Machine learning examples in Hyper-Kamiokande

Sergio Suárez

(suarezsergio@uniovi.es)

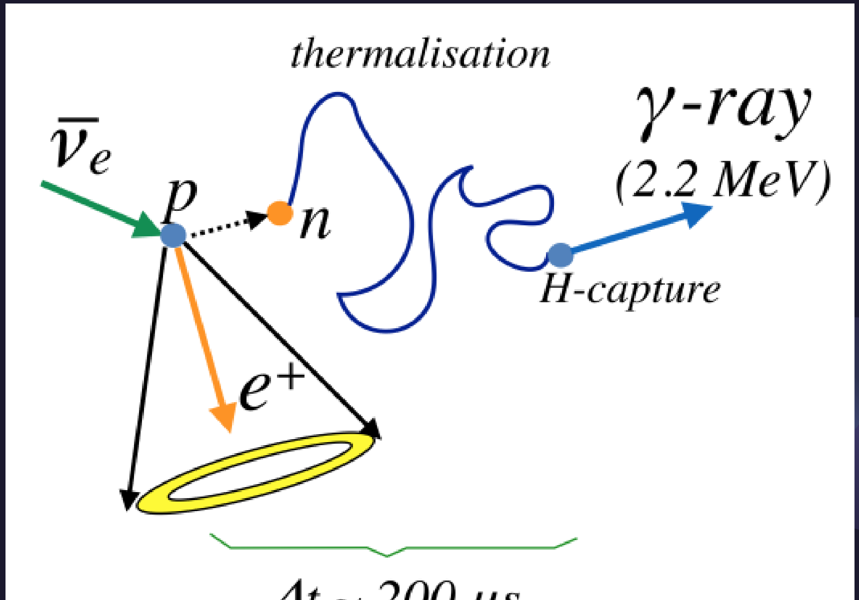
Summary

- Neutron tagging
- Event reconstruction
- Tau identification

Introduction

Neutrons produced in the detector thermalize and are captured by hydrogen
After the capture a 2.2 MeV photon is emitted with a mean half-live of 200 µs

$$n + p \rightarrow D^{*} \rightarrow +D + \gamma(2.2MeV)$$



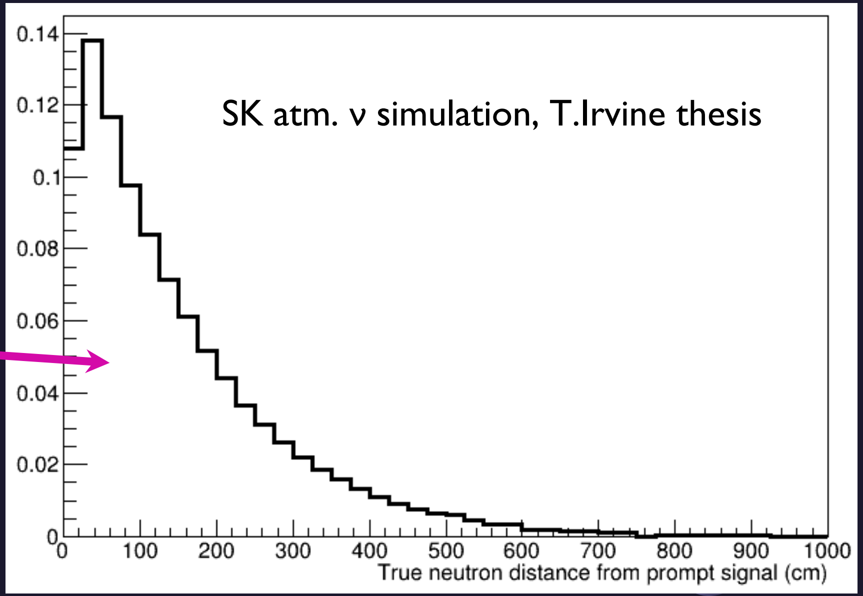
IBD sketch

H has a sizeable cross-section for capturing thermal neutrons, 0.329 barn
Given the amount of H in HK, it is guaranteed that all thermalised neutrons will be captured by hydrogen

Being this such a small signal (below the lowest energy trigger), only neutrons produced by in a previous triggered interaction are looked for

Simulation

- Following Izumiyama-san's guidance, we simulated 2.2 MeV γ uniformly distributed in the HK ID using WCSim (hybridPMT version)
- For the time being, we are using the nominal HK configuration, i.e. 20k B&L PMTs with noise levels of 4.2 kHz
- As the reconstructed prompt signal is used to compute the TOF in the search for neutron's delayed signal, it is important to account for the distance travelled by the neutron during the thermalization
  - this depends on the energy of the generated neutrons
  - conservatively we are assuming neutrons generated from atmospheric neutrinos
  - this is included after the WCSim simulation in the TOF calculation



Selection of neutron candidates with Neural Networks

UP TO NOW

- Data sorted in time windows (10 ns latest works from Izumiyama-san)
- Threshold in the number of hits in 10 ns window to be considered as candidate
    - Results around ~58% True Positive Rate (TPR) and ~3% False Positive Rate (FPR)

OUR APPROACH

- Data sorted in 30 ns windows.
- Classification of whether the data of each PMT hit comes from signal or from background noise.
- We are planning to solve both detection of the signal, and a further step, identificate inside the signal which PMT measurements correspond to signal.

Selection of neutron candidates with Neural Networks

OUR APPROACH

- As inputs for the NN we use not only the timing information but also the charge of the hit PMTs
  - A small issue! Sometimes, several hits happen in the same ns. We choose to add the intensities to give the network all the information.
    - We checked adding the intensities actually improves significantly the performance of the method
- Flag each PMT hit whether it comes from noise or signal

```
t-TOF (ns) | Q (pe) | Label (1 signal, 0 bkg)
126.599  1.10166 0
129.3098 2.00506 1
146.43456 1.38051 1
146.73961 0.813318 0
147.06379 0.47011 1
148.03917 1.06994 1
```

Neutron Tagging with Neural Networks

SOME ISSUES WE HAD TO CONSIDER

- Selection of an adequate time window
- Structure of the network and the training for this problem.
- How to give a simultaneous answer to both detection of a signal and its correspondent PMT hits identification problem

