



XVIII Conference on Resistive Plate Chambers and Related Detectors

14–18 September 2026



Machine Learning-Based Suppression of Secondary Hit Background in Low-Resistive RPCs

Souvik Chattopadhyay, Zubayer Ahammed

speaker: Souvik Chattopadhyay

Senior Research Fellow

Variable Energy Cyclotron Centre, DAE, India

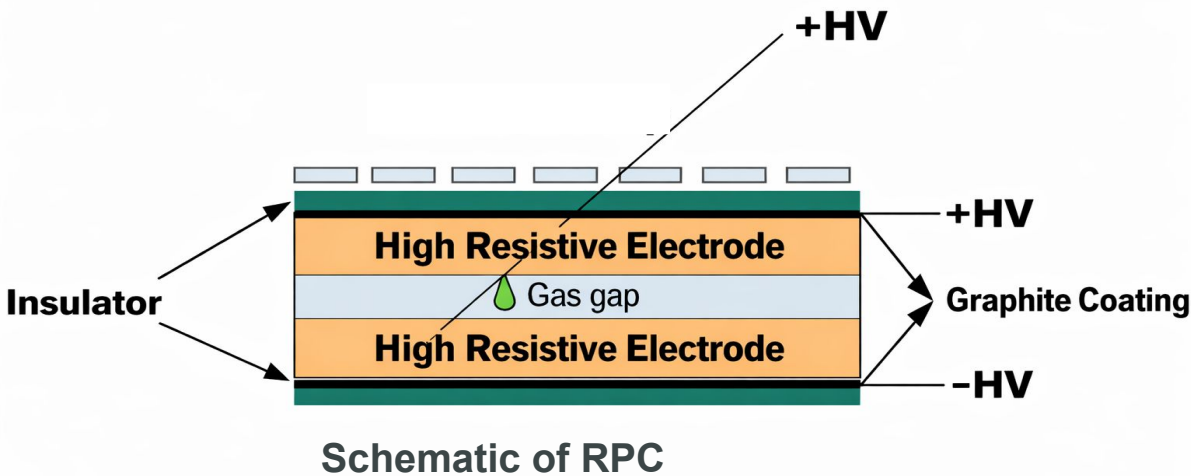
Introduction

- Resistive Plate Chambers (RPCs) provide efficient particle detection with fast timing and economical large-area coverage.
- High particle flux causes a voltage drop across the resistive electrodes, reducing the effective gas-gap field and detection efficiency.
- The voltage drop across the electrodes under particle flux (Φ) is given by:

$$V_d = \rho t \langle Q \rangle \Phi,$$

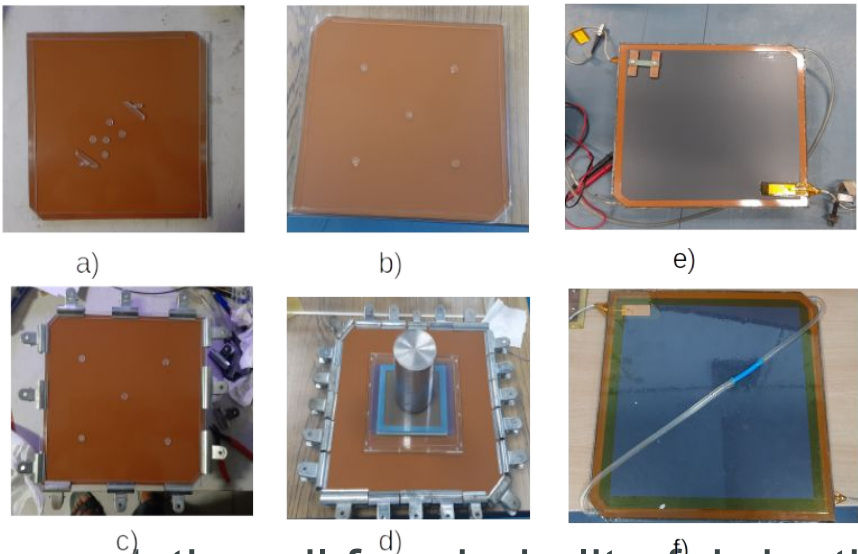
- Low-resistivity bakelite electrodes reduce this voltage drop and shorten local recovery time, improving rate capability.
- In the low-resistivity RPC, we observed number of **secondary hits increases with applied voltage**.
- Secondary hits in these RPCs can broaden timing and spatial hit distributions[1,2].
- **This work:** use cluster-level features and supervised machine learning to distinguish signal-like clusters from background in free-streaming RPC data.

Detector Components and Fabrication



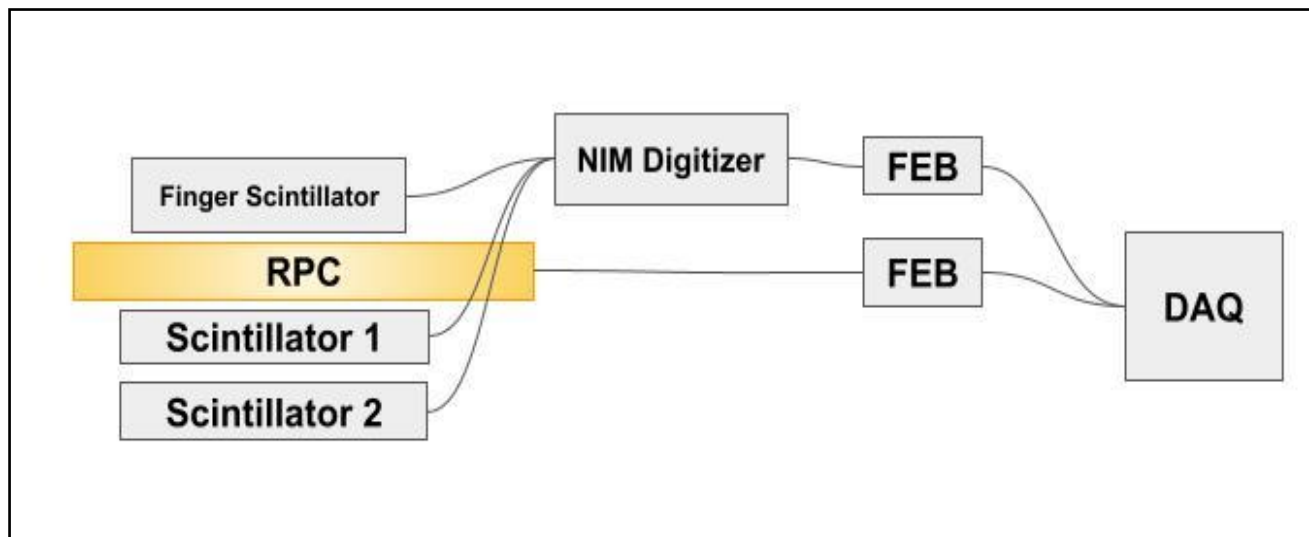
Detector specifications

parameter	Value
Gas Gap	Single
dimension	30 cm* 30 cm
Electrode thickness	1.2 mm
Gas Gap thickness	2mm
Bulk resistivity	$(3 \cdot 10^9 - 1 \cdot 10^{10}) \Omega \text{cm}$
Surface resistivity	10-100 k Ω



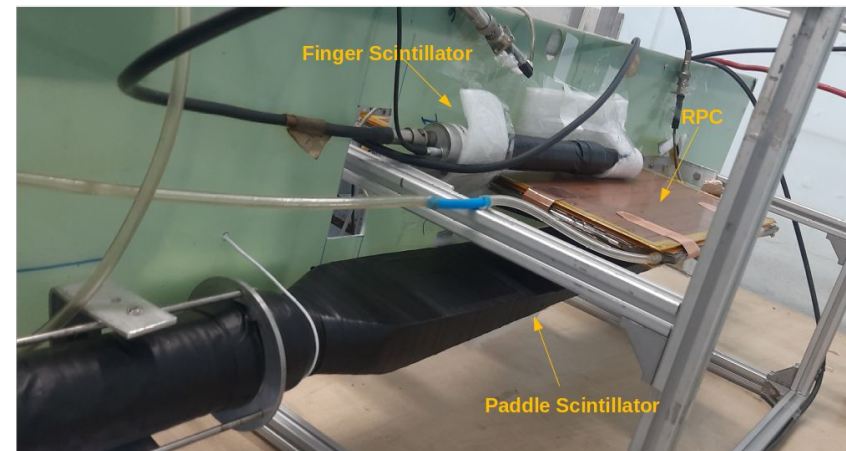
Low resistive oil free bakelite fabrication steps

Experimental Setup:

