



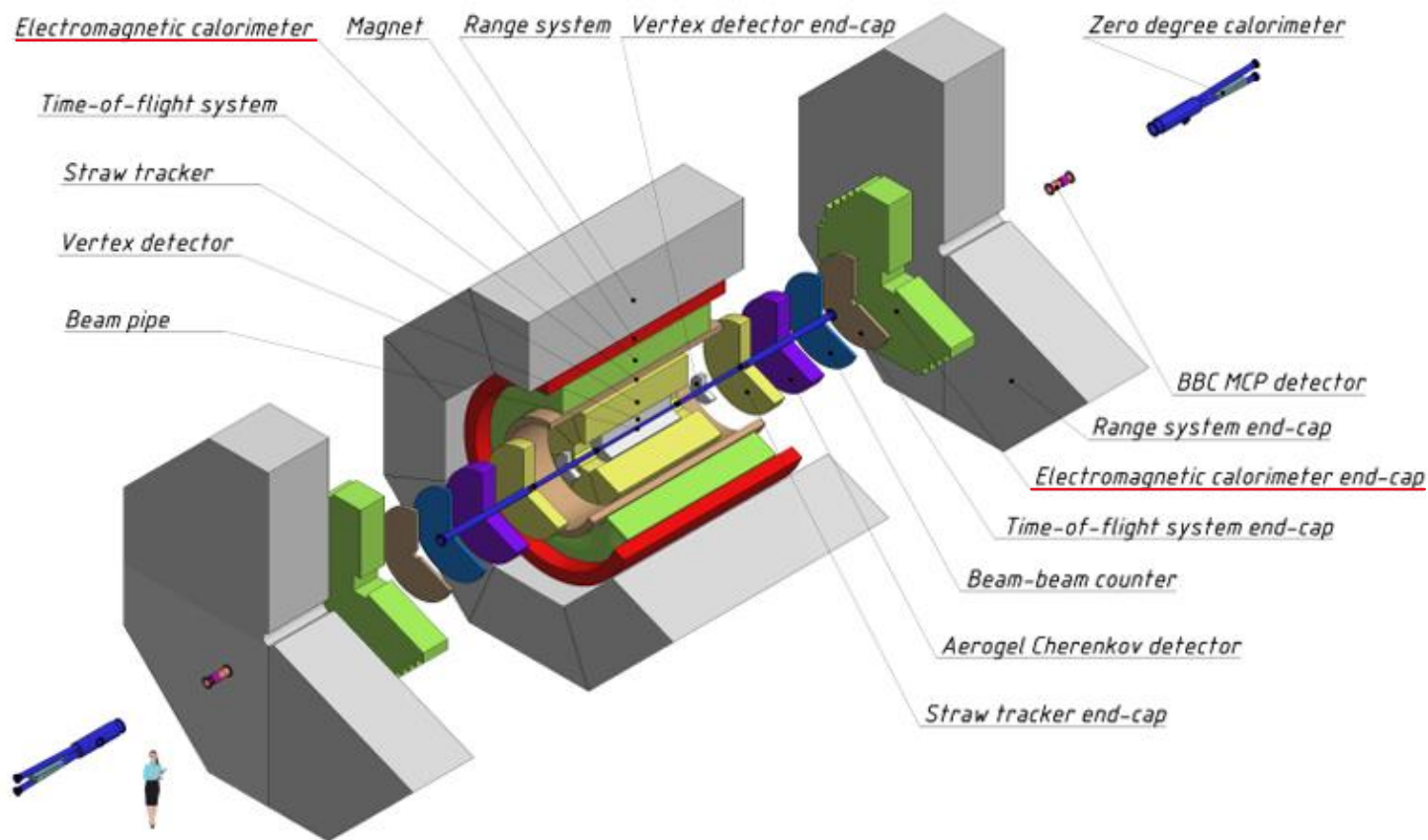
# Разработка алгоритма идентификации фотонов в электромагнитном калориметре эксперимента SPD

Студент: Петров Г. Е. Б22-102

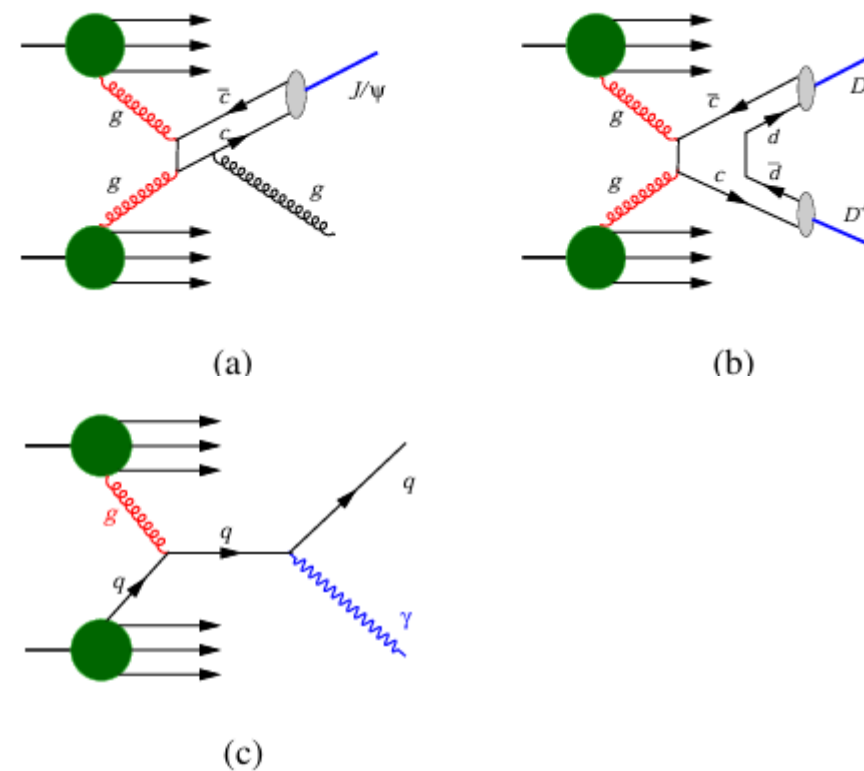
Научный руководитель: Солдатов Е. Ю.



# Об эксперименте SPD



Детекторы второй стадии эксперимента



Диаграммы, иллюстрирующие процессы: рождение (a) чармония, (b) открытого чарма и (c) прямых фотонов.

# Задача разделения фотонных и $\pi^0 \rightarrow \gamma + \gamma$ событий

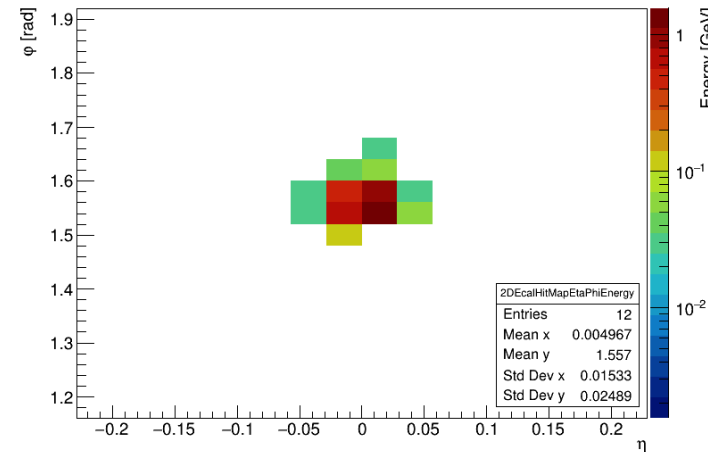
## Постановка задачи:

В связи с физикой распада фотоны из распада  $\pi^0$  могут попадать в один кластер из-за малого угла разлета, поэтому существует 2 принципиальных случая:

- 1. оба фотона от  $\pi^0$  попали в один кластер**  
– тогда такой кластер имеет специфическую форму
- 2. Фотоны от  $\pi^0$  попали в разные кластеры**  
– такие кластеры можно идентифицировать из кинематики

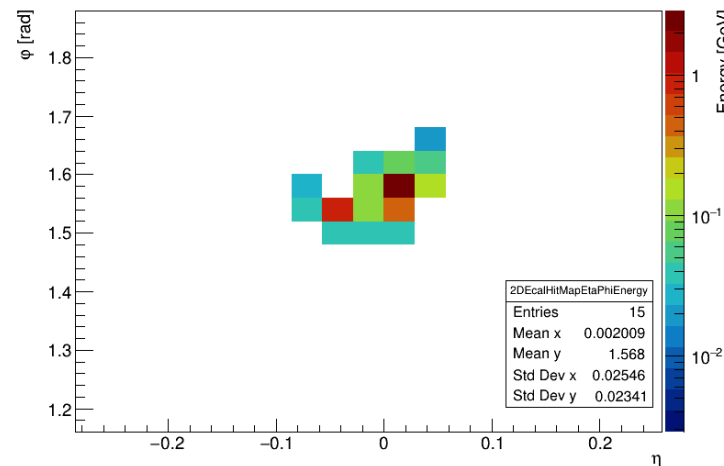
В данной работе **в качестве фона рассмотрены кластеры 1го типа** – именно такие кластеры отделялись от «фотонных кандидатов»: кластеров 2го типа и кластеров от прямых фотонов.

ECAL Energy Hit Map ( $\eta$  vs  $\phi$ )



Пример  
кластера  $\gamma$  (5 ГэВ)

ECAL Energy Hit Map ( $\eta$  vs  $\phi$ )



Пример  
кластера  $\pi^0$  (5 ГэВ)

# Задача разделения фотонных и $\pi^0 \rightarrow \gamma + \gamma$ событий

## Подходы к решению задачи сепарации $\gamma / \pi^0$

Существуют два основных класса методов:

- **Классический подход:**
  - фиксированные отборы по физическим переменным
- **ML-подход:**
  - модели машинного обучения, обучающиеся форме ливня
  - (MLP, CNN, GNN, BDT и др.)