Neutron Tagging with Neural Networks

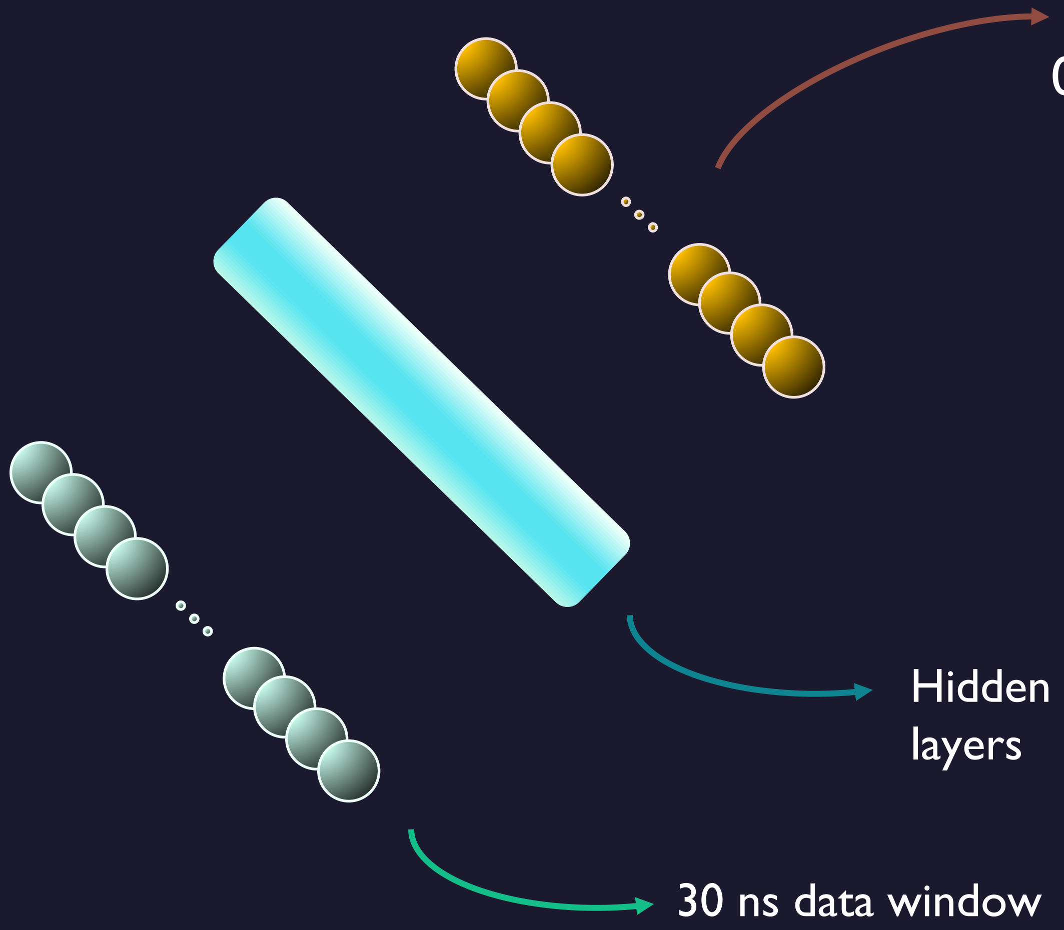
Improving traditional approach for detection only

Using what we learned from the problem to set a network that identifies the nature of each ns data in a window of 30ns

Setting a rule to define if there is a signal in the data window, as a consequence of each ns identification data

A conservative rule: at least 3 identifications are required to consider a detection

Neutron Tagging with Neural Networks



30 outputs, indentify the PMT of each ns with hit
0 or 1 corresponding to if the hit is from bkg or sgn

If 3 or more sgn identifications

Detection of the sample as signal

Neutron Tagging with Neural Networks

Input (Intensities in 30 ns window)

	0	1	2	3
11469	0.0	0.0	0.0	0.0
11470	1.0	2.0	1.0	1.0
11471	0.0	0.0	0.0	0.0

	28	29	30
11469	0.0	0.0	0.0
11470	0.0	0.0	0.0
11471	0.0	0.0	0.0

NN output

	0	1	2	3
11469	0.080491014	8.849915E-4	5.9369346E-4	8.642116E-4
11470	0.98310167	0.97287685	0.9955674	0.9927126
11471	0.18850504	0.011148328	0.011107495	0.22130708

	28	29	30
11469	6.8074017E-4	6.9838244E-4	3.9816546E-4
11470	6.439832E-4	1.1712634E-4	4.48175E-4
11471	5.7975438E-5	3.0369007E-5	1.4106388E-4

Results over simulated data

SOME NUMERIC DATA RESULTS OVER SEVERAL ROC CUTS

IDENTIFICATION	10% CUT	30% CUT
TPR	0.8207	0.6175
TNR	0.966	0.9907
FPR	0.034	0.0093

DETECTION	60% CUT	90% CUT
TPR	0.7537	0.6294
TNR	0.9726	0.9893
FPR	0.0274	0.0107

Comparison with previous results ~3% of FPR, that achieved ~58% of TPR

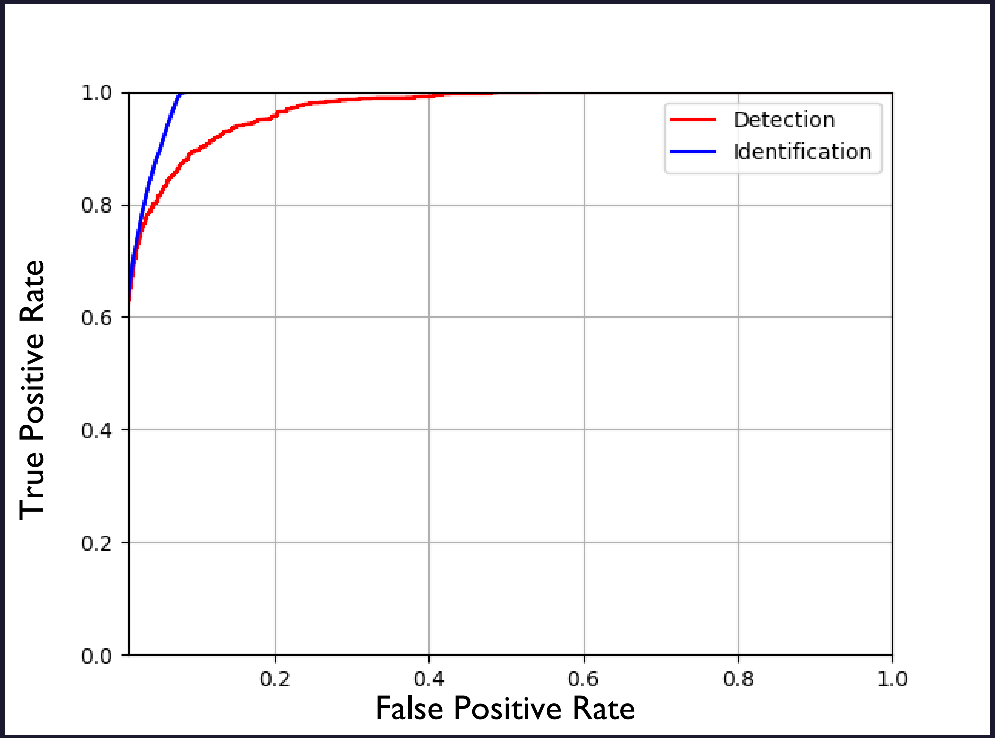
At ~1% of FPR, we are achieving up to ~62% of TPR

