

Machine Learning approach to CMS RPC HV scan data analysis

(Автоматизиране на процеса на определяне на работното напрежение на камерите със съпротивителна плоскост на експеримента **CMS** чрез използване на методи от машинното обучение)



Михаела Пехливанова

специалност “Физика на ядрото и елементарните частици”

научен ръководител: доц. д-р. Пейчо Петков

*Катедра “Атомна физика”
Софийски университет “Св. Климент Охридски”*

CMS (Компактен мюонен соленоид)

CMS DETECTOR

Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T

STEEL RETURN YOKE
12,500 tonnes

SILICON TRACKERS
Pixel ($100 \times 150 \mu\text{m}$) $\sim 16\text{m}^2 \sim 66\text{M}$ channels
Microstrips ($80 \times 180 \mu\text{m}$) $\sim 200\text{m}^2 \sim 9.6\text{M}$ channels

SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying $\sim 18,000\text{A}$

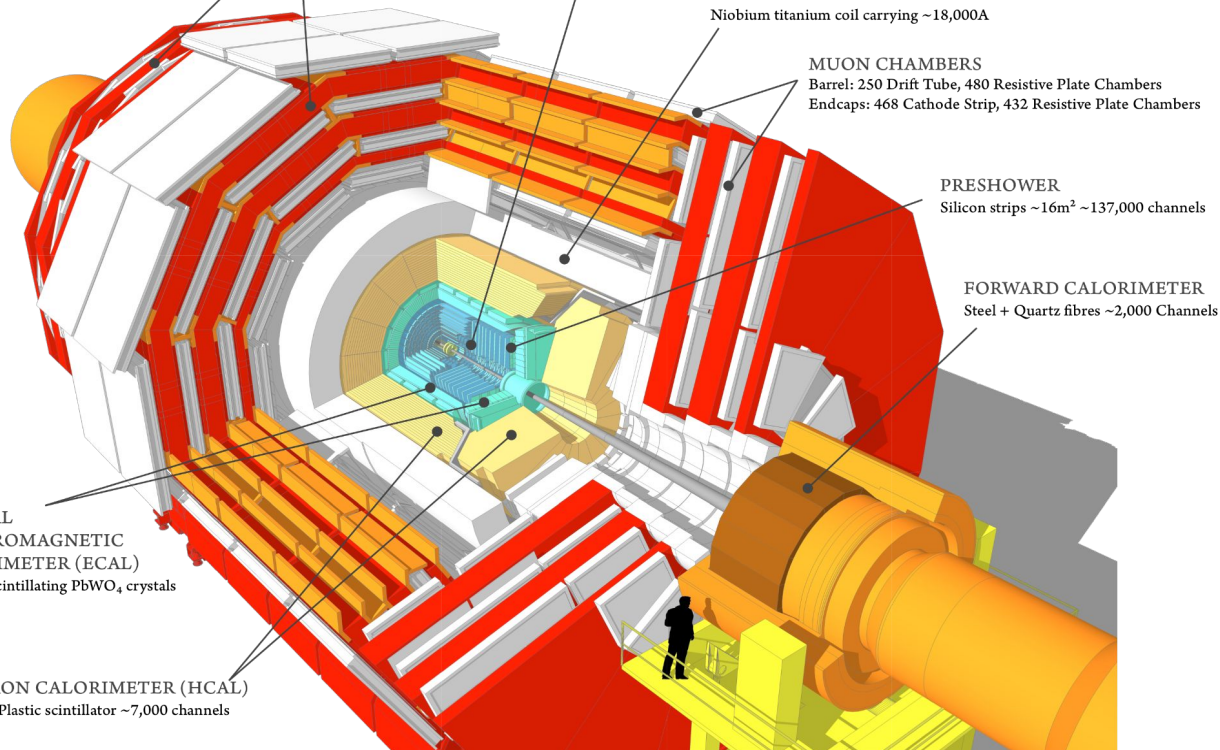
MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 468 Cathode Strip, 432 Resistive Plate Chambers

PRESHOWER
Silicon strips $\sim 16\text{m}^2 \sim 137,000$ channels

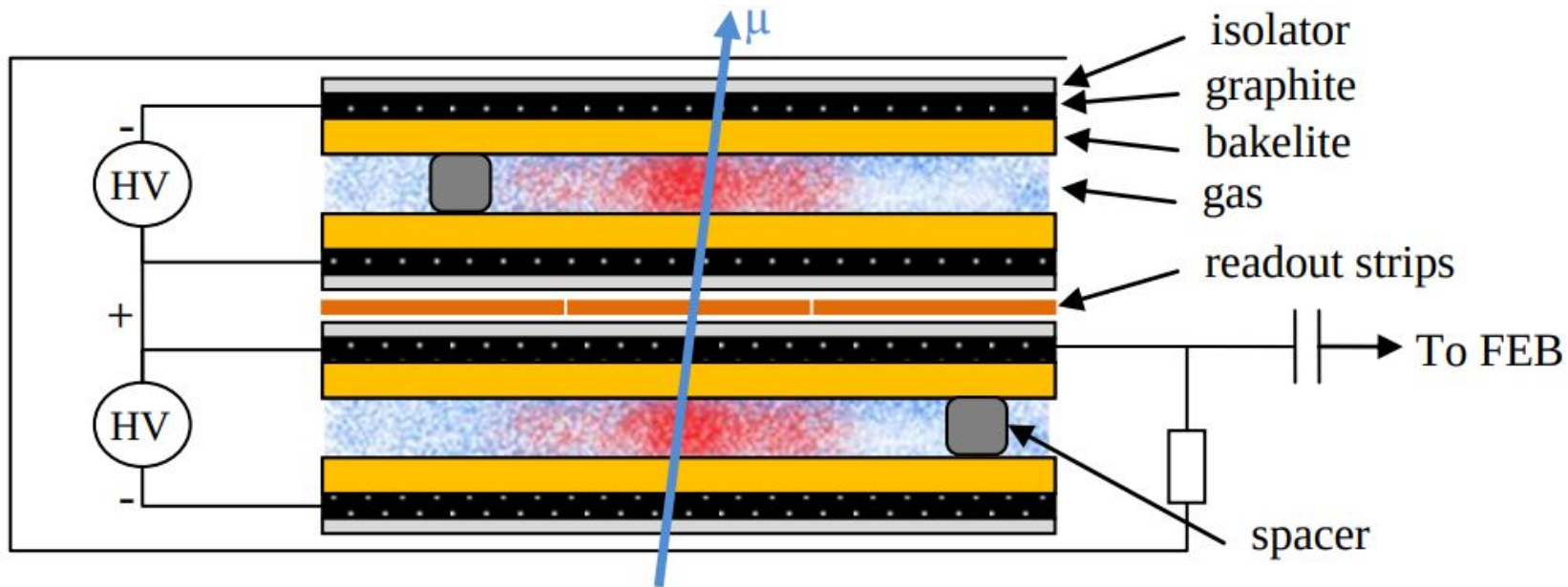
FORWARD CALORIMETER
Steel + Quartz fibres $\sim 2,000$ Channels

CRYSTAL
ELECTROMAGNETIC
CALORIMETER (ECAL)
 $\sim 76,000$ scintillating PbWO_4 crystals

HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator $\sim 7,000$ channels

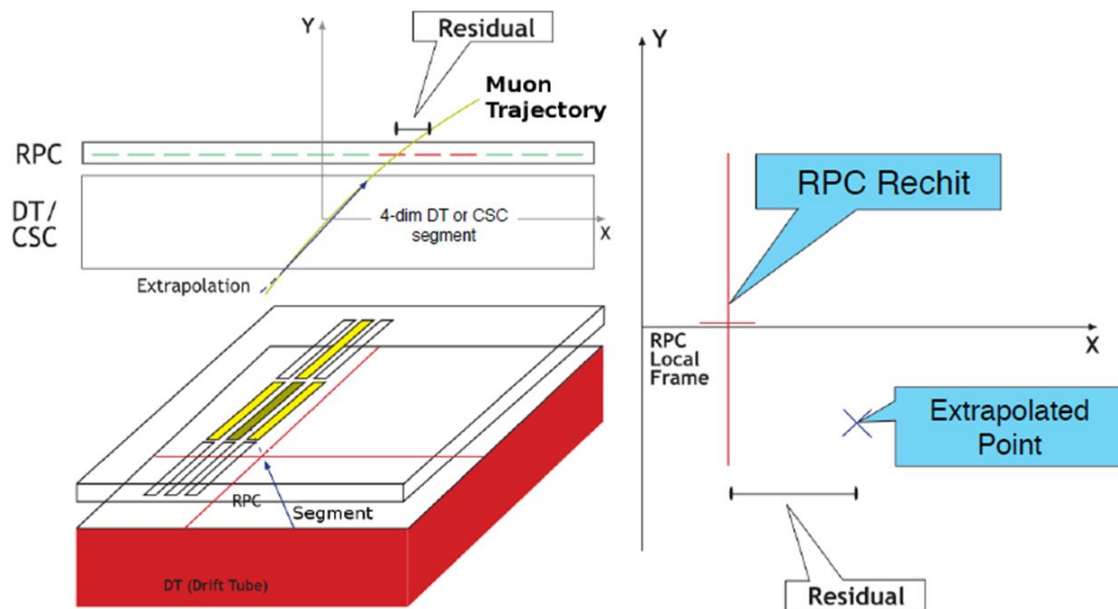


RPC (Камери със съпротивителна плоскост)



газова смес от 95.2% $\text{C}_2\text{H}_2\text{F}_4$ + 4.5% iC_4H_{10} + 0.3% SF_6