## Предыдущие исследования:

- ранее изучался **MLP-подход** для данной задачи
- получена **90%  $\pi^0$  режекция при 80%  $\gamma$  эффективности** в сценарии одиночных запусков  $\pi^0$  и  $\gamma$  равномерно по телесному углу

## Цель данной работы:

- Расширенное исследование данной задачи: исследование более широкого круга параметров и альтернативных методов (**BDT** – boosted decision trees и **cut-based** – классические отборы)

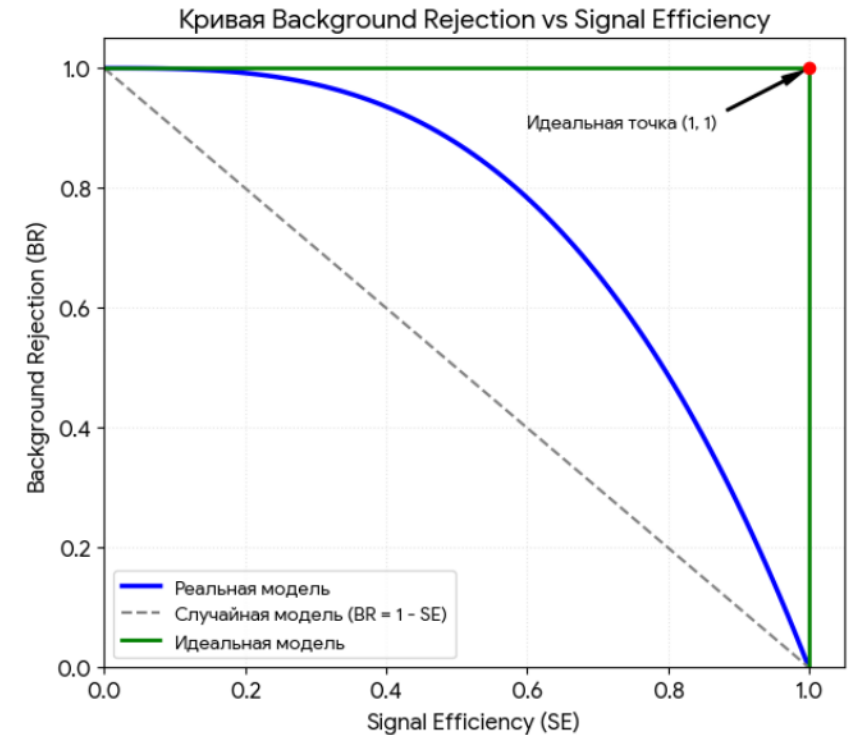
## Ключевая задача:

- разработать **универсальный инструмент**, позволяющий:
  - применять **как ML**, так и **классические методы**

# Задача разделения фотонных и $\pi^0 \rightarrow \gamma\gamma$ событий

## Методология

- False Positive Rate (FPR) – доля фона, принятого за сигнал среди всего фона (соответственно режекция фона – доля правильно отброшенного фона среди всего фона)
- True Positive Rate (TPR) – доля правильно отобранного сигнала среди всего сигнала
- Основная **метрика** оценки качества модели, которая использовалась в данной работе: **площадь под ROC кривой** – режекции фона (1 - FPR) от эффективности сигнала (TPR)

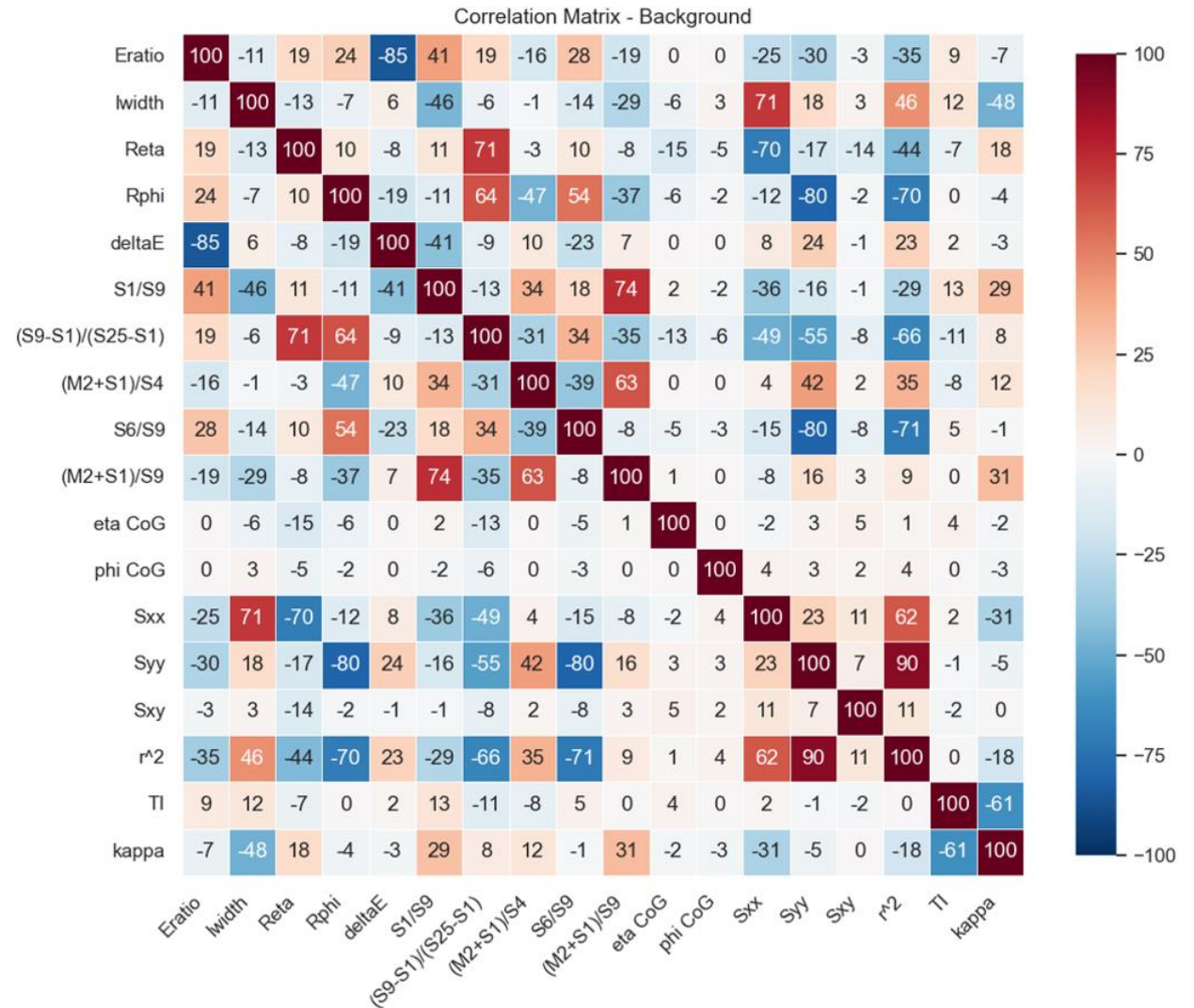


- Примеры ROC-кривых

# Адаптированные параметры

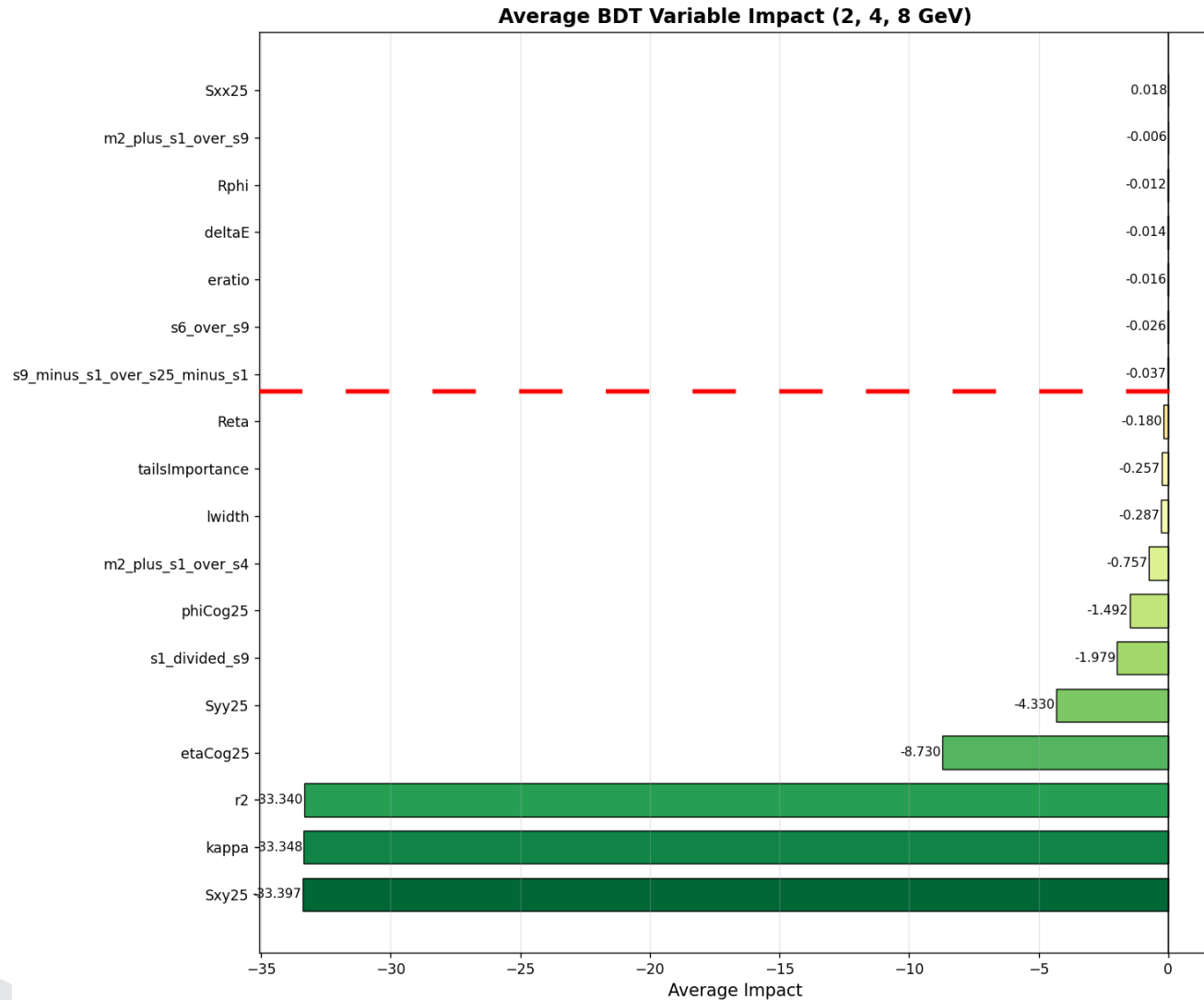
**Используемые параметры кластеров (комбинация переменны, применяемых в ATLAS, и переменных из предыдущего исследования решения задачи MLP-подходом)**

