

Neutral particles identification in the LHCb calorimeter using Machine Learning techniques



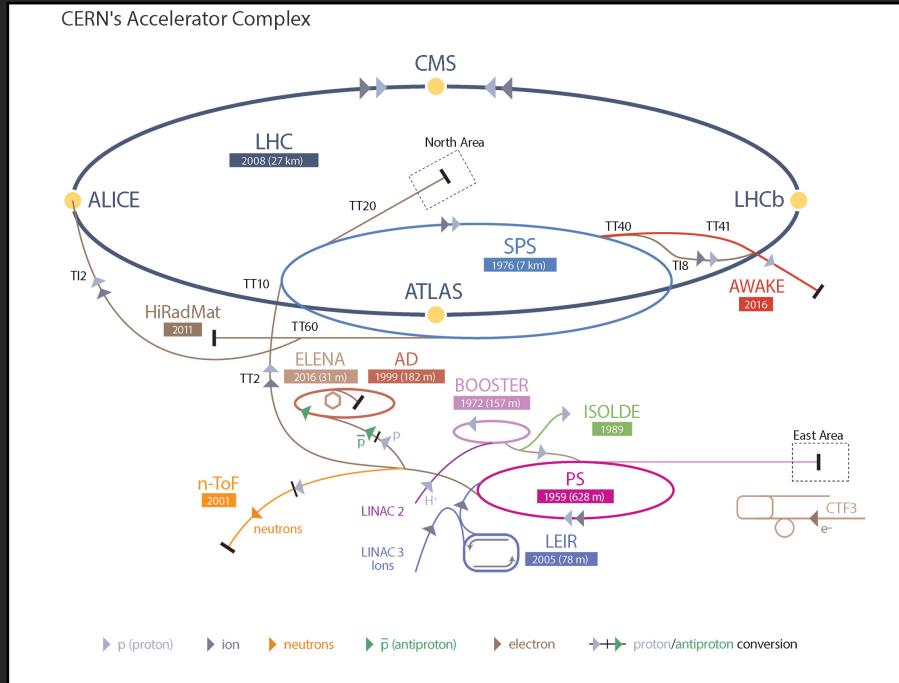
ICERN

- Organització Europea per a la Recerca Nuclear.
- 17500 persones.
- 23 estats membres.