Определяне на ефективността на **RPC**

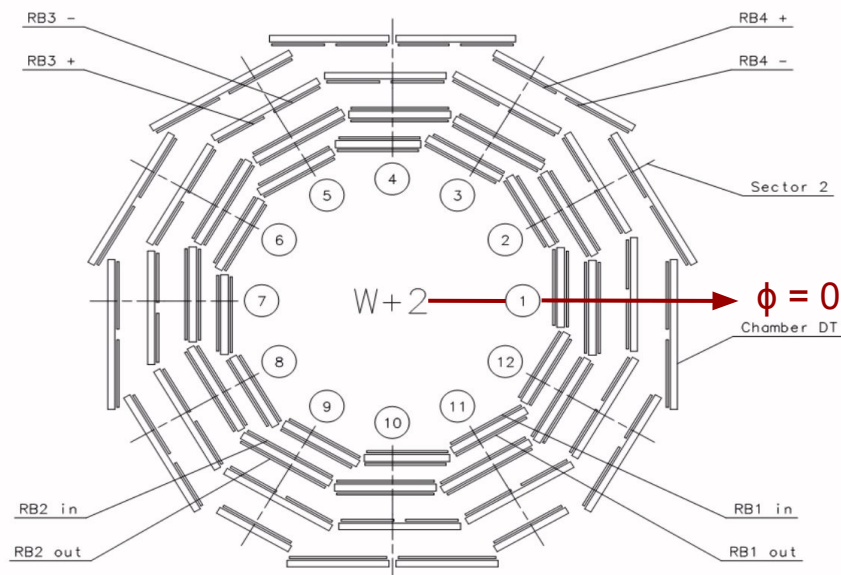
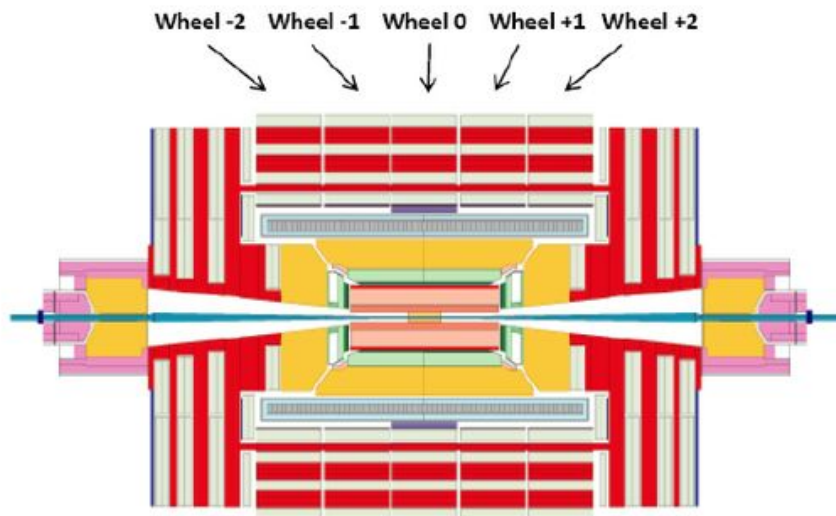


$$Efficiency = \frac{\#MuonsAtRPC}{\#CSC\ DT\ Muons\ Extrap}$$

Segment Extrapolation Method

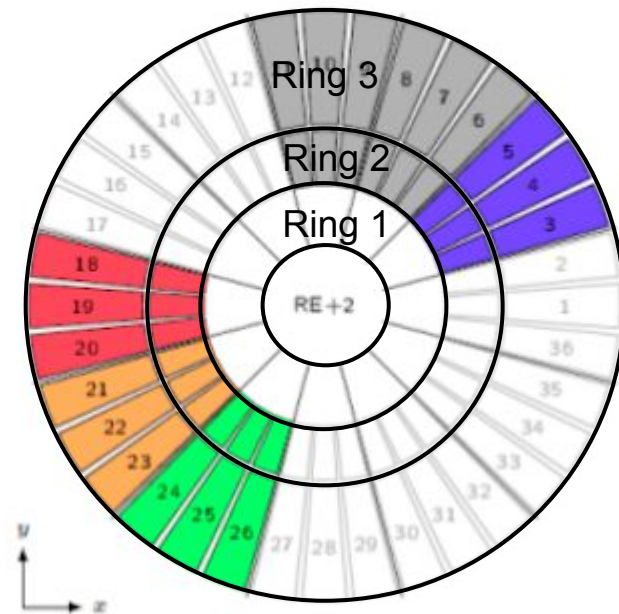
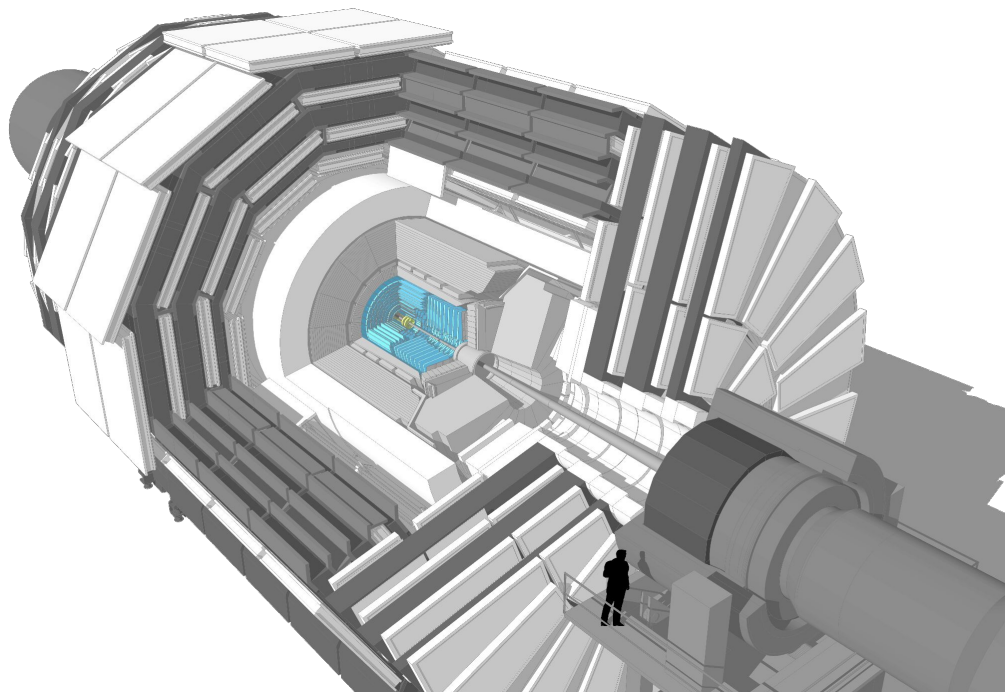
[JINST 5 \(2010\) T03017, DOI: 10.1088/1748-0221/5/03/T03017](#)

Камери със съпротивителна плоскост в цилиндричната част на **CMS**



[CMS-DP-2017-053](#)

Камери със съпротивителна плоскост в затварящите части на **CMS**



[CMS-DP-2017-053](#)

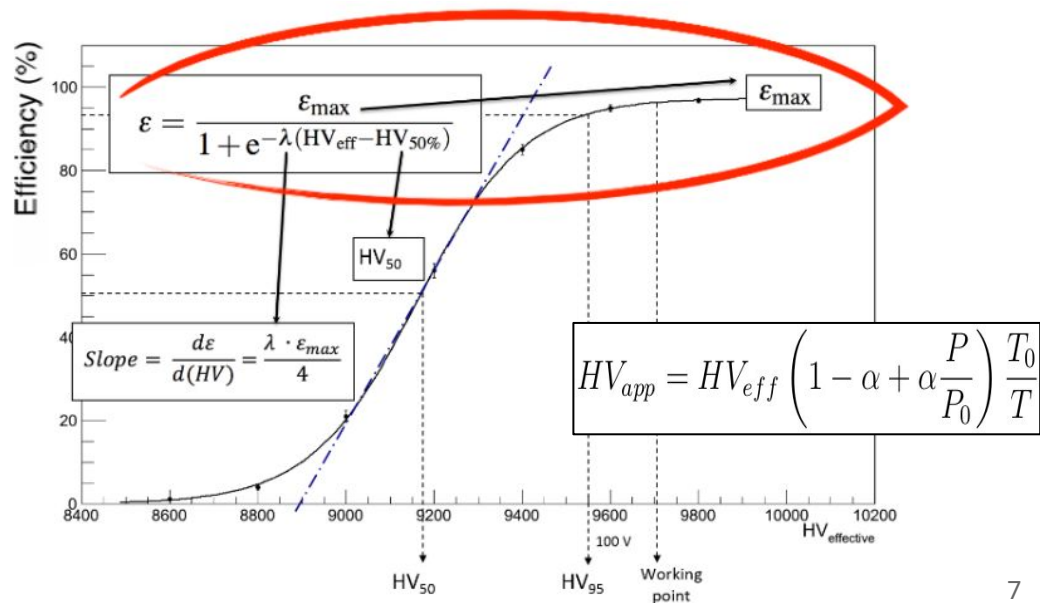
HV-Scan

Цел: определяне на работни напрежения на камерите със съпротивителна плоскост (1054 камери, захранвани от над 700 HV канала)

Последователни колекции от събития за 11 точки на напрежение в диапазона [8.6,9.8] kV

$$HV_{WP} = HV_{knee} + \begin{cases} 100 \text{ V (Barrel)} \\ 150 \text{ V (Endcap)} \end{cases}$$

$$HV_{knee} = \frac{\lambda HV_{50\%} - \ln\left(\frac{1}{0.95} - 1\right)}{\lambda}$$



