

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО
ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В.ЛОМОНОСОВА»**

Физический факультет

Кафедра физики частиц и космологии

**Разработка фотонно-протонного
классификатора для обсерватории
Telescope Array**

**Дипломная работа
студента 6 курса, 643 группы
Пискунова Максима Сергеевича**

**Научный руководитель
к.ф.-м.н. Г. И. Рубцов**

**Заведующий кафедрой
физики частиц и космологии
д.ф.-м.н., проф. В. А. Рубаков**

Москва, 2015 г.

Аннотация

В настоящей дипломной работе разработан классификатор между протонами и фотонами для обсерватории Telescope Array, а также разработан алгоритм, вычисляющий верхнее ограничение на поток фотонов. Методы основаны на результатах Монте-Карло моделирования. Верхняя граница потока фотонов составляет $\frac{\Delta\gamma}{p} = 0.0014$ для частиц с энергией выше $E_{\min} = 10$ ЭэВ, $\frac{\Delta\gamma}{p} = 0.0077$ для частиц с энергией выше $E_{\min} = 31.6$ ЭэВ и $\frac{\Delta\gamma}{p} = 0.037$ для частиц с энергией выше $E_{\min} = 100$ ЭэВ.

Содержание

1 Введение	4
2 Метод	7
2.1 Реконструкция событий Telescope Array	7
2.2 Отбор событий	9
2.3 Вычисление выборки фотонов с согласующимся распределением	10
2.4 Выбор тренировочной и тестирующей выборки . . .	14
2.5 Алгоритмы классификации	15
2.6 Выбор наилучшей границы классификации	16
2.7 Вычисление фотонного потока	18
3 Результаты	21

1. Введение

Широкий атмосферный ливень (ШАЛ) – это каскад частиц, образующихся в атмосфере при взаимодействии с ней космических лучей – первичных высокоэнергетичных частиц. Изучение космических лучей представляет большой интерес в частности потому, что их происхождение достоверно не установлено (хотя известно, что они частично производятся в сверхновых [1]). Одной из интересных задач этой области является определение химического состава космических лучей сверх-высоких энергий – в частности, определение соотношения между количеством фотонов и протонов.

На данный момент работают три обсерватории космических лучей сверх-высоких энергий: Якутская комплексная установка ШАЛ [2], Pierre Auger Observatory [3] и Telescope Array [4].

Массив по изучению широких атмосферных ливней в Якутске расположен на реке Лене около деревни Октемцы (50 км к юго-западу от Якутска, $61.7^{\circ}N$, $129.40^{\circ}E$, в 100 м над уровнем моря). За время работы обсерватории ее максимальная площадь достигала 17 км^2 . В центральной части телескопа детекторы расположены на расстоянии от 100 м до 250 м друг от друга. За все время, было зарегистрировано около 10^6 широких атмосферных ливней с исходной энергией выше 30 ПэВ, а также три события с энергией более 100 ЭэВ с осью симметрии внутри площади массива и зенитным углом меньше 60° .

Обсерватория имени Пьера Оже (Pierre Auger) находится в западной Аргентине ($35.2068^{\circ}S$, $69.3160^{\circ}W$). Pierre Auger использует два типа детекторов для наблюдения за космическими лучами. Один тип детекторов представляет собой контейнер, наполненный 12000 л

воды, и обнаруживающий частицы с ней взаимодействующие. Массив состоит из 1600 контейнеров, размещенных в узлах гексагональной решетки с интервалом 1.5 км друг от друга. Также Pierre Auger использует 24 флюоресцентных детектора, образующие 4 флюоресцентные станции.

Telescope Array – эксперимент, находящийся в штате Юта, США ($39.2969^{\circ}N, 112.9086^{\circ}W$), также содержит два типа детекторов. Массив включает в себя 500 наземных сцинтилляционных детекторов, покрывающих большую площадь (данные которых используются в данной работе). Кроме того, инструмент включает в себя 38 флюоресцентных детекторов, собранных в 3 станции.

Обычно для определения состава используется следующий метод: для каждого события вычисляется определенный параметр, чувствительный к составу, строится распределение значений этого параметра для данных и фотонного Монте-Карло и производится сравнение. Например, по тому насколько близко распределение параметра, построенное на основе данных, находится к распределению параметра из моделирования, можно установить ограничение на наблюдаемый поток фотонов.

На настоящий момент используются несколько таких наблюдаемых, самыми распространенными из которых являются следующие:

- Атмосферная глубина, на которой продольная эволюция ливня достигает максимального количества частиц, X_{max} .
- Мюонная плотность, регистрируемая подземным детектором, ρ_{μ} .

