

**ЗАДАЧА ПОДАВЛЕНИЯ АДРОННОГО ФОНА ПРИ
РЕГИСТРАЦИИ ГАММА-КВАНТОВ ВЫСОКИХ
ЭНЕРГИЙ АТМОСФЕРНЫМ ИМИДЖЕВЫМ
ЧЕРЕНКОВСКИМ ТЕЛЕСКОПОМ (АЧТ) В
ЭКСПЕРИМЕНТЕ TAIGA.**

Васютина Мария Романовна

Цель работы

Целью работы является улучшения метода подавления фона адронных атмосферных ливней в эксперименте TAIGA, нацеленном на регистрацию ливней, образованных первичными гамма-квантами ТэВных энергий, приходящих из космоса.

Широкий атмосферный ливень

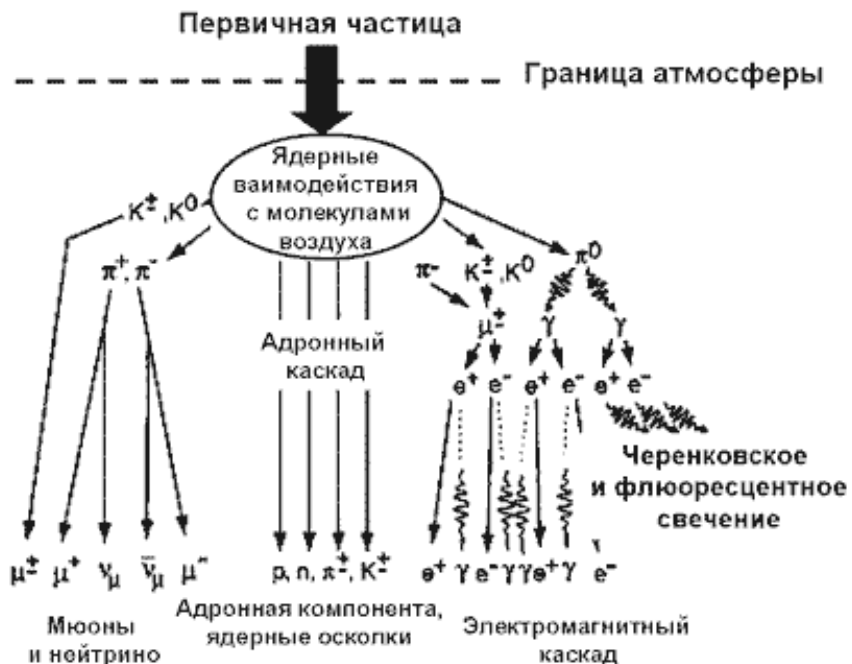


Рис. 1: Каскадный процесс взаимодействия ПКЛ с атмосферой – рождение вторичной компоненты космических лучей (ШАЛ)

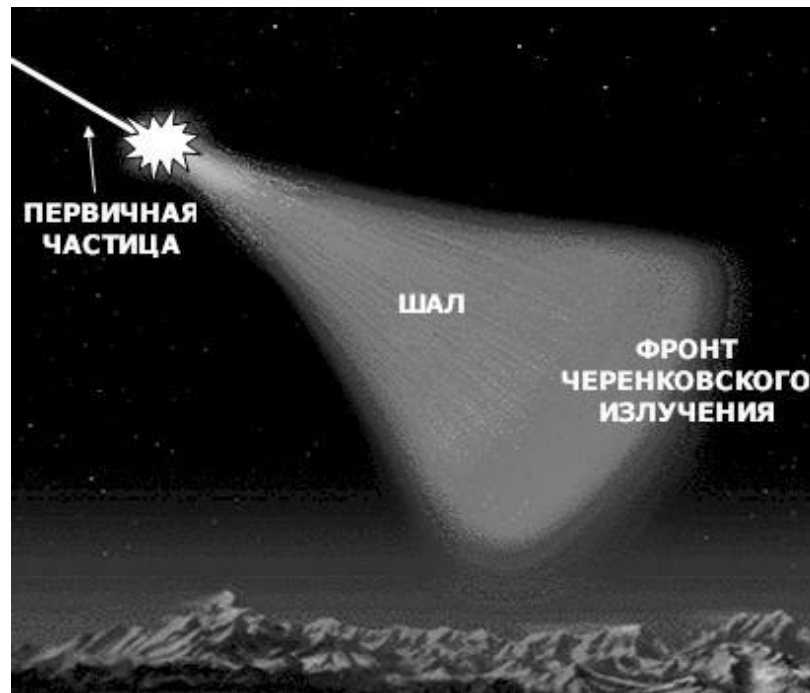


Рис. 2: Принцип регистрации космических лучей методом измерения черенковского света.

