Thesis Presentation

Multiclass Classification Of Leptons In Proton-Proton Collisions At $\sqrt{s}$=13 TeV Using Machine Learning

Kristoffer Langstad

University of Oslo, Department of Physics

July 2, 2021

2021-07-02

Multiclass Classification Of Leptons In Proton-Proton Collisions At $\sqrt{s}$=13 TeV Using Machine Learning

└Thesis Presentation

Thesis Presentation

Multiclass Classification Of Leptons In Proton-Proton Collisions At $\sqrt{s}$=13 TeV Using Machine Learning

Kristoffer Langstad

University of Oslo, Department of Physics

July 2, 2021

Outline

1. Introduction
   (i) Particle physics model
   (ii) Machine Learning
2. Multiclass Classification
3. Results
4. Summary, Conclusion and Outlook

Outline

1. Introduction
   (i) Particle physics model
   (ii) Machine Learning
2. Multiclass Classification
3. Results
4. Summary, Conclusion and Outlook

Standard Model

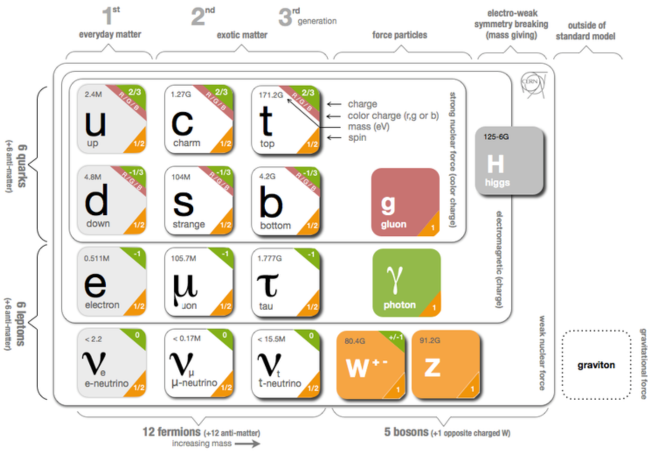
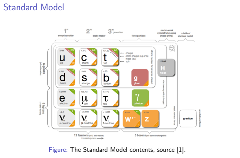


Figure: The Standard Model contents, source [1].



SM explain with great precision- fundamental particles in figure w/ charge, spin and mass - 17 particles - fermions and bosons. Does not explain graviton - non-zero mass of neutrino. Know from neutrino oscillations from the Sun - they change flavor and must have mass.

Introduce Inverse seesaw mechanism w/ heavy neutrinos and right-handed neutrinos. Only left-handed neutrinos, LH and RH for other SM particles. LH means direction of spin and motion are opposite. ISS-> trilepton final state with a neutrino through p-p collisions and decay through W-boson and heavy pseudo-Dirac neutrino. Continue to next slide with model->

Trilepton Final State

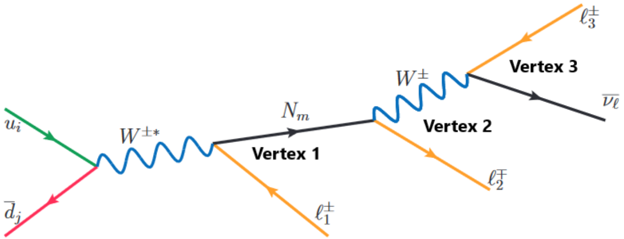


Figure: The Born diagram for the charged current Drell-Yan process of the proton-proton collision (on the left) producing a heavy pseudo-Dirac neutrino N in the inverse seesaw mechanism model, leading to a trilepton plus missing transverse energy (a light neutrino) final state. Figure is taken from ref. Pascoli et al. [2].

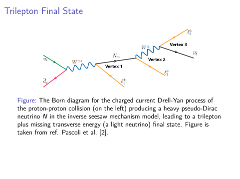


Figure of P-P collisions to trilepton final state and decays. Charges. At LHC and CERN, detected by ATLAS but neutrinos are not. Only MET since conservation of energy.

Called charged current Drell-Yan process, same model we look at for neutrinos as by Pascoli et al. [2]. Gives almost conserved lepton number and consider only electrons and muons. Amount of SS and OS events vertex 1 and 2 differ from normal seesaw. Allows LFV for vertex 1 and 2 for e and mu. Study two simulated neutrino signals, mass 150 GeV and 450 GeV, expect different LFV for different neutrino mass models.