- Атмосферная глубина, на которой достигается максимальное число мюонов, X_{max}^μ .
- Асимметрия времени от 10% до 50% интегрированного сигнала [5].

Несмотря на то, что эти параметры позволяют установить определенные ограничения на поток частиц, они по-отдельности, тем не менее, не позволяют надежно различать различные частицы, а следовательно ограничения и измерения, полученные с помощью таких методов, могут быть улучшены, если использовать параметры, различающие первичные частицы более надежно. В данной работе используются методы машинного обучения для классификации частиц, наблюдаемых обсерваторией Telescope Array, используя одновременно несколько параметров. На основе такого классификатора строится параметр, равный вероятности классификации события в качестве определенного типа частицы (например, фотона). Так как распределения такого параметра значительно отличаются для различных типов частиц, можно установить гораздо лучшее ограничение на поток фотонов. В некотором смысле, используемый в данной работе метод объединяет сильные стороны ранее использованных методов (использующих ранее перечисленные параметры по-отдельности).

2. Метод

2.1. Реконструкция событий Telescope Array

В качестве входных данных для классификации используются результаты реконструкции (версии Moscow.i12) Монте-Карло моделирования событий Telescope Array [6]. Используются следующие параметры:

1. Координата x ядра;
2. Координата y ядра;
3. F800;
4. Зенитный угол направления прихода первичной частицы θ ;
5. Азимутальный угол направления прихода первичной частицы ϕ ;
6. Реконструированное значение энергии в предположении первичного протона;
7. Приемная башня связи (BRM, LR или SK);
8. Количество сработавших детекторов;
9. Количество сработавших детекторов на границе установки;
10. Количество насыщенных детекторов;
11. Количество детекторов с двумя сигналами, разделенными по времени;

12. Количество нулевых детекторов, включенных в реконструкцию;
13. Качество фита, $\frac{\chi^2}{N}$;
14. η , параметр наклона функции пространственного распределения;
15. Параметр кривизны Линсли [7];
16. Наибольший сигнал;
17. Сумма сигналов на всех детекторах;
18. Расстояние от оси ливня до границы установки;
19. Размер симплекса итогового фита в пространстве параметров;
20. Количество детекторов, исключенных из фита фронта ШАЛ;
21. Количество пиков во временной развертке сигнала всех детекторов. Для пика необходимо значение сигнала выше определенного порога, большее 3-х соседей слева и 3-х соседей справа;
22. Полный сигнал, используемый для асимметрии;
23. Асимметрия верхнего / нижнего слоев;
24. Количество пиков во временной развертке детектора с максимальным сигналом;
25. Координаты детектора с максимальным сигналом;
26. Число асимметричных пиков (верхний > нижний);

27. Число асимметричных пиков (верхний < нижний);
28. Число детекторов с $r > 600$ м (не насыщенных, не исключенных геометрически);
29. Параметр, определенный как $S_b = \sum_{i=1}^N \left[S_i \times \left(\frac{r_i}{r_0} \right)^b \right]$, где сумма производится по всем N сработавшим детекторам, $r_0 = 1000$ м – эталонное расстояние, S_i – сигнал, измеренный на станции № i и r_i – расстояние от данного детектора до оси ливня [8].

Всего используются 4505300 Монте-Карло событий фотонов и 3369284 событий протонов.

2.2. Отбор событий

Среди событий, описанных в предыдущей секции, есть события, вызванные случайным одновременным попаданием случайных мюонов. Кроме того, качество реконструкции для многих событий не позволяет точно восстановить параметры первичной частицы. Поэтому мы выполнили отбор событий, используя следующие условия:

- $\frac{\chi^2}{N} < 5$ (параметр №13);
- Количество детекторов включенных в геометрический фит (разность параметров №8 и №20) не меньше 7;
- Расстояние до границы установки (параметр №18) больше, чем 1.2 км.

После применения приведенного выше фильтра, остаются 2569412 Монте-Карло событий фотонов и 249309 событий протонов.

2.3. Вычисление выборки фотонов с согласующимся распределением

Полученный набор событий, однако, пока нельзя использовать для классификации, т.к. спектр фотонов, полученных из Монте-Карло моделирования, сильно отличается от спектра протонов и железа, а значит классификатор может использовать энергию и зависящие от энергии параметры в качестве параметра отбора, следовательно точность классификации может быть существенно завышена.

Кроме того, набор фотонных событий Монте-Карло не изотропен, т.к. в него включены дополнительные события с выделенных направлений. Наконец, границы изменения параметра θ различны для фотонов и протонов.

