Status of Deep Machine Learning on $\overline{\text{P}}$ANDA Software Trigger

PANDA Collaboration Meeting 20/3, 26-30 Oct, 2020

P.Jiang[1,2 a] K.Goetzen[1] R.Kliemt[1] F.Nerling[1] K.Peters[1]

[1] GSI Helmholtz Center for Heavy Ion Research GmbH(GSI), Darmstadt, Germany

[2] Institute of Modern Physics, Chinese academy of sciences(IMP,CAS), Lanzhou, China

[a] Supported by China and Germany Postdoctoral Exchange Program

Outline

1. Introduction
2. Previous Status
3. Deep Machine Learning Methods
4. Physics Channels and Data
5. Results
6. Conclusion and Outlook
7. Acknowledgement

1 Introduction

2 Previous Status

3 Deep Machine Learning Methods

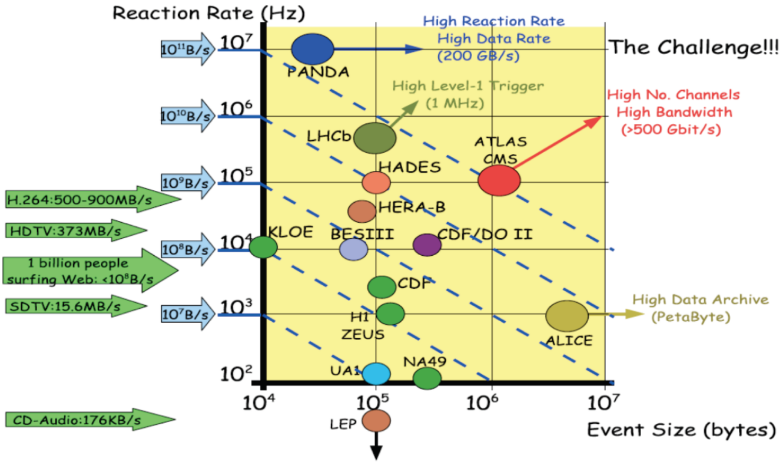
4 Physics Channels and Data

5 Results

6 Conclusion and Outlook

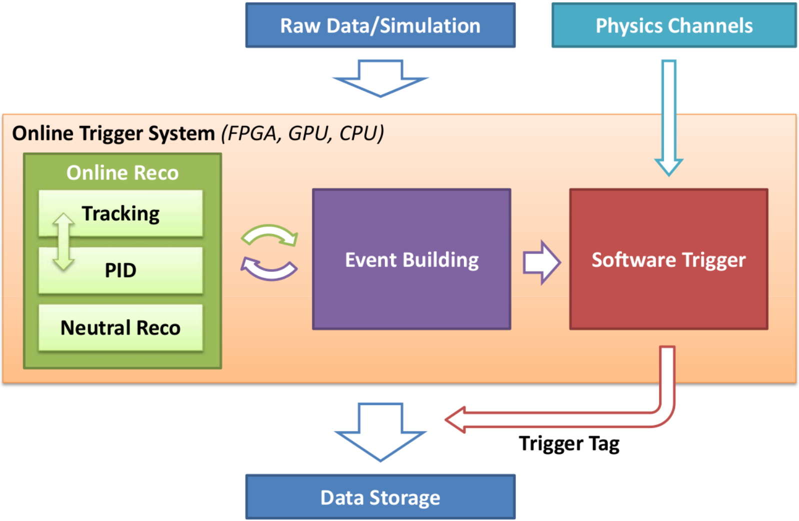
7 Acknowledgement

Motivation



Reaction Rate (Hz)
High Reaction Rate
High Data Rate
(200 GB/s)
The Challenge!!!
PANDA
High Level-1 Trigger
(1 MHz)
High No. Channels
High Bandwidth
(>500 Gbit/s)
LHCb
ATLAS
CMS
HADES
HERA-B
KLOE
BESIII
CDF/DO II
CDF
H1
ZEUS
ALICE
High Data Archive
(PetaByte)
NA49
UA1
LEP
10^{11}B/s
10^{10}B/s
10^{9}B/s
10^{8}B/s
10^{7}B/s
10^7
10^6
10^5
10^4
10^3
10^2
10^4
10^5
10^6
10^7
Event Size (bytes)
H.264:500-900MB/s
HDTV:373MB/s
1 billion people surfing Web: <10^8B/s
SDTV:15.6MB/s
CD-Audio:176KB/s

