UiO : Fysisk institutt
Det matematisk-naturvitenskapelige fakultet

Application of Supervised Machine Learning to the Search for New Physics in ATLAS data

A Study of Ordinary Dense, Parameterized and Ensemble Networks and their Application to High Energy Physics

William Hirst **May 20, 2023**

Outline

Overview

1. Study individual attributes of a set of supervised methods
2. Compare expected sensitivity between methods on a subset of data
3. Attempt to increase sensitivity via feature reduction (PCA)
4. Compare the expected limits achieved by best performing methods to previous ATLAS analysis

Outline

Why apply machine learning to HEP problems?

How do we search for new physics?

Outline

A summary of the applied methods

Three neural network variants

- Ordinary dense neural network
- Ensemble networks utilizing Local-Winner-Takes-All (LWTA) layers
- Parameterized neural networks (PNN)

One boosted decision tree method

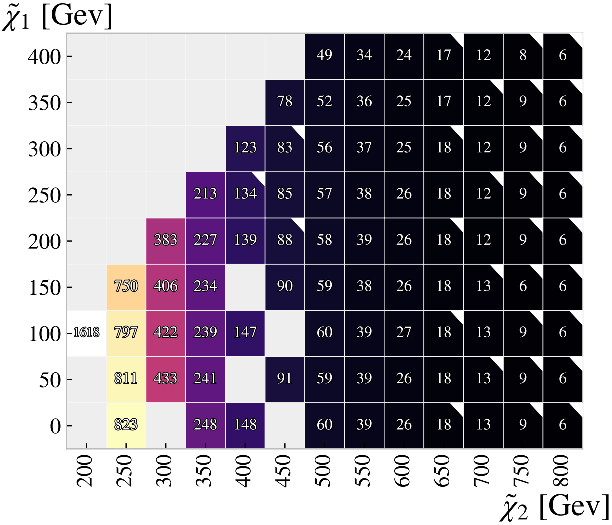
- XGBoost using default settings

How are the methods compared?

Training strategy

Mass combinations of the chargino-neutralino pair

- Full signal grid
  - 89 mass combinations
- Original signal set: white corners
  - 30 mass combinations
- The smaller the masses, the larger the contribution

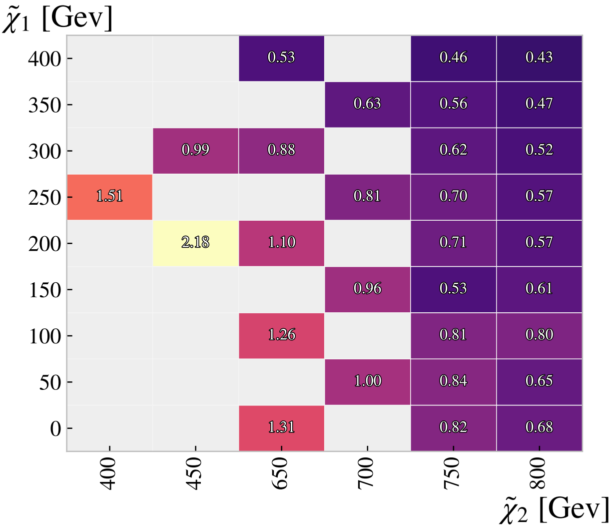


Outline

An introduction and study of each method

Boosted decision trees - XGBoost

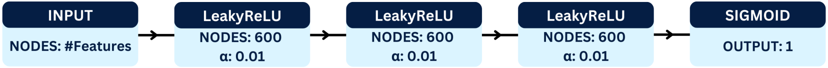
- Used as benchmark
- Minimal time spent tuning BDT
- Trained on original signal set
- Displayed better performance on lower masses