- Пространственное распределение энергии в кластере
- Поперечная ширина
- Асимметрия энергосвыделения и наличие нескольких локальных максимумов
- Положение центра тяжести кластера
- Вторые моменты и ковариации распределения энергии
- Энергетические отношения в различных областях кластера
- Параметры компактности, вытянутости и «хвостов» ливня



- Корреляционная матрица для фона ( $\pi_0$ ) (фиксированная энергия и направление)

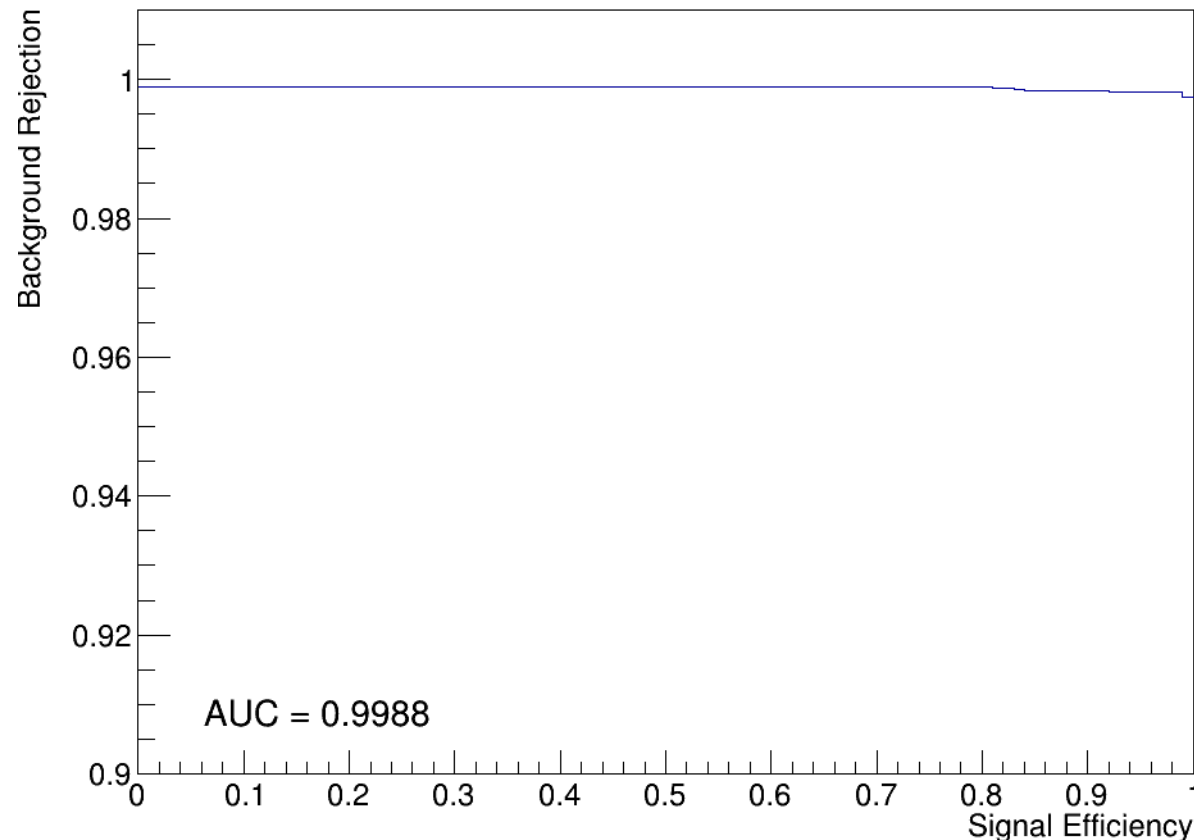
# Оценка влияния переменных на качество классификации



- Для классификатора фиксированных отборов отобраны переменные: kappa, Rphi, Reta, lwidth, deltaE, s6\_over\_s9
- Для BDT классификатора: были отброшены Eratio, deltaE, Rphi,  $(m2+s1)/s4$ , Sxx25, s6/s9,  $(s9-s1)/(s25-s1)$

# Результаты BDTG (~15000 гамма и $\pi^0$ перпендикулярно баррелю ( $p_x=p_z=0$ ): равномерное распределение энергии 1-10GeV)

Background Rejection vs Signal Efficiency



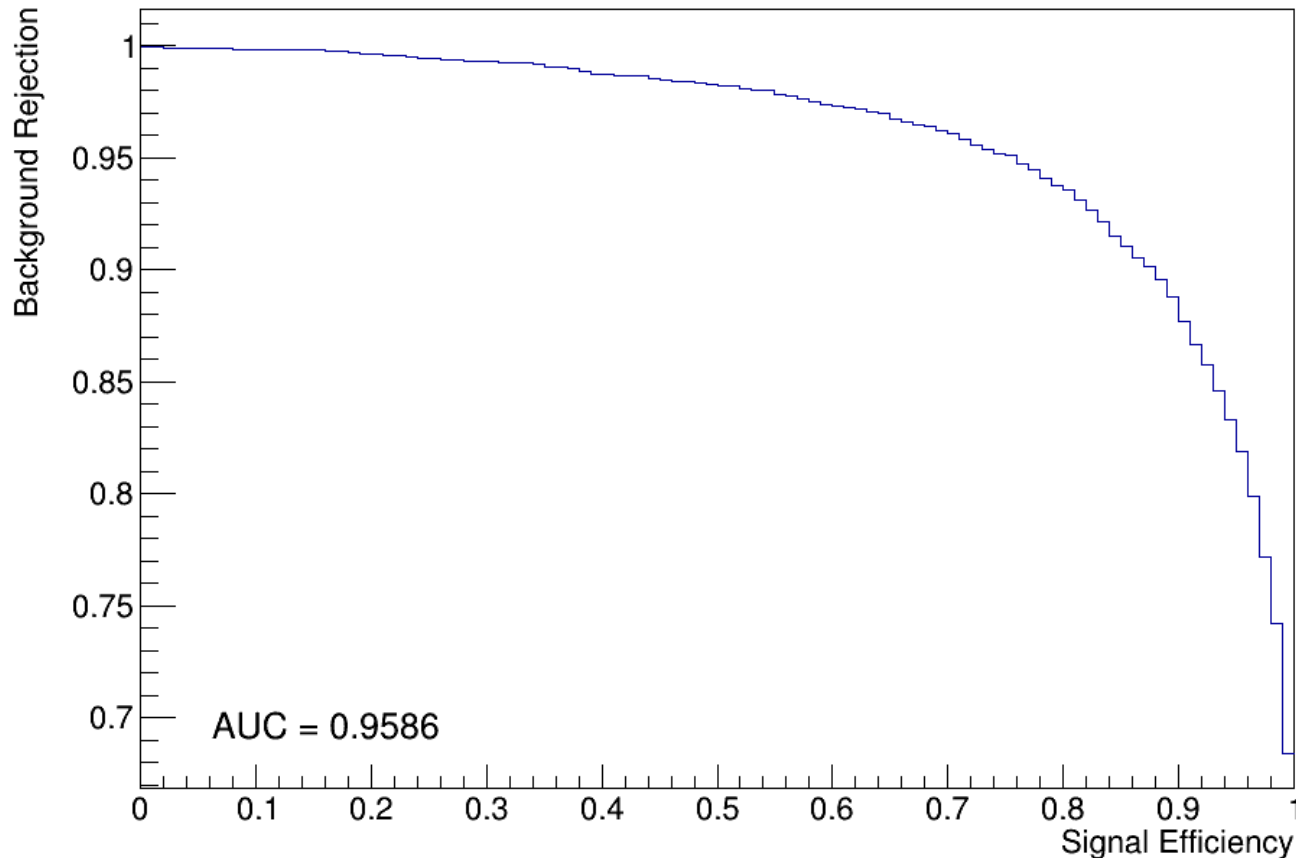
- Фон: кластеры вида [ $\pi^0$ , gamma, gamma] (train: 7457 кластеров)
- Сигнал: кластеры вида [gamma]. (train: 13867 кластеров)

Параметры лучшего BDT классификатора:

- Ntrees = 800
- MaxDepth = 5
- nCuts = 40

# Результаты BDTG (~18000 гамма и pi0: равномерное распределение по телесному углу, равномерное распределение по Pt [0, 10] GeV)