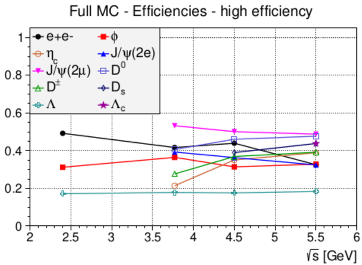
Schematic Overview

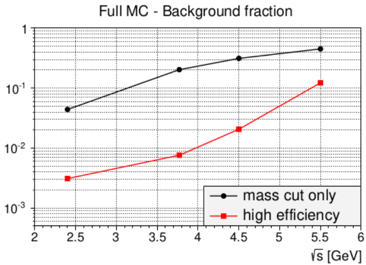


1. Introduction
2. Previous Status
3. Deep Machine Learning Methods
4. Physics Channels and Data
5. Results
6. Conclusion and Outlook
7. Acknowledgement

Conservative Results - High-Efficiency Selection

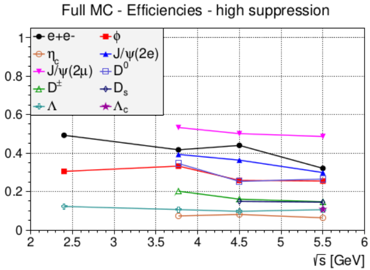
$\sqrt{s}$ (GeV)	ee	Phi	Etac	J2e	J2mu	D0	Dch	Ds	Lam	Lamc	DPM
2.4	49.21	31.20	-	-	-	-	-	-	17.16	-	0.31
	97.23	*87.59*	–	–	–	–	–	–	*86.71*	–	*7.05*
3.77	41.67	36.44	21.41	39.27	53.24	40.64	27.66	-	17.84	-	0.76
	96.19	*83.29*	*49.90*	*90.23*	*94.31*	*82.02*	*67.91*	–	*84.55*	–	*3.74*
4.5	43.95	31.41	35.07	36.31	50.07	46.06	36.90	39.00	17.58	-	2.06
	96.81	*72.32*	*66.67*	*84.62*	*90.92*	*83.76*	*72.74*	*73.12*	*81.46*	–	*6.58*
5.5	32.54	32.85	38.91	32.43	48.76	47.68	39.18	43.81	18.40	43.79	12.24
	89.54	*79.21*	*66.86*	*68.40*	*83.91*	*82.91*	*71.00*	*75.13*	*80.00*	*71.78*	*27.14*

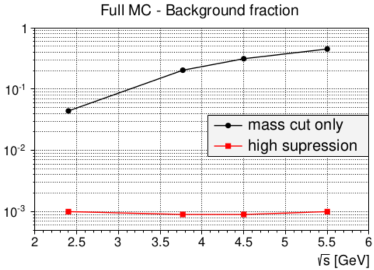




Conservative Results - High-Suppression Selection

$\sqrt{s}$ (GeV)	ee	Phi	Etac	J2e	J2mu	D0	Dch	Ds	Lam	Lamc	DPM
2.4	49.21	30.48	-	-	-	-	-	-	12.27	-	0.10
	97.23	*85.57*	–	–	–	–	–	–	*62.00*	–	*2.27*
3.77	41.67	33.21	7.35	39.26	53.23	34.55	20.19	-	10.65	-	0.09
	96.19	*75.91*	*17.13*	*90.21*	*94.30*	*69.73*	*49.57*	–	*50.47*	–	*0.44*
4.5	43.95	25.80	8.10	36.30	50.06	25.27	16.04	14.88	9.73	-	0.09
	96.81	*59.41*	*15.40*	*84.60*	*90.90*	*45.95*	*31.62*	*27.90*	*45.09*	–	*0.29*
5.5	32.05	25.47	6.37	29.85	48.59	26.51	14.68	14.51	10.60	10.81	0.10
	88.19	*61.42*	*10.95*	*62.96*	*83.62*	*46.10*	*26.60*	*24.88*	*46.09*	*17.72*	*0.22*