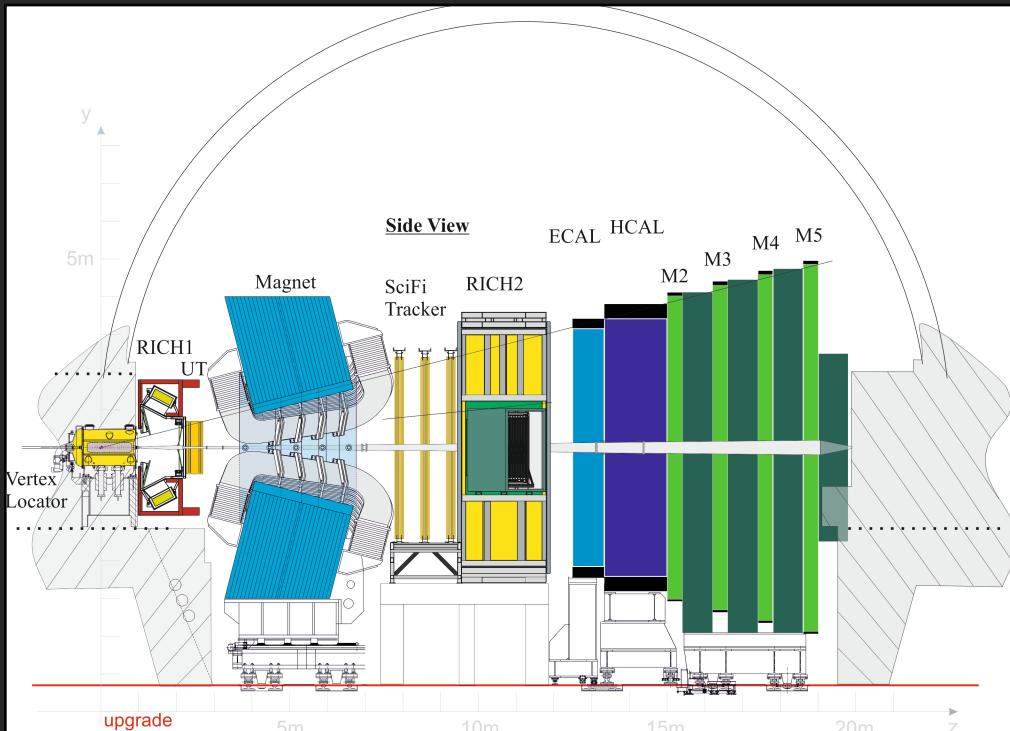
LHC

- Large Hadron Collider
- Experiments
 - CMS
 - Alice
 - Atlas
 - LHCb



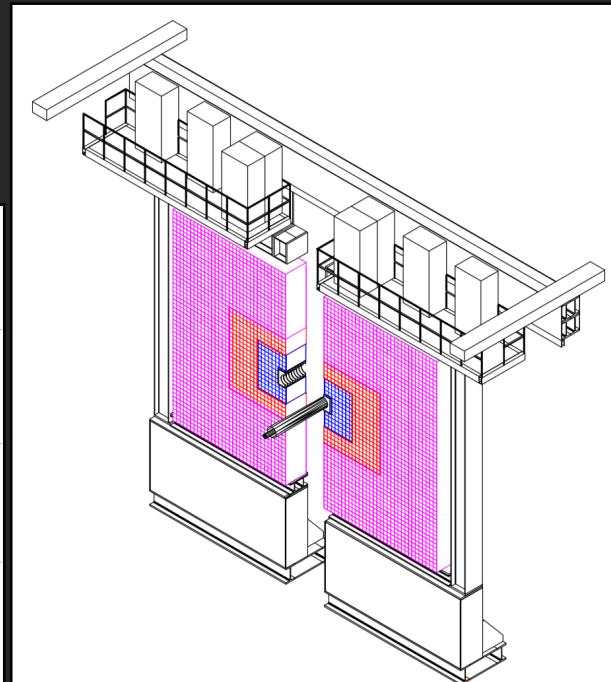
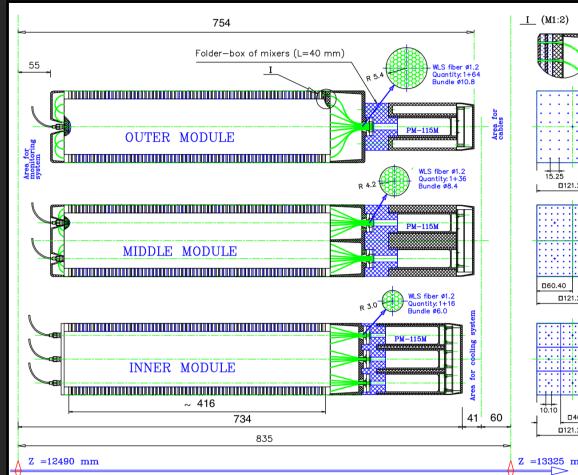
LHCb

- Sistema de reconstrucció de traces
- Sistema d'identificació de partícules



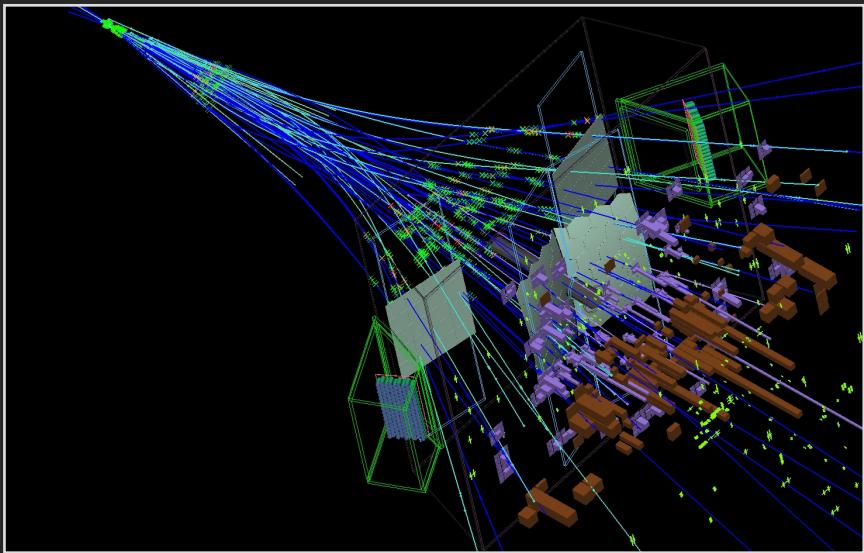
| Calorímetre électromagnétique

- 7.8 x 6.3 x 0.84 m.
- 6016 cel·les.
- Alterna plaques centellejadores i de plom (Shashlik).
- 7M€.



| Monte Carlo

- Simula el detector LHCb, partícules i com interactuen.
- Mètode per obtenir "la veritat".
- Ajuda als científics a entendre el comportament de les partícules observades.
- Permet mesurar l'eficiència el detector extraient informació d'una col·lisió.



PROBLEMA I DADES



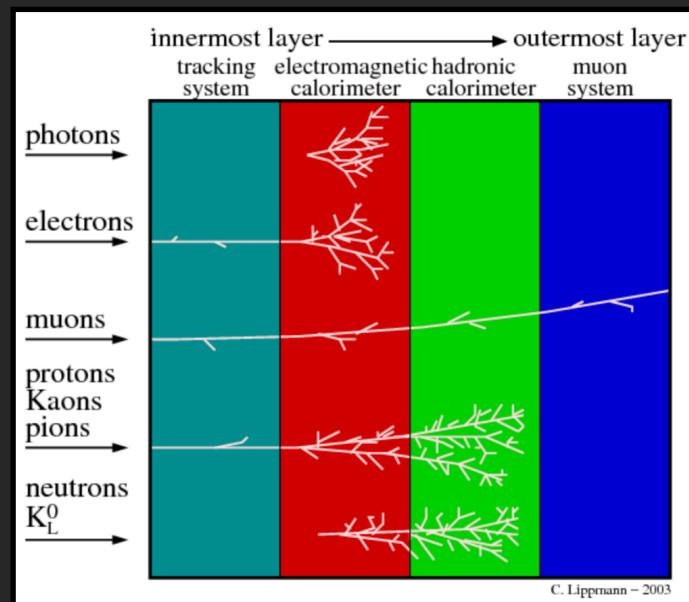
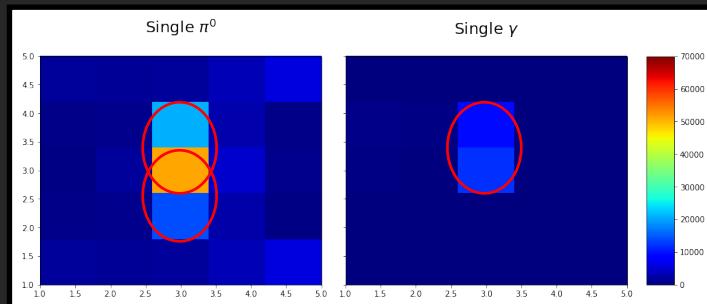
Data