Identification of each ns hit has awesome results, ~3% of FPR with ~82% of TPR

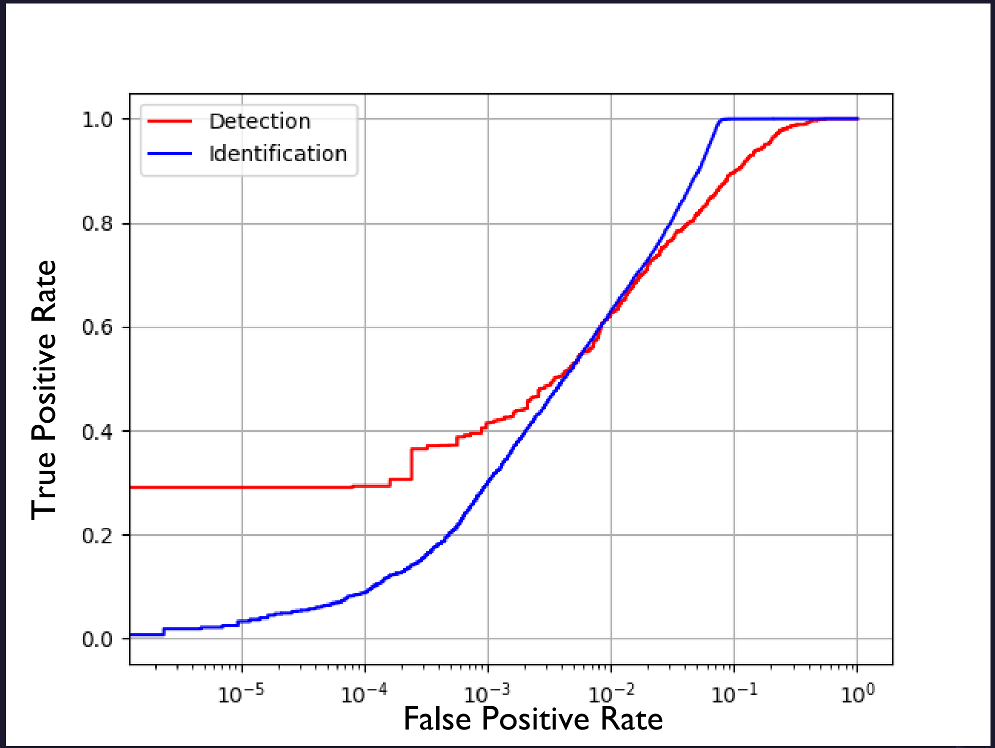
A more conservative cut in Identification of each ns hit, allows ~1% of FPR with ~62% of TPR

Results over simulated data

ROC CURVE



SAME ROC CURVE WITH LOGARITHMIC AXES



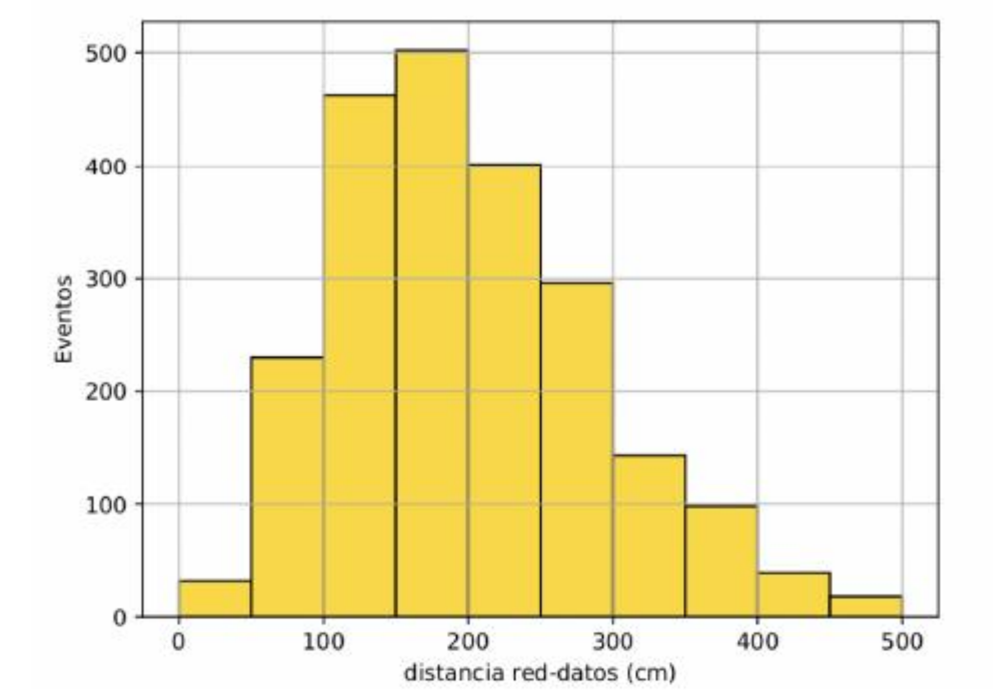
Reconstruction of high energy neutrinos using artificial intelligence techniques for the Hyper-Kamiokande experiment

TFG by Pedro Pablo Reyes Riera

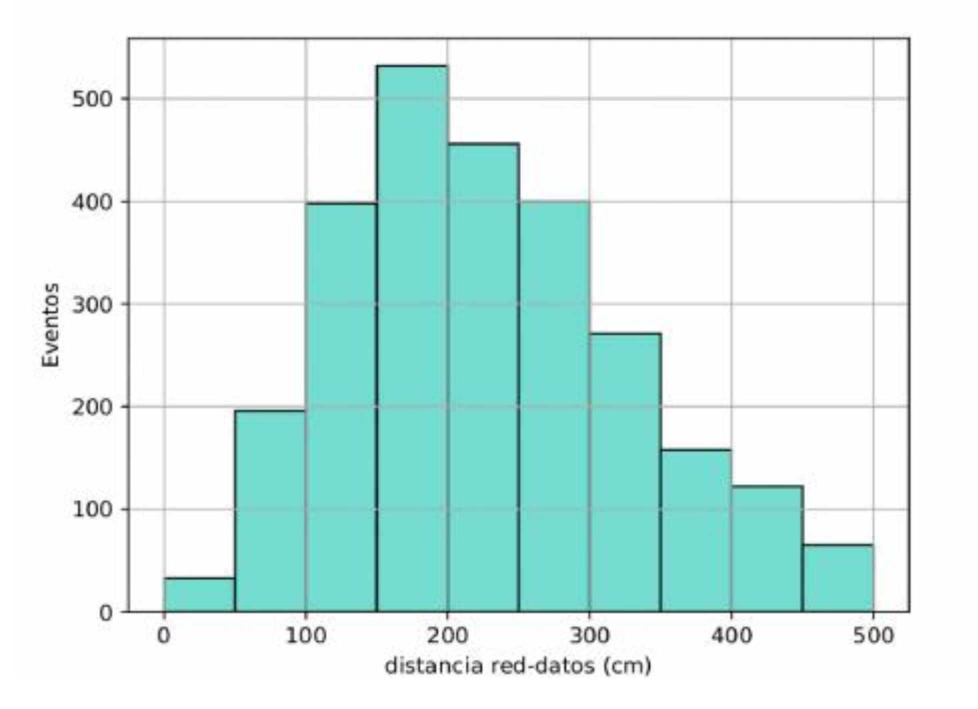
Supervised by Sergio and Pablo

- Information given by each of the PMTs was used to recover the position where neutrinos decay and thus identify their flavor, distinguishing between muonic and electronic neutrinos.
- the intensity measured by each photomultiplier and the time when each photon produced by the Cherenkov radiation reaches the PMT's were simulated.
- As a first approach, we only used the information given by those PMT's located in the cylinder's body, ignoring the cylinder's heads.
- Two tasks were developed: a classifier for neutrino classification and two additional networks for the position determination;
- Due to muons and electrons independently and each of them is used depending on the classifier output. This approach is hinted as the different nature of both kinds of particles suggest different interaction models.

