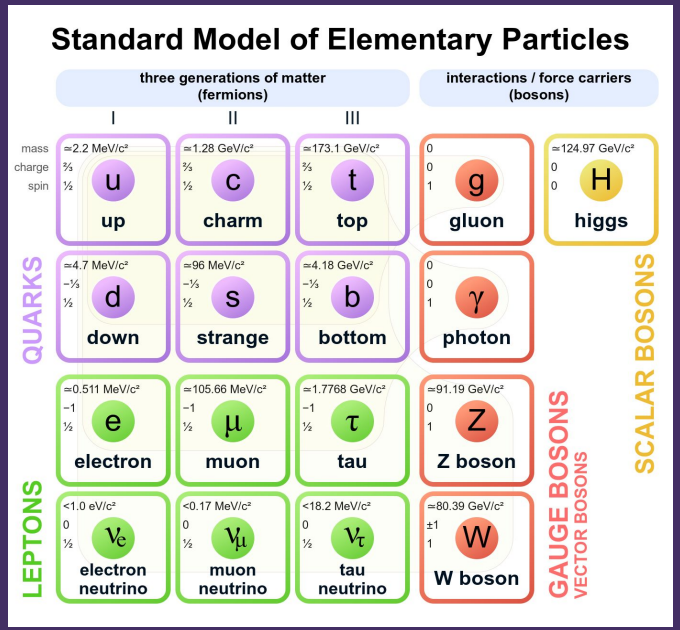
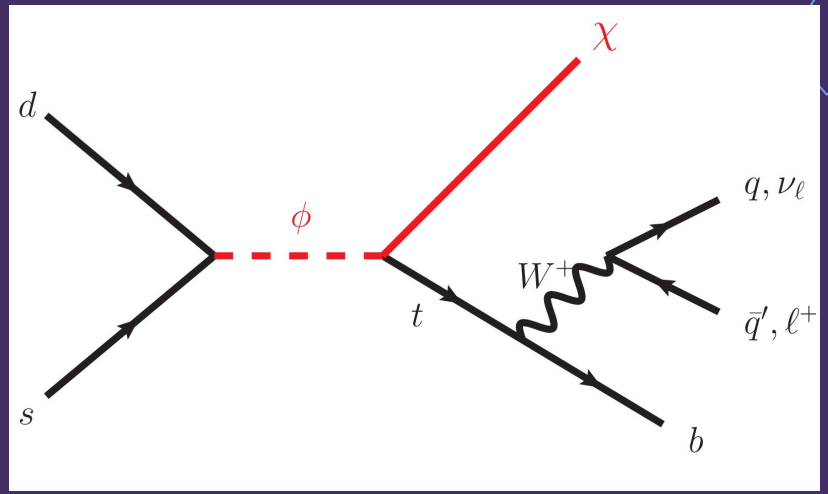
Advanced data analysis in the search for dark matter associated with quark top on the LHC at CERN