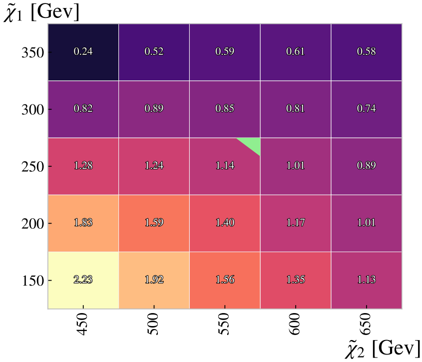
Ordinary dense neural network

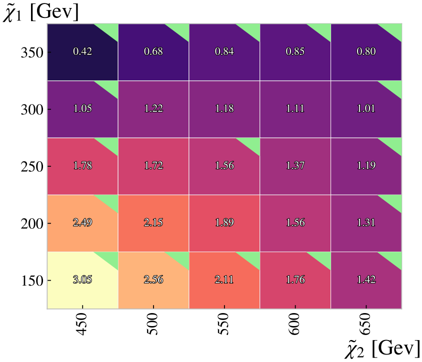
- First network variant tested
- 'Traditional' neural network
- 3 hidden layers with 600 nodes each
  - Deep

NN:



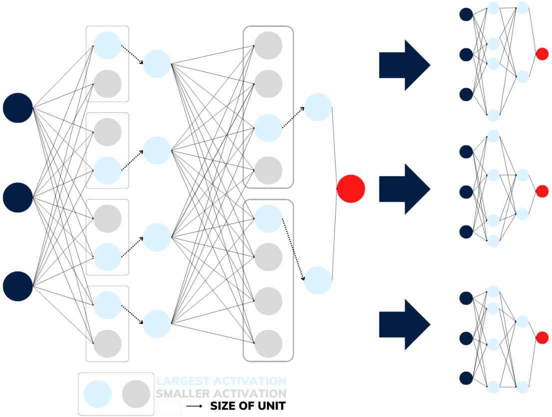
Compare one-mass approach to several-masses approach





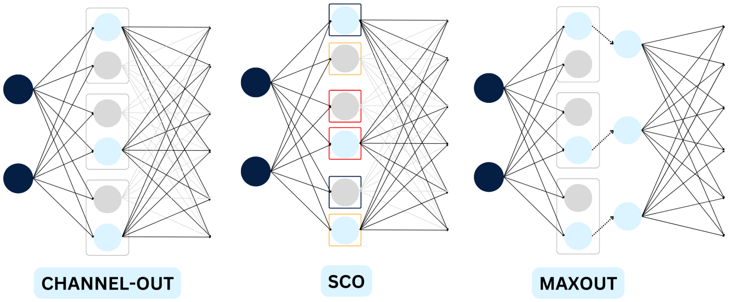
Ensemble methods - LWTA

- Dropout
- What is LWTA?
- Competing nodes - Units
- Encode information in pattern specific pathways



Channel-Out, SCO and Maxout

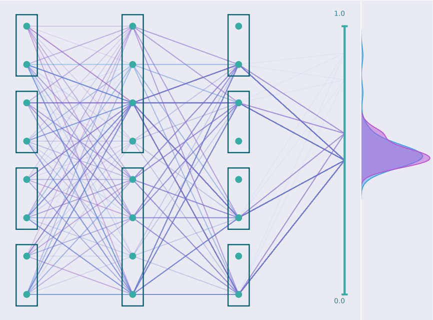
Layer	Separate Weights & Biases	Static Units
Channel-Out	*Yes*	*Yes*
SCO	*Yes*	*No*
Maxout	*No*	*Yes*



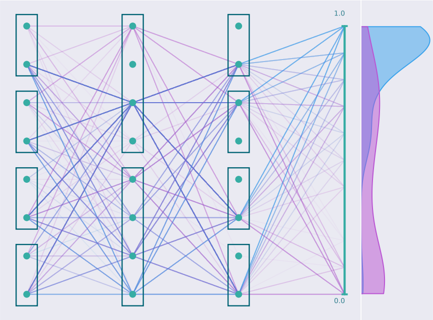
Visualization and study of sparse pathways

- A study of the implementation and effect of LWTA layers
- Visualize the activation and paths of 100 randomly sampled events
  - 50 background
  - 50 signal
- The bolder the line the more frequently the path is used.