Reconstruction of high energy neutrinos using artificial intelligence techniques for the Hyper-Kamiokande experiment



Histogram of electronic events: distance from the network to the actual event



Histogram of muonic events: distance from the network to the actual event

Note: In perspective, the maximum possible error corresponds to the distance of two opposed sides of the tank ~105m, which means that we are achieving a **very preliminary** 2% raw error

Application of artificial intelligence techniques for the identification of Tau neutrinos in the experiment Super-Kamiokande.

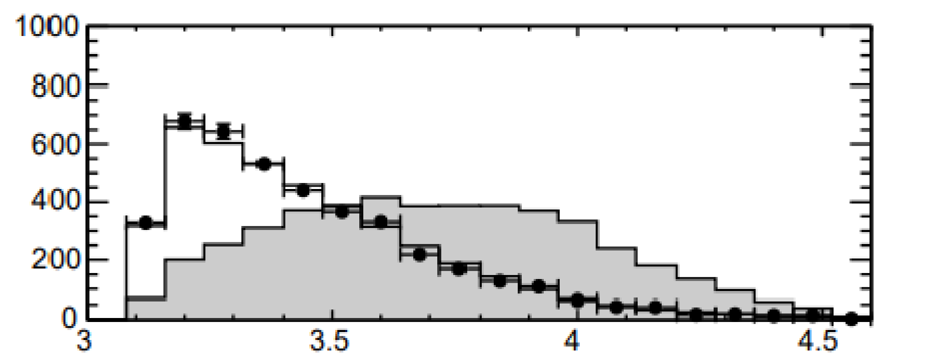
TFG by Paula Obladen Aguilera

Supervised by Sergio and Pablo

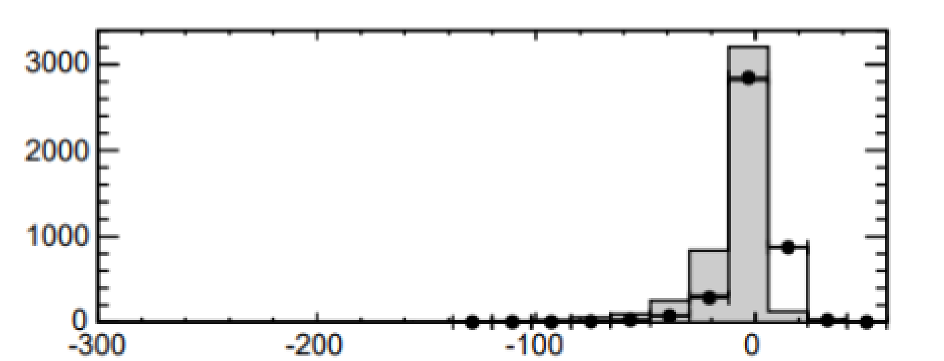
- Detection can only be placed if charged-current neutrino interactions take place. Also, its large mass and short lifetime result in a suppression of its Charged-current cross section, which makes a direct detection of these particles very difficult.

- Moreover, atmospheric neutrinos are mainly electron or muon flavoured, making the abundance of these heavier particles less probable, being neutrino flavour oscillations the main process through which Tau neutrinos are present.

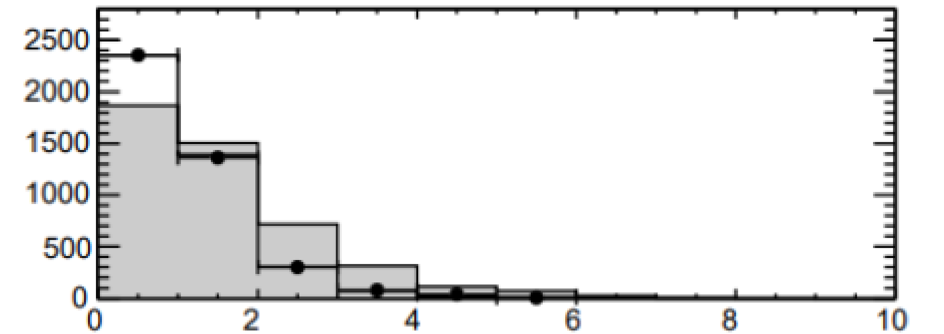
-The objective is then to separate between background and tau signal. An MLP is used in the analysis process. It takes the input directly from the MC simulation