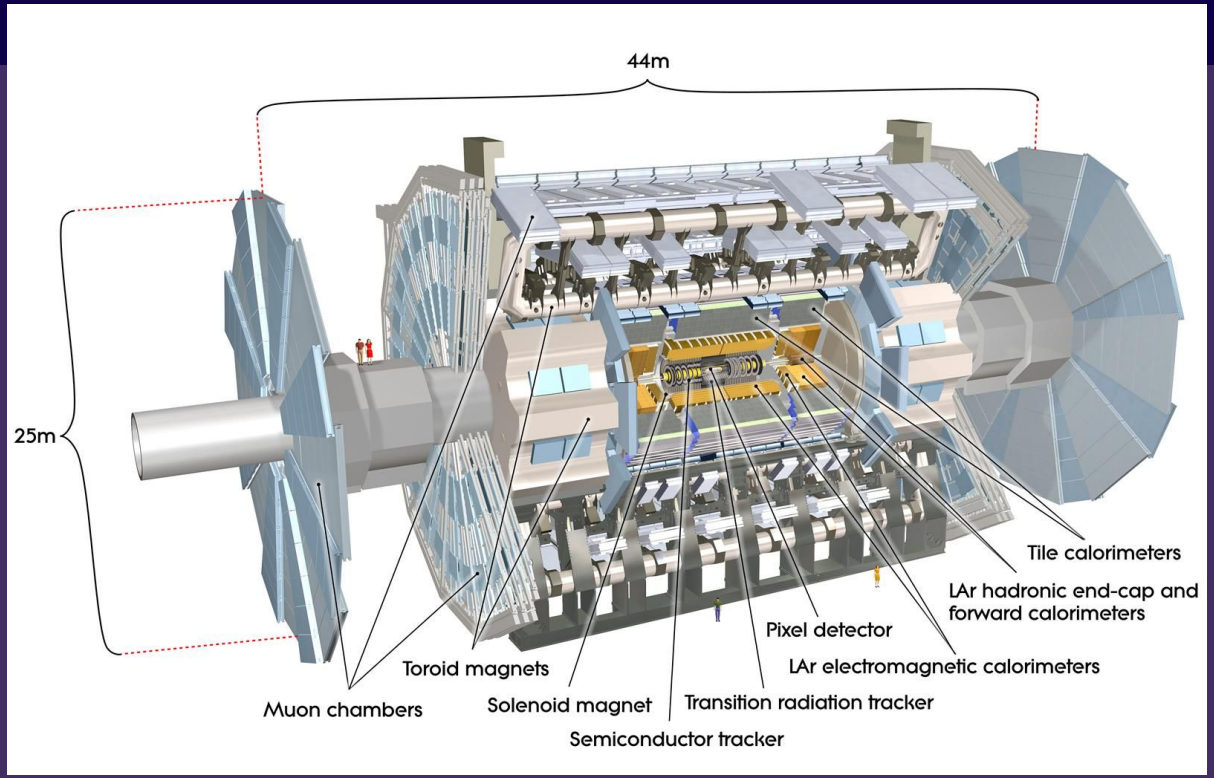
By Miguel Peixoto
1º year student - MIEFIS,UM

Introduction





Introduction – ATLAS



http://atlas.cern/discover/detector

Data Exploration

“The first step to training a neural net is to not touch any neural net code at all and instead begin by thoroughly inspecting your data.”

- Andrej Karpathy

Data Exploration

	GenJet02_Multi	GenJet021_PT	GenJet022_PT	GenJet023_PT	GenJet024_PT	GenJet025_PT	GenJet021_Mass	GenJet022_Mass	GenJet023_Mass	GenJet024_Mass	...
count	750000.000000	750000.000000	750000.000000	750000.000000	750000.000000	750000.000000	750000.000000	750000.000000	750000.000000	750000.000000	...
mean	3.585093	218.988492	114.924698	56.917692	23.268296	8.619090	13.029389	8.281990	5.031884	2.363067	...
std	1.293321	91.608875	56.953163	42.773353	30.644702	19.050968	6.831858	4.276064	3.640029	3.059804	...
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-0.000037	-0.000065	-0.000086	-0.000106	...
25%	3.000000	158.540897	76.096834	30.780883	0.000000	0.000000	8.396796	5.474636	2.776319	0.000000	...
50%	3.000000	211.244629	110.186356	52.686663	0.000000	0.000000	11.539050	7.559322	4.887335	0.000000	...
75%	4.000000	269.106812	146.515869	80.774464	40.902470	0.000000	16.014464	10.325954	7.075413	4.442586	...
max	14.000000	2293.890869	2292.538330	720.844849	505.682037	314.503204	181.089874	168.190033	52.145088	49.039822	...

8 rows × 463 columns

Statistics of an non processed signal sample

	CaloJet08_Multi	CaloJet081_PT	CaloJet082_PT	CaloJet083_PT	CaloJet084_PT	CaloJet085_PT	CaloJet081_Mass	CaloJet082_Mass	CaloJet083_Mass	CaloJet084_Mass	...	E
0	2	453.641968	45.273438	0.000000	0.000000	0.0	166.171173	7.901662	0.000000	0.00000	...	
1	3	174.681198	65.354713	33.970352	0.000000	0.0	51.266632	8.033169	12.794903	0.00000	...	
2	2	362.785095	40.609589	0.000000	0.000000	0.0	143.886475	17.862165	0.000000	0.00000	...	
3	3	193.785385	115.942780	46.830433	0.000000	0.0	68.541824	29.208618	16.798588	0.00000	...	
4	4	275.712830	137.585297	37.642029	30.241388	0.0	57.443779	21.555725	9.159877	1.56638	...	