Формирование изображения

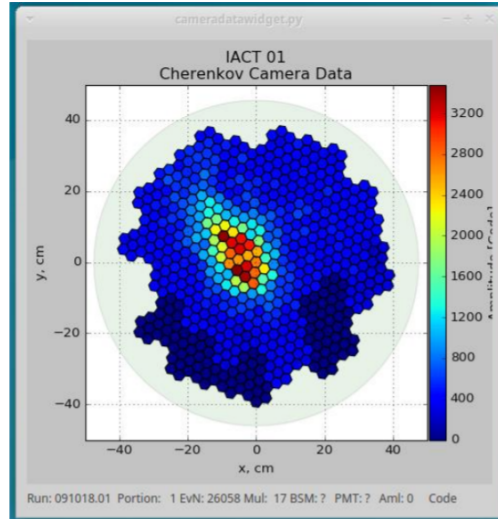


Рис. 3: Общий вид телескопа TAIGA-IACT (а) и регистрируемые изображения (б) атмосферных ливней камерой телескопа

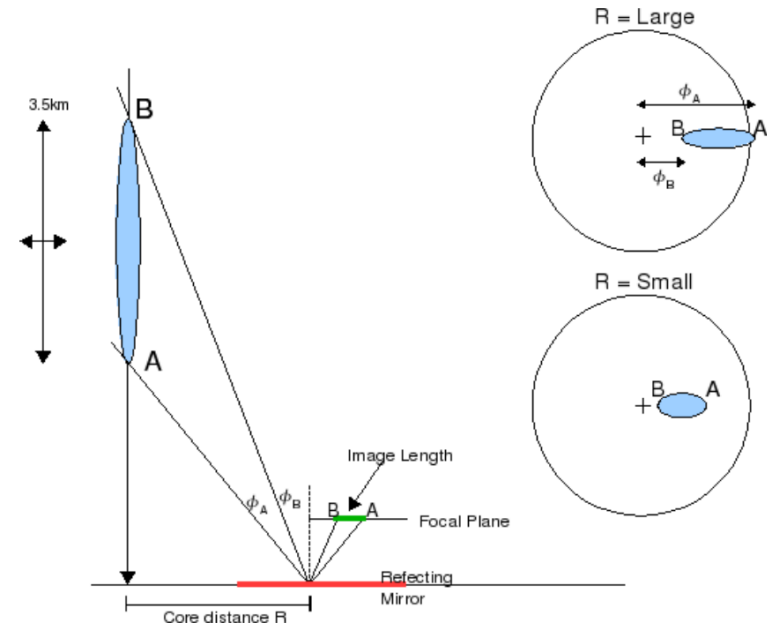


Рис. 4: Отображение ливня на плоскость камеры через отражающее зеркало, направленное на источник. Угол ϕ_B представляет собой угол между светом из верхней части ливня, точка B, и оптической осью отражающего зеркала, угол ϕ_A представляет собой угол между светом из точки A ливня и оптической осью отражающего зеркала. Изображение камеры показано для двух значений R.

Параметры Хилласа

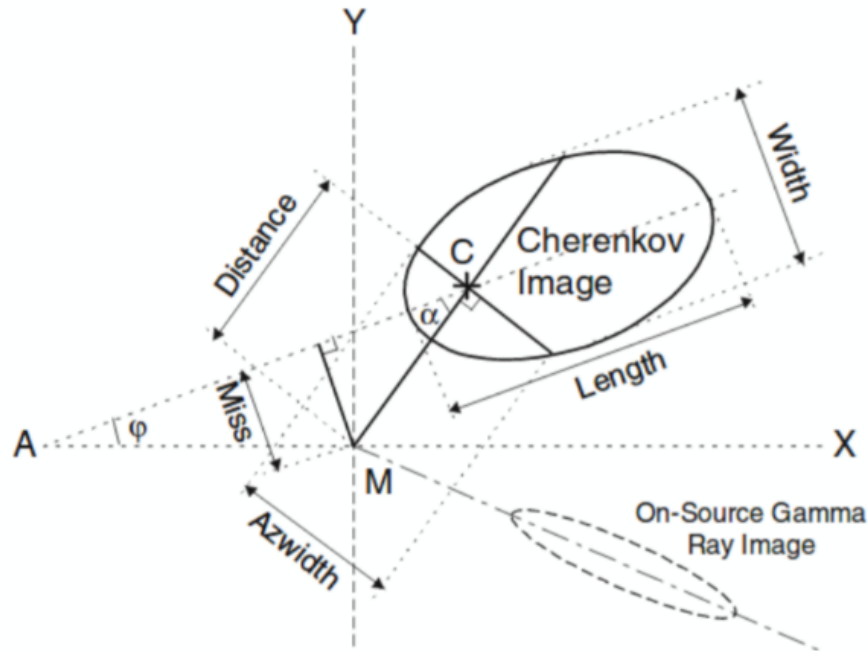


Рис. 5: Схема эллиптического изображения ШАЛ, сформированное на матрице IACT, где C — это центр тяжести эллипса.

