

# Pruning for Deep Neural Network compression: Constant vs. polynomial schedules

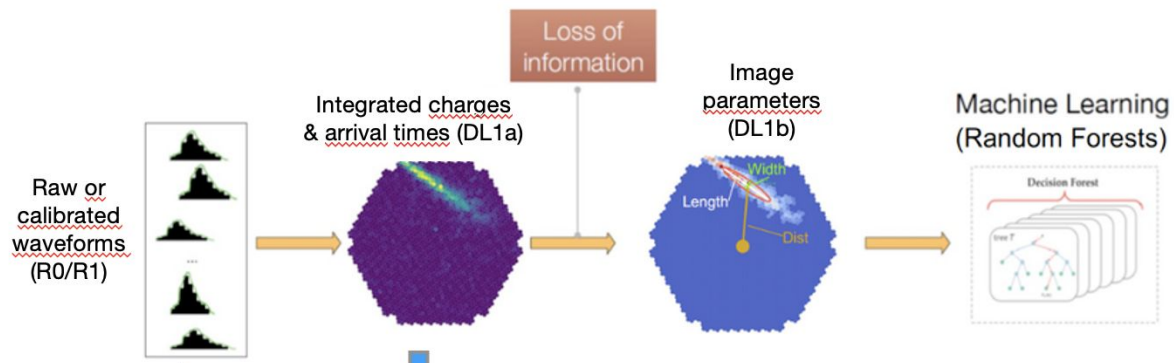
Dafne Martín Domínguez ([dafmar02@ucm.es](mailto:dafmar02@ucm.es))

D. Nieto, J. Buces, A. Cerviño,  
UCM-GAE

4th November 2025

# State of the Art: Event reconstruction

- Supervised Learning
- Hillas parameters
- Random Forest



*Modificado de (Grespan et al., 2022)*

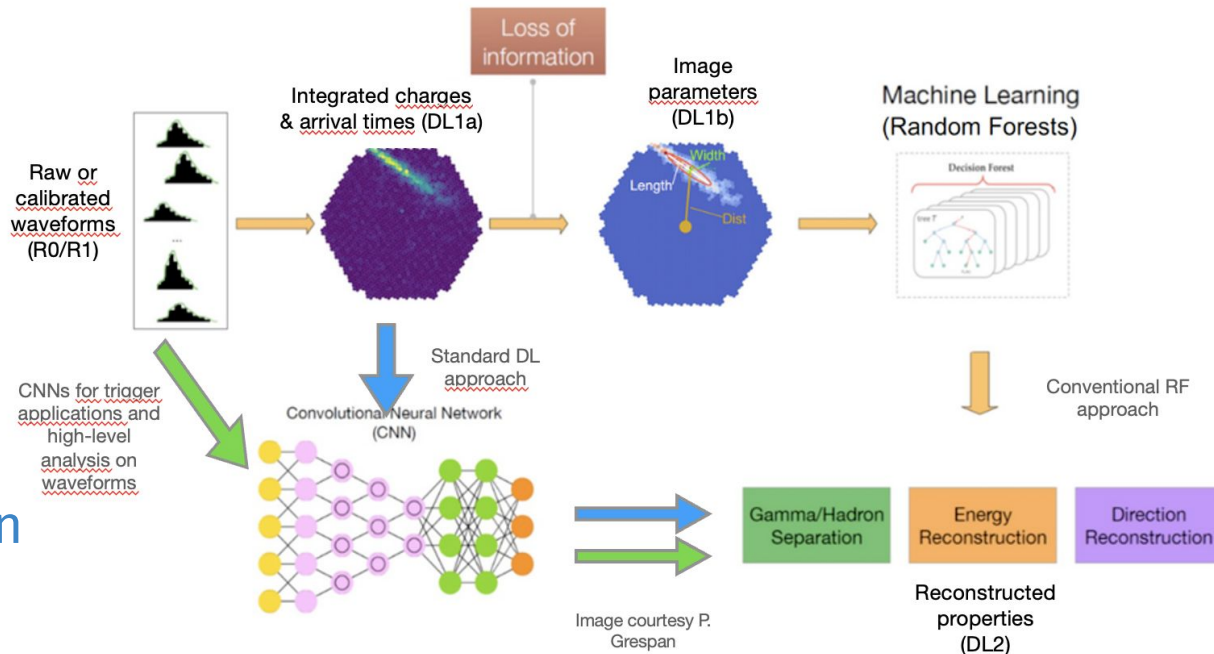
Loss of information



# State of the Art: Event reconstruction

- Deep Learning
- CNN-based architecture

Use all the information



(Grespan et al., 2022)

# Hypothesis and objectives: Model compression

## *Explore model compression for ResNet-36*

### *Objectives*

- Identify and explore model compression techniques
- Implement:
  - Quantisation
  - Pruning
- Evaluate degradation in performance
  - Metrics
  - IRFs
- Evaluate model size reduction

# Hypothesis and objectives: Model compression

## *Explore model compression for ResNet-36*

### *Objectives*

- Explore model compression techniques
- Implement:
  - Pruning ✓
  - Quantisation
- Evaluate degradation in performance
  - Metrics
  - IRFs
- Evaluate model size reduction

# Hypothesis and objectives: Model compression

## *Explore model compression for ResNet-36*

### *Objectives*

- Explore model compression techniques
- Implement:
  - Pruning ✓
  - Quantisation
- Evaluate degradation in performance
  - Metrics
  - IRFs
- Evaluate model size reduction

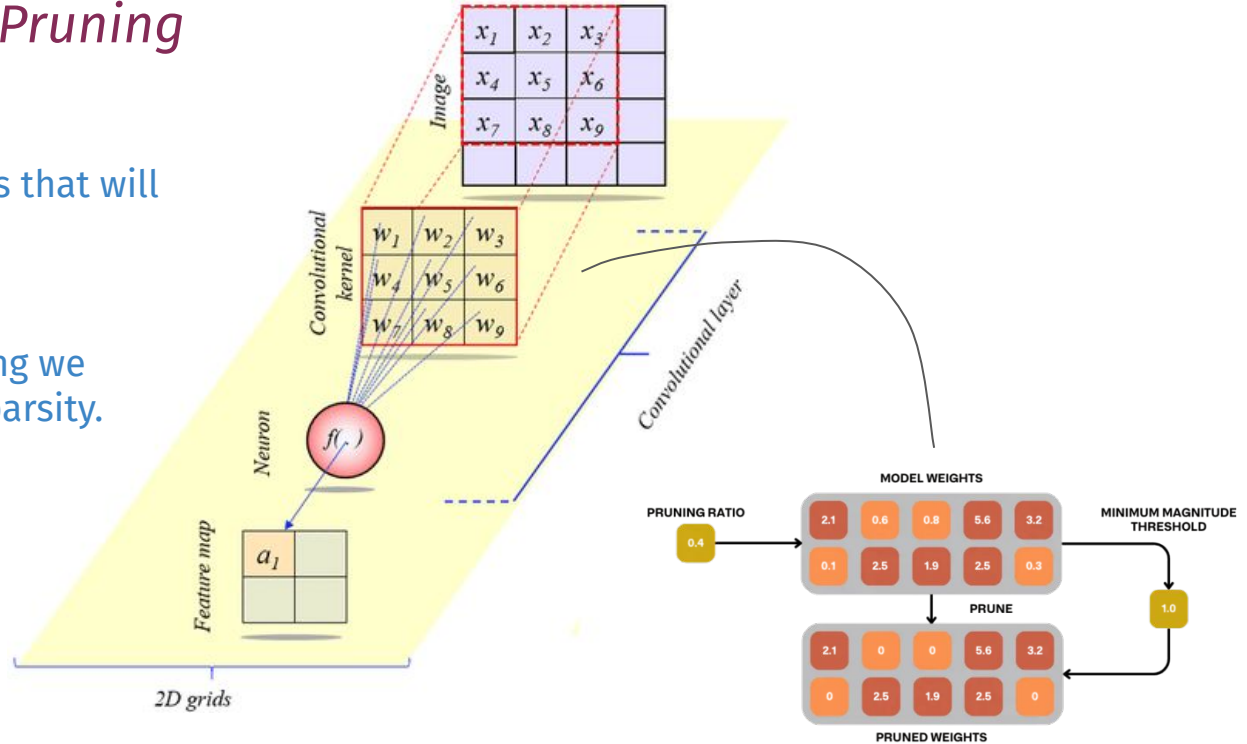
### *Note!*

Currently on offline model: extension to trigger will come soon

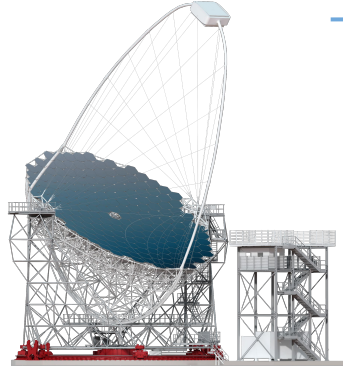
# Methodology

## Magnitude- based Pruning

- Sparsity = How much?  
Percentage of weights that will be zeroed out.
- Schedule = How?  
How quickly in training we obtain the desired sparsity.
  - Constant
  - Polynomial

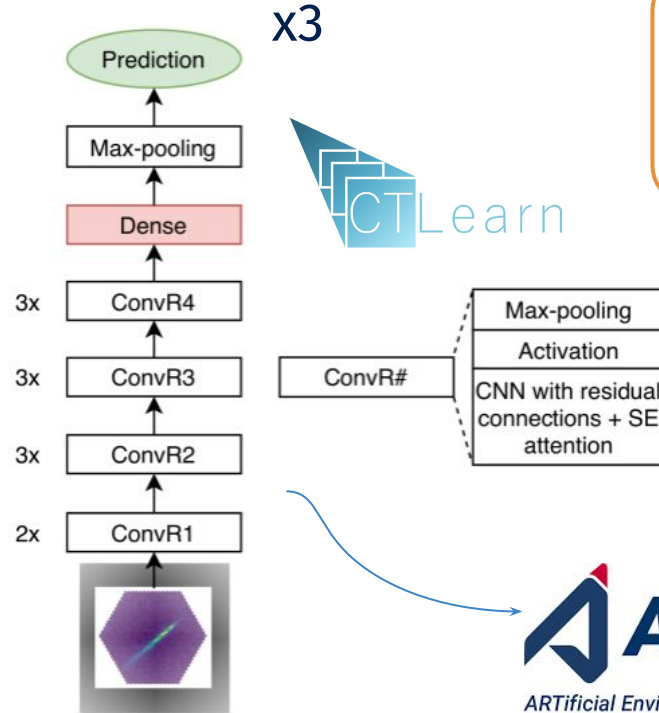
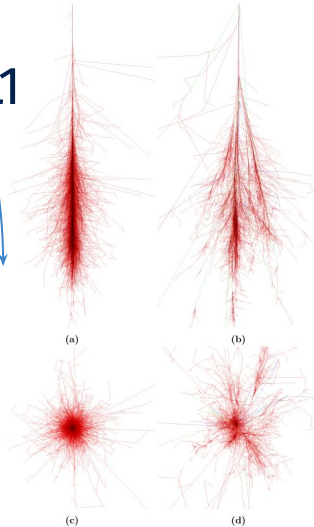


# Methodology



Implement pruning  
based magnitude

DL1



Evaluation:

- Metrics
- IRFs

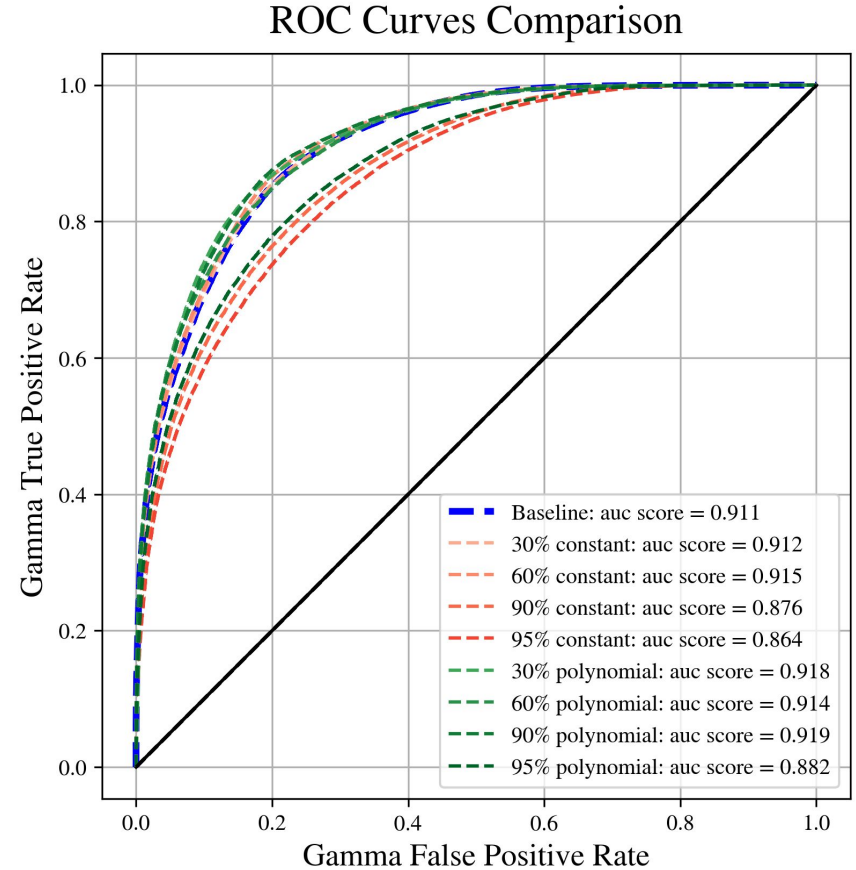
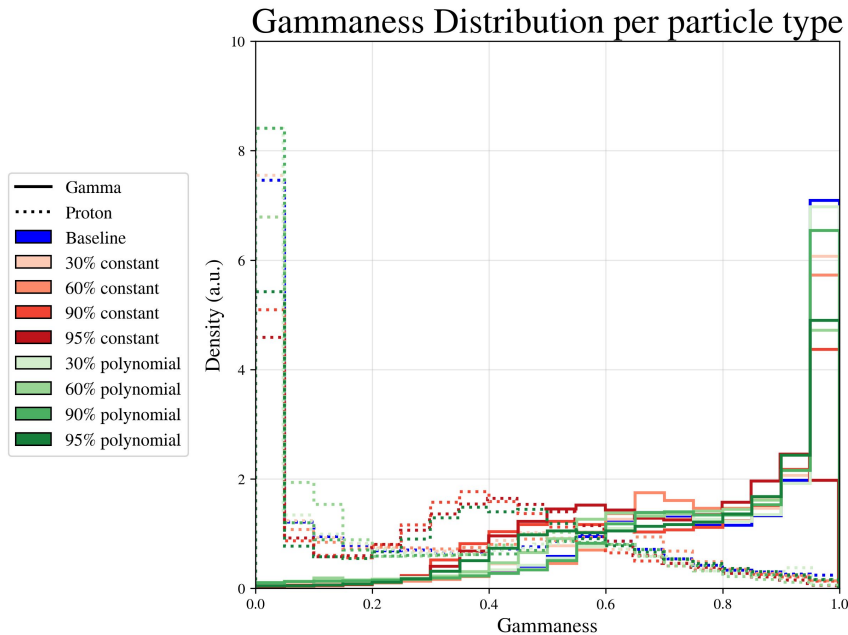


## Results: Classification task

- Compare with Baseline
- Efficiency cut on 70%
- *Gammaness* histograms
- ROC curves: performance of a binary classifier as a function of gammaness
- Area under curve (AUC)

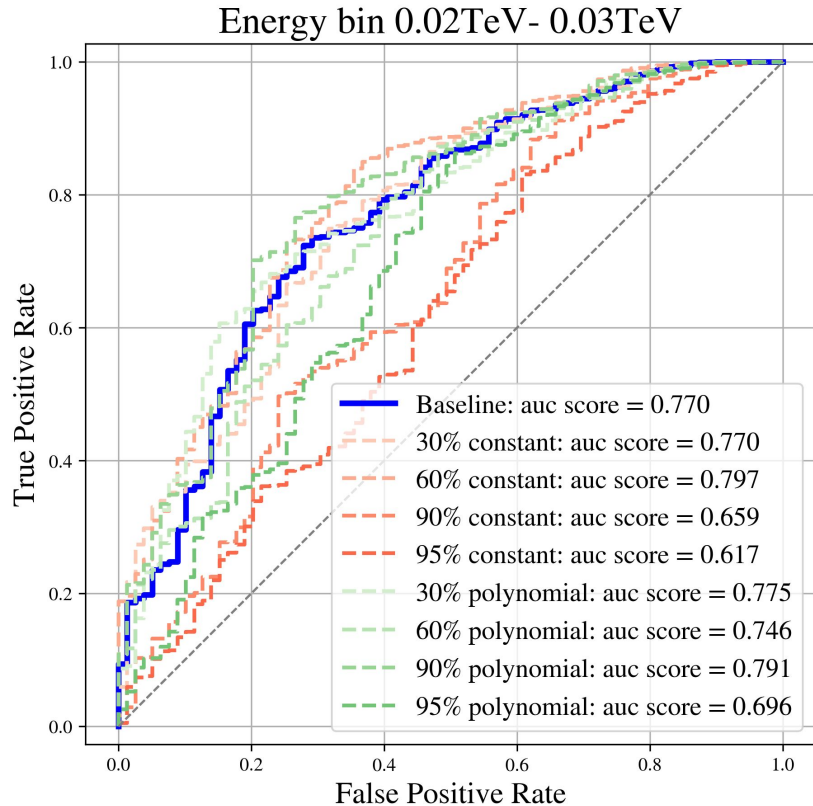


# Results: Classification task

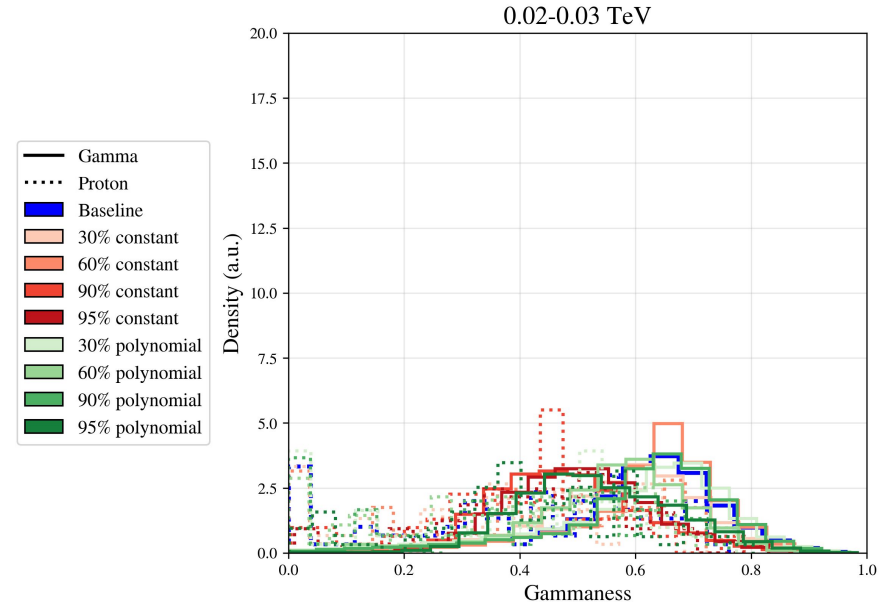




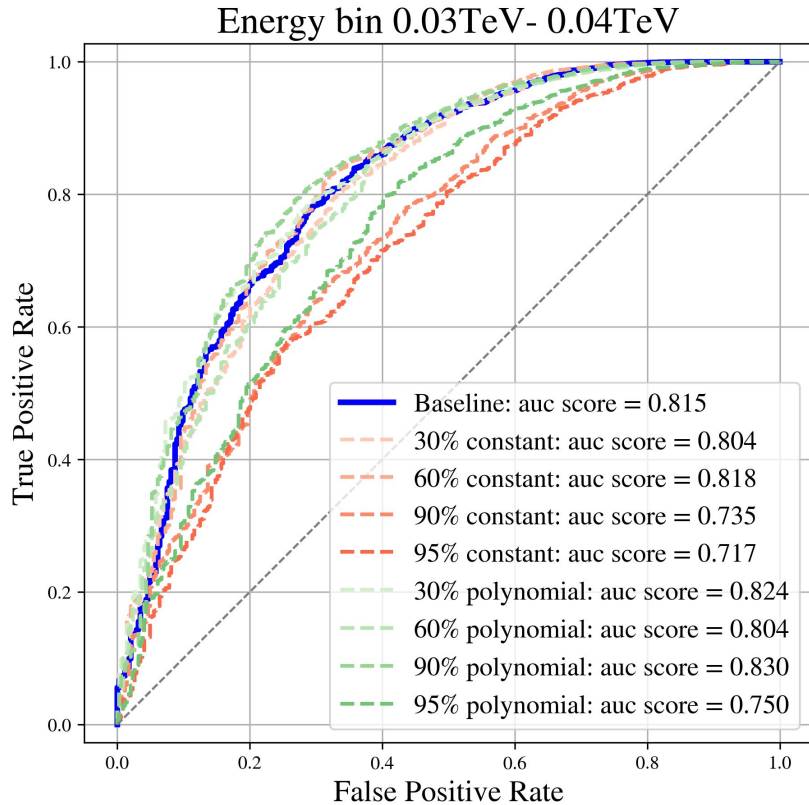
# Results: Classification task



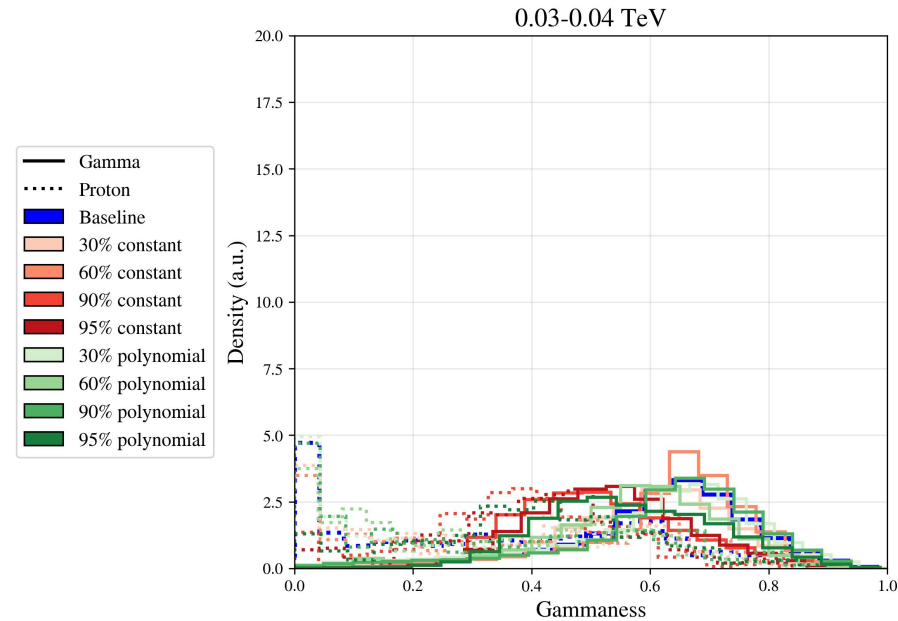
## Gammaness Distribution Comparison per Energy Bin



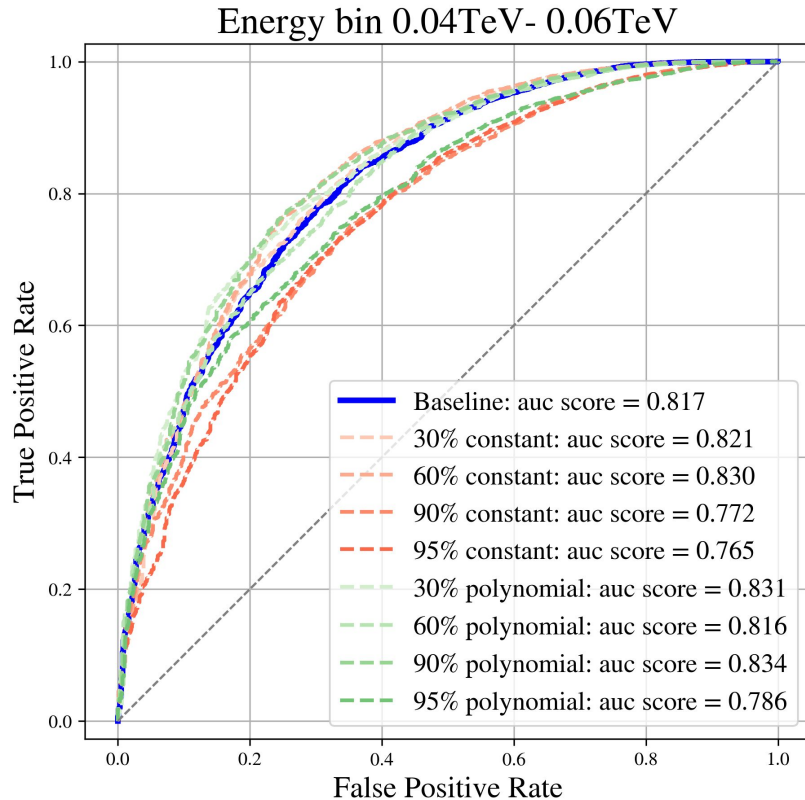
# Results: Classification task



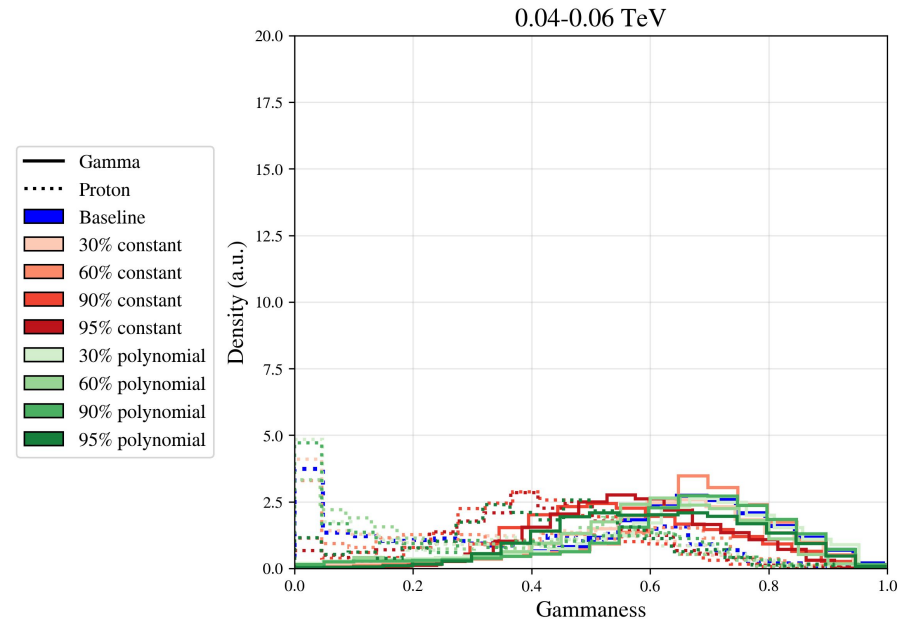
## Gammaness Distribution Comparison per Energy Bin



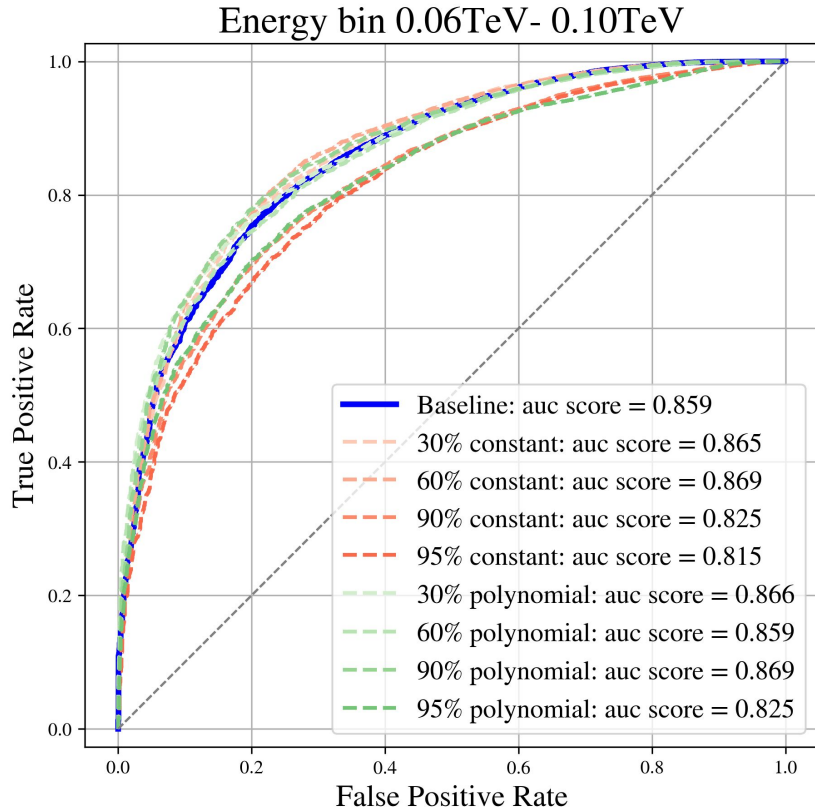
# Results: Classification task



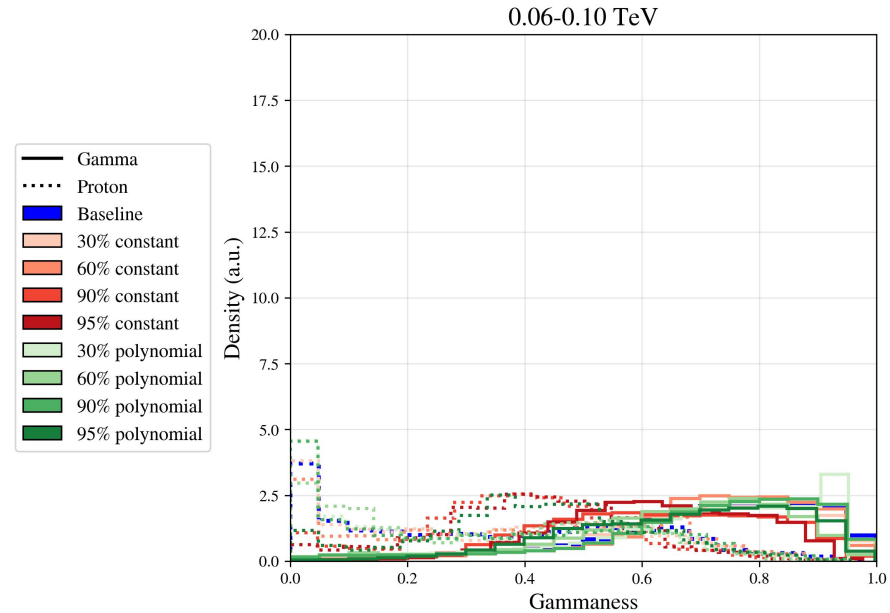
## Gammaness Distribution Comparison per Energy Bin



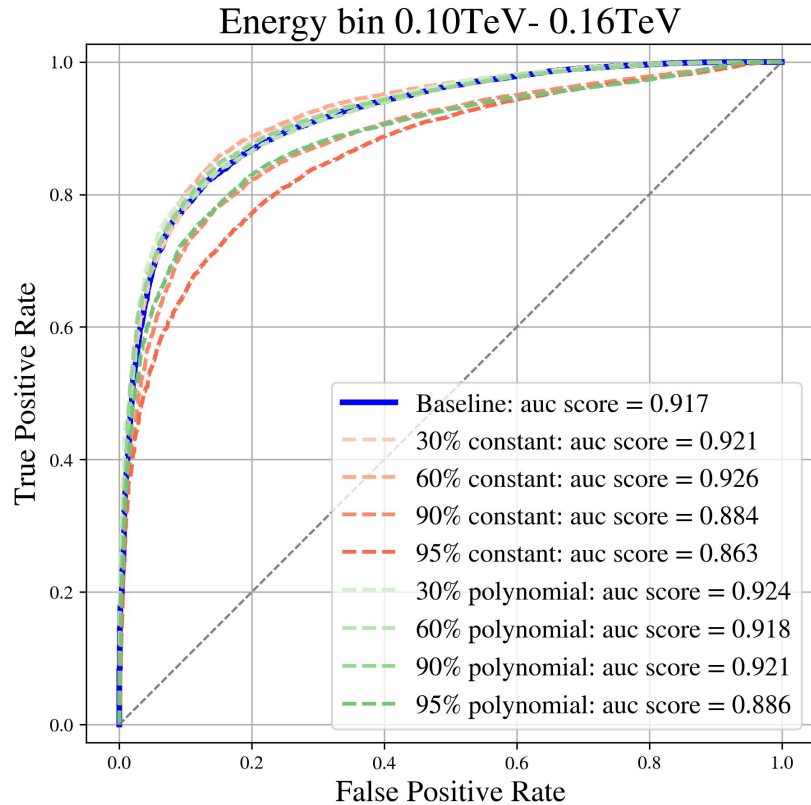
# Results: Classification task



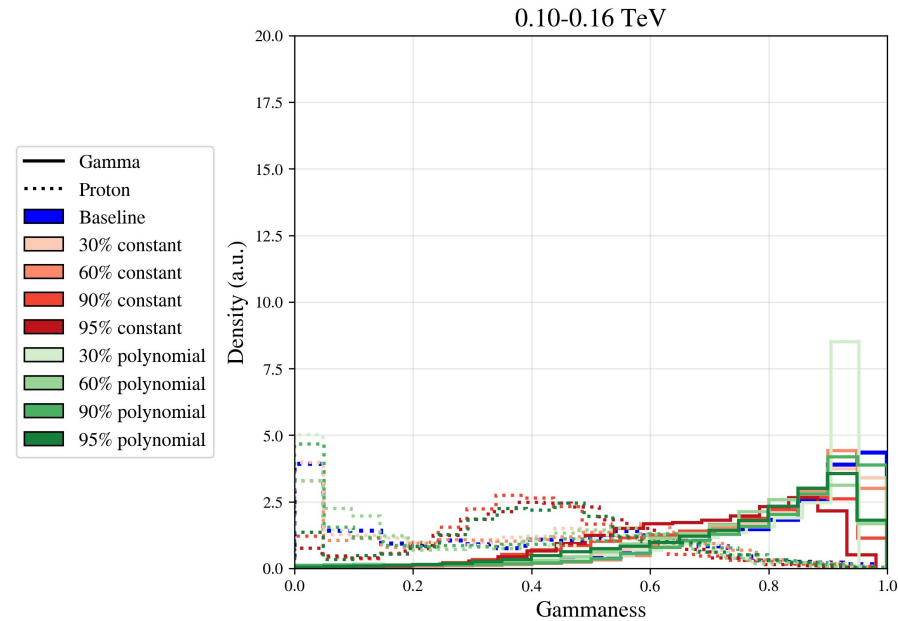
## Gammaness Distribution Comparison per Energy Bin