Lepton Flavor Distributions

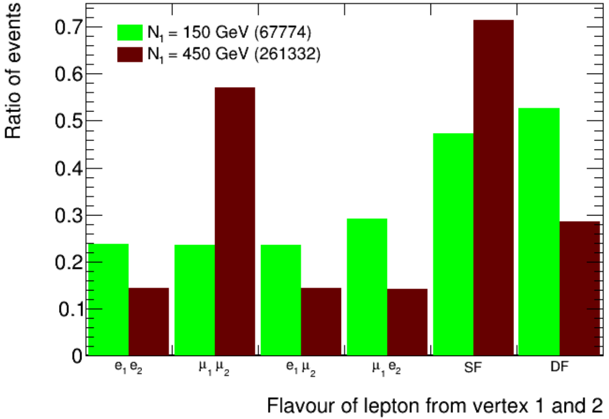


Figure: Distribution of the lepton flavor between vertex 1 and 2 for the 150 and 450 GeV signals where we either have two electrons, or two muons (SF) or an electron and a muon (DF).

Lepton Flavor Distributions

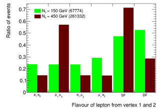


Figure: Distribution of the lepton flavor between vertex 1 and 2 for the 150 and 450 GeV signals where we either have two electrons, or two muons (SF) or an electron and a muon (DF).

Distributions of lepton flavor - event ratios lepton 1 and 2 - electrons and muons - two neutrino signals - SF: ee or mumu - DF: emu or mue. 450 - more SF events - 150 - barely more DF events.

Proton-Proton Collision Data

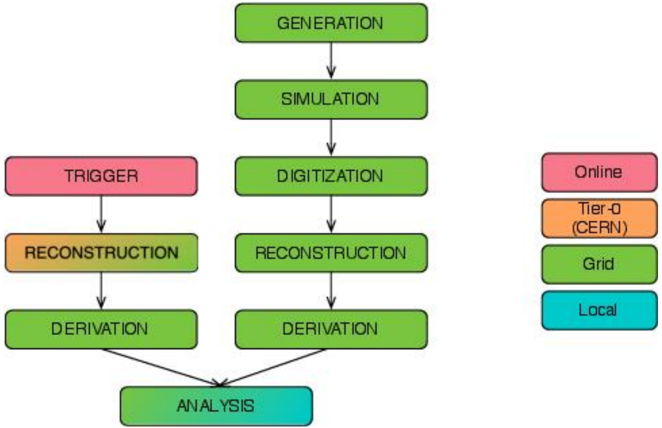
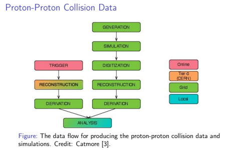


Figure: The data flow for producing the proton-proton collision data and simulations. Credit: Catmore [3].



MC simulated backgrounds, simulated neutrinos and p-p data at 13 TeV from LHC 2018 - data flow left side - not interesting. MC and signals - right side - ML.

Signal sim - train ML - after Generation - truth. MC - best represent all prod-mech with trilepton+MET. MC + signals - classif after Reconstruction.

Data Features

Original dataset features:

- Flavor, Charge, η, ϕ, p_T, E_T^{miss} (MET).

New and added features:

- p_x, p_y, p_z, θ, E.
- $\Delta\phi$, ΔR, m_{ll}, m_{3l}.

Target features:

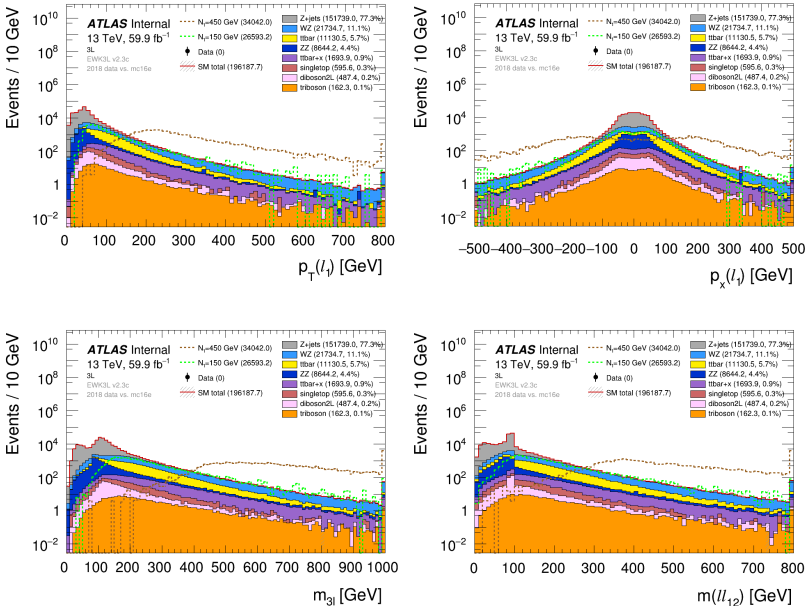
Vtx perm	Vtx 1	Vtx 2	Vtx 3
123	p_T^1	p_T^2	p_T^3
132	p_T^1	p_T^3	p_T^2
213	p_T^2	p_T^1	p_T^3
231	p_T^2	p_T^3	p_T^1
312	p_T^3	p_T^1	p_T^2
321	p_T^3	p_T^2	p_T^1