1 Introduction

2 Previous Status

3 Deep Machine Learning Methods

4 Physics Channels and Data

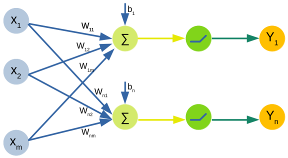
5 Results

6 Conclusion and Outlook

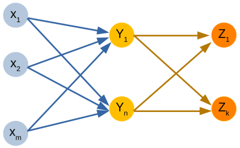
7 Acknowledgement

Neural Network

- The essence of Deep Machine Learning is Neural Network(NN).
- A NN is a fitting function.
- Linear fitting
  Given k samples
  {input m-dimensional vector x; output n-dimensional vector y},
  there is a linear fitting function $Y = w \cdot X + b$,
  where w, ***weight***, is a $n \times m$ matrix; b, ***bias***, is a n-dimensional vector.
- Nonlinearity
  In order to increase nonlinear performance, a nonlinear function, activation function $y = f(x)$, is introduced: $Y = f(w \cdot X + b)$



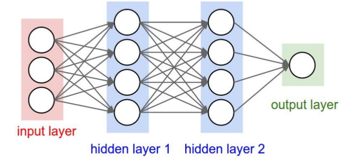
(a) Neurons



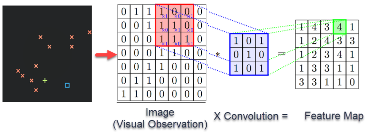
(b) Multi-layer

- Multiple layers
- Simplest NN
  $Y = w \cdot X$: Equivalent to Singular Value Decomposition(SVD)

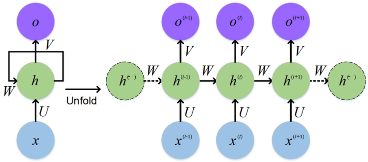
Neural Networks



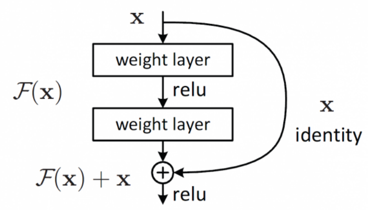
(a) DNN



(b) CNN



(c) RNN



(d) Residual network

1. Introduction
2. Previous Status
3. Deep Machine Learning Methods
4. Physics Channels and Data
5. Results
6. Conclusion and Outlook
7. Acknowledgement

Physics Channels and Data

Physics topic	Reaction channel	Code	Trigger	Tag
Electromagnetic	$p\bar{p} \to e^+e^-$	ee	$p\bar{p} \to e^+e^-$	e^+e^-
Exotics	$p\bar{p} \to \phi_{(1)}\phi_{(2)}$; $\phi_{(1)} \to$ trigger, $\phi_{(2)} \to X$	Phi	$\phi \to K^+K^-$	ϕ
Charmonium	$p\bar{p} \to \eta_c\pi^+\pi^-$; $\eta_c \to$ trigger	Etac	$\eta_c \to K_S K^-\pi^+$	η_c
	$p\bar{p} \to J/\psi\,\pi^+\pi^-$; $J/\psi \to$ trigger	J2e	$J/\psi \to e^+e^-$	$J/\psi(2e)$
	$p\bar{p} \to J/\psi\,\pi^+\pi^-$; $J/\psi \to$ trigger	J2mu	$J/\psi \to \mu^+\mu^-$	$J/\psi(2\mu)$
Open charm	$p\bar{p} \to D^0\overline{D^0}$; $D^0 \to$ trigger; $\overline{D^0} \to X$	D0	$D^0 \to K^-\pi^+$	D^0
	$p\bar{p} \to D^+D^-$; $D^+ \to$ trigger, $D^- \to X$	Dch	$D^+ \to K^-\pi^+\pi^+$	D^+
	$p\bar{p} \to D_s^+D_s^-$; $D_s^+ \to$ trigger, $D_s^- \to X$	Ds	$D_s^+ \to K^+K^-\pi^+$	D_s^+
Baryons	$p\bar{p} \to \Lambda\overline{\Lambda}$; $\Lambda \to$ trigger; $\overline{\Lambda} \to X$	Lam	$\Lambda \to p\pi^-$	Λ
	$p\bar{p} \to \Lambda_c\overline{\Lambda}_c$; $\Lambda_c \to$ trigger; $\overline{\Lambda}_c \to X$	Lamc	$\Lambda_c \to p\mathrm{K}^-\pi^+$	Λ_c
Background	$p\bar{p}$ generic (DPM)	DPM	–	–