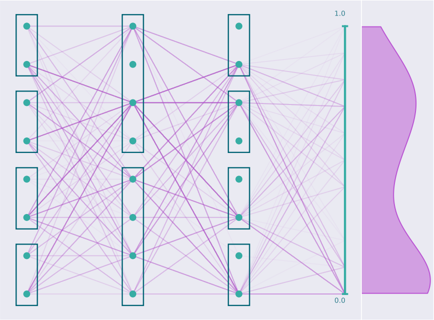
Before training



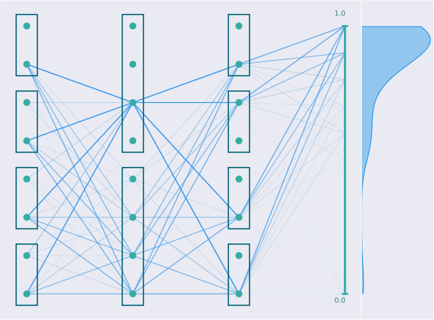
After training



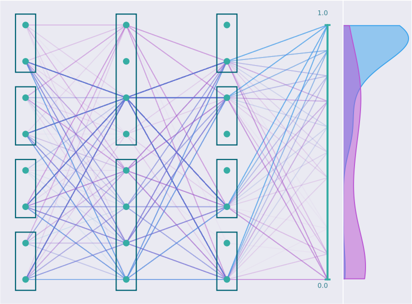
Background



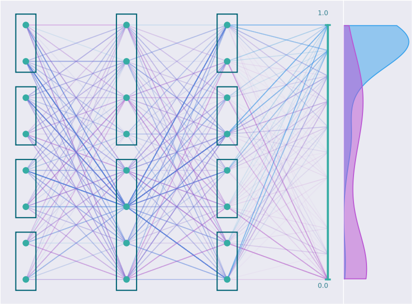
Signal



Comparing activation of Maxout with SCO



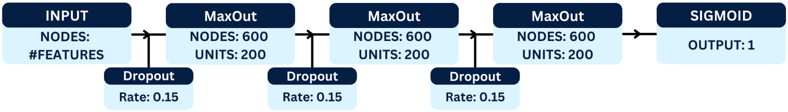
Maxout



SCO

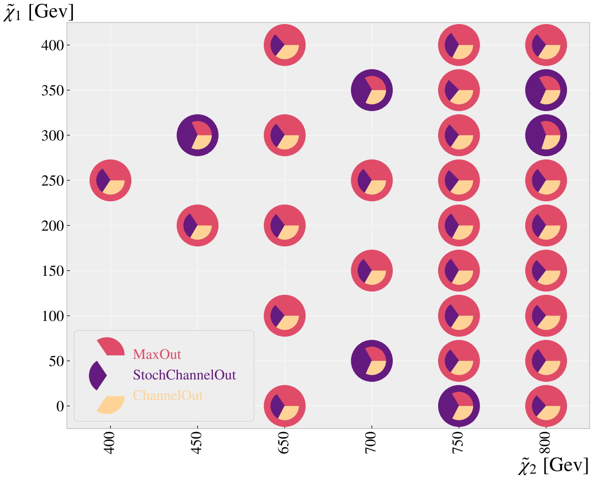
Ensemble network architecture

MaxOut:



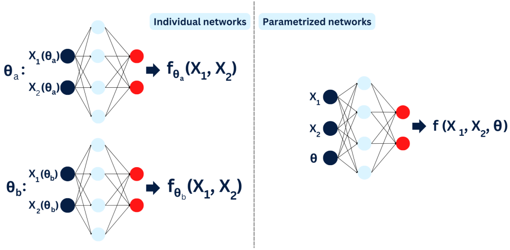
Comparing sensitivity of channel-out, SCO and maxout

- Maxout: 23/30
- SCO: 7/30
    - No trend for preferred masses
    - Possibly improve without layer on prediction