Analysis

Idea

| Repte

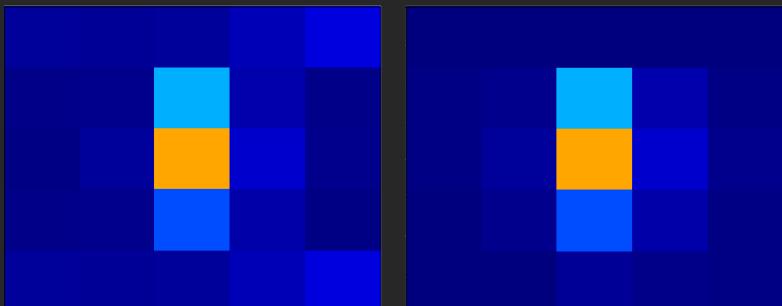
- Distingir entre fotons i pions neutres.
- Replicar els mètodes amb millors resultats existents.
- Estudiar possibles nous mètodes que superin els resultats existents.



| Les Dades (5x5)

- Processat

+ Processat



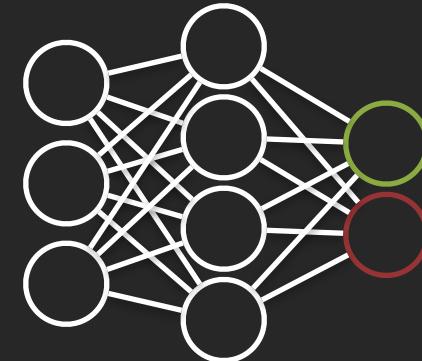
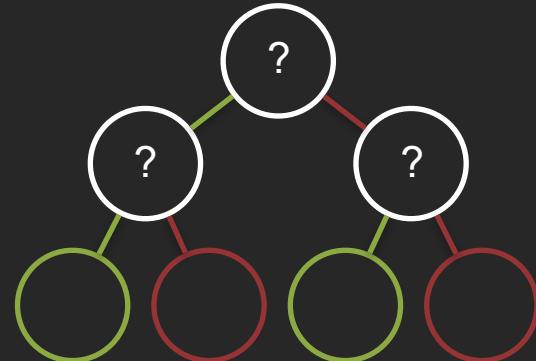
Sxx	830.161691
Sxy	-31.312723
Syy	597.362768

isPhr2	883.27343
isPhr2r4	0.660999
isPhasym	0.021731
isPhkappa	0.303136
isPhEseed	0.486411
isPhE2	0.782828