Эти параметры включают в себя:

- информацию о форме эллипса (большая полуось эллипса — длина (Length), малая полуось - ширина (Width) и азимутальная ширина (Azwidth));
- местоположение и ориентацию эллипса относительно центра камеры (центр тяжести эллипса, расстояние от центра камеры (Distance), ошибка (Miss), угол α и угол φ);
- Параметры, характеризующие полное число фотоэлектронов в имидже и их распределение по эллипсу (размер изображения (Size), концентрация (Concentration) и коэффициент асимметрии);

Различия параметров Хилласа для разных частиц

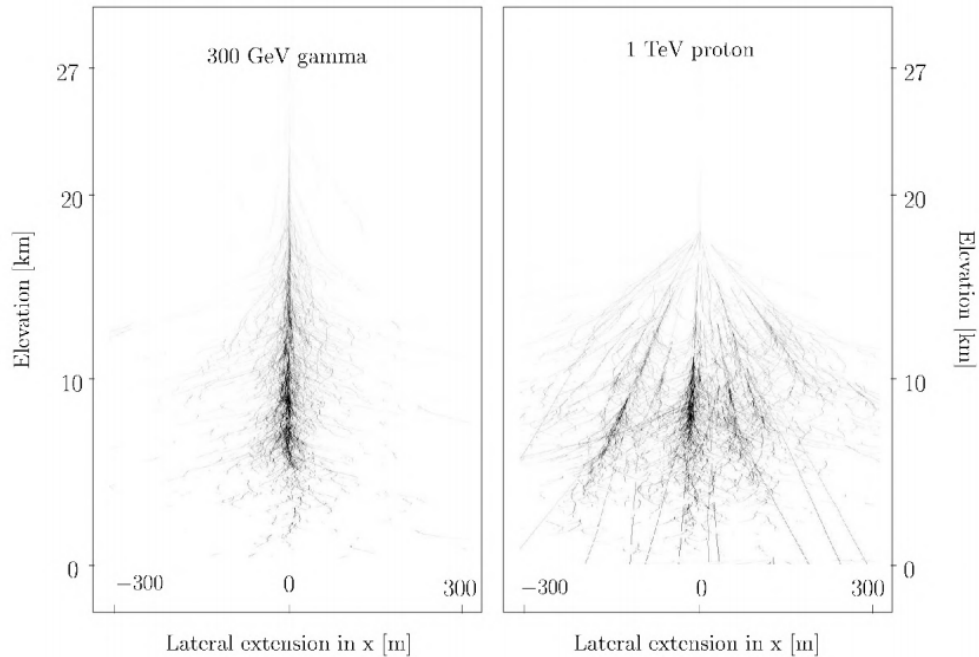


Рис. 6: Схема развития ШАЛ от первичных гамма-кванта и протона

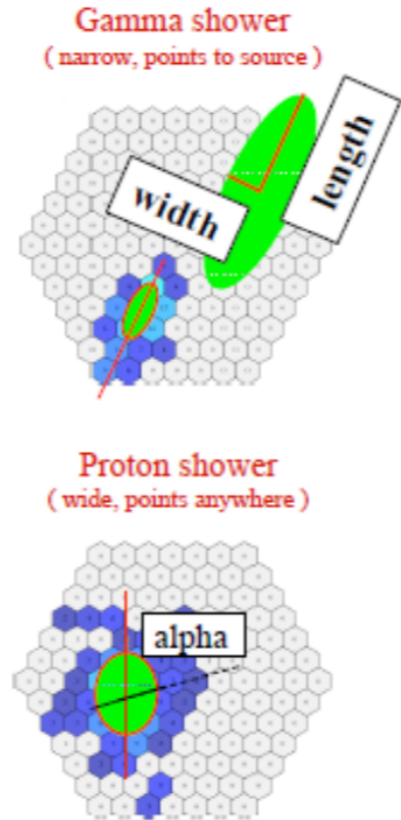


Рис. 7: Имиджи от гамма-кванта и протона

Используемые параметры

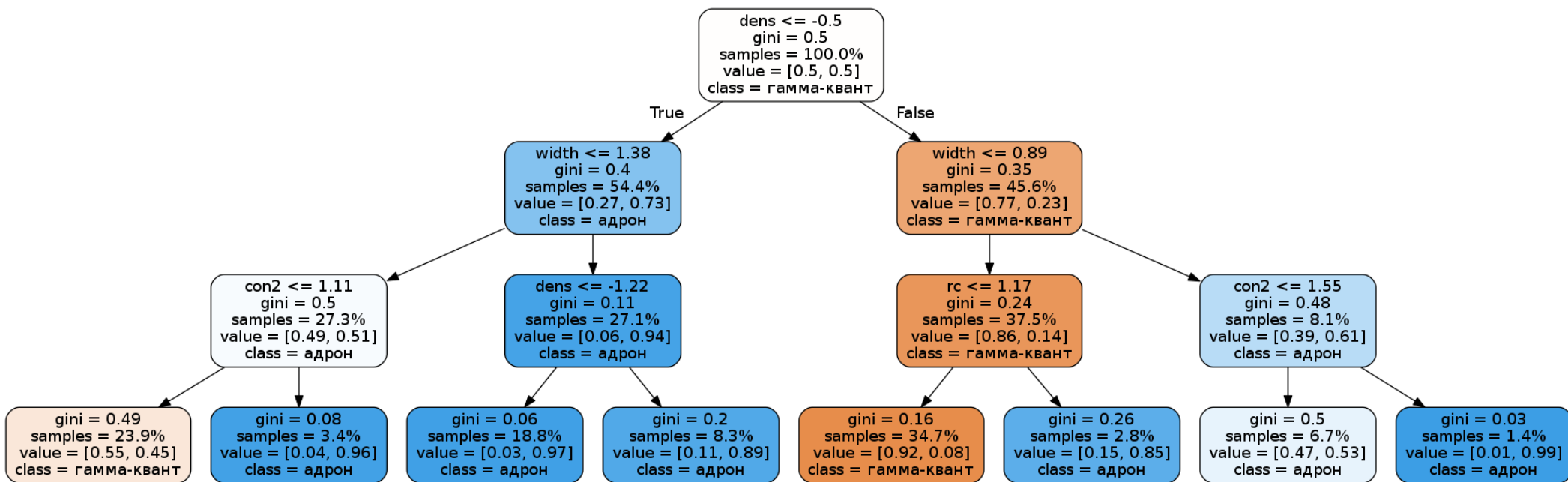
- R_c — расстояние до взвешенного центра имиджа от центра камеры с координатами $[0,0]$;
- $Width$ — параметр Хилласа;
- $Con2$ — концентрация света в пикселях с максимальной амплитудой относительно всего сайза;