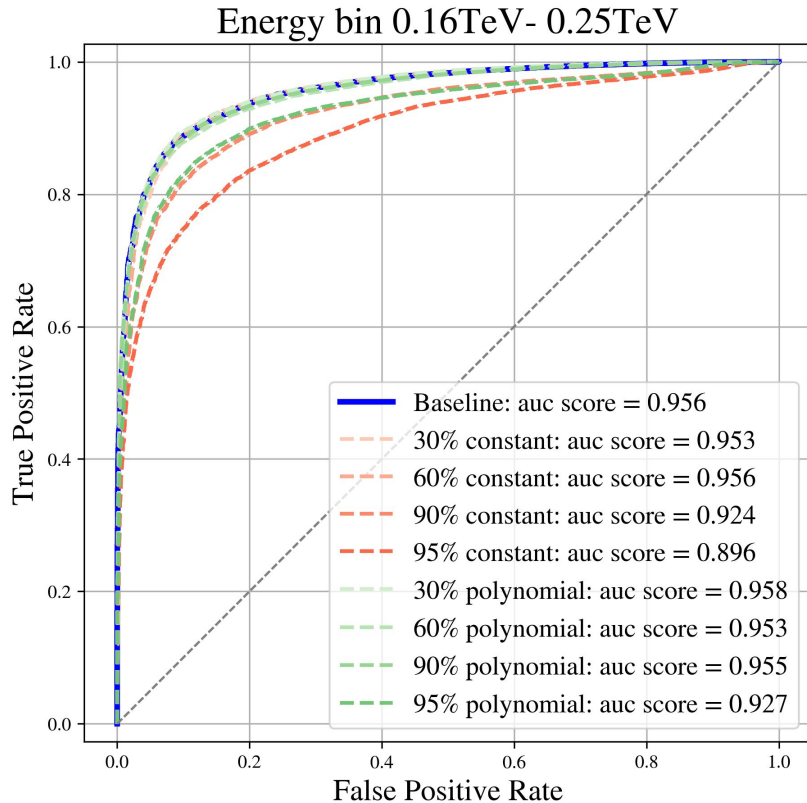
# Results: Classification task



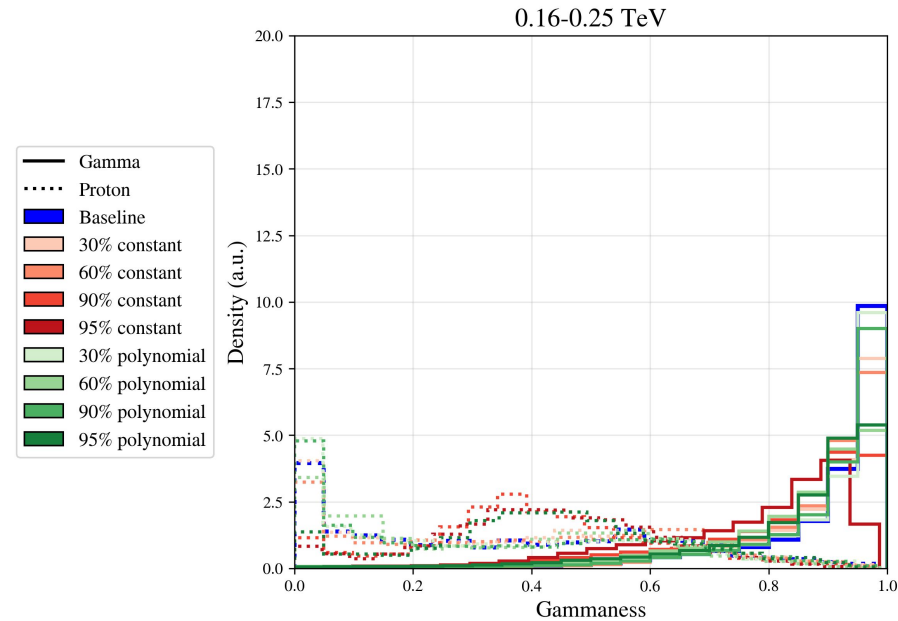
## Gammaness Distribution Comparison per Energy Bin



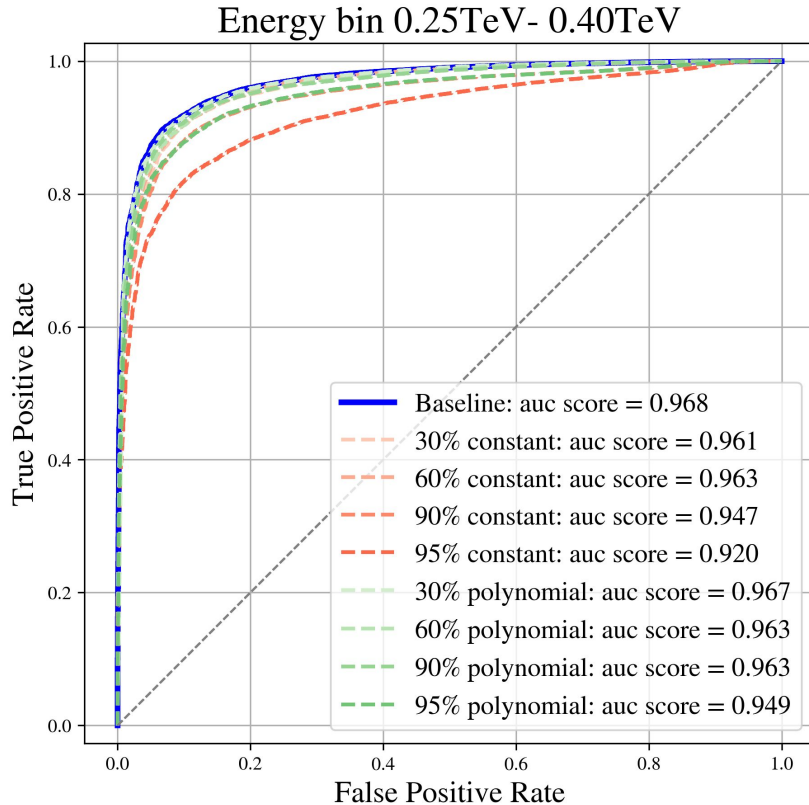
# Results: Classification task



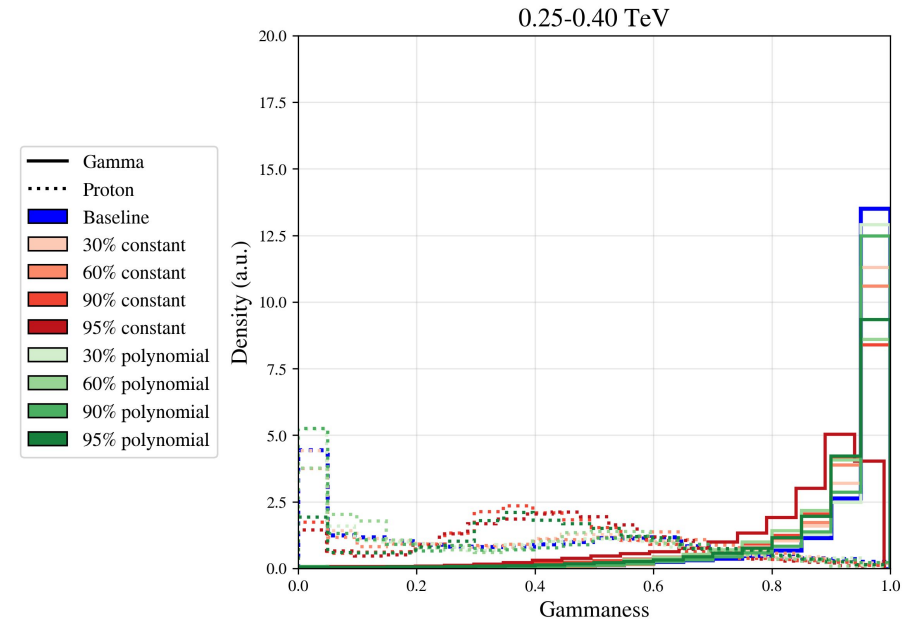
## Gammaness Distribution Comparison per Energy Bin



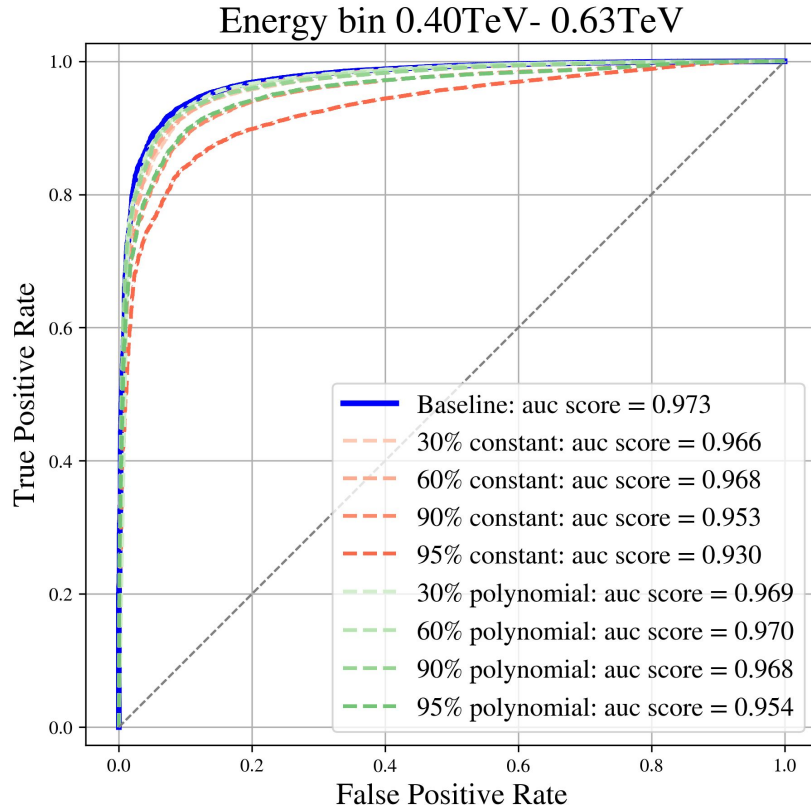
# Results: Classification task



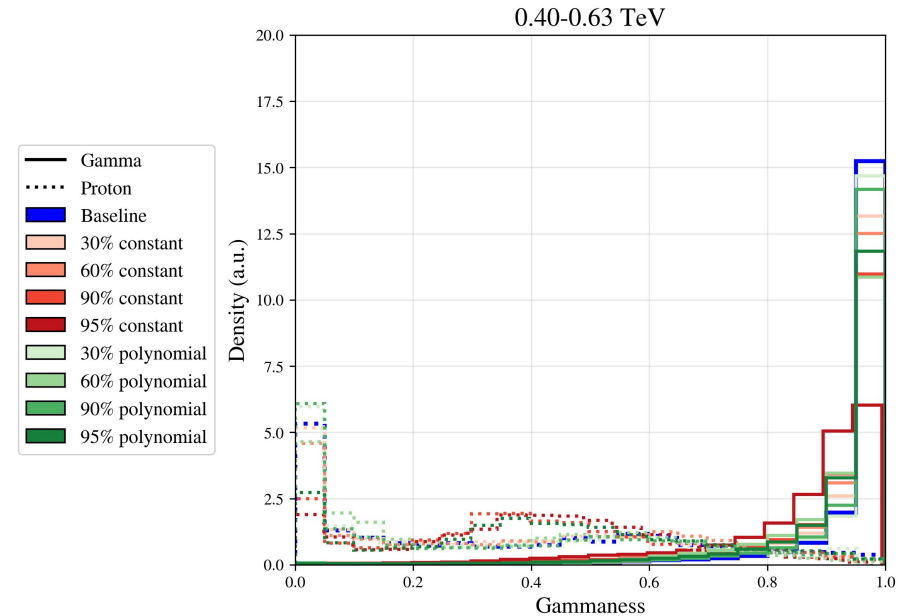
## Gammaness Distribution Comparison per Energy Bin



# Results: Classification task

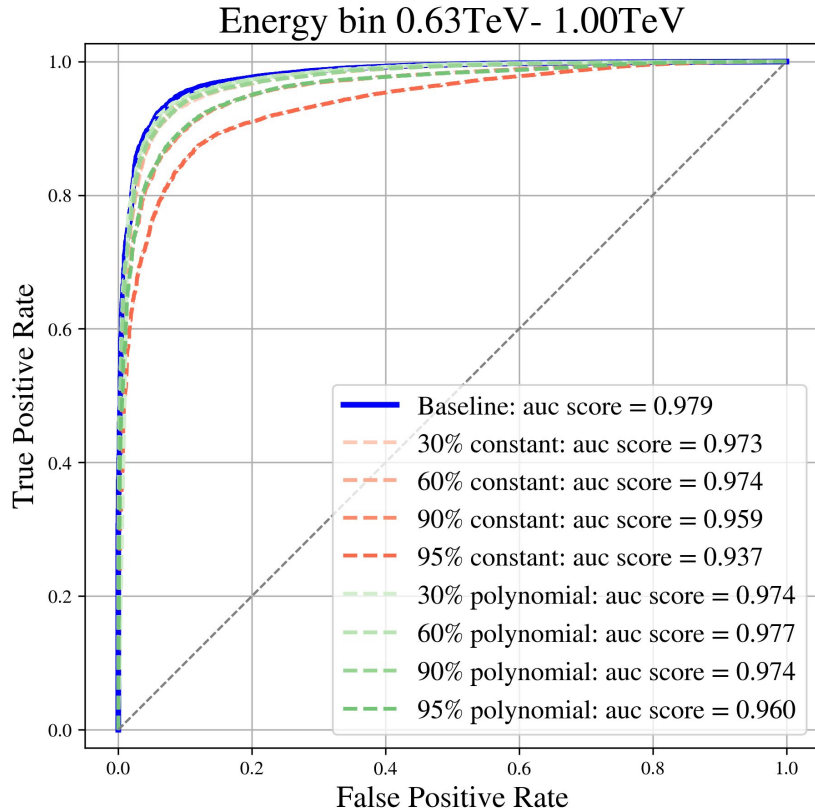


## Gammaness Distribution Comparison per Energy Bin

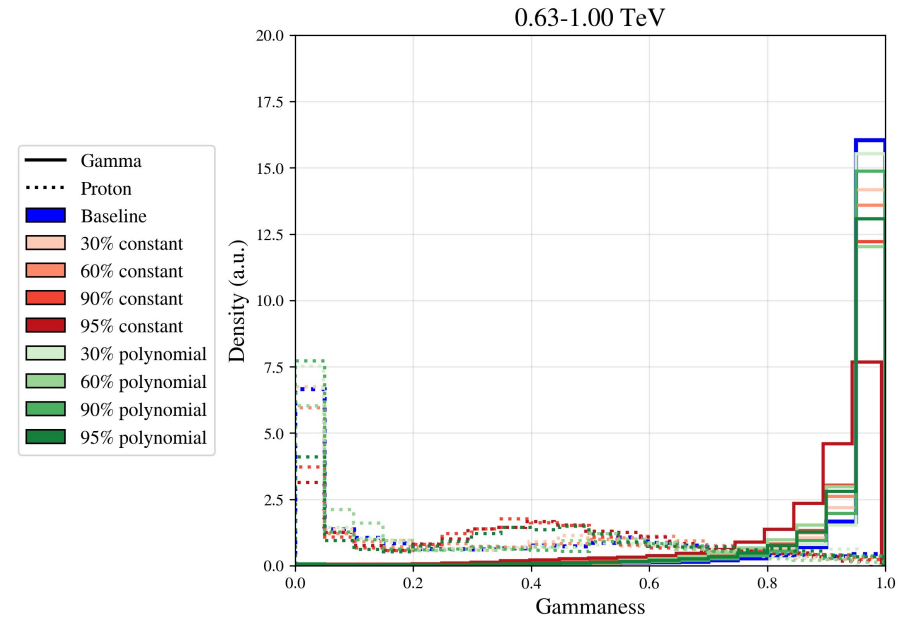




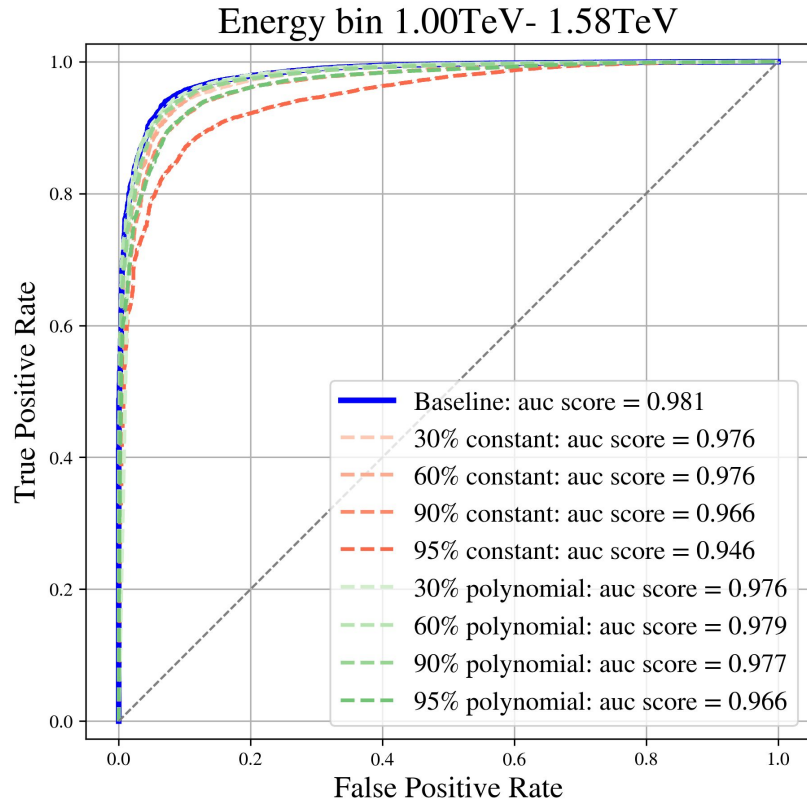
# Results: Classification task



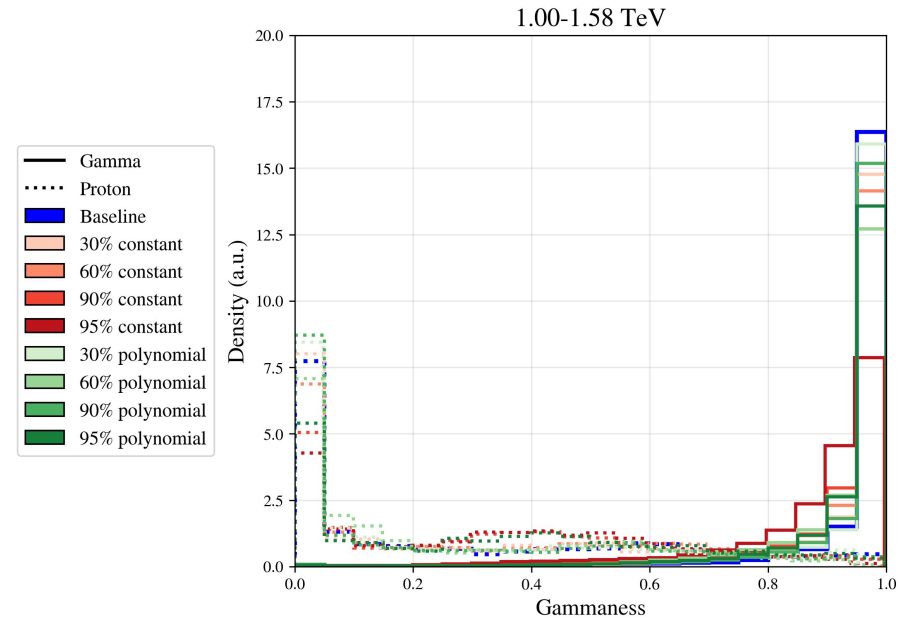
## Gammaness Distribution Comparison per Energy Bin



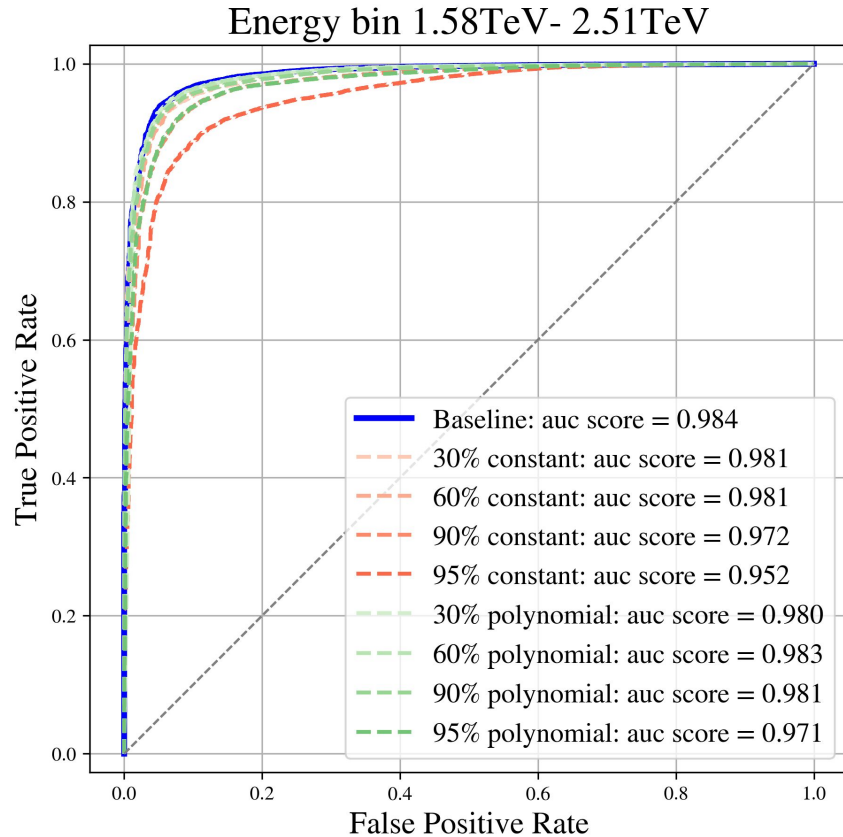
# Results: Classification task



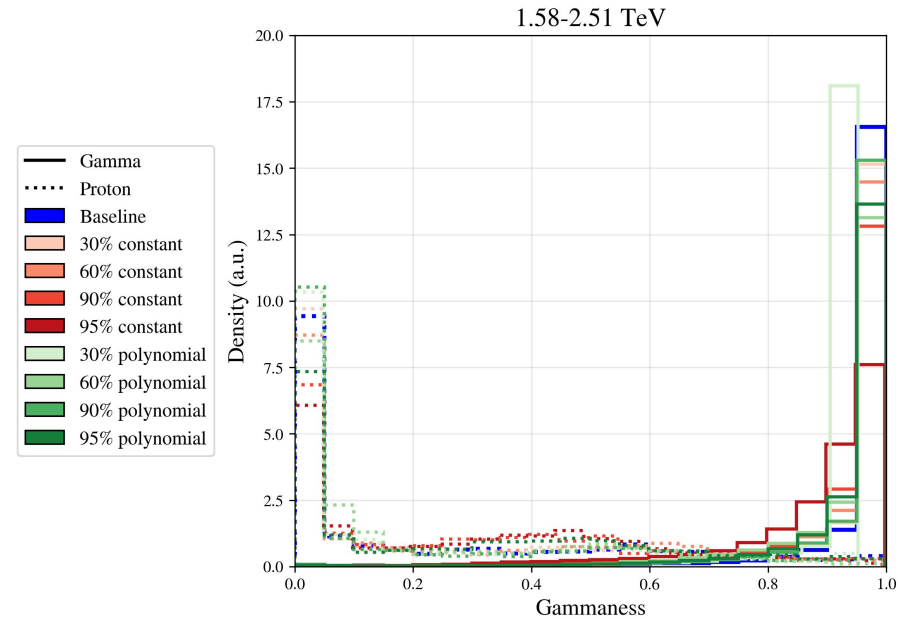
## Gammaness Distribution Comparison per Energy Bin



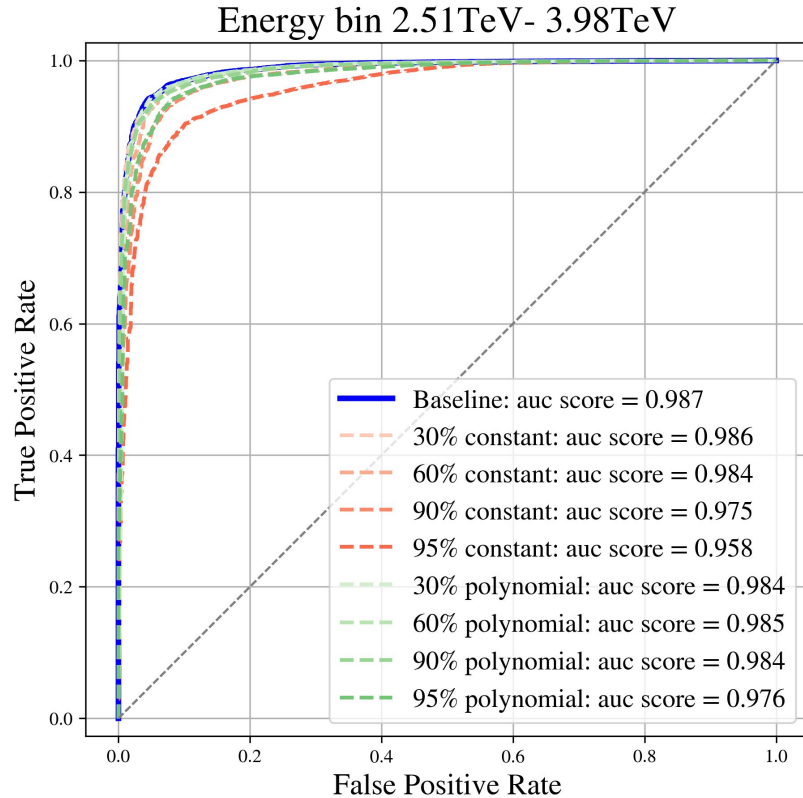
# Results: Classification task



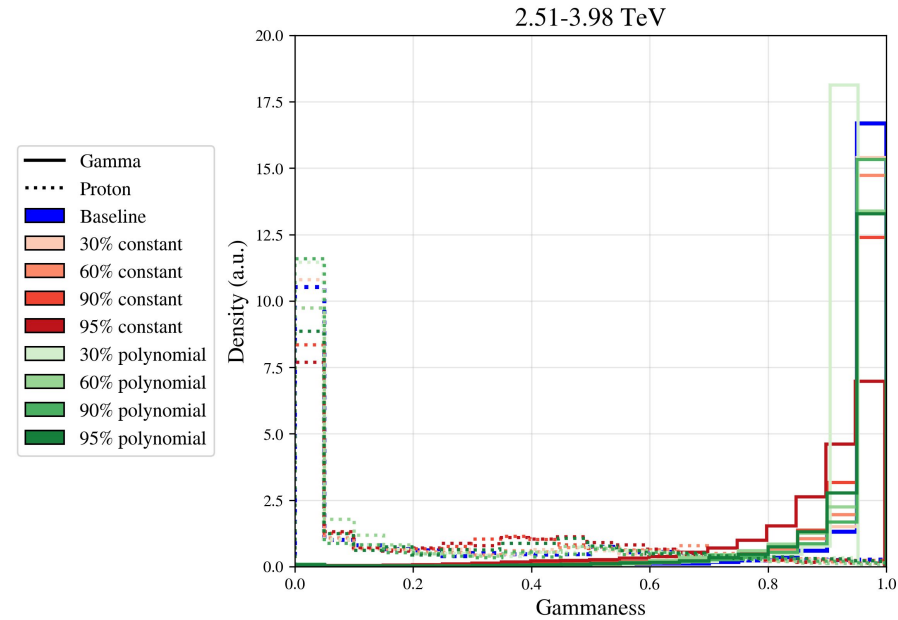
## Gammaness Distribution Comparison per Energy Bin



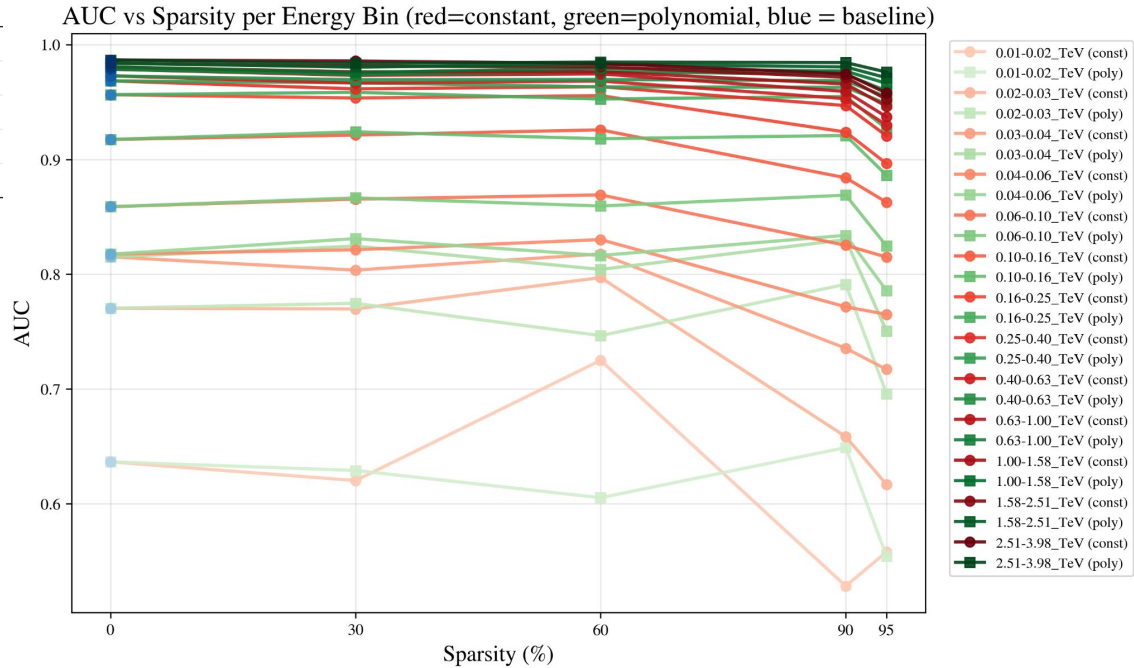
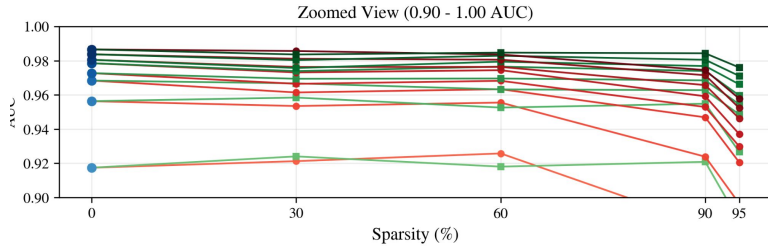
# Results: Classification task