Data Features
Original dataset features:
- Flavor, Charge, η, ϕ, p_T, E_T^{miss} (MET).
New and added features:
- p_x, p_y, p_z, θ, E.
- $\Delta\phi$, ΔR, m_{ll}, m_{3l}.
Target features:
Vtx perm, Vtx 1, Vtx 2, Vtx 3
123 p_T^1 p_T^2 p_T^3
132 p_T^1 p_T^3 p_T^2
213 p_T^2 p_T^1 p_T^3
231 p_T^2 p_T^3 p_T^1
312 p_T^3 p_T^1 p_T^2
321 p_T^3 p_T^2 p_T^1

P-P collisions simulated, and measure properties like momentum, transverse momentum and coord-angles. Make new variables dPhi, dR and mll, m3l.

Make target classes - vertex permutations - three vertices, model in slide 4. Leading, subleading, subsubleading - wrt. pT - leading lepton = lepton 1, highest pT - six possible permutations to classify w/ML.

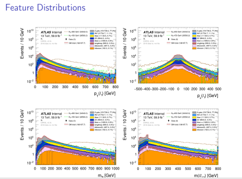
Feature Distributions



Multiclass Classification Of Leptons In Proton-Proton Collisions At $\sqrt{s}$=13 TeV Using Machine Learning

2021-07-02

└─Feature Distributions



Short explanation of what we see.

Machine Learning Process

What we want to do:

Use machine learning to identify lepton vertices in simulated backgrounds and signals.

How to do it:

I Use supervised learning and multiclass classification.
II Train and optimize machine learning algorithms
III Evaluate which model that best predicts the vertices.
IV Predict lepton vertices for simulated backgrounds and signals.

End goal:

I Look for lepton flavor violation between the classified leptons 1 and 2.
II Compare with a more standard analysis by Pascoli et al. [2].

Machine Learning Process
What we want to do:
Use machine learning to identify lepton vertices in simulated backgrounds and signals.
How to do it:
I Use supervised learning and multiclass classification.
II Train and optimize machine learning algorithms
III Evaluate which model that best predicts the vertices.
IV Predict lepton vertices for simulated backgrounds and signals.
End goal:
I Look for lepton flavor violation between the classified leptons 1 and 2.
II Compare with a more standard analysis by Pascoli et al. [2].

Machine learning - train to identify the particles vertices - pattern recognition in particle properties. Truth simulated neutrino signals - know the origins. Train various ML algorithms - use best performing model to predict simulated backgrounds and signals. Six lepton vertex permutations - multiclass classification case with six classes - not been studied much previously in particle physics.

Supervised learning - multiclass classification - Need good, fast algorithms for classification - lot of data - preprocessing of data before classification - train and tune ML algorithms/models - evaluate performance of the models for predicting classes/vertices - export best, skip training - predict lepton vertices sim backgrounds and signals.

Signal region cuts - study LFV for lepton 1 and 2 - distributions - compare with a more standard analysis by Pascoli et al. [2].

Preprocessing of Data

(i) Feature correlations
(ii) Resampling
(iii) Splitting into data sets
(iv) Scaling

Preprocessing of Data
(i) Feature correlations
(ii) Resampling
(iii) Splitting into data sets
(iv) Scaling

Check correlations - strong correlations=strong linear dependence - remove one feature=better results. Mutual information - want higher values=features have more information regarding the classes - statistical dependence between variables.

Imbalanced data=more data for some classes (bias)=bad predictions of minority classes - resampling techniques to balance classes (number of events).

Split data - training, validation, test - training=train models - validation=tune model hyperparameters and check models - tune w/randomized search - cross-validation - hyperparameters. Test=check final performance of (tuned) models.

Scale - transform values - standardization=0 mean, 1 std.dev- avoid weighted favoring of some classes - only for features - want categorical classes not distributions.