Процедура за анализ на данни от **HV-Scan**

1. Изчисляване на $\epsilon_{\max}$, $HV_{50\%}$ и $\text{slope}_{50\%}$ за всеки детекторен елемент
2. Детекторните елементи се разделят на две групи: една, която удовлетворява условията за $\epsilon_{\max}$, и друга, която не ги удовлетворява
3. Определя се нова работна точка (WP) за всеки детекторен елемент, удовлетворяващ условията за $\epsilon_{\max}$
4. Определя се WP за всеки канал на захранващата система на RPC
5. **Ръчно** се определя WP за детекторните елементи, които не изпълняват условията за $\epsilon_{\max}$
6. Подготвя се конфигурационен файл за DCS с новите работни точки

Цели и задачи на дипломната работа:

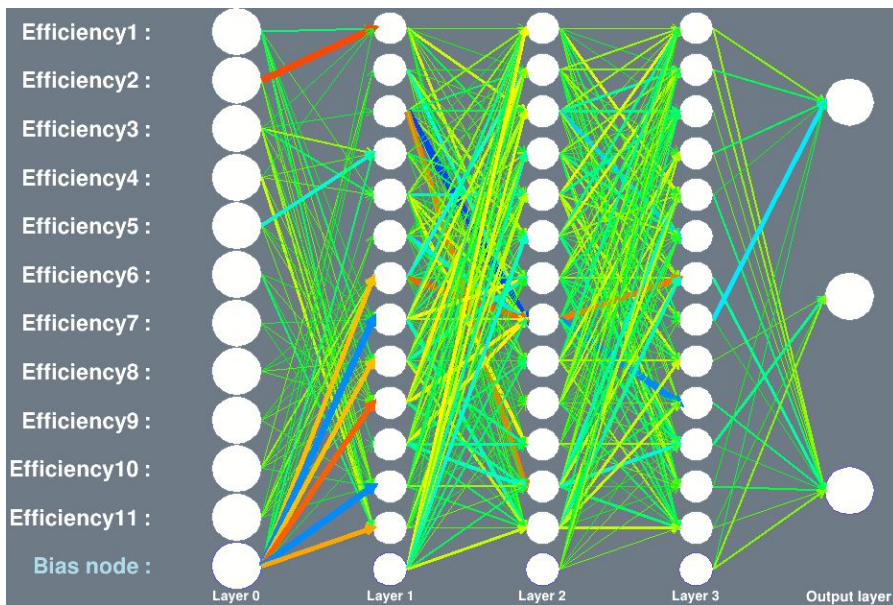
Автоматизация на процеса на анализ на данните чрез използване на методи от машинното обучение

Задачи:

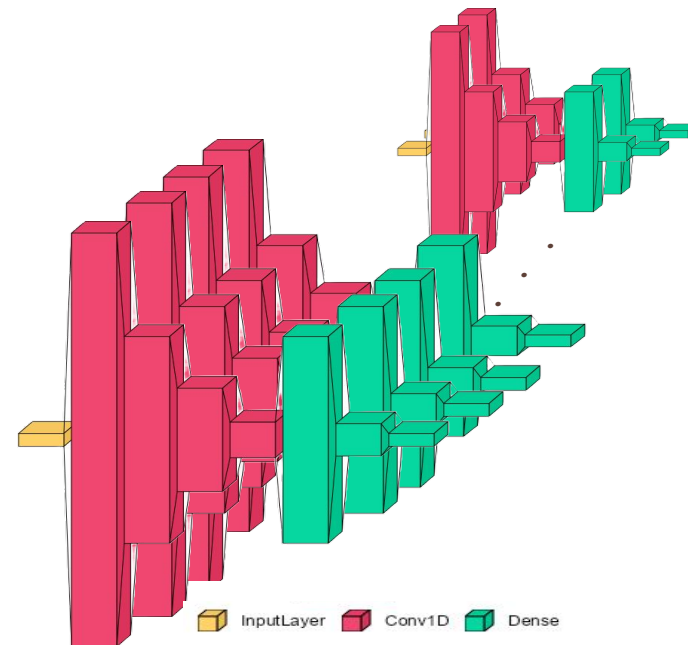
- **Разработване** на модели от машинното обучение
- **Повишаване на точността** на оценките за ефективността
- **Разработване на софтуер за анализа на данни:** да включва модул за ML задачата, модул за апроксимация на данните със сигмоидална функция и извличането на параметрите ѝ, модул за определяне на работните точки на камерите, създаването на списъци с “добри” и “лоши” камери, създаване на конфигурационен файл за контролната система на CMS (CMS DCS)