## Gammaness Distribution Comparison per Energy Bin



# Results: Classification task



# Results: Regression task

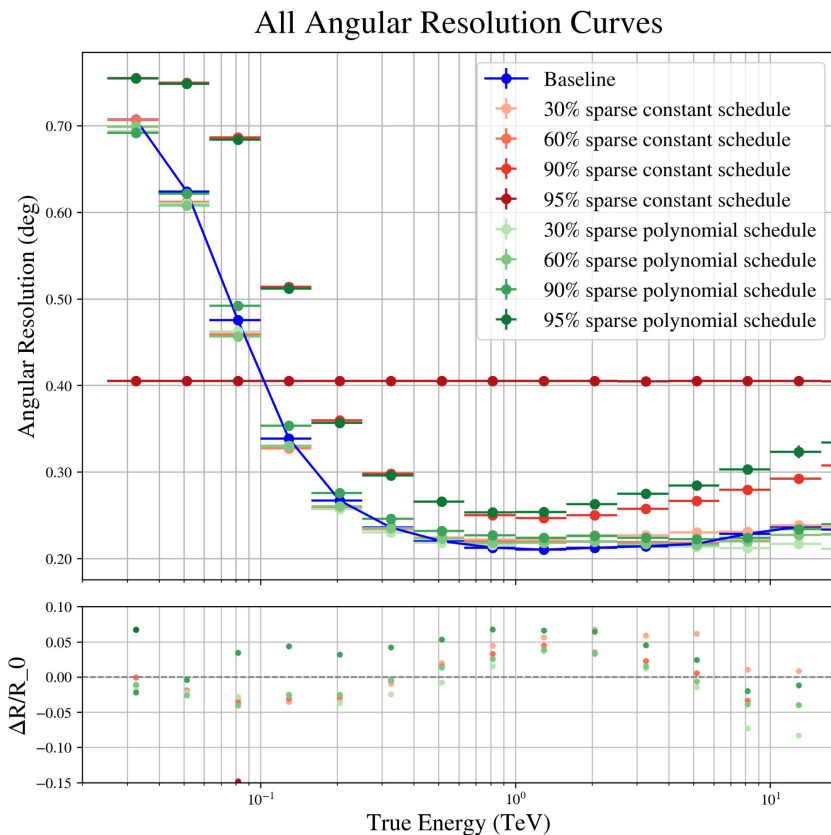


## Resolution Curves

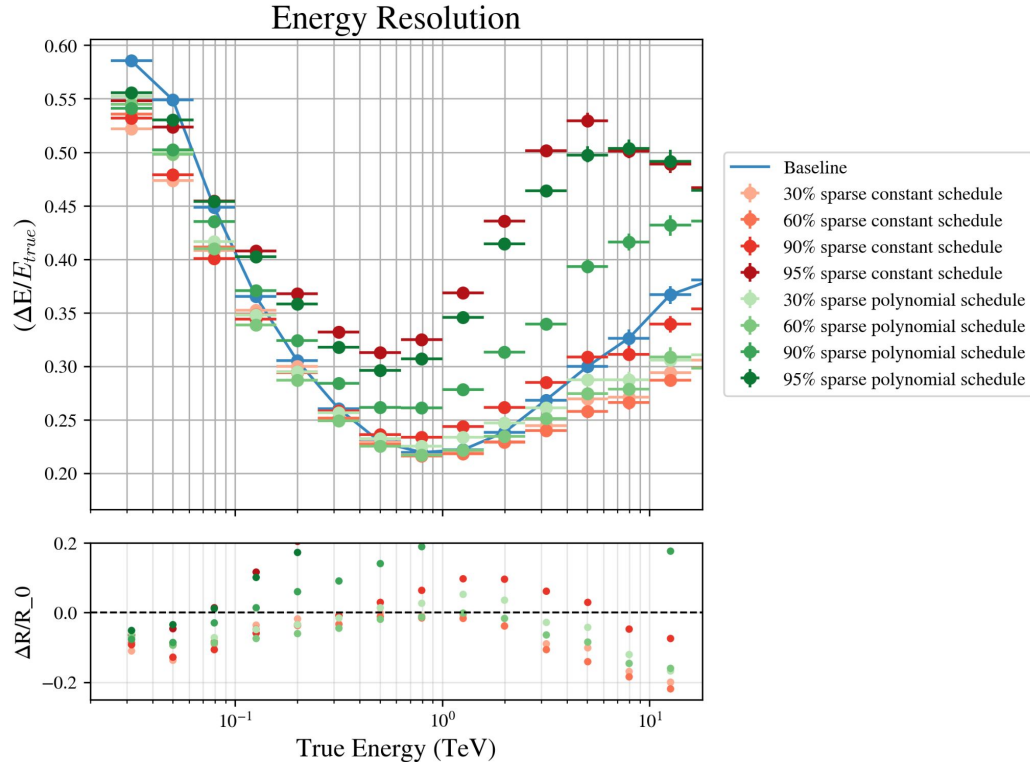
- Angular resolution
- Energy resolution

# Results: Regression task

- Polynomial: Similar to baseline until sparsities up to 90%
- Polynomial outperforms constant schedule
- Possible improvement in some energy ranges
- Relative deviation with respect to baseline



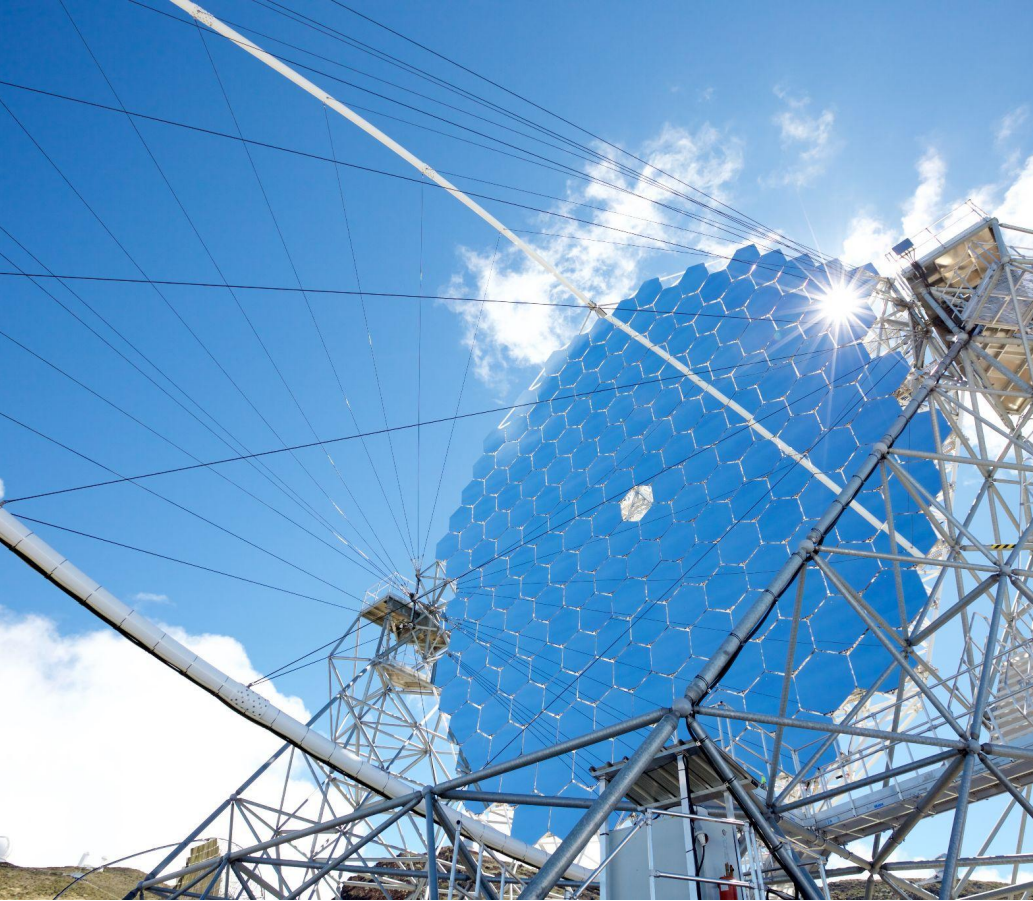
# Results: Regression task



- Polynomial: Degradation at 90% sparsity
- Possible improvement in some energy ranges
- Relative deviation with respect to baseline



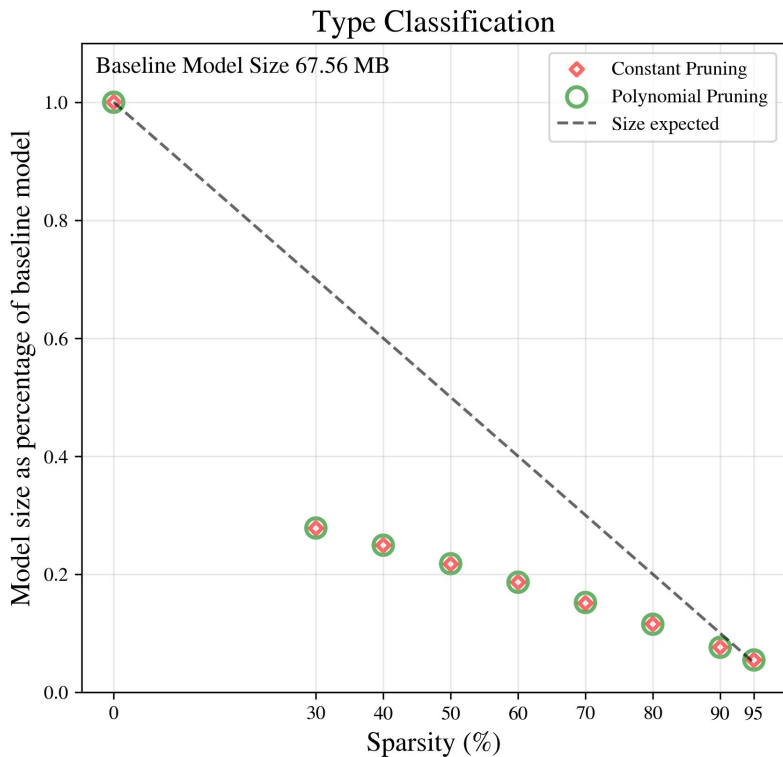
## Results: Model size



Evaluate the size of  
the model as a  
function of sparsity

# Results: Reduction in size

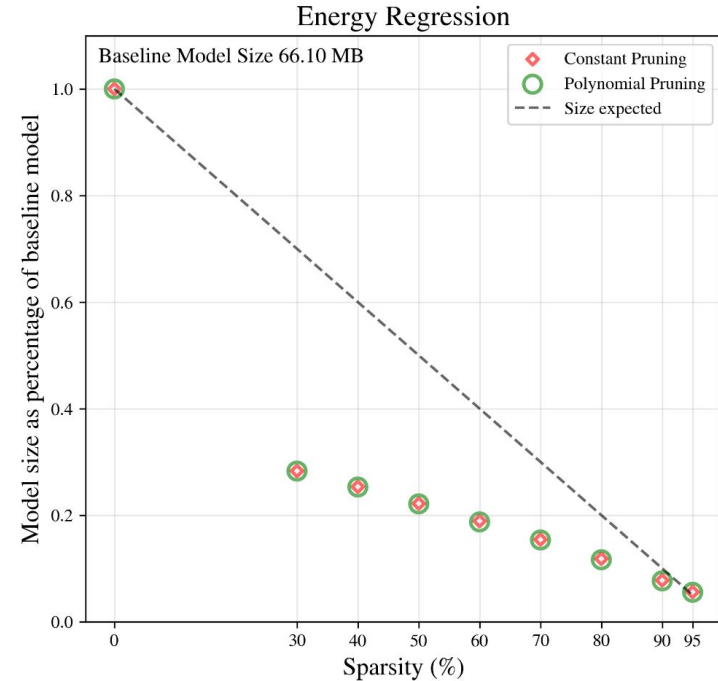
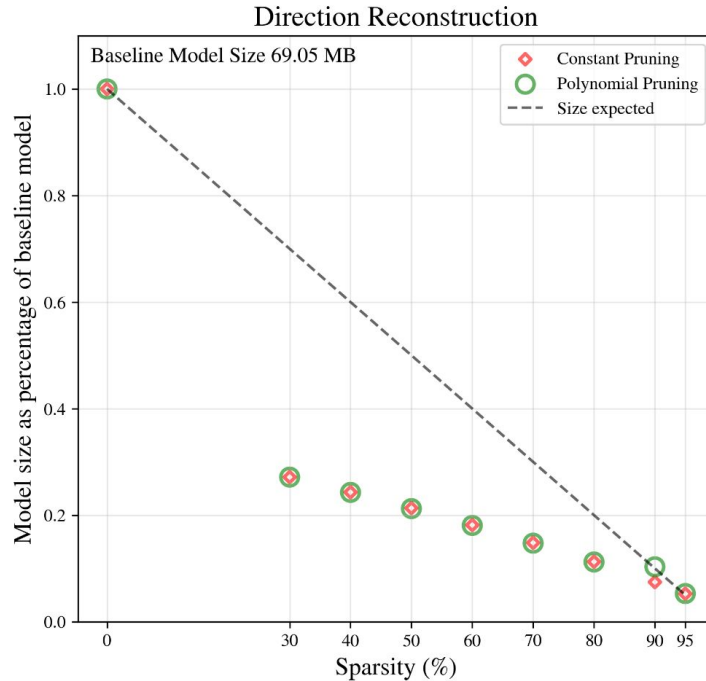
## Model Size vs Sparsity for Different Tasks



- Model sizes after compression to [.zip](#)
- Smaller size than linear reduction
- 30% sparsity: 28% of original size
- 95% sparsity: 5% of original size

# Results: Reduction in size

## Model Size vs Sparsity for Different Tasks



# Conclusions & Future Work

## Conclusions

- **Polynomial sparsity** outperforms **constant sparsity**
- **Classification: polynomial schedule** without significant degradation up to 90-95%.
- **Energy polynomial schedule** without significant degradation at 90%.
- **Direction: polynomial schedule** without significant degradation up to 90%.
- Significant reduction in model **size**

## Ongoing Work

- NSB vs. cosmics classification
- L1 over flowers with 2155 threshold
- Use simpler CNN model
- Sparsity-aware quantisation implementation

# Agradecimientos



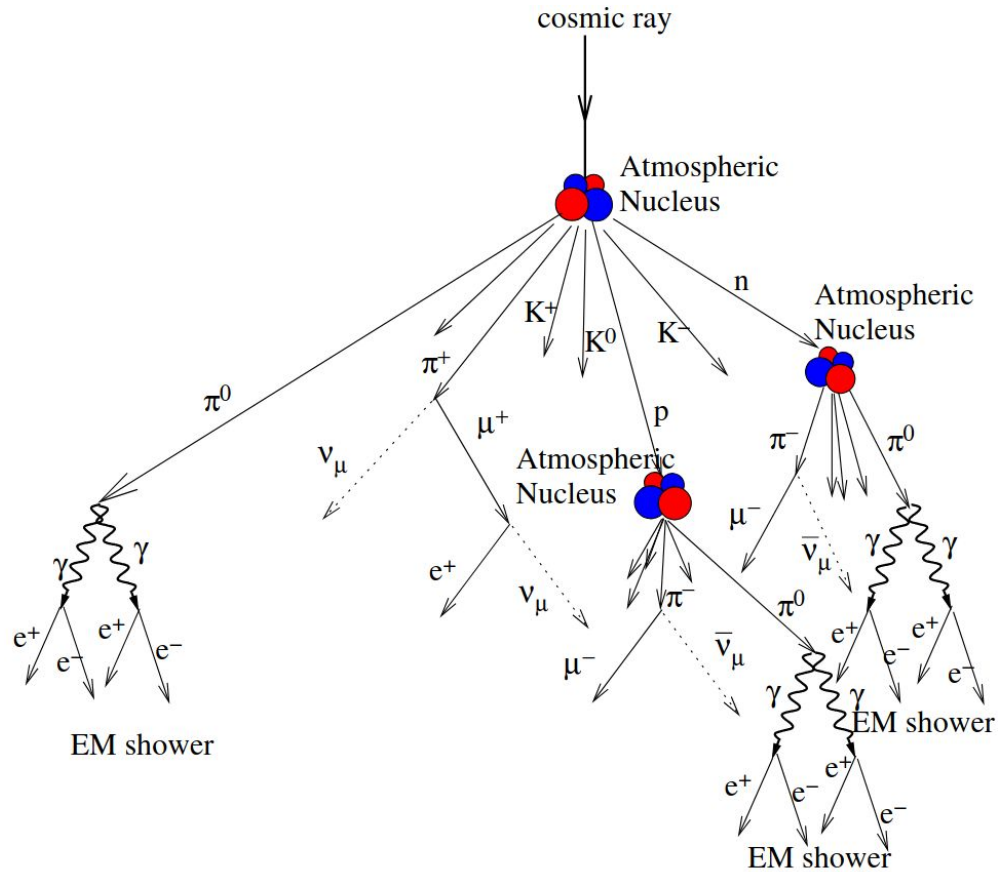
The research here presented has been partially supported by the MICIU/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR, under grant PDC2023-145839-I00



UNIVERSIDAD  
COMPLUTENSE  
MADRID

The author(s) gratefully acknowledges the computer resources at Artemisa, funded by the European Union ERDF and Comunitat Valenciana as well as the technical support provided by the Instituto de Física Corpuscular, IFIC (CSIC-UV)





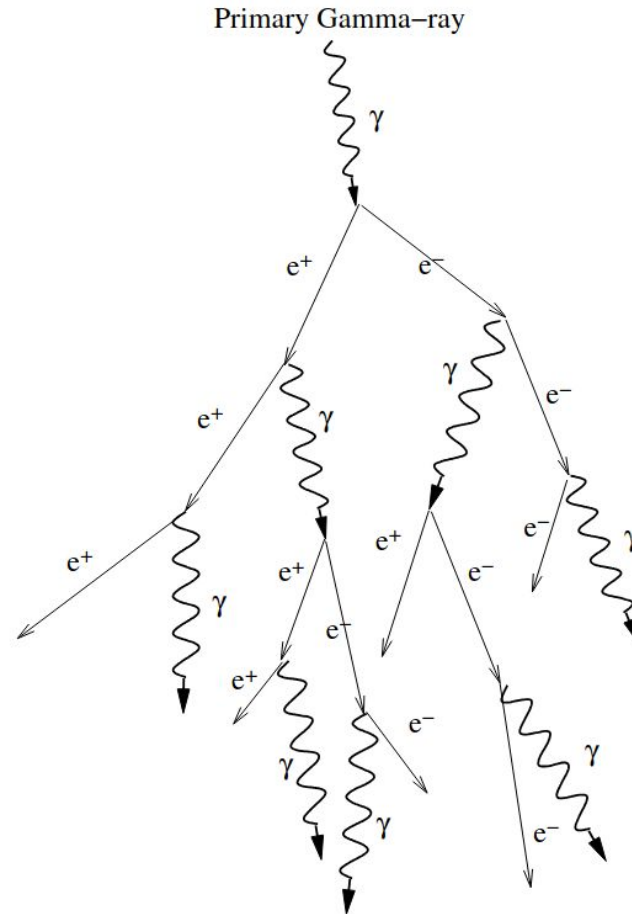
## Cascada Hadrónica:

- Protón
- Núcleos de Helio ( $\alpha$ )

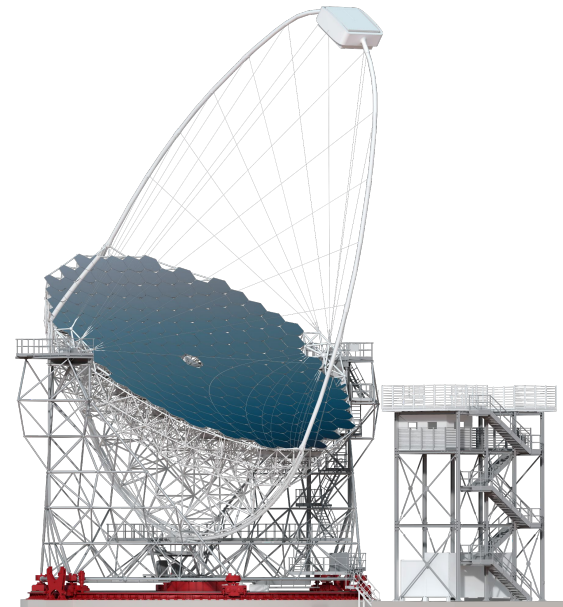
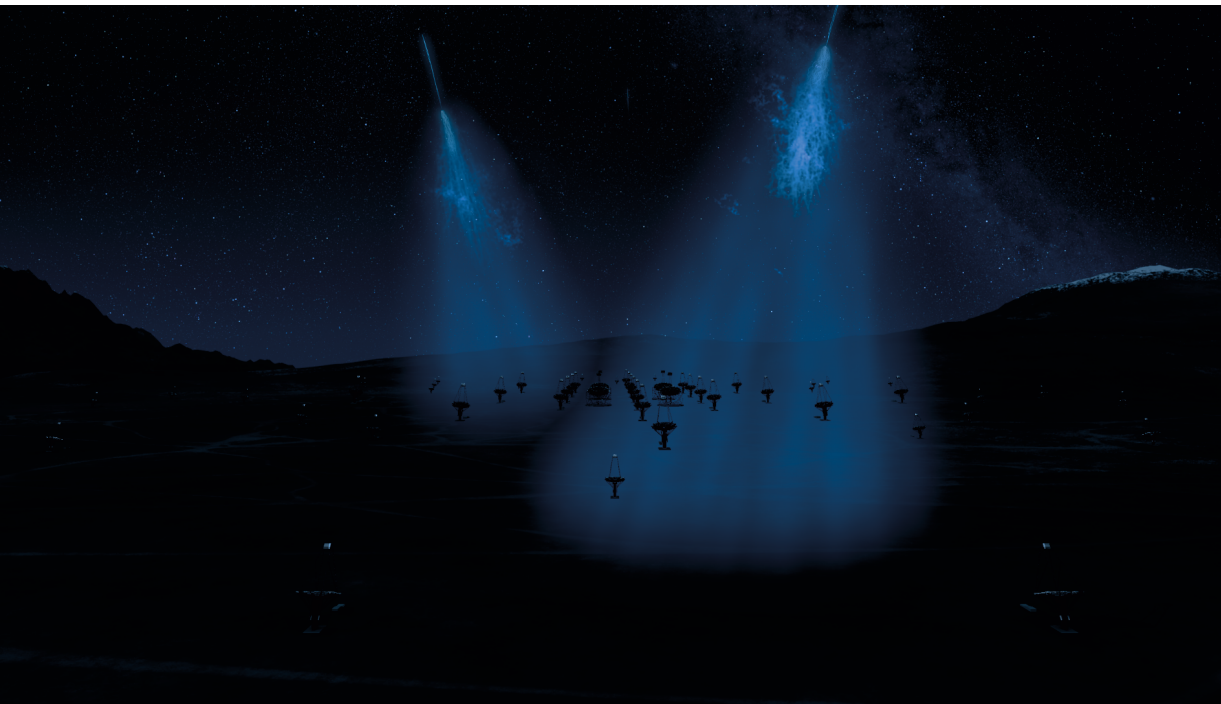


## Cascada Electromagnética:

- Rayo Gamma ( $\gamma$ )

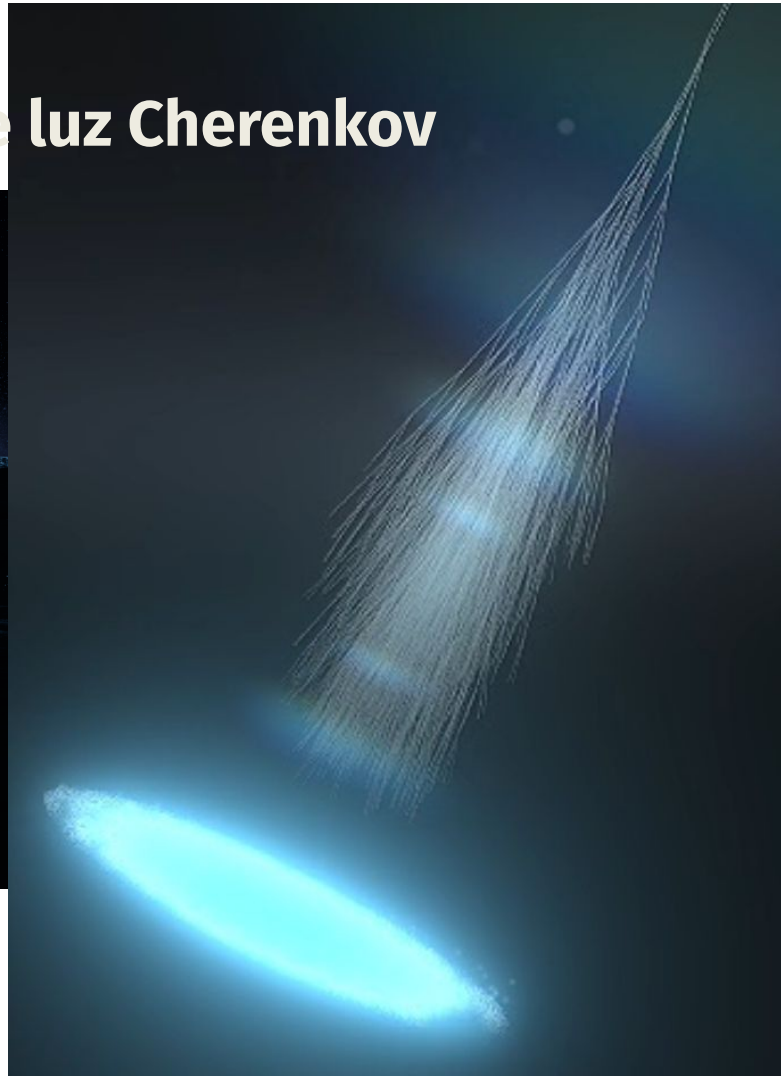
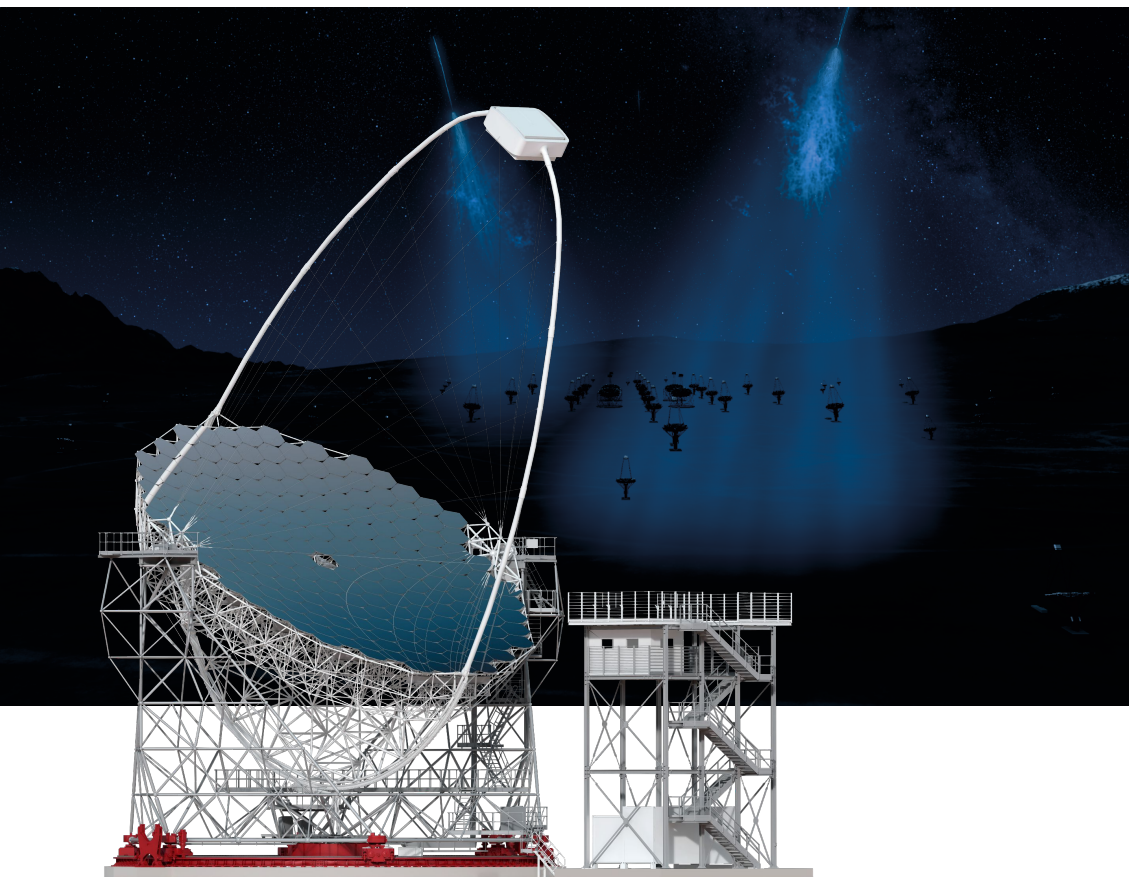


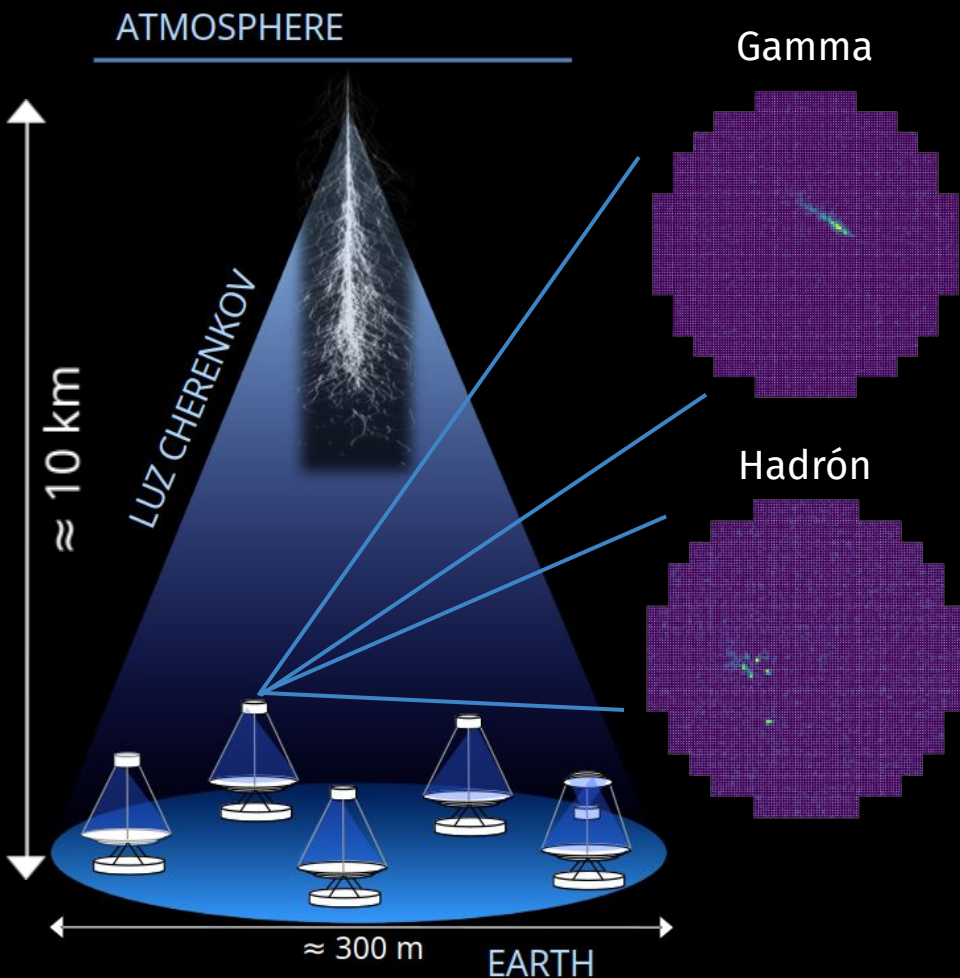
# Contextualización: Detección de luz Cherenkov





# Contextualización: Detección de luz Cherenkov





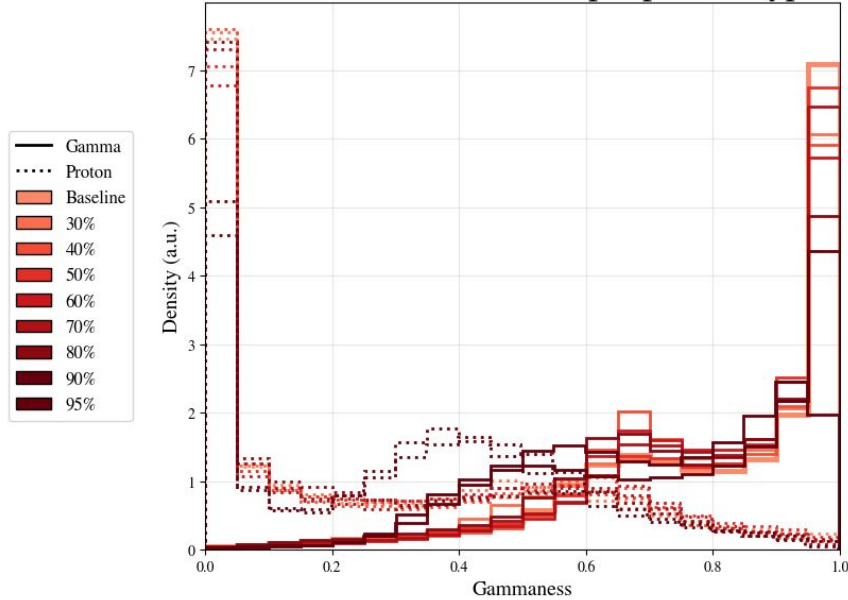
# Introducción

## Reconstrucción de eventos

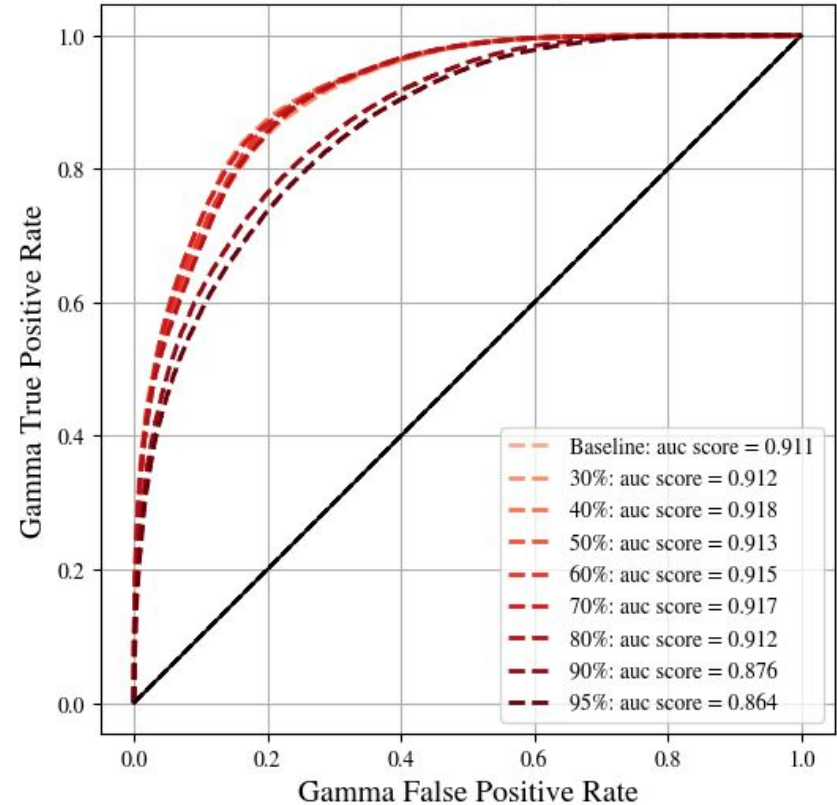
- Clasificación:
  - Rayo cósmico
  - Rayo gamma
- Regresión:
  - Dirección
  - Energía
- Algoritmos AA