- $Dens$ — отношение полного числа фотоэлектронов в имидже ($Size$) к количеству пикселей в имидже;

Критерий Джини

$$I_G(n) = 1 - \sum_{i=1}^J (p_i)^2$$

Загрязнение Джини узла n равно 1 минус сумма отношений класса к общему количеству образцов p_i , возведённых в квадрат, для каждого из множества классов J (в нашем случае 2 класса).

Визуализация дерева решений (максимальная глубина = 3)



Возможные пути решения проблемы переобучения

- 1) Ограничение максимального значения различных параметров дерева (уменьшение вариативности, увеличение погрешности)
- 2) Увеличение количества деревьев (уменьшение общей вариативности, без увеличения погрешности)

$$l_{tree\ i, term. - node\ j} = \frac{N_h}{N_h + N_g}$$

$$h = \frac{\sum_{i=1}^{n_{tree}} l_i}{n_{tree}}$$

Увеличение точности

- Использование масштабированных параметров;
- Использование выборки с возвращением.

$$\omega_{i, \text{scaled}} = \frac{\omega_i - \bar{\omega}_i}{\sigma_{\omega_i}}$$

Q-фактор

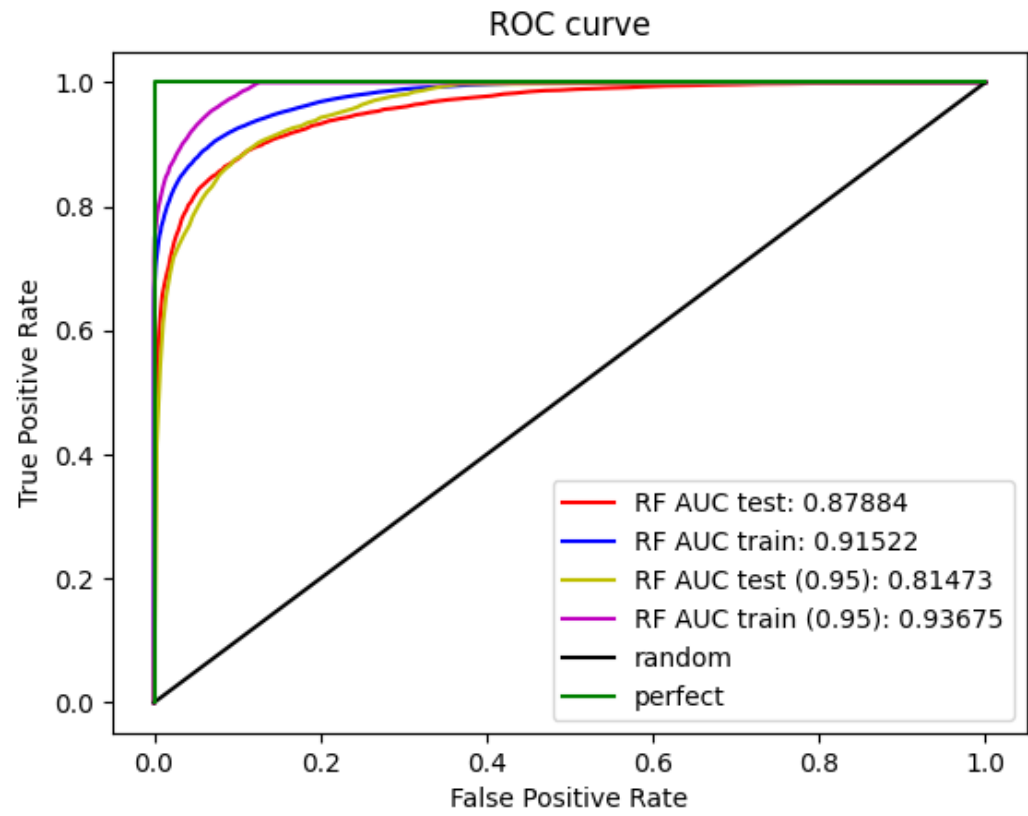
$$Q = \frac{\varepsilon_g}{\sqrt{\varepsilon_h}}$$