Для избежания этих проблем необходимо произвести случайную выборку событий из множества фотонов таким образом, чтобы ее спектр совпадал со спектром протонов.

Для этого используется алгоритм схожий с фильтрацией, рассмотренной в предыдущей секции, выбираются события удовлетворяющие следующим условиям:

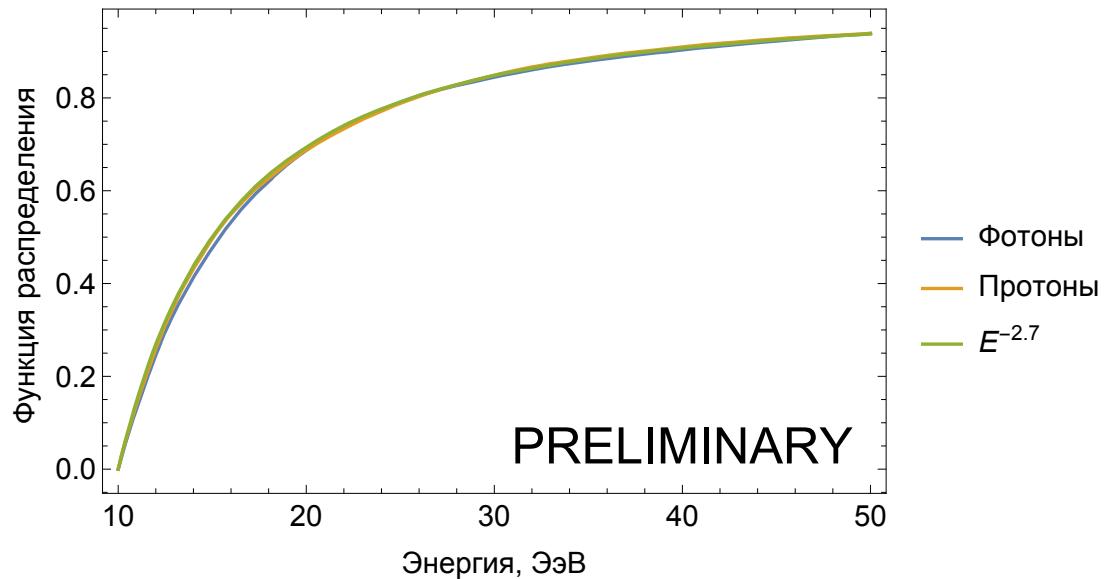
- Пара углов (θ, ϕ) не содержится в списке выделенных направлений, для которых проводилось отдельное моделирование;
- Угол θ не превосходит $\theta_{\max} = 60^\circ$;
- Реконструированное значение энергии (параметр №6) не превышает E_{\min} . В данной работе рассматриваются 3 значения E_{\min} : 10 ЭэВ, 31.6 ЭэВ, 100 ЭэВ;

- $r \leq \left(\frac{E}{E_{\min}}\right)^p$, где r – случайное число, равномерно распределенное от 0 до 1; E, E_{\min} – значения реконструированной и минимальной энергии из предыдущего пункта; p – разница между степенным индексом исходного распределения фотонов и необходимого их распределения (который совпадает со степенным индексом распределения протонов). p зависит от E_{\min} и для исследуемых результатов Монте-Карло моделирования составляет, $p = -2.2$ для $E_{\min} = 10$ ЭэВ, $p = -1.3$ для $E_{\min} = 31.6$ ЭэВ и $p = 0.0$ для $E_{\min} = 100$ ЭэВ.