# Resultados: Tarea de clasificación

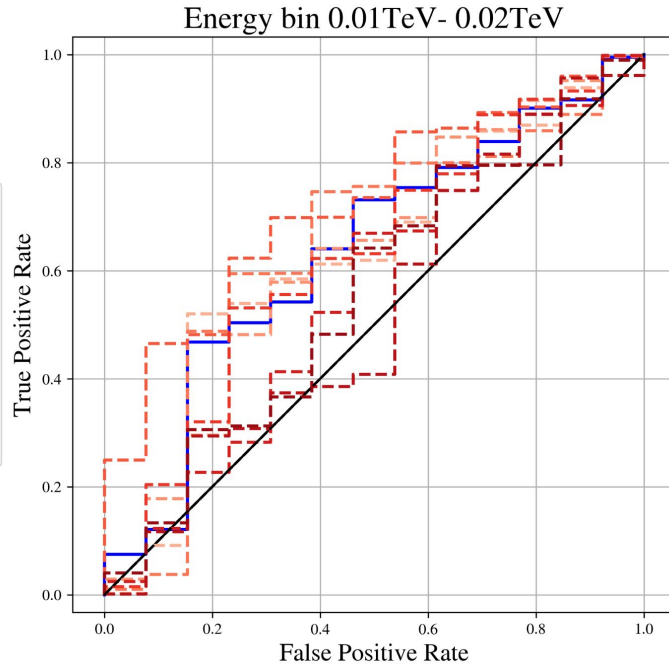
Gammaness Distribution per particle type



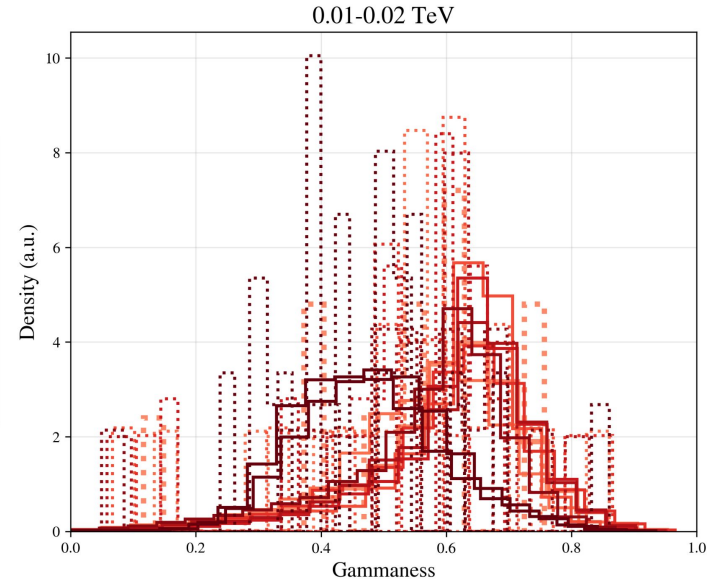
ROC Curves Comparison



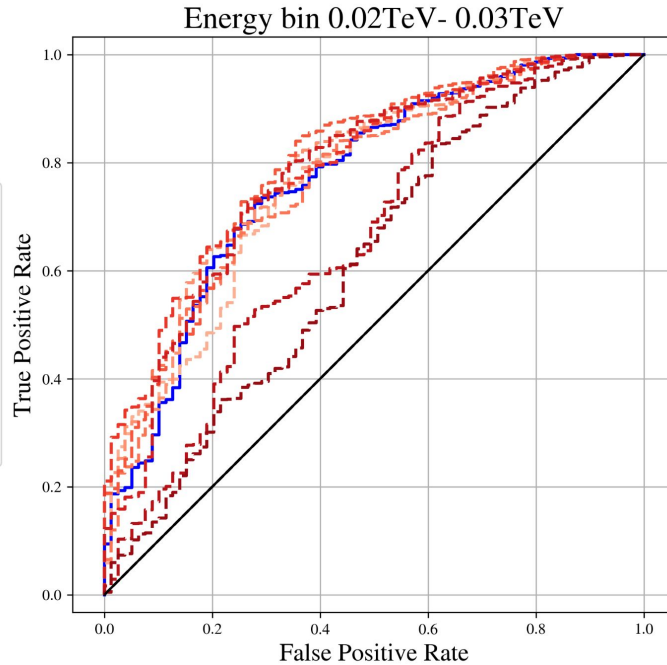
# Resultados: Tarea de clasificación



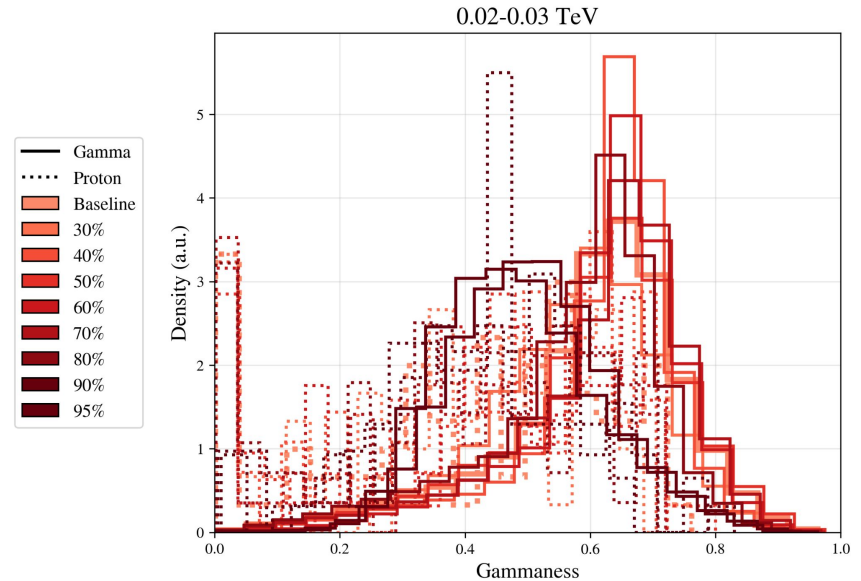
## Gammaness Distribution Comparison per Energy Bin



# Resultados: Tarea de clasificación

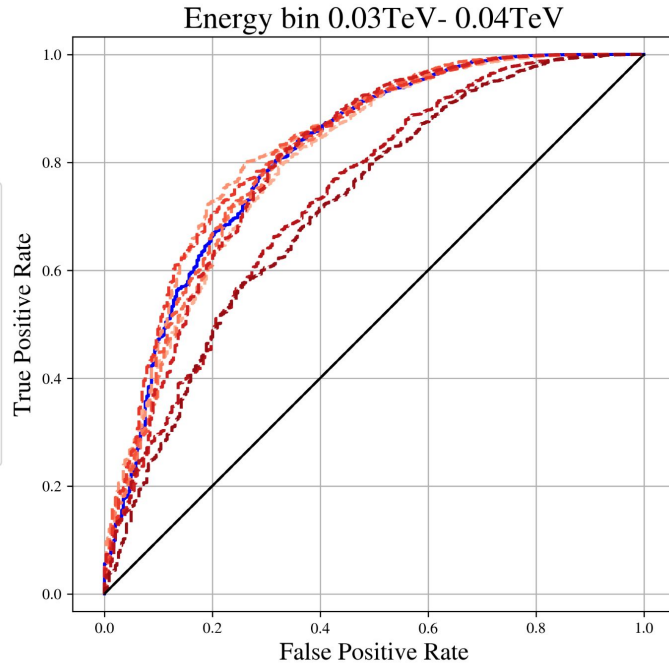


## Gammaness Distribution Comparison per Energy Bin

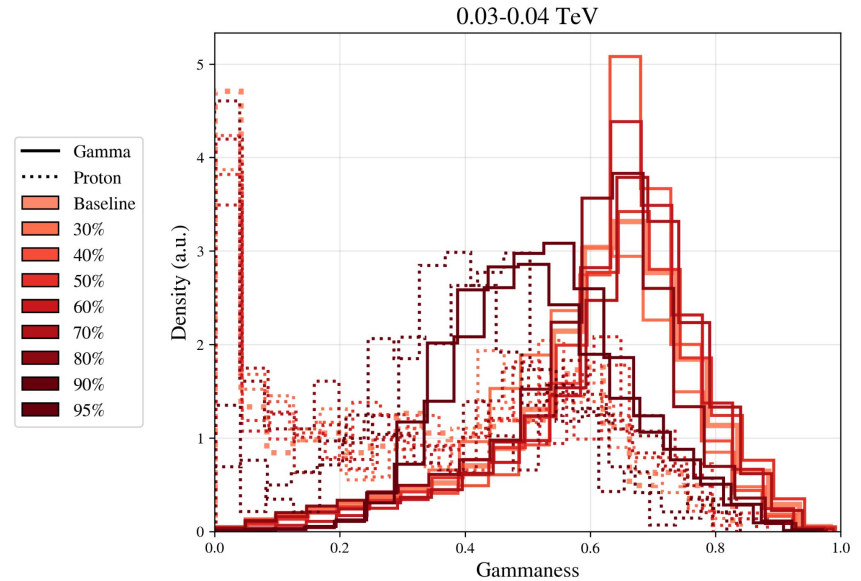




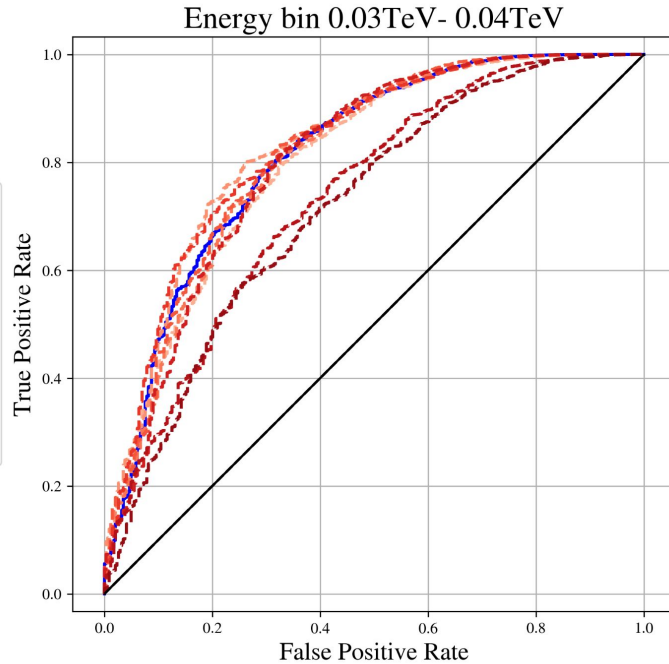
# Resultados: Tarea de clasificación



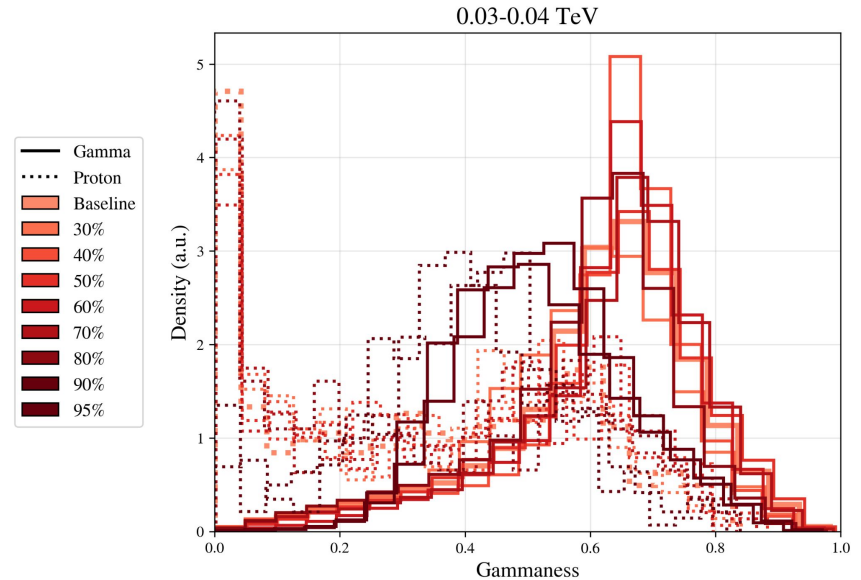
Gammaness Distribution Comparison per Energy Bin



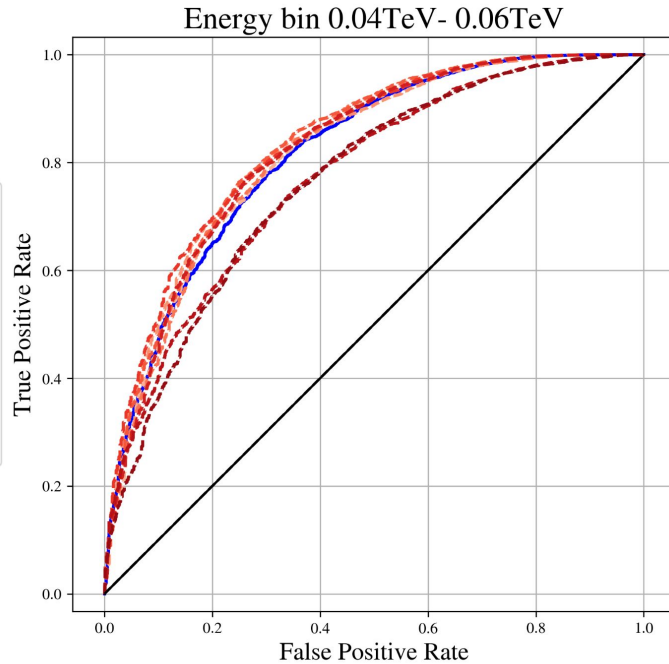
# Resultados: Tarea de clasificación



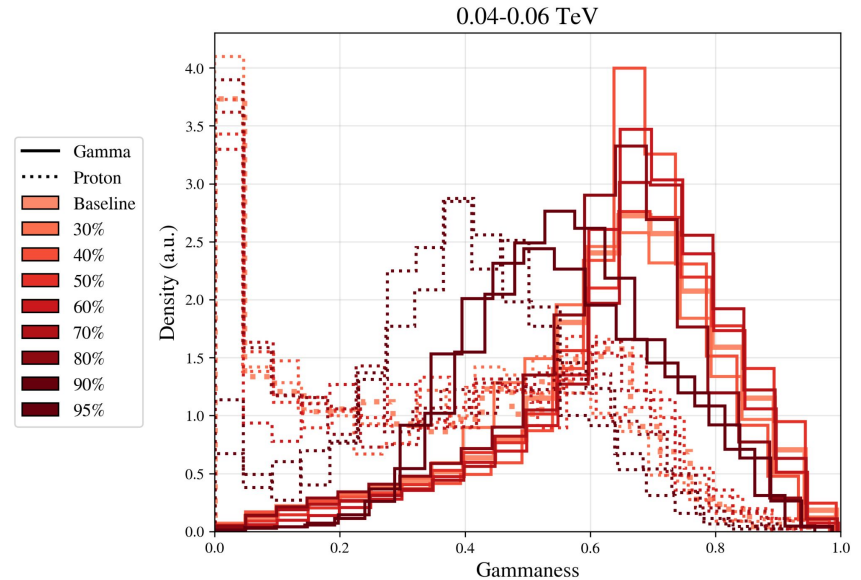
Gammaness Distribution Comparison per Energy Bin



# Resultados: Tarea de clasificación

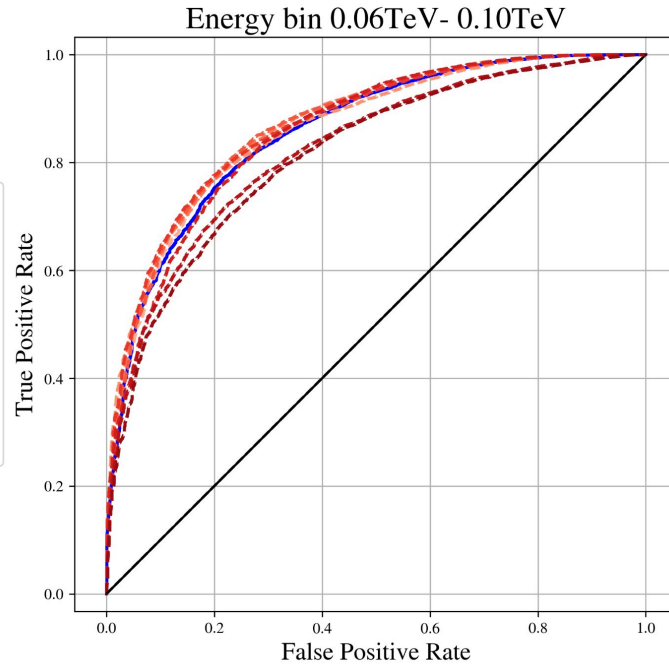


Gammaness Distribution Comparison per Energy Bin

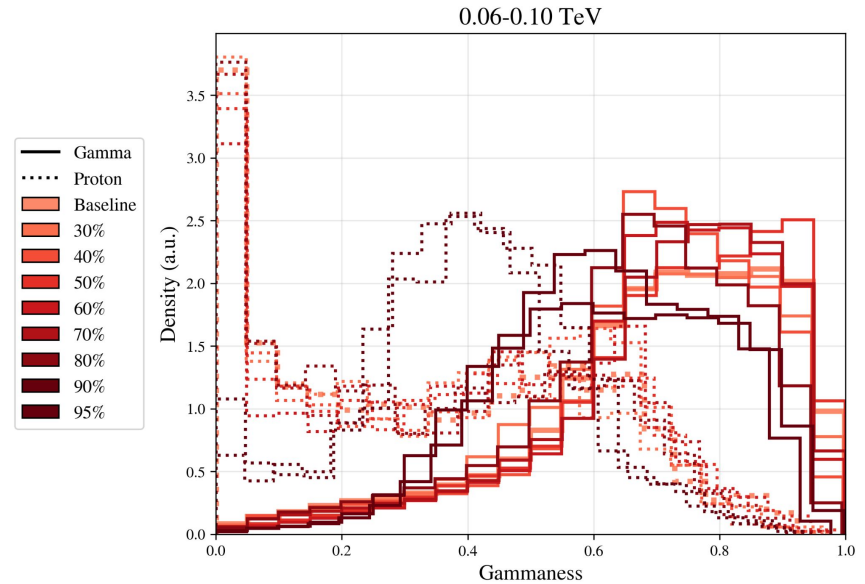




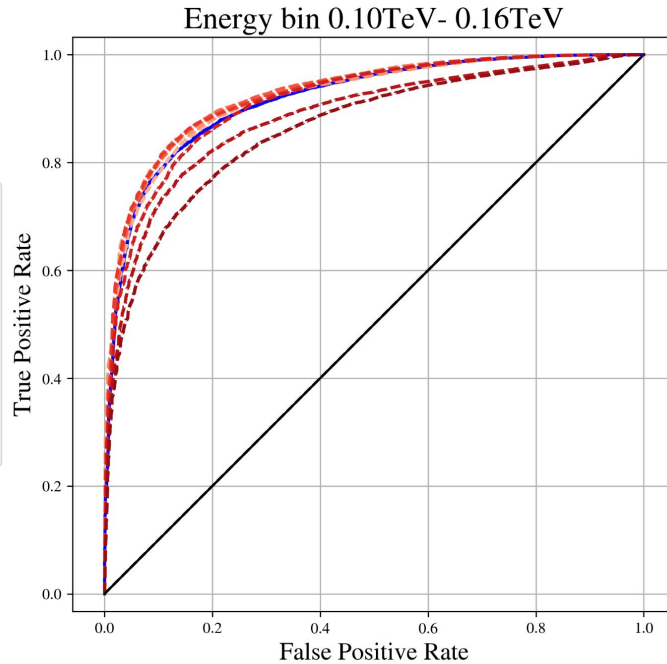
# Resultados: Tarea de clasificación



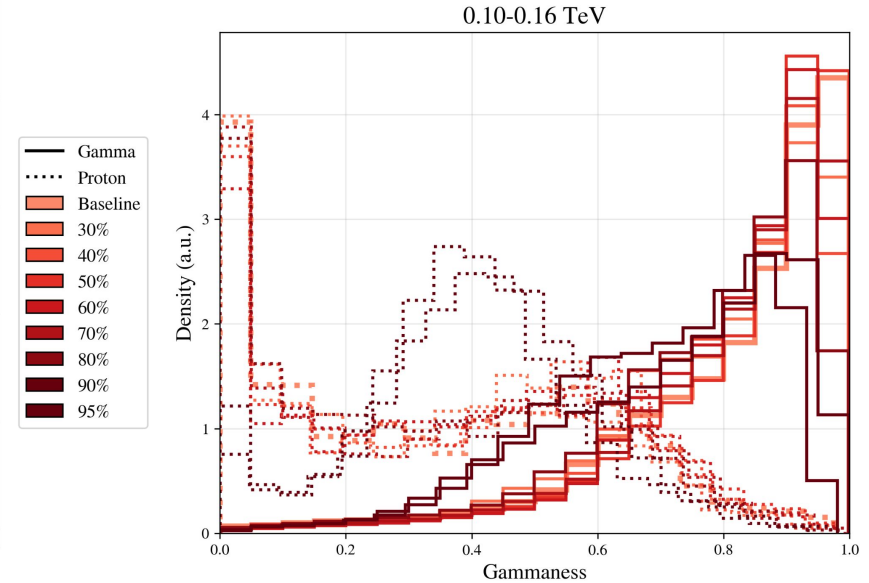
## Gammaness Distribution Comparison per Energy Bin



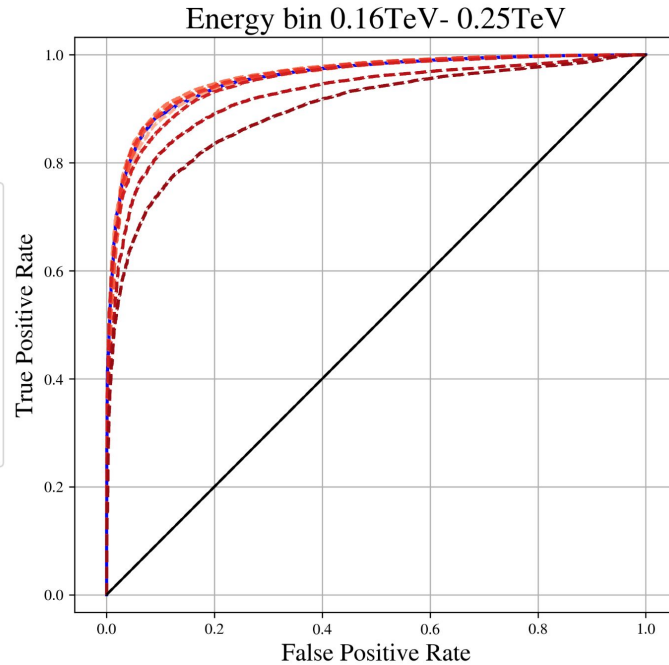
# Resultados: Tarea de clasificación



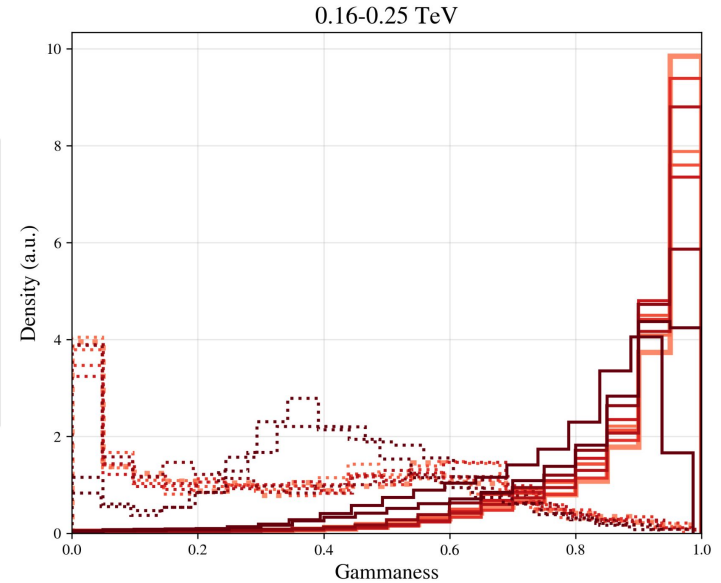
## Gammaness Distribution Comparison per Energy Bin



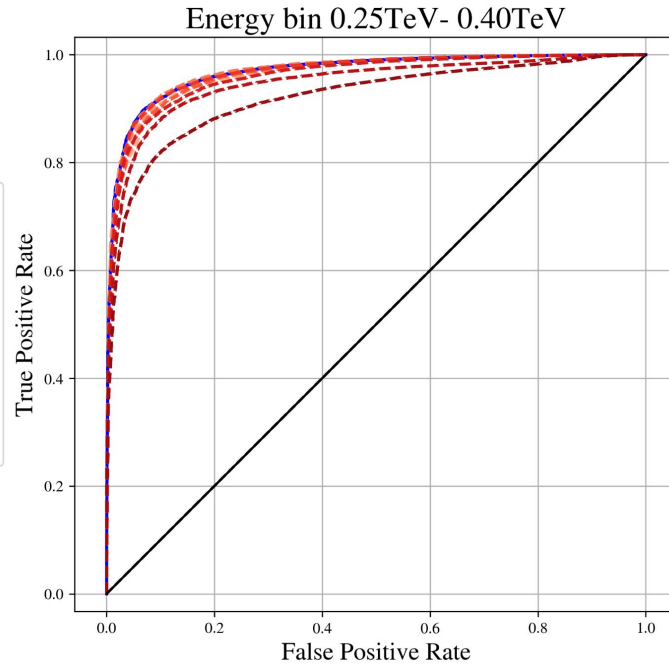
# Resultados: Tarea de clasificación



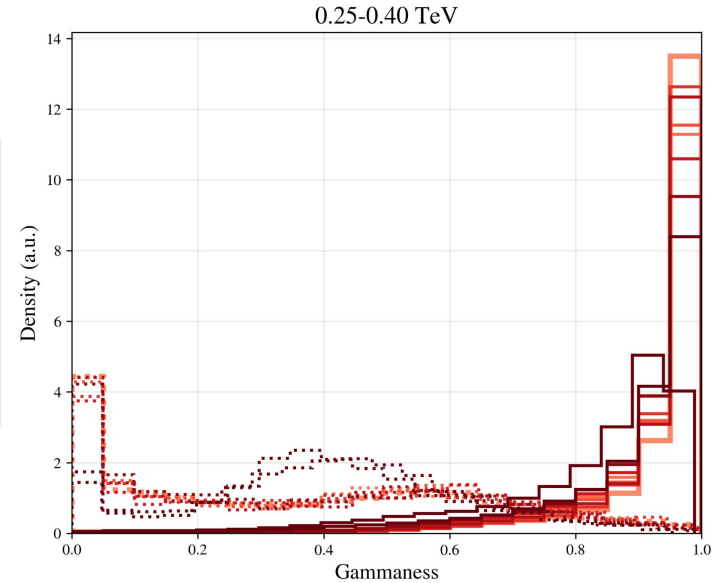
## Gammaness Distribution Comparison per Energy Bin



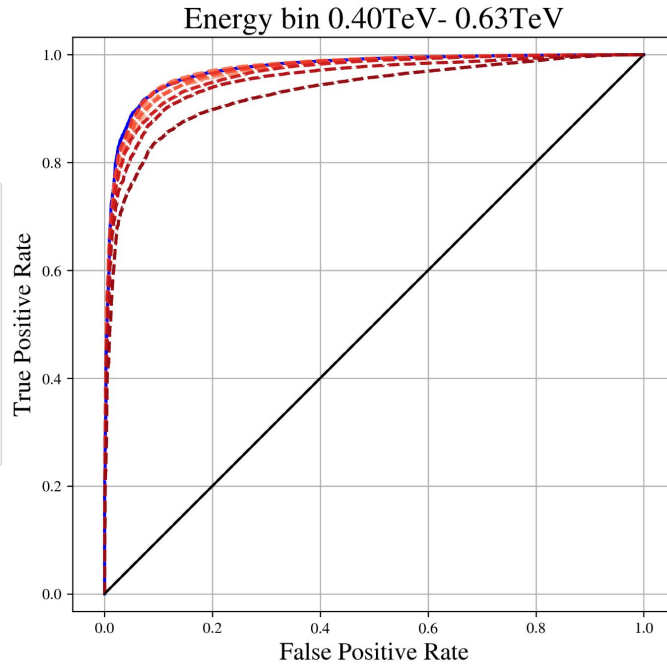
# Resultados: Tarea de clasificación



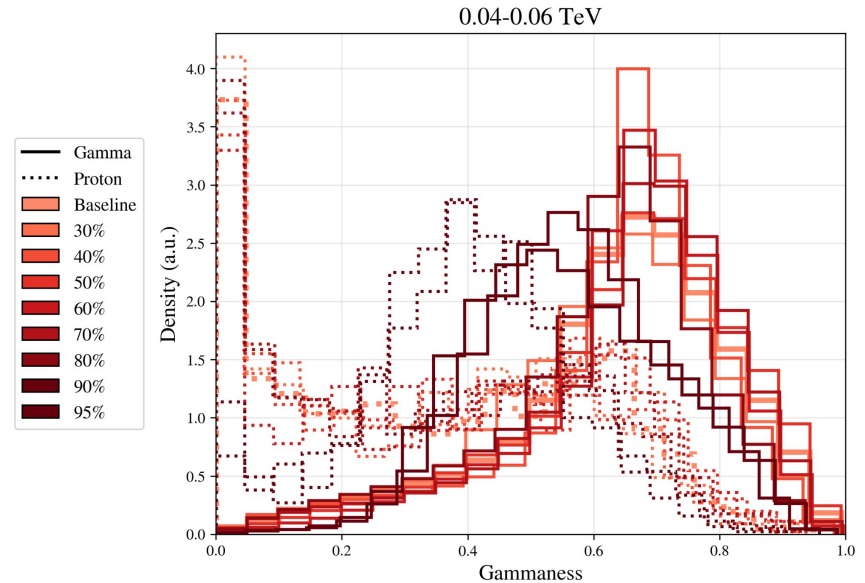
## Gammaness Distribution Comparison per Energy Bin



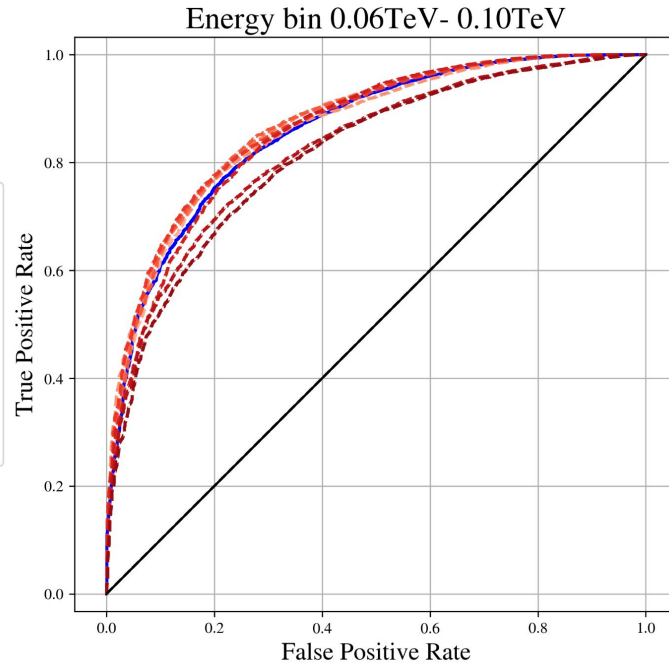
# Resultados: Tarea de clasificación



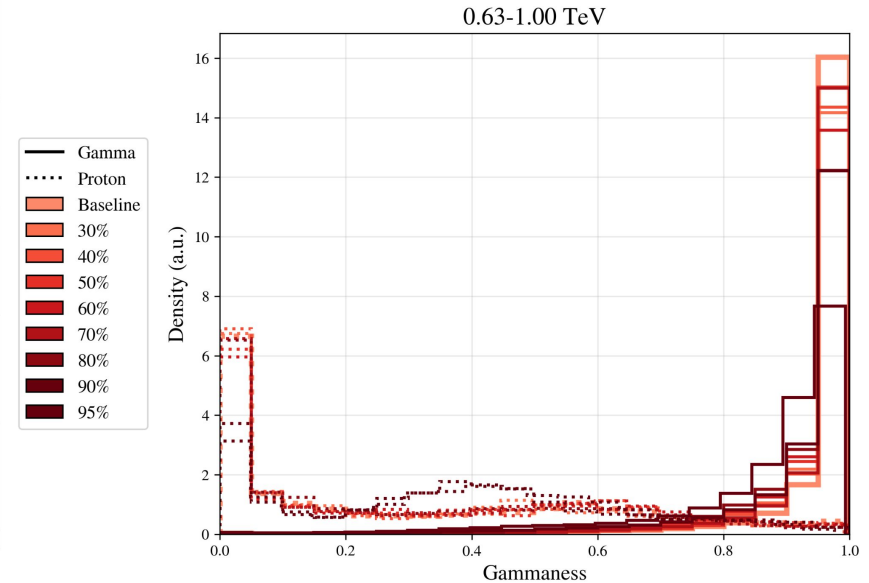
## Gammaness Distribution Comparison per Energy Bin