$$\varepsilon_g = \frac{N(g \rightarrow g)}{N(g \rightarrow g) + N(g \rightarrow h)}$$

$$\varepsilon_h = \frac{N(h \rightarrow g)}{N(h \rightarrow h) + N(h \rightarrow g)}$$

-suppression efficiency

Рос - кривая или кривая ошибок



N samples	Test (gam > 0.5)		Train (gam > 0.5)		Test (gam > 0.95)		Train (gam > 0.95)		Excess	Significance
	q-factor	roc auc	q-factor	roc auc	q-factor	roc auc	q-factor	roc auc		
100	2.81397	0.84056	2.40324	0.87765	10.45664	0.69786	10.48839	0.84529	117	4.58559
1000	2.84066	0.86911	2.61009	0.89773	11.31265	0.77784	12.47343	0.90178	141	4.75041
7000	2.87298	0.87884	2.82552	0.91522	12.20136	0.81473	89.96418	0.93675	135	4.41026
10000	2.88691	0.87979	2.92381	0.92089	11.18433	0.81913	inf	0.94329	137	4.33449
35000	2.89160	0.88135	4.75275	0.96969	10.83857	0.83120	inf	0.98349	127	3.82745
50000	2.89830	0.88057	10.9107	0.99448	10.11680	0.83259	inf	0.99768	114	3.30192
65000	2.9143	0.8822	59.813	0.9997	9.7717	0.838	inf	0.9998	130	3.64789

Подбор количества деревьев

N trees	Test (gam > 0.5)		Train (gam > 0.5)		Test (gam > 0.95)		Train (gam > 0.95)		Excess	Significance
	q-factor	roc auc	q-factor	roc auc	q-factor	roc auc	q-factor	roc auc		
10	2.63656	0.87358	2.62518	0.90651	7.07890	0.84696	10.87656	0.94249	154	3.96833415
100	2.87298	0.87884	2.82552	0.91522	12.20136	0.81473	89.96418	0.93675	135	4.41026
1000	2.88151	0.87856	2.84043	0.91541	11.76454	0.81667	inf	0.93706	112	3.92560
10000	2.89670	0.87900	2.84986	0.91568	12.01202	0.81645	inf	0.93647	117	4.14435

График адронности

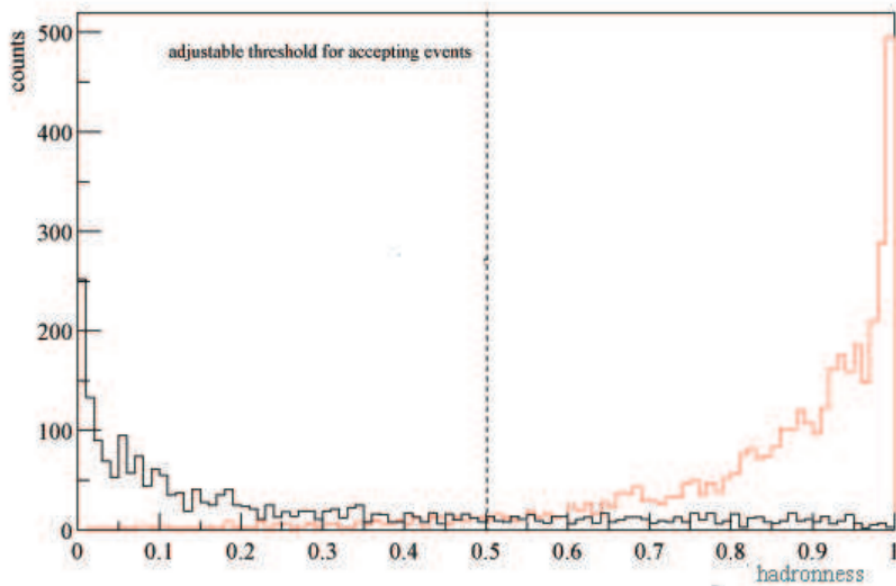


Рис. 8: График, показывающий среднюю адронность для двух тестовых выборок гамма-квантов (черная кривая) и адронов (красная кривая) в эксперименте MAGIC, где уже использовался данный метод классификации RF.