5 rows × 71 columns

The first events of a signal sample

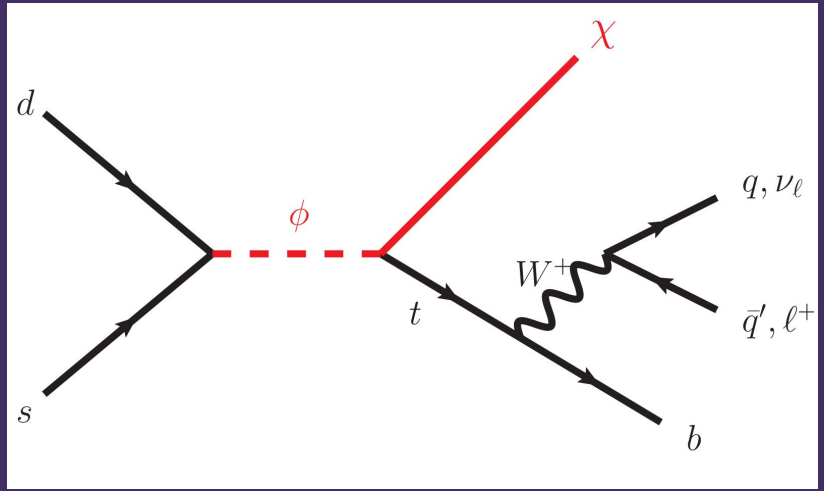
Data Pre-Processing - Data Cuts

Background Data:

- –0 Leptons drop ratio: 0.3438 %
- –At least 1 B-Tag: 81.2528 %
- –At least 1 FatJet: 2.2760 %
- –Total drop rate: 81.3892 %

Signal Data:

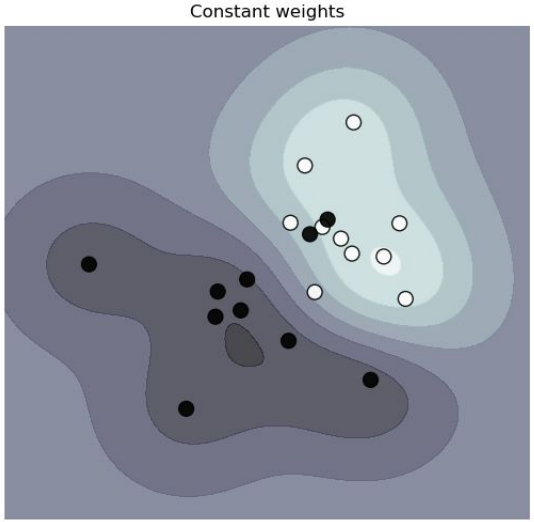
- –0 Leptons drop ratio: 0.5268 %
- –At least 1 B-Tag: 34.2679 %
- –At least 1 FatJet: 0.0074 %
- –Total drop rate: 34.4409 %

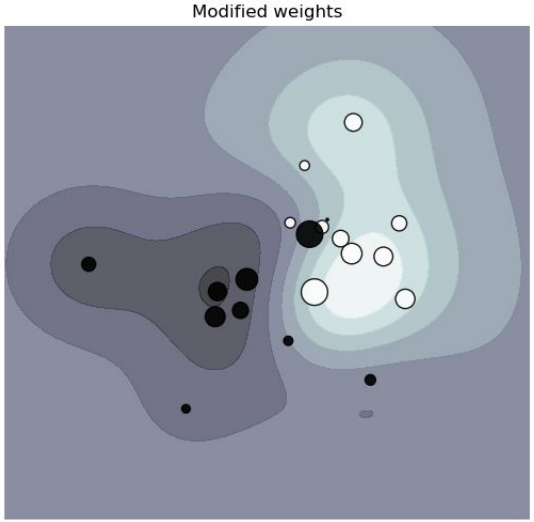


Data Pre-Processing - Sample and Class Weights

Bring balance to the force!