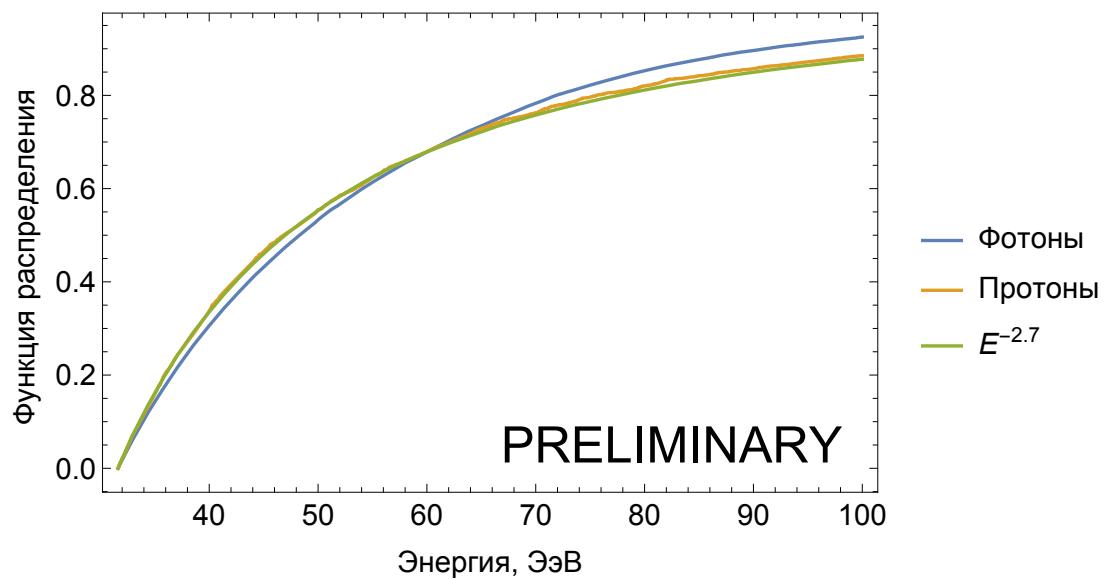
Распределение энергий фотонов и протонов для вышеописанной выборки показано на Рис. 1. Из графика видно, что спектры энергий фотонов и протонов в новой выборке совпадают с достаточно высокой точностью. В результате, классификатор не сможет различать типы частиц исключительно по их энергии. После описанного выше фильтра, выборка событий готова для классификации.

После построения данных выборок остаются следующие количества событий в зависимости от энергии:

1. $E_{\min} = 10$ ЭэВ. Остаются 98006 Монте-Карло событий фотонов и 29351 событий протонов.
2. $E_{\min} = 31.6$ ЭэВ. Остаются 123837 Монте-Карло событий фотонов и 3992 событий протонов.
3. $E_{\min} = 100$ ЭэВ. Остаются 85749 Монте-Карло событий фотонов и 459 событий протонов.



(a) Минимальная энергия $E_{\min} = 10$ ЭэВ



(b) Минимальная энергия $E_{\min} = 31.6$ ЭэВ

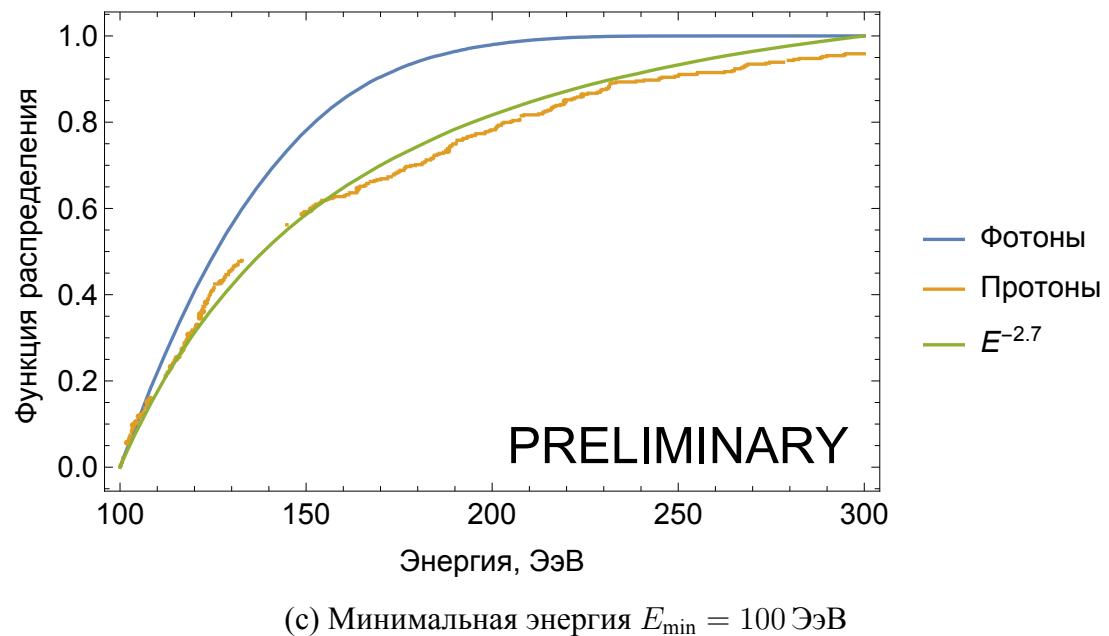


Рис. 1: Спектры наборов фотонов и протонов для различных значений минимальной энергии E_{\