| Estudis previs

Atributs simples

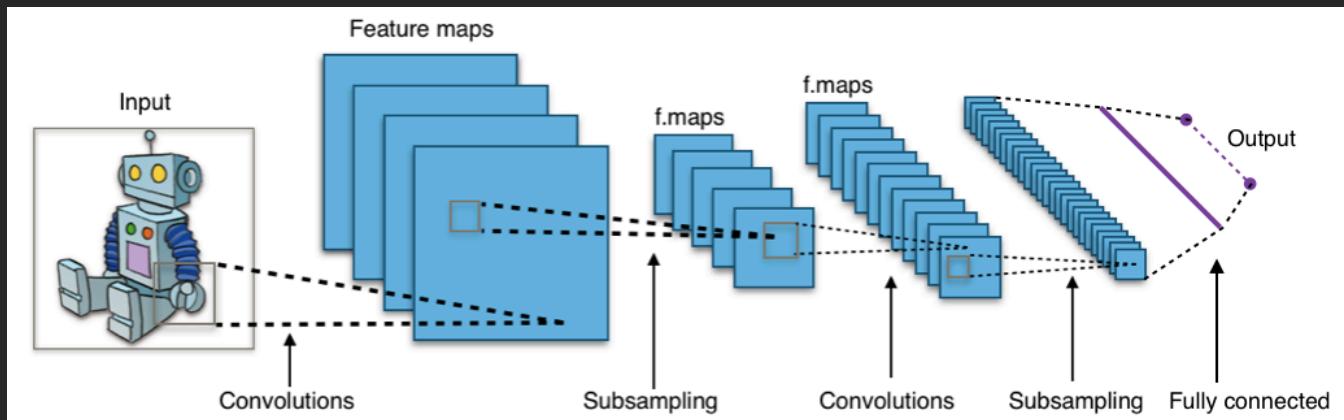
Atributs complexos



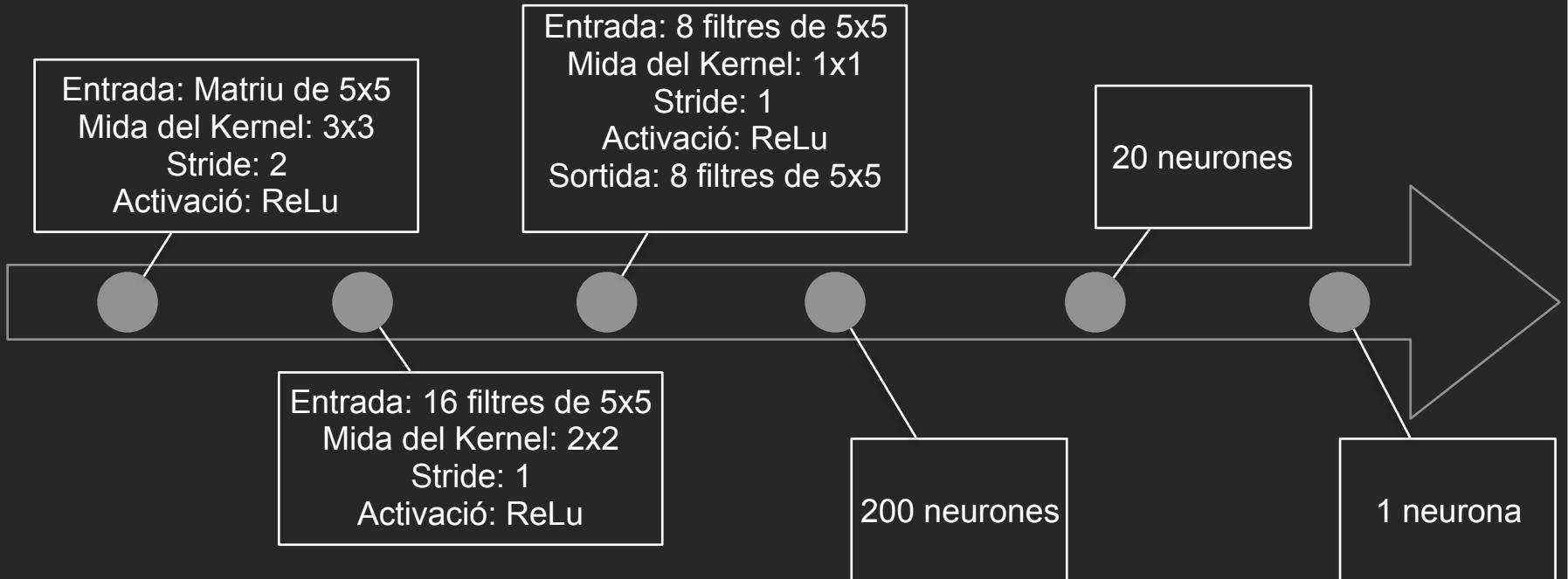
Algoritmes, consistència i mètriques

ICNN

- Xarxa neuronal convolucional.
- Inspirada en el còrtex visual dels animals.
- Molt utilitzades en visió artificial.
- Capa convolucional.
- Capa de reducció.
- Capa totalment connectada (ANN).



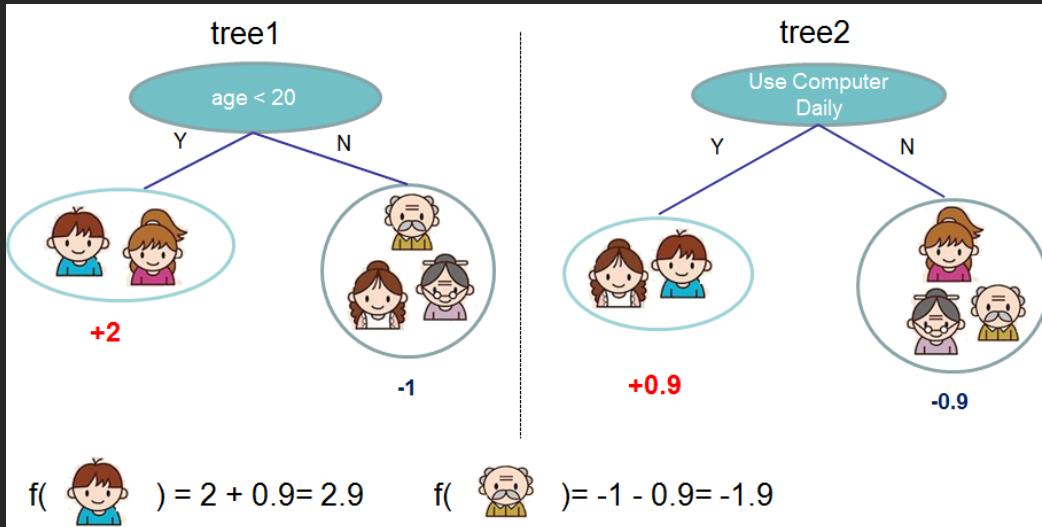
Arquitectura de la CNN



Llibreria Keras · Llenguatge Python

XGBoost

- Extreme Gradient Boosting
- Entrenament ràpid
- Aprendentatge supervisat
- Conjunt d'arbres de decisió
- CART
 - Classificació
 - Regressió
 - Binari
 - "Cost-complexity Prunning"
- Tree Boosting