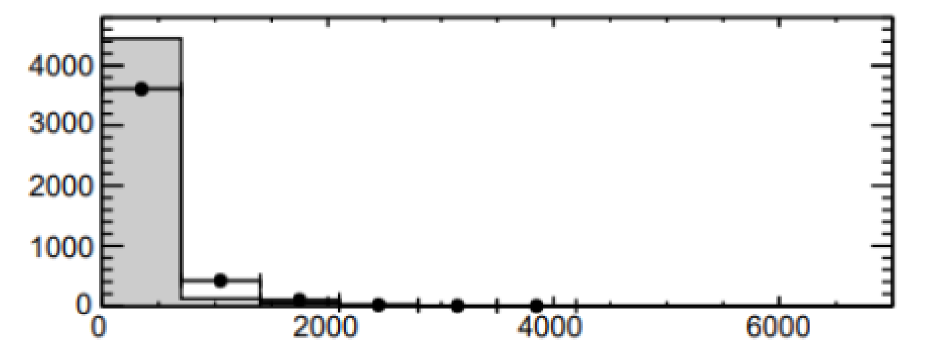
(1) Log of Visible Energy



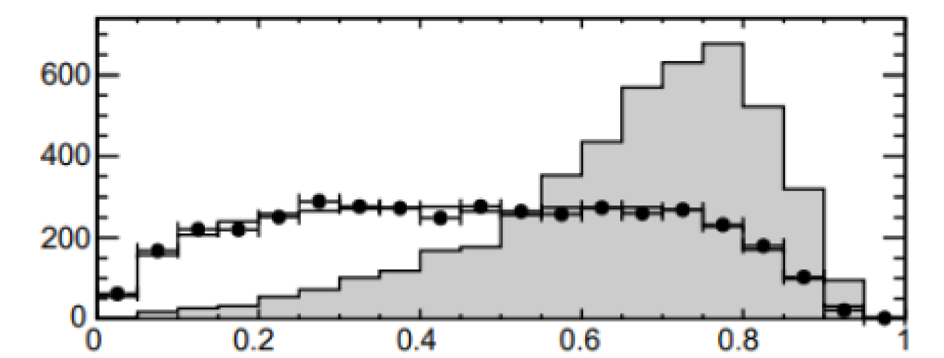
(2) ID of Maximum Energy Ring



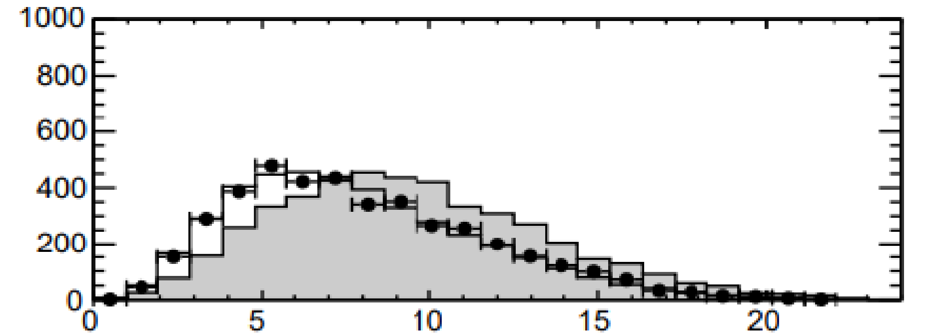
(3) Number of Decay Electrons



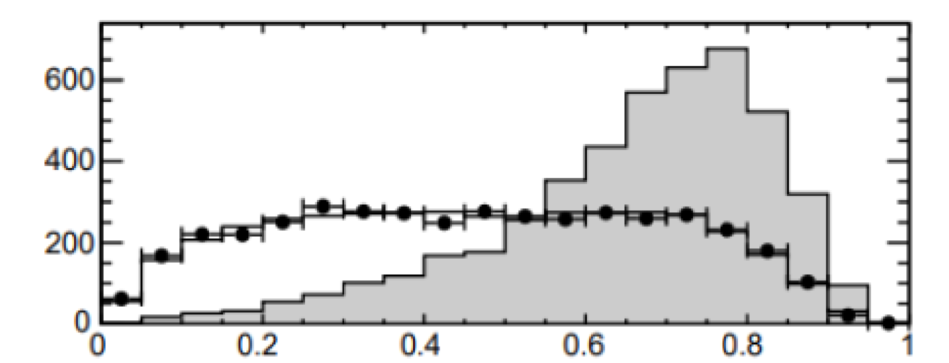
(4) Max Distance to Decay-e



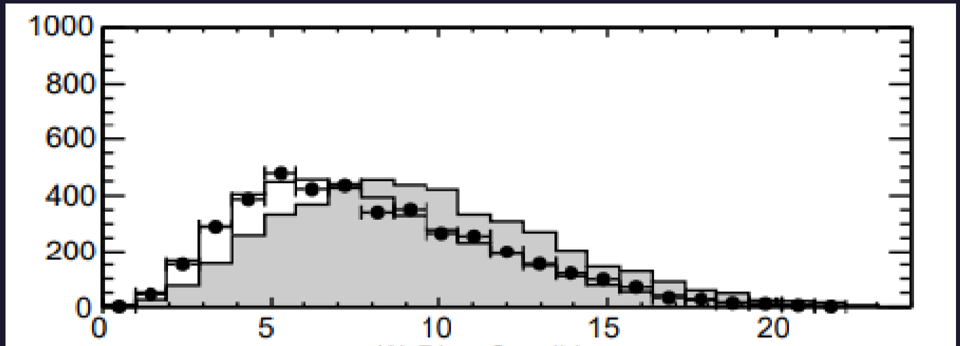
(5) Clustered Sphericity



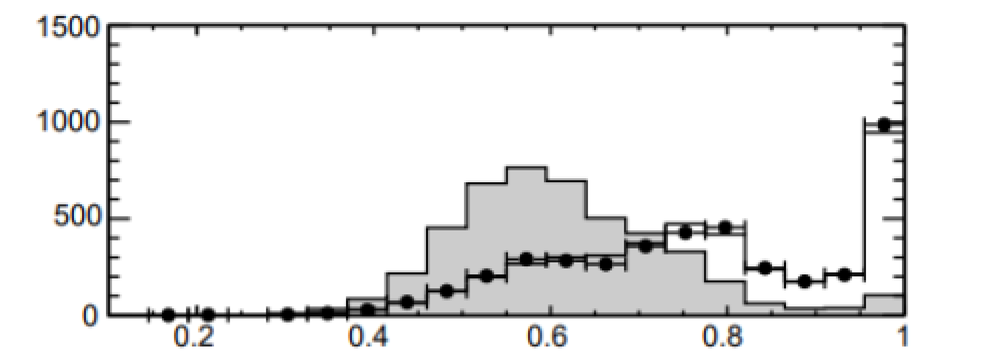
(6) Ring Candidates



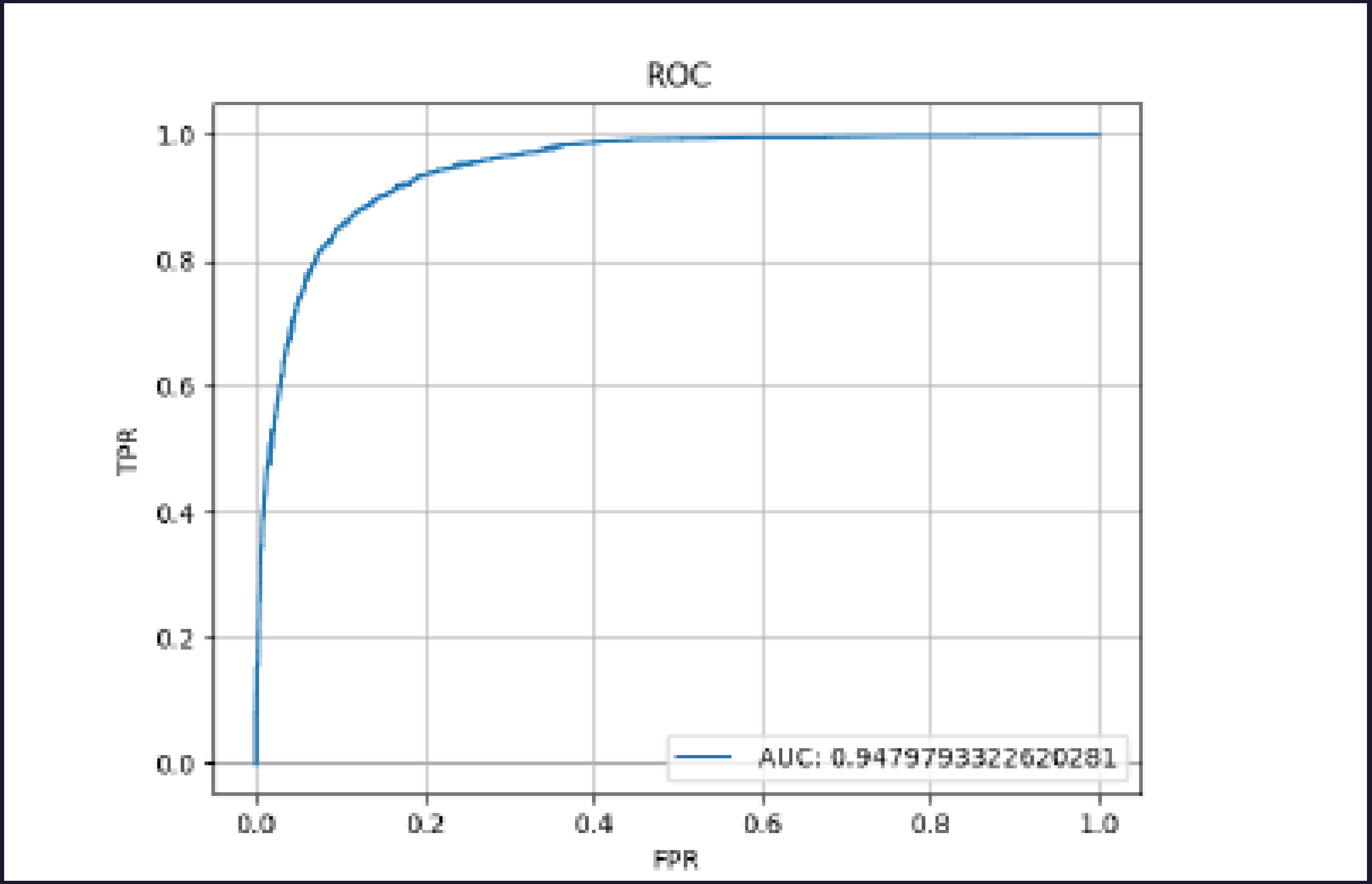
(5) Clustered Sphericity



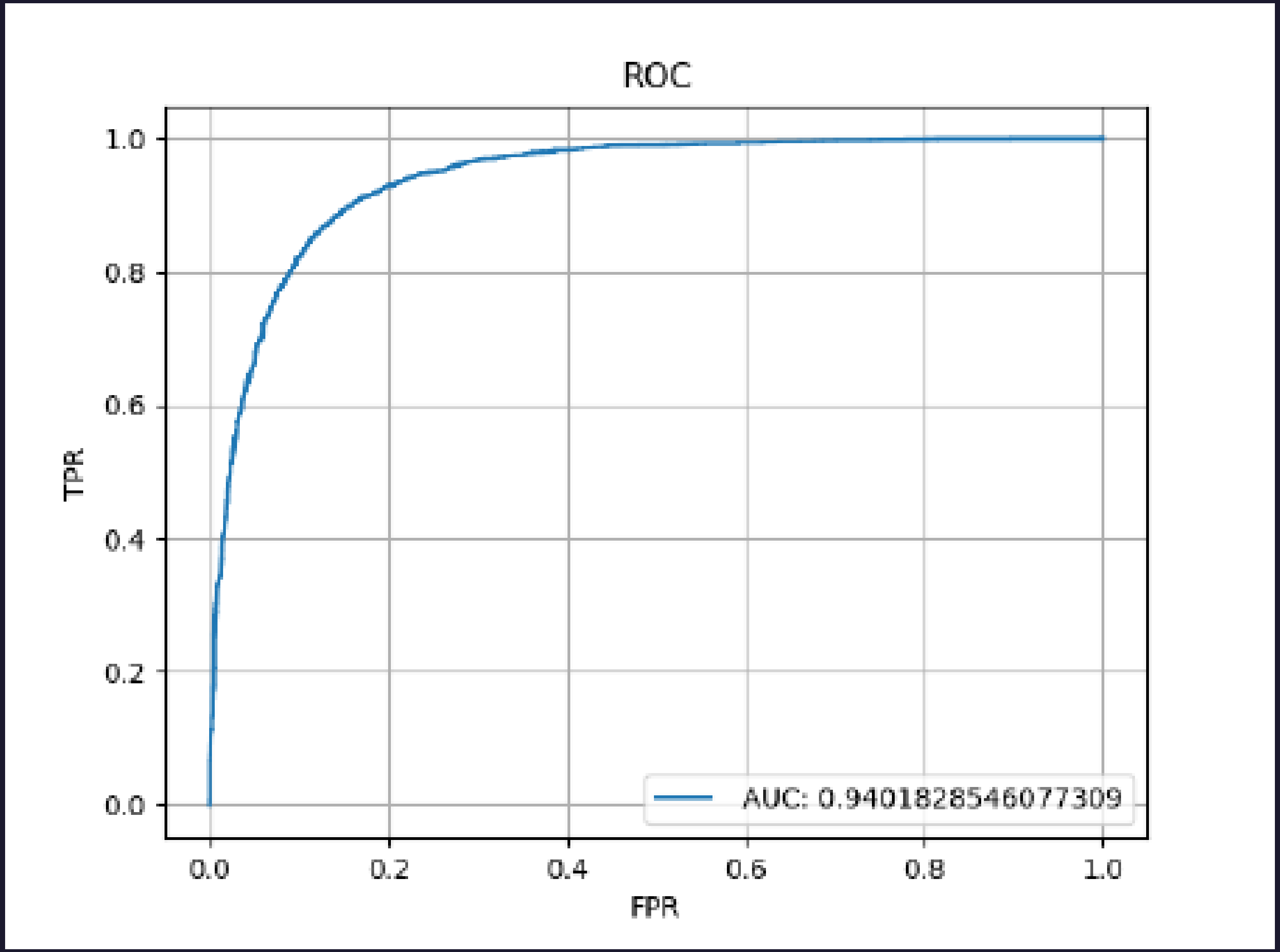