(a) Physics channels

$\sqrt{s}$ [GeV]	$p_{\bar{p}}$ [GeV/c]	ee	Phi	Etac	J2e	J2mu	D0	Dch	Ds	Lam	Lamc	DPM
2.4	1.91	X	X	–	–	–	–	–	–	X	–	X
3.77	6.57	X	X	X	X	X	X	X	–	X	–	X
4.5	9.81	X	X	X	X	X	X	X	X	X	–	X
5.5	15.15	X	X	X	X	X	X	X	X	X	X	X

(b) Data

Observables

Short cut	Description
p	momentum p of reconstructed candidate (lab)
pt	transverse momentum p_t of reconstructed candidate
pcm	momentum p of reconstructed candidate (cms)
e	energy e of reconstructed candidate (lab)
ecm	energy e of reconstructed candidate (cms)
tht	polar angle θ of reconstructed candidate (lab)
thtcm	polar angle θ of reconstructed candidate (cms)
mmiss	missing mass of reconstructed candidate
di{p,pt,tht}	kinematic variables from i-th daughter of candidate
d$i$pidk, d$i$pidp	Kaon/Proton PID probability of i-th daughter
pmax	maximum particle momentum in event (cms)
ptmax	maximum transvers particle momentum in event
sumpc	sum of momenta of charged particles in event (cms)
sumptc	sum of transverse momenta of charged particles in event (cms)
detemcsum	sum of cluster energies in EMC
detemcmax	maximum cluster energy in EMC
lnpide	Number of loose ($P > 0.25$) electron candidates
lnpidmu	Number of loose ($P > 0.25$) muon candidates
lnpidpi	Number of loose ($P > 0.25$) pion candidates
lnpidk	Number of loose ($P > 0.25$) kaon candidates
lnpidp	Number of loose ($P > 0.25$) proton candidates
thr	Event shape: Magnitude of thrust of event (cms)
apl	Event shape: Aplanarity of event (cms)
fw1	Event shape: 1. Fox-Wolfram Moment $R_1 = H_1/H_0$ (cms)
fw2	Event shape: 2. Fox-Wolfram Moment $R_2 = H_2/H_0$ (cms)
fw3	Event shape: 3. Fox-Wolfram Moment $R_3 = H_3/H_0$ (cms)
fw4	Event shape: 4. Fox-Wolfram Moment $R_4 = H_4/H_0$ (cms)
fw5	Event shape: 5. Fox-Wolfram Moment $R_5 = H_5/H_0$ (cms)

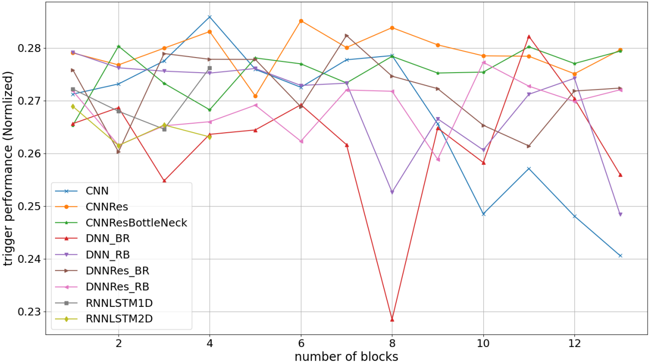
1 Introduction
2 Previous Status
3 Deep Machine Learning Methods
4 Physics Channels and Data
5 Results
6 Conclusion and Outlook
7 Acknowledgement