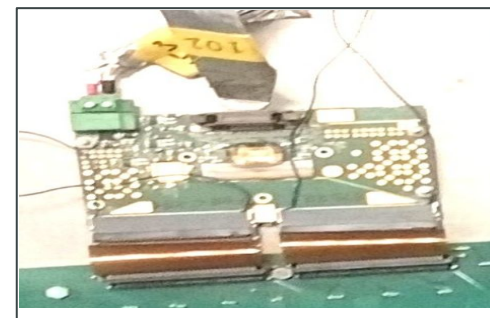


Schematic of the laboratory experimental setup

- FEB Specifications: MuCH-XYTER
- Preamplifier: Low-noise Charge Sensitive Amplifier (CSA)
- Channels: 128 channels per ASIC
- Signal processing: Two parallel paths per channel
 - Timing path: fast shaper ($\tau = 30$ ns) + discriminator $\rightarrow$ 14-bit timestamp
 - Amplitude path: slow shaper ($\tau = 280$ ns) + 5-bit flash ADC



testing set-up in Lab



Time resolution: 3.125 ns

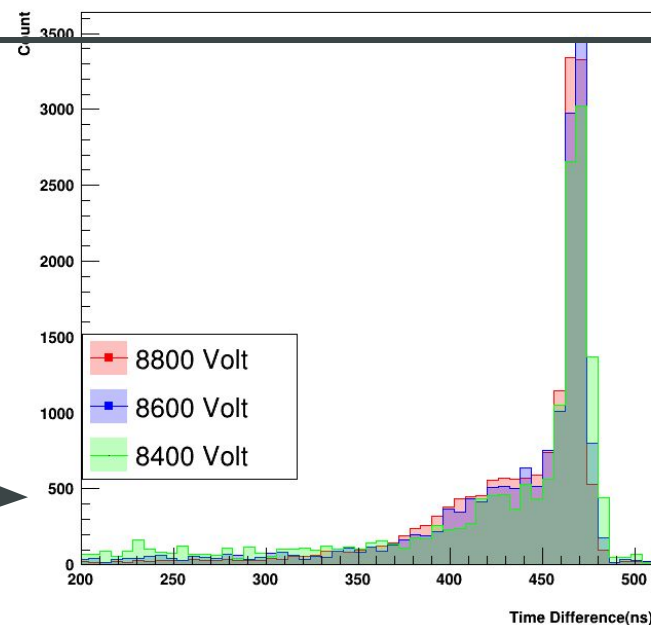
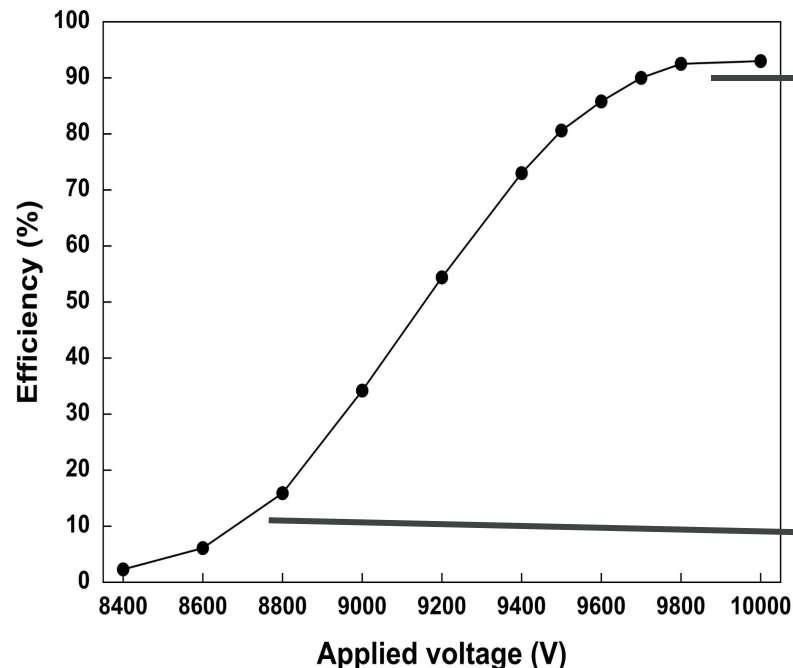
ADC resolution: 5 fC (15 fC- 150fC)

2d readout pad with pad size=2cm* 2 cm

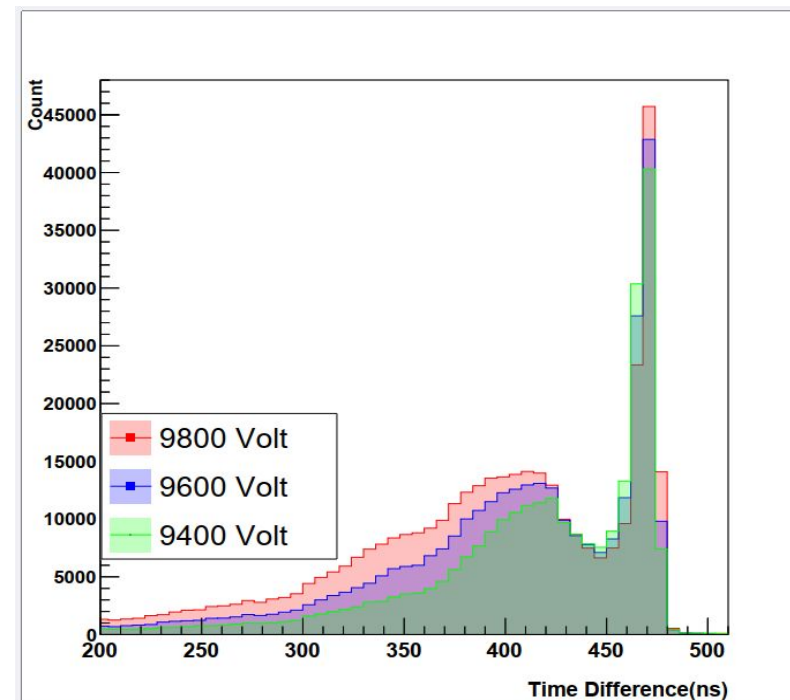
Time Correlation with Master Trigger:

$$\text{Time-difference} = \text{Time}_{\text{Trigger}} - \text{Time}_{\text{RPC}}$$

Efficiency Scan



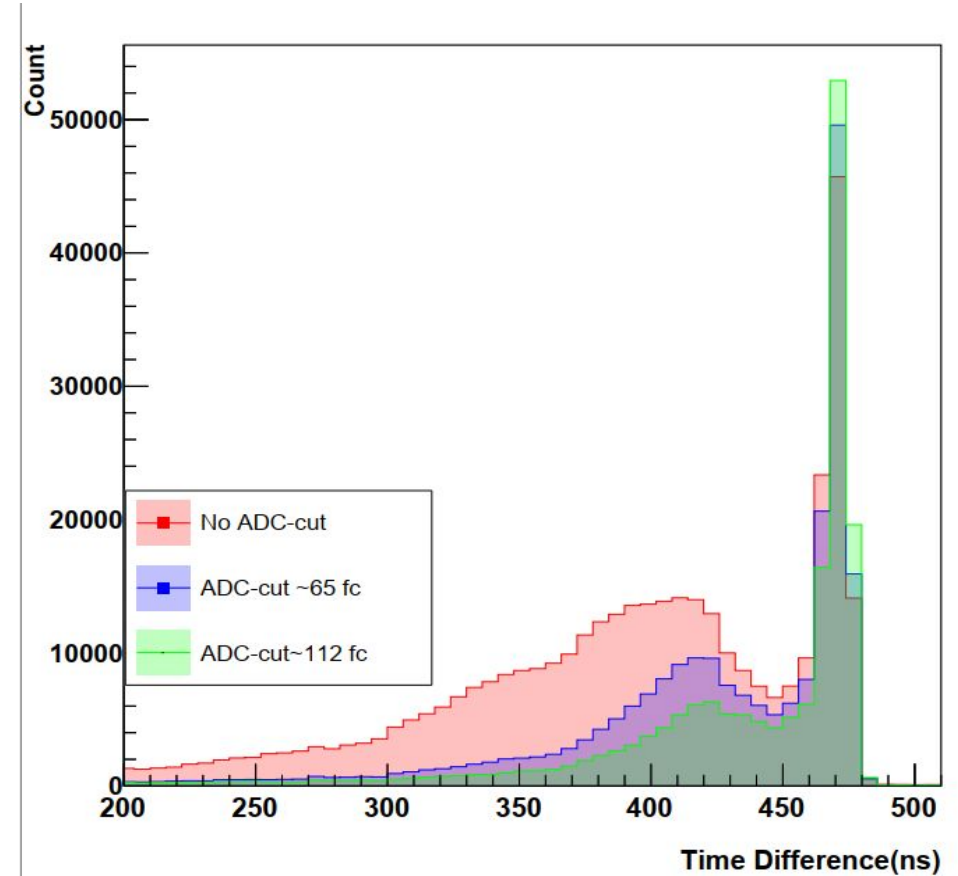
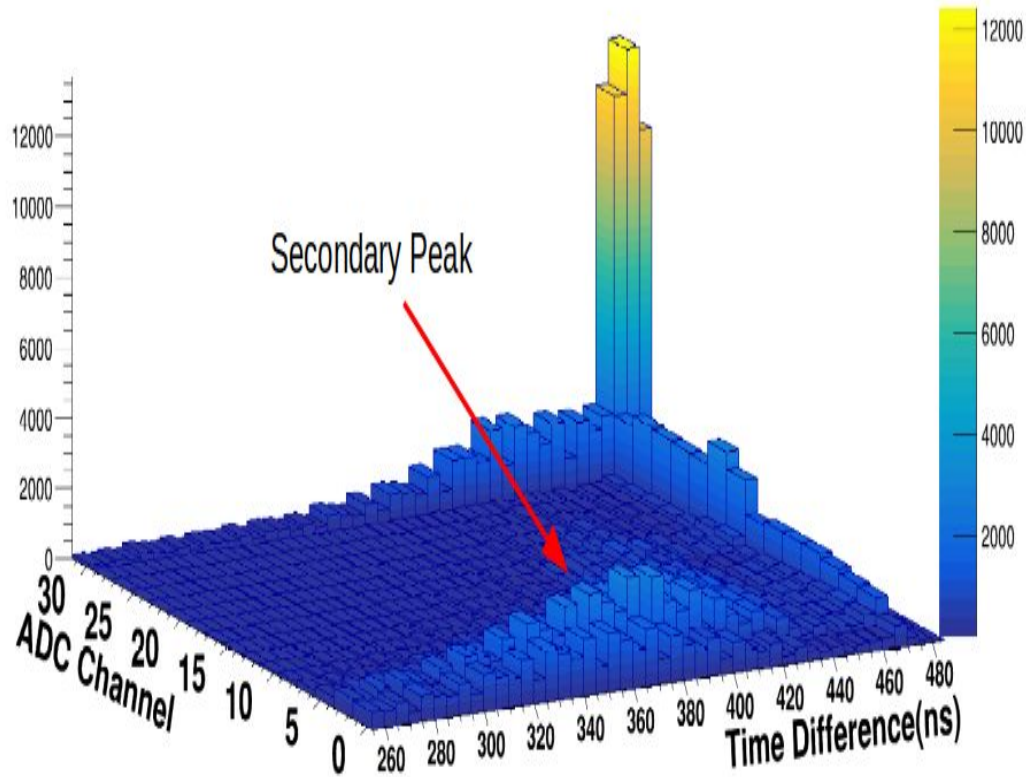
Time-Correlation at different lower Voltages