(6) Ring Candidates



(7) Fraction carried in First Ring



Classification as tau with a precision of an AUC ~0,948



Classification with only six variables;
excluding log10(Evis) with precision of an AUC ~0,940

Summary

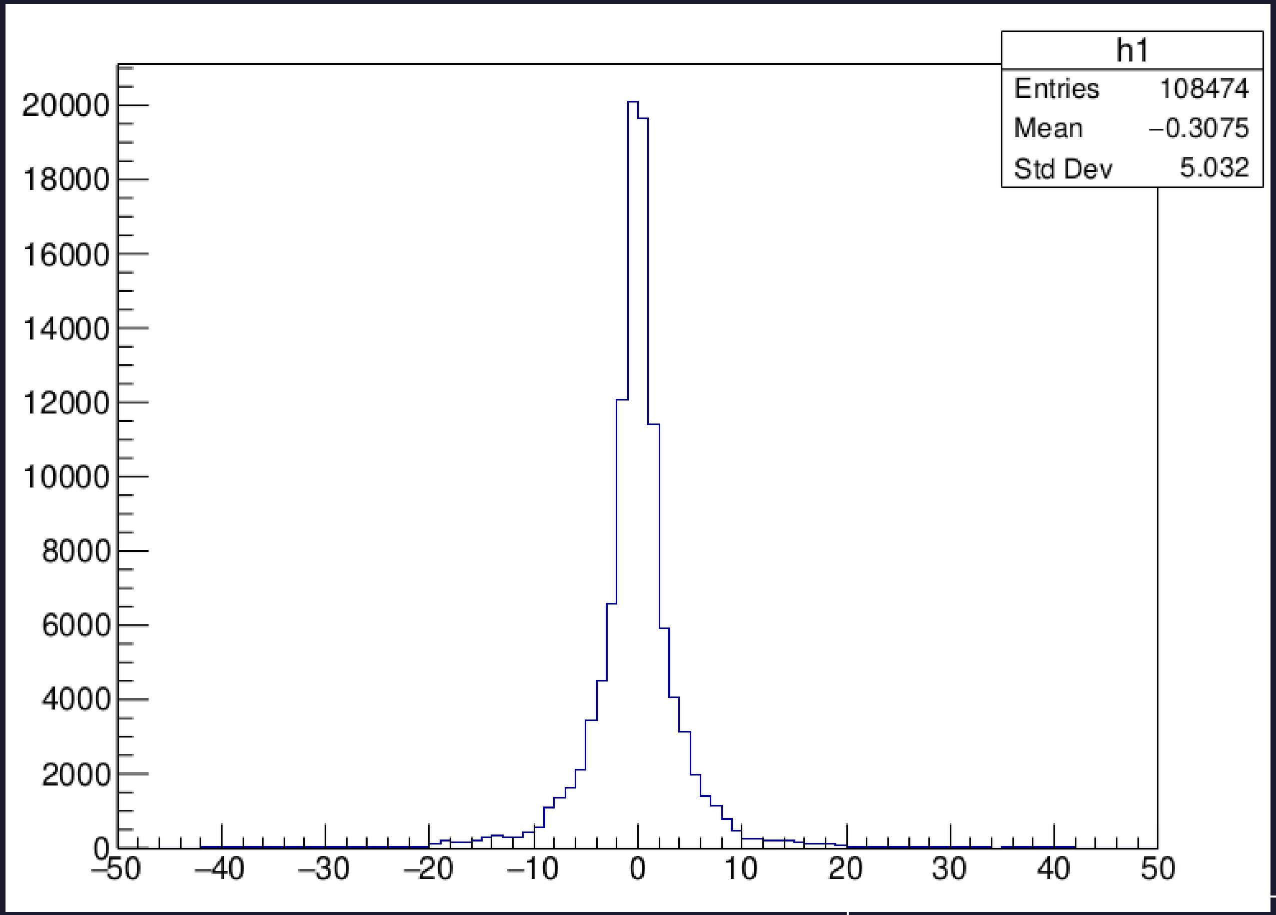
- Detection of neutron singal has increased, at ~3% of FPR, from ~58% of TPR to ~75%
- Identification of which PMTs hits are from a positive signal are classified, at ~3% of FPR, with ~82% of TPR
- Promising first steps both in recontruction and tau identification.