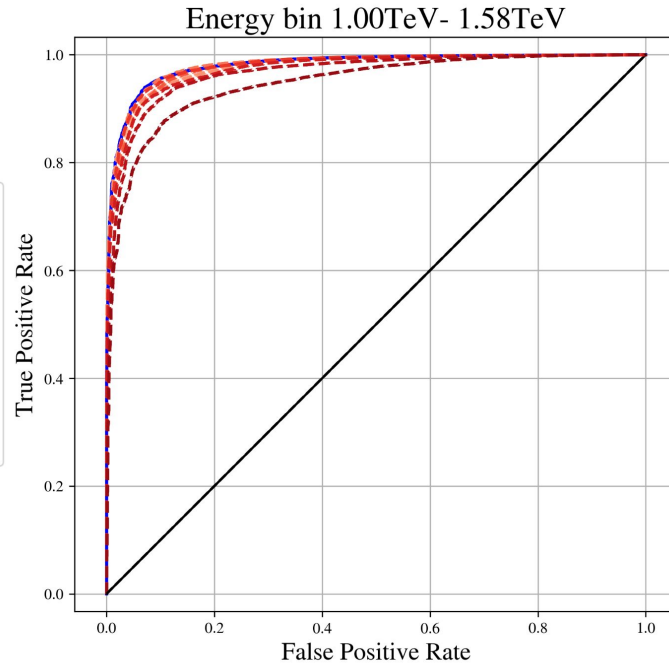
# Resultados: Tarea de clasificación



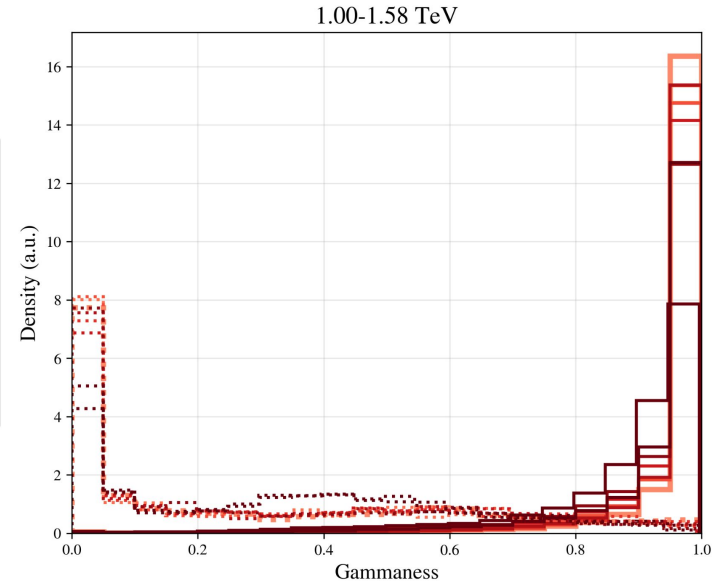
## Gammaness Distribution Comparison per Energy Bin



# Resultados: Tarea de clasificación

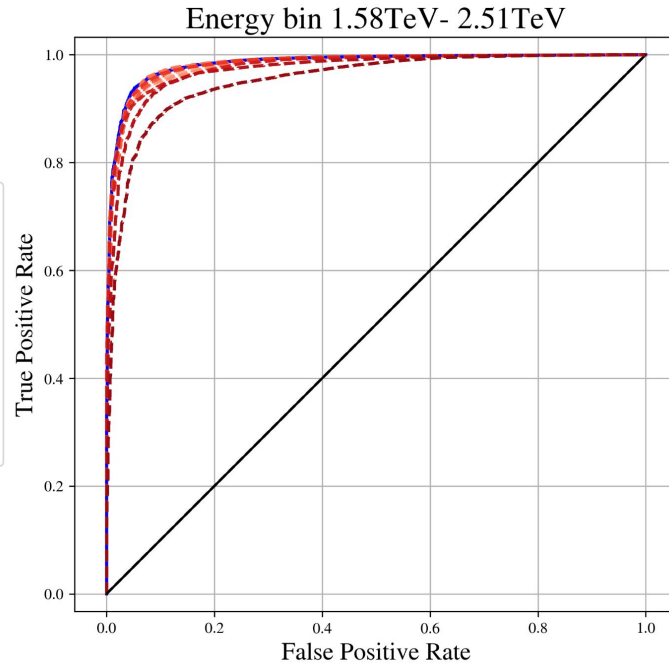


## Gammaness Distribution Comparison per Energy Bin

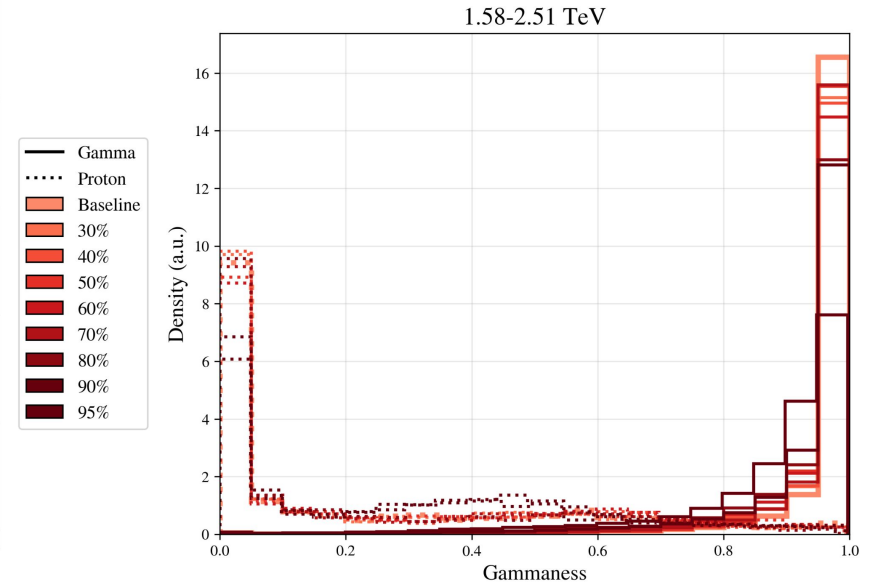




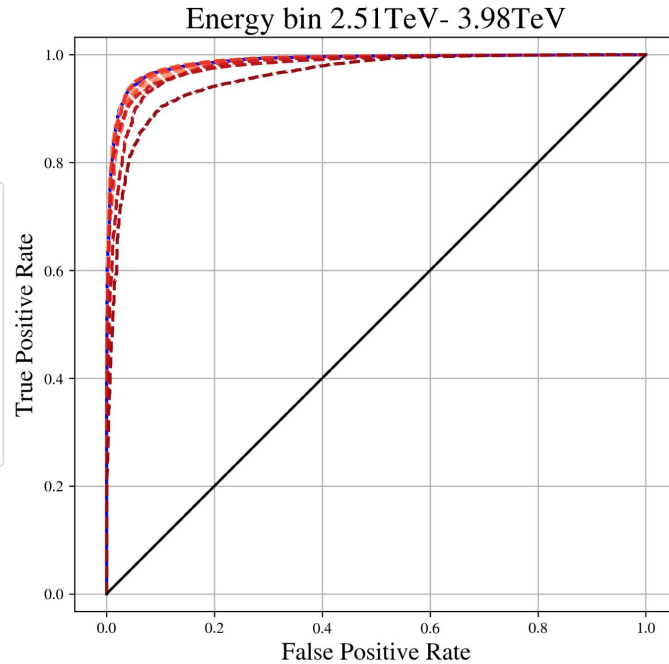
# Resultados: Tarea de clasificación



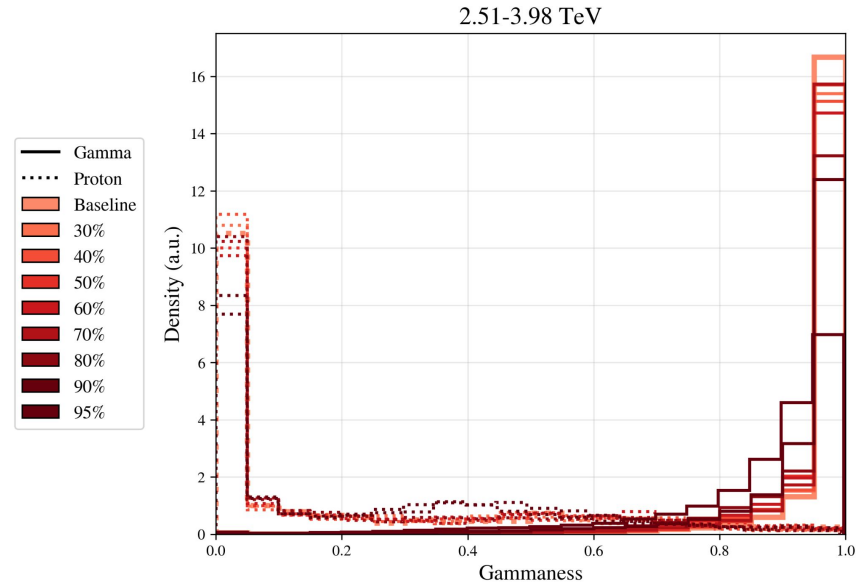
## Gammaness Distribution Comparison per Energy Bin



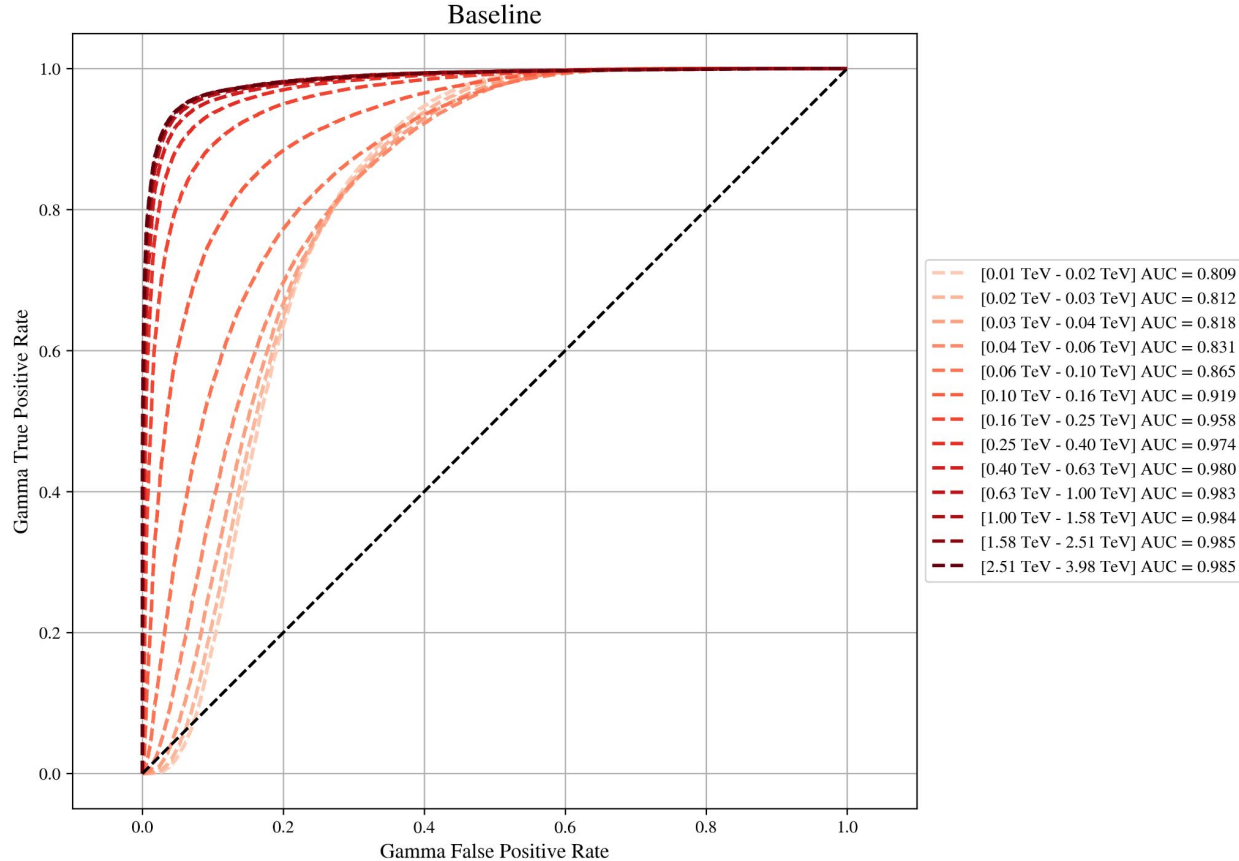
# Resultados: Tarea de clasificación



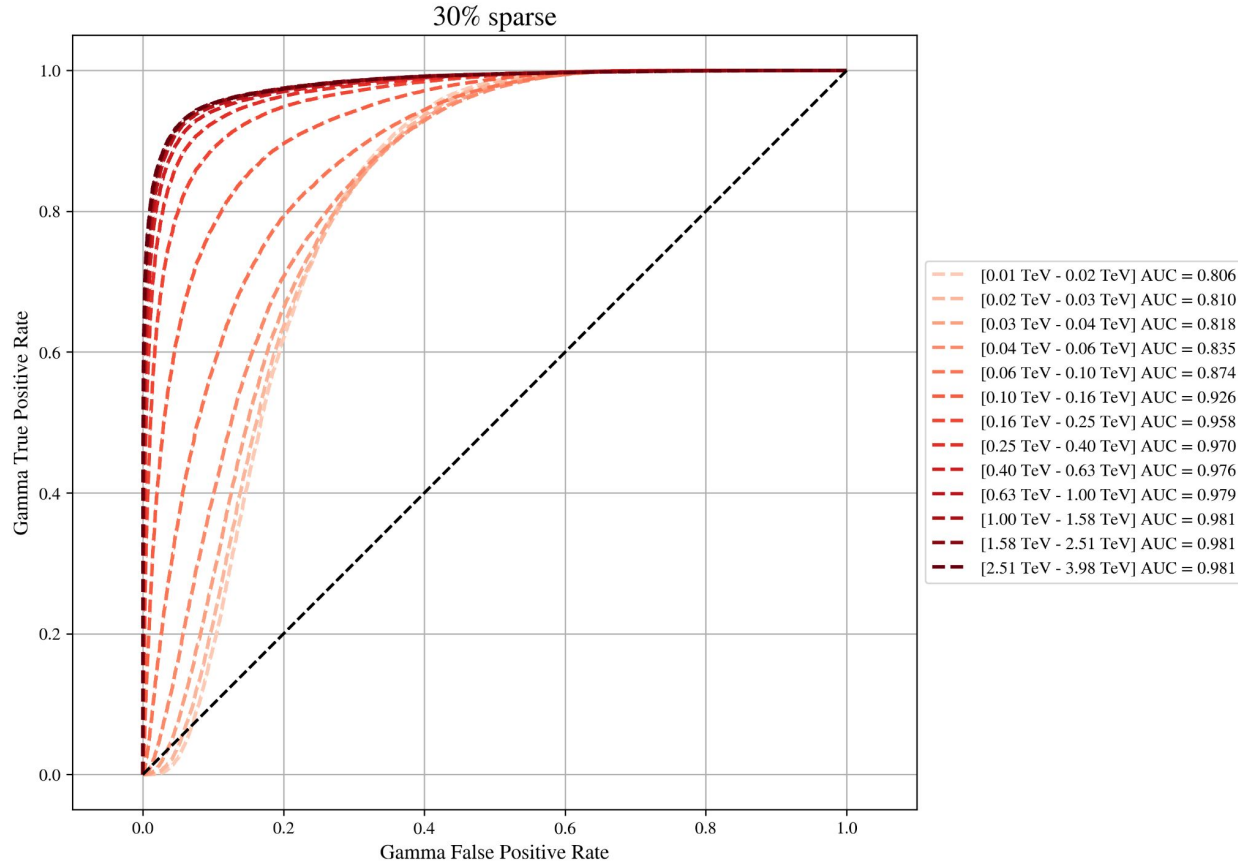
## Gammaness Distribution Comparison per Energy Bin