Time-Correlation at different High Voltages

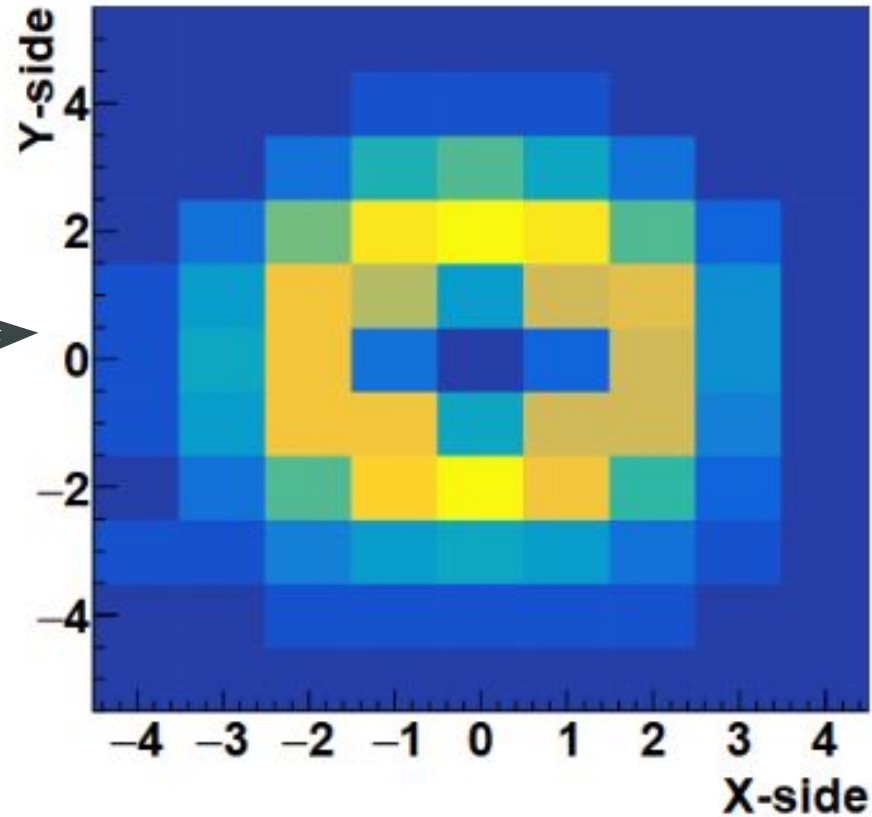
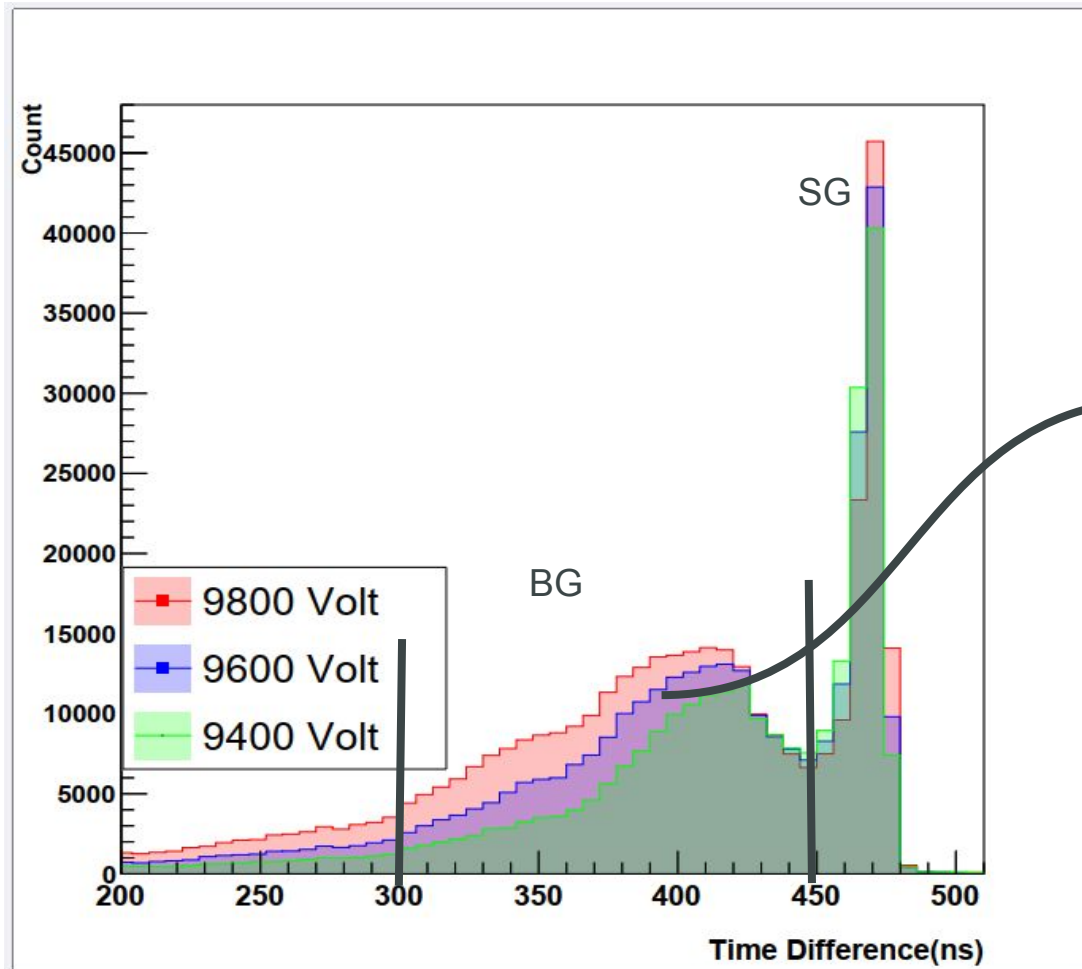
Gas Mixture: R134a, iC4H10, SF6:: 95.0% : 4.5% : 0.5%

ADC-Distribution and Effect of ADC-cut



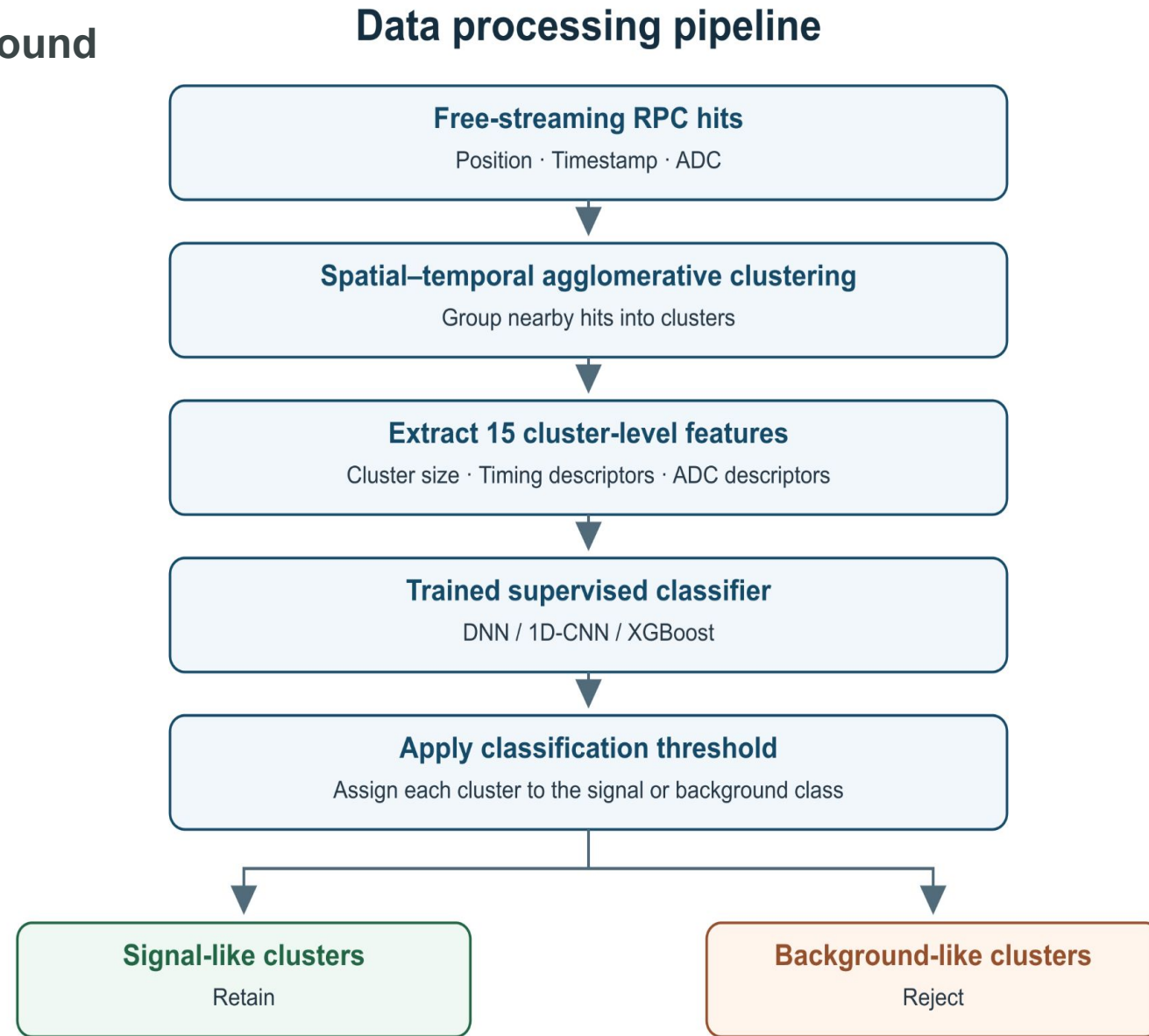
1. At higher ADC-cut also secondary peak still exist.
2. Secondary peak do not contribute to the detector's efficiency
3. But it reduce the spatial and temporal resolution.