Bias-Variance Tradeoff

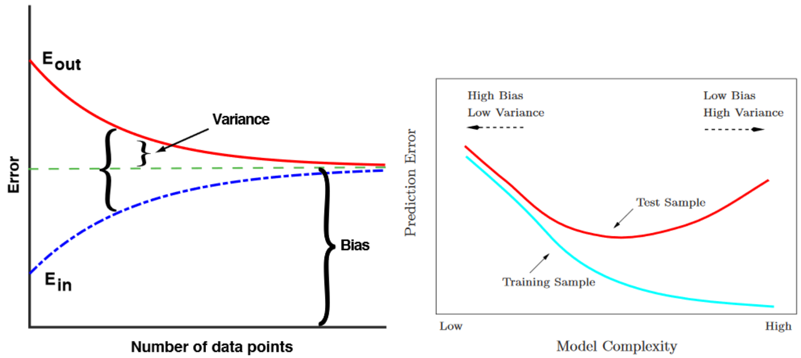
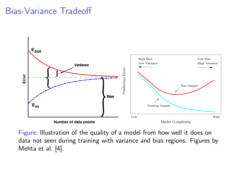


Figure: Illustration of the quality of a model from how well it does on data not seen during training with variance and bias regions. Figures by Mehta et al. [4].



Supervised learning problem - balance between variance and bias - best compromise=best model for number of data points/training set size and model complexity. Left: Out-of-sample error - test set error - in-sample error - training set error. Higher number of data points - lower var, higher bias, lower total error.

Right: Low complexity=high bias,low var=underfitting - high complexity=low bias,high var=overfitting. Overfitting more normal today. Figure - optimal at test minimum. Quality of model - on data not seen during training.

Classification Algorithms

Types of classification algorithms used:

(i) Logistic regression
(ii) Multi-layer perceptron (MLP)
(iii) Trees
  (1) Decision Tree
  (2) Random Forest
(iv) Boosters
  (1) AdaBoost
  (2) Gradient Boosting (HGBC)
  (3) Extreme Gradient Boost (XGBoost)
  (4) Light Gradient Boosting Machine (LGBM)
(v) Multiclassifiers
  (1) One-Vs-Rest
  (2) One-Vs-One

Classification Algorithms
Types of classification algorithms used:
(i) Logistic regression
(ii) Multi-layer perceptron (MLP)
(iii) Trees
(1) Decision Tree
(2) Random Forest
(iv) Boosters
(1) AdaBoost
(2) Gradient Boosting (HGBC)
(3) Extreme Gradient Boost (XGBoost)
(4) Light Gradient Boosting Machine (LGBM)
(v) Multiclassifiers
(1) One-Vs-Rest
(2) One-Vs-One

Binary classifiers to multiclass: LR - linear regression with a logistic function to predict. MLP - neural network - input, hidden, output layers - weights, biases, non-linear activation func to output - hyperparameters, regularization control overfitting. DTC - simpler, single tree model with features - criterion for value splits - control hyperparm for overfit. RF - ensemble of trees - increase accuracy, decease variance. Boost - iteration weights - sequential models built - better classifier.
AdaBoost - adaptive boost - weights adapt each iteration - majority vote -> better classifier. GradientBoost - tree boost approx Ada - weighted gradient in loss func - HistGradient larger data sets - less time, higher accuracy - bins. XGB - opt hist dist grad boost alg - accurate fast parallel - scalable - dist of features - complex w/hyperparameters. LGBM - distributed gradient boost - faster, memory efficient, accurate - large data sets - information gain - drop feat threshold.

Multiclass: Multiclass to binary cases techniques.

Classification Results

Model	Signal models			
	150 GeV		450 GeV	
	Accuracy	Accuracy_train	Accuracy	Accuracy_train
AdaBoost	0.8519	1.0000	0.9385	1.0000
MLP	0.8227	0.9492	0.9350	0.9606
HGBC	0.7863	0.8999	0.9280	0.9571
XGBoost	0.8631	0.9998	0.9509	0.9999
LGBM	**0.8779**	0.9999	**0.9541**	0.9999

Table: Accuracy scores of the highest performing classification models trained on the 150 GeV and 450 GeV validation and training sets.

Classification Results

Model	Signal models			
	150 GeV		450 GeV	
	Accuracy	Accuracy_train	Accuracy	Accuracy_train
AdaBoost	0.8519	1.0000	0.9385	1.0000
MLP	0.8227	0.9492	0.9350	0.9606
HGBC	0.7863	0.8999	0.9280	0.9571
XGBoost	0.8631	0.9998	0.9509	0.9999
LGBM	**0.8779**	0.9999	**0.9541**	0.9999

Table: Accuracy scores of the highest performing classification models trained on the 150 GeV and 450 GeV validation and training sets.

Highest scoring - both signals - validation and training scores. Tendency to overfit - 450 scores closer. XGB and LGBM best - LGBM better and faster - LGBM chosen.