Next Steps

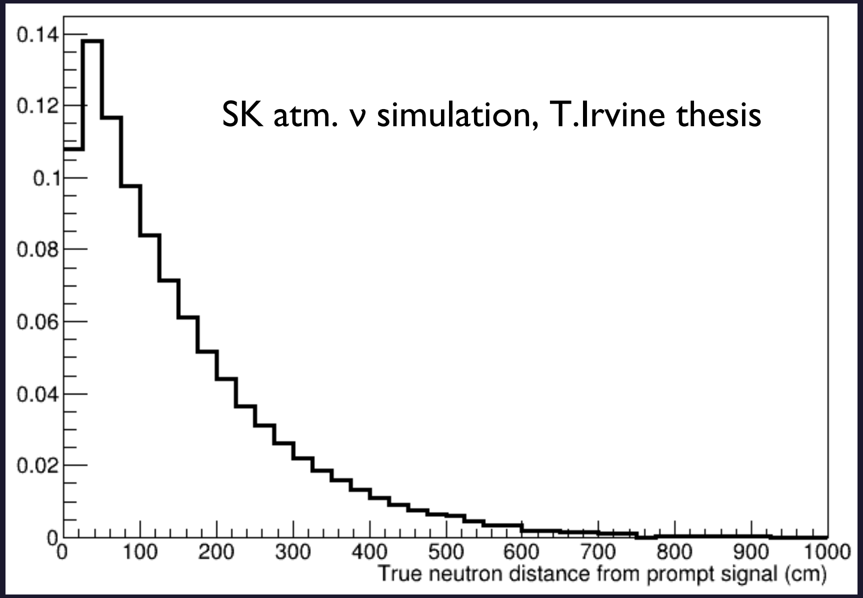
- Try out more HK configurations (mPMTs, afterpulse, increase dark rate...)
- Converge on first selection method
- WCSim python interface for better integration of machine learning methods
- Post-production code to recover timing information and merge split signals.
- Continue with the last two topics presented, hopefully, with new students for their doctoral tesis.

Back up

Simulation



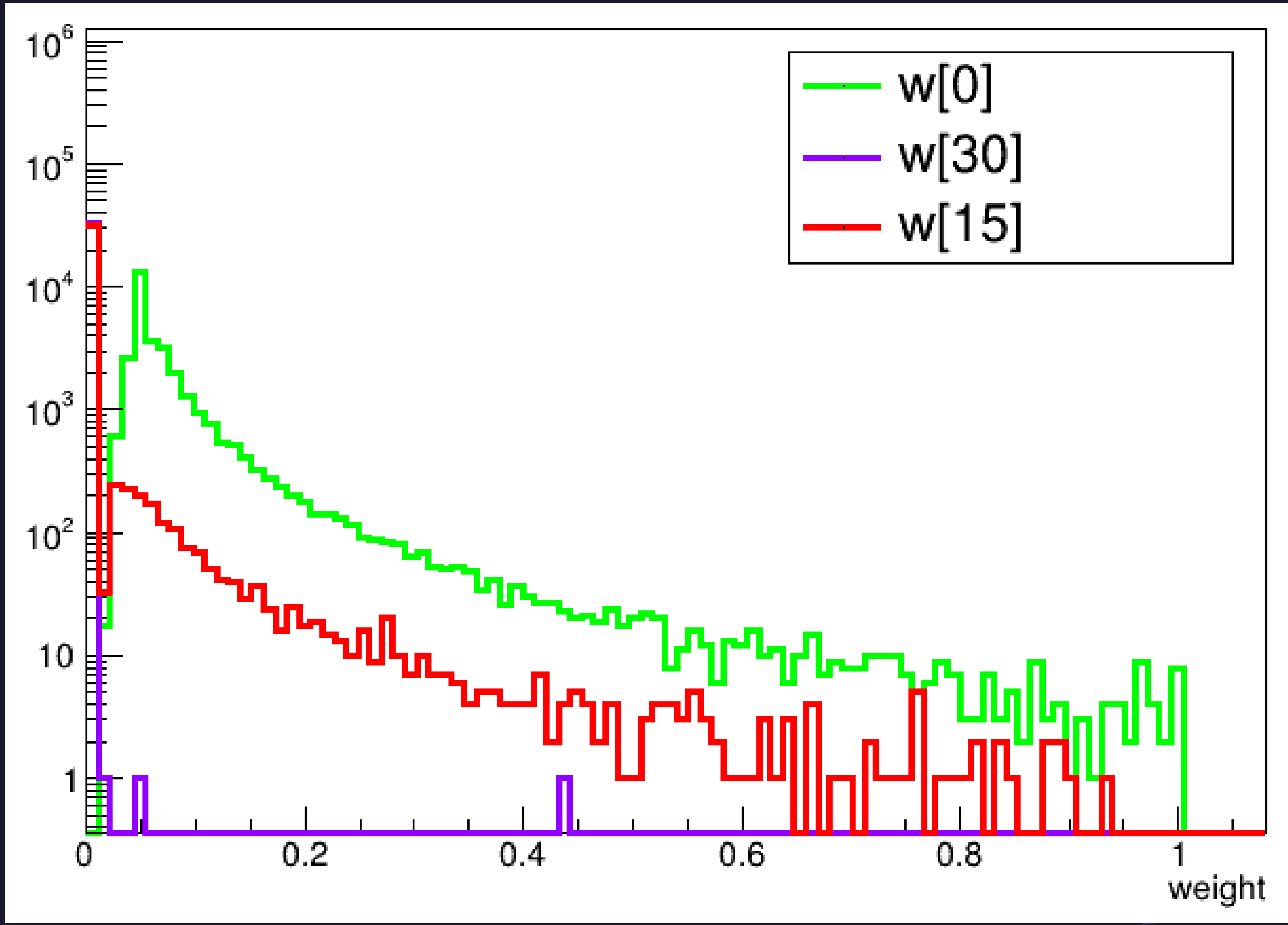
ΔTOF (true-prompt) [ns]