Secondary peak and its pad distribution:

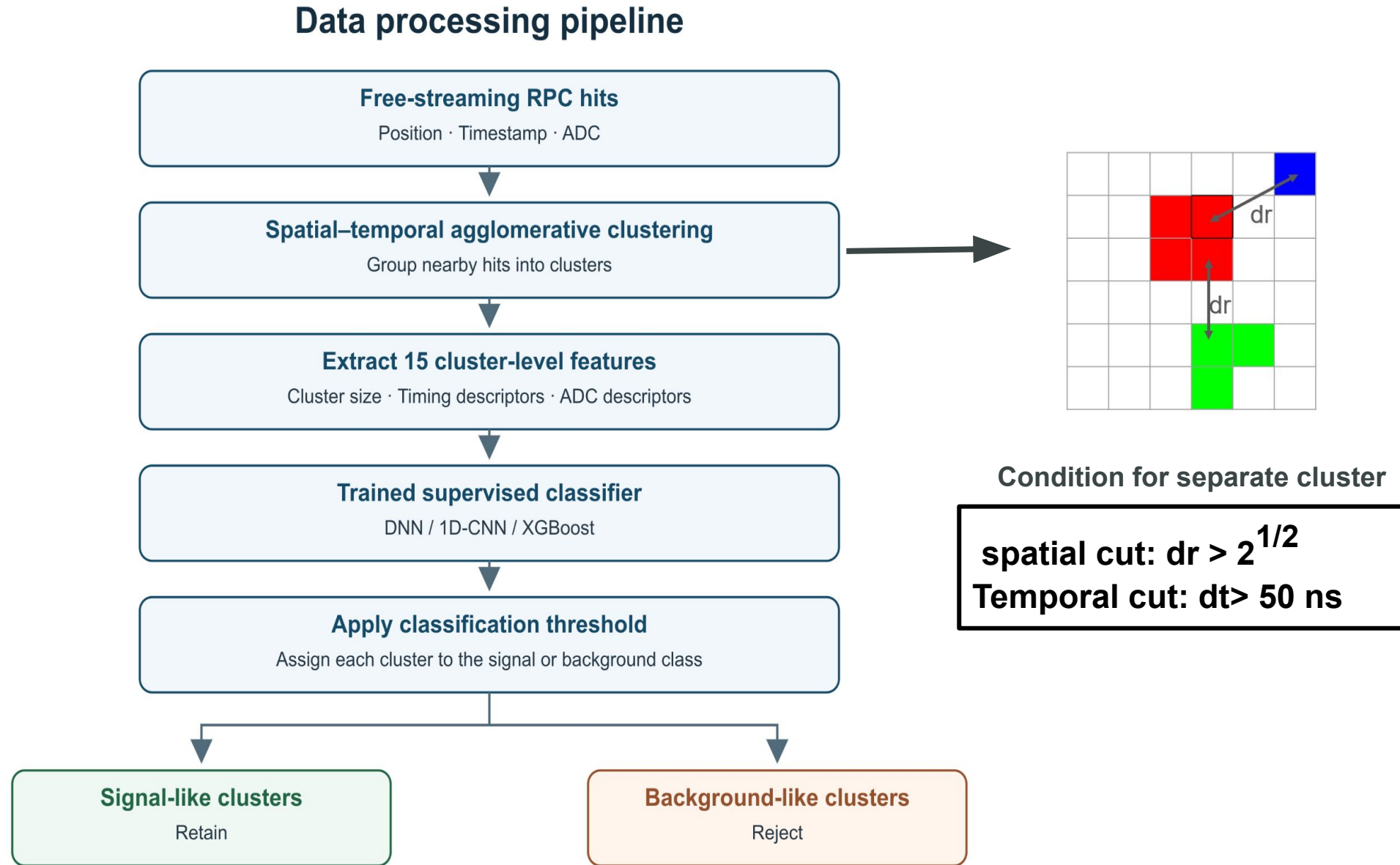


1. Bad Time resolution
2. Bad spatial resolution

Pipeline for signal and background classification:

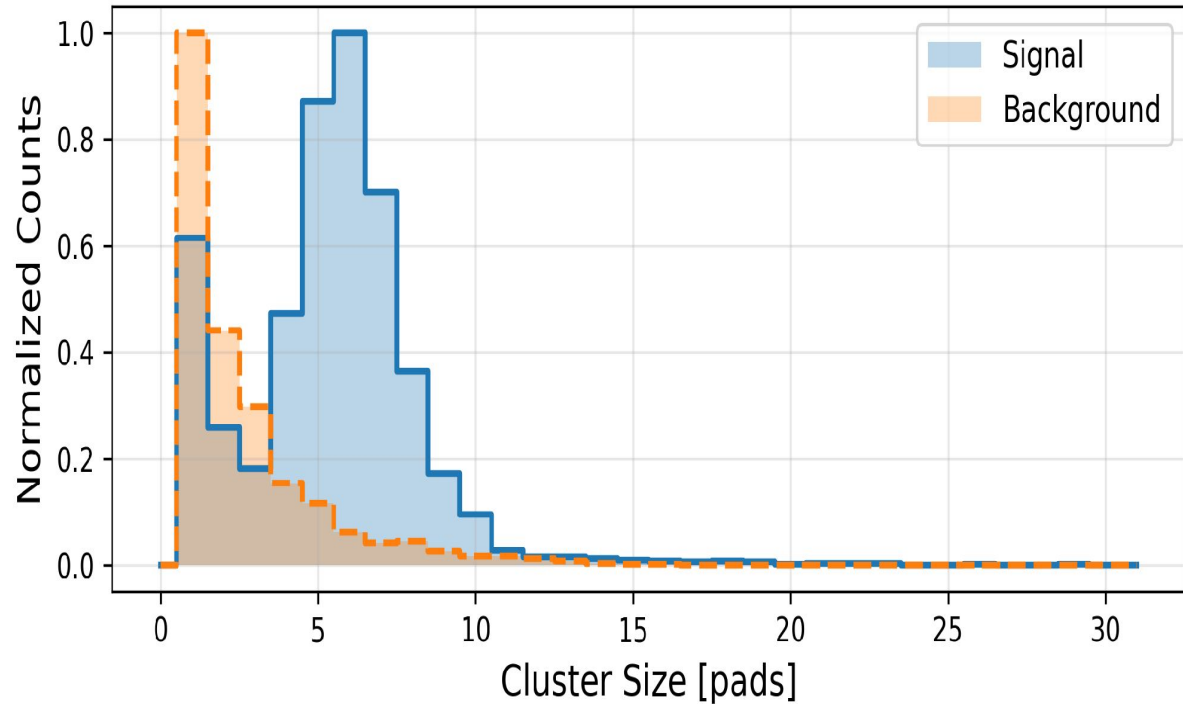


Pipeline:



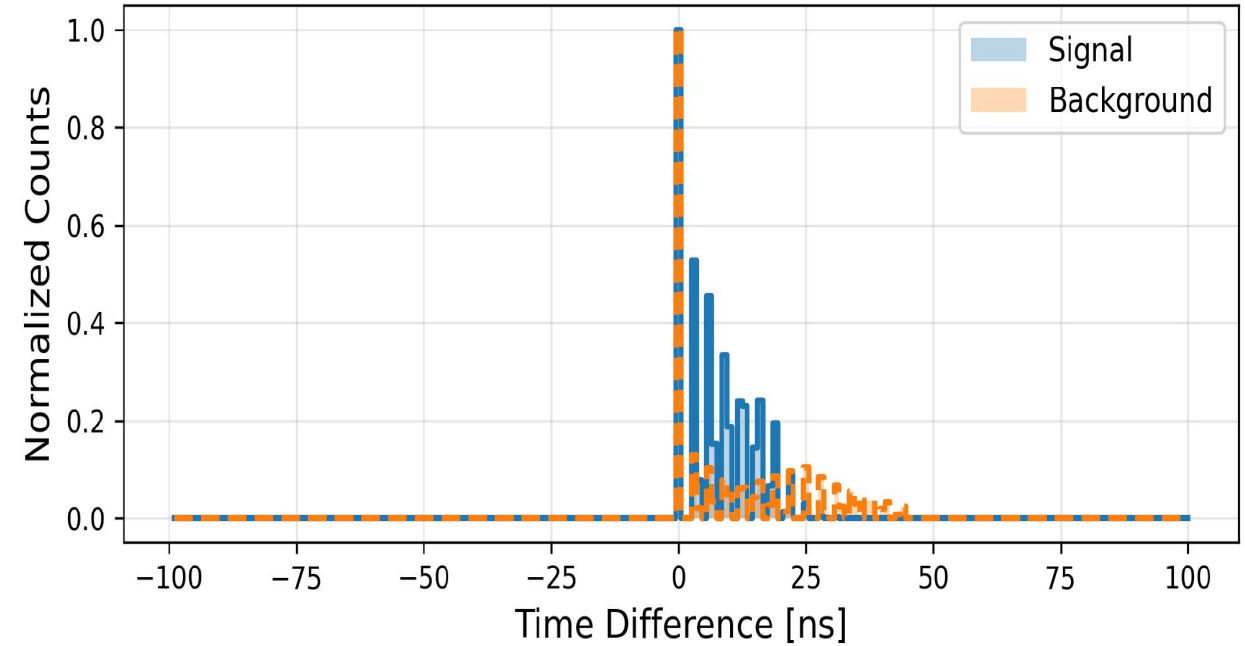
Signal vs. Background: Cluster and Intra-Cluster Time Distributions