Data Pre-Processing – Sample and Class Weights

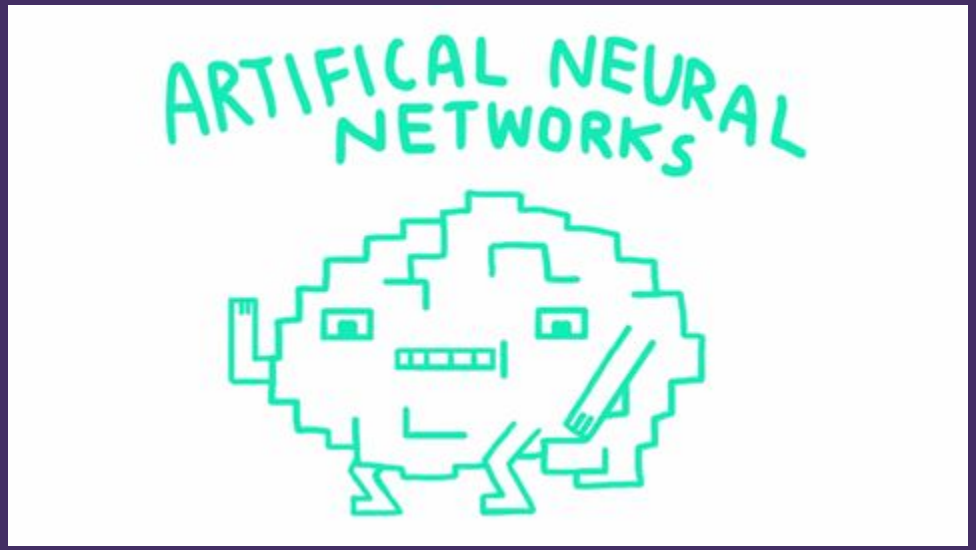




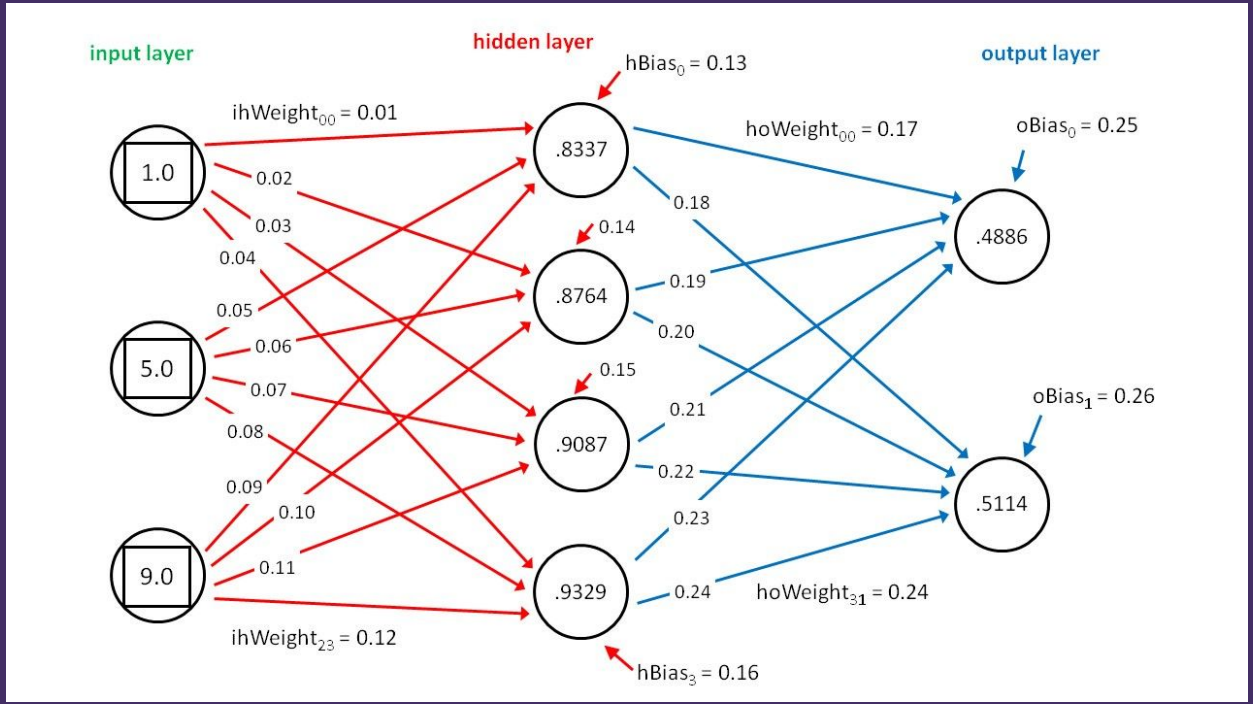
Data Pre-Processing – Training, Validation and Test Set

- Training Set - Will be used to train the model,
- Validation Set - Will provide an unbiased evaluation of a model fit on the training dataset using the loss and metrics.
- Test Set - is used to provide an unbiased evaluation of how well the model generalizes to new data.

