Lorentz Boost Network

Richard Barton

Abstract

- Utilizing both high-level and low-level variables has been shown to improve the performance of NN in particle physics applications. The Lorentz Boost Network was developed with the intention to autonomize the process of finding suitable variables that describe the main characteristics of a particle physics task. To test it performance, the LBN will be inserted at the front end of an existing NN being used for the $H^+ \to \tau\nu$ analysis.

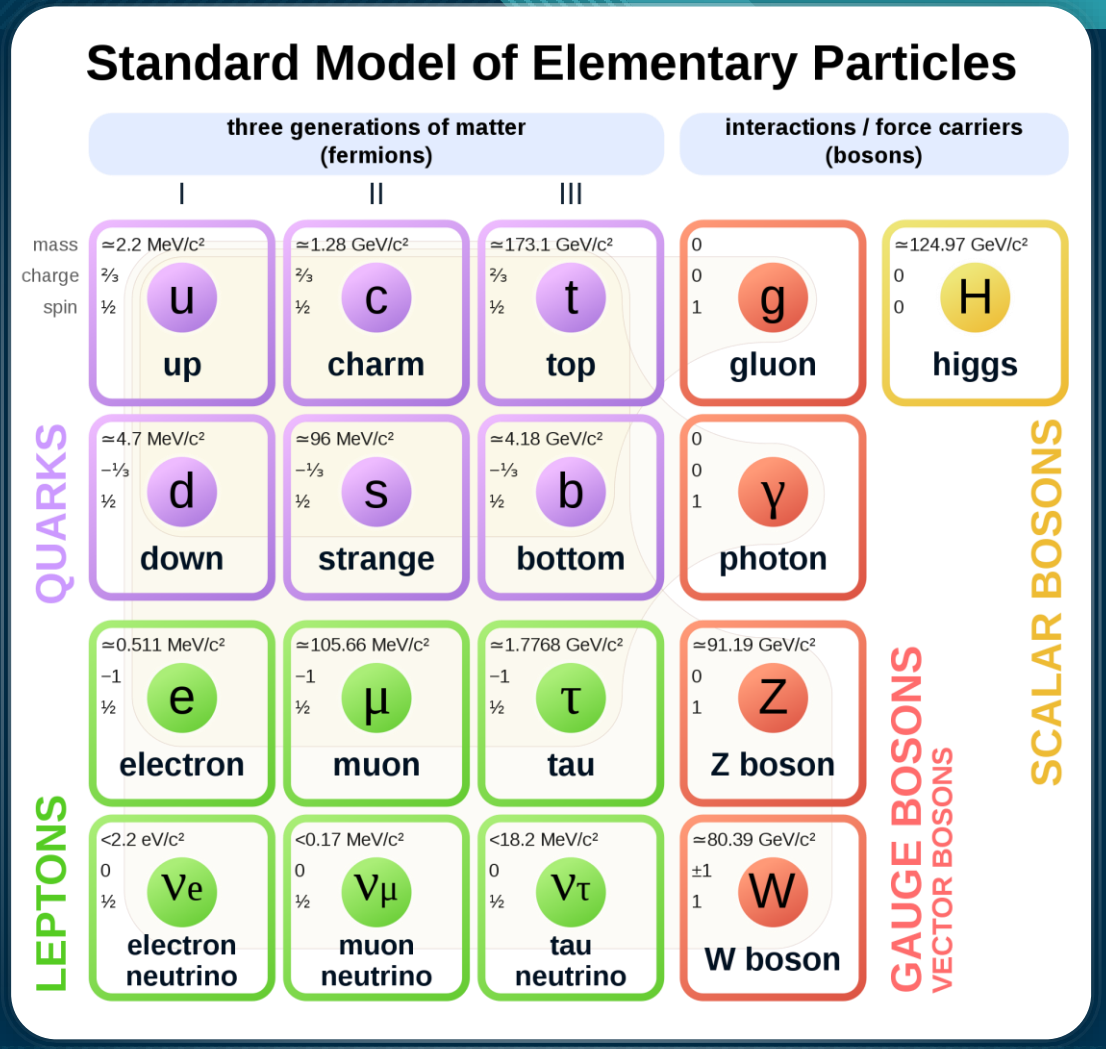
Introduction

Fermions

- Generations
- Matter
- ½ integer spin
- Pauli Exclusion Principle
  - Unique (n, ℓ, m_ℓ, m_s)

Bosons

- Force
- Integer Spin
- Can occupy same quantum state



Introduction

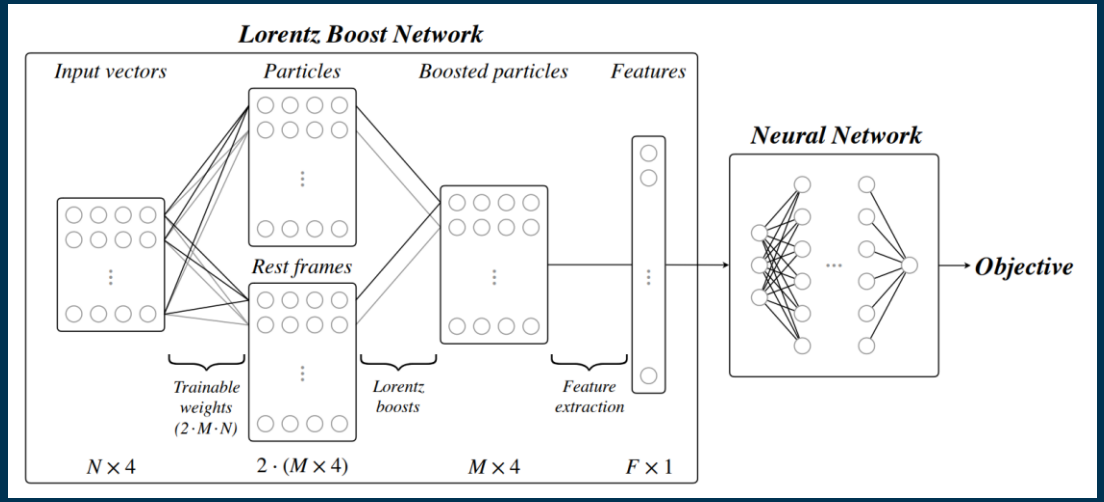
Beyond Standard Model

- Dark matter candidates
- Hierarchy problem
  - Comparative strengths of the fundamental forces
- Gravity (Graviton G)
- Unification of forces

Looking for new particles

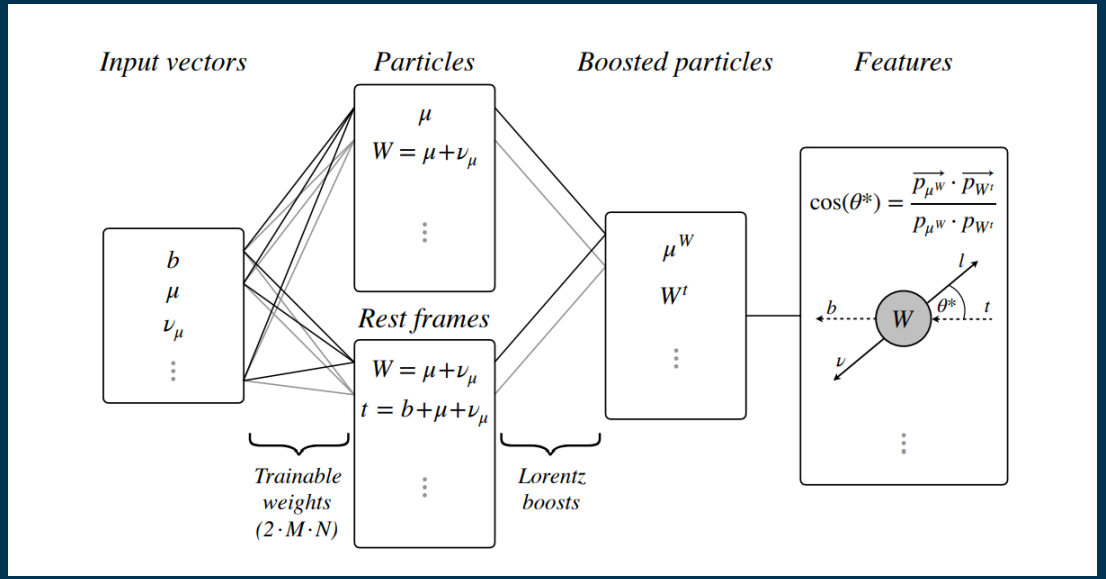
Motivation

- In HEP many particles are short lived and decay before they can reach the detector and must be found indirectly.
- Traditionally high-level features were fed into a neural network.
- Many studies have shown neural networks perform better with both high- and low-level features being used.
- The LBN uses these low-level variables to generate a set of features that can be fed into a Neural Network to solve specific physics tasks.