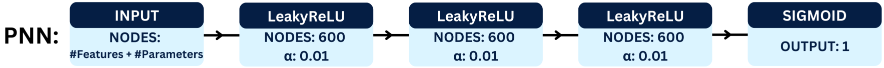


Parameterized neural network

- For diverse data set, X, dependent on a parameter, $X(\theta)$
  - Classical approach: One model for each parameter
  - PNN approach: Include θ as feature in feature set
- Signal events using masses $\{A, B\}_{GeV}$ to generate event during simulation will include the parameters A and B in feature set
- Background assigned parameters randomly using same distribution as signal
- Motivation
  - Network will associate parameters with trends in the data

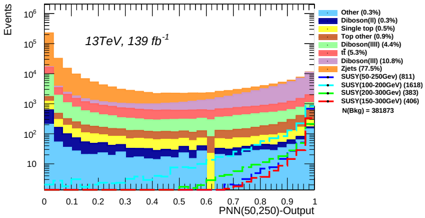


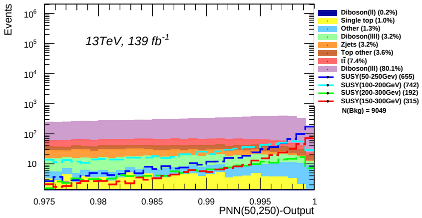
PNN architecture

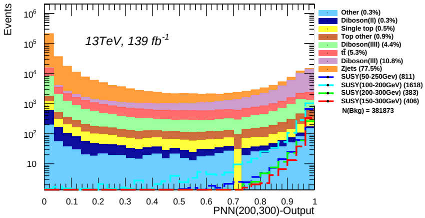


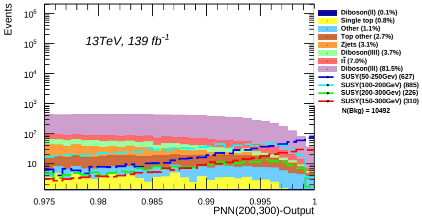
Study the effect of the parameters in the PNN

- Study if the parameters effect the training as intended
- Test: Manually assign all the events, both background and signal, the same parameters (mass combinations) thereby assigning most of the signal the wrong parameters
- Hypothesis: PNN performs better when events are assigned correct parameters
- First test: All events are given parameters $\{50, 250\}_{GeV}$
- Second test: All events are given parameters $\{200, 300\}_{GeV}$


Events
13TeV, 139 fb-1
Other (0.3%)
Diboson(II) (0.3%)
Single top (0.5%)
Top other (0.9%)
Diboson(IIII) (4.4%)
tt̄ (5.3%)
Diboson(III) (10.8%)
Zjets (77.5%)
SUSY(50-250Gev) (811)
SUSY(100-200GeV) (1618)
SUSY(200-300Gev) (383)
SUSY(150-300GeV) (406)
N(Bkg) = 381873
PNN(50,250)-Output


Events
13TeV, 139 fb-1
Diboson(II) (0.2%)
Single top (1.0%)
Other (1.3%)
Diboson(IIII) (3.2%)
Zjets (3.2%)
Top other (3.6%)
tt̄ (7.4%)
Diboson(III) (80.1%)
SUSY(50-250Gev) (655)
SUSY(100-200GeV) (742)
SUSY(200-300Gev) (192)
SUSY(150-300GeV) (315)
N(Bkg) = 9049
PNN(50,250)-Output