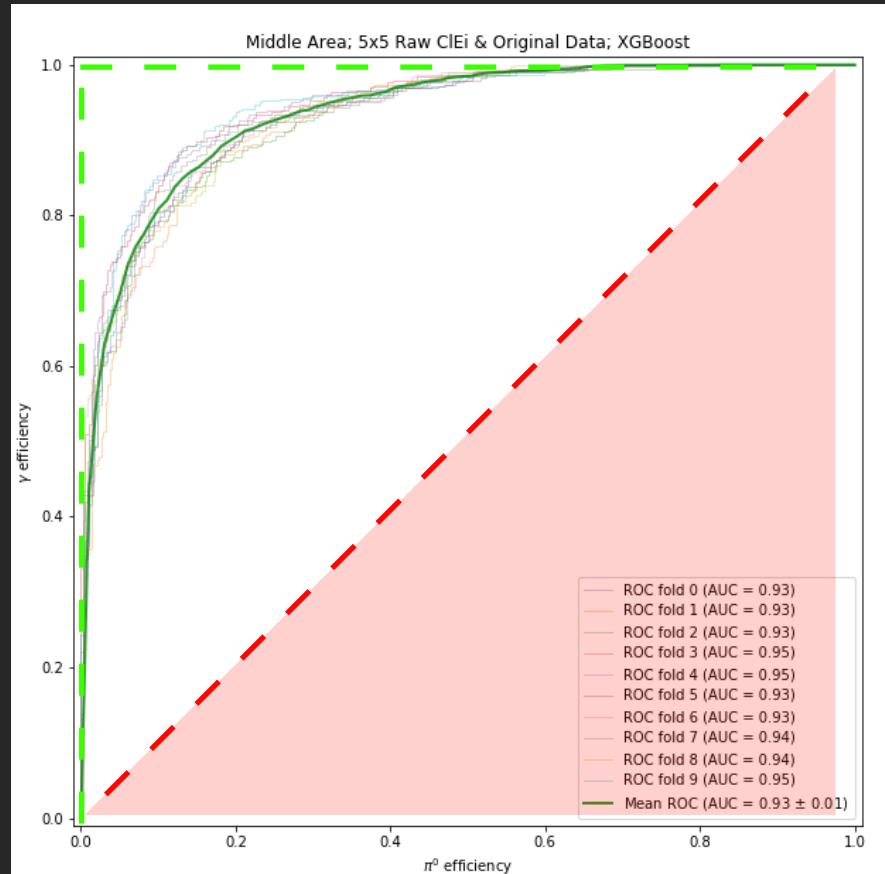
$$y'_i = \sum_{k=1}^K f_k(x_i), f_k \in \mathcal{F}$$