Background Rejection vs Signal Efficiency

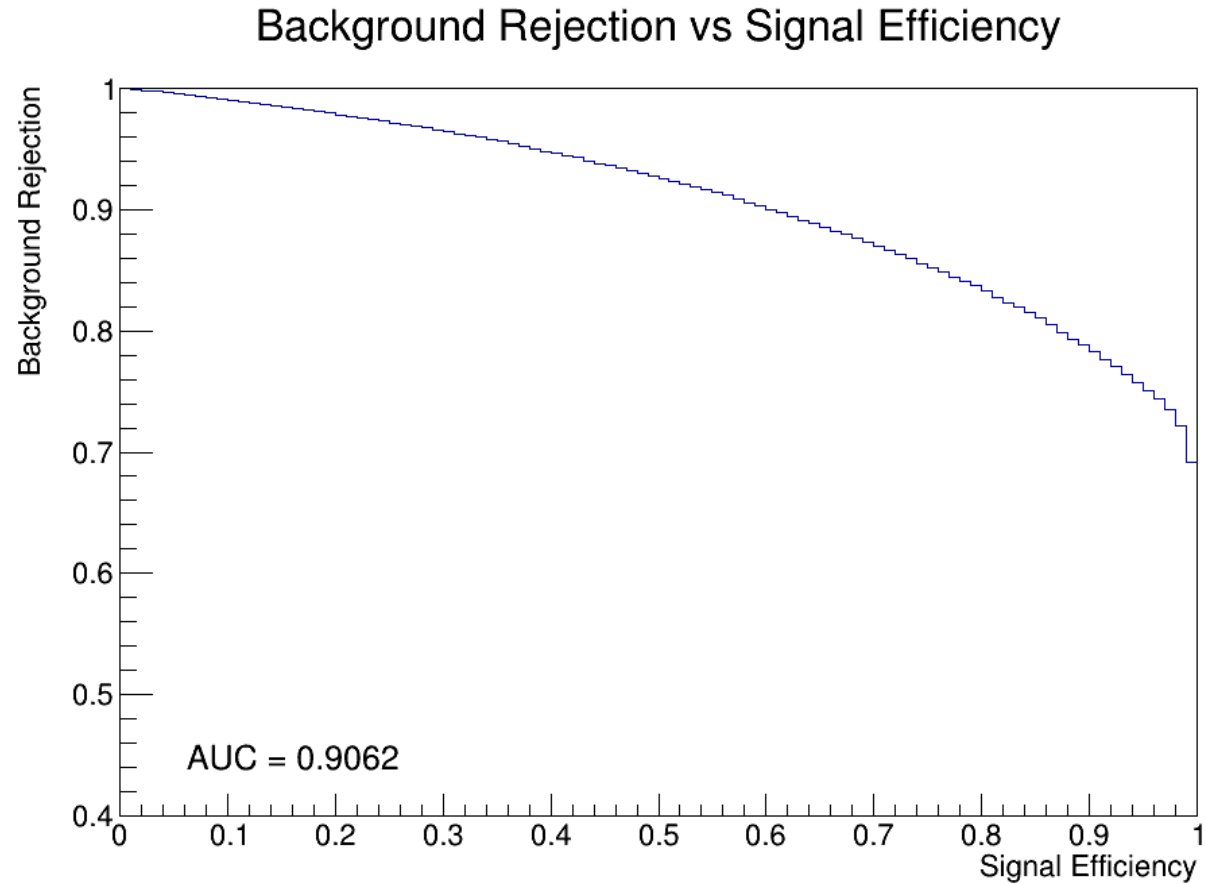


- Фон: кластеры вида [pi0, gamma, gamma] (train: 7149 кластеров)
- Сигнал: кластеры вида [gamma]. (train: 10139 кластеров)
- Фильтр: энергия кластера > 1 GeV

Параметры лучшего BDT классификатора:

- Ntrees = 800
- MaxDepth = 5
- nCuts = 40

# Результаты BDTG (minbias: 4000000 событий, pp-столкновения, 27 GeV)

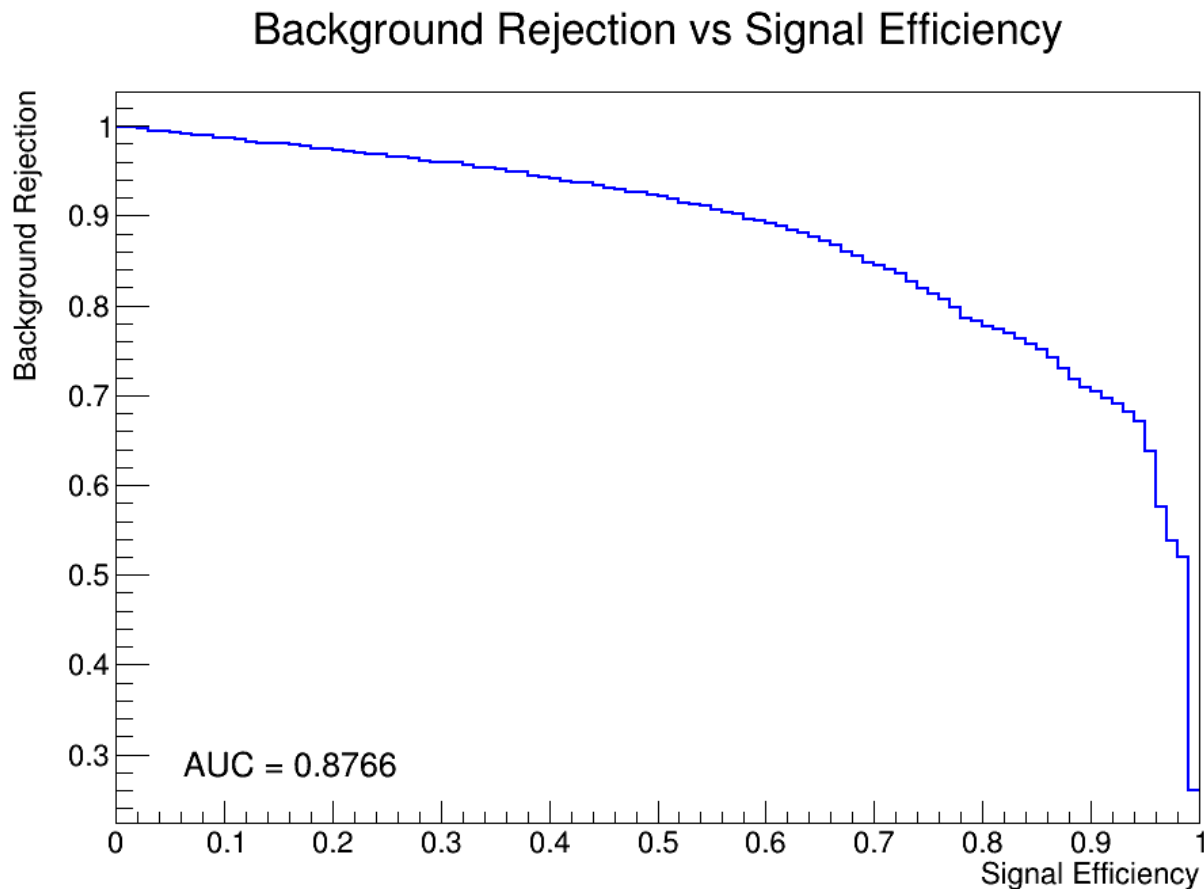


- Фон: кластеры вида [pi0, gamma, gamma] (train: 119319 кластеров)
- Сигнал: кластеры вида [gamma] (train: 29746 кластеров)
- Фильтр: энергия кластера > 1 GeV

Параметры лучшего BDT классификатора:

- Ntrees = 800
- MaxDepth = 5
- nCuts = 40

# Фиксированные отборы - GA (minbias: 4000000 событий, pp-столкновения, 27 GeV)



- Фон: кластеры вида [pi0, gamma, gamma] (train: 119319 кластеров)
- Сигнал: кластеры вида [gamma] (train: 29746 кластеров)
- Фильтр: энергия кластера > 1 GeV

Параметр	Отбор
$lwidth$	$> -2.0 \times 10^{-4}$
$R_\eta$	$> 0.137$
$R_\phi$	$> 0.771$
$\Delta E$	$< 0.100$
$s6/s9$	$> -3.5 \times 10^{-3}$
$\kappa$	$< 1.007$

Отборы при 80% эффективности сигнала

# Сравнение методов

Метрика	MLP	BDT	Cut-based
$\pi^0$ режекция при 80% $\gamma$ эффективности (равномерное распределение по телесному углу, разные энергии)	90%	94%	79%
ROC AUC: равномерное распределение по телесному углу, разные энергии	-	0.958	0.878
ROC AUC: Minbias, Ecluster>1GeV	-	0.906	0.877

- Сценарий разделения кластеров вида: [pi0, gamma, gamma] ( $\pi^0$ ) и [gamma] ( $\gamma$ )

# Заключение



## Основные результаты

- Адаптированы переменные электромагнитного ливня, применяемые в эксперименте ATLAS, к геометрии калориметра SPD и реализованы в анализе
- Проведен анализ расширенного набора переменных, исключены переменные, имеющие наименьшее влияние на метрики
- Выполнено сравнение методов классификации кластеров  $\gamma/\pi^0$  в различных сценариях: от фиксированных энергий и направления до разыгранных **minbias** событий
- На **minbias** достигнуто **лучшее значение ROC AUC 0.906**, полученное на основе BDT классификатора

## Дальнейшие планы

- Оценка возможности использования множества классификаторов в различных пространственных и энергетических областях для увеличения качества идентификации
- Разработка алгоритмов идентификации других типов кластеров и других частиц
- Интеграция алгоритмов в основную цепочку реконструкции SPDROOT/SAMPO
- Исследование альтернативных ML-подходов для дальнейшего повышения качества фотонной идентификации