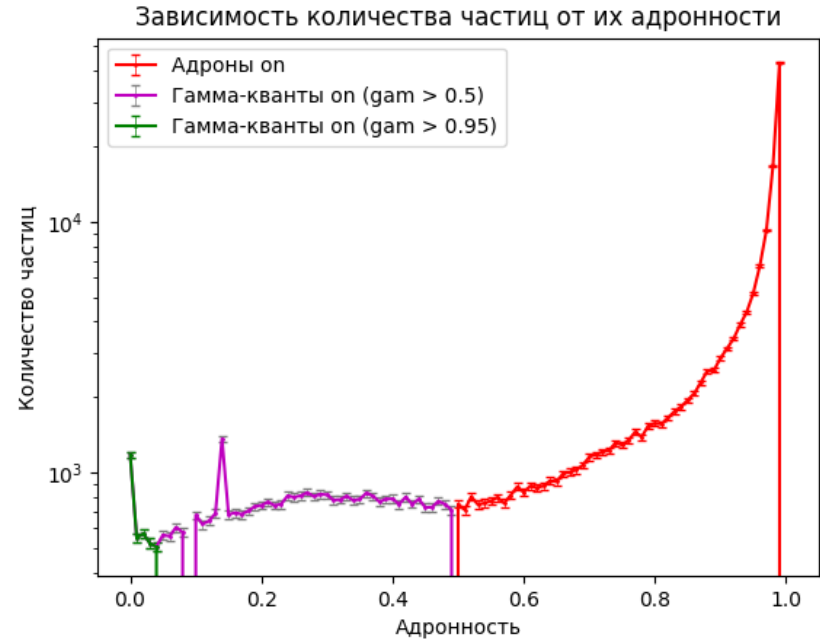


Рис. 9: График зависимости адронности для тестовых выборок гамма-квантов (черная линия) и адронов (красная линия), полученные в данной работе

Зависимость от параметра альфа

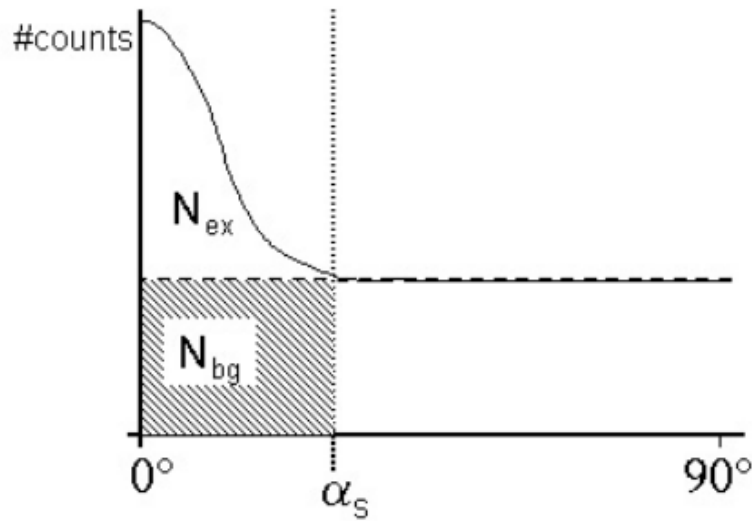


Рис. 10: Схематический график, показывающий расчет N_{ex} и N_{bg} . Горизонтальная пунктирная линия, представляющая уровень фона, в простейшем случае может считаться постоянной.

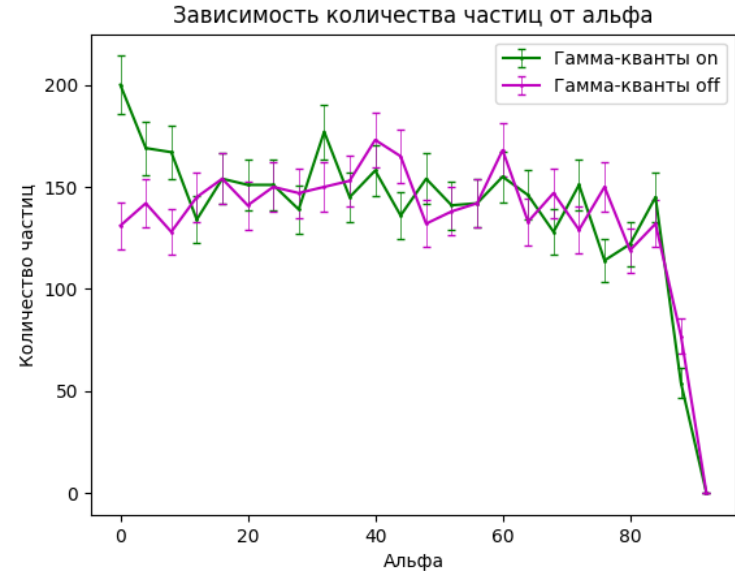


Рис. 11: Экспериментальный результат для зависимости количества гамма-квантов от альфа. Зеленая линия представляет собой зависимость для ON событий, фиолетовая — для OFF.