Multiclass Classification VS Binary Classification

Channels	Multiclass Classification[%]	Binary Classification[%]
ee	99.996	-
Phi	87.931	-
Etac	48.059	76.182
J2e	98.604	-
J2mu	99.876	99.999
D0	69.932	93.348
Dch	48.427	78.552
Ds	32.683	73.661
Lam	48.093	94.683

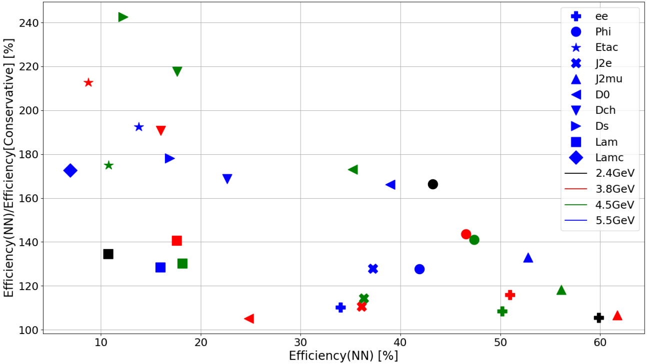
Conclusion: Used binary classification (one NN for individual physics channel) instead of multiclass classification (one NN for multiple physics channels).

Neural Network Selection



Conclusion: CNN with 4 residual blocks was the final choice.

Comparison



Conclusion: Deep machine learning optimized all physics channels from 105% to 245%. The better the performance of the channel, the smaller the gain from deep machine learning.