Искусственные нейронные сети

ANN

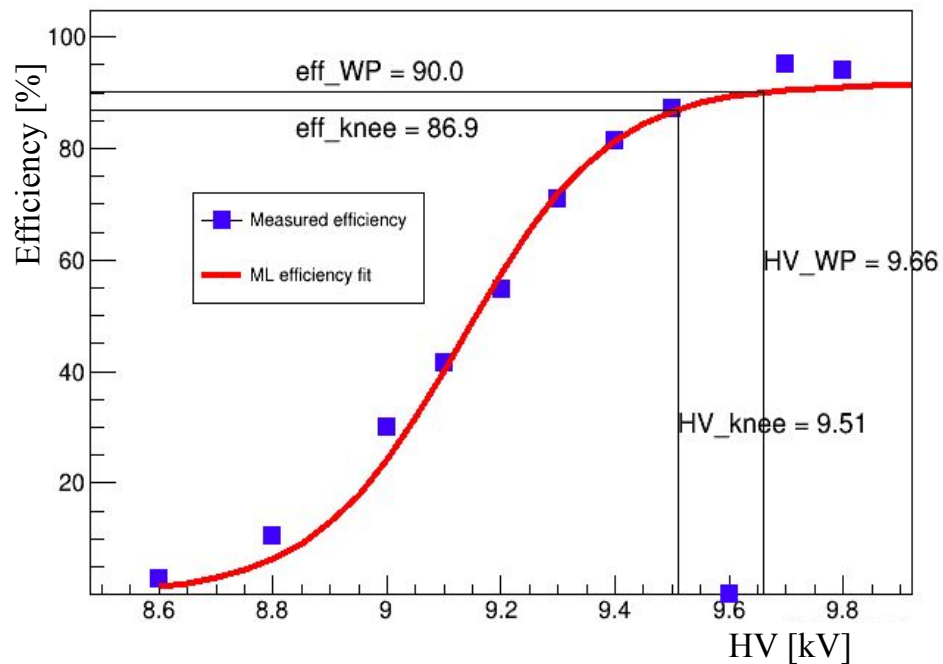


1D CNN

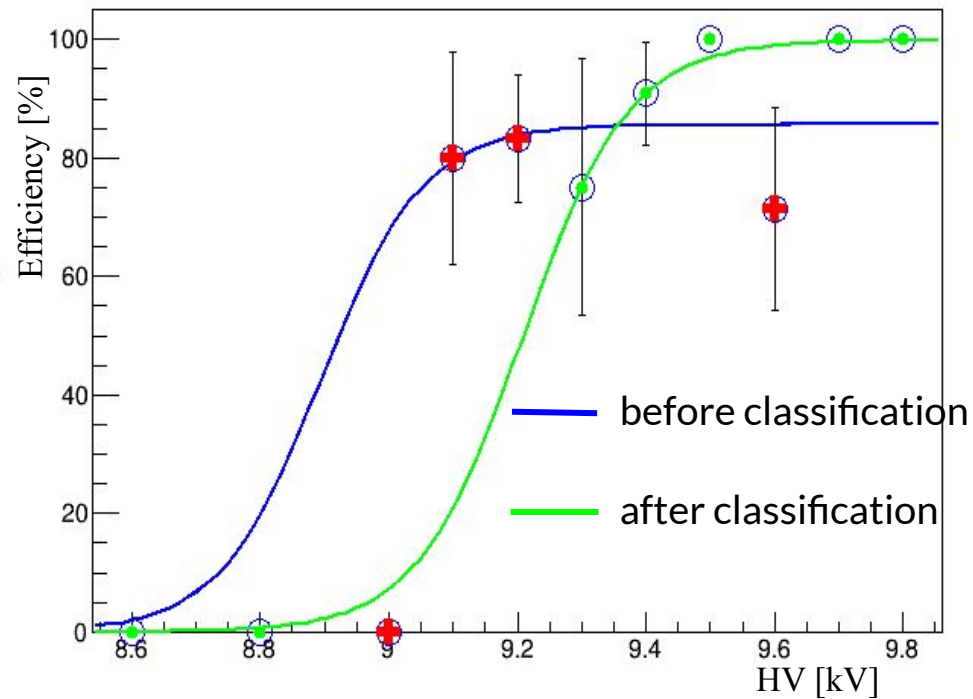


Добри резултати

ANN

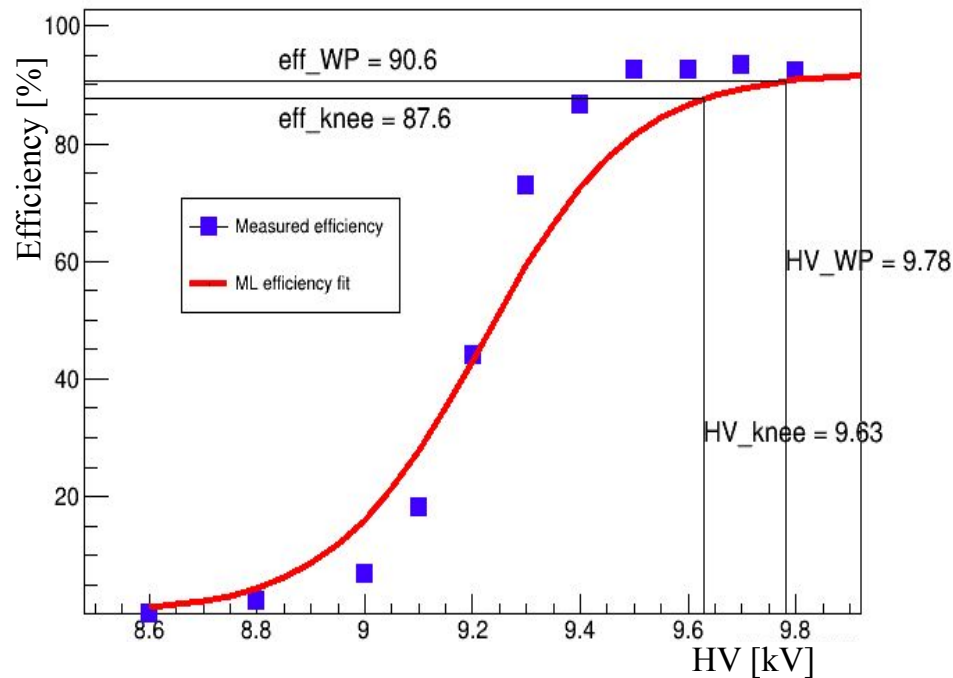


1D CNN

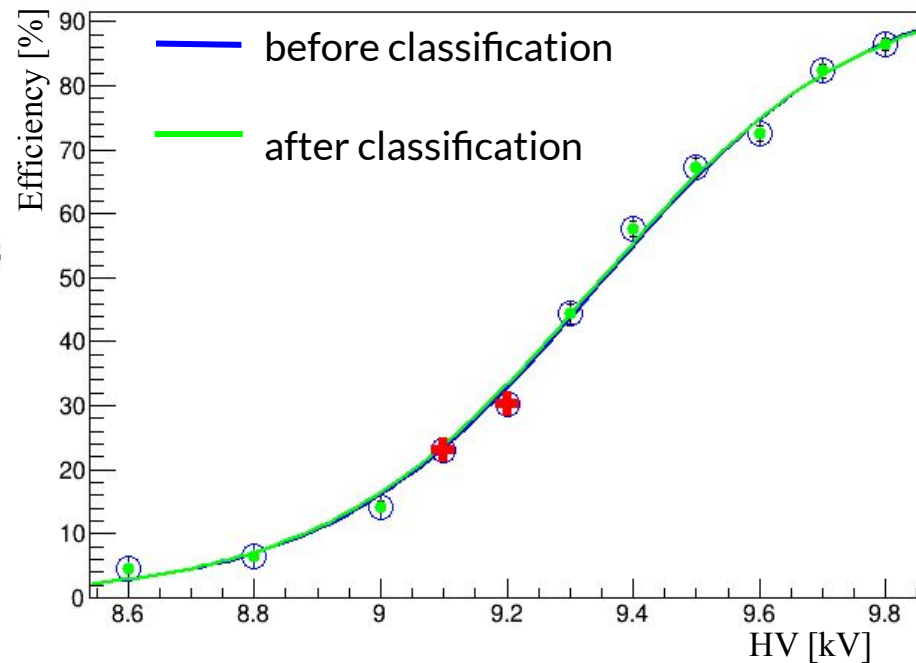


Не толкова добри резултати

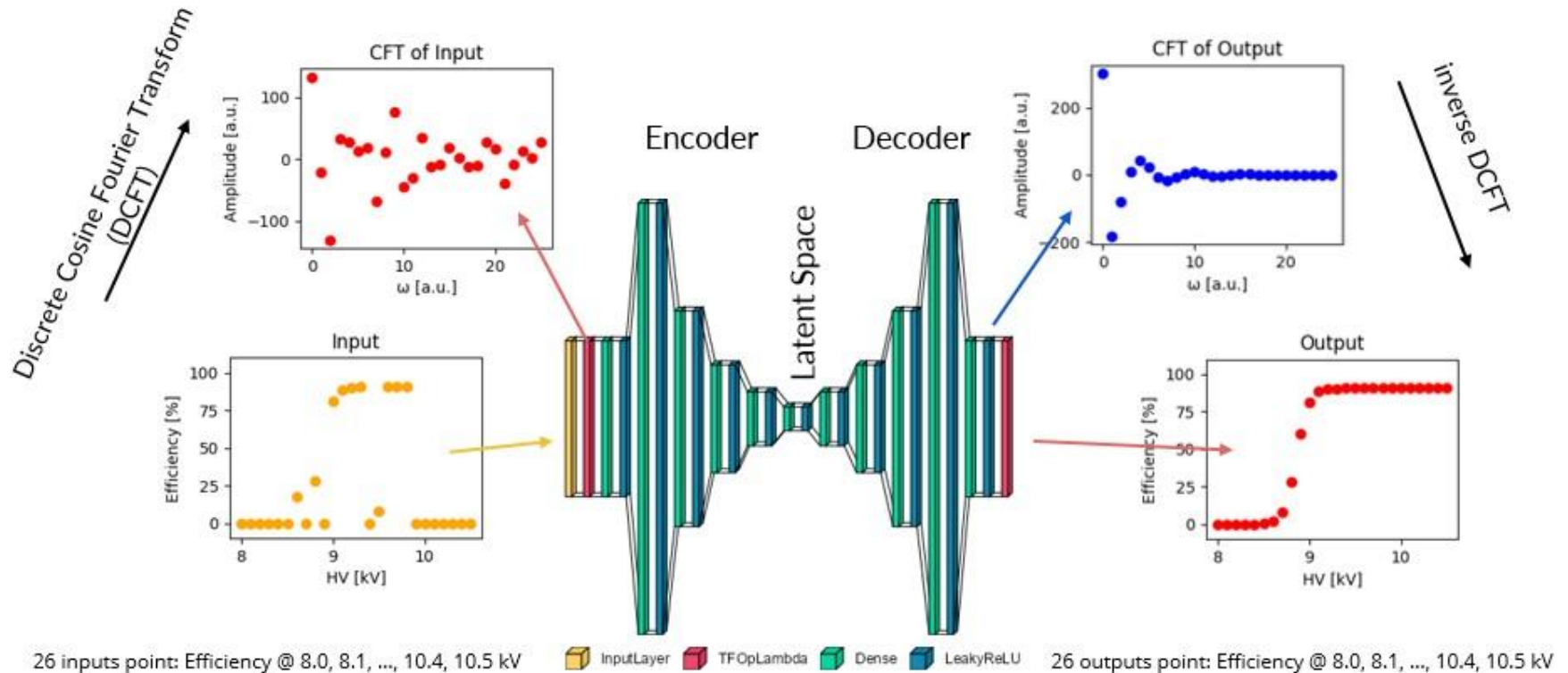
ANN



1D CNN



Fourier Space Autoencoder (FSAC)

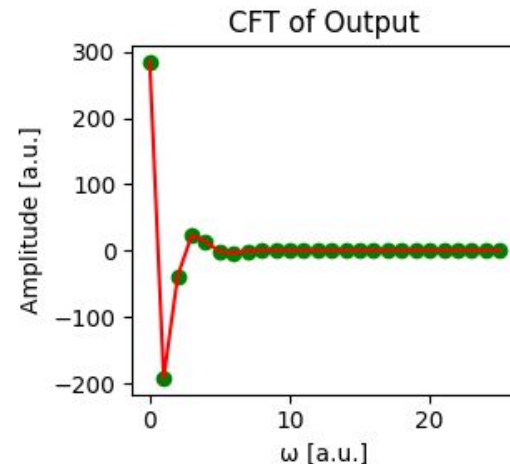
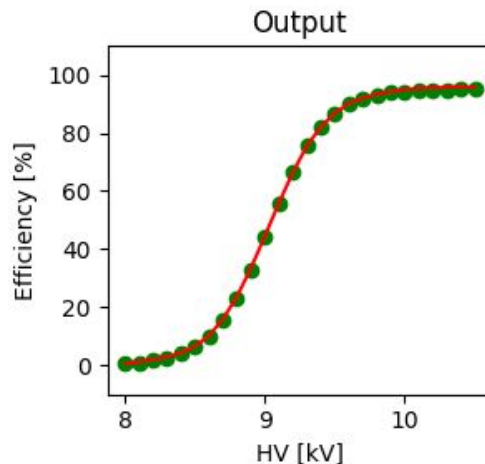
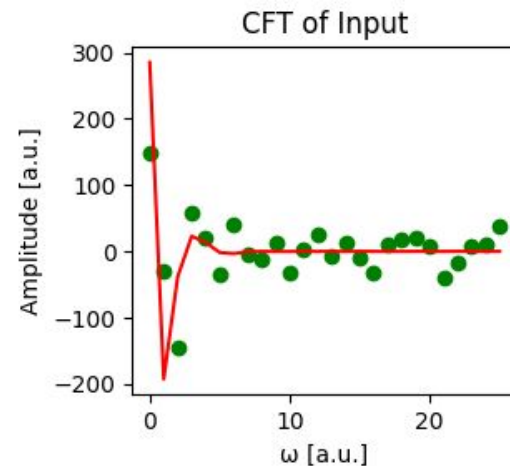
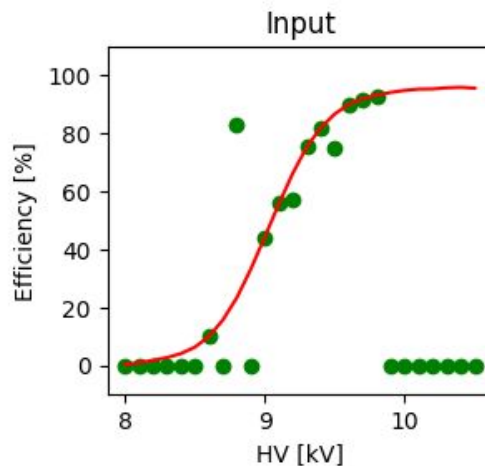