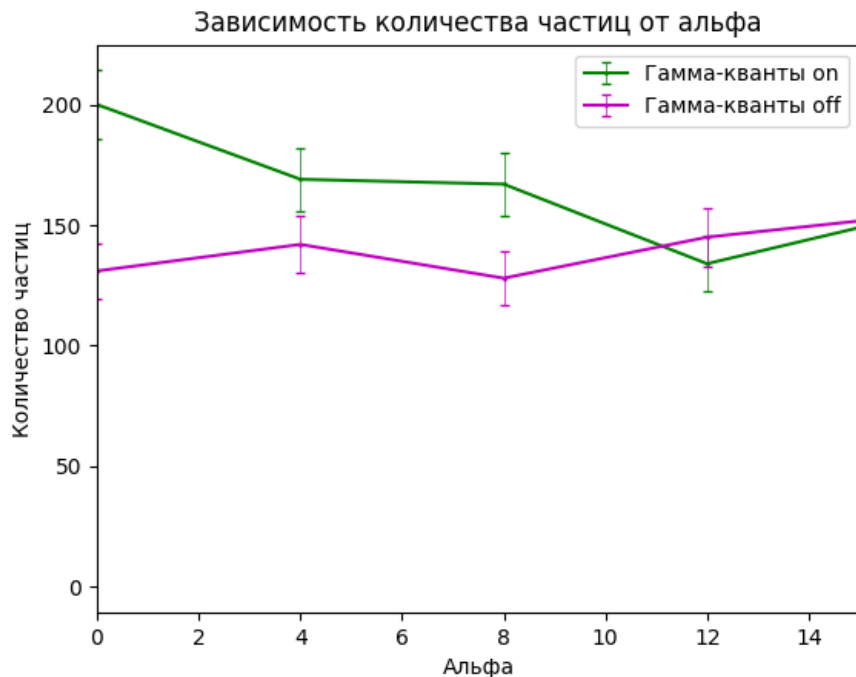


Рис. 12: Экспериментальный результат для зависимости количества гамма-квантов от альфа. Область альфа < 15 градусов.

100 tree, 7000 samples, (gam>0.95):

Для адронов:

Non = 19640;

Noff = 19495;

Для гамма-квантов:

Non = 536;

Noff = 401;

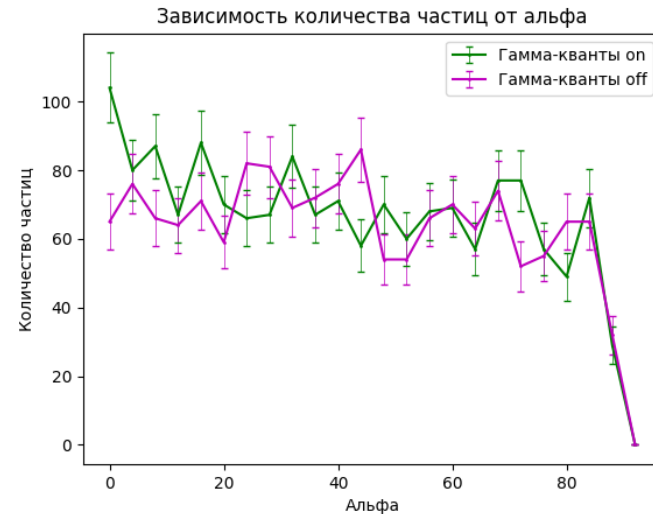
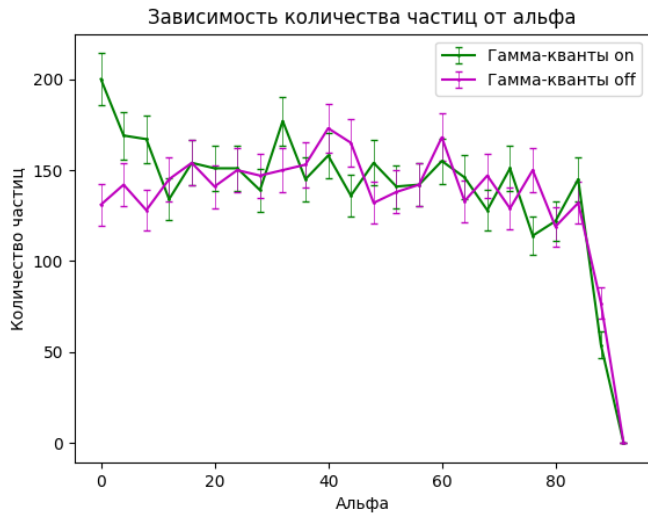
Excess = 135;