Normal results - High Suppression Efficiency

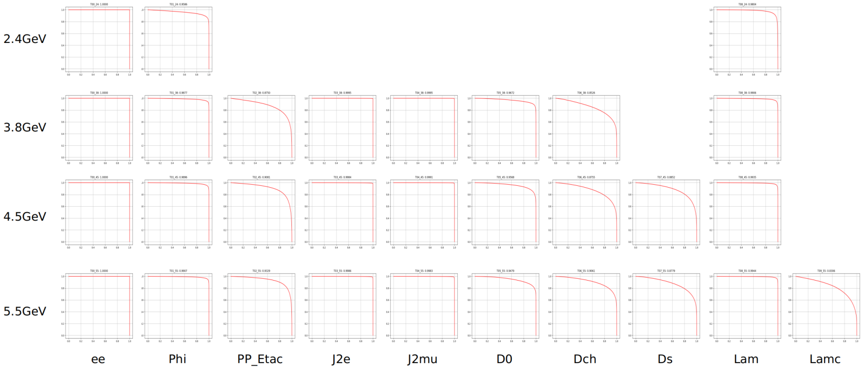
eff[%]	ee	Phi	pp_Etac	J2e	J2mu	D0	Dch	Ds	Lam	Lamc	DPM
Individual Absolute Reconstruction Efficiencies											
2.4GeV	59.86	67.85	---	---	---	---	---	---	36.31	---	1.59
3.8GeV	40.33	56.75	34.69	35.46	61.57	57.31	63.78	---	29.95	---	0.32
4.5GeV	37.26	56.50	34.73	35.73	55.47	60.58	60.27	66.74	27.20	---	0.25
5.5GeV	26.90	44.68	20.72	36.38	51.10	56.90	50.23	56.81	21.59	52.79	0.16
Individual Relative Trigger Efficiencies											
2.4GeV	99.99	73.82	---	---	---	---	---	---	31.41	---	2.10
3.8GeV	100.00	84.89	27.03	99.76	99.88	73.27	30.93	---	68.69	---	3.91
4.5GeV	100.00	83.90	24.79	99.22	99.62	62.54	31.16	20.87	74.05	---	4.39
5.5GeV	100.00	85.08	35.04	98.84	99.04	67.38	41.82	27.25	77.95	17.37	6.45
Individual Absolute Trigger Efficiencies											
2.4GeV	59.86	50.09	---	---	---	---	---	---	11.40	---	0.03
3.8GeV	40.33	48.18	9.38	35.37	61.50	41.99	19.73	---	20.57	---	0.01
4.5GeV	37.26	47.40	8.61	35.45	55.25	37.89	18.78	13.93	20.14	---	0.01
5.5GeV	26.90	38.02	7.26	35.96	50.61	38.34	21.01	15.48	16.83	9.17	0.01
Total Absolute Reconstruction Efficiencies											
2.4GeV	61.25	71.79	---	---	---	---	---	---	38.35	---	32.03
3.8GeV	57.50	73.04	80.23	74.34	81.54	80.69	83.00	---	32.66	---	38.25
4.5GeV	55.06	71.75	81.15	73.31	77.69	80.73	84.42	83.61	29.33	---	43.37
5.5GeV	37.22	63.74	75.99	70.21	74.73	77.00	79.91	79.63	22.78	73.87	44.81
Total Relative Trigger Efficiencies											
2.4GeV	97.73	69.79	---	---	---	---	---	---	30.04	---	0.31
3.8GeV	89.45	70.90	13.30	49.88	76.47	53.91	25.03	---	65.10	---	0.26
4.5GeV	91.82	70.81	16.16	50.68	73.18	49.34	25.84	18.96	71.28	---	0.23
5.5GeV	92.86	70.41	20.61	56.98	72.51	54.11	33.53	25.67	77.08	13.41	0.22
Total Absolute Trigger Efficiencies											
2.4GeV	59.86	50.10	---	---	---	---	---	---	11.52	---	0.10
3.8GeV	51.43	51.79	10.67	37.08	62.36	43.50	20.77	---	21.26	---	0.10
4.5GeV	50.56	50.81	13.12	37.15	56.85	39.83	21.82	15.86	20.91	---	0.10
5.5GeV	34.56	44.88	15.66	40.00	54.19	41.67	26.79	20.44	17.56	9.91	0.10

Individual: Cross trigger line was not taken into consideration.
Total: Cross trigger line was taken into consideration.

Absolute: triggered events / simulated events
Relative: triggered events / reconstructed events

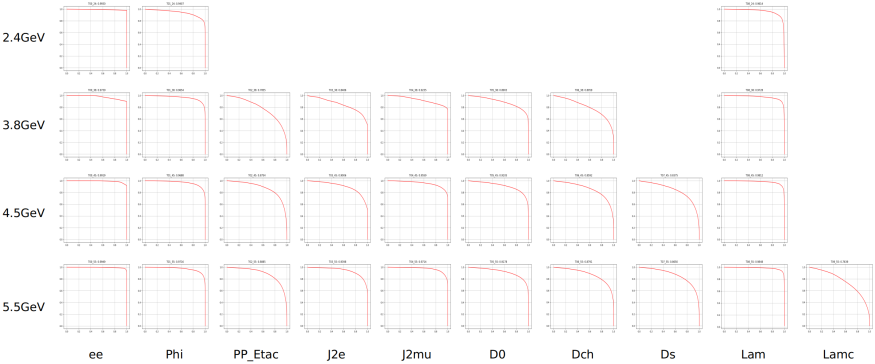
Reconstruction: Reconstructed events before trigger.
Trigger: Triggered events after trigger.

Normal results - ROCs - Individual



Conclusion: AUCs of ”easier” channels were almost 1.

Normal results - ROCs - Total



Conclusion: It is the result of multiple neural networks acting on multiple datasets.