NN features

First nanosecond (w[0]) of the 30 ns interval shows larger values of the NN output than average of the rest

Last nanosecond (w[30]) of the 30 ns interval shows lower values of the NN output than average of the rest



Full results

```
IDENTIFICATION
['Measure ', '10% ', '20% ', '30% ', '40% ', '50% ', '60% ', '70% ', '80% ', '90% ', '95% ']
TN [407400, 415679, 417813, 418935, 419616, 420135, 420562, 420919, 421260, 421445]
FN [1824, 3242, 3892, 4388, 4831, 5222, 5674, 6174, 6935, 7609]
TP [8350, 6932, 6282, 5786, 5343, 4952, 4500, 4000, 3239, 2565]
FP [14349, 6070, 3936, 2814, 2133, 1614, 1187, 830, 489, 304]
TPR [0.8207, 0.6813, 0.6175, 0.5687, 0.5252, 0.4867, 0.4423, 0.3932, 0.3184, 0.2521]
TNR [0.966, 0.9856, 0.9907, 0.9933, 0.9949, 0.9962, 0.9972, 0.998, 0.9988, 0.9993]
PPV [0.3679, 0.5331, 0.6148, 0.6728, 0.7147, 0.7542, 0.7913, 0.8282, 0.8688, 0.894]
NPV [0.9955, 0.9923, 0.9908, 0.9896, 0.9886, 0.9877, 0.9867, 0.9855, 0.9838, 0.9823]
FPR [0.034, 0.0144, 0.0093, 0.0067, 0.0051, 0.0038, 0.0028, 0.002, 0.0012, 0.0007]
FNR [0.1793, 0.3187, 0.3825, 0.4313, 0.4748, 0.5133, 0.5577, 0.6068, 0.6816, 0.7479]
FDR [0.3679, 0.5331, 0.6148, 0.6728, 0.7147, 0.7542, 0.7913, 0.8282, 0.8688, 0.894]

AUC 0.9848837807700453
```

Full results

```
DETECTION
['Measure ', '10% ', '20% ', '30% ', '40% ', '50% ', '60% ', '70% ', '80% ', '90% ', '95% ']
TN [8209, 10570, 11309, 11677, 11858, 11987, 12070, 12125, 12177, 12193]
FN [18, 112, 199, 287, 344, 396, 447, 502, 572, 596]
TP [1590, 1496, 1409, 1321, 1264, 1212, 1161, 1106, 1036, 1012]
FP [4116, 1755, 1016, 648, 467, 338, 255, 200, 148, 132]
TPR [0.9888, 0.9303, 0.8762, 0.8215, 0.7861, 0.7537, 0.722, 0.6878, 0.6443, 0.6294]
TNR [0.666, 0.8576, 0.9176, 0.9474, 0.9621, 0.9726, 0.9793, 0.9838, 0.988, 0.9893]
PPV [0.2787, 0.4602, 0.581, 0.6709, 0.7302, 0.7819, 0.8199, 0.8469, 0.875, 0.8846]
NPV [0.9978, 0.9895, 0.9827, 0.976, 0.9718, 0.968, 0.9643, 0.9602, 0.9551, 0.9534]
FPR [0.334, 0.1424, 0.0824, 0.0526, 0.0379, 0.0274, 0.0207, 0.0162, 0.012, 0.0107]
FNR [0.0112, 0.0697, 0.1238, 0.1785, 0.2139, 0.2463, 0.278, 0.3122, 0.3557, 0.3706]
FDR [0.2787, 0.4602, 0.581, 0.6709, 0.7302, 0.7819, 0.8199, 0.8469, 0.875, 0.8846]

AUC 0.9675772758923437
```

Other examples of results

	0	1	2	3	4
11793	0.0	0.0	0.0	0.0	0.0
11794	0.0	0.0	0.0	0.0	0.0
11795	0.0	0.0	0.0	0.0	0.0
11796	0.0	0.0	0.0	0.0	0.0
11797	0.0	0.0	0.0	0.0	0.0
11798	0.0	0.0	0.0	0.0	0.0
11799	0.0	0.0	0.0	0.0	0.0
11800	0.0	0.0	0.0	0.0	0.0
11801	0.0	0.0	0.0	0.0	0.0
11802	4.0	1.0	1.0	0.0	1.0
11803	0.0	0.0	0.0	0.0	0.0
11804	1.0	1.0	0.0	0.0	0.0
11805	1.0	0.0	0.0	0.0	0.0
11806	0.0	0.0	0.0	0.0	0.0
11807	0.0	0.0	0.0	0.0	0.0
11808	1.0	0.0	0.0	0.0	0.0
11809	0.0	0.0	0.0	0.0	0.0
11810	0.0	0.0	0.0	0.0	0.0
11811	0.0	0.0	0.0	0.0	0.0
11812	0.0	0.0	0.0	0.0	0.0
11813	1.0	0.0	0.0	0.0	0.0
11814	0.0	0.0	0.0	0.0	0.0
11815	1.0	0.0	0.0	0.0	0.0

	0	1	2	3	4
11794	0.083714396	8.1488333E-4	5.496421E-4	8.1905595E-4	3.4955452E-4
11795	0.8199326	0.007131189	0.0022878905	0.28128293	0.3023508
11796	0.09419119	7.071172E-4	4.8430564E-4	7.595913E-4	2.9458167E-4
11797	0.23282889	0.027997203	0.02558214	0.11785879	6.633524E-4
11798	0.055467397	5.4094114E-4	6.4544316E-4	4.946162E-4	4.8521918E-4
11799	0.07731779	0.0017017192	0.0013359981	0.0016848961	9.915272E-4
11800	0.07881349	0.0012573206	8.9498697E-4	0.0011985314	6.506785E-4
11801	0.0781834	0.0010863101	7.4060826E-4	0.0010254116	5.172088E-4
11802	0.9997677	0.99420124	0.9640394	0.06081807	0.9107287
11803	0.07158054	5.438972E-4	5.0892436E-4	0.0025030035	4.2825335E-4
11804	0.18753129	0.1597616	1.016847E-4	0.0014957279	3.6454803E-4
11805	0.27851027	6.6052744E-4	3.2303002E-4	8.63345E-4	1.520352E-4
11806	0.0076360577	0.0021975322	7.1700924E-6	2.2044066E-4	5.964592E-5
11807	0.037995726	6.7366415E-4	5.472012E-4	1.1884519E-4	1.6830143E-4
11808	0.07816413	0.0010452878	7.075307E-4	9.884735E-4	4.8796504E-4
11809	0.100373015	0.0018965065	2.1902483E-4	7.013739E-4	1.9724933E-4
11810	0.13230506	0.0021328572	4.489344E-4	0.0010254708	1.9543622E-4
11811	0.078205325	0.0010957019	7.4842223E-4	0.0010341587	5.241001E-4
11812	0.093234025	0.0018602188	0.0023249427	0.0018919639	1.9239933E-4
11813	0.1440331	6.058978E-4	6.772329E-4	3.7883522E-4	3.7465536E-4
11814	0.08448966	5.0795724E-4	2.1865884E-4	7.4196147E-4	4.1070266E-4
11815	0.13115253	0.0027069473	4.3344093E-4	0.0011535102	3.391787E-4