Significance = 4.41026.

$$Excess = N_{ON} - N_{OFF}$$

$$Significance = \frac{N_{ON} - N_{OFF}}{\sqrt{N_{ON} + N_{OFF}}}$$

	Test (gam > 0.5)		Train (gam > 0.5)		Test (gam > 0.95)		Train (gam > 0.95)		Exce ss	Signifi cance
	q-factor	roc auc	q-factor	roc auc	q-factor	roc auc	q-factor	roc auc		
With dens	2.872 98	0.878 84	2.825 52	0.915 22	12.20 136	0.814 73	89.96 418	0.9367 5	135	4.4102 6
Without dens	2.690 42	0.863 36	2.453 73	0.894 15	9.871 51	0.717 95	139.8 89582	0.8997 3	64	2.9272 9



Графики зависимости кол-ва частиц от альфа

Итоговый файл

	A	B	C	D	E	F	G	H	I	J
1	Nsob	data	P(0)	P(1)	particle	alfa	rc	size		
2	76923.0	311019.0	0.0727	0.9273	1.0	87.43	4.69	190.0		
3	78064.0	311019.0	0.029	0.971	1.0	81.67	0.27	165.0		
4	78156.0	311019.0	0.0006	0.9994	1.0	14.58	3.66	1740.0		
5	81562.0	311019.0	0.4722	0.5278	1.0	64.02	1.97	172.0		
6	81573.0	311019.0	0.1063	0.8937	1.0	5.69	-0.13	177.0		
7	85742.0	311019.0	0.1564	0.8436	1.0	16.21	-0.95	265.0		
8	85789.0	311019.0	0.9868	0.0132	0.0	62.28	-0.86	149.0		
9	85852.0	311019.0	0.0283	0.9717	1.0	10.12	-0.29	171.0		
10	85991.0	311019.0	0.099	0.901	1.0	62.04	-0.01	1050.0		
11	86015.0	311019.0	0.1414	0.8586	1.0	7.11	-1.62	2140.0		
12	86103.0	311019.0	0.963	0.037	0.0	44.28	-0.46	140.0		
13	86146.0	311019.0	0.069	0.931	1.0	67.86	3.15	272.0		
14	86250.0	311019.0	0.0663	0.9337	1.0	42.69	-0.46	155.0		
15	86414.0	311019.0	0.0	1.0	1.0	41.94	5.24	1140.0		
16	86424.0	311019.0	0.0031	0.9969	1.0	3.66	4.22	237.0		
17	86445.0	311019.0	0.0015	0.9985	1.0	29.01	3.19	938.0		
18	86485.0	311019.0	0.2731	0.7269	1.0	24.37	-0.0	441.0		
19	86572.0	311019.0	0.0004	0.9996	1.0	60.31	6.02	2340.0		
20	86656.0	311019.0	0.0203	0.9797	1.0	7.84	1.97	6050.0		
21	86746.0	311019.0	0.0028	0.9972	1.0	31.56	3.85	651.0		
22	86778.0	311019.0	0.0003	0.9997	1.0	60.31	2.5	432.0		
23	86788.0	311019.0	0.007	0.993	1.0	68.44	6.31	5980.0		
24	86910.0	311019.0	0.1524	0.8476	1.0	42.25	-1.85	173.0		

Выводы

- В данной работе освоена программа, включающая в себя алгоритм машинного обучения RANDOM FOREST, и приспособлена для выполнения задачи подавления адронного фона в эксперименте TAIGA.
- На выборках, полученных из данных Монте-Карло для первичных частиц, исследованы и получены оптимальные настройки программы для данной задачи, такие как число деревьев, максимальное число событий в одном дереве и т.п., выбраны параметры имиджей. Показано, что метод выдает стабильные результаты и устойчив к входным параметрам. Получены ожидаемые из вводимых выборок точности классификации.
- Аналогичная оптимизация проведена на экспериментальных выборках ON (когда телескоп направлен на источник) и OFF (когда телескоп направлен на фон неба). Зависимость от параметра альфа показала возможность выделения событий, инициированных гамма-квантами, от фона адронов.
- Полученный результат сравнивается с методами подавления фона полуэмпирическим подбором параметров, применяемыми в эксперименте ТАЙГА до настоящего времени. По сравнению с эмпирическим подходом, метод приспособливается к имеющемуся многомерному пространству с минимальным вмешательством человека: настройка количества деревьев, глубины роста дерева, минимальные узлы. Это более простое управление и настройка процедуры выделения ливней от гамма-квантов рассматриваются как общее преимущество метода RF.



Спасибо за внимание