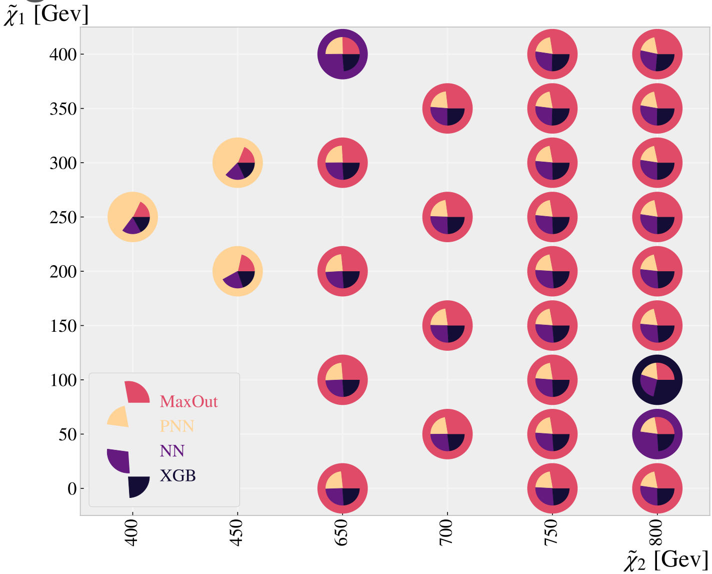

Events
13TeV, 139 fb-1
Other (0.3%)
Diboson(II) (0.3%)
Single top (0.5%)
Top other (0.9%)
Diboson(IIII) (4.4%)
tt̄ (5.3%)
Diboson(III) (10.8%)
Zjets (77.5%)
SUSY(50-250Gev) (811)
SUSY(100-200GeV) (1618)
SUSY(200-300Gev) (383)
SUSY(150-300GeV) (406)
N(Bkg) = 381873
PNN(200,300)-Output


Events
13TeV, 139 fb-1
Diboson(II) (0.1%)
Single top (0.8%)
Other (1.1%)
Top other (2.7%)
Zjets (3.1%)
Diboson(IIII) (3.7%)
tt̄ (7.0%)
Diboson(III) (81.5%)
SUSY(50-250Gev) (627)
SUSY(100-200GeV) (885)
SUSY(200-300Gev) (226)
SUSY(150-300GeV) (310)
N(Bkg) = 10492
PNN(200,300)-Output

Efficiency table

Channel / Parameters	$(50, 250)$	$(100, 200)$	$(150, 300)$	$(200, 300)$
$(50, 250)$	**80.8**%	45.8%	**77.5**%	50.1%
$(200, 300)$	77.3%	**54.6**%	76.3%	**59.0**%

Comparing the sensitivity on a subset of the signal

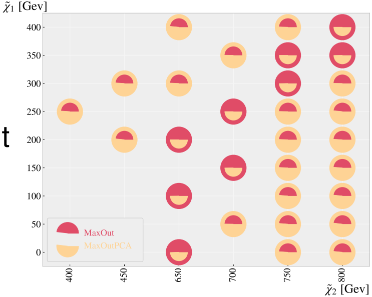


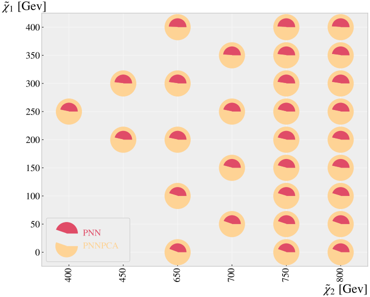
Increasing sensitivity through a PCA

- What is PCA?
- Dimensionality reduction
- Creates new features using linear combination of original features
- Ranks from most to least variance
- This analysis
  - Demand conservation of 99.9% of variance/spread
  - 5 features removed