Обучение на модела

Синтетични данни, генерирани
от сигмоидални функции

зелените маркери:
генерираните данни

червените криви: първични
функции

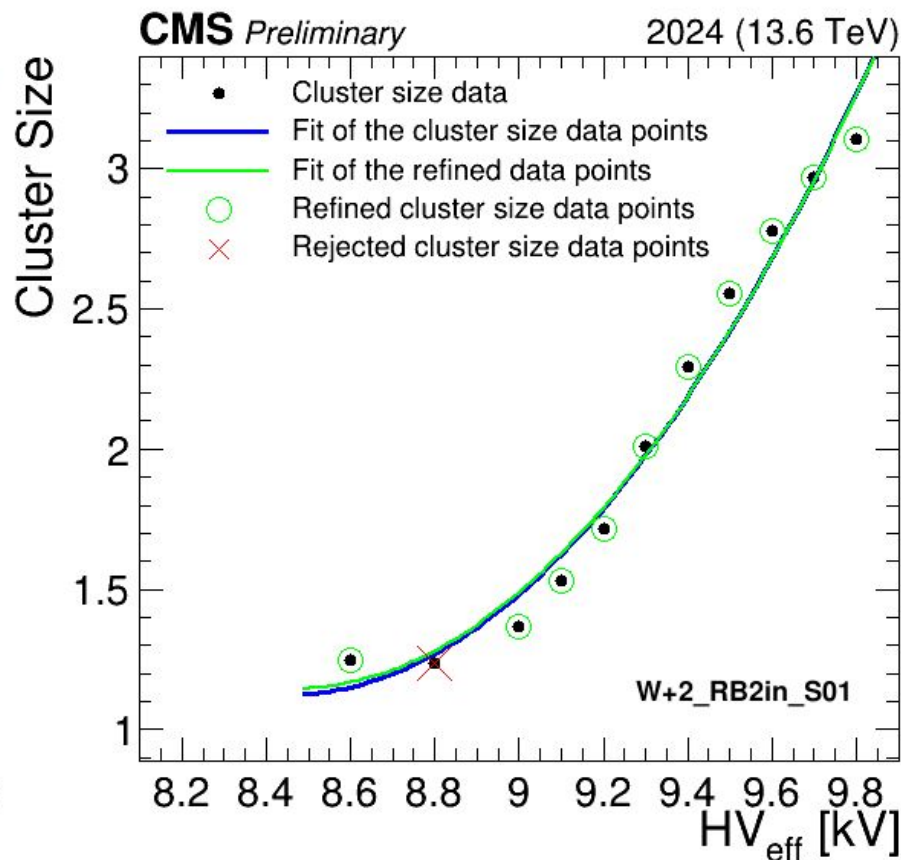
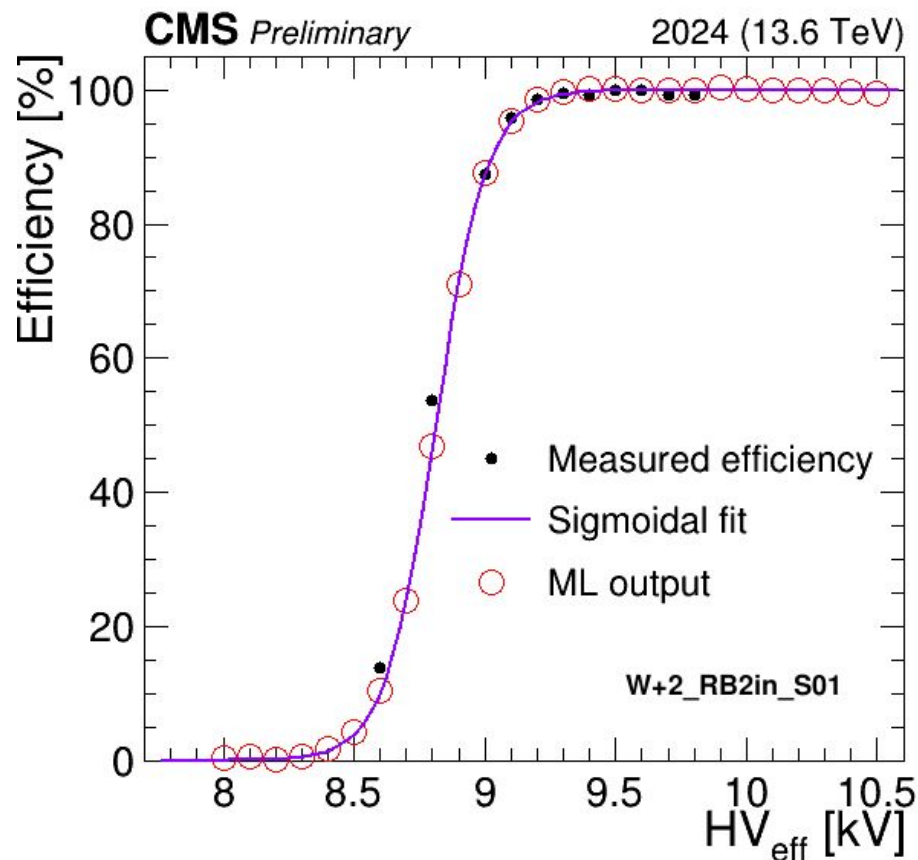


Обработване на данните чрез **FSAC**

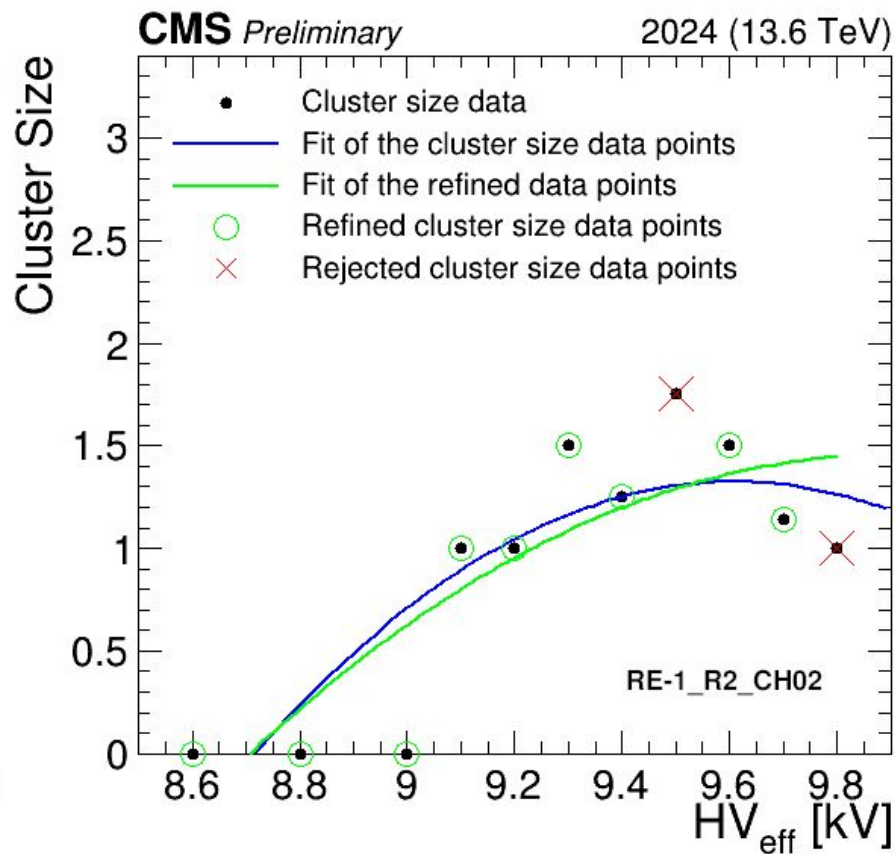
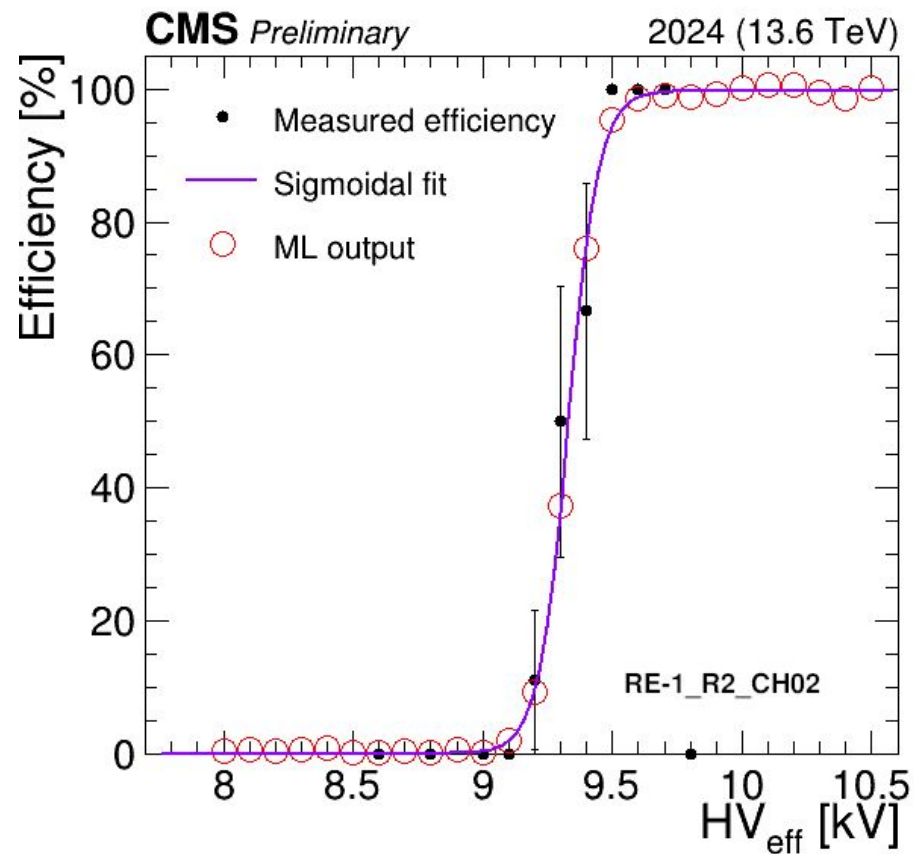
1. Входните данни за FSAC: задава се ефективността за известните HV точки, а останалите ефективности = 0
2. Прилагане на FSAC
3. Изчисляване на параметрите $\epsilon_{\max}$, $HV_{50\%}$ и $\text{slope}_{50\%}$ за предсказаните от модела ефективности
4. Отстраняване на несъвместимите точки от данните за размер на клъстера
5. Апроксимиране на зависимостта на размера на клъстера от напрежението с полином от втора степен