| Corba ROC i KFold

- Corbes ROC
 - AUC

		Realitat	
		1	0
Predicció	1'	TP	FP
	0'	FN	TN

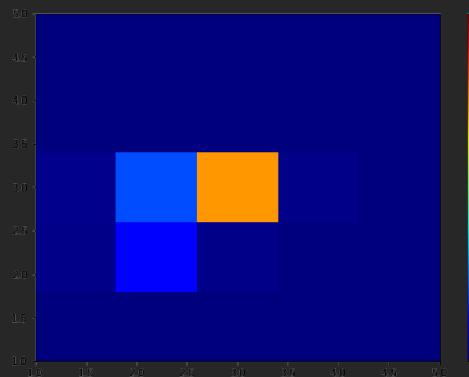
- Validació Creuada
 - KFold



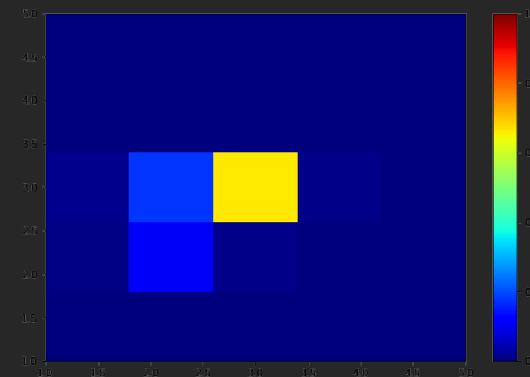
Resultats de la classificació

Processat de les dades

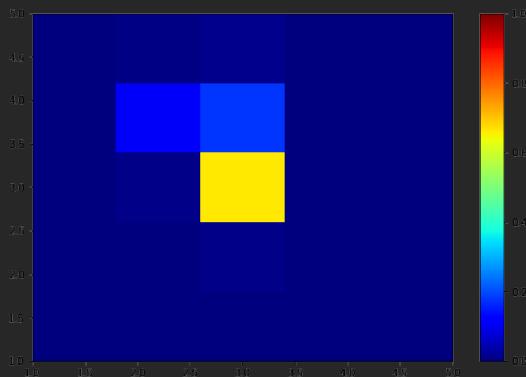
Dipòsit d'energia 5x5



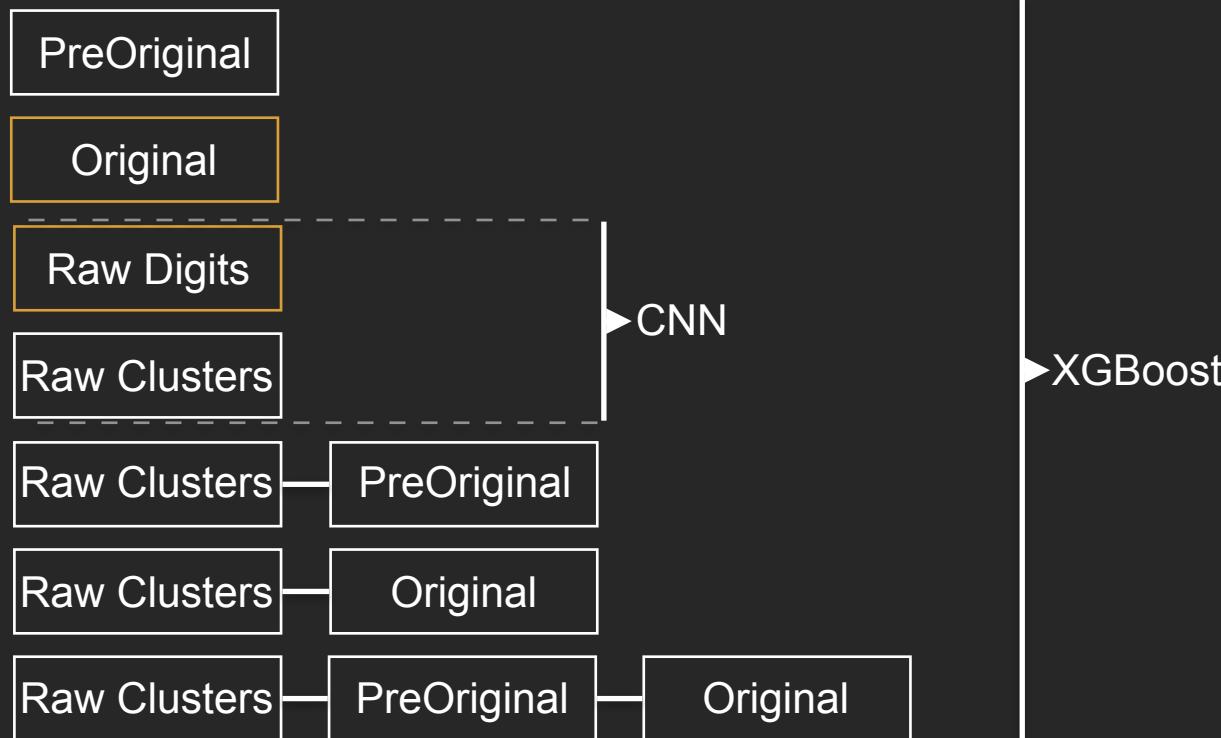
Dipòsit d'energia normalitzat



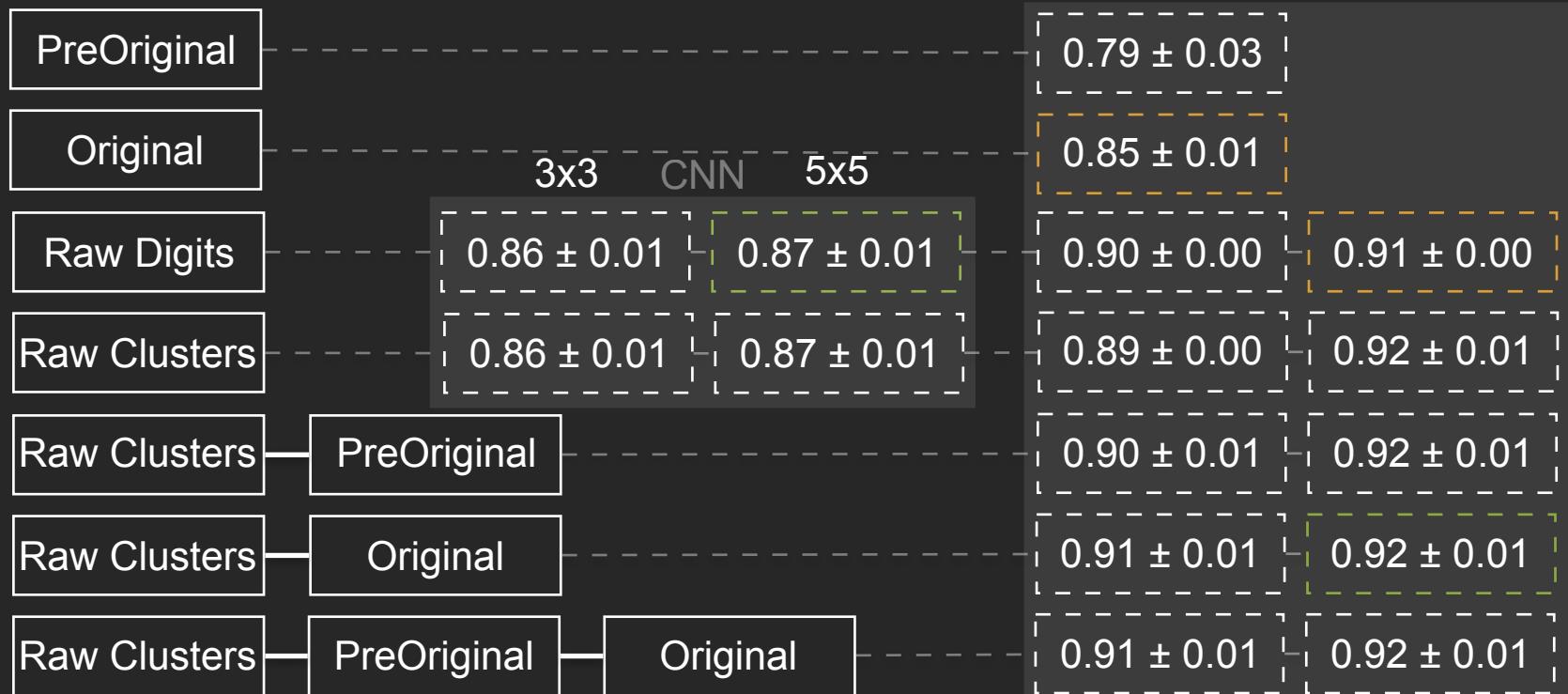
Dipòsit d'energia rotat