Light Gradient Boosting Machine

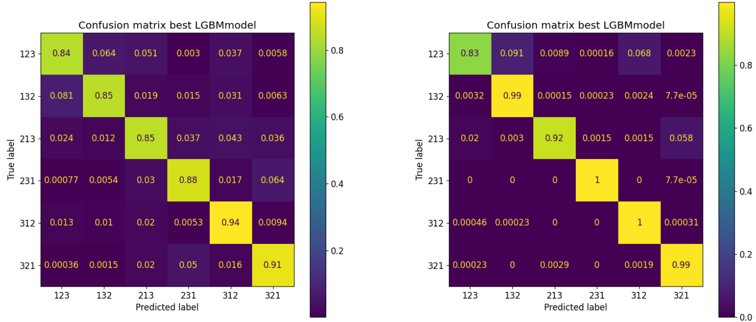
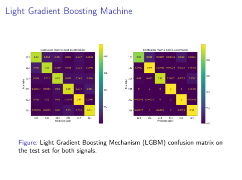


Figure: Light Gradient Boosting Mechanism (LGBM) confusion matrix on the test set for both signals.



LGBM on test data - no tuning - more metrics. Confusion matrices - accuracy of each class (diagonal) - true class versus predicted class - normalized horizontally. Diagonal between 0.8-1.0 - good predictions. 450 GeV more accurate - 132, 231, 312, 321 better accuracy.

Scores

Signal [GeV]	Accuracy	Accuracy_train	CKS	LogLoss
150	0.8793	0.9999	0.8551	0.3335
450	0.9570	0.9999	0.9484	0.1120

Table: Accuracy score of both the test and training sets, Cohen Kappe score and logloss for both the 150 and 450 GeV signal models.

Signal [GeV]	ROC Curve		Precision-Recall
	Micro AUC	Macro AUC	Micro AUC
150	0.99	0.99	0.949
450	1.0	1.0	0.994

Table: Micro and macro area under the curve (AUC) scores for both Receiver Operating Characteristic (ROC) and precision-recall curves.

Scores

Signal [GeV]	Accuracy	Accuracy_train	CKS	LogLoss
150	0.8793	0.9999	0.8551	0.3335
450	0.9570	0.9999	0.9484	0.1120

Table: Accuracy score of both the test and training sets, Cohen Kappe score and logloss for both the 150 and 450 GeV signal models.

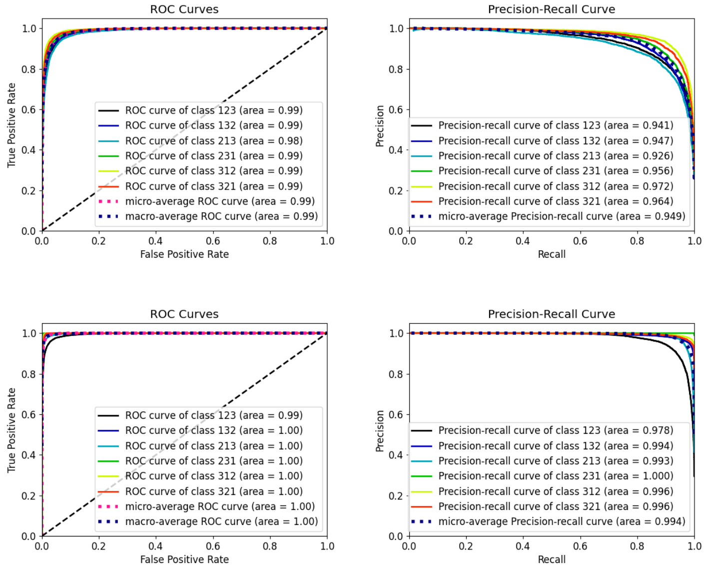
Signal [GeV]	ROC Curve		Precision-Recall
	Micro AUC	Macro AUC	Micro AUC
150	0.99	0.99	0.949
450	1.0	1.0	0.994

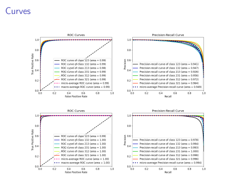
Table: Micro and macro area under the curve (AUC) scores for both Receiver Operating Characteristic (ROC) and precision-recall curves.

Table 1: Accuracy scores, CKS and log loss (error) of LGBM - CKS accounts uncertainties, random model vs LGBM - logloss, error of probabilities of classes. Validation and test similar accuracy - good for different data - NB: same original data set. 450 GeV still better - high acc - low log loss.

Table 2: Micro and macro AUC for ROC and precision-recall - show overall performance - AUC over 0.8=good model - both values over 0.9. LGBM shows great promise on predicting the vertices on these test set.

Curves





Short explanation of what we see.

Classify Simulated Data

What to classify with the LGBM:

(i) Simulated background production-mechanisms with trilepton final states plus MET.