Тестване

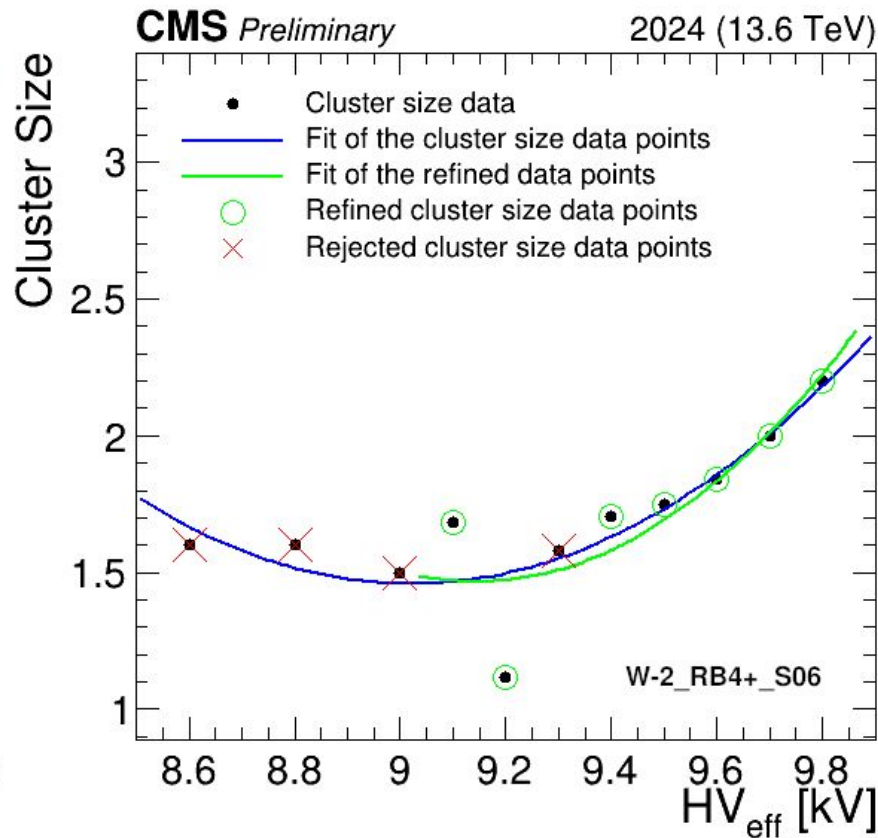
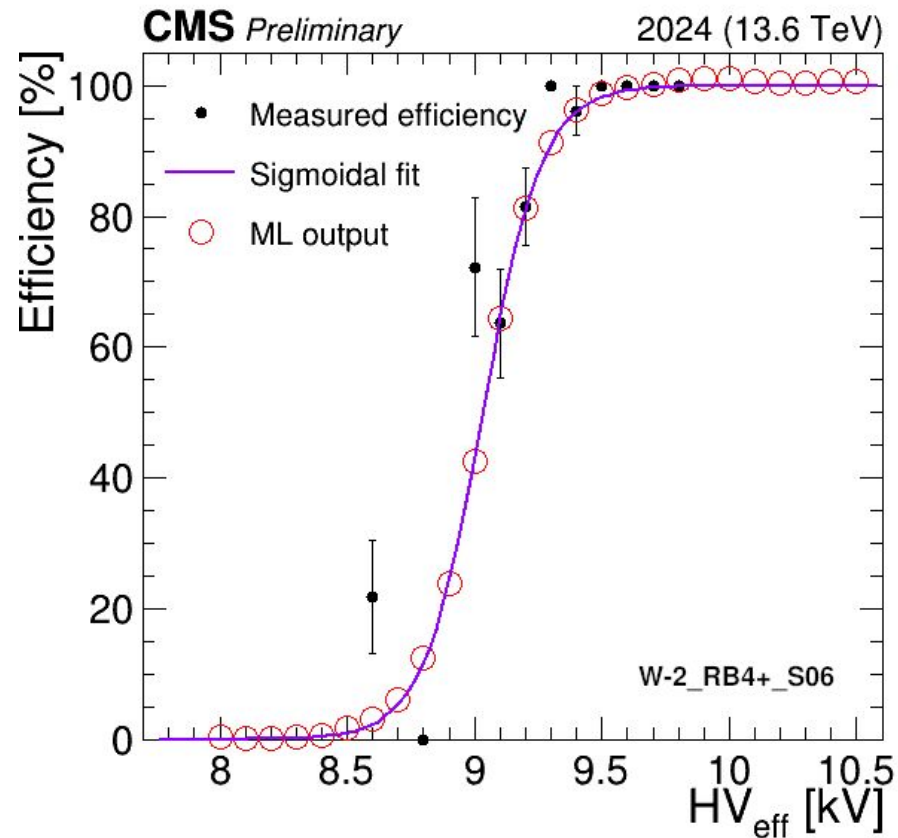
Пример за данни, при които точките са разположени без значителни отклонения върху кривата.



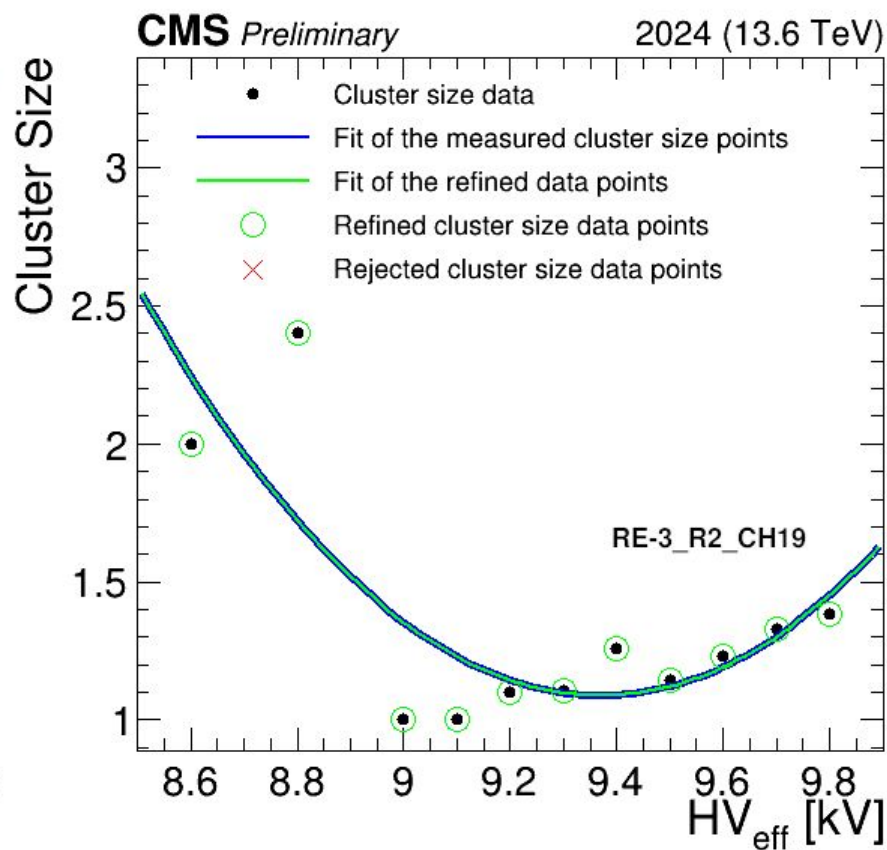
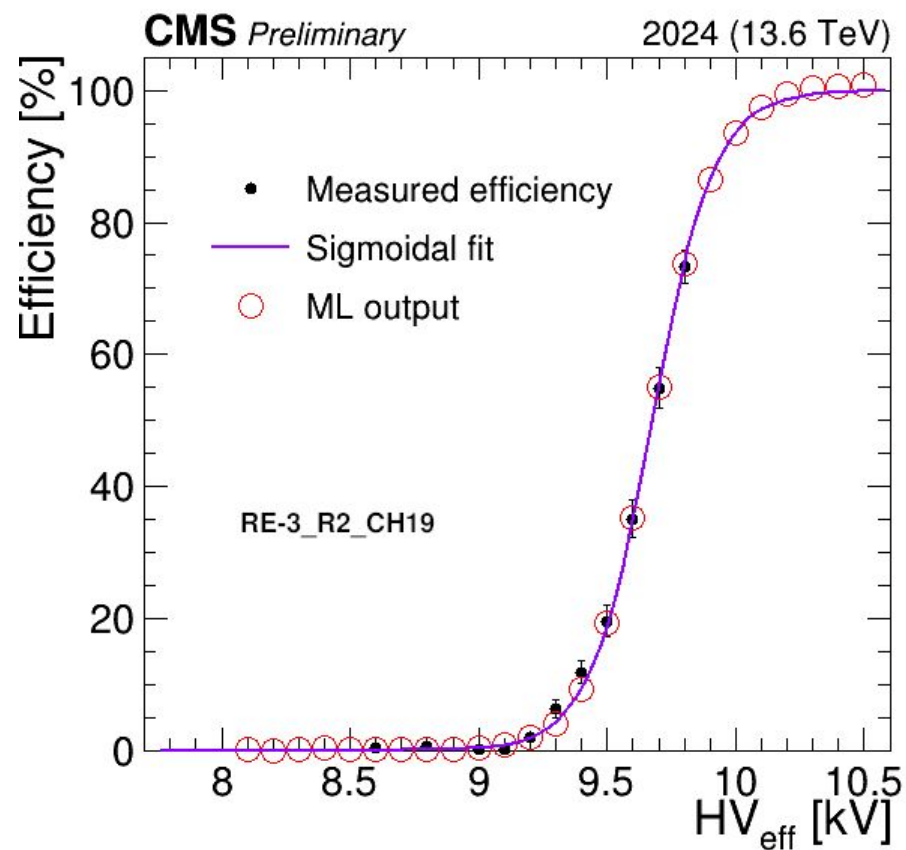
Пример за данни, при които последната точка има нулева измерена ефективност.