(a) Cluster Size Distribution



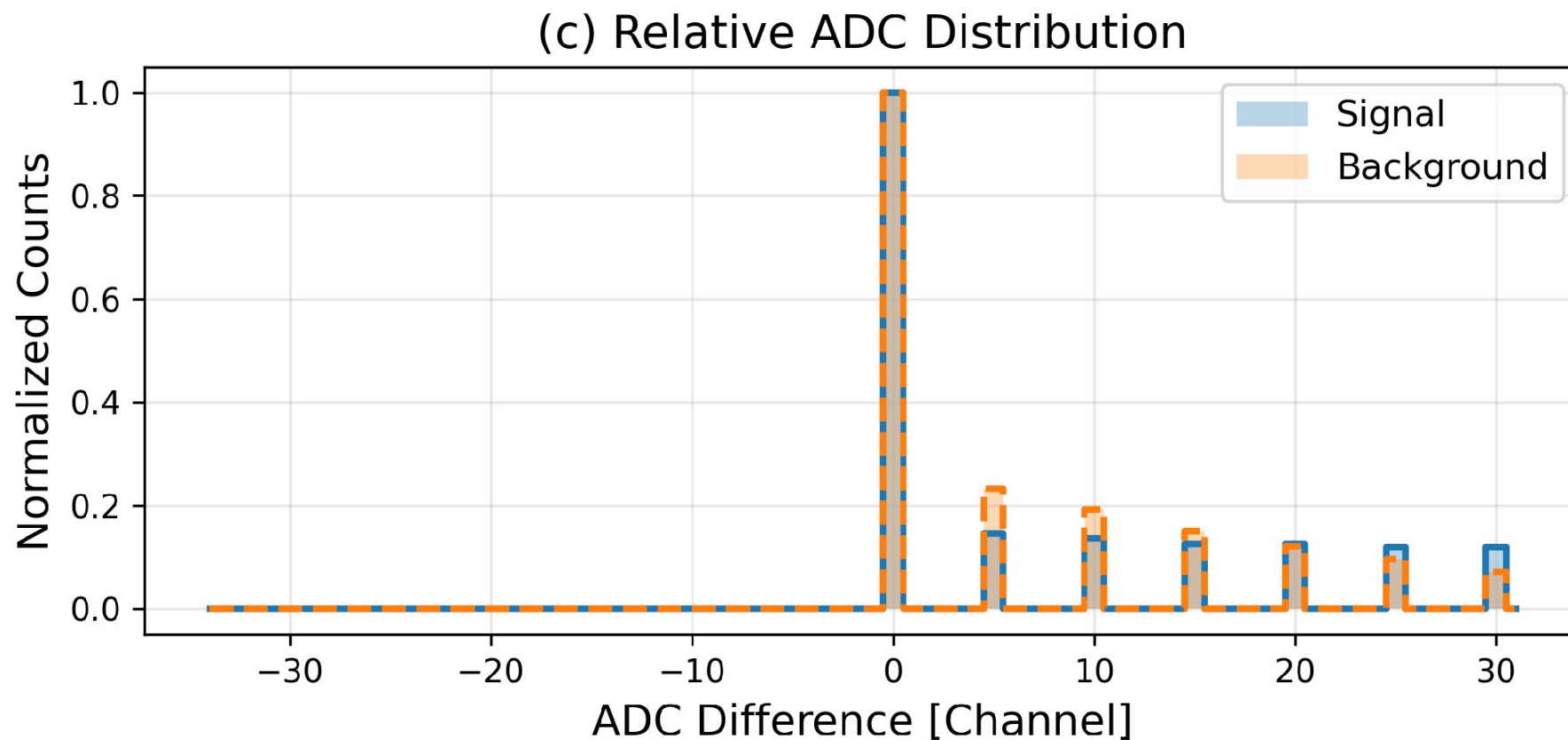
Signal: generally larger clusters

(b) Relative Time Distribution



Signal: narrower time spread

Intra-cluster ADC distribution



ΔADC is measured relative to the minimum ADC in each cluster

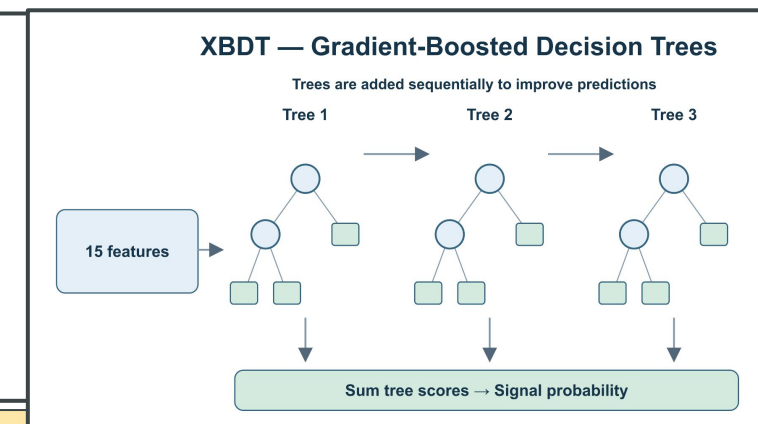
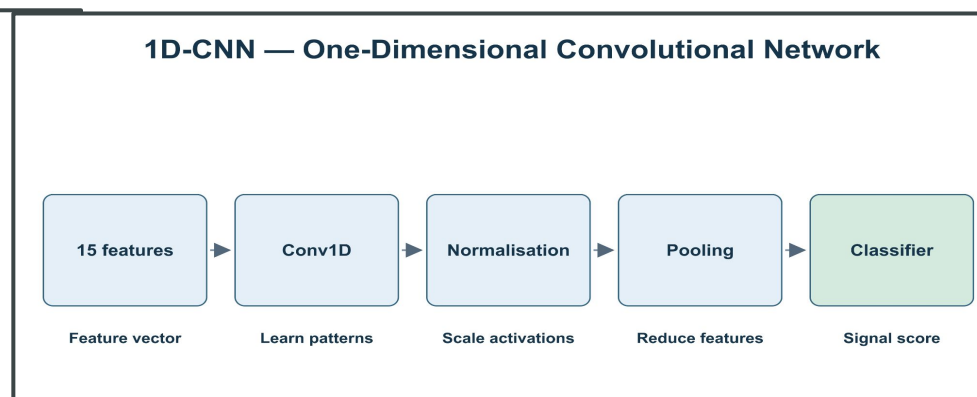
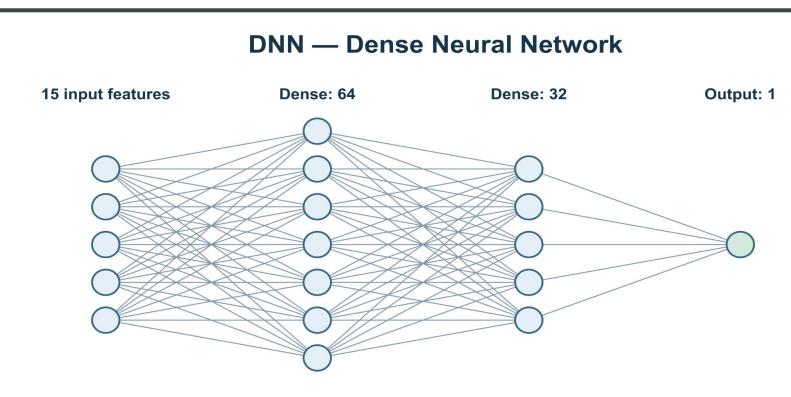
cluster-level input features:

Feature group	Descriptors	No. of Features
Cluster size	Number of fired readout elements	1
Time distribution	Histogram mean and σ ; Gaussian mean, σ , amplitude; fit χ^2 and NDF	7
ADC distribution	Histogram mean and σ ; Gaussian mean, σ , amplitude; fit χ^2 and NDF	7