(ii) Two reconstructed neutrino signals, with neutrino masses of 150 and 450 GeV.

- Use classified vertex permutations to define new signal regions with opposite sign and same flavor or different flavor for lepton 1 and 2.
- Compare signal regions with benchmark analysis.

Classify Simulated Data

What to classify with the LGBM:

(i) Simulated background production-mechanisms with trilepton final states plus MET.

(ii) Two reconstructed neutrino signals, with neutrino masses of 150 and 450 GeV.

- Use classified vertex permutations to define new signal regions with opposite sign and same flavor or different flavor for lepton 1 and 2.
- Compare signal regions with benchmark analysis.

LGBM - performs well on truth neutrino signals - classify similar datasets - simulated backgrounds and signals - same neutrino masses.

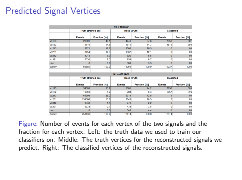
Expected - most events with highest p_T from N_1 production - loose momentum after decaying - 123 and 132 vertices most both - few 213 also for MC - enough for further analysis.

Predicted Signal Vertices

N1 = 150GeV						
	Truth (trained on)		Reco (truth)		Classified	
	Events	Fraction [%]	Events	Fraction [%]	Events	Fraction [%]
vtx123	26801	40.1	4241	37.6	7358	60.1
vtx132	9716	14.5	1615	14.3	4879	39.9
vtx213	12871	19.2	2308	20.5	0	0.0
vtx231	8454	12.6	1362	12.1	0	0.0
vtx312	4013	6.0	620	5.5	0	0.0
vtx321	5030	7.5	754	6.7	0	0.0
uncl	0	0.0	369	3.3	0	0.0
sumev	66885	100.0	11269	100.0	12237	100.0

N1 = 450 GeV						
	Truth (trained on)		Reco (truth)		Classified	
	Events	Fraction [%]	Events	Fraction [%]	Events	Fraction [%]
vtx123	34303	13.2	3005	24.2	7856	60.8
vtx132	10863	4.2	784	6.3	5057	39.2
vtx213	65308	25.2	5319	42.8	1	0.0
vtx231	139686	53.8	2043	16.5	0	0.0
vtx312	3938	1.5	279	2.2	0	0.0
vtx321	5338	2.1	438	3.5	0	0.0
uncl	0	0.0	546	4.4	0	0.0
sumev	259436	100.0	12414	100.0	12914	100.0

Figure: Number of events for each vertex of the two signals and the fraction for each vertex. Left: the truth data we used to train our classifiers on. Middle: The truth vertices for the reconstructed signals we predict. Right: The classified vertices of the reconstructed signals.



Extra: Check on signal predictions - see table for vertex events - 150 and 450 predicted with respective model - see number of events for each vertex + fractions. Left: original truth data - training ML - Middle: reconstructed signals we classify - Right: classified vertices.

Interesting to mention - truth after Generation and recon 150 GeV - similar event fractions - 450 GeV have not as close. Events disappears after reconstruction - between left to middle. Classif - 123 and 132 predicted - one 213 with 450 - predictions does not fit truth recon - Why? - LGBM trained on signals before recon in data flow - not sure if true, a guess - resampling leading to misclassification? - further analysis needed?

Signal Regions

The signal regions are for the predicted vertices:

- vtx123 with lepton 1 and 2 having SF/DF + OS
- vtx132 with lepton 1 and 3 having SF/DF + OS
- vtx213 with lepton 1 and 2 having SF/DF + OS

Benchmark "Standard" Analysis at $\sqrt{s} = 14$ TeV:
$m_{l_i,l_j} > 10$ GeV, $\|m_{l_i,l_j} - M_Z\| > 15$ GeV, $\|m_{3l} - M_Z\| > 15$ GeV, $p_T^{l_1} > 55$ GeV, $p_T^{l_2} > 15$ GeV, $m_{3l} > 80$ GeV

Table: Cuts used for a benchmark analysis. The combinations of $l_i l_j$ are for l_1, l_2 and l_3. $M_Z = 91.2$ GeV is the mass of the Z-boson and m_{3l} is the invariant mass of the three lepton system. Reference: Table 6 in Pascoli et al. [2].