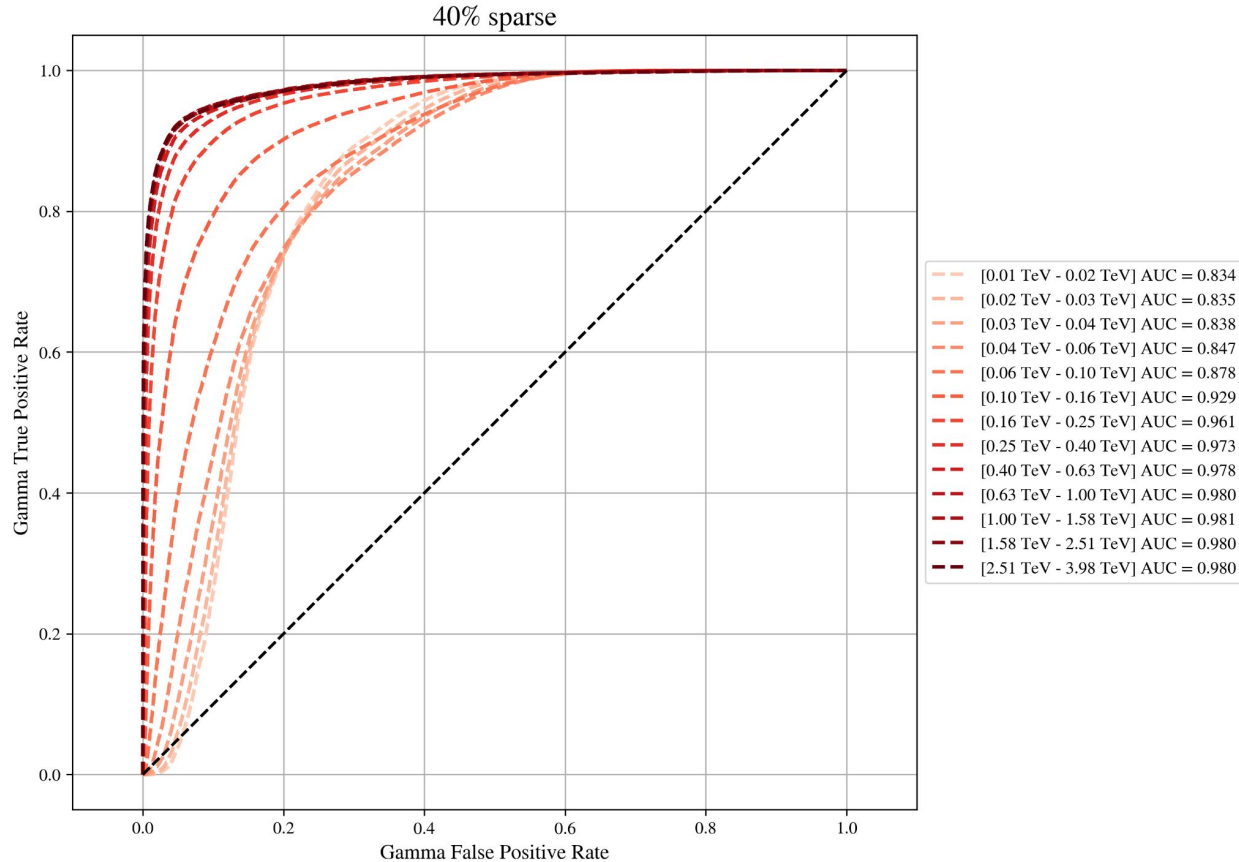
# Resultados: Tarea de clasificación



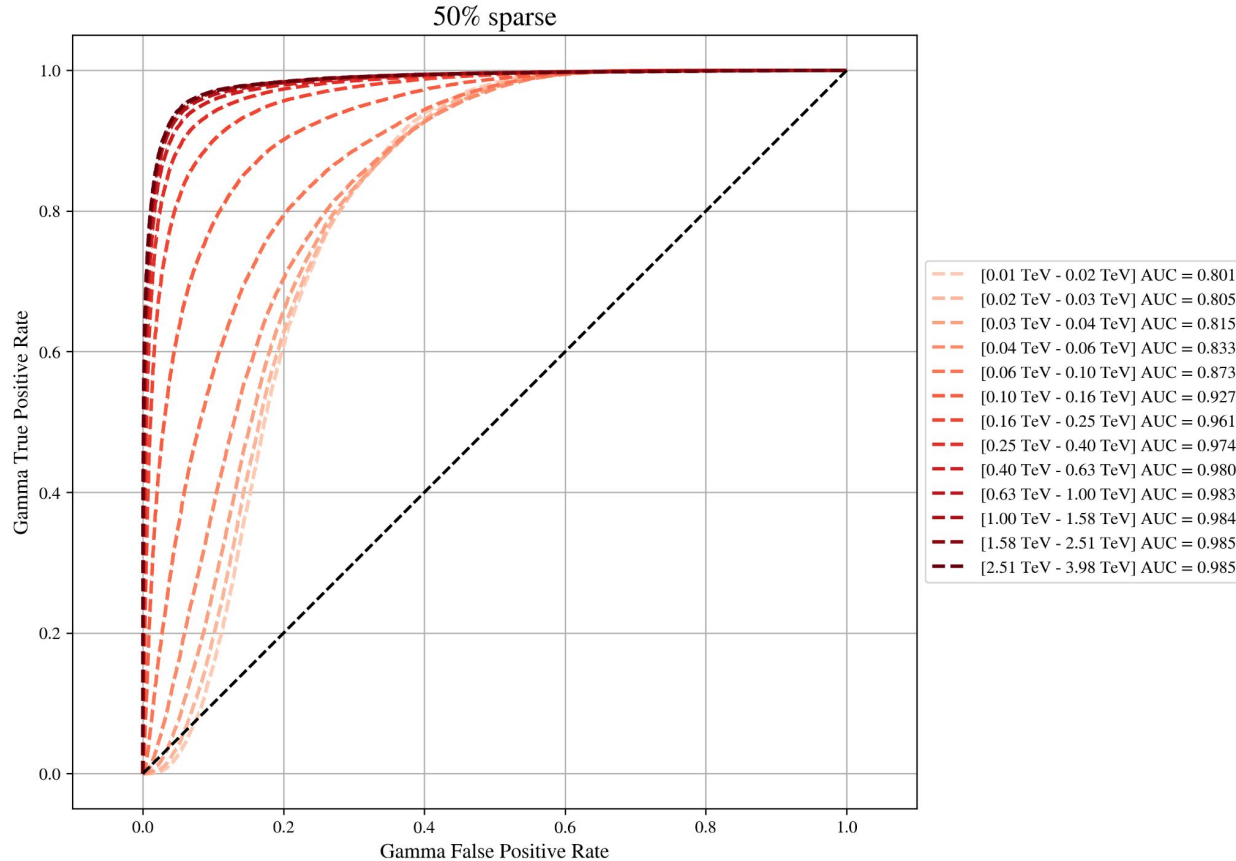
# Resultados: Tarea de clasificación



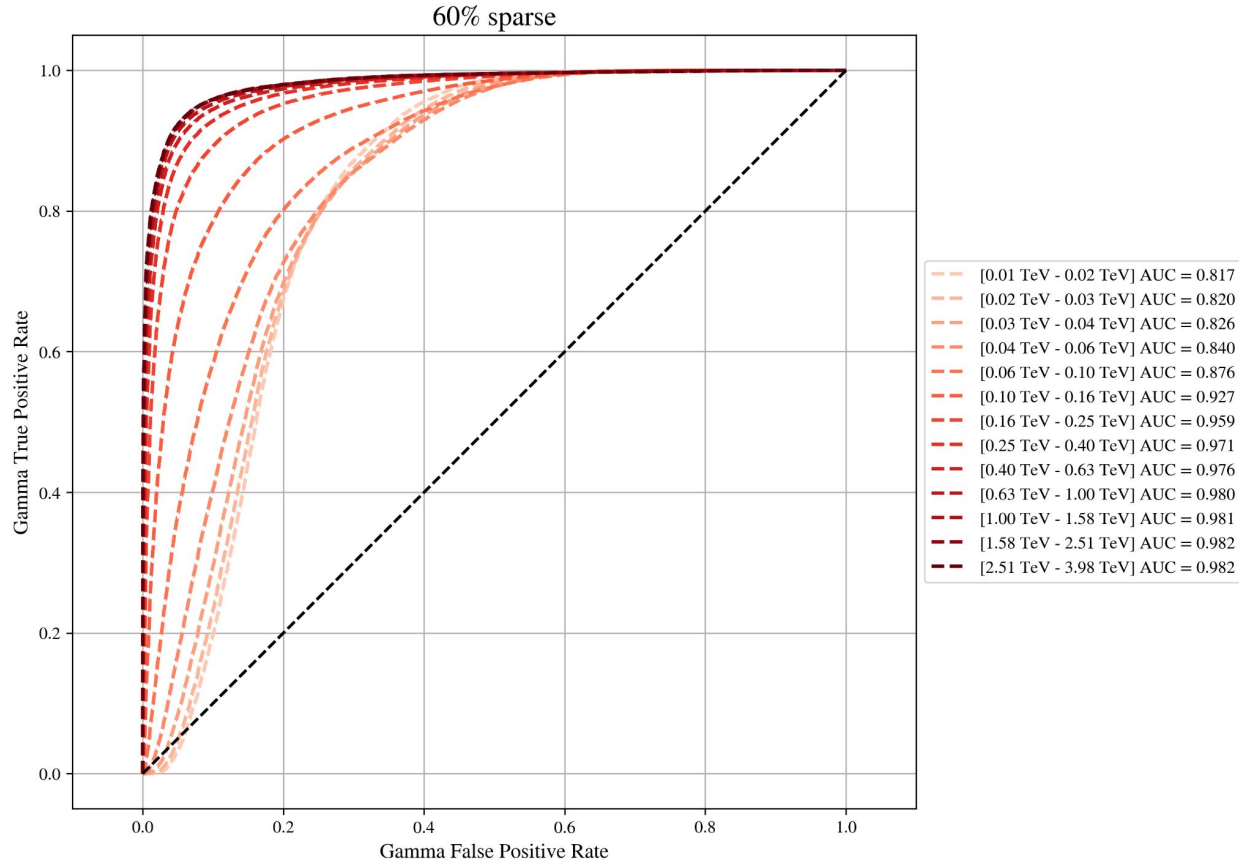
# Resultados: Tarea de clasificación



# Resultados: Tarea de clasificación

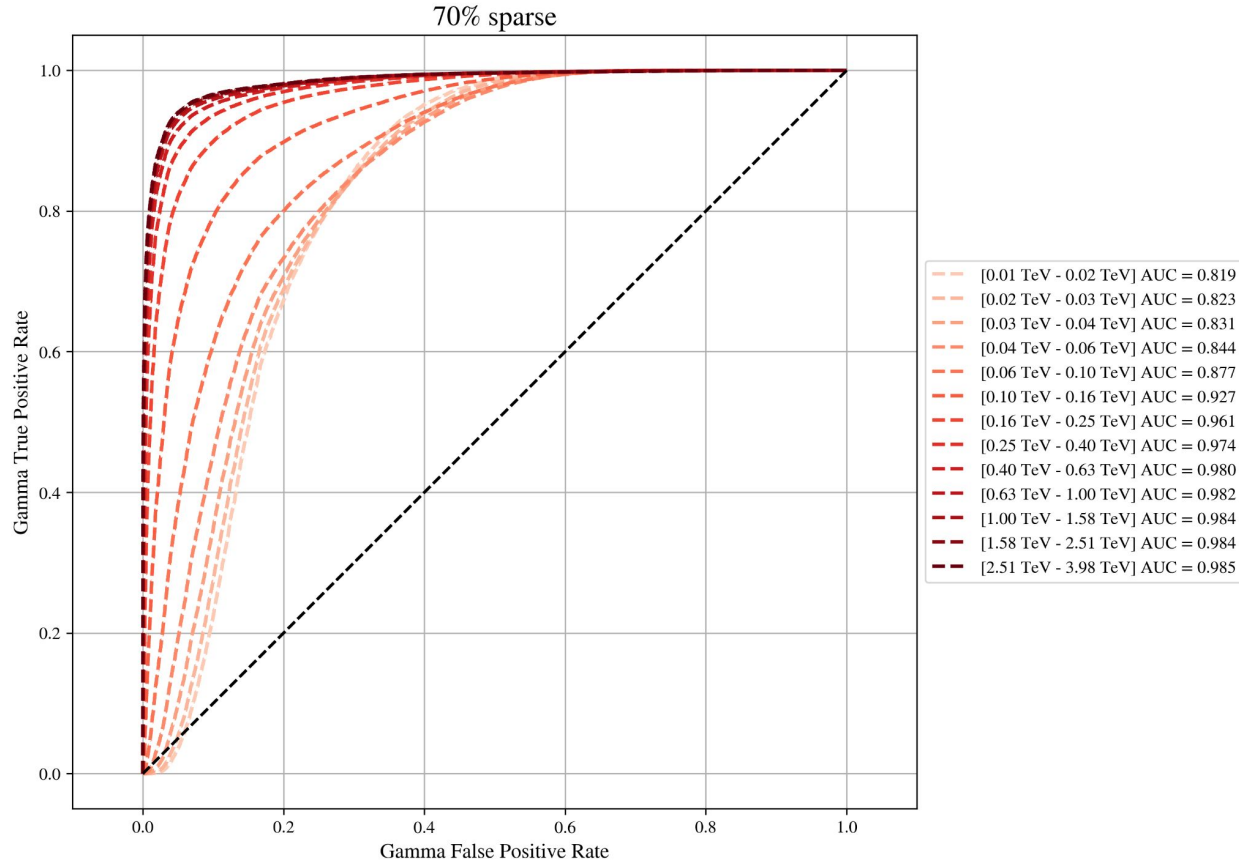


# Resultados: Tarea de clasificación

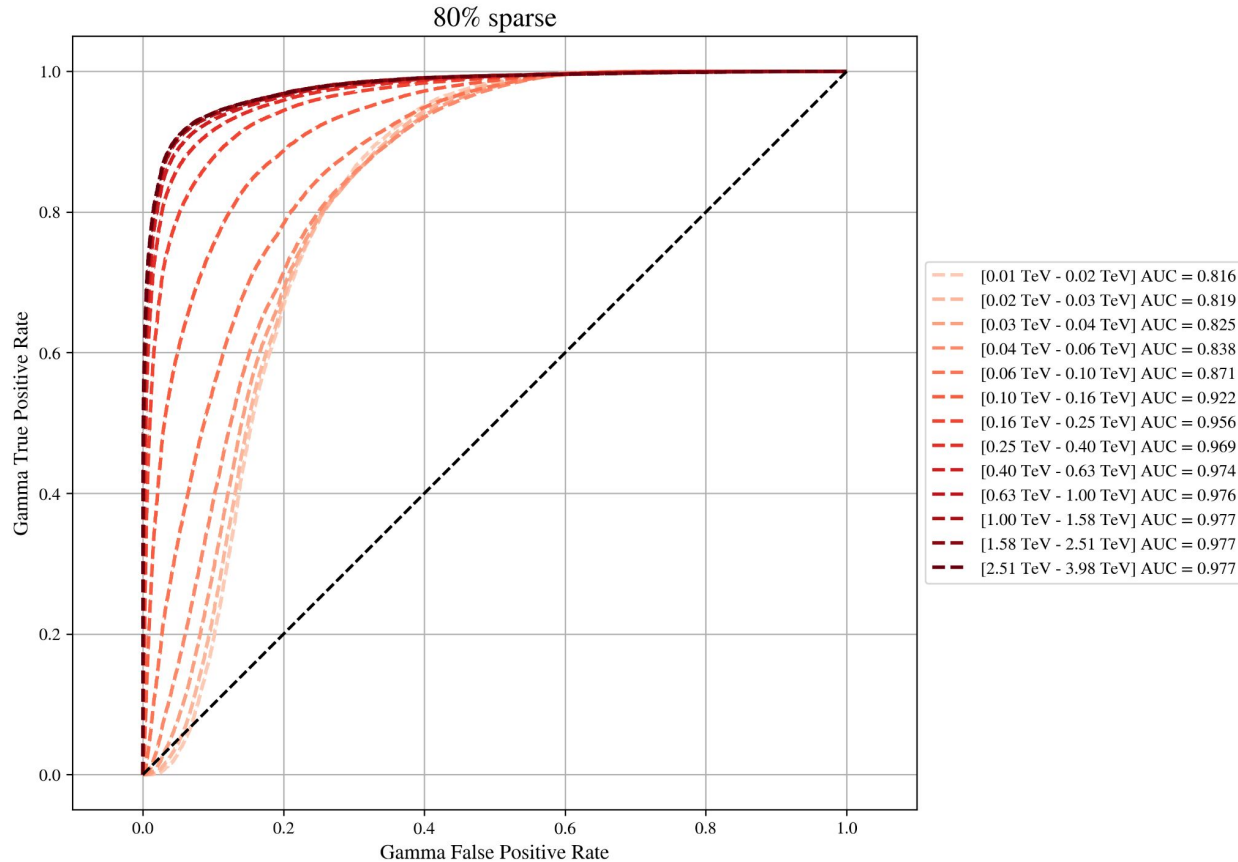




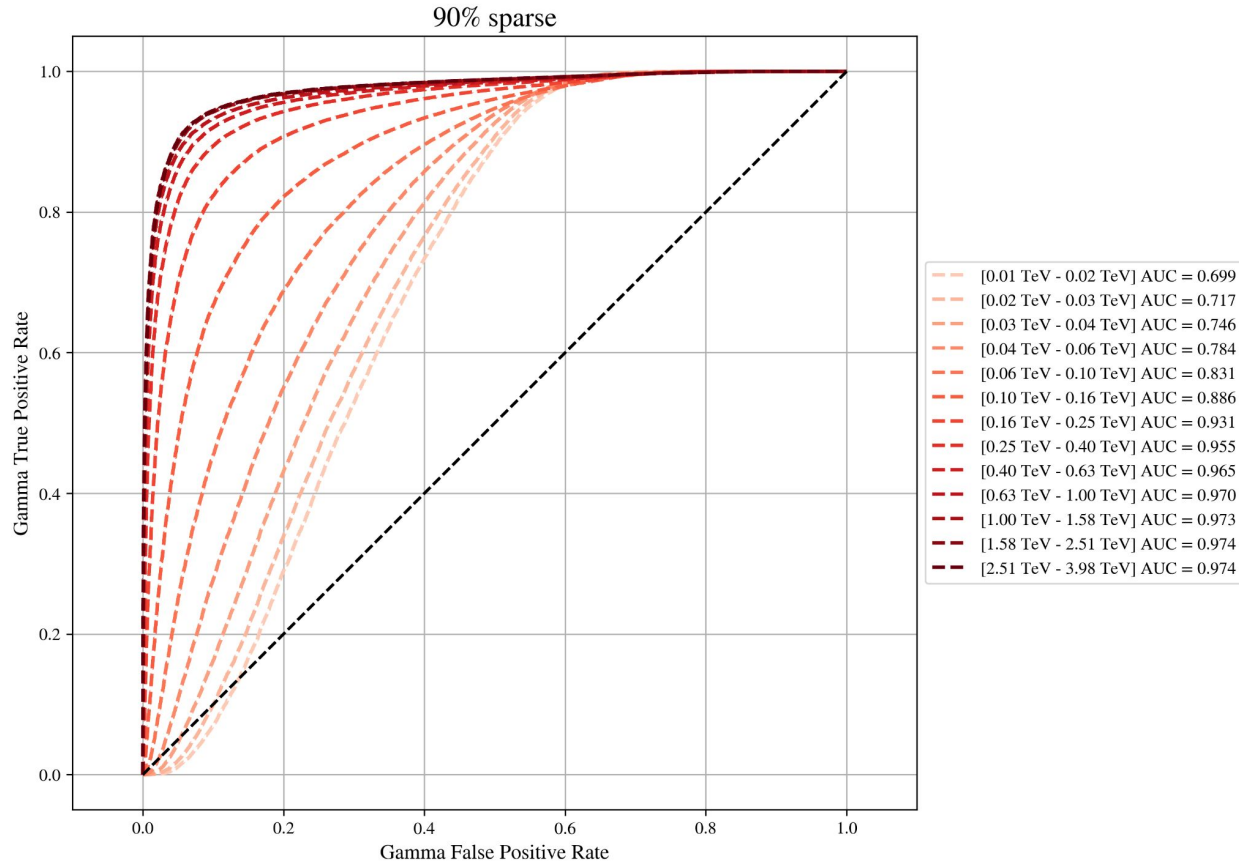
# Resultados: Tarea de clasificación



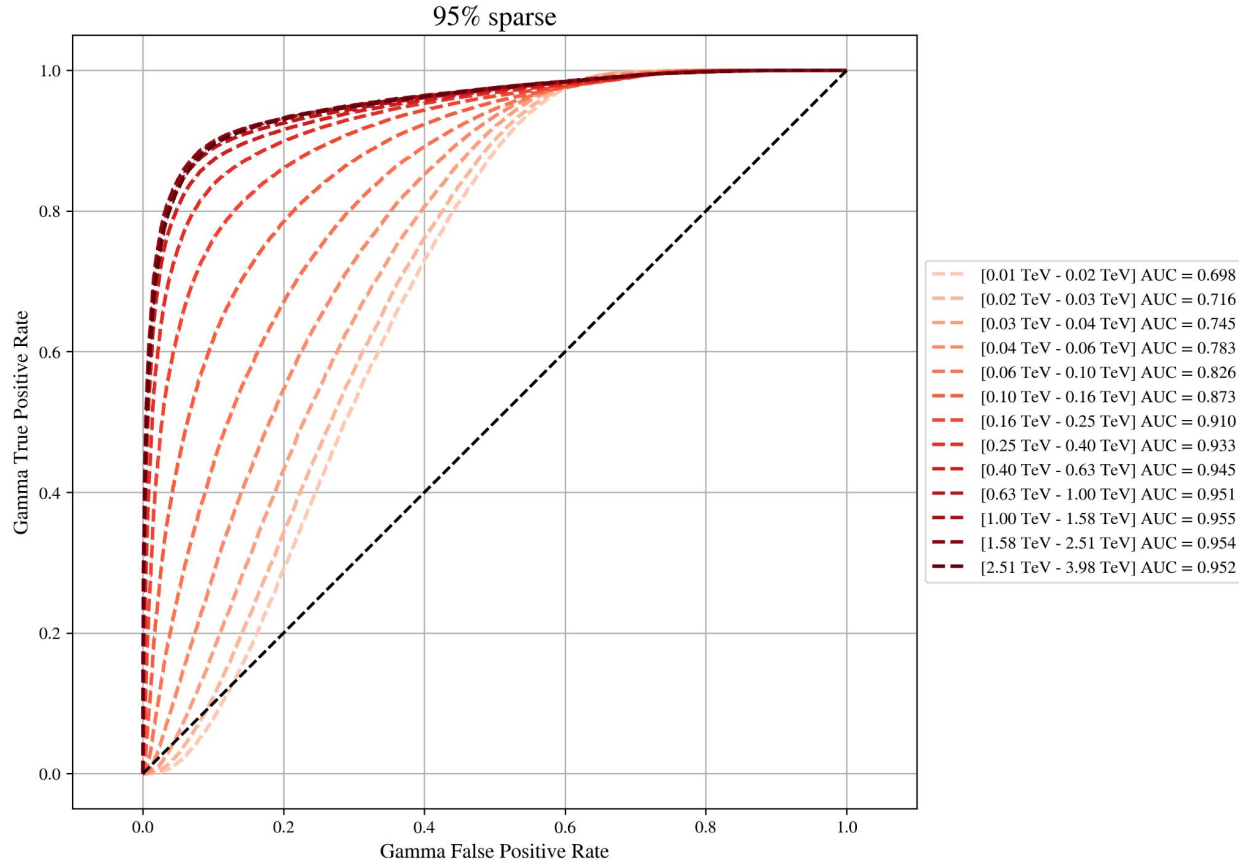
# Resultados: Tarea de clasificación



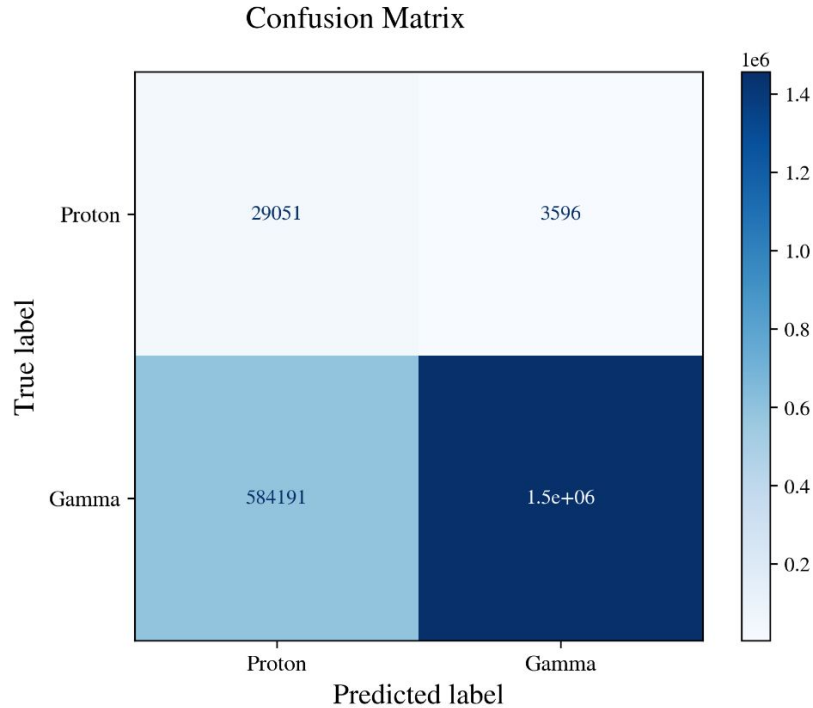
# Resultados: Tarea de clasificación



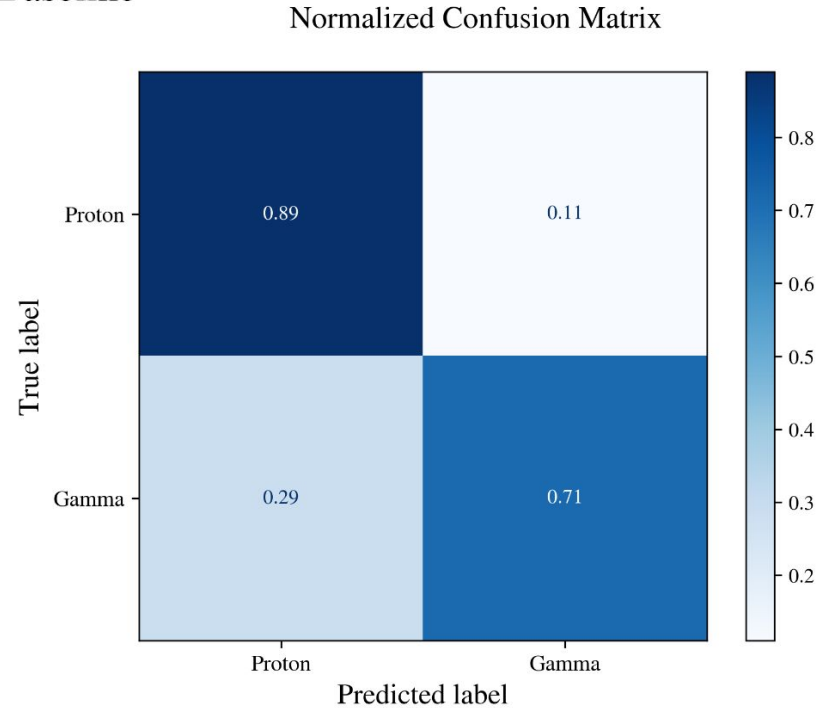
# Resultados: Tarea de clasificación



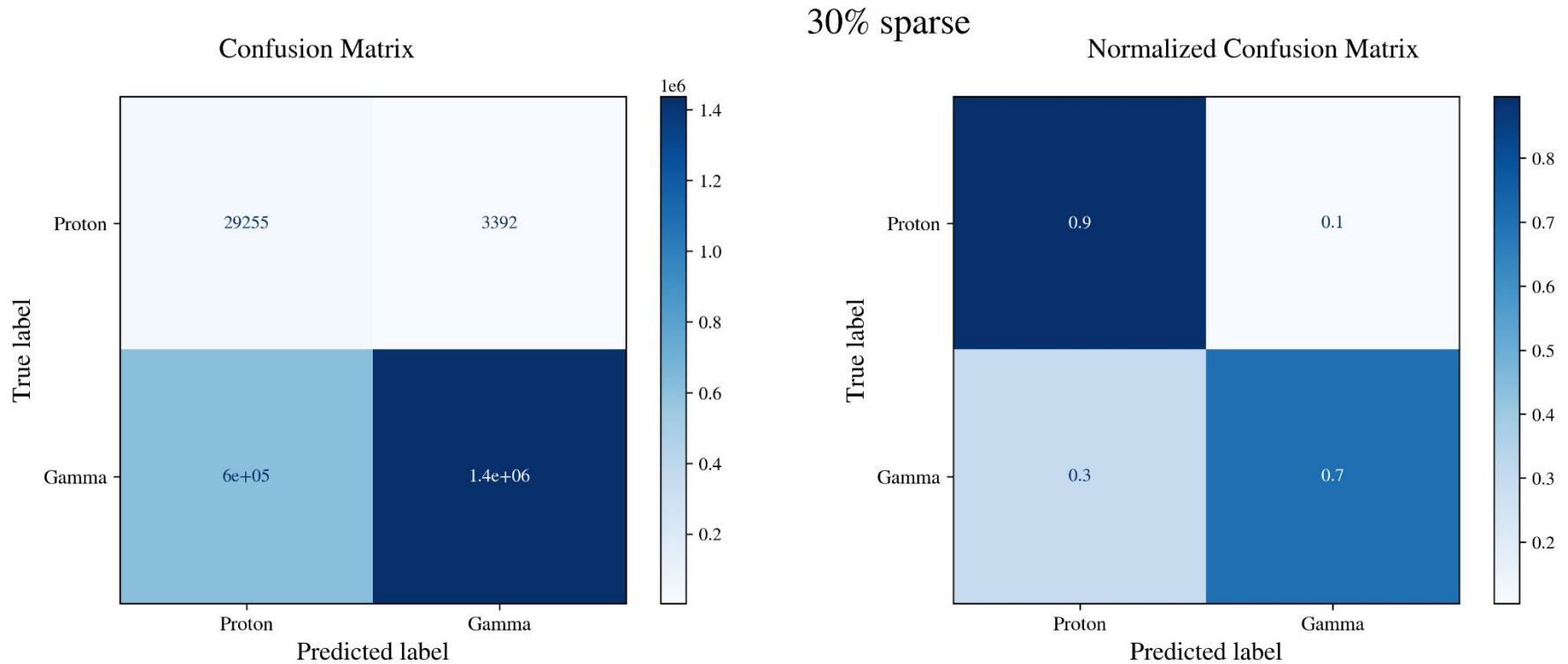
# Resultados: Tarea de clasificación



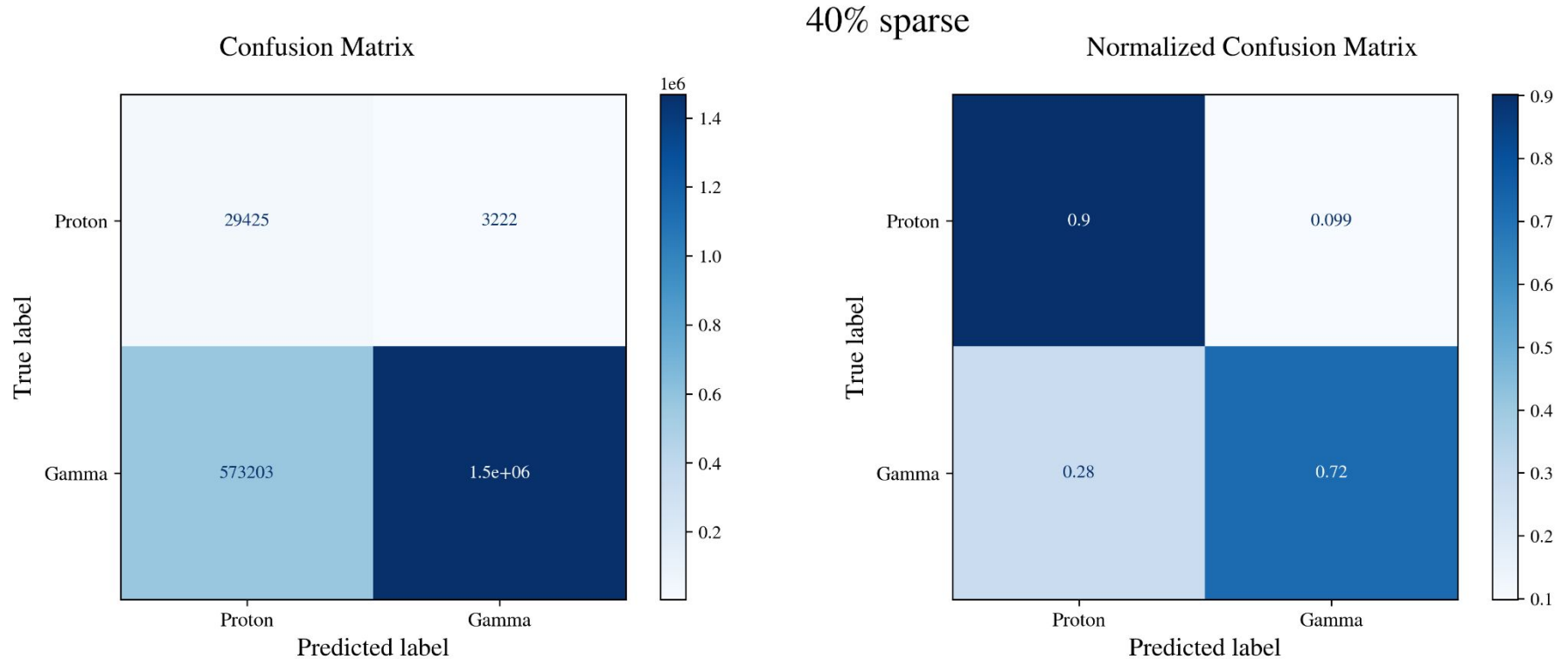
Baseline



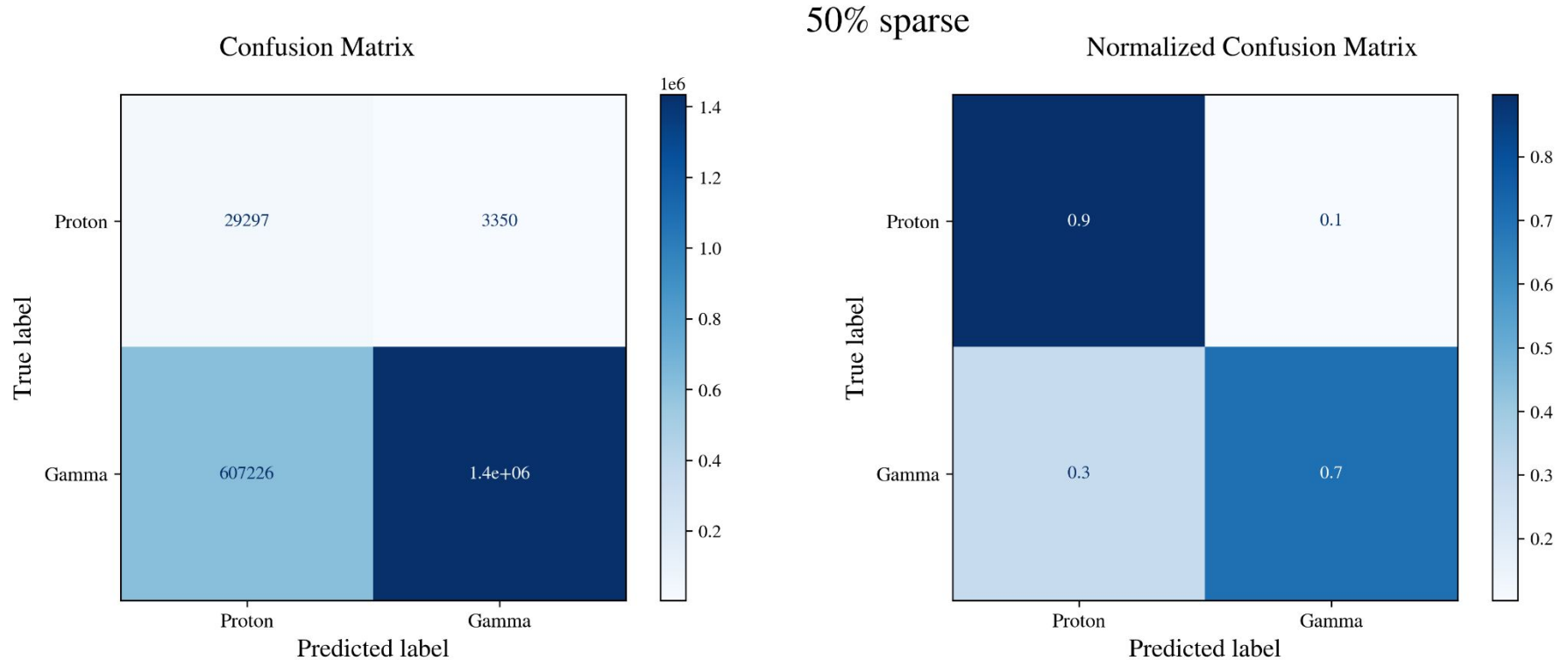
# Resultados: Tarea de clasificación



# Resultados: Tarea de clasificación

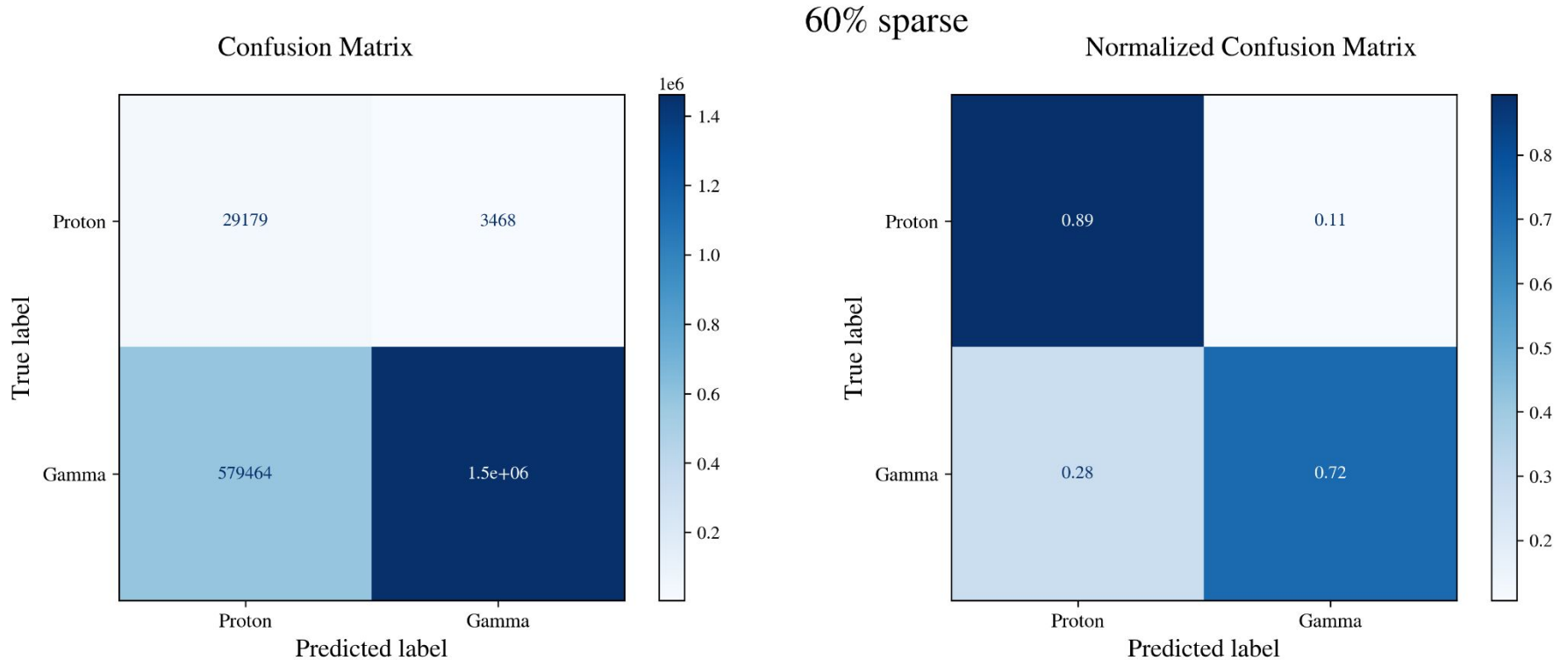


# Resultados: Tarea de clasificación

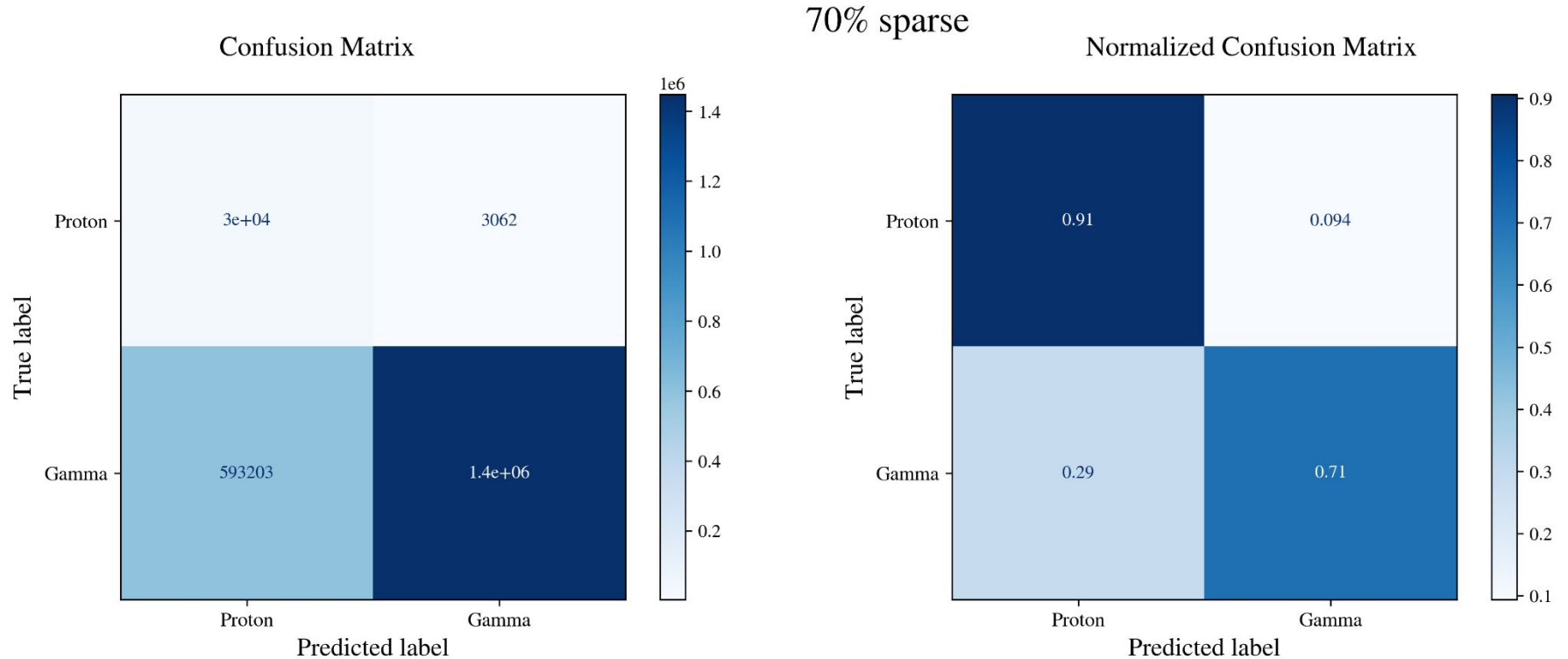




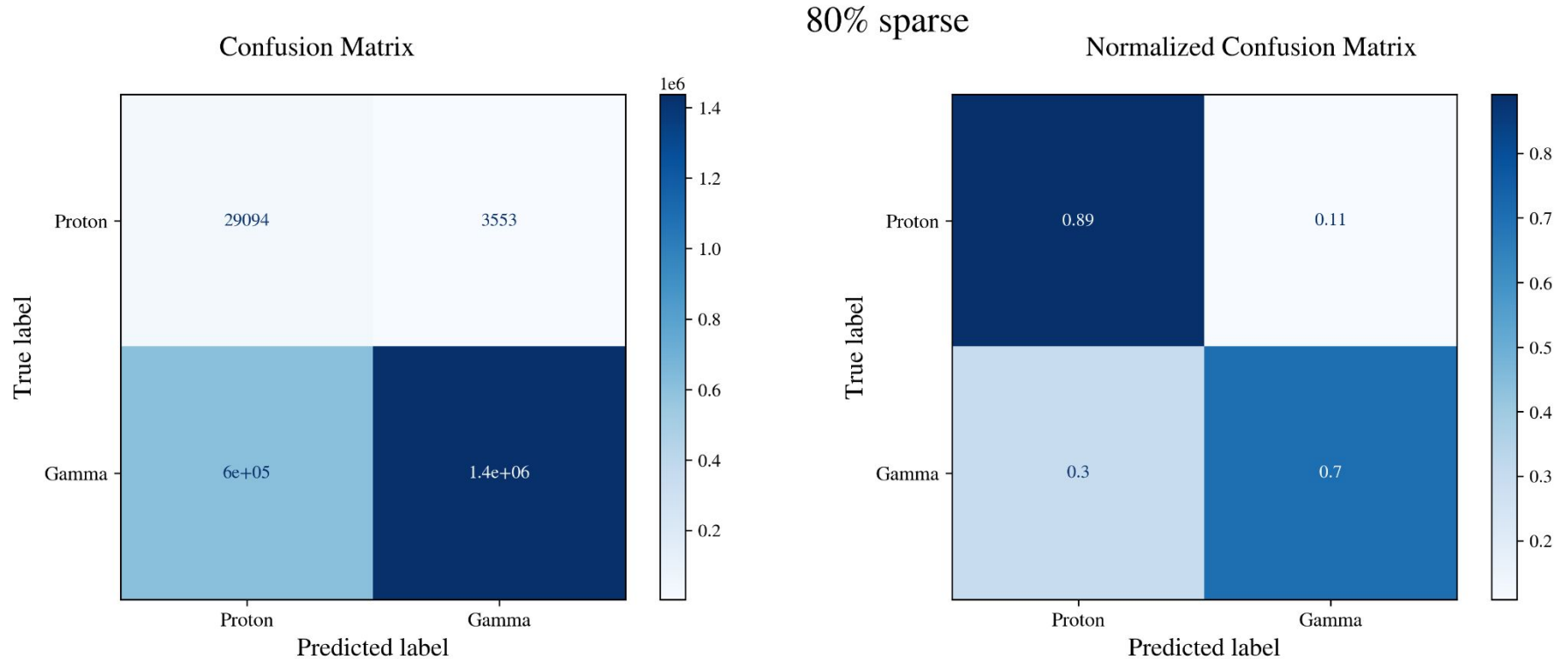
# Resultados: Tarea de clasificación



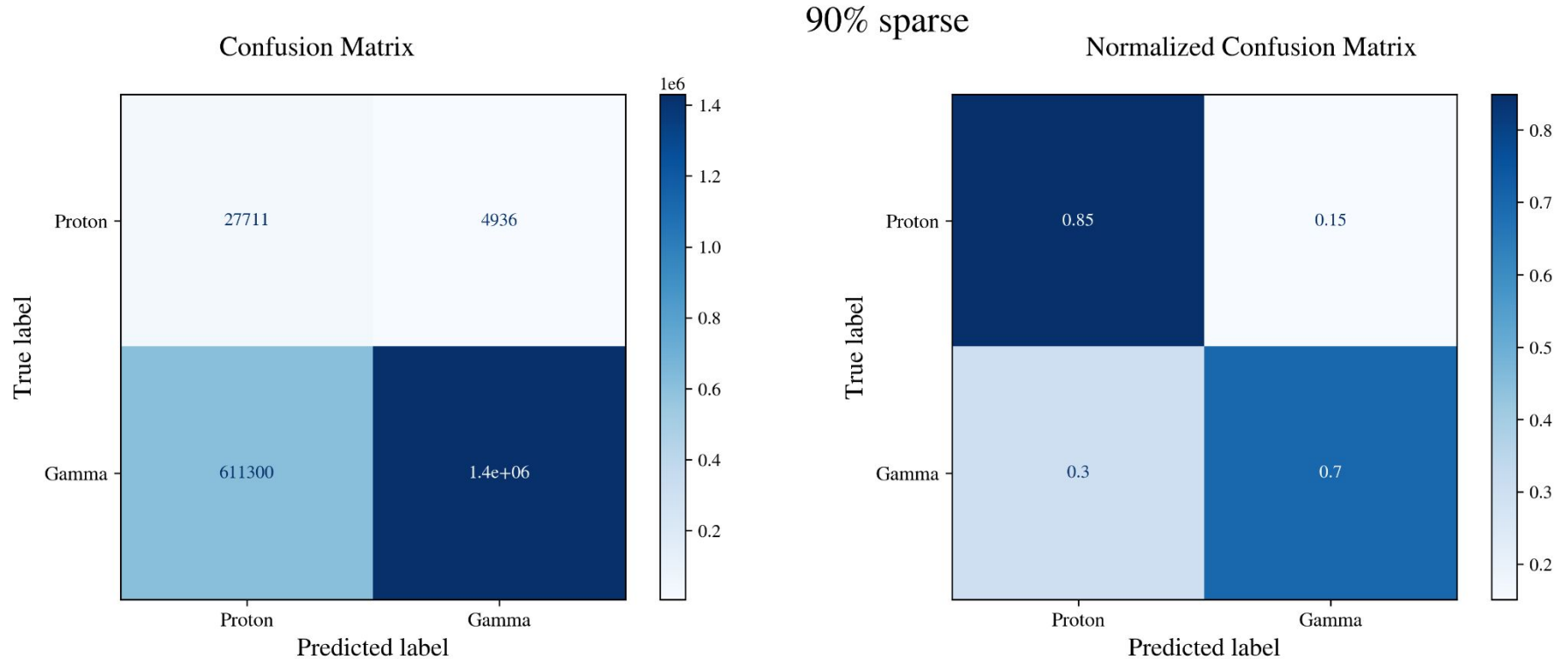
# Resultados: Tarea de clasificación



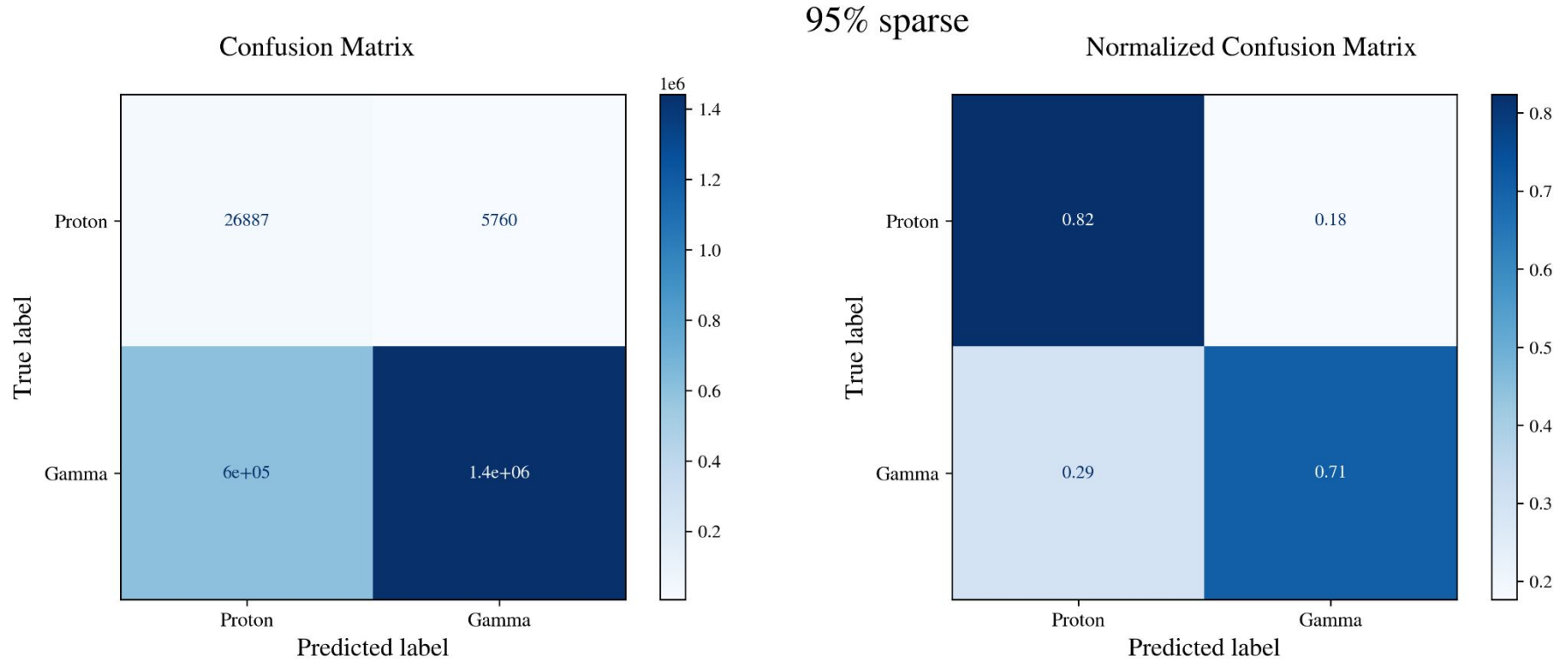
# Resultados: Tarea de clasificación



# Resultados: Tarea de clasificación



# Resultados: Tarea de clasificación



# Resultados: Tarea de clasificación

---

Datos:

Train:

- Gamma: 1,039,893
- Hadrónicos: 1,042,951

10% del conjunto de train se usa para validación.