The same feature set is used for DNN, 1D-CNN and XGBoost

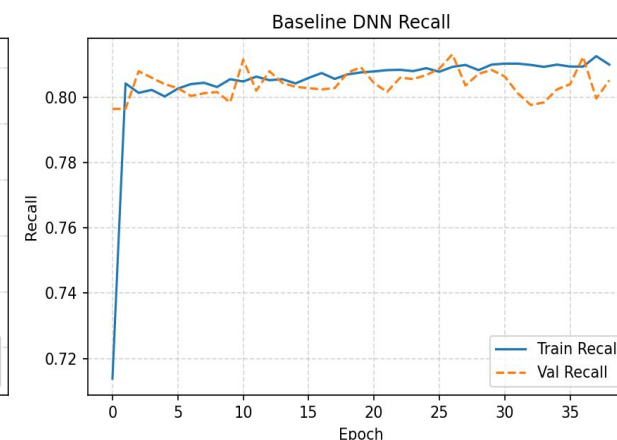
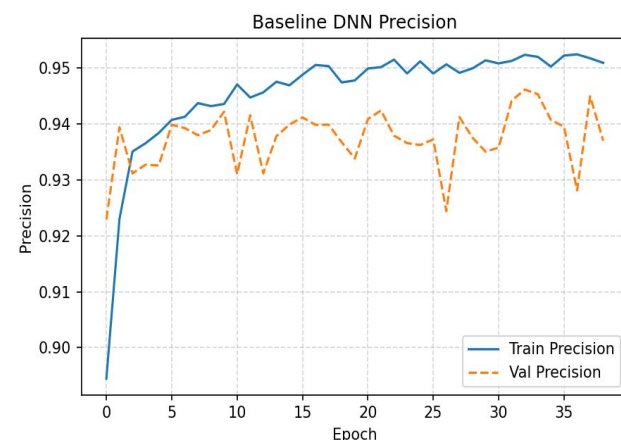
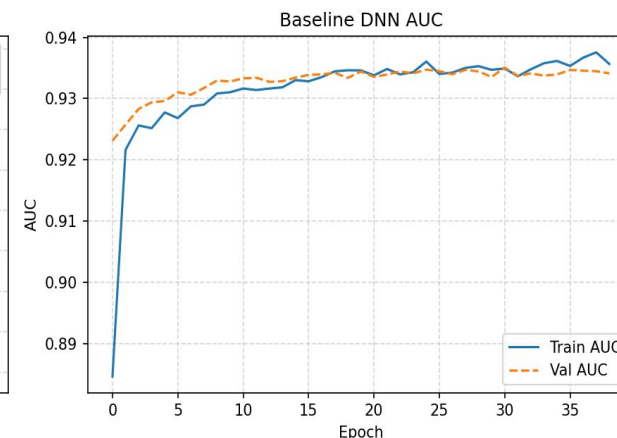
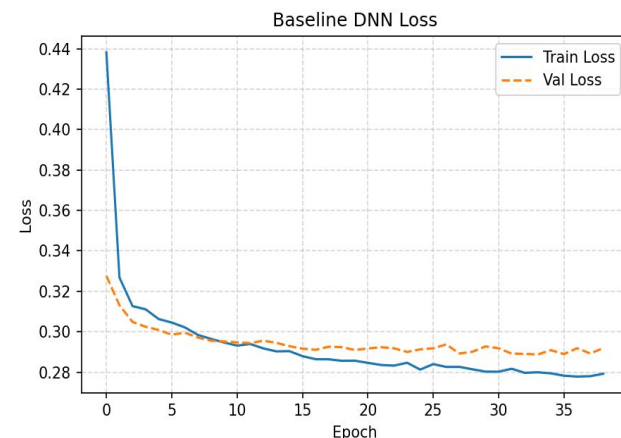
Three classifiers, the same feature set

Model	Configuration	Training / tuning
DNN	Dense 64 → 32 → 1(ReLU,Sigmoid) 3,137 parameters	Adam, binary cross-entropy Early stopping on validation AUC
1D-CNN	Conv1D, pooling(ReLU,Sigmoid) 17,857 total parameters	Early stopping on validation AUC
XGBoost	200 trees, maximum depth 5	Learning rate 0.05 Row subsampling 0.7



DNN classifier and Training History

		Predicted class	
		Signal	Background
True class	Signal	40.67% True positive (TP)	9.48% False negative (FN)
	Background	2.32% False positive (FP)	47.53% True negative (TN)



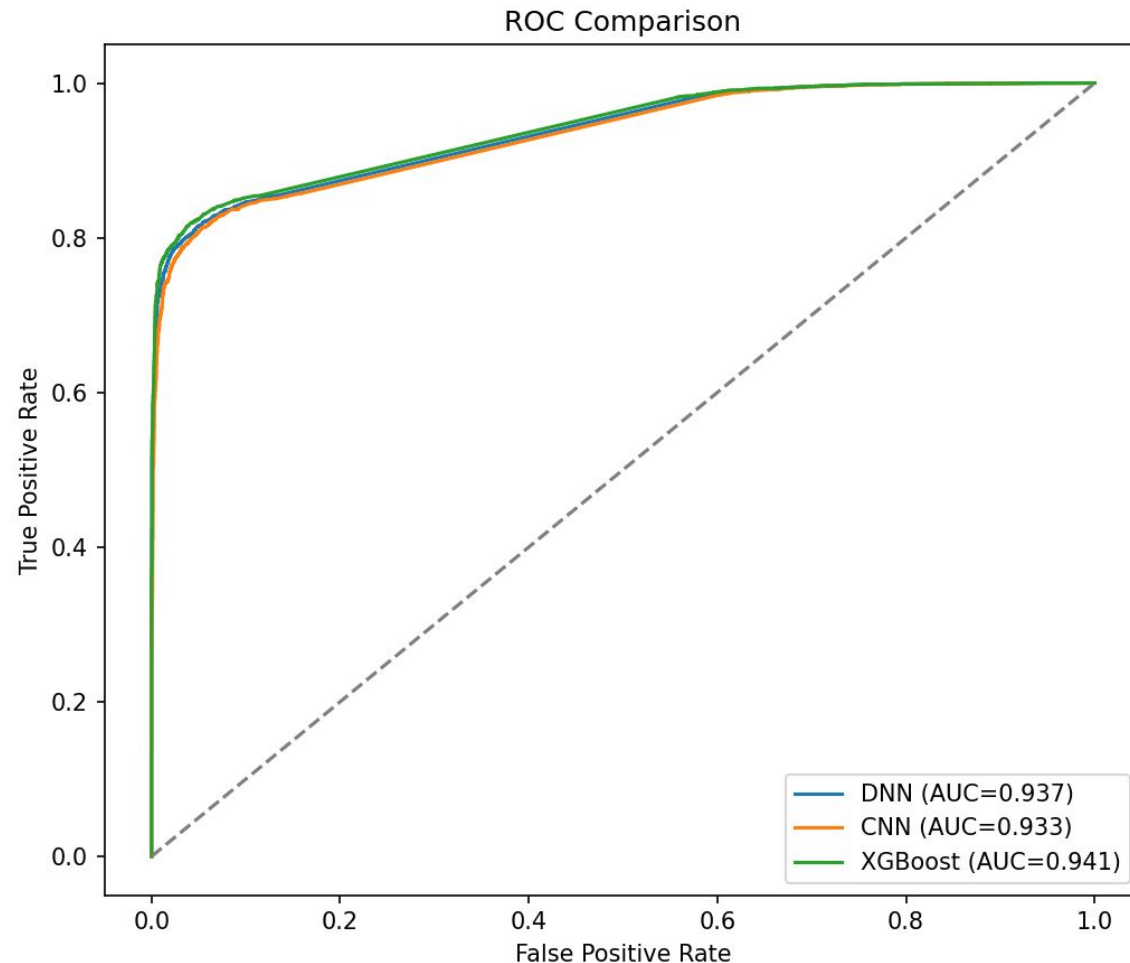
Precision = $TP / (TP + FP)$

Recall = $TP / (TP + FN)$

Loss: Measures the error in the model's predictions

AUC: Measures how well the model separates signal from background across classification

ALL THREE MODELS ACHIEVE AUC ABOVE 0.93



Model	Bootstrap AUC
-------	---------------

DNN	0.9367 ± 0.0021
-----	---------------------

1D-CNN	0.9333 ± 0.0022
--------	---------------------

XGBoost	0.9405 ± 0.0021
---------	---------------------

XGBoost is marginally highest; the models have comparable performance within the reported uncertainties.

$$\text{FPR} = \text{FP} / (\text{FP} + \text{TN})$$

$$\text{TPR} = \text{TP} / (\text{TP} + \text{FN})$$

PERFORMANCE AT THE REPORTED OPERATING POINT

Model	Accuracy	Precision	Recall	F1
DNN	0.882	0.946	0.811	0.873
1D-CNN	0.879	0.942	0.807	0.870
XGBoost	0.888	0.954	0.816	0.880