Normal - Observable Importance Ranking

ee | Phi | ... | Lam | Lamc

ee branch	ee proportion	Phi branch	Phi proportion	Lam branch	Lam proportion	Lamc branch	Lamc proportion
xd1emcecal	0.034372	xpt	0.023953	xpt	0.022313	xpcm	0.014811
esprapmax	0.031086	esfw2	0.020678	esnchrg	0.019366	xd0tht	0.014777
esfw4	0.029378	xd1sttdedx	0.019516	xd0sttdedx	0.019308	xd1pt	0.014674
xd1tofbeta	0.027659	essumenl	0.019145	mmiss	0.019065	essumetn	0.014408
xd1stthits	0.027595	xd1stthits	0.018924	xtht	0.019063	xd0pt	0.014032
xd0emcnx	0.025817	esnchrg	0.018905	essumen	0.018288	xpt	0.013983
xd0emcecal	0.025039	xd1mvdhits	0.018883	essumetnl	0.018153	xecm	0.013976
esnchrg	0.024598	xd0sttdedx	0.018726	xd1pt	0.017902	xd0stthits	0.013649
esnpart	0.023382	xd0pt	0.018370	xd0stthits	0.017719	xd1pidk	0.013423
xd1muonlay	0.023113	xd1tht	0.018345	xd0mvdhits	0.017715	essumenl	0.013362
mmiss	0.023036	mmiss	0.018285	essumenl	0.017313	xd1tht	0.013054
esdetemcmax	0.022775	xd0tht	0.018086	xd1mvdhits	0.017247	xd2pt	0.012903
xd0tofbeta	0.022118	xd0mvdhits	0.017929	esfw4	0.017125	eslnpidk	0.012765
xd0stthits	0.021728	xd1drcthtc	0.017617	xd0pt	0.017110	esnchrg	0.012661
xthtcm	0.021585	xtht	0.017285	esthr	0.016890	xd1tofbeta	0.012103
esapl	0.021350	xd0stthits	0.017092	espla	0.016777	xd0mvdhits	0.012089
eslnpide	0.021106	essumetnl	0.016752	xd0tht	0.016091	xd1stthits	0.011994
xd0muonlay	0.020605	essumen	0.016698	esfw1	0.015982	essumen	0.011810
xd0pidpi	0.020540	xd0drcthtc	0.016664	xd0mvddedx	0.015362	xd2tht	0.011774
escir	0.020500	esprapmax	0.016491	xd0drcthtc	0.015228	xd1p	0.011665
espmin	0.020301	xd1mvddedx	0.016390	xd0emcecal	0.014805	xd0dscthtc	0.011639
xd1pidpi	0.020027	xd0pidk	0.016285	esfw3	0.014672	xd1pidp	0.011631
esl1npide	0.019869	xd1tofbeta	0.016045	xd1sttdedx	0.014648	esfw1	0.011628
xd0dscthtc	0.019630	xd0tofbeta	0.015630	xd0tofbeta	0.014521	xd1dscthtc	0.011356
esfw1	0.019074	xd1emcecal	0.015122	esnneut	0.014252	xd1drcthtc	0.011336
essumetnl	0.018277	esfw5	0.015054	esprapmax	0.013959	xd0tofbeta	0.011321

Conclusion: Some observables were chosen based on this table for dimensionality reduction.

1 Introduction

2 Previous Status

3 Deep Machine Learning Methods

4 Physics Channels and Data

5 Results

6 Conclusion and Outlook

7 Acknowledgement

Conclusion and Outlook

- High-suppression selection was used for software trigger according to previous study.
- Binary classification was used instead of multiclass classification.
- Nine NN structures were compared and CNN with 4 residual blocks was the final choice.
- Deep machine learning optimized all physics channels from 105% to 245%.
- The final high-suppression efficiency was obtained.
- Individual and total ROCs were plotted.
- Observable importance ranking was calculated.

- Is it possible to add physics in?
- How to optimize the NN parameters slightly based on real experimental data?

1 Introduction

2 Previous Status

3 Deep Machine Learning Methods

4 Physics Channels and Data

5 Results

6 Conclusion and Outlook

7 Acknowledgement

Acknowledgement

Many thanks to Dr. Y. Liang for the explaination of high energy physics and particle detectors.

Many thanks to Dr. M. Kunze for importance ranking.

Many thanks for your attention!