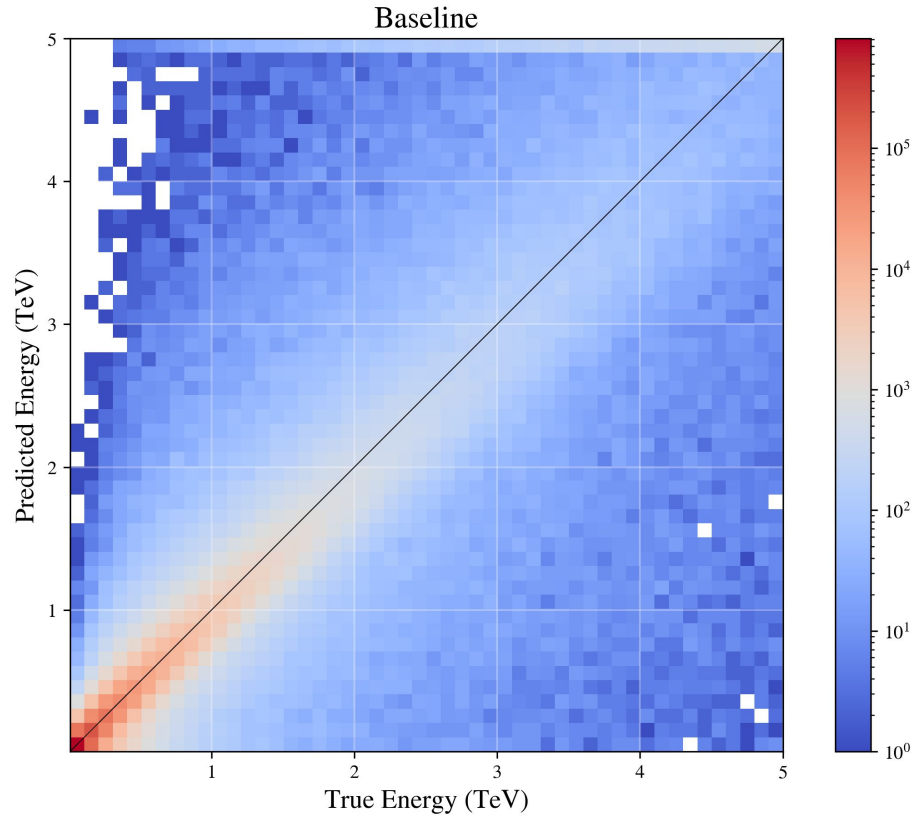
Test:

- Gamma: 2,041,124
- Hadrónicos: 32,647

# Resultados: Tarea de clasificación

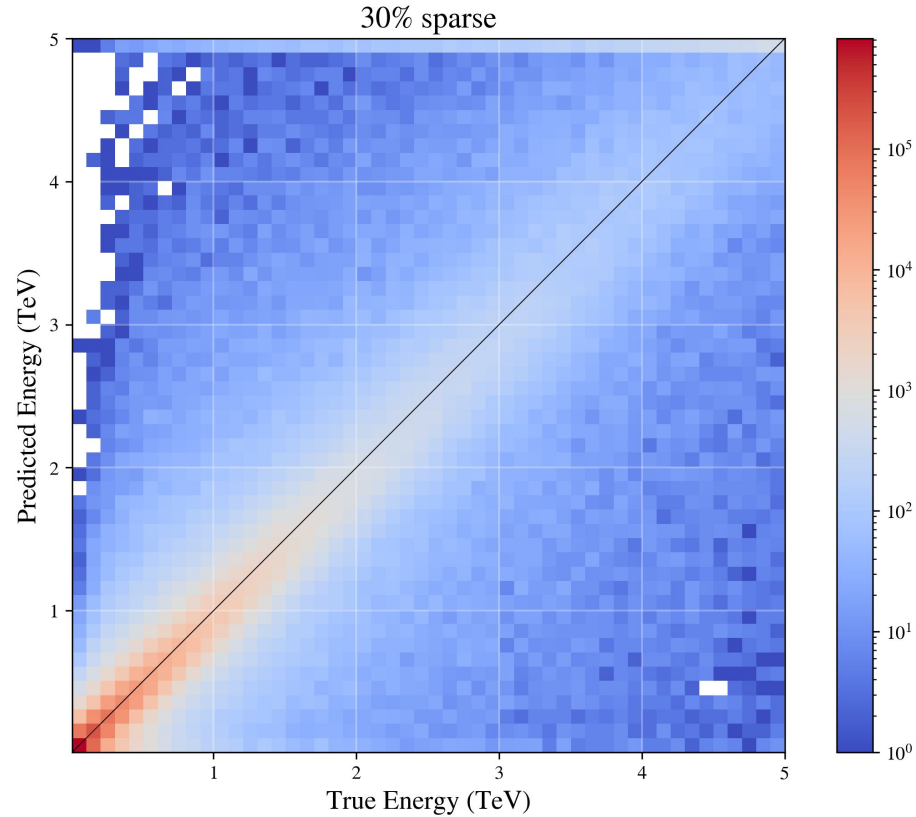
|   | <b>Metric</b>          | <b>Baseline</b> | <b>30%<br/>sparse</b> | <b>40%<br/>sparse</b> | <b>50%<br/>sparse</b> | <b>60%<br/>sparse</b> | <b>70%<br/>sparse</b> | <b>80%<br/>sparse</b> | <b>90%<br/>sparse</b> | <b>95%<br/>sparse</b> |
|---|------------------------|-----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 0 | Accuracy               | 71.66%          | 70.72%                | 72.20%                | 70.56%                | 71.89%                | 71.25%                | 70.71%                | 70.28%                | 70.78%                |
| 1 | Precision<br>(Class 0) | 0.05            | 0.05                  | 0.05                  | 0.05                  | 0.05                  | 0.05                  | 0.05                  | 0.04                  | 0.04                  |
| 2 | Precision<br>(Class 1) | 1.00            | 1.00                  | 1.00                  | 1.00                  | 1.00                  | 1.00                  | 1.00                  | 1.00                  | 1.00                  |
| 3 | Recall<br>(Class 0)    | 0.89            | 0.90                  | 0.90                  | 0.90                  | 0.89                  | 0.91                  | 0.89                  | 0.85                  | 0.82                  |
| 4 | Recall<br>(Class 1)    | 0.71            | 0.70                  | 0.72                  | 0.70                  | 0.72                  | 0.71                  | 0.70                  | 0.70                  | 0.71                  |
| 5 | F1 Score               | 0.83            | 0.83                  | 0.84                  | 0.82                  | 0.83                  | 0.83                  | 0.83                  | 0.82                  | 0.83                  |
| 6 | Support<br>(Class 0)   | 32647.0         | 32647.0               | 32647.0               | 32647.0               | 32647.0               | 32647.0               | 32647.0               | 32647.0               | 32647.0               |
| 7 | Support<br>(Class 1)   | 2041124.0       | 2041124.0             | 2041124.0             | 2041124.0             | 2041124.0             | 2041124.0             | 2041124.0             | 2041124.0             | 2041124.0             |
| 8 | AUC                    | 0.91            | 0.91                  | 0.92                  | 0.91                  | 0.91                  | 0.92                  | 0.91                  | 0.88                  | 0.86                  |

# Resultados: Tarea de regresión

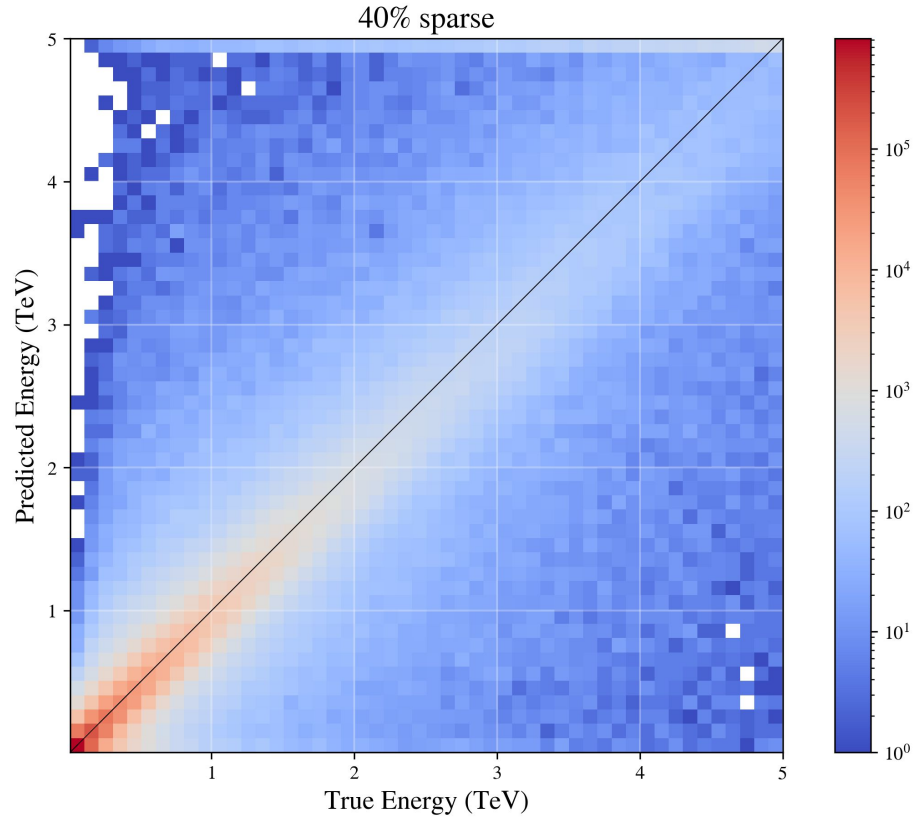




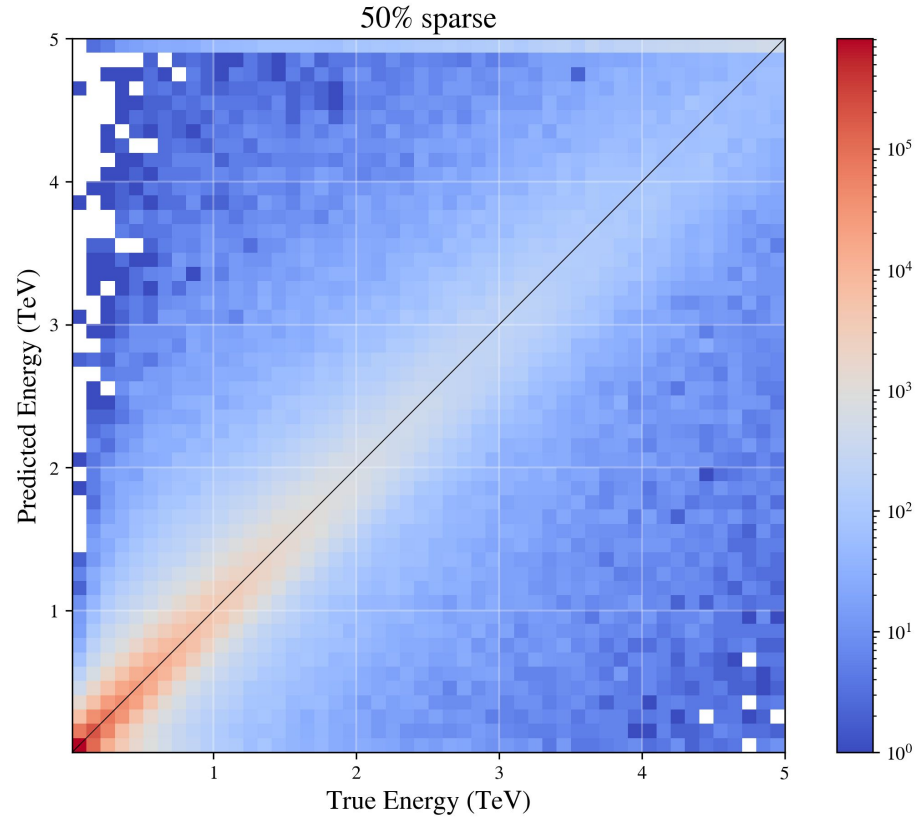
# Resultados: Tarea de regresión



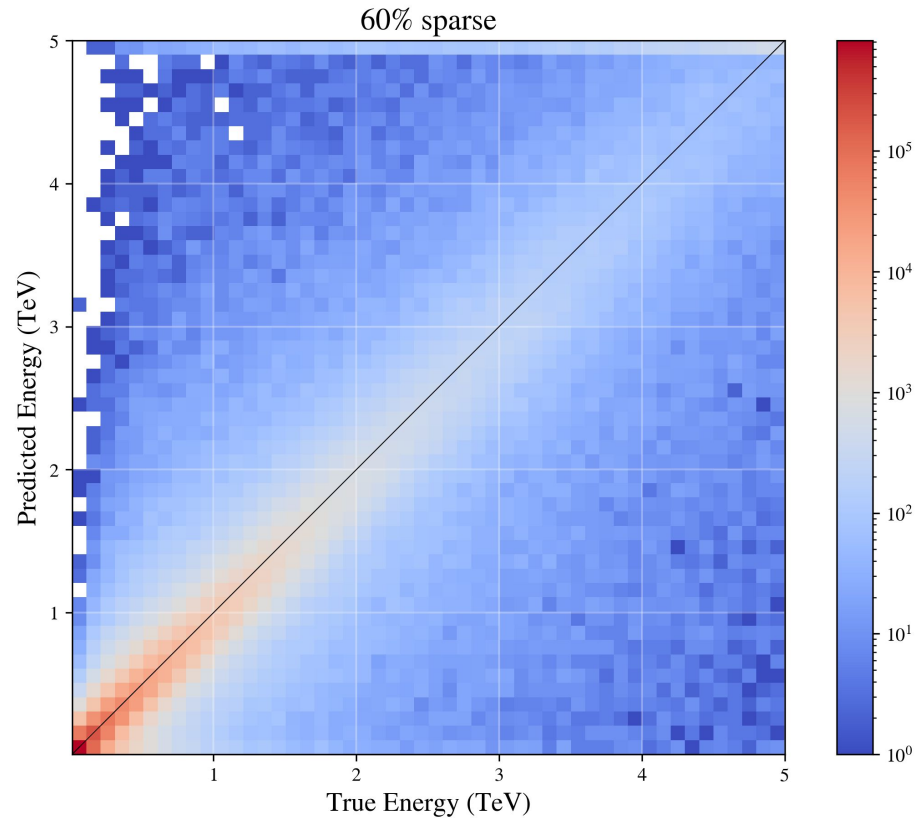
# Resultados: Tarea de regresión



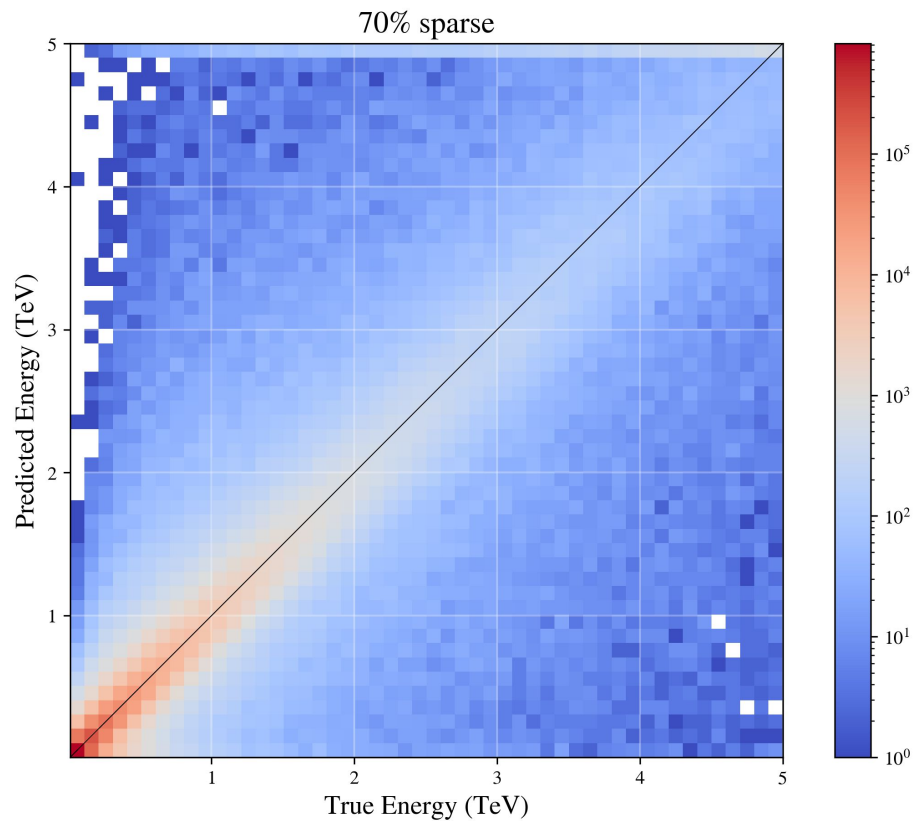
# Resultados: Tarea de regresión



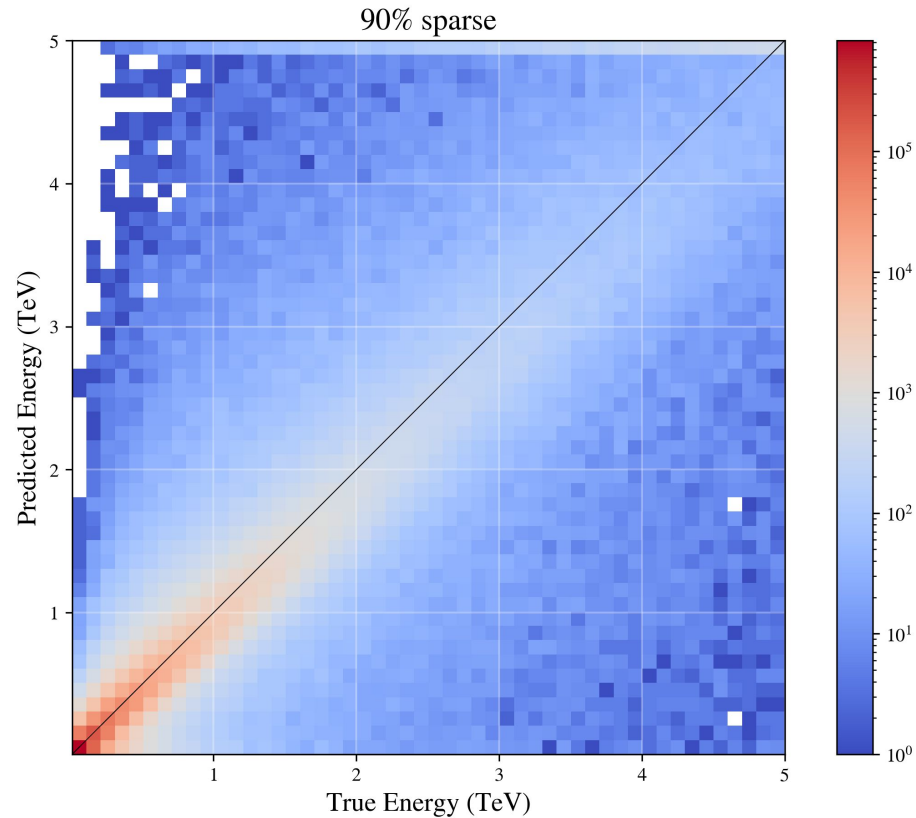
# Resultados: Tarea de regresión



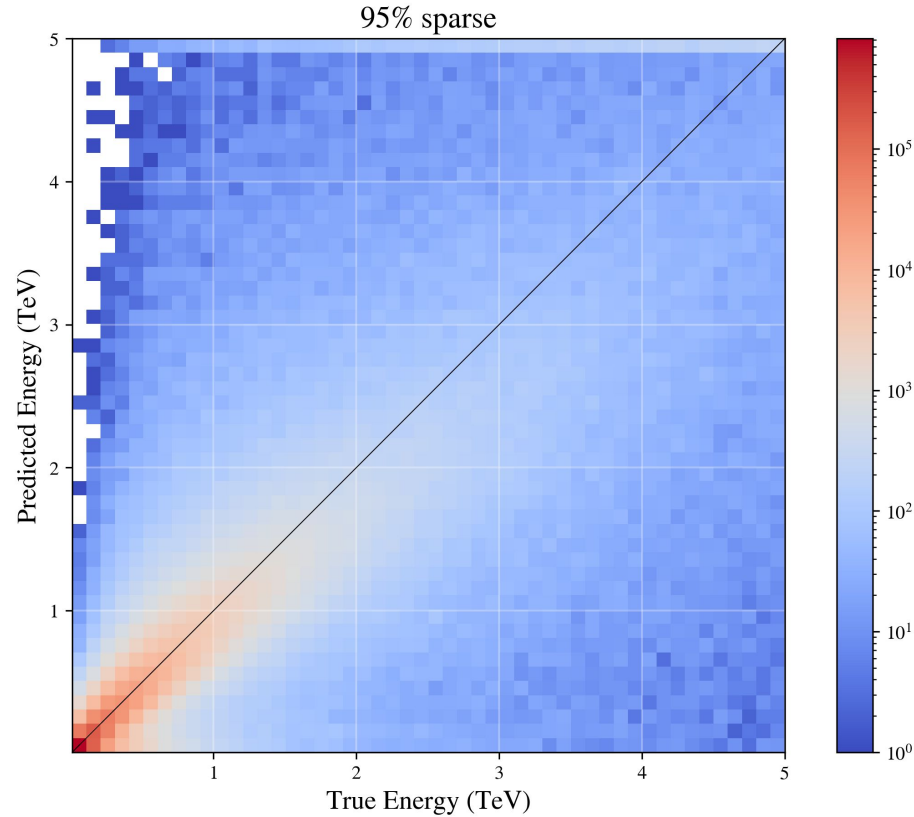
# Resultados: Tarea de regresión



# Resultados: Tarea de regresión



# Resultados: Tarea de regresión



## Resultados: Tarea de regresión

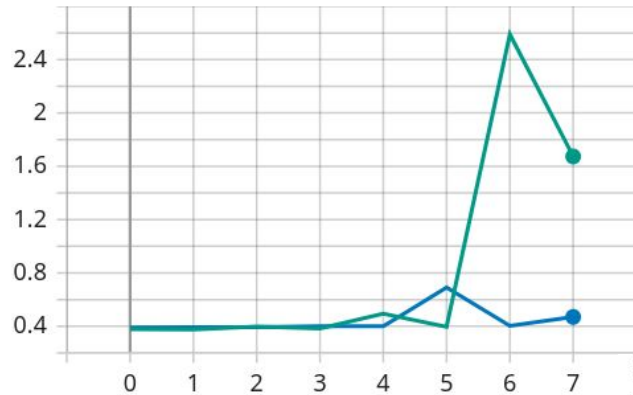
---

| Model size in (MB) | Baseline | 30%   | 40%   | 50%   | 60%   | 70%   | 80%  | 90%  | 95%  |
|--------------------|----------|-------|-------|-------|-------|-------|------|------|------|
| Type of shower     | 67.56    | 18.77 | 16.82 | 14.67 | 12.59 | 10.19 | 7.80 | 5.15 | 3.68 |
| Direction          | 69.05    | 18.74 | 16.78 | 14.72 | 12.53 | 10.23 | 7.77 | 5.14 | 3.62 |
| Energy             | 66.10    | 18.70 | 16.74 | 14.66 | 12.49 | 10.19 | 7.81 | 5.13 | 3.68 |

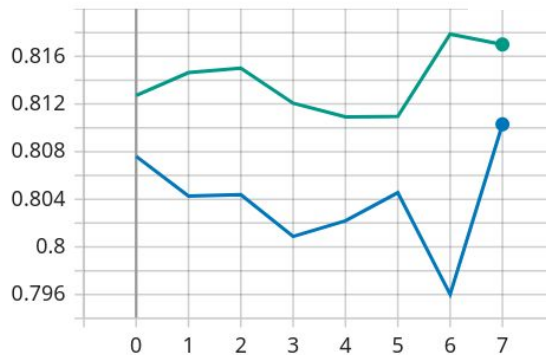


# Monitorización clasificación: TensorBoard

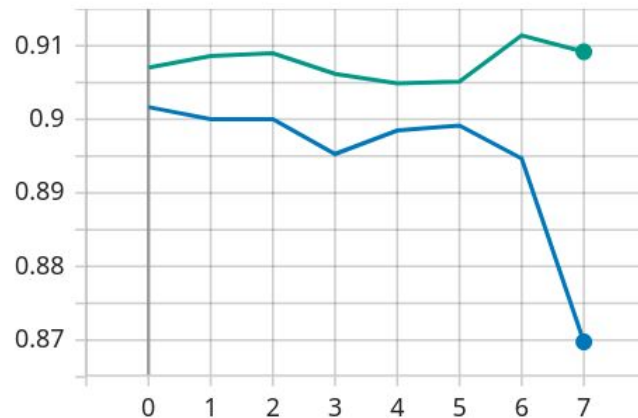
epoch\_loss  
tag: epoch\_loss



epoch\_accuracy  
tag: epoch\_accuracy

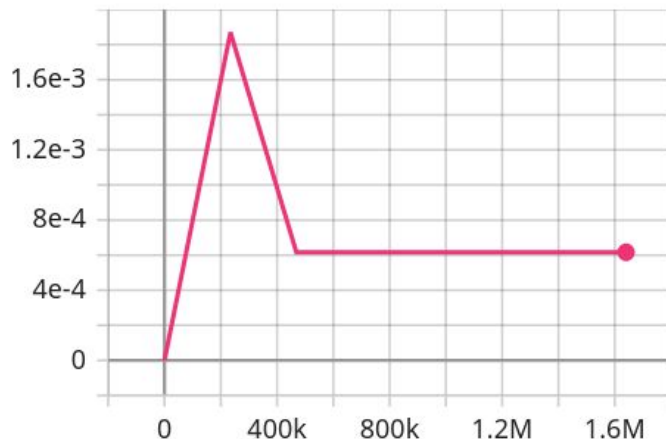


epoch\_auc  
tag: epoch\_auc

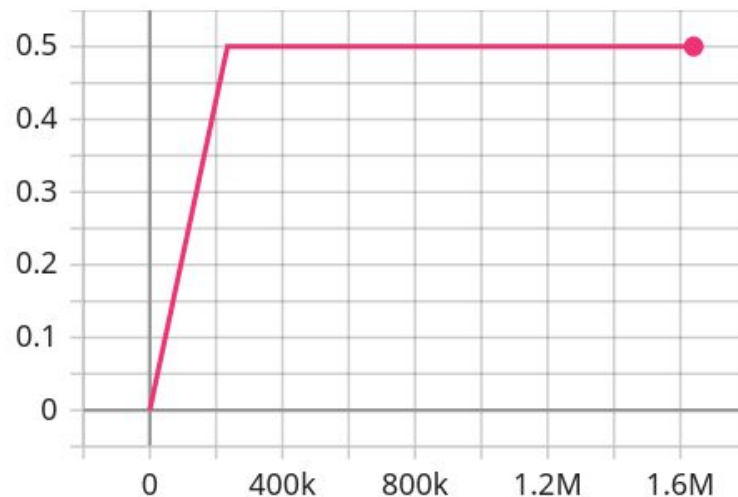


# Monitorización clasificación: TensorBoard

prune\_low\_magnitude\_type/threshold:0/  
threshold  
tag: prune\_low\_magnitude\_type/threshold:0/  
threshold

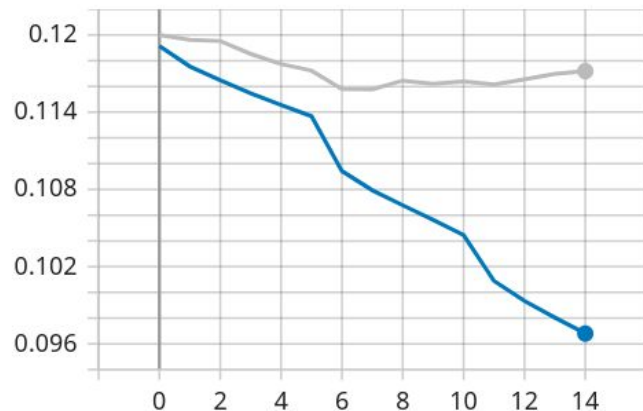


prune\_low\_magnitude\_type/mask:0/sparsity  
tag: prune\_low\_magnitude\_type/mask:0/sparsity

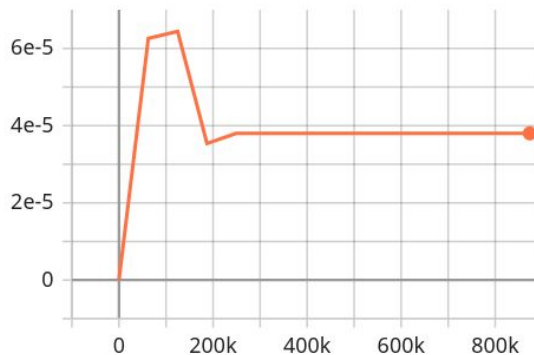


# Monitorización: TensorBoard

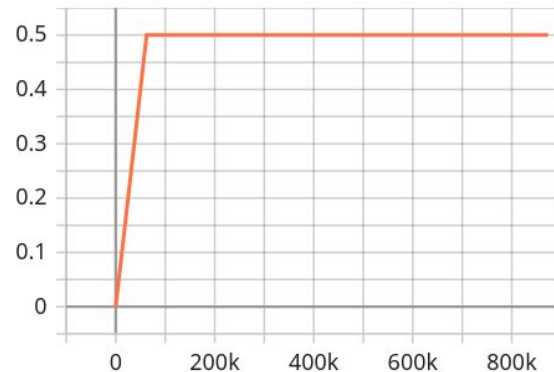
epoch\_loss  
tag: epoch\_loss



prune\_low\_magnitude\_cameradirection/  
threshold:0/threshold  
tag: prune\_low\_magnitude\_cameradirection/  
threshold:0/threshold

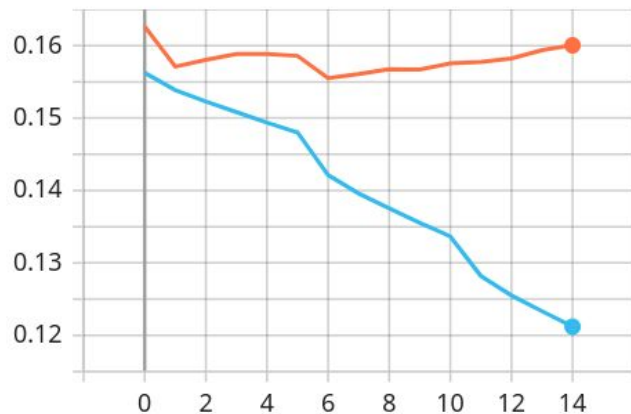


prune\_low\_magnitude\_cameradirection/mask:0/  
sparsity  
tag: prune\_low\_magnitude\_cameradirection/  
mask:0/sparsity