$$F_1 = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN}$$

F1 score is particularly useful for imbalanced datasets

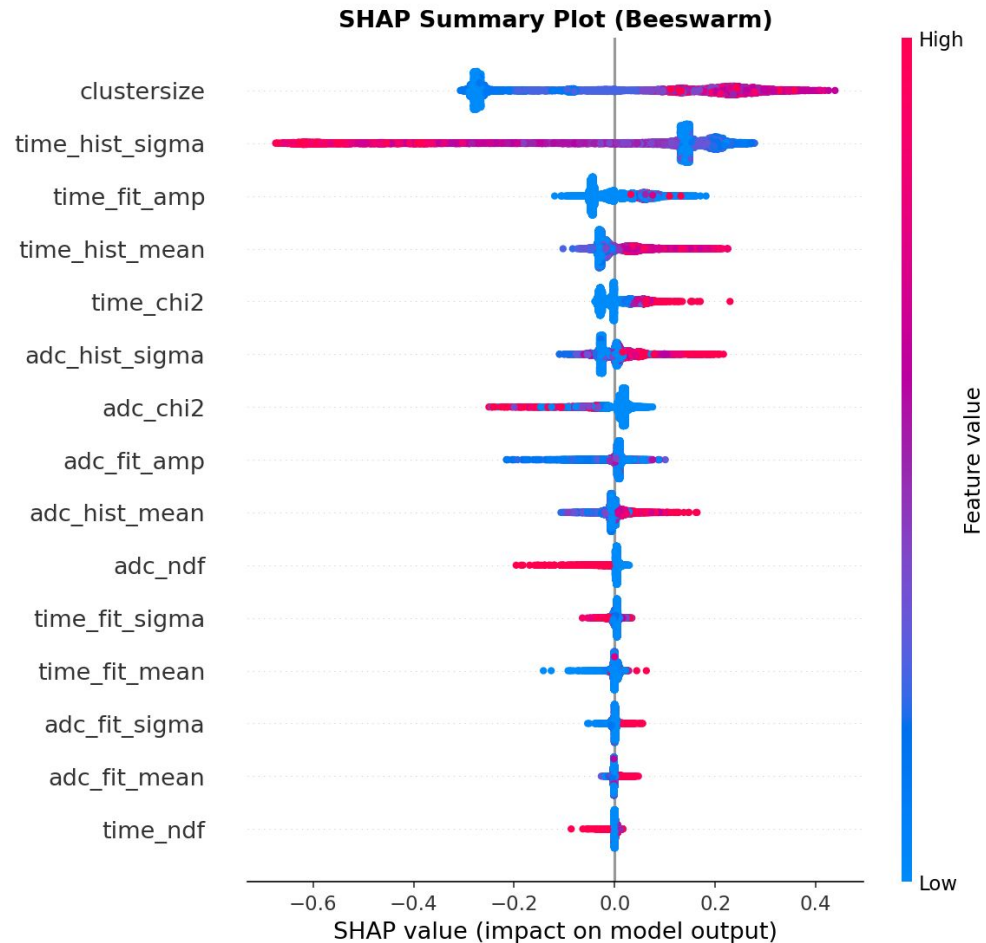
XGBoost: 95.4% precision with 81.6% recall

Feature importances from the trained XGBoost classifier (average gain).

Feature Description	Gain	Share (%)
Cluster size	0.4177	41.8
Standard deviation of time distribution	0.2290	22.9
Amplitude from Gaussian time fit	0.0936	9.4
Standard deviation of ADC distribution	0.0529	5.3
χ^2 of ADC fit	0.0335	3.3
χ^2 of time fit	0.0254	2.5
NDF of ADC fit	0.0232	2.3
Mean ADC value across hits	0.0243	2.4
Mean time of hits in cluster	0.0171	1.7
Amplitude from ADC Gaussian fit	0.0182	1.8
Mean of Gaussian time fit	0.0148	1.5
Sigma of Gaussian ADC fit	0.0145	1.5
Sigma of Gaussian time fit	0.0144	1.4
Mean of Gaussian ADC fit	0.0094	0.9
NDF of time fit	0.0119	1.2
Total	1.0000	100.0

- **Cluster size:** most influential feature, contributing **>40%** of total gain.
- **Temporal-shape descriptors:** important for signal–background separation.
- **Time-histogram width + time-fit amplitude:** together contribute **>30%** of total gain.

Shap value for different features



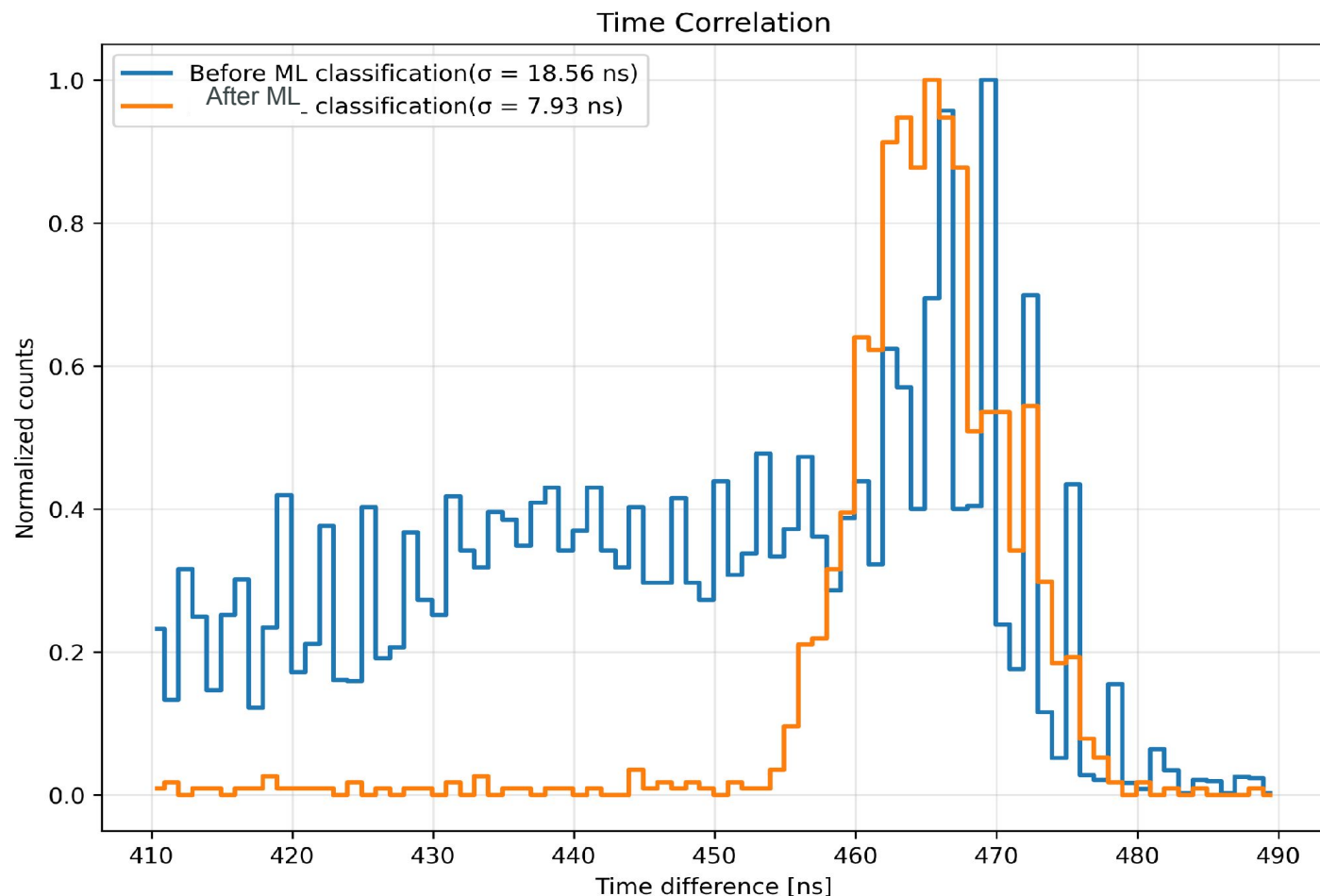
Larger cluster size favors signal

Larger time spread favors background

XGBoost gain shares
Cluster size: 41.8%
Time histogram σ : 22.9%

Positive SHAP values favor signal; negative values favor background

Time correlation before and after background Suppression

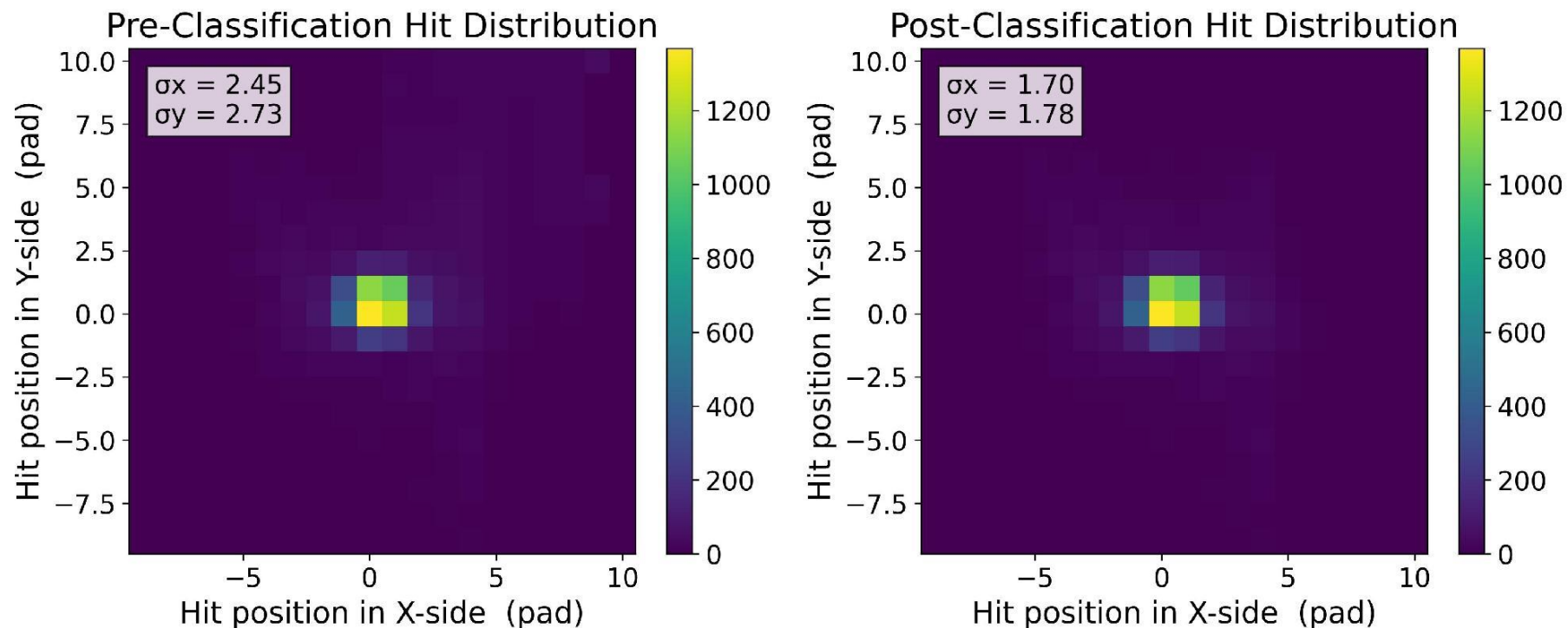


18.56 $\rightarrow$ 7.93 ns

57% smaller spread

Total system distribution;
intrinsic RPC timing is not
isolated.