| Proves (3x3 i 5x5)



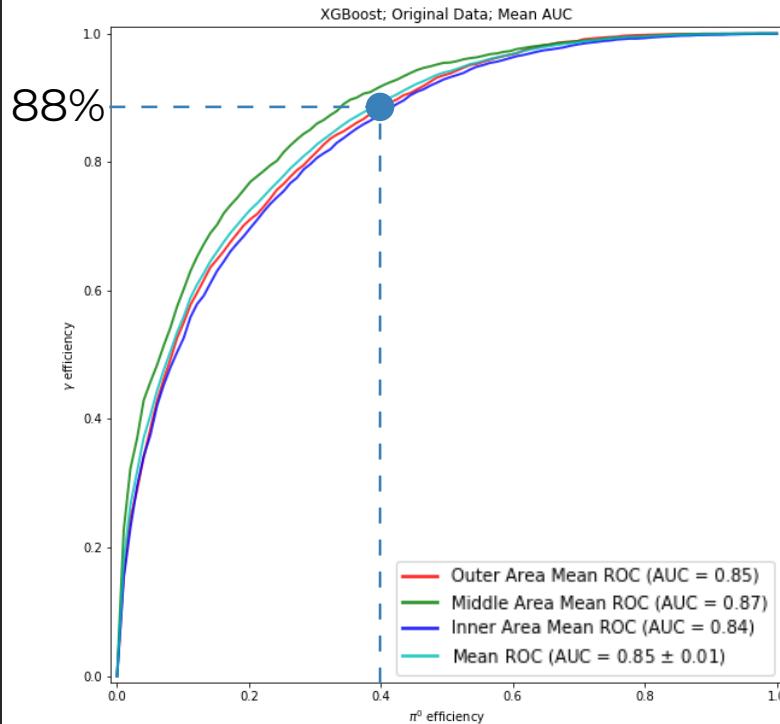
Mitjana dels resultats



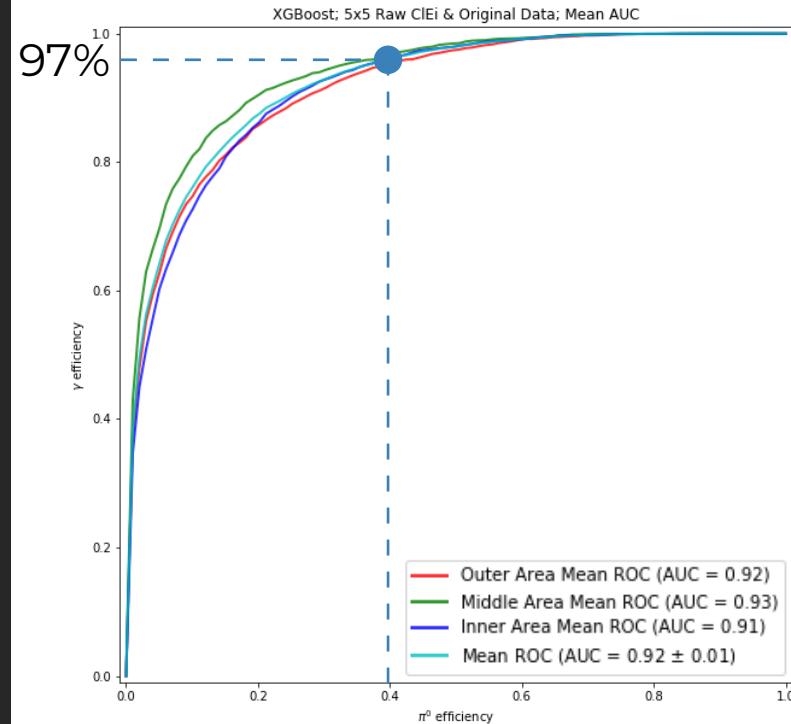
Comparació dels resultats

	Original	Raw Digits	Raw Digits	Raw Clusters
Area del Calorímetre	XGBoost	XGBoost	CNN	XGBoost
Exterior	0.85 ± 0.00	0.91 ± 0.01	0.85 ± 0.01	0.92 ± 0.01
Mig	0.87 ± 0.00	0.92 ± 0.01	0.88 ± 0.01	0.93 ± 0.01
Interior	0.84 ± 0.01	0.91 ± 0.00	0.88 ± 0.01	0.91 ± 0.00
Mitjana	0.85 ± 0.01	0.91 ± 0.00	0.87 ± 0.01	0.92 ± 0.01