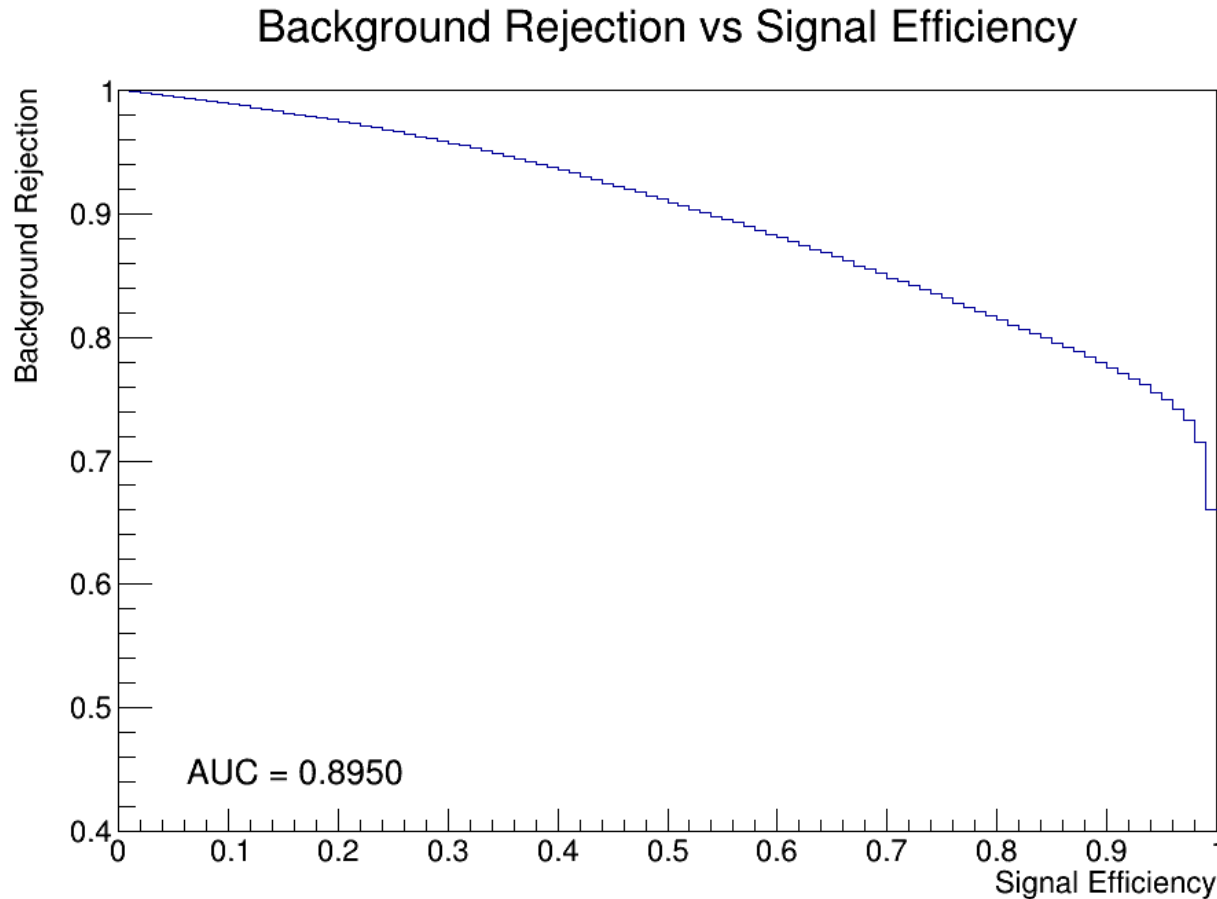


**МИФИ**

Национальный  
исследовательский  
ядерный университет

**Спасибо за  
внимание**

# Результаты BDTG (minbias: 4000000 событий, pp-столкновения, 27 GeV)



- Фон: кластеры вида [pi0, gamma, gamma] (train: 119319 кластеров)
- Сигнал (фотонные кандидаты): кластеры вида [gamma], [pi0, gamma] (train: 1169614 кластеров)
- Фильтр: энергия кластера > 1 GeV

Параметры лучшего BDT классификатора:

- Ntrees = 800
- MaxDepth = 5
- nCuts = 40

# Описание всех адаптированных параметров (из эксперимента ATLAS)

- $w_{\eta 2}$ : поперечная ширина электромагнитного ливня в ячейках кластера

$$w_{\eta 2} = \sqrt{\frac{\sum_i E_i \eta_i^2}{\sum_i E_i} - \left(\frac{\sum_i E_i \eta_i}{\sum_i E_i}\right)^2}$$

сумма по окну  $3 \times 5$  ячеек по  $(\eta, \phi)$ ;

- $R_\phi$ : отношение энергии, выделенной в окне  $3 \times 3$  ячеек, к энергии в окне  $3 \times 7$  ячеек по  $(\eta, \phi)$ ;
- $E_{\text{ratio}}$ : отношение

$$E_{\text{ratio}} = \frac{E_{\text{max1}} - E_{\text{max2}}}{E_{\text{max1}} + E_{\text{max2}}},$$

где  $E_{\text{max1}}$  и  $E_{\text{max2}}$  — наибольшее и второе по величине энерговыделения в ячейках калориметра;

- $\Delta E$ : разность между энергией второго локального максимума в ячейках кластера и минимальной энергией, реконструированной в прямоугольнике между первым и вторым максимумами. В случае отсутствия второго максимума параметр принимает значение 0.
- $R_\eta$ : отношение энергии, выделенной в окне  $3 \times 7$  ячеек, к энергии в окне  $7 \times 7$  ячеек по  $(\eta, \phi)$ ;

# Описание всех адаптированных параметров

- Смещение центра тяжести относительно максимума  $\eta/\varphi$ :

$$x_{\text{cog}} = \frac{1}{S_{25}} \sum_{i=1}^{25} E_i X_i^{\text{rel}}, \quad y_{\text{cog}} = \frac{1}{S_{25}} \sum_{i=1}^{25} E_i Y_i^{\text{rel}},$$

где  $S_{25} = \sum_{i=1}^{25} E_i$ ,  $X_i^{\text{rel}}$ ,  $Y_i^{\text{rel}}$  — координаты  $i$ -го хита относительно максимума.

- Вторые моменты:

$$\langle r^2 \rangle = S_{XX} + S_{YY} = \frac{\sum_{i=1}^N e_i [(x_i - x_c)^2 + (y_i - y_c)^2]}{\sum_{i=1}^N e_i},$$

где  $N=25$ ,  $e_i$  — энергия  $i$ -го хита,  $(x_c, y_c)$  — координаты центра тяжести кластера ( $\eta_c = \frac{\sum_{i=1}^N E_i \eta_i}{\sum_{i=1}^N E_i}$ ,  $\phi_c = \arctan\left(\frac{\sum_{i=1}^N E_i \sin \phi_i}{\sum_{i=1}^N E_i \cos \phi_i}\right)$ ).

$$S_{XX} = \frac{\sum_{i=1}^N e_i (x_i - x_c)^2}{\sum_{i=1}^N e_i}, \quad S_{YY} = \frac{\sum_{i=1}^N e_i (y_i - y_c)^2}{\sum_{i=1}^N e_i},$$

$$S_{XY} = S_{YX} = \frac{\sum_{i=1}^N e_i (x_i - x_c)(y_i - y_c)}{\sum_{i=1}^N e_i}.$$

где  $N=25$ ,  $e_i$  — энергия  $i$ -го хита,  $(x_c, y_c)$  — координаты центра тяжести кластера.

- Параметр формы кластера:

$$\kappa = \sqrt{1 - \frac{S_{XX}S_{YY} - S_{XY}^2}{(S_{XX} + S_{YY})^2}} = \sqrt{1 - \frac{4 \det S}{\text{Tr}(S)^2}},$$

где  $\det S = S_{XX}S_{YY} - S_{XY}^2$ ,  $\text{Tr}(S) = S_{XX} + S_{YY}$ .

- Также на распределениях с нефиксированной энергией и направлением использовались энергия кластера и координаты его центра в качестве переменных BDT.

- Энергетические отношения:

$$\frac{S_1}{S_9}, \quad \frac{S_9 - S_1}{S_{25} - S_1}, \quad \frac{M_2 + S_1}{S_4}, \quad \frac{S_6}{S_9}, \quad \frac{M_2 + S_1}{S_9}.$$

где  $S_1$  — максимальная энергия в ячейке кластера,  $M_2$  — вторая по величине энергия в ячейке кластера,  $S_4, S_6, S_9, S_{25}$  — суммы энергий в окнах  $2 \times 2, 3 \times 2, 3 \times 3, 5 \times 5$  ячеек соответственно по  $(\eta, \phi)$ .

- Важность хвостов (Tail Importance):

$$\text{TI} = \frac{\langle r^4 \rangle}{\langle r^2 \rangle^2} = 1 - \frac{\langle r^2 \rangle^2}{\langle r^4 \rangle}.$$

где

$$\langle r^4 \rangle = \frac{\sum_{i=1}^N e_i [(x_i - x_c)^2 + (y_i - y_c)^2]^2}{\sum_{i=1}^N e_i}.$$