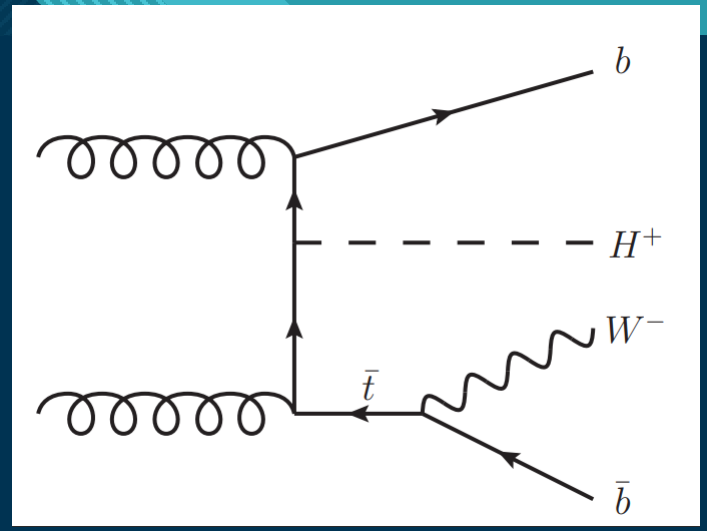
LBN

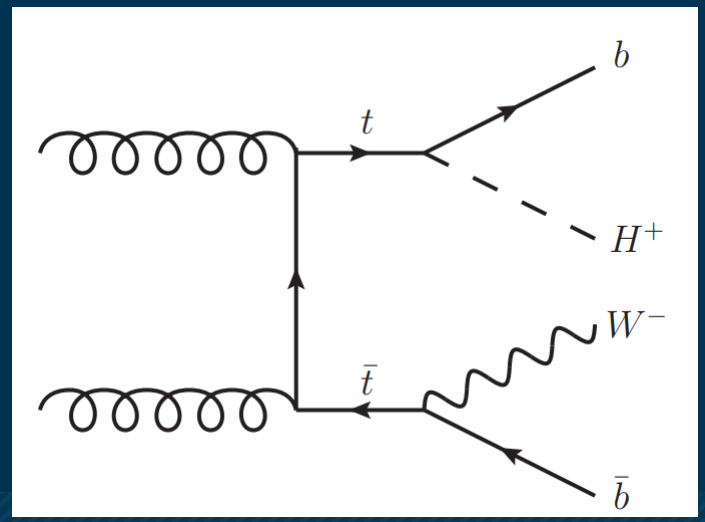
- The LBN takes in low-level variables in the form of particle 4-momenta vectors
  - (E, P_x, P_y, P_z)
- The LBN makes linear combinations from the 4-momenta vectors
  - Combined Particle
  - Combined Rest frames
- The combined particles are boosted to the rest frames
- Features are then calculated from the boosted particles
  - Single features
    - Mass
    - Transverse momentum
    - Pseudorapidity
  - Multiparticle features
    - Cos of special angular difference
    - Distance in η-ϕ plane
    - Distance in Minkowski space
    - Invariant quantities measured from combined restframe



Overview

- Definition of the tasks / challenge
  - The task is to separate Charged Higgs to Tau Nu signal events from background events such as TTbar, QCD, SingleTop, Wtaunu, Ztautau, and DiBoson.
- Approach
  - The approach is to take an existing neural network, insert the LBN layer, and see if the performance of the network noticeably improves.
- Summary of the performance achieved
  - The network has yet to be tested on an appropriately sized data set to show any significant performance improvements over the original HPANA model.





Preprocessing / Cleanup

- Access to the data used for the HPANA neural network required CERN grid space certification.
- Symlinks were created to access the ntuples on the CERN grid space.
- The HPANA code used to run the network (train-classifier.py) generates a pickle file that includes a dataframe of the requested signal and background data points. A subset of this data has been saved to PtEtaPhiE.csv for conversion.
- For the initial setup, the signal events only include the Hplus200 mass point, and the background events only includes DiBoson.
- The required 4-momenta vectors for the LBN layer were produced by converting the PtEtaPhiE vectors to EPxPyPz vectors in ROOT using the LBN_Vector_Conversion.py script.

```
file_name = "./TRAIN_DATA.pkl"
objects = pd.read_pickle(file_name)
objects
```

		event_number	tau_0_n_charged_tracks	tau_0_pt	tau_0_eta	tau_0_phi	tau_0_E	lep_0_pt	lep_0_eta	lep_0_phi	lep_0_E	...
DiBoson	0	51501305	1	61.196223	0.597434	-1.977407	72.446254	301.454469	0.172820	-1.838867	-1429.018434	...
	1	51840911	3	50.513039	0.735092	0.232820	64.786385	64.118320	2.174325	1.505945	193.673980	...
	2	51843660	1	65.117453	0.790218	-1.260932	86.528807	53.415461	0.405626	-2.715441	-82.565721	...
	3	51843204	1	66.388078	1.580813	1.166146	168.118262	90.245320	1.002049	-0.213378	-172.916212	...
	4	51742600	3	41.767187	0.771101	-2.810712	54.812115	87.059977	-0.170727	0.485725	14.393968	...
...	...	...	...	...	...	...	...	...	...	...	...	...
Hplus3000	3389	119212	1	596.653438	0.126293	-1.263547	601.418059	64.177766	-0.198520	2.394535	-90.250865	...
	3390	110937	1	258.894719	2.105565	2.336684	1078.752247	221.454000	0.973893	-2.148421	-85.222468	...
	3391	119615	1	621.742500	-0.437455	-1.300927	682.187816	41.723176	-0.124043	-0.690370	-143.414827	...
	3392	110288	1	1191.337625	-0.845155	0.362451	1642.753473	89.843414	0.188039	-0.349078	-12.475519	...
	3393	119768	3	752.355750	0.671846	-2.122138	928.637679	51.295402	-2.115407	-1.871007	-266.834575	...