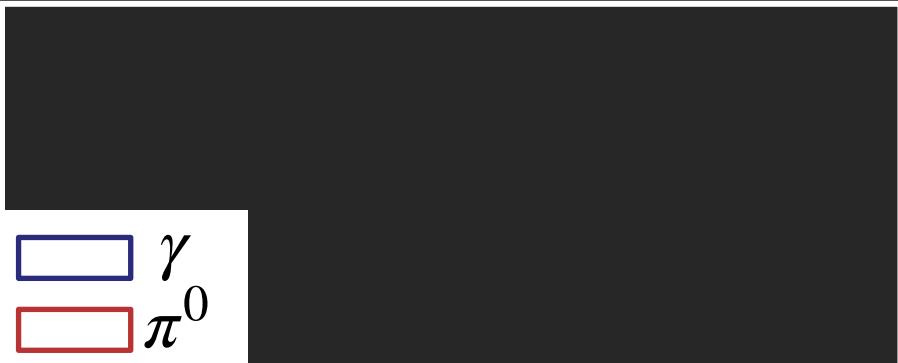
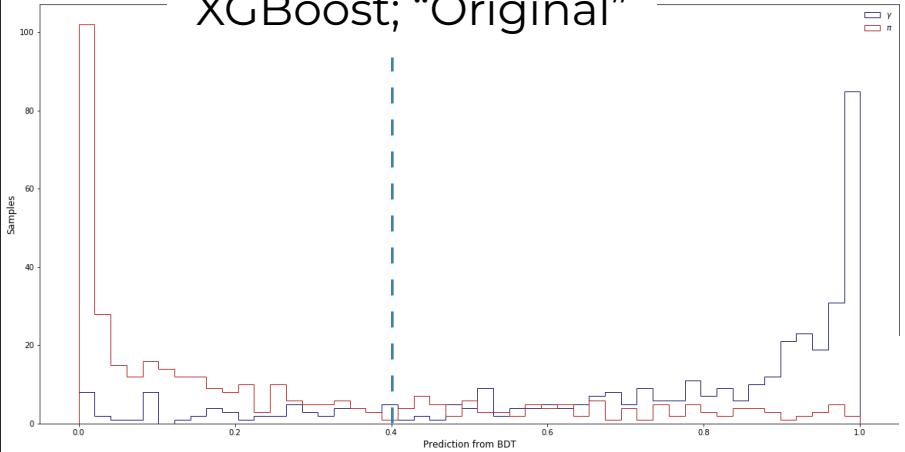
XGBoost; "Original"



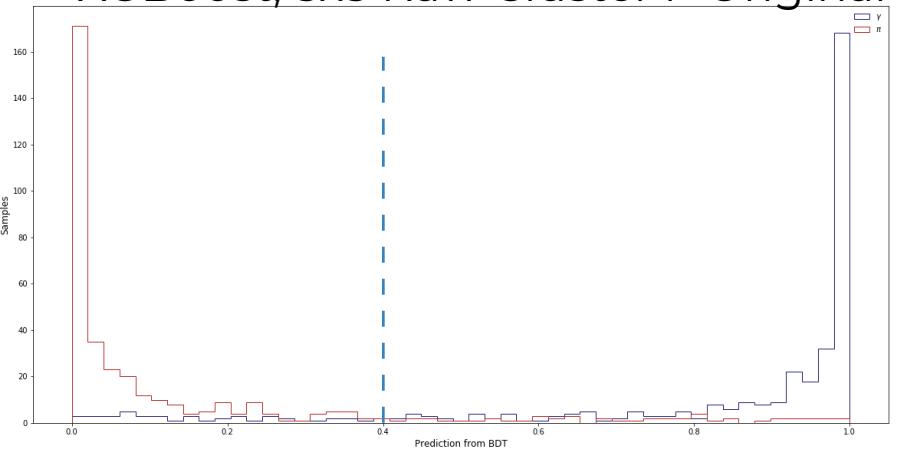
XGBoost; 5x5 Raw Cluster i "Original"



XGBoost; “Original”



XGBoost; 5x5 Raw Cluster i “Original”



| Conclusions

- L'àrea del mig obté els millors resultats de classificació.
- En general l'eficiència de classificació entre l'àrea exterior i interior és similar.
- En l'àrea interior es produeix una superposició major en comparació amb la del mig.
- En l'àrea exterior la resolució de les cel·les del calorímetre és menor.
- Hi ha una balança clara entre resolució i superposició.

| Conclusions i línies de futur

- Un augment de resolució en totes les àrees milloraria els resultats de classificació.
- En els clústers de 5x5 s'obté millor resultats que en els de 3x3.
- Es confirma els resultats dels investigadors de Yandex en el detector millorat.
- La classificació millora si s'utilitzen conjuntament les dades "Raw Data" i "Original".
- Es millora significantment l'eficiència de classificació versus el mètode "Original".



Gràcies per la vostra atenció!
Torn de preguntes