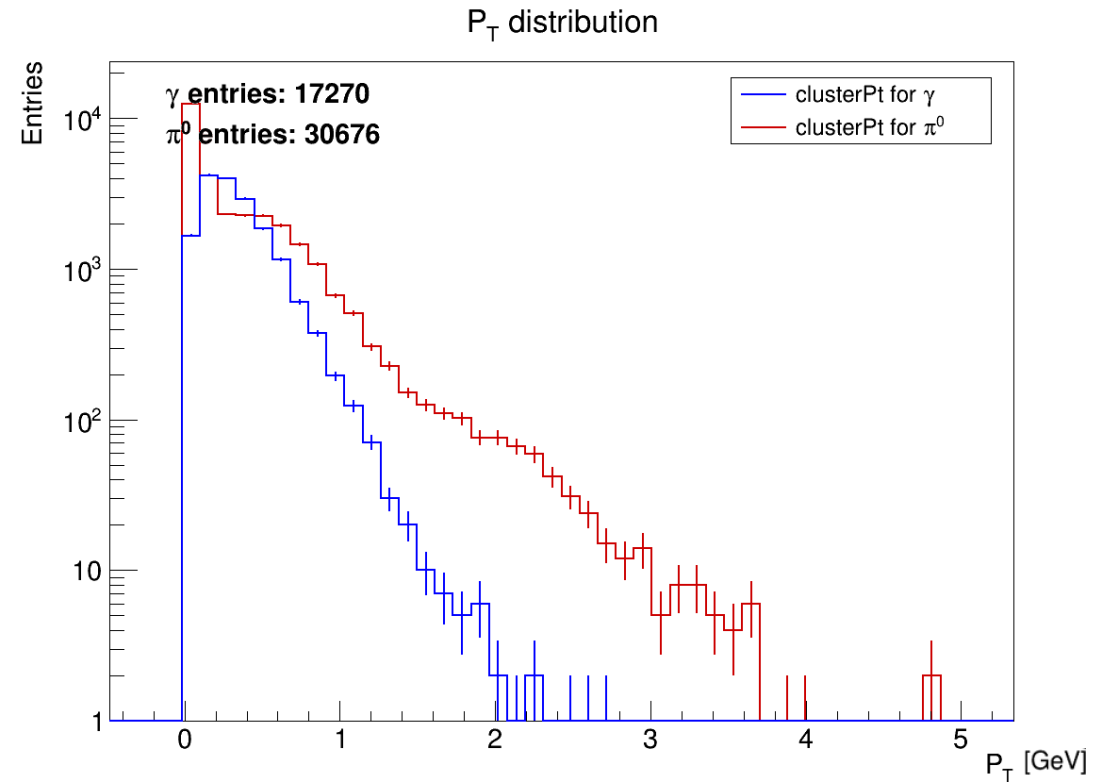
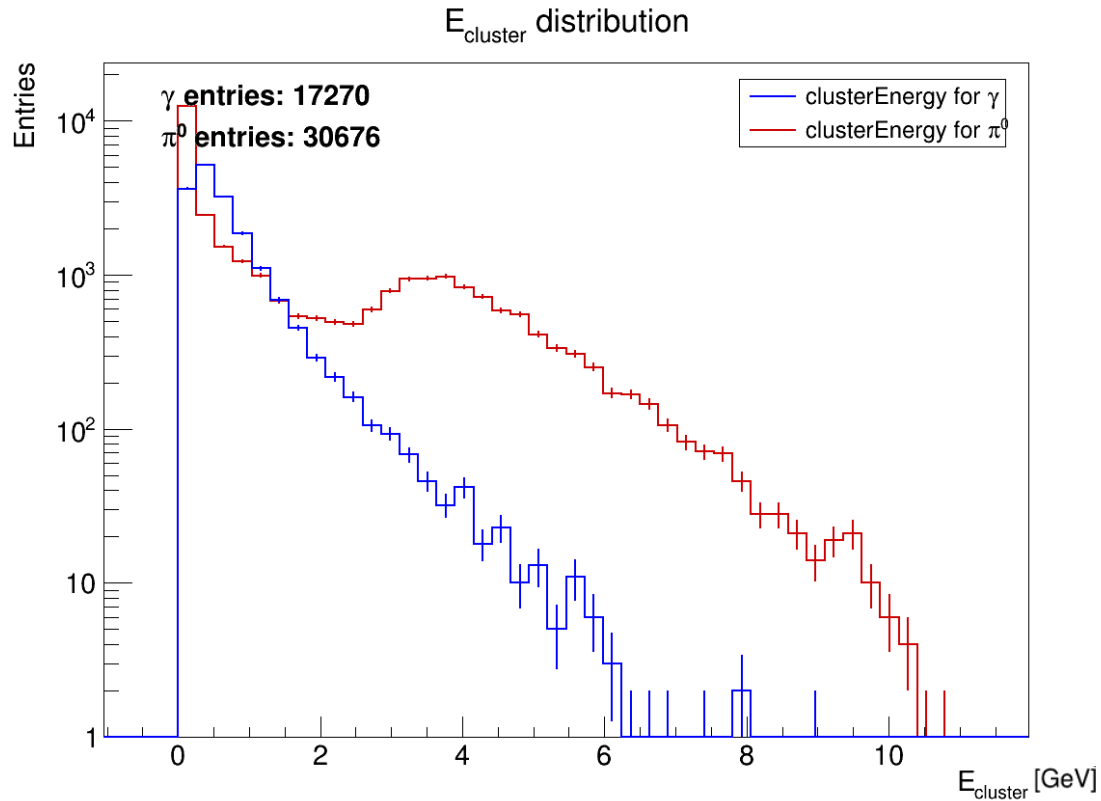
# minbias: 400000 событий (самые частые кластеры)

```
=== PDG Cluster Particles Combination Counts (Top 20) ===
 1. [gamma, pi0] (22,111): 1304565 entries
 2. [e-, gamma, pi0] (11,22,111): 56583 entries
 3. [e+, gamma, pi0] (-11,22,111): 52465 entries
 4. [gamma, gamma, pi0] (22,22,111): 31151 entries
 5. [e+, e-, gamma, pi0] (-11,11,22,111): 22914 entries
 6. [gamma] (22): 16795 entries
 7. [gamma, gamma, pi0, pi0] (22,22,111,111): 15931 entries
 8. [e-, gamma, gamma, pi0] (11,22,22,111): 15237 entries
 9. [e+, e-, gamma, gamma, pi0] (-11,11,22,22,111): 14782 entries
10. [e+, gamma, gamma, pi0] (-11,22,22,111): 13862 entries
11. [e+, e-, gamma, gamma, gamma, pi0] (-11,11,22,22,22,111): 8049 entries
12. [e-, gamma, gamma, gamma, pi0] (11,22,22,22,111): 6697 entries
13. [gamma, gamma, pi0, eta] (22,22,111,221): 6367 entries
14. [e+, gamma, gamma, gamma, pi0] (-11,22,22,22,111): 6254 entries
15. [gamma, gamma, gamma, pi0] (22,22,22,111): 5613 entries
16. [e+, e-, gamma, gamma, gamma, gamma, pi0] (-11,11,22,22,22,22,111): 4040 entries
17. [e-, gamma, gamma, gamma, gamma, pi0] (11,22,22,22,22,111): 2667 entries
18. [e+, gamma, gamma, gamma, gamma, pi0] (-11,22,22,22,22,111): 2558 entries
19. [e-, pi0] (11,111): 2351 entries
20. [e+, pi0] (-11,111): 2335 entries
```

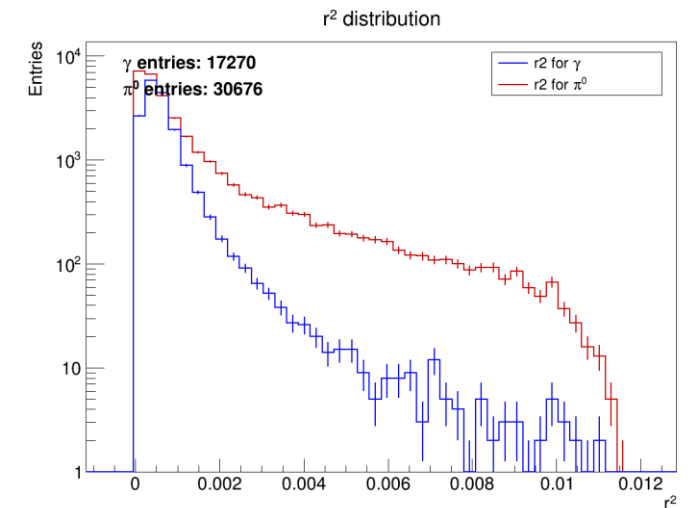
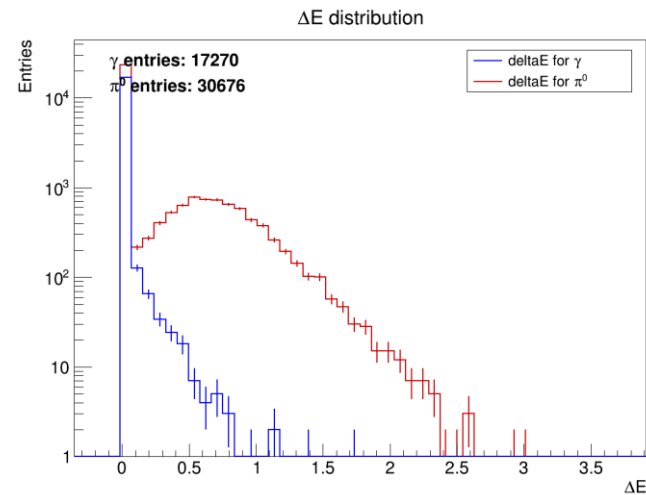
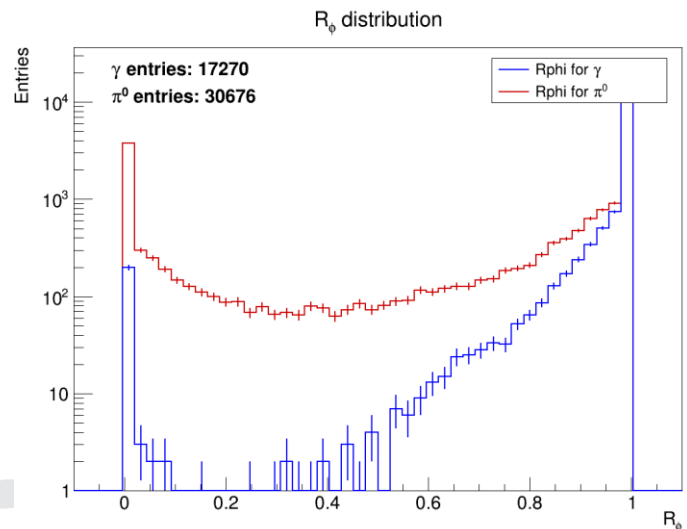
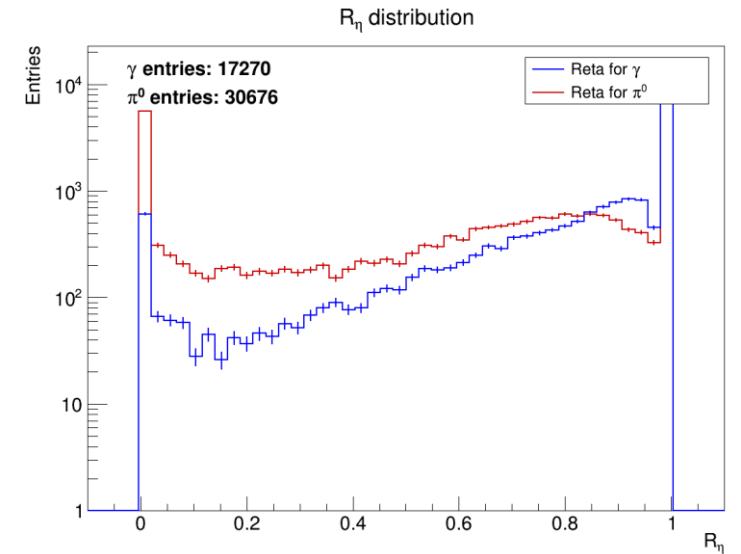
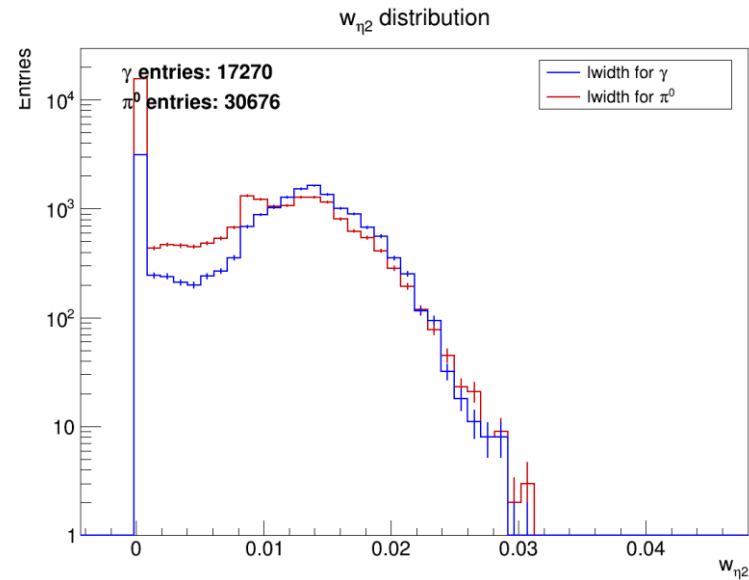
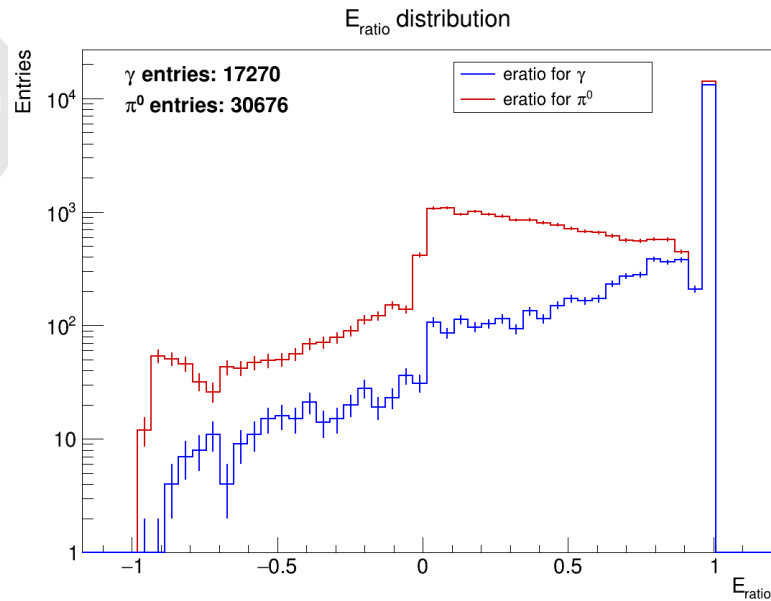
# ~15000 событий $\gamma$ и $\pi^0$ перпендикулярно баррелю (самые частые кластеры)

```
=== PDG Cluster Particles Combination Counts (Top 20) ===
 1. [ $\gamma$ ] (22): 19752 entries
 2. [ $\gamma$ ,  $\gamma$ ,  $\pi^0$ ] (22,22,111): 10614 entries
 3. [ $\gamma$ ,  $\pi^0$ ] (22,111): 2873 entries
 4. [ $e^-$ ,  $\gamma$ ] (11,22): 206 entries
 5. [ $e^-$ ,  $\gamma$ ,  $\gamma$ ,  $\pi^0$ ] (11,22,22,111): 148 entries
 6. [ $\gamma$ ,  $\gamma$ ] (22,22): 73 entries
 7. [ $e^-$ ,  $\gamma$ ,  $\pi^0$ ] (11,22,111): 63 entries
 8. [ $e^+$ ,  $\gamma$ ,  $\pi^0$ ] (-11,22,111): 61 entries
 9. [ $e^+$ ,  $\gamma$ ,  $\gamma$ ,  $\pi^0$ ] (-11,22,22,111): 43 entries
10. [ $\gamma$ ,  $\gamma$ ,  $\gamma$ ,  $\pi^0$ ] (22,22,22,111): 22 entries
11. [ $e^+$ ,  $e^-$ ,  $\gamma$ ] (-11,11,22): 15 entries
12. [ $e^+$ ,  $e^-$ ,  $\gamma$ ,  $\gamma$ ,  $\pi^0$ ] (-11,11,22,22,111): 15 entries
13. [ $e^+$ ,  $\gamma$ ] (-11,22): 8 entries
14. [ $e^+$ ,  $e^-$ ,  $e^-$ ,  $\gamma$ ] (-11,11,11,22): 3 entries
15. [ $e^+$ ,  $\gamma$ ,  $\gamma$ ,  $\gamma$ ,  $\pi^0$ ] (-11,22,22,22,111): 3 entries
16. [ $e^-$ ,  $\gamma$ ,  $\gamma$ ] (11,22,22): 2 entries
17. [ $e^-$ ,  $e^-$ ,  $\gamma$ ] (11,11,22): 2 entries
18. [ $e^-$ ,  $\gamma$ ,  $\gamma$ ,  $\gamma$ ,  $\pi^0$ ] (11,22,22,22,111): 2 entries
19. [ $e^+$ ,  $e^-$ ,  $\gamma$ ,  $\pi^0$ ] (-11,11,22,111): 2 entries
20. [ $e^-$ ,  $e^-$ ,  $\gamma$ ,  $\gamma$ ,  $\pi^0$ ] (11,11,22,22,111): 1 entries
```