Signal Regions

The signal regions are for the predicted vertices:

- vtx123 with lepton 1 and 2 having SF/DF + OS
- vtx132 with lepton 1 and 3 having SF/DF + OS
- vtx213 with lepton 1 and 2 having SF/DF + OS

Benchmark "Standard" Analysis at $\sqrt{s} = 14$ TeV:

$m_{l_i,l_j} > 10$ GeV, $|m_{l_i,l_j} - M_Z| > 15$ GeV, $|m_{3l} - M_Z| > 15$ GeV, $p_T^{l_1} > 55$ GeV, $p_T^{l_2} > 15$ GeV, $m_{3l} > 80$ GeV

Table: Cuts used for a benchmark analysis. The combinations of $l_i l_j$ are for l_1, l_2 and l_3. $M_Z = 91.2$ GeV is the mass of the Z-boson and m_{3l} is the invariant mass of the three lepton system. Reference: Table 6 in Pascoli et al. [2].

SF, DF and OS between classified lepton 1 and 2.

Compare with standard analysis - missing one b-tagged cut - not available in datasets.

Analysis Results

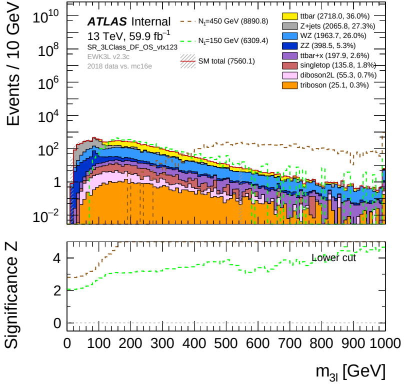
Look at:

- Invariant mass of three lepton system, m_{3l}, and MET in the mentioned signal regions.
- Event distributions and significance.
- SF versus DF.
- ML versus benchmark.

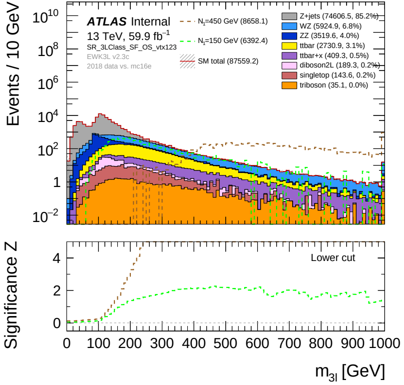
Analysis Results
Look at:
- Invariant mass of three lepton system, m_{3l}, and MET in the mentioned signal regions.
- Event distributions and significance.
- SF versus DF.
- ML versus benchmark.

Invariant mass three lepton system and MET - event distributions and significance - high significance=high sensitivity - where to cut on variable to maximize sensitivity.

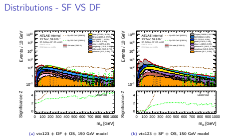
Distributions - SF VS DF



(a) vtx123 + DF + OS, 150 GeV model



(b) vtx123 + SF + OS, 150 GeV model

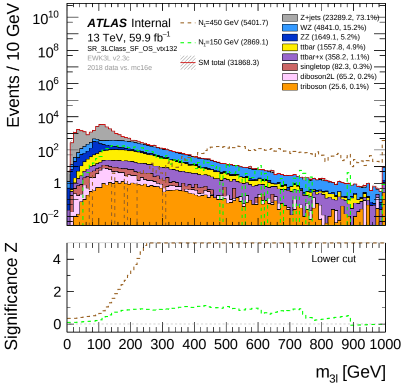


Much less events after cuts - more events for SF - difference from Z not decaying into electron-muon events - less events for large MC (DF) like WZ and Z+jets - same number events for signals SF and DF.

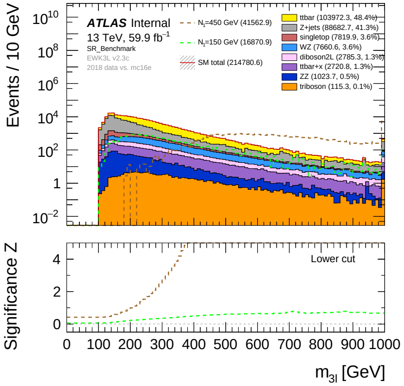
213 - no significance, no signal - few MC events - little information from 213 distributions here.