Deep Neural Networks (DNN)



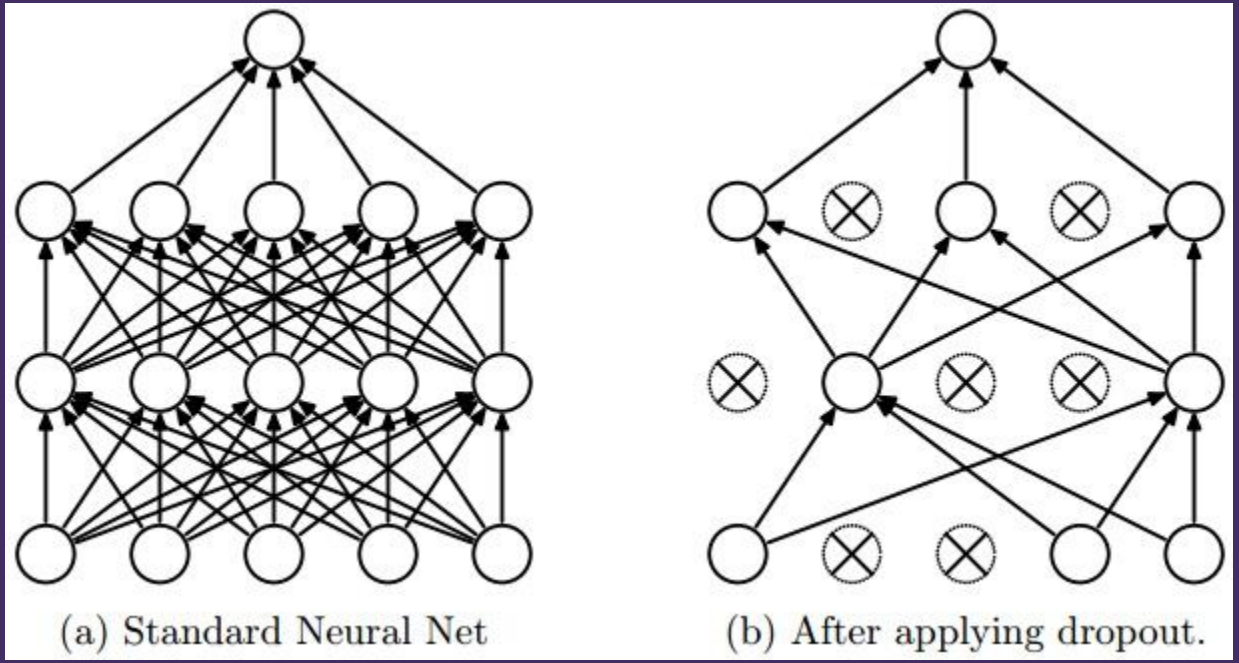
Deep Neural Networks (DNN)



The Model

```
Model: "functional_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 69)]              0
_________________________________________________________________
dense (Dense)                (None, 100)               7000
_________________________________________________________________
dropout (Dropout)            (None, 100)               0
_________________________________________________________________
dense_1 (Dense)              (None, 100)               10100
_________________________________________________________________
dropout_1 (Dropout)          (None, 100)               0
_________________________________________________________________
dense_3 (Dense)              (None, 30)                3030
_________________________________________________________________
dropout_3 (Dropout)          (None, 30)                0
_________________________________________________________________
dense_4 (Dense)              (None, 1)                 31
=================================================================
Total params: 20,161
Trainable params: 20,161
Non-trainable params: 0
```

The Model – MC Dropout

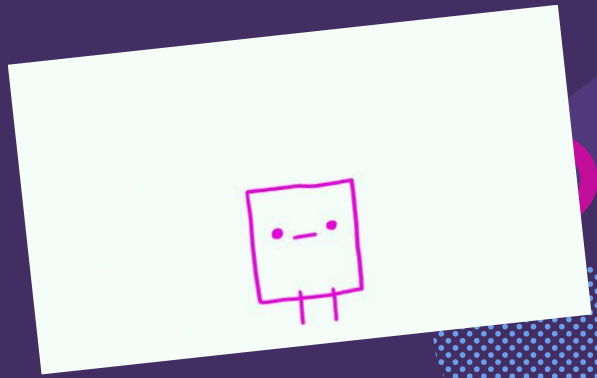


(a) Standard Neural Net

(b) After applying dropout.