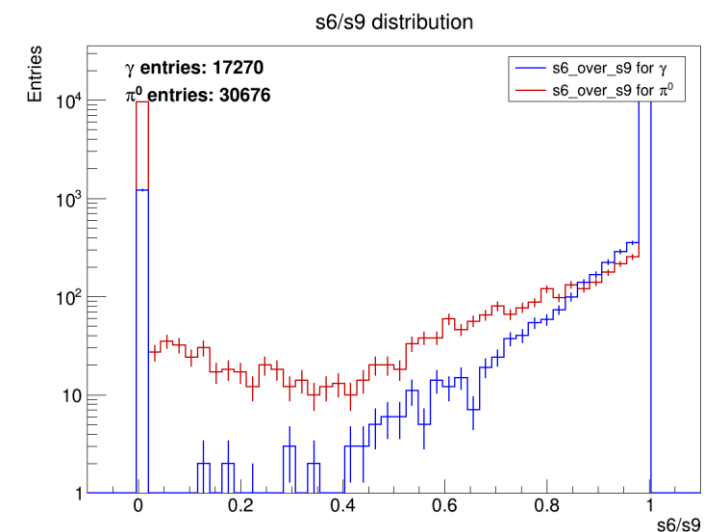
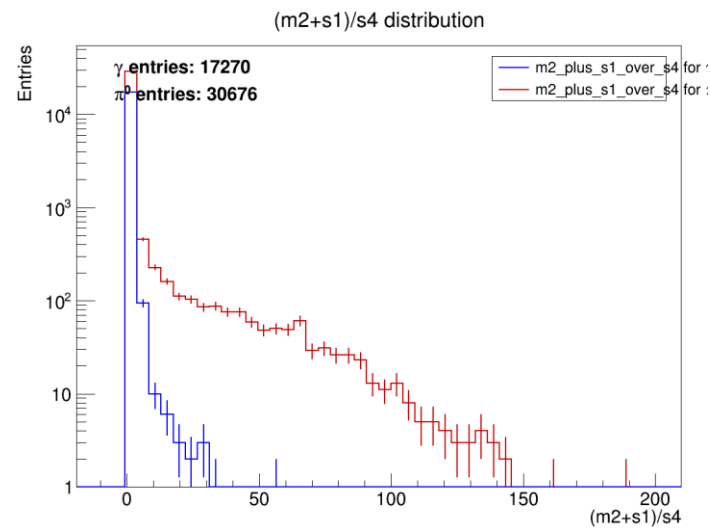
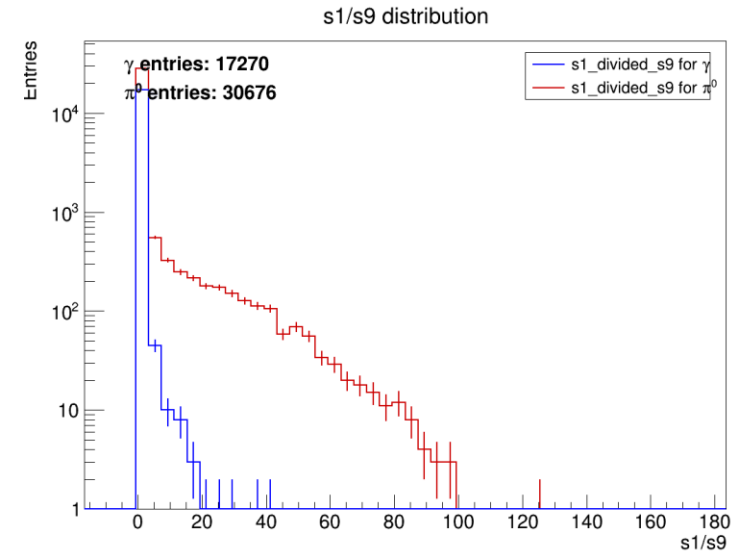
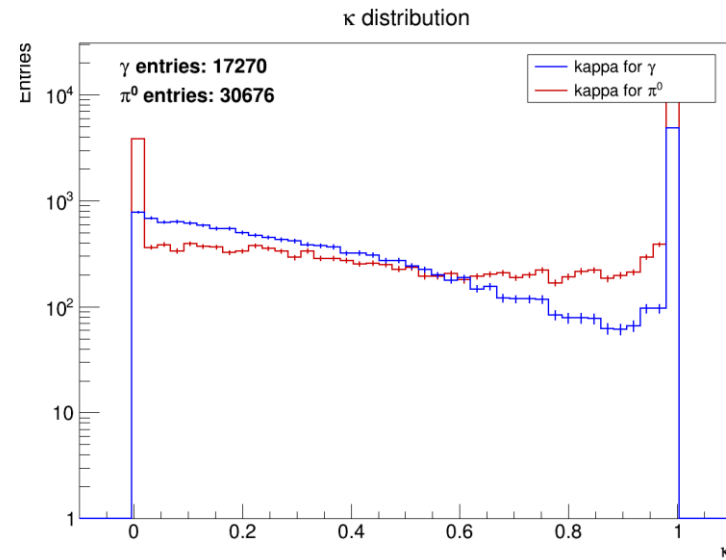
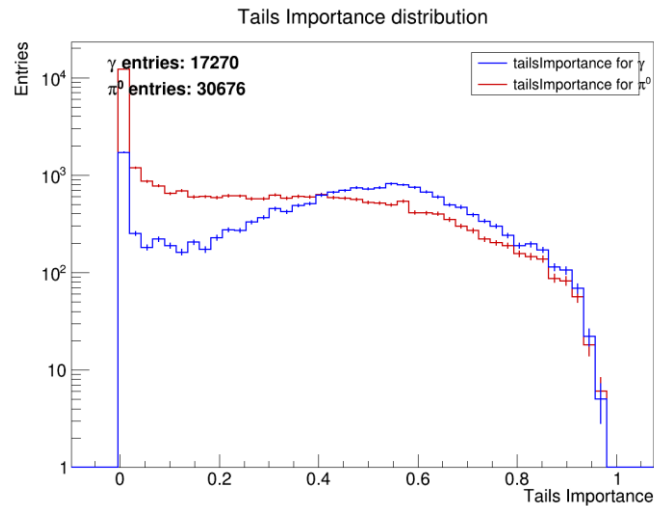
# Гистограммы minbias (400000 событий, pp-столкновения, 27 GeV): только кластеры вида [гамма], [гамма, гамма, пи0]