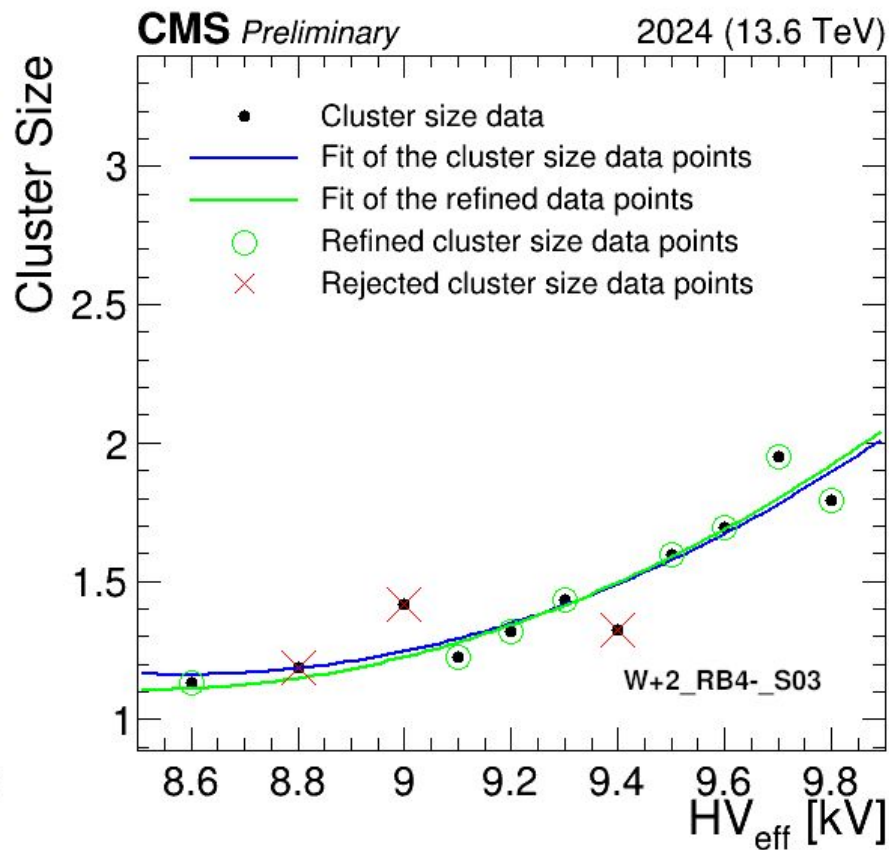
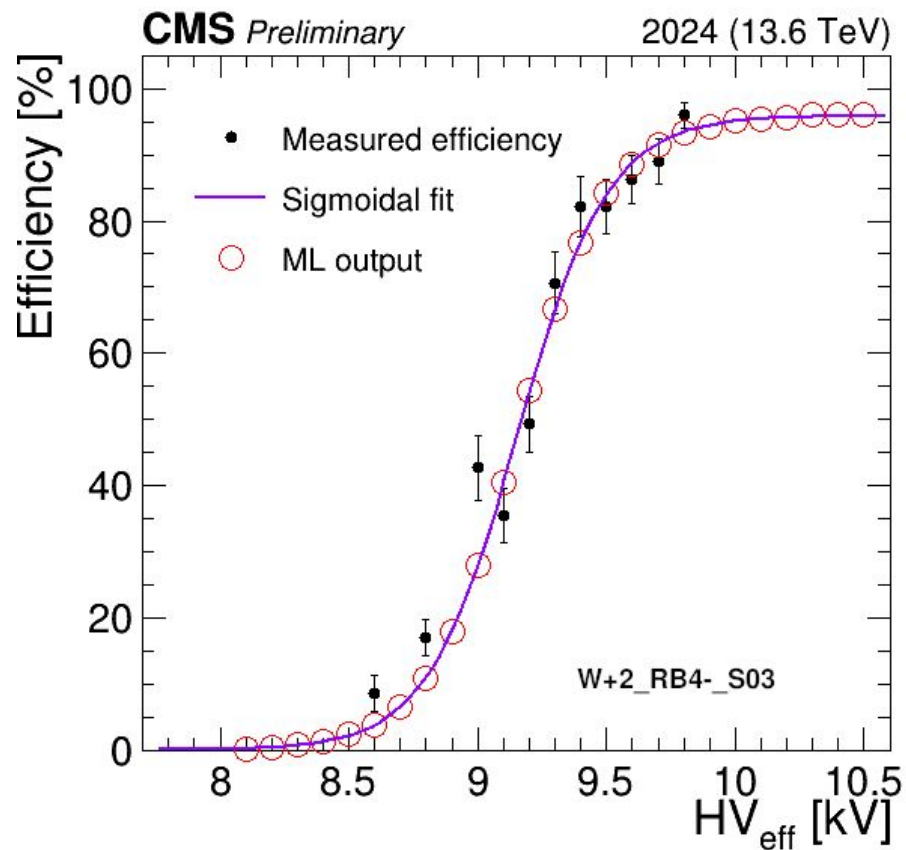
Пример за данни, при които има наличие на точки със значителни отклонения от кривата.



Пример за данни, при които има липса на плато.

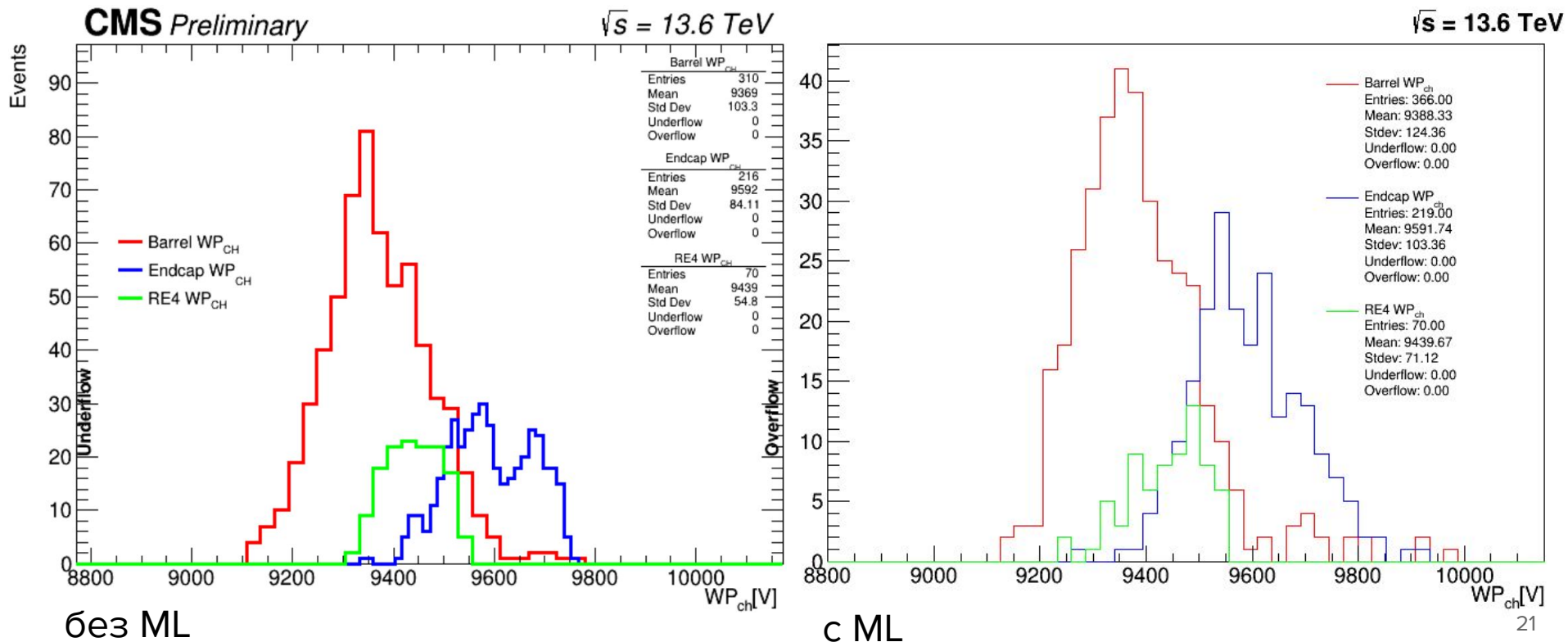


Пример за данни, при които има и наличие на точки с отклонения от кривата, и липсва добре дефинирано плато.



