Computer software]. https://doi.org/10.5281/zenodo.1201549

hls4ml details

hls4ml in this experiment?

- Reconstructed via RAM as this greatly reduced resource usage
- All internal layers use 16 total bits with 8 integer bits
- Output Layers use 7 total bits, with 2 integer bits
- FPGA part used: XCVU080-FFVB2104-2-E

Comparison and Synthesis of Model on FPGAs

Synthesis/Performance

- Latency and resource usage
- Bit restriction can hurt performance slightly
- hls4ml predict function vs keras predict function

Resource Definitions/Jargon

- LUT (Look Up Table): Basic logic of FGPA, generic functions that build the algorithm
- FF (Flip Flops): Build the pipeline of data with the clock pulse
- DSP (Digital Signal Processor): Performs arithmetic in the FPGA
- BRAM (Block RAM): Additional Memory usage