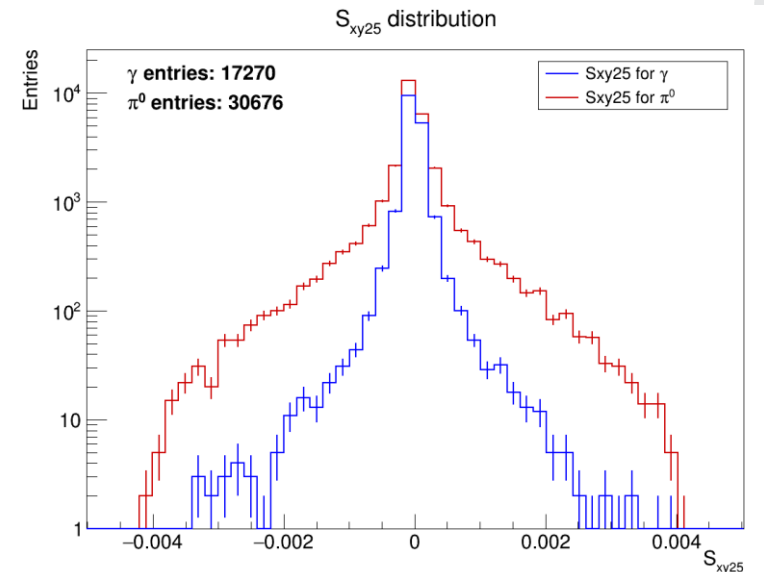
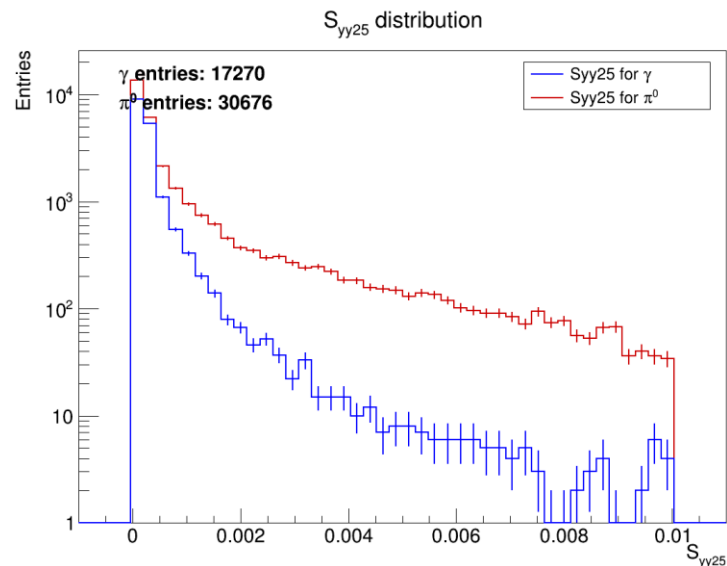
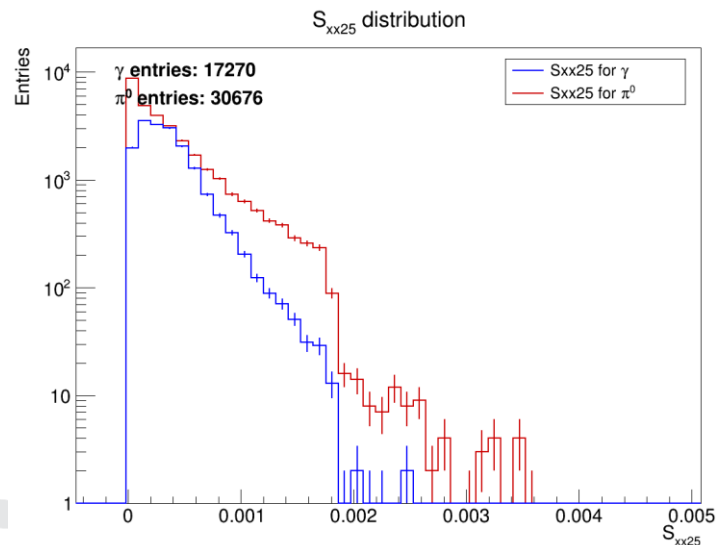
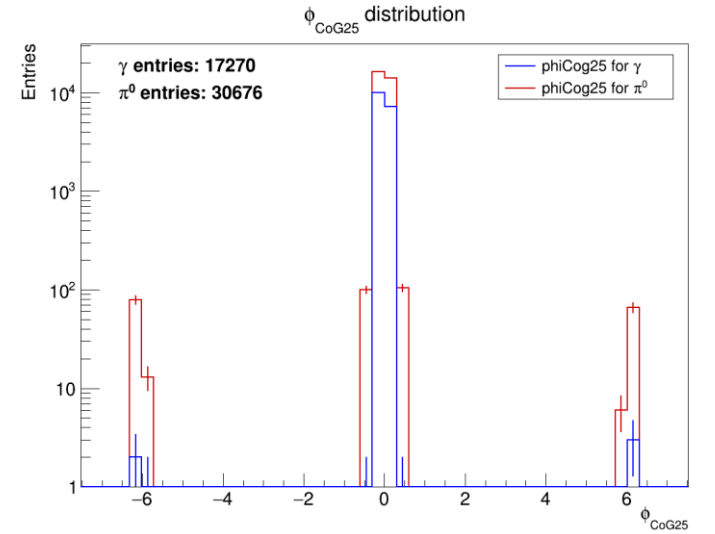
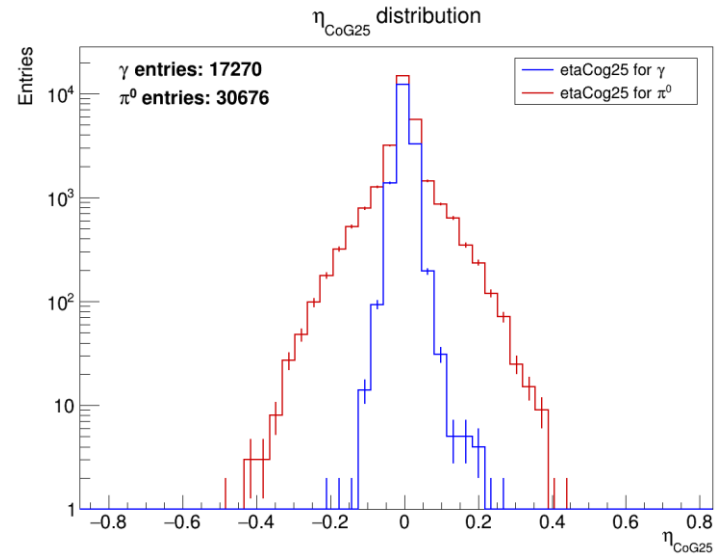
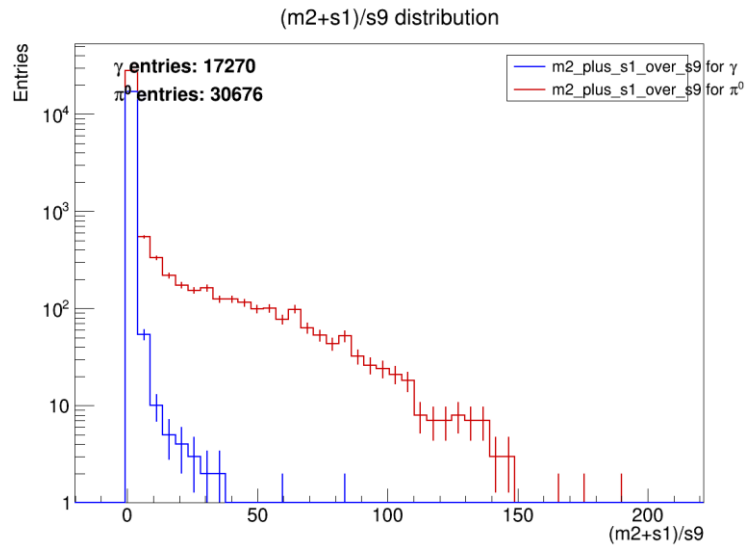
# Гистограммы minbias (400000 событий, pp-столкновения, 27 GeV): только кластеры вида [гамма], [гамма, гамма, пи0]



# Гистограммы minbias (400000 событий, pp-столкновения, 27 GeV): только кластеры вида [гамма], [гамма, гамма, пи0]



# Гистограммы minbias (400000 событий, pp-столкновения, 27 GeV): только кластеры вида [гамма], [гамма, гамма, пи0]



# Версия minbias

Production name/ID	Status	Description					Software type/version	Short description (for datasets naming)
		Stage	Collision type	Geometry type	Energy	Polarization		
PROD2025-020	Done	S2	pp	DSSD, TS, TOF, ECal, FARICH, RS, BBC, ZDC (sketch)	27 GeV	UU	spdroot-dev-4.1.7.4	minbias-P8-spdroot4174-dev