Spatial hit distributions before and after classification



σ_x : 2.45 $\rightarrow$ 1.70 pads

σ_y : 2.73 $\rightarrow$ 1.78 pads

Conclusions

- Developed an **ML-based pipeline** to suppress background in free-streaming RPC data.
- Used **spatial–temporal clustering** and **15 cluster-level features** for classification.
- **DNN, CNN and XGBoost** achieved comparable performance, with XGBoost giving the highest **ROC AUC ≈ 0.94** .
- **Cluster size and timing descriptors** provided the strongest discrimination.
- Classification reduced background and narrowed the observed **time and spatial distributions**.
- The approach supports **triggerless RPC data analysis**.

Thank you

Thank You!

References

1. **S. Chattopadhyay et al.**, “Performance of a real-size, low resistivity resistive plate chamber at GIF++ using self-trigger electronics for the Muon Chamber of the CBM Experiment,” *Journal of Instrumentation* **20** (2025) P03009.
DOI: [10.1088/1748-0221/20/03/P03009](https://doi.org/10.1088/1748-0221/20/03/P03009)
2. **M. Mondal et al.**, “Performance of a prototype bakelite RPC at GIF++ using self-triggered electronics for the CBM experiment at FAIR,” *Nuclear Instruments and Methods in Physics Research A* **1025** (2022) 166042.
DOI: [10.1016/j.nima.2021.166042](https://doi.org/10.1016/j.nima.2021.166042)

BACKUP | COMPLETE FEATURE CATALOGUE

Group	Variables
Size (1)	clustersize
Time histogram (2)	time_hist_mean, time_hist_sigma
Time Gaussian fit (3)	time_fit_mean, time_fit_sigma, time_fit_amp
Time fit quality (2)	time_chi2, time_ndf
ADC histogram (2)	adc_hist_mean, adc_hist_sigma
ADC Gaussian fit (3)	adc_fit_mean, adc_fit_sigma, adc_fit_amp
ADC fit quality (2)	adc_chi2, adc_ndf

Backup:benchmark

100,000 events in 4.46 s

≈1.5 million hits · ≈330 MB memory
AMD EPYC 7713, single-threaded execution

Stage	μs / event
Clustering	19.4
Feature extraction	2.4
XGBoost inference	14.6
Full pipeline	44.6

BACKUP | INTERPRETATION AND VALIDATION SCOPE

Established by this study

Laboratory classification with trigger-derived labels

Independent-test metrics and bootstrap AUC uncertainties

Narrower selected distributions and a CPU benchmark

Validation still needed

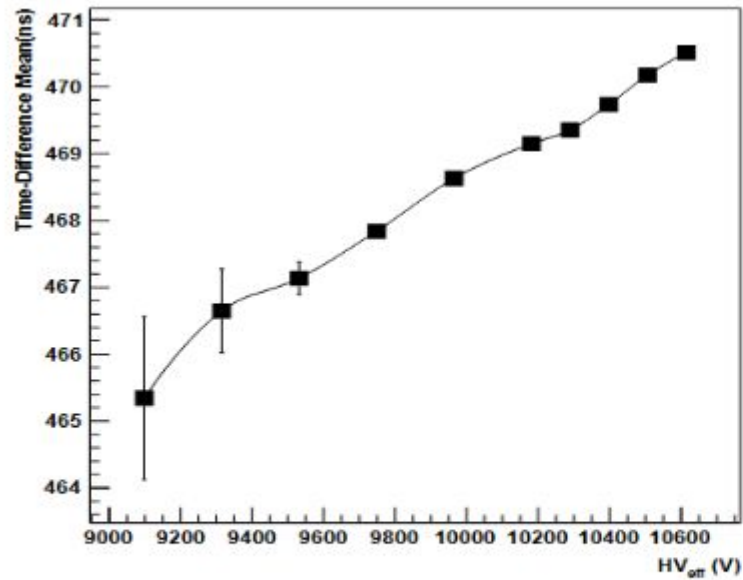
Run-separated tests and changes in rate, voltage or environment

Background rejection at a fixed signal-retention target

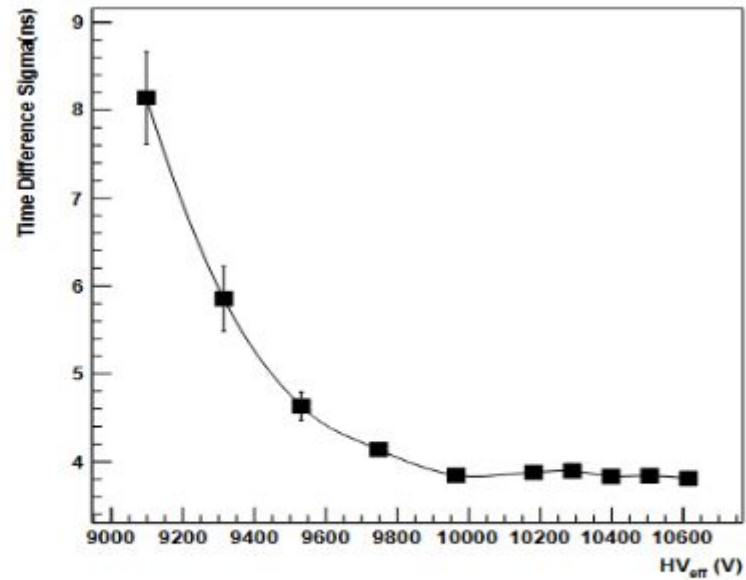
Intrinsic resolution measurements and full online integration

Exact split sizes, label windows and classifier thresholds are not specified in the paper

Time difference mean and Sigma

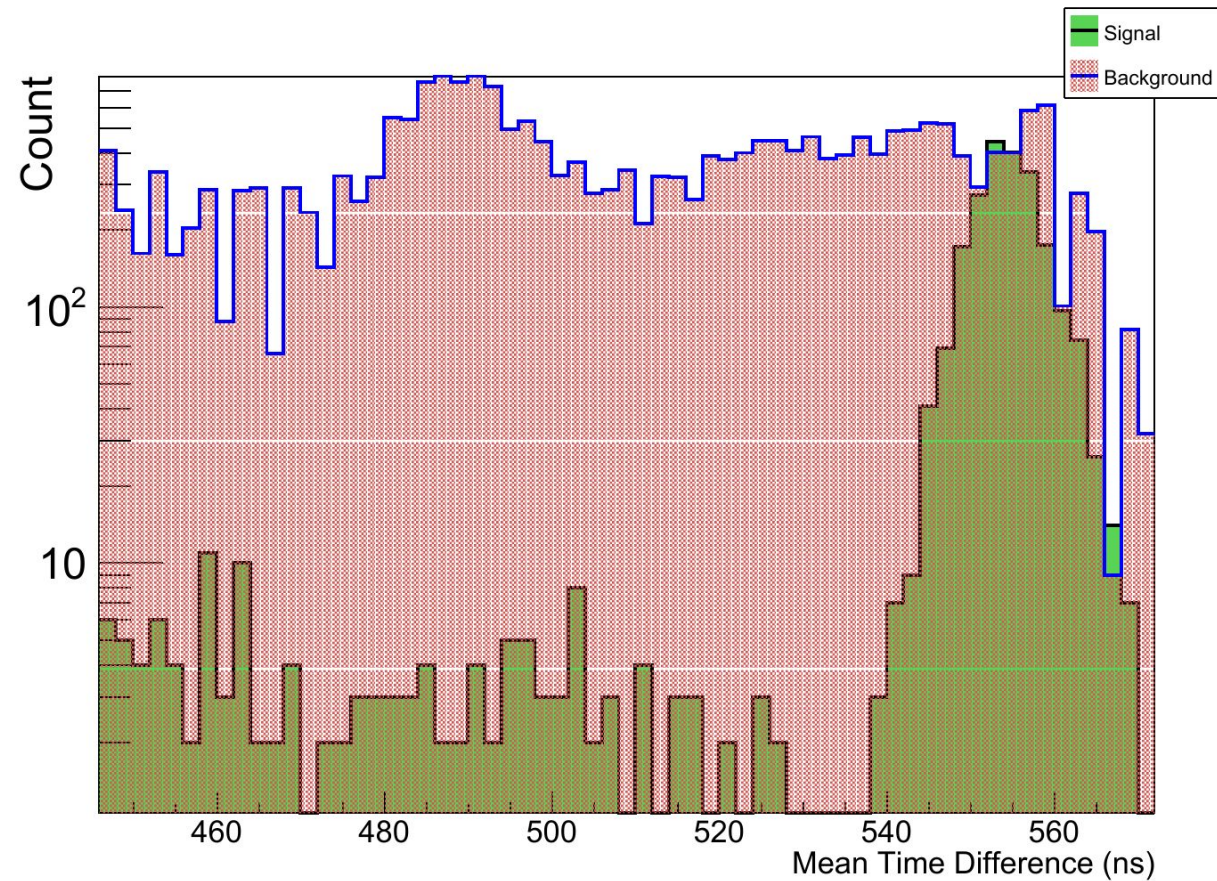


Time-Difference Mean as a function of effective high voltages



Time-Difference Sigma as a function of effective high voltages

DNN SELECTION SEPARATES THE TIMING COMPONENTS



The broader component is mostly classified as background.