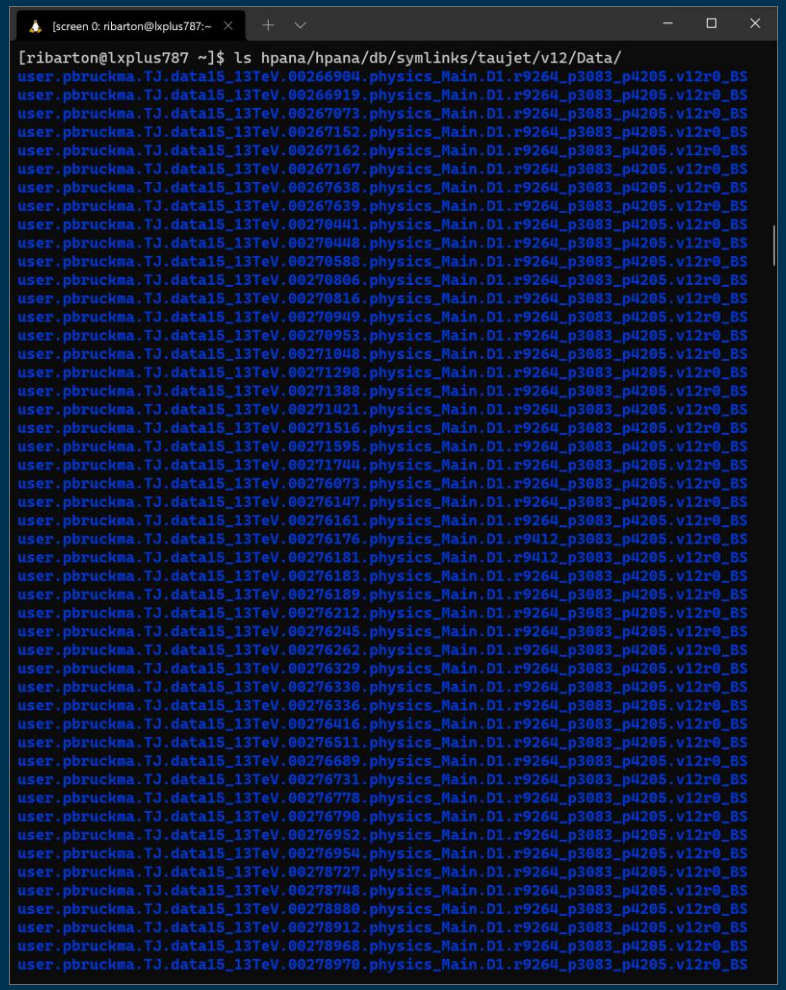
194347 rows × 24 columns

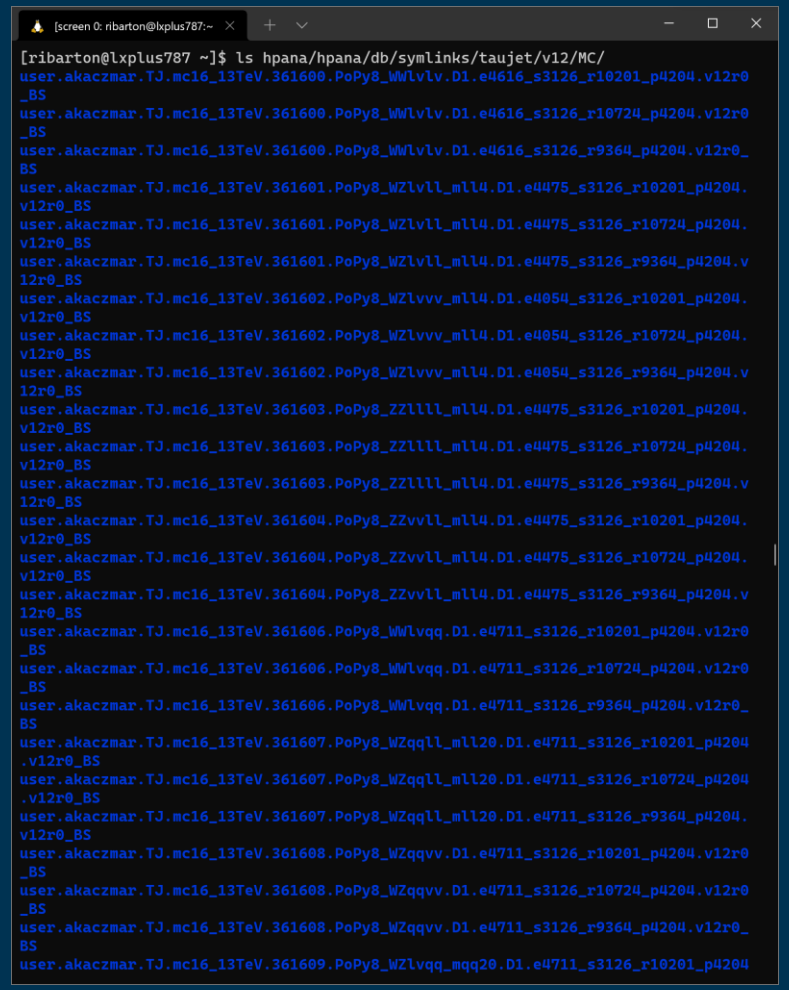
File structure

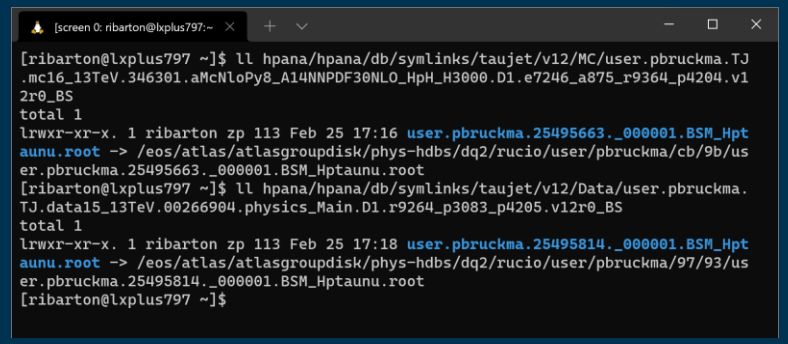
```
lsetup                  lsetup <tool1> [ <tool2> ...] (see lsetup -h):
 lsetup asetup           (or asetup) to setup an Athena release
 lsetup atlantis         Atlantis: event display
 lsetup eiclient         Event Index
 lsetup emi              EMI: grid middleware user interface
 lsetup ganga            Ganga: job definition and management client
 lsetup lcgenv           lcgenv: setup tools from cvmfs SFT repository
 lsetup panda            Panda: Production ANd Distributed Analysis
 lsetup pyami            pyAMI: ATLAS Metadata Interface python client
 lsetup root             ROOT data processing framework
 lsetup rucio            distributed data management system client
 lsetup views            Set up a full LCG release
 lsetup xcache           XRootD local proxy cache
 lsetup xrootd           XRootD data access
advancedTools           advanced tools menu
diagnostics             diagnostic tools menu
helpMe                  more help
printMenu               show this menu
showVersions            show versions of installed software

[rabarton1988@master ~]$ ssh ribarton@lxplus.cern.ch
Password:
Last login: Mon Feb 21 16:27:15 2022 from cpe-70-119-176-2.tx.res.rr.com
* ****************************************************************
* Welcome to lxplus787.cern.ch, CentOS Linux release 7.9.2009 (Core)
* Archive of news is available in /etc/motd-archive
* Reminder: you have agreed to the CERN
*   computing rules, in particular OC5. CERN implements
*   the measures necessary to ensure compliance.
*   https://cern.ch/ComputingRules
* Puppet environment: qa, Roger state: production
* Foreman hostgroup: lxplus/nodes/login
* Availability zone: cern-geneva-b
* LXPLUS Public Login Service - http://lxplusdoc.web.cern.ch/
* A CS8 based lxplus8.cern.ch is now available
* Please read LXPLUS Privacy Notice in http://cern.ch/go/TpV7
* ****************************************************************
[ribarton@lxplus787 ~]$
```

```
[ribarton@lxplus787 ~]$ ls
hpana  myCertificate.p12  private  public  PythonPackags
[ribarton@lxplus787 ~]$ ls hpana/
aux  bin  hpana  notebooks  README.md  requirements.txt  setup.sh  Test_Run
[ribarton@lxplus787 ~]$ ls hpana/hpana/
analysis.py     containers.py      lumi.pyc          trigger.py
analysis.pyc    containers.pyc     mem_branches.py   trigger.pyc
categories.py   cxxmacros          mem_branches.pyc  utils.py
categories.pyc  dataset_hists.py   mva               utils.pyc
cluster         dataset_hists.pyc  plotting          variables.py
cmd.py          db                 rqcd.py           variables.pyc
cmd.pyc         __init__.py        samples           weights.py
config.py       __init__.pyc       systematics.py    weights.pyc
config.pyc      lumi.py            systematics.pyc
[ribarton@lxplus787 ~]$ ls hpana/hpana/db/
DataBase                     decorators.py           samples
datasets_config_cedar.yml    decorators.pyc          symlinks
datasets_config_rnnTest.yml  grid_datasets.py        xsec
datasets_config_v09.yml      grid_datasets.pyc       yaml_utils.py
datasets_config.yml          __init__.py             yaml_utils.pyc
datasets.py                  __init__.pyc
datasets.pyc                 old_datasets_config.yml
[ribarton@lxplus787 ~]$ ls hpana/hpana/db/symlinks/
taujet  taulep
[ribarton@lxplus787 ~]$ ls hpana/hpana/db/symlinks/taujet/
v12
[ribarton@lxplus787 ~]$ ls hpana/hpana/db/symlinks/taujet/v12/
Data  MC
[ribarton@lxplus787 ~]$
```

File structure



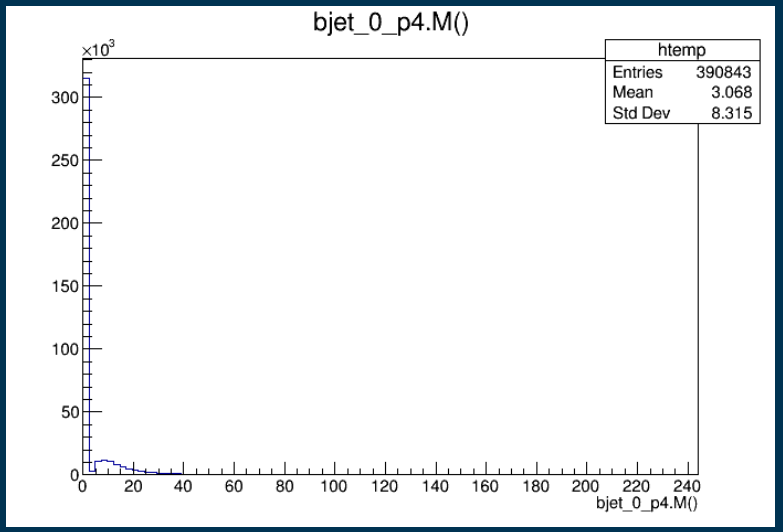




File structure

```
[ribarton@lxplus797 ~]$ ls hpana/hpana/db/symlinks/taujet/v12/Data/user.pbruckma.
TJ.data18_13TeV.00364292.physics_Main.D1.f1002_m2037_p4205.v12r0_BS/
user.pbruckma.25498539._000001.BSM_Hptaunu.root
user.pbruckma.25498539._000002.BSM_Hptaunu.root
[ribarton@lxplus797 ~]$ root -b hpana/hpana/db/symlinks/taujet/v12/Data/user.pbru
ckma.TJ.data18_13TeV.00364292.physics_Main.D1.f1002_m2037_p4205.v12r0_BS/user.pbr
uckma.25498539._000001.BSM_Hptaunu.root
   ------------------------------------------------------------------
  | Welcome to ROOT 6.24/06                        https://root.cern |
  | (c) 1995-2021, The ROOT Team; conception: R. Brun, F. Rademakers |
  | Built for linuxx8664gcc on Sep 02 2021, 14:20:23                 |
  | From tags/v6-24-06@v6-24-06                                      |
  | With c++ (GCC) 4.8.5 20150623 (Red Hat 4.8.5-44)                 |
  | Try '.help', '.demo', '.license', '.credits', '.quit'/'.q'       |
   ------------------------------------------------------------------

root [0]
Attaching file hpana/hpana/db/symlinks/taujet/v12/Data/user.pbruckma.TJ.data18_13
TeV.00364292.physics_Main.D1.f1002_m2037_p4205.v12r0_BS/user.pbruckma.25498539._0
00001.BSM_Hptaunu.root as _file0...
(TFile *) 0x2b5efa0
root [1] .ls
TFile**         hpana/hpana/db/symlinks/taujet/v12/Data/user.pbruckma.TJ.data18_1
3TeV.00364292.physics_Main.D1.f1002_m2037_p4205.v12r0_BS/user.pbruckma.25498539._
000001.BSM_Hptaunu.root
 TFile*         hpana/hpana/db/symlinks/taujet/v12/Data/user.pbruckma.TJ.data18_1
3TeV.00364292.physics_Main.D1.f1002_m2037_p4205.v12r0_BS/user.pbruckma.25498539._
000001.BSM_Hptaunu.root
  KEY: TObjString       baseTreeName;1  Collectable string class
  KEY: TObjString       nominalTreeName;1       Collectable string class
  KEY: TObjString       metadataDirName;1       Collectable string class
  KEY: TDirectoryFile   xCompression;1  xCompression
  KEY: TTree    NOMINAL;1       NOMINAL
  KEY: TH1D     cutflow_muon_NOMINAL;1  cutflow_muon_NOMINAL
  KEY: TH1D     cutflow_ele_NOMINAL;1   cutflow_ele_NOMINAL
  KEY: TH1D     cutflow_pho_NOMINAL;1   cutflow_pho_NOMINAL
  KEY: TH1D     cutflow_tau_NOMINAL;1   cutflow_tau_NOMINAL
  KEY: TH1D     cutflow_jet_NOMINAL;1   cutflow_jet_NOMINAL
  KEY: TH1D     h_metadata;1
  KEY: TH1D     h_metadata_theory_weights;1
root [2] NOMINAL->Print()
******************************************************************************
*Tree    :NOMINAL   : NOMINAL                                                *
*Entries :   390843 : Total =       628841864 bytes  File  Size =  202581797 *
*        :          : Tree compression factor =   3.10                       *
******************************************************************************
*Br    0 :HLT_3j175 : HLT_3j175/i                                            *
*Entries :   390843 : Total  Size=    1565079 bytes  File Size  =       8381 *
*Baskets :       14 : Basket Size=     251392 bytes  Compression= 186.67     *
*.............................................................................*
```

```
*         | tau_0_trig_HLT_tau160_mediumRNN_tracktwoMVA_L1TAU100/i          *
*Entries :   390843 : Total  Size=    1565853 bytes  File Size  =      69315 *
*Baskets :       14 : Basket Size=     251904 bytes  Compression=  22.58     *
*.............................................................................*
*Br  201 :tau_0_trig_HLT_tau200_medium1_tracktwoEF_L1TAU100 :                 *
*         | tau_0_trig_HLT_tau200_medium1_tracktwoEF_L1TAU100/i              *
*Entries :   390843 : Total  Size=    1565799 bytes  File Size  =      64603 *
*Baskets :       14 : Basket Size=     251904 bytes  Compression=  24.23     *
*.............................................................................*
*Br  202 :tau_0_trig_HLT_tau200_mediumRNN_tracktwoMVA_L1TAU100 :             *
*         | tau_0_trig_HLT_tau200_mediumRNN_tracktwoMVA_L1TAU100/i          *
*Entries :   390843 : Total  Size=    1565853 bytes  File Size  =      65604 *
*Baskets :       14 : Basket Size=     251904 bytes  Compression=  23.86     *
*.............................................................................*
*Br  203 :tau_0_trig_HLT_tau80_medium1_tracktwo :                            *
*         | tau_0_trig_HLT_tau80_medium1_tracktwo/i                         *
*Entries :   390843 : Total  Size=    1565583 bytes  File Size  =       8772 *
*Baskets :       14 : Basket Size=     251392 bytes  Compression= 178.39     *
*.............................................................................*
*Br  204 :tau_0_trig_HLT_tau80_medium1_tracktwo_L1TAU60 :                    *
*         | tau_0_trig_HLT_tau80_medium1_tracktwo_L1TAU60/i                 *
*Entries :   390843 : Total  Size=    1565727 bytes  File Size  =       8886 *
*Baskets :       14 : Basket Size=     251904 bytes  Compression= 176.12     *
*.............................................................................*
*Br  205 :tau_0_trig_trigger_matched : tau_0_trig_trigger_matched/i          *
*Entries :   390843 : Total  Size=    1565385 bytes  File Size  =      79094 *
*Baskets :       14 : Basket Size=     251392 bytes  Compression=  19.78     *
*.............................................................................*
*Br  206 :tau_0_type : tau_0_type/I                                          *
*Entries :   390843 : Total  Size=    1565097 bytes  File Size  =       8395 *
*Baskets :       14 : Basket Size=     251392 bytes  Compression= 186.36     *
*.............................................................................*
*Br  207 :tau_0_upsilon_energy_based : tau_0_upsilon_energy_based/F          *
*Entries :   390843 : Total  Size=    1565385 bytes  File Size  =    1169913 *
*Baskets :       14 : Basket Size=     251392 bytes  Compression=   1.34     *
*.............................................................................*
*Br  208 :tau_0_upsilon_pt_based : tau_0_upsilon_pt_based/F                  *
*Entries :   390843 : Total  Size=    1565313 bytes  File Size  =    1425695 *
*Baskets :       14 : Basket Size=     251392 bytes  Compression=   1.10     *
*.............................................................................*
*Br  209 :useEvent  : useEvent/i                                             *
*Entries :   390843 : Total  Size=    1565061 bytes  File Size  =       9175 *
*Baskets :       14 : Basket Size=     251392 bytes  Compression= 170.51     *
*.............................................................................*
root [3] NOMINAL->Draw("bjet_0_p4.M()")
Info in <TCanvas::MakeDefCanvas>:  created default TCanvas with name c1
root [4] c1->Print("/afs/cern.ch/user/r/ribarton/hpana/bjet_0_p4_M.png")
Info in <TCanvas::Print>: png file /afs/cern.ch/user/r/ribarton/hpana/bjet_0_p4_M
.png has been created
root [5]
```



Datasets

```
Channels: ['taujet', 'taulep’]

Data Streams: ['2015', '2016', '2017', '2018']

Signals: ['Hplus80', 'Hplus90', 'Hplus100', 'Hplus110', 'Hplus120', 'Hplus130’,
'Hplus140', 'Hplus150', 'Hplus160', 'Hplus170', 'Hplus180’,
'Hplus190','Hplus200’, 'Hplus225', 'Hplus250', 'Hplus275', 'Hplus300',
'Hplus350', 'Hplus400', 'Hplus500’, 'Hplus600', 'Hplus700', 'Hplus800',
'Hplus900', 'Hplus1000', 'Hplus1200’, 'Hplus1400', 'Hplus1600', 'Hplus1800',
'Hplus2000', 'Hplus2500', 'Hplus3000’]

Backgrounds: ['TTbar', 'QCD', 'SingleTop', 'Wtaunu', 'Ztautau', 'DiBoson']
```

Datasets

```
train-classifier --channel taulep --data-streams 2018 \
--db-version v12 --train-nn --bin-scheme ALL --train-data TRAIN_DATA.pkl \
--parallel --ncpu 1 --bkg DiBoson --outdir myOutDirClf
```

Vector Conversion

- The LBN require 4-momenta vectors as inputs.
- LBN_Vector_Conversion.py script takes PtEtaPhiE.csv and outputs EPxPyPz.csv for LBN training.

```
for i in range(len(PtEtaPhiE)):

    print("Event: {}".format(i))
    print('PtEtaPhiE[i][0]: {}'.format((PtEtaPhiE[i][0])))

    tau_0 = ROOT.TLorentzVector()
    lep_0 = ROOT.TLorentzVector()
    bjet_0 = ROOT.TLorentzVector()
    met = ROOT.TLorentzVector()
    jet_0 = ROOT.TLorentzVector()
    jet_1 = ROOT.TLorentzVector()

    tau_0.SetPtEtaPhiE(PtEtaPhiE[i][0],PtEtaPhiE[i][1],PtEtaPhiE[i][2],PtEtaPhiE[i][3])
    lep_0.SetPtEtaPhiE(PtEtaPhiE[i][4],PtEtaPhiE[i][5],PtEtaPhiE[i][6],PtEtaPhiE[i][7])
    bjet_0.SetPtEtaPhiE(PtEtaPhiE[i][8],PtEtaPhiE[i][9],PtEtaPhiE[i][10],PtEtaPhiE[i][11])
    met.SetPtEtaPhiE(PtEtaPhiE[i][12],PtEtaPhiE[i][13],PtEtaPhiE[i][14],PtEtaPhiE[i][15])
    jet_0.SetPtEtaPhiE(PtEtaPhiE[i][16],PtEtaPhiE[i][17],PtEtaPhiE[i][18],PtEtaPhiE[i][19])
    jet_1.SetPtEtaPhiE(PtEtaPhiE[i][20],PtEtaPhiE[i][21],PtEtaPhiE[i][22],PtEtaPhiE[i][23])

    EPxPyPz.append([tau_0.E(),tau_0.Px(),tau_0.Py(),tau_0.Pz(), \
                    lep_0.E(),lep_0.Px(),lep_0.Py(),lep_0.Pz(), \
                    bjet_0.E(),bjet_0.Px(),bjet_0.Py(),bjet_0.Pz(), \
                    met.E(),met.Px(),met.Py(),met.Pz(), \
                    jet_0.E(),jet_0.Px(),jet_0.Py(),jet_0.Pz(), \
                    jet_1.E(),jet_1.Px(),jet_1.Py(),jet_1.Pz(), \
                    PtEtaPhiE[i][24]])
```

Data

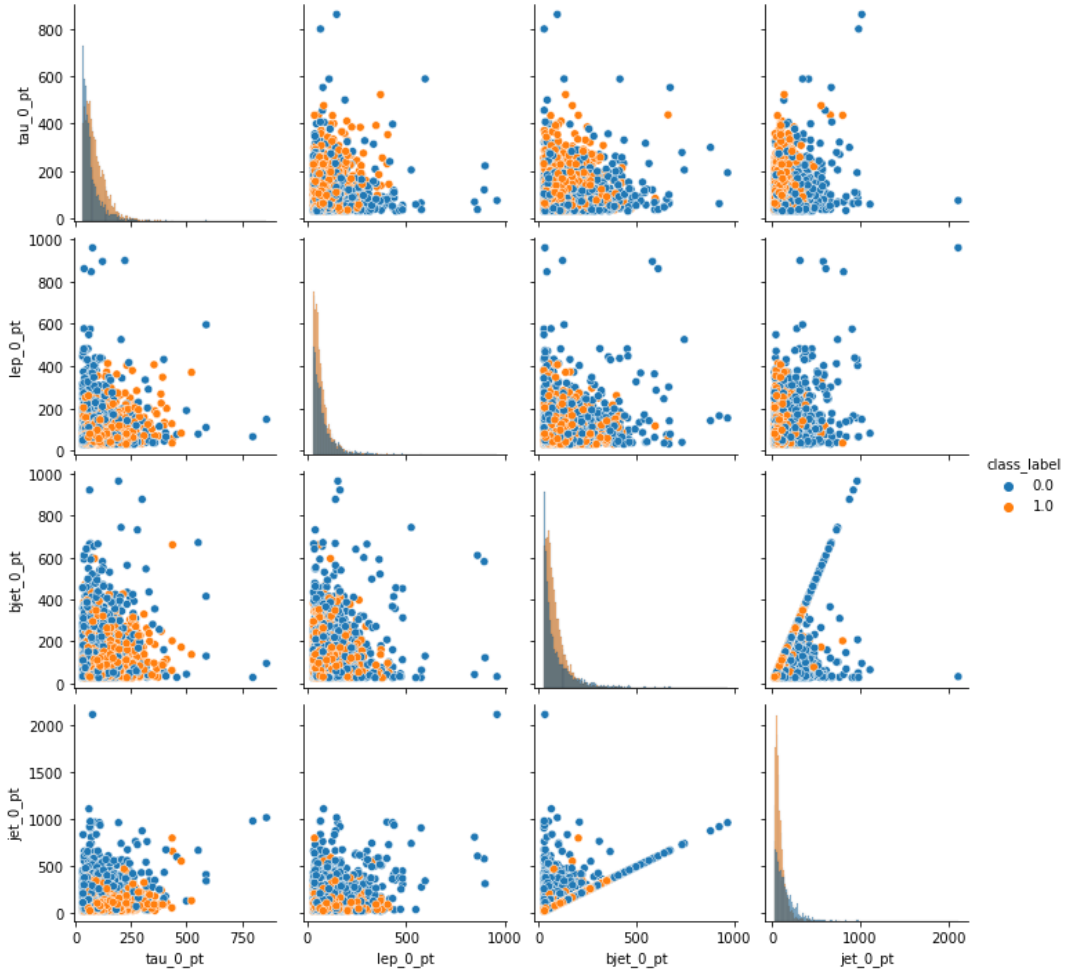
- Type: Ntuples taken from Monte Carlo signal/background and run data.
  - HPANA Input: Data frame of particle events including high- and low-level variables.
  - LBN Input: CSV of particle 4-momenta
  - Output: signal/background
- Size: 7433 events
- Instances (Train, Test, Validation Split):
  - Training: 5000 events
  - Testing: 2433 events

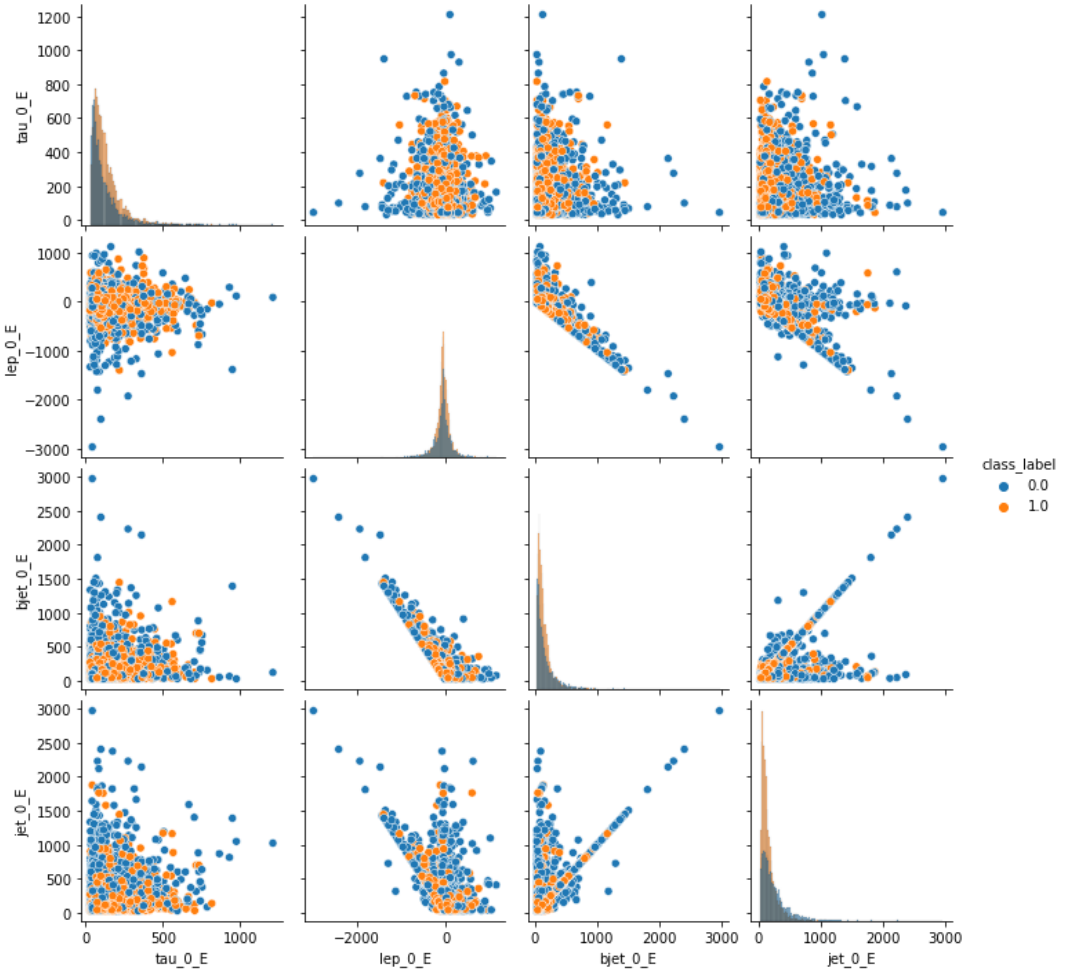
```
df2 = pd.read_csv('EPxPyPz.csv', header=None, names=df2variables)
df2
```

	tau_0_E	tau_0_Px	tau_0_Py	tau_0_Pz	lep_0_E	lep_0_Px	lep_0_Py	lep_0_Pz	bjet_0_E	bjet_0_Px	...
0	36.553666	-29.165498	-19.011040	-11.141122	21.390974	3.425496	47.384285	-35.876886	38.141930	-1.223022	...
1	117.187997	89.148620	-41.041988	64.039874	-54.726424	-72.815231	-18.457950	-7.763543	130.244883	-53.356551	...
2	243.011801	156.880646	-27.624827	183.521299	-204.676655	-41.323151	18.359706	-39.405644	204.676655	-13.014949	...
3	145.974976	-40.080426	70.531011	-121.357445	-118.946446	73.472068	-61.114372	-71.763554	238.458496	45.939602	...
4	130.586211	26.636359	-69.311486	107.420579	-64.557645	-30.880374	-7.971840	177.343715	64.557645	-55.315097	...
...	...	...	...	...	...	...	...	...	...	...	...
7428	113.056460	-25.976380	35.579702	-104.120486	-69.770719	-57.552105	-4.063886	-36.646056	138.120587	40.151658	...
7429	35.126350	-33.506106	-6.179581	8.544828	276.658221	-88.029132	137.653409	-294.789010	60.385041	-21.113074	...
7430	140.716003	73.434103	-4.622980	119.946047	-226.236410	-14.997524	-26.571417	-8.615940	226.236410	25.637152	...
7431	59.333343	-14.377494	-29.145287	49.641571	-77.785994	-49.018382	15.250505	51.416802	77.785994	42.396560	...
7432	301.766230	-59.060368	161.136978	248.212822	-459.321221	-66.102542	-62.129010	123.684405	612.707672	7.699276	...

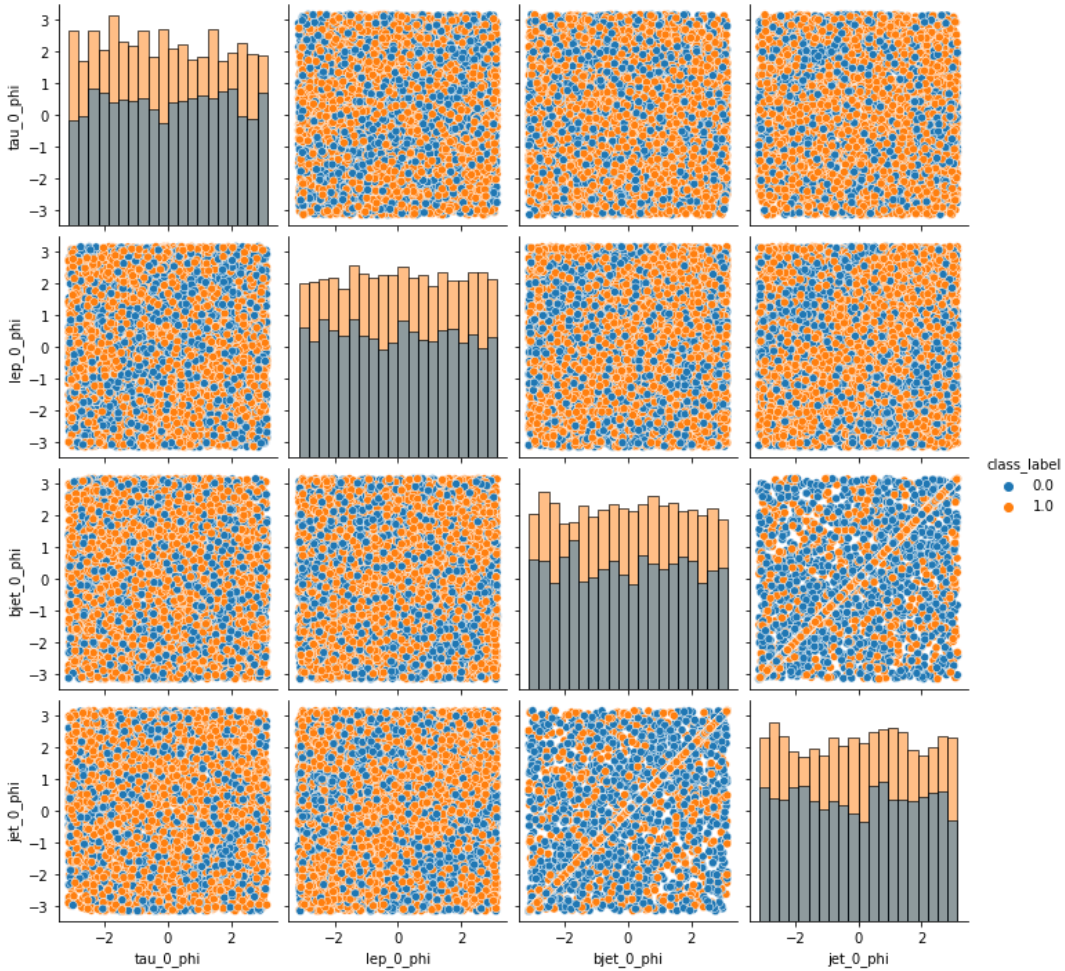
7433 rows × 25 columns

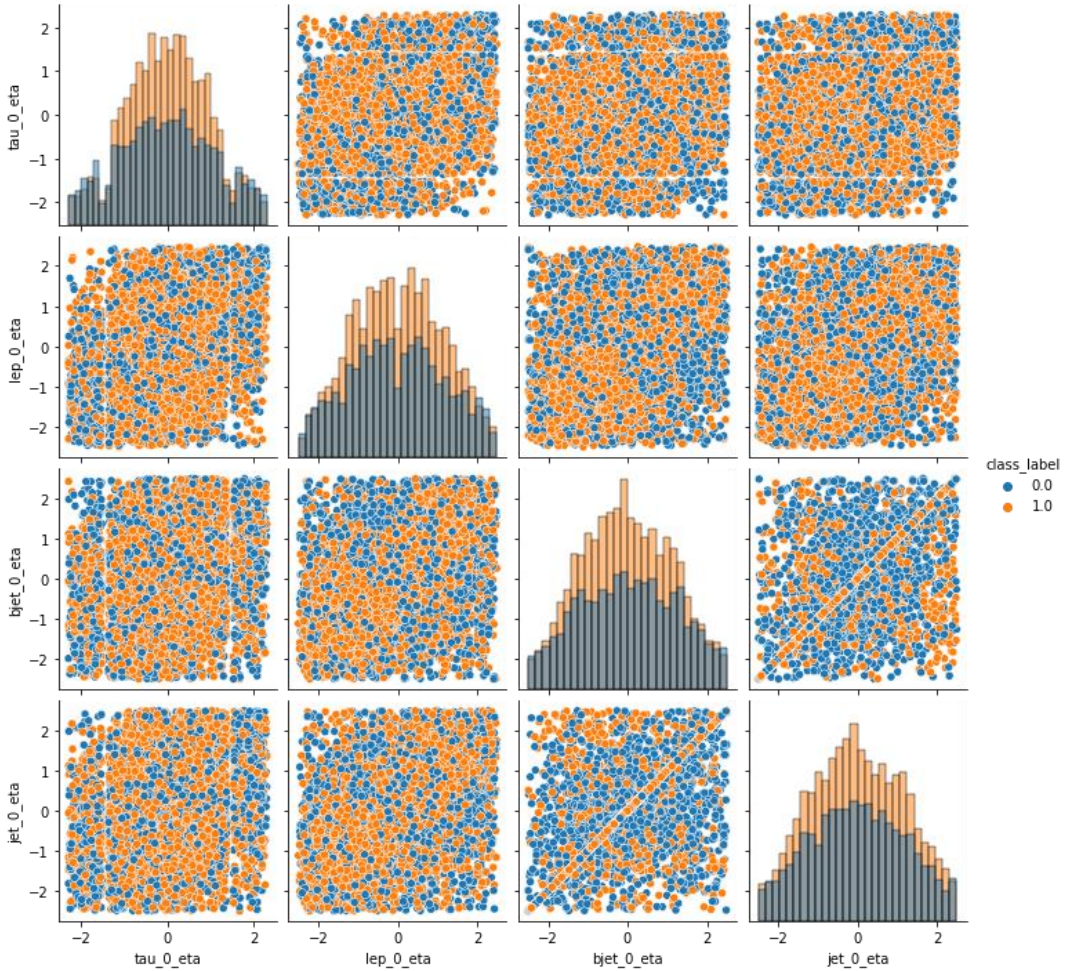
Data Visualization





Data Visualization





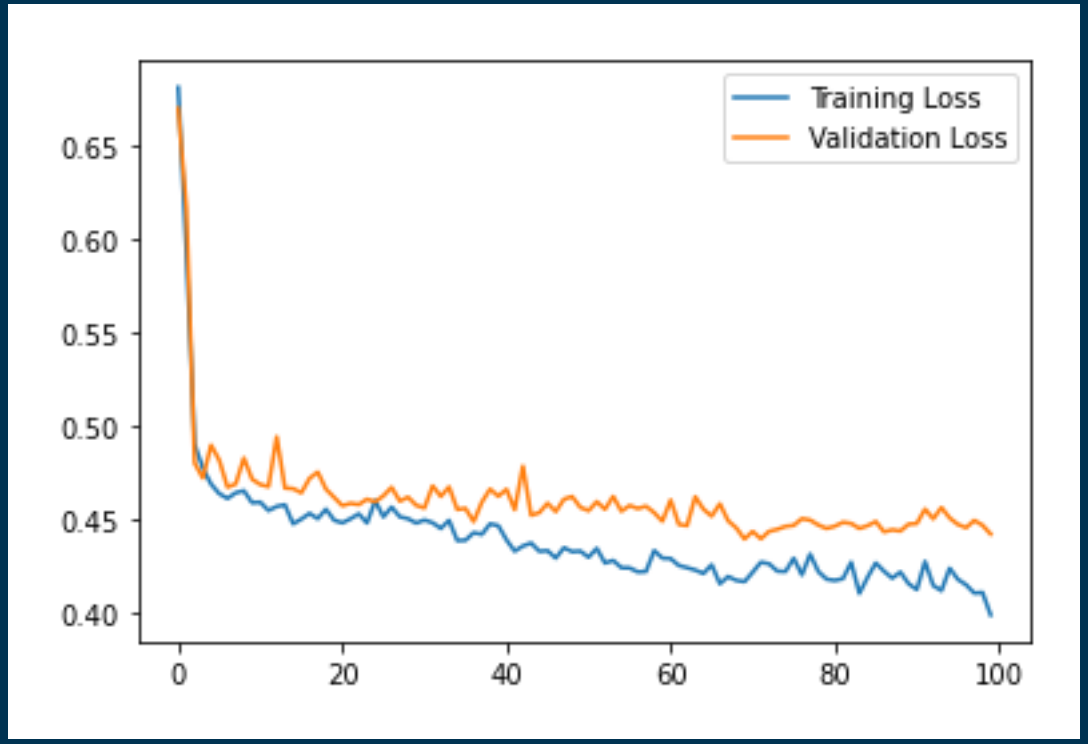
Problem Formulation

- Models:
  - HPANA model:
    - Sequential model with Dense and Dropout layers
  - LBN model:
    - Same as HPANA, but with the addition of the LBN layer.
- Loss: binary_crossentropy
- Optimizer: adam
- Hyperparameters: The LBN contains the M hyperparameters. This should be selected for the number of intermediate particles and rest frames.

```
Keras_model = Sequential()
input_shape = (6, 4)
Keras_model.add(LBNLayer(input_shape, 13, boost_mode=LBN.PAIRS))
Keras_model.add(BatchNormalization())
Keras_model.add(Dense(64, activation="sigmoid"))
Keras_model.add(Dropout(0.1))
Keras_model.add(Dense(64, activation="sigmoid"))
Keras_model.add(Dropout(0.1))
Keras_model.add(Dense(64, activation="sigmoid"))
Keras_model.add(Dropout(0.1))
Keras_model.add(Dense(64, activation="sigmoid"))
Keras_model.add(Dropout(0.1))
Keras_model.add(Dense(1,activation="sigmoid"))
```

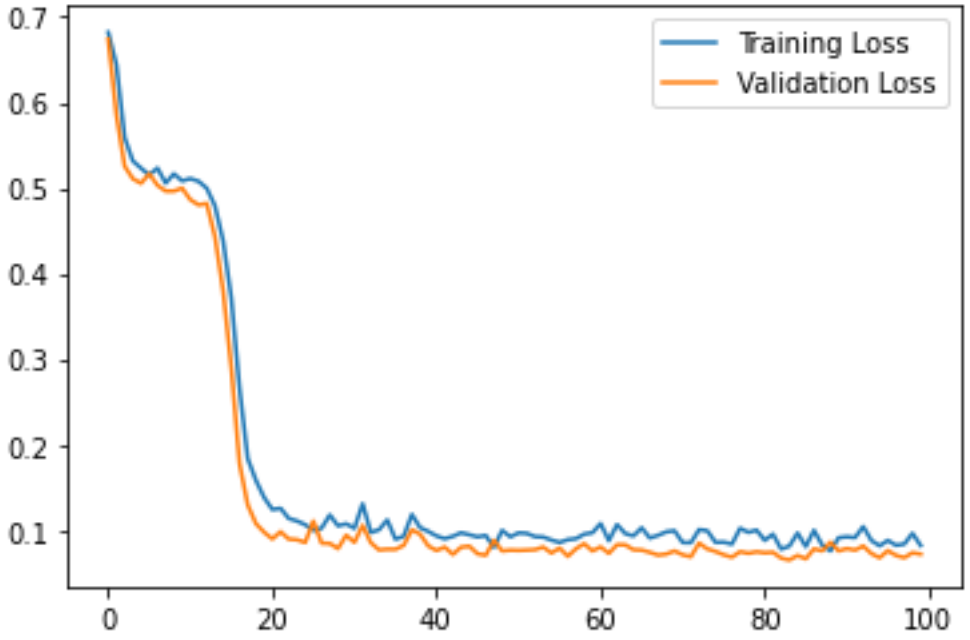
Training

- The training was done locally on a laptop
- Training on the small 7433 event sample takes roughly 3 minutes to complete.
- Training duration was determined by finding the point when overtraining begins and stopping prior.

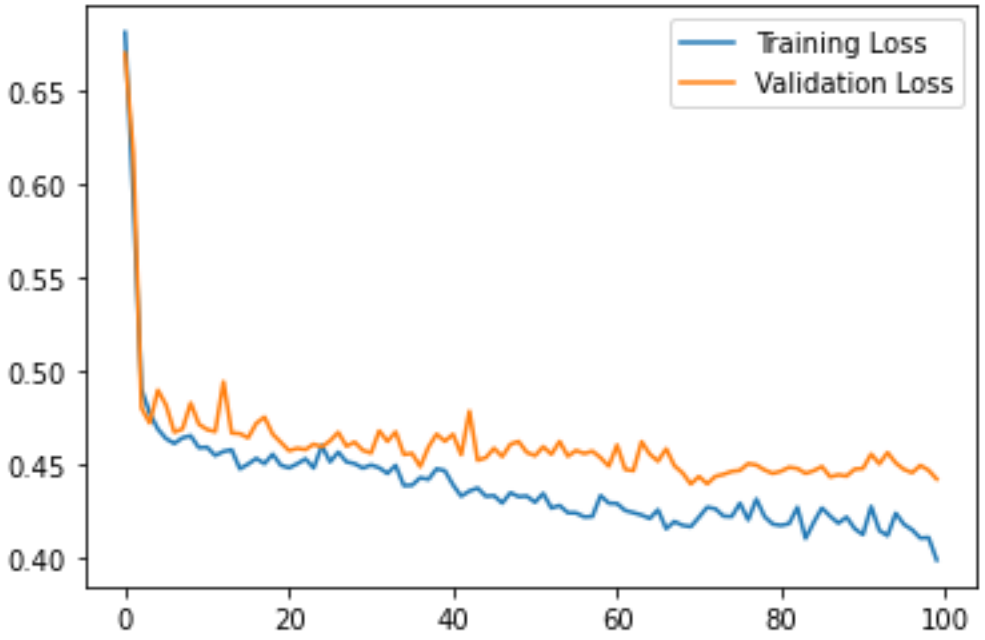


Loss

HPANA

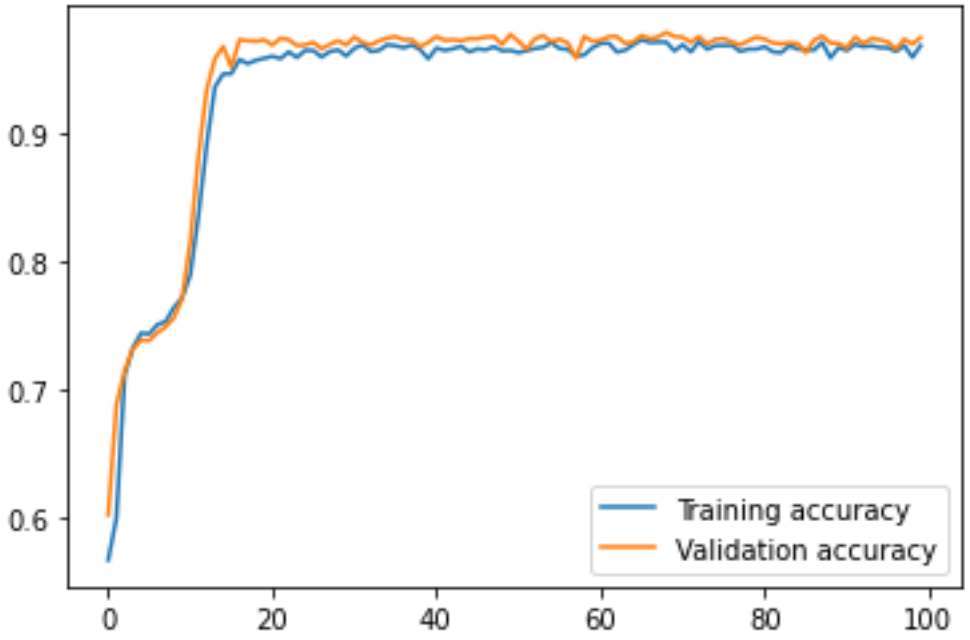


LBN

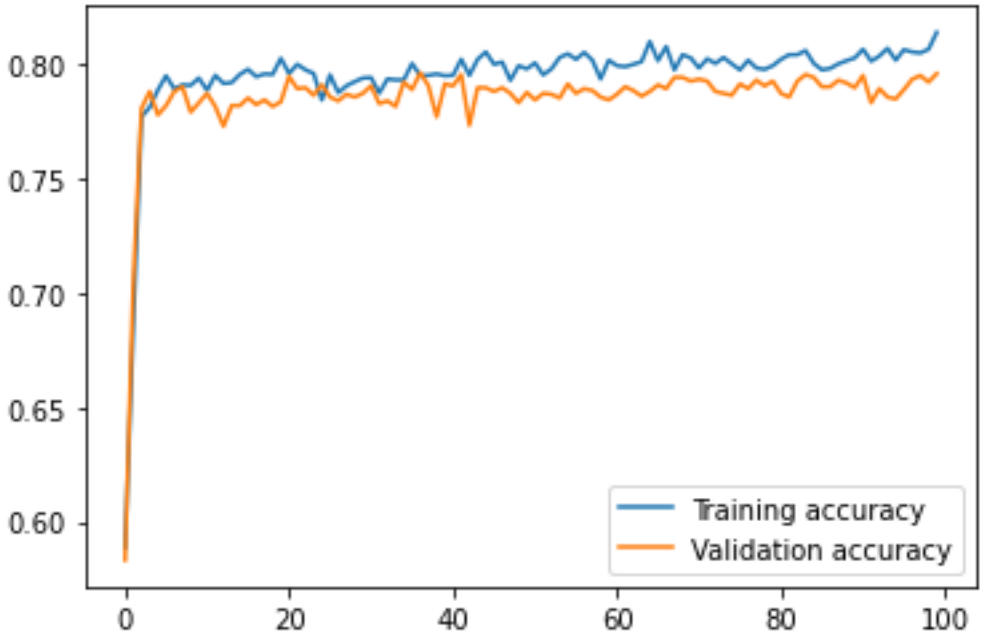


Accuracy

HPANA

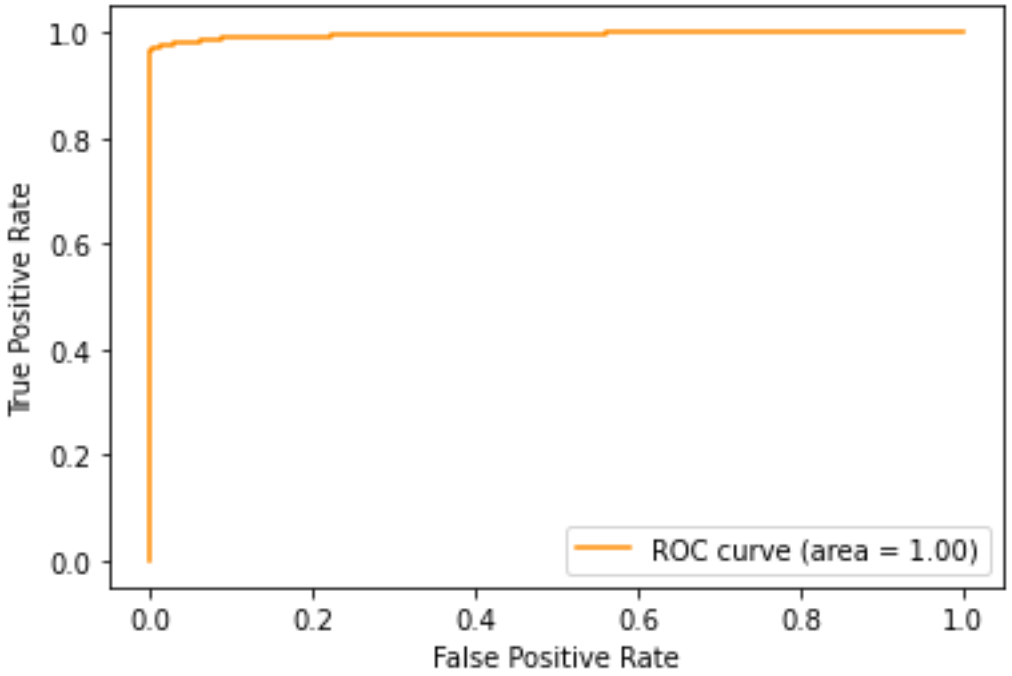


LBN

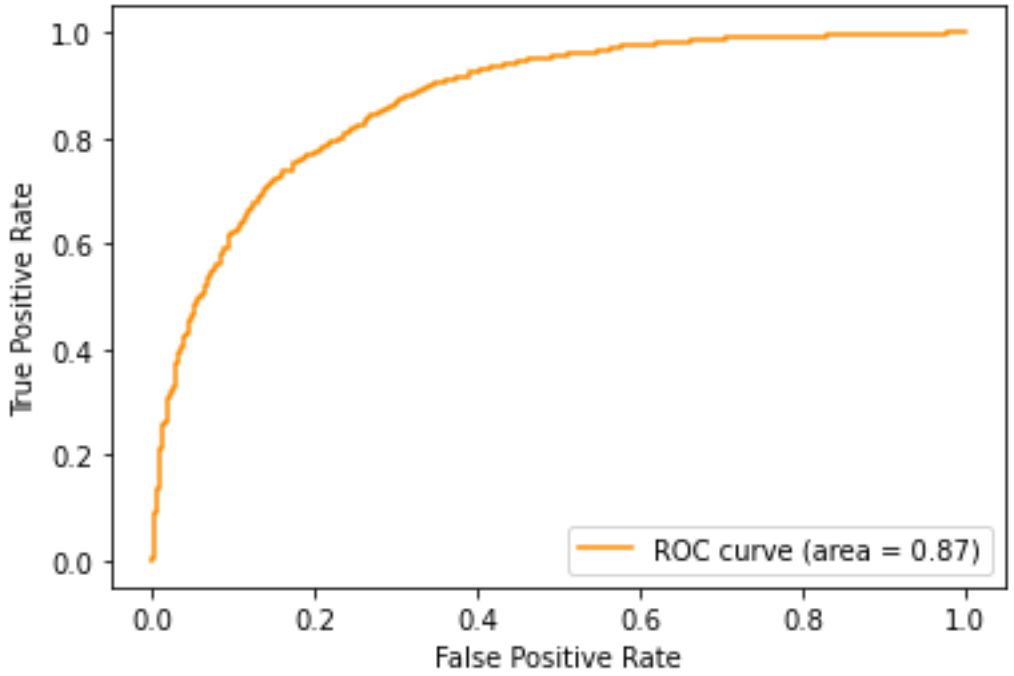


ROC

HPANA

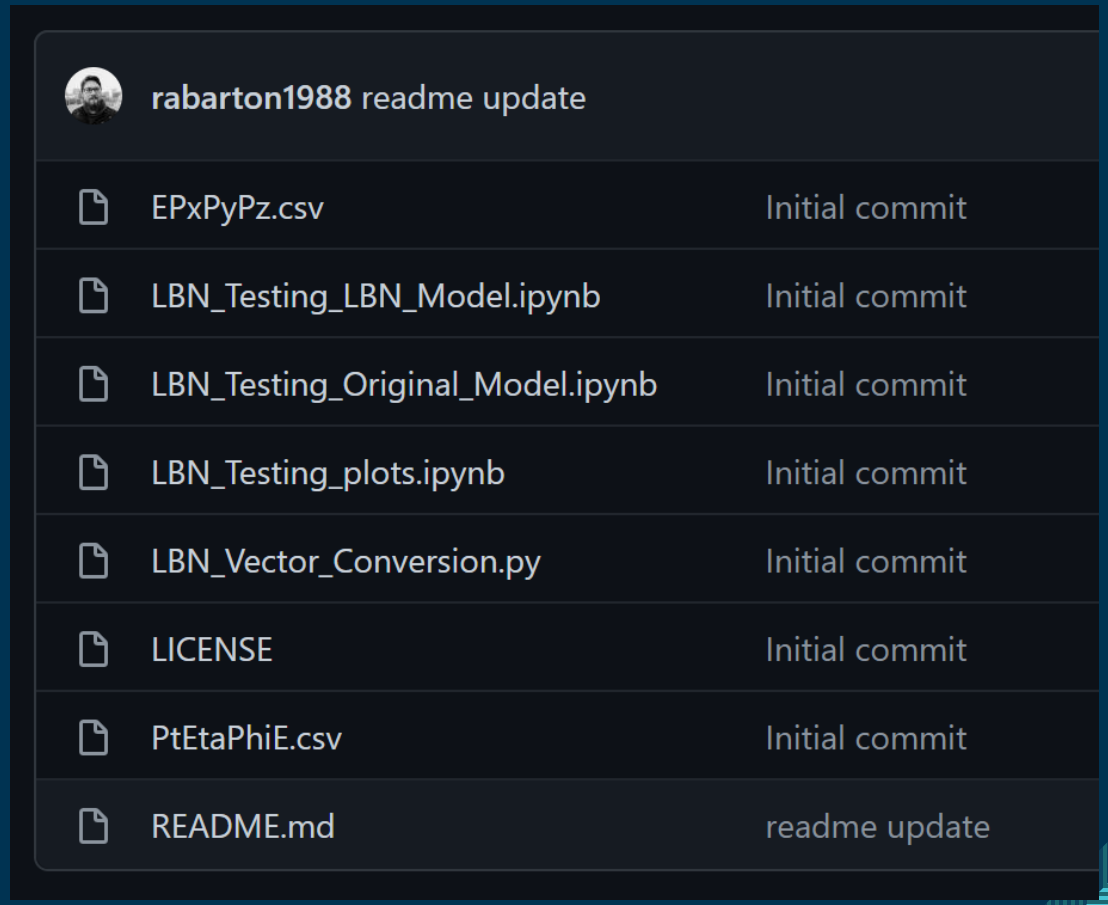


LBN



Overview of files in repository

- LBN_Testing_plots.ipynb
  - Plots of the data from PtEtaPhiE.csv and EPxPyPz.csv
- LBN_Testing_Original_Model.ipynb
  - Original - HPANA - model that trains on the dataframe from the pickle file
- LBN_Testing_LBN_Model.ipynb
  - Model including the - LBN - layer that trains on EPxPyPz.csv
- LBN_Vector_Conversion.py
  - Script used to convert PtEtaPhiE vectors to EPxPyPz vectors
- Generated data files
  - EPxPyPz.csv
  - PtEtaPhiE.csv



Software Setup / Citations / Conclusions

- Software:
  - Python
  - Numpy
  - Tensorflow
  - ROOT
- Citations:
  - https://github.com/riga/LBN
  - https://gitlab.cern.ch/atlas-hbsm-charged-higgs-taunu/hpana
  - [1812.09722] Lorentz Boost Networks: Autonomous Physics-Inspired Feature Engineering (arxiv.org)

- The results of the LBN implementation are inconclusive. The dataset used is incomplete and too small to see any significant results.
- The next step is to get a larger data set that includes all the backgrounds and Hplus mass points.