Разпределения на работните напрежения на захранващите канали за двете процедури

Използваните данни са от **2022** г.



Софтуерна имплементация

- Python
- <https://gitlab.cern.ch/ppetkov/ml-playground>
- `./hvscan.py -i /eos/cms/store/group/dpg_rpc/comm_rpc/Run-III/HVscan2024/Input_files -b
../maps/detIdBarrelScan2022.txt -e ../maps/detIdEndCapScan2022.txt -y ../maps/map2.yaml -m
nfo.keras -l 2024err -p params_df_2024err.csv`

Заключение:

1. Бяха изследвани три различни методи от машинното обучение за цел да се автоматизира анализа на кривите на ефективност на RPC
2. Fourier Space Autoencoder беше избран като най-подходящ метод и беше тестван върху данни от HV-Scan от 2022 г. и 2024 г.
3. За цялостното реализиране на анализа беше създаден софтуер, с помощта на който времето беше съкратено от няколко месеца на по-малко от час
4. Разработеният софтуер беше използван за актуализиране на работните напрежения на RPC системата
5. През април 2025 г. ще бъдат измерени ефективностите при така зададените напрежения за финално сертифициране на алгоритъма



Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment



Volume 1075, June 2025, 170367

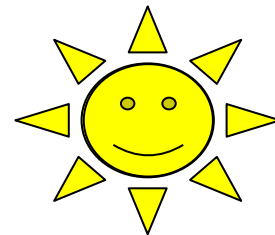
Machine learning approach to CMS RPC HV scan data analysis

M. Pehlivanova⁶  , M. Tytgat^{1 a}, K. Mota Amarilo^{2 b}, A. Samalan^{2 c},
K. Skovpen², G.A. Alves³, E. Alves Coelho³, F. Marujo da Silva³,
M. Barroso Ferreira Filho⁴, E.M. Da Costa⁴, D. De Jesus Damiao⁴,
S. Fonseca De Souza⁴, R. Gomes De Souza⁴, L. Mundim⁴, H. Nogima⁴,
J.P. Pinheiro⁴, A. Santoro⁴, M. Thiel⁴, A. Aleksandrov⁵, R. Hadjiiska⁵...F. Fienga³⁴

<https://doi.org/10.1016/j.nima.2025.170367>

IV-ТИ НАЦИОНАЛЕН КОНГРЕС
ПО ФИЗИЧЕСКИ НАУКИ

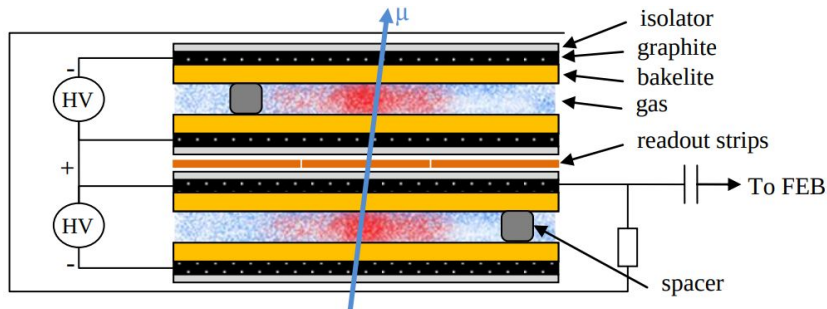
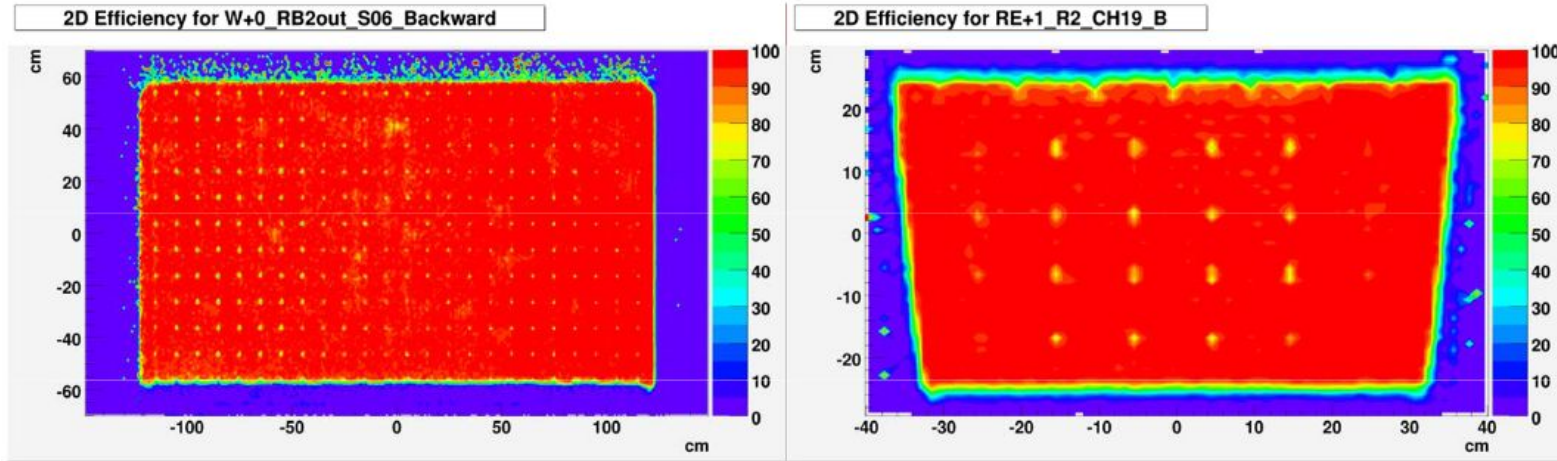
СОФИЯ, 07.10.– 09.10.2024 Г., ИНТЕР ЕКСПО ЦЕНТЪР



Благодаря за вниманието!

1. Какво представлява величината **HV_{knee}** ?

2. Защо има разлика при дефинирането на работното напрежение?



3. На базата на кои критерии е направен изборът за определяне на най-подходящият метод за машинно обучение сред всички тествани?

Функция на загубите:

$$l(w_i) = \frac{1}{n} \sum (y_{pred_{w_i}} - y_i)^2 + \alpha \frac{1}{n} \sum (dy_{pred_{w_i}} + dy_i)^2$$

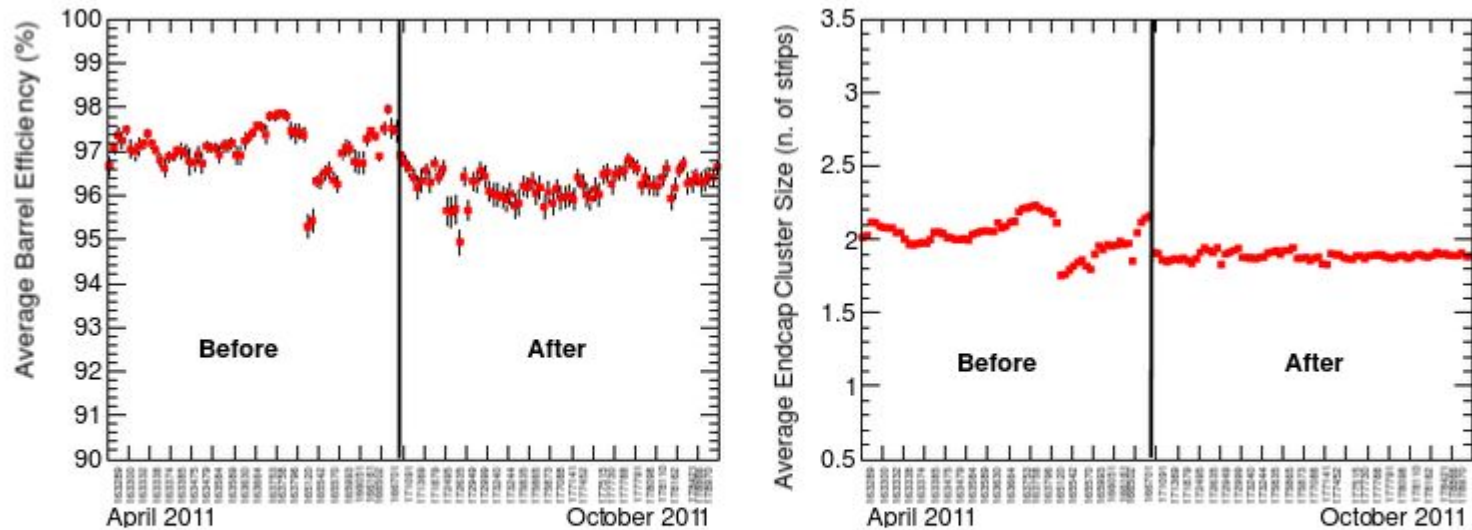


Figure 4: Left: Barrel efficiency (in %) as a function of the run number, *i.e.* as a function of time, for runs taken between April 2011 and October 2011. The two regions, before and after the automatic $HV_{\text{eff}}(P, T)$ correction, are shown in the plot. Right: Endcap cluster size (in number of strips), as a function of the run number, for the same run range between April and October 2011.