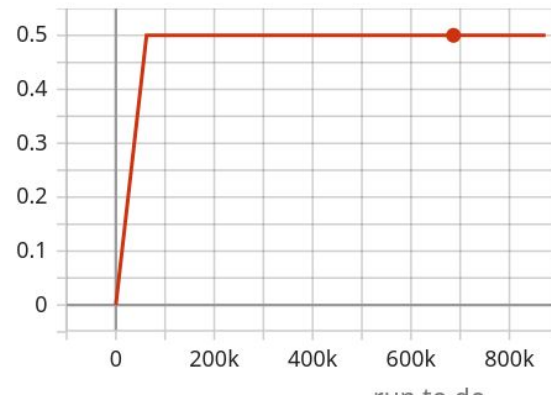


# Monitorización: TensorBoard

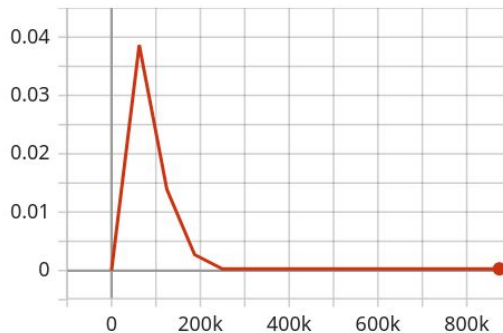
epoch\_loss  
tag: epoch\_loss



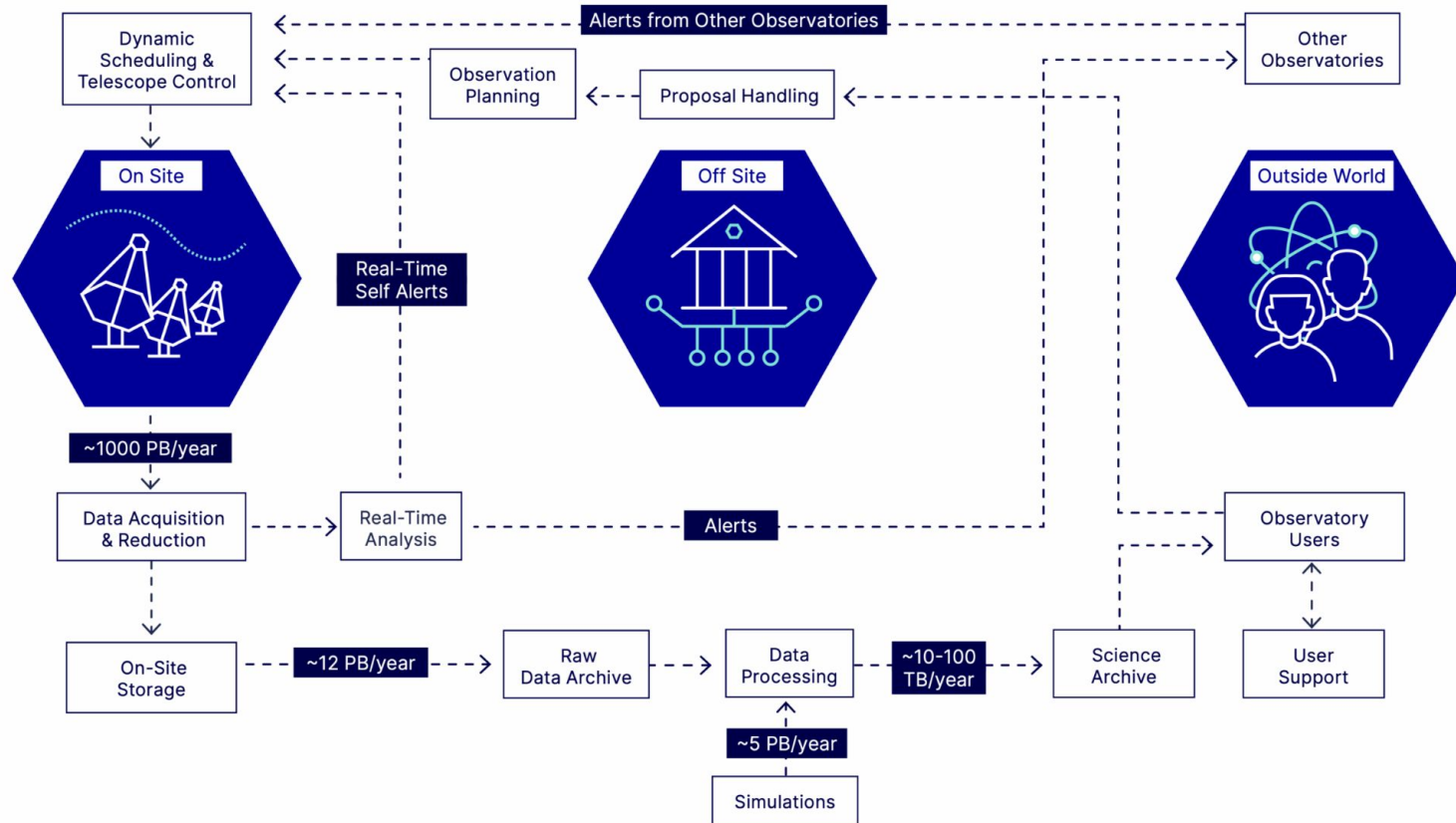
prune\_low\_magnitude\_energy/mask:0/sparsity  
tag: prune\_low\_magnitude\_energy/mask:0/  
sparsity



prune\_low\_magnitude\_energy/threshold:0/  
threshold  
tag: prune\_low\_magnitude\_energy/threshold:0/  
threshold



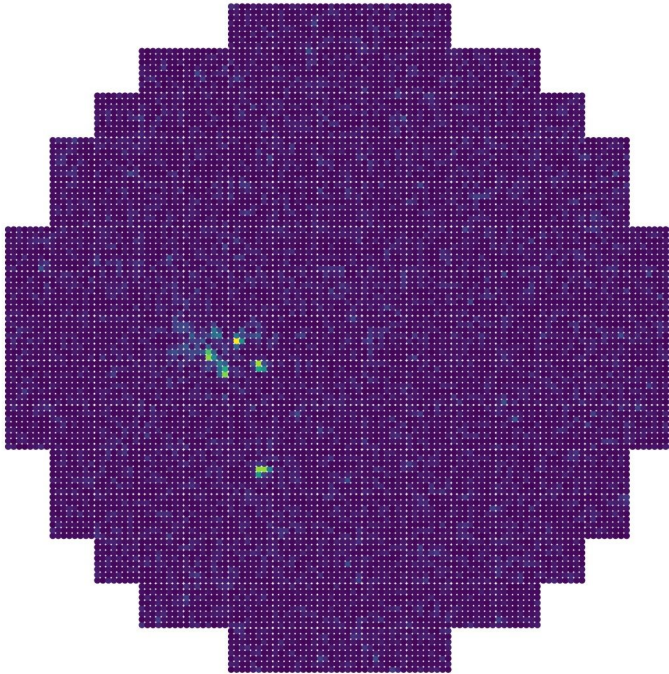
# Data Rates



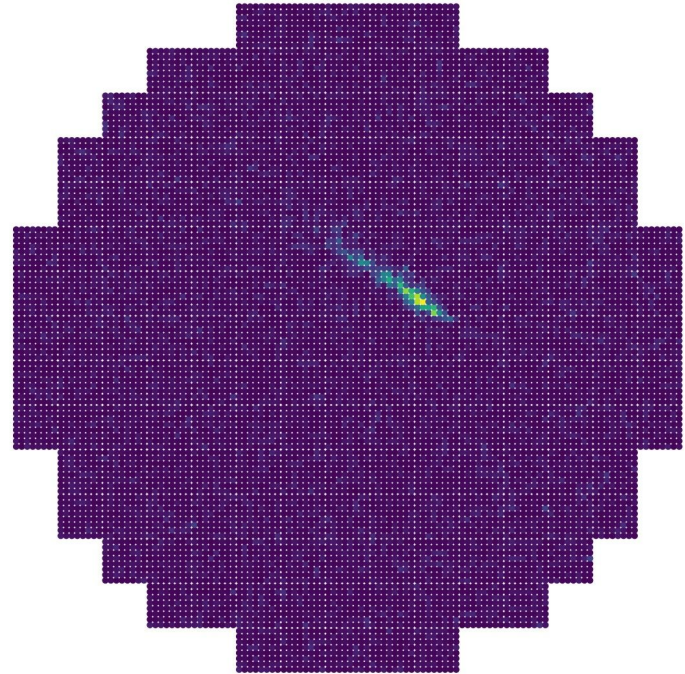
# Cascada procesada

---

Hadrónica



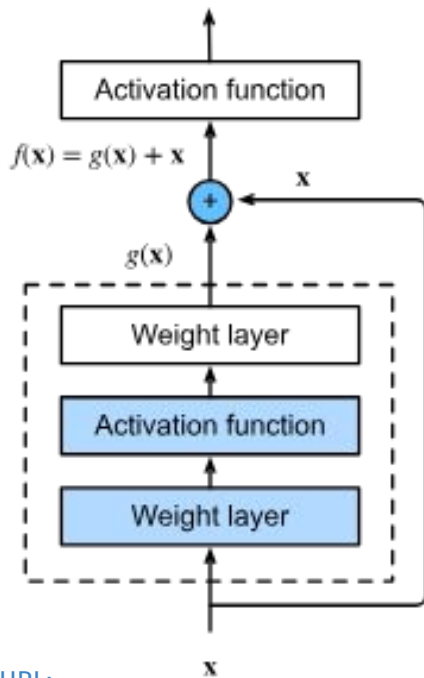
Gamma



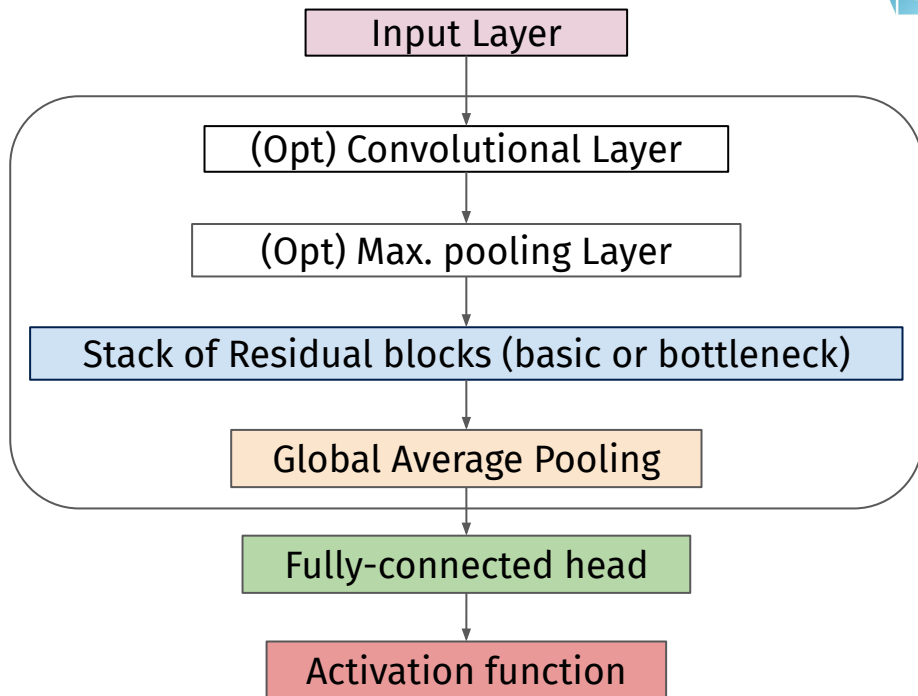
# CTLearn: Arquitectura del Modelo backup



## Bloque Residual

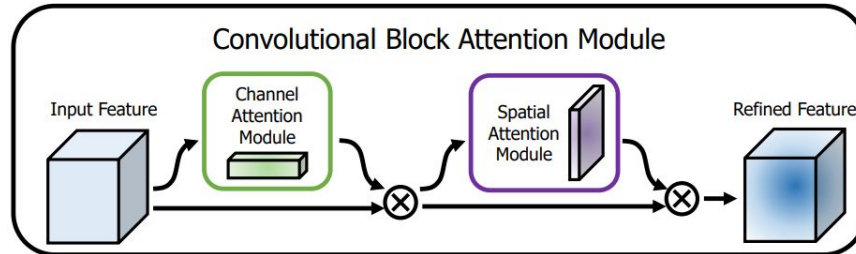
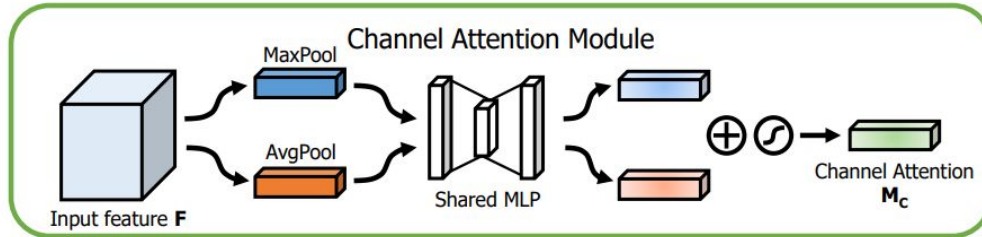
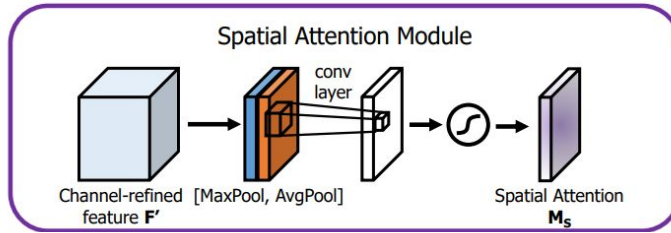


## ResNet

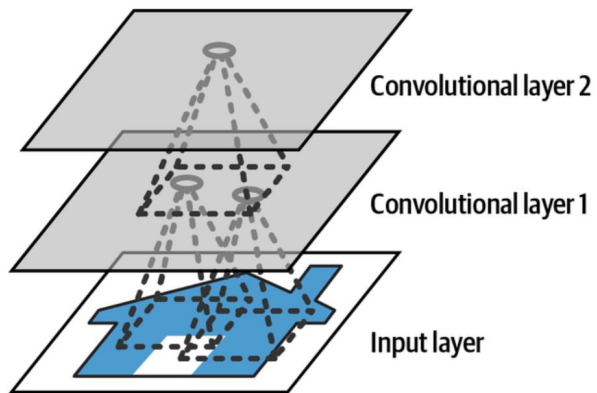




# Mecanismo de atención *squeeze-excite* dual

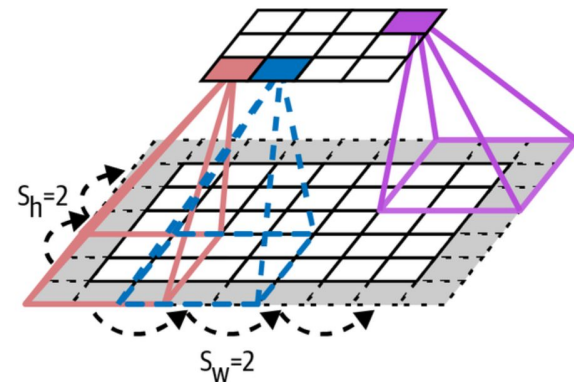




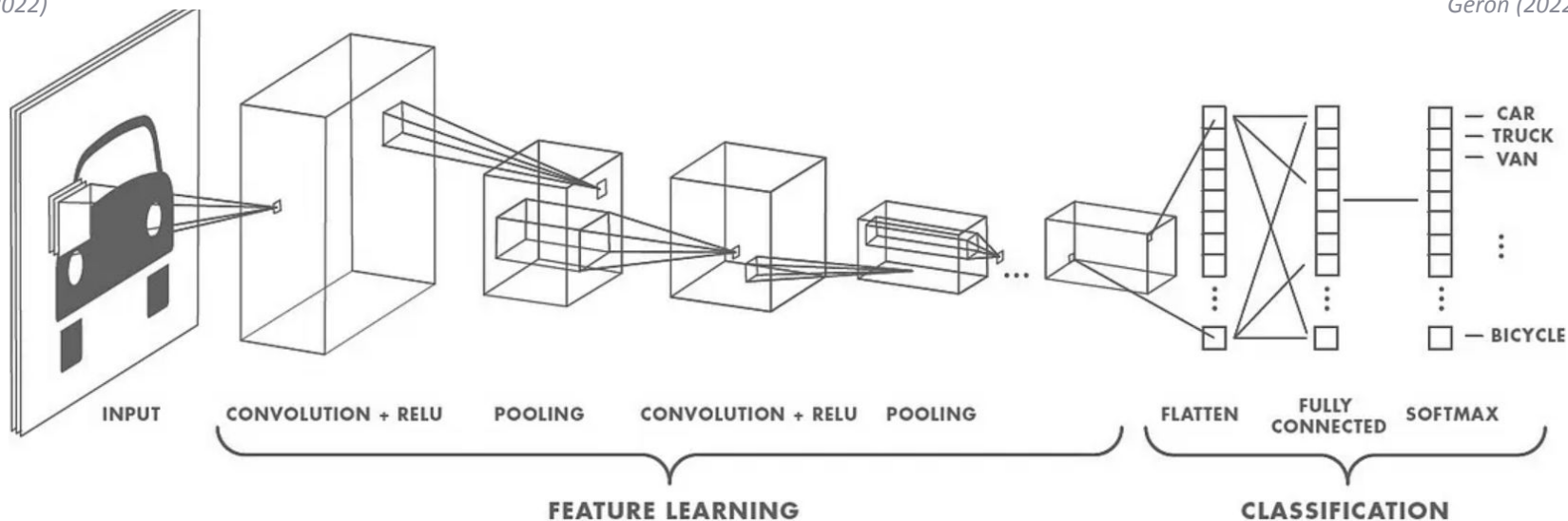


Geron (2022)

## Red convolutional

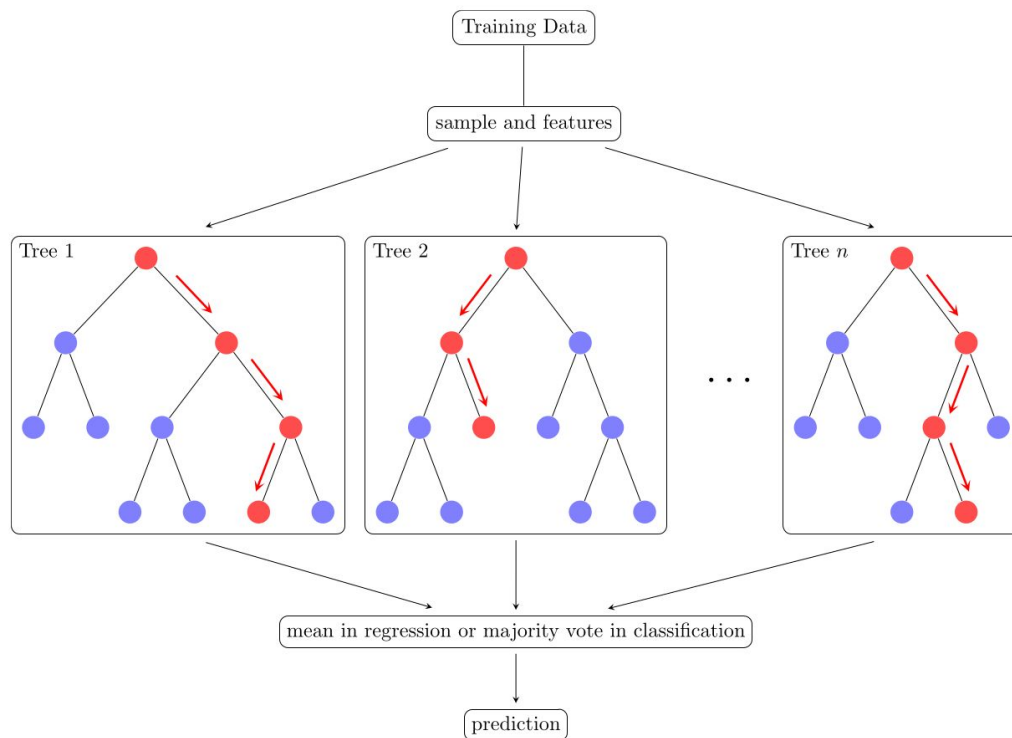


Geron (2022)



|                                                   |                                                                        |                              |                                                   |                            |                                                    |                            |                                          |                    |                                             |                    |
|---------------------------------------------------|------------------------------------------------------------------------|------------------------------|---------------------------------------------------|----------------------------|----------------------------------------------------|----------------------------|------------------------------------------|--------------------|---------------------------------------------|--------------------|
| <div> <div>QUARKS</div> <div>LEPTONS</div> </div> | mass → $\approx 2.3 \text{ MeV}/c^2$<br>charge → $2/3$<br>spin → $1/2$ | $u$<br>up                    | $\approx 1.275 \text{ GeV}/c^2$<br>$2/3$<br>$1/2$ | $c$<br>charm               | $\approx 173.07 \text{ GeV}/c^2$<br>$2/3$<br>$1/2$ | $t$<br>top                 | $0$<br>$0$<br>$1$                        | $g$<br>gluon       | $\approx 126 \text{ GeV}/c^2$<br>$0$<br>$0$ | $H$<br>Higgs boson |
|                                                   |                                                                        |                              |                                                   |                            |                                                    |                            |                                          |                    |                                             |                    |
|                                                   |                                                                        |                              |                                                   |                            |                                                    |                            |                                          |                    |                                             |                    |
|                                                   |                                                                        |                              |                                                   |                            |                                                    |                            |                                          |                    |                                             |                    |
|                                                   | $\approx 4.8 \text{ MeV}/c^2$<br>$-1/3$<br>$1/2$                       | $d$<br>down                  | $\approx 95 \text{ MeV}/c^2$<br>$-1/3$<br>$1/2$   | $s$<br>strange             | $\approx 4.18 \text{ GeV}/c^2$<br>$-1/3$<br>$1/2$  | $b$<br>bottom              | $0$<br>$0$<br>$1$                        | $\gamma$<br>photon |                                             |                    |
|                                                   |                                                                        |                              |                                                   |                            |                                                    |                            |                                          |                    |                                             |                    |
|                                                   |                                                                        |                              |                                                   |                            |                                                    |                            |                                          |                    |                                             |                    |
|                                                   |                                                                        |                              |                                                   |                            |                                                    |                            |                                          |                    |                                             |                    |
|                                                   | $0.511 \text{ MeV}/c^2$<br>$-1$<br>$1/2$                               | $e$<br>electron              | $105.7 \text{ MeV}/c^2$<br>$-1$<br>$1/2$          | $\mu$<br>muon              | $1.777 \text{ GeV}/c^2$<br>$-1$<br>$1/2$           | $\tau$<br>tau              | $91.2 \text{ GeV}/c^2$<br>$0$<br>$1$     | $Z$<br>Z boson     | <div>GAUGE BOSONS</div>                     |                    |
|                                                   |                                                                        |                              |                                                   |                            |                                                    |                            |                                          |                    |                                             |                    |
|                                                   |                                                                        |                              |                                                   |                            |                                                    |                            |                                          |                    |                                             |                    |
|                                                   |                                                                        |                              |                                                   |                            |                                                    |                            |                                          |                    |                                             |                    |
|                                                   | $< 2.2 \text{ eV}/c^2$<br>$0$<br>$1/2$                                 | $\nu_e$<br>electron neutrino | $< 0.17 \text{ MeV}/c^2$<br>$0$<br>$1/2$          | $\nu_\mu$<br>muon neutrino | $< 15.5 \text{ MeV}/c^2$<br>$0$<br>$1/2$           | $\nu_\tau$<br>tau neutrino | $80.4 \text{ GeV}/c^2$<br>$\pm 1$<br>$1$ | $W$<br>W boson     |                                             |                    |

# Random Forest



# TensorFlow: Pruning schedule polynomial

---

$$s(t) = s_f + (s_i - s_f) \cdot \left(1 - \frac{t - t_i}{t_f - t_i}\right)^{\text{exponent}}$$

$$t_f = \frac{\text{np.ceil}(N_{\text{images}})}{\text{batchsize}} * n_{\text{epochs}} - c$$

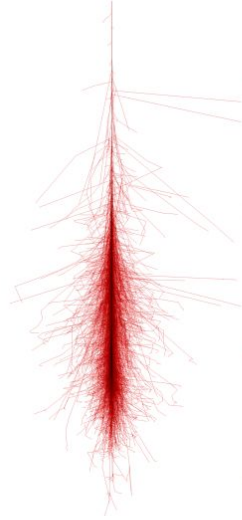
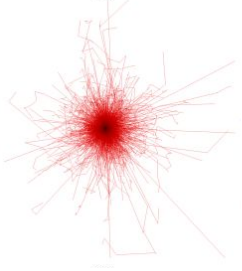
# Full integer quantization of weights and activations

## Dynamic range quantisation

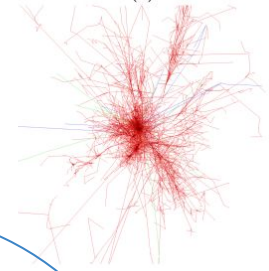
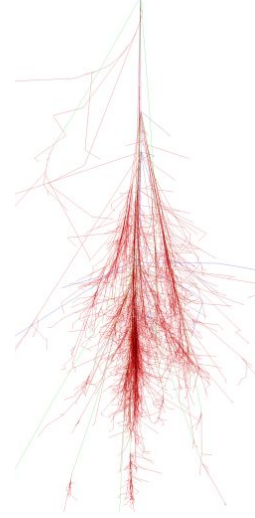
The activations are always stored in floating point.

activations are quantized to 8 bits of precision dynamically prior to processing and are de-quantized to float precision after processing

Cascada  
Electromagnética  
(Rayo  $\gamma$ )



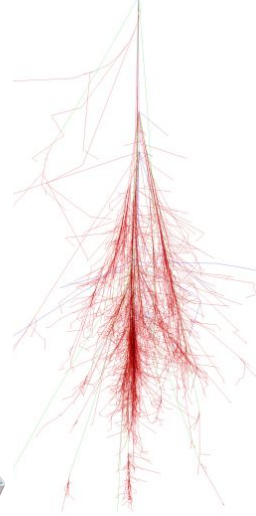
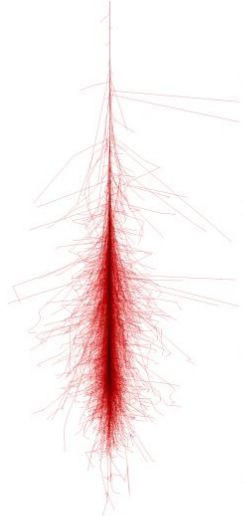
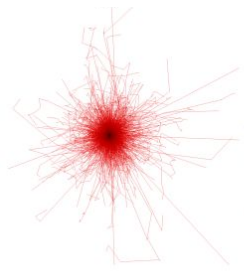
Cascada Hadrónica  
(Protón,  $\alpha$ )



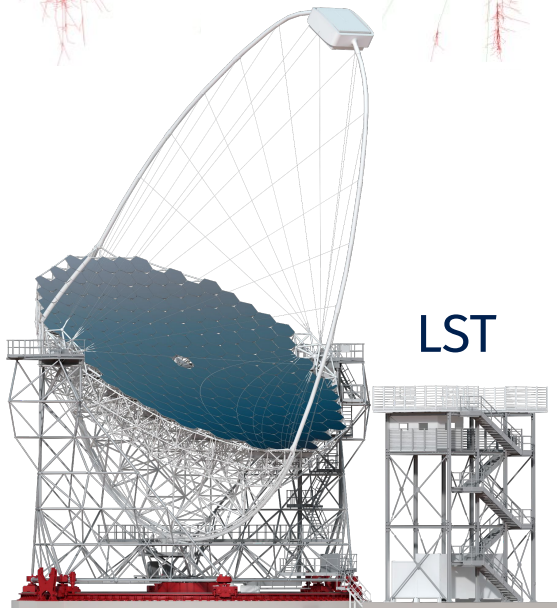
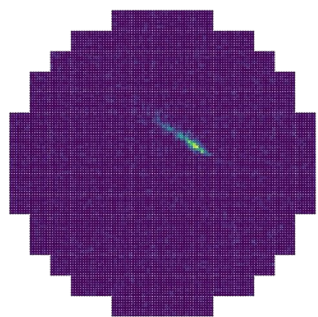
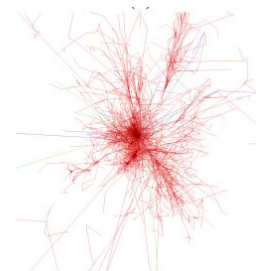
Datos simulados

¡Cargadas!

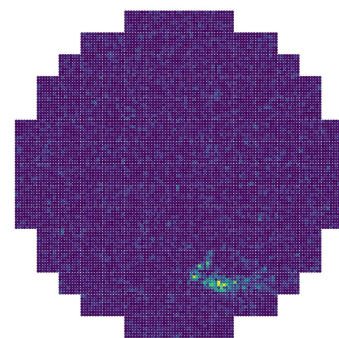
Cascada  
Electromagnética  
(Rayo  $\gamma$ )



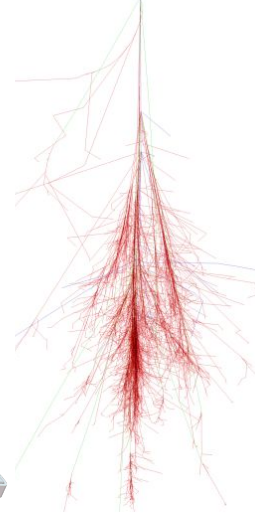
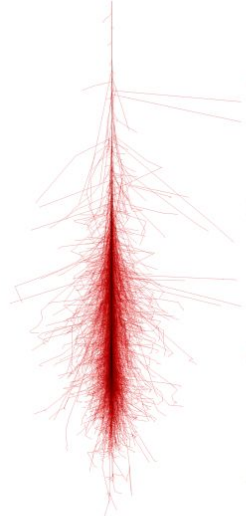
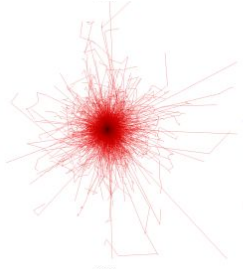
Cascada Hadrónica  
(Protón,  $\alpha$ )



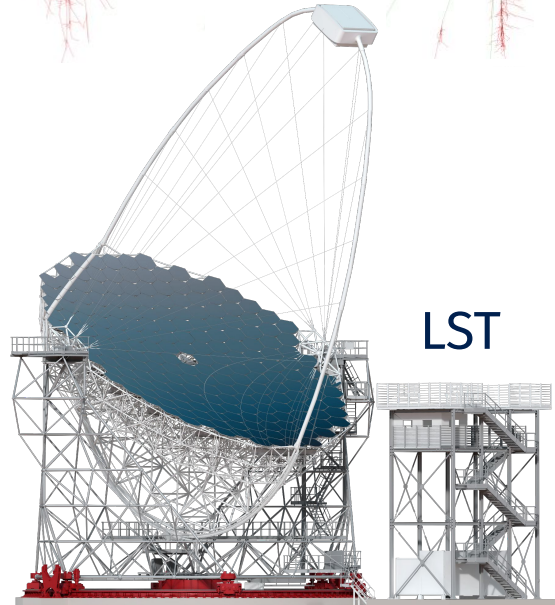
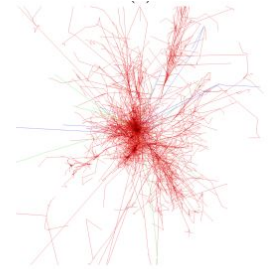
LST



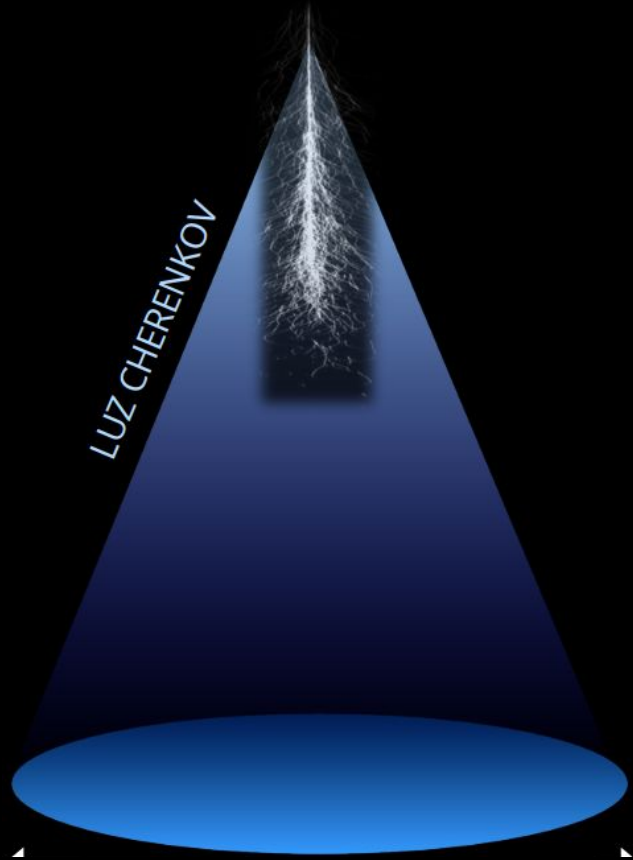
## Cascada Electromagnética (Rayo $\gamma$ )



## Cascada Hadrónica (Protón, $\alpha$ )







## Introducción

---

- Cascada atmosférica
- Radiación Cherenkov