The Model – Callbacks

- **Early Stopping** - Max of 500 epochs and a patience monitoring the loss function of 30
- **Tensorboard** - Enables the logging of training metrics and see their improvement though out training.
- **Model Checkpoint** - only the best performing model iteration on the validation data would be keptas our final model.



Model Evaluation – No Dropout vs Dropout vs MC Dropout

	ROC Score	Improvement
Without Dropout	0.8990490403416795	-
With Dropout	0.9934078794963115	10,5%
With MC Dropout	0.9947038946272739	10,6%

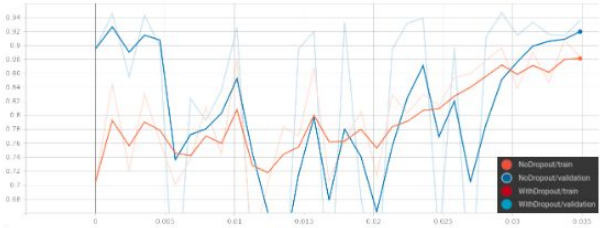


Figure 4: Evolution of the area under the ROC curve for the model without dropout.

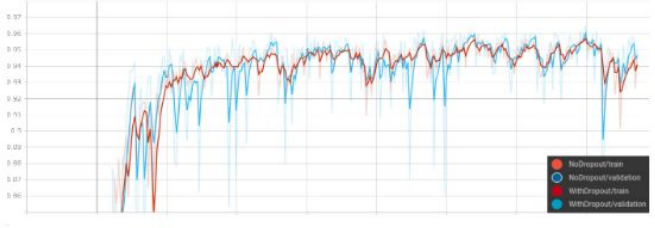
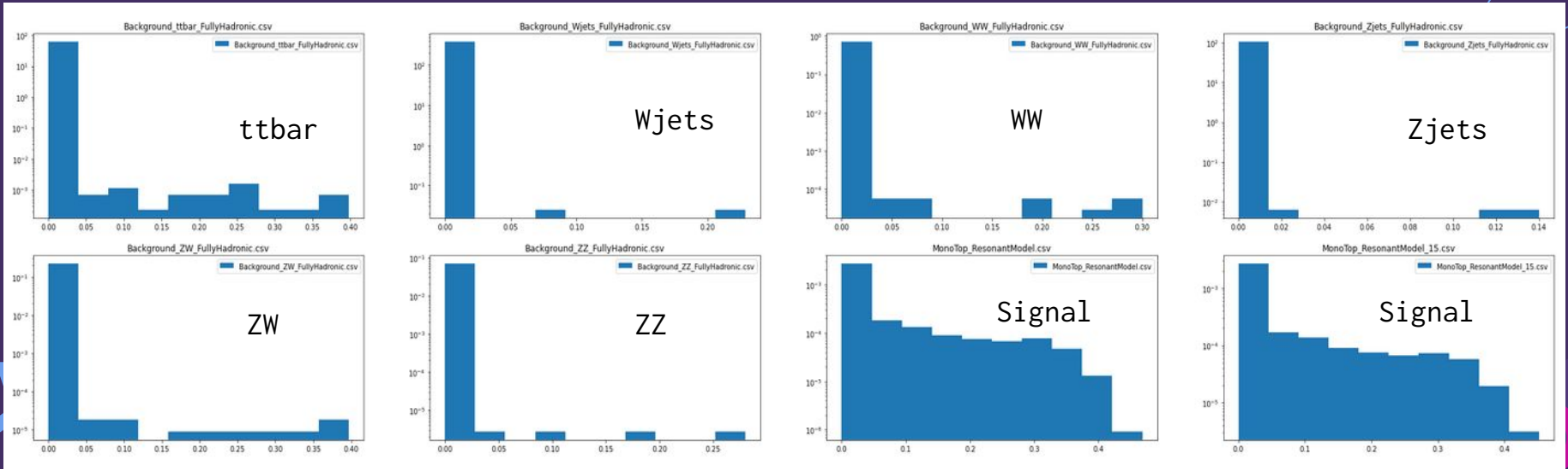


Figure 5: Evolution of the area under the ROC curve for the model with dropout implementation

Model Evaluation – Standard Deviation



Conclusion

Thanks!

Any questions?

You can find me at miguelpeixoto457@gmail.com or miguelpeixoto.net

Link to my repository