Compare methods with and without PCA

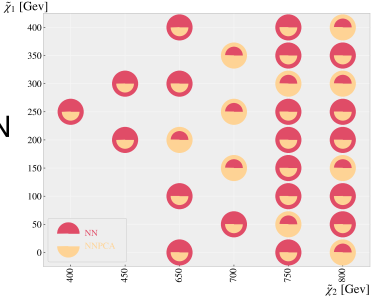
Maxout





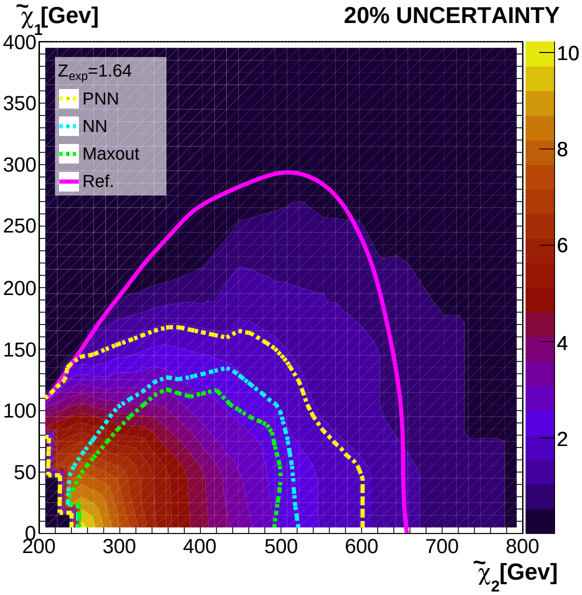
PNN

NN

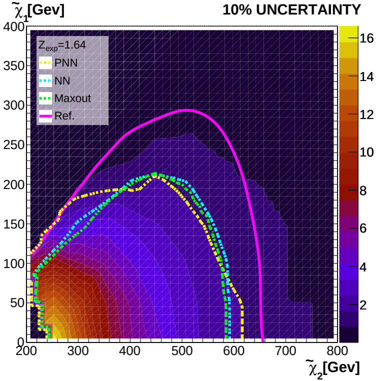


Comparing the methods to previous analysis

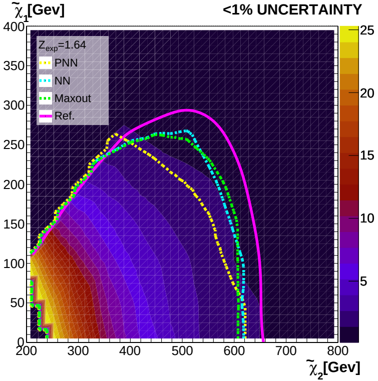
- Compare the expected limits of three best models to analysis made by ATLAS in 2021 [1]
- Introduce flat uncertainty for realistic comparison (20%, 10%, $< 1\%$)
- Include top performing methods
  - Maxout model with PCA
  - PNN with PCA
  - Ordinary dense neural network withou PCA



$\tilde{\chi}_1$[Gev]
20% UNCERTAINTY
Z_{exp}=1.64
PNN
NN
Maxout
Ref.
400
350
300
250
200
150
100
50
0
200
300
400
500
600
700
800
10
8
6
4
2
$\tilde{\chi}_2$[Gev]



$\tilde{\chi}_1$[Gev]
10% UNCERTAINTY
Z_{exp}=1.64
PNN
NN
Maxout
Ref.
$\tilde{\chi}_2$[Gev]



$\tilde{\chi}_1$[Gev]
<1% UNCERTAINTY
Z_{exp}=1.64
PNN
NN
Maxout
Ref.
$\tilde{\chi}_2$[Gev]

Outline

Conclusion & Outlook

1. Including a diverse signal set can improve performance
2. The LWTA layers improve long-term memory via pattern specific pathways
3. All network variants outperformed default settings of XGBoost
4. PCA increased sensitivity of PNN and maxout model in original signal set
5. None of the networks extended expected limit past previous ATLAS analysis
6. PNN exhibited bias towards lower masses, wheras maxout model achieved a more balanced sensitivity
7. LWTA layer's increase in long-term memory is promising in future analysis where higher masses are studied

References I

ATLAS Collaboration.
'Search for chargino–neutralino pair production in final states with three leptons and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector'.
http://arxiv.org/abs/2106.01676

UiO : Fysisk institutt
Det matematisk-naturvitenskapelige fakultet

William Hirst

Application of Supervised Machine Learning to the Search for New Physics in ATLAS data

A Study of Ordinary Dense, Parameterized and Ensemble Networks and their Application to High Energy Physics