450 GeV sim signal - easier to differentiate with MC - masses above 400-500 GeV for m_{3l}.

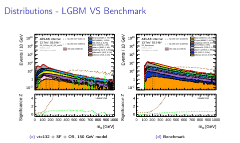
Distributions - LGBM VS Benchmark



(c) vtx132 + SF + OS, 150 GeV model

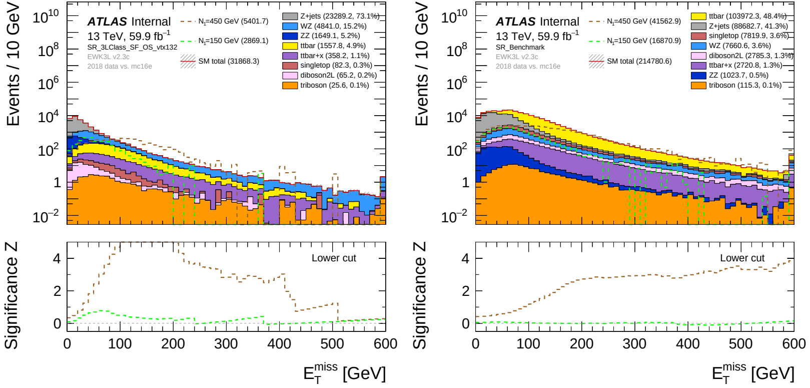


(d) Benchmark



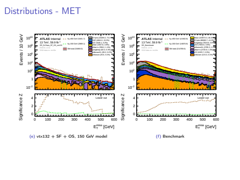
Significance higher for 450 GeV comp 150 GeV - higher for inv mass vs met. Benchmark similar to vtx132, SF, OS - significance 450 GeV above 4σ - 150 GeV below 1σ. Differentiate bkgs vs 450 GeV signal similar above 500 GeV - significance and number of events bkgs vs 450 signal shows LGBM model in general better than benchmark for differentiate bkgs and 450 signal. Benchmark not optimized to our model.

Distributions - MET



(e) vtx132 + SF + OS, 150 GeV model

(f) Benchmark



Comparing same signal regions - MET - signals and backgrounds more alike - MET does not discriminate well signals and backgrounds.

Summary and Outlook

We have shown:

- Multiclass classification is well suited for predicting from which vertex a lepton comes from for two types of neutrino mass models.
- Lepton flavor violation with different number of events for SF and DF.
- 450 GeV signal easier to differentiate against backgrounds with significance above 5σ.

Future and improvements:

- Test other classifiers.
- Train more models with other parameters.
- Train on data after reconstruction.
- Other neutrino mass models.

Summary and Outlook
We have shown:
- Multiclass classification is well suited for predicting from which vertex a lepton comes from for two types of neutrino mass models.
- Lepton flavor violation with different number of events for SF and DF.
- 450 GeV signal easier to differentiate against backgrounds with significance above 5σ.
Future and improvements:
- Test other classifiers.
- Train more models with other parameters.
- Train on data after reconstruction.
- Other neutrino mass models.

Shown - use multiclass classification and ML - predict and classify - sim of subseq decay leptons - from p-p collisions - final state model w/three leptons + neutrino.
Lepton Flavor violation - SF more events - lep 1 and 2 - predicted vertices - high significance in some signal regions.
Implement/construct framework multiclass classification - if excess observed - understand sign and flavor predicted by neutrino models. E.g. excess CMS - 2.8 significance in eejj - not in mmjj - eejj had SS/OS event ratio 1/14 - not consistent w/LRSM theory - classifier to understand neutrino models.
No time to study sensitivity - would discover 450 model long time ago - significance above 5σ - understand discovery rather than used to discover.
Future: Test other and/or better suited models. Train more models with more parameters. Train on data after reconstruction. Particle aspects future - other neutrino masses - SS vs OS - include detector data from CERN.

Codes

Codes found at the following GitHub-repository:
https://github.com/krilangs/ComPhys—Master

Codes
Codes found at the following GitHub-repository:
https://github.com/krilangs/ComPhys—Master

References

[1] Andrew Purcell. Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules. page 10, Aug 2012. URL `https://cds.cern.ch/record/1473657`.

[2] Silvia Pascoli, Richard Ruiz, and Cedric Weiland. Heavy neutrinos with dynamic jet vetoes: multilepton searches at $\sqrt{s}$= 14, 27, and 100 TeV. *Journal of High Energy Physics*, 2019(6):49, 2019.

[3] James Catmore. The atlas data processing chain: from collisions to papers. *University of Oslo, presentation slides*, 2020.

[4] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre GR Day, Clint Richardson, Charles K Fisher, and David J Schwab. A high-bias, low-variance introduction to machine learning for physicists. *Physics reports*, 810:1–124, 2019.

References

[1] Andrew Purcell. Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules. page 10, Aug 2012. URL https://cds.cern.ch/record/1473657.

[2] Silvia Pascoli, Richard Ruiz, and Cedric Weiland. Heavy neutrinos with dynamic jet vetoes: multilepton searches at $\sqrt{s}$= 14, 27, and 100 TeV. *Journal of High Energy Physics*, 2019(6):49, 2019.

[3] James Catmore. The atlas data processing chain: from collisions to papers. *University of Oslo, presentation slides*, 2020.

[4] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre GR Day, Clint Richardson, Charles K Fisher, and David J Schwab. A high-bias, low-variance introduction to machine learning for physicists. *Physics reports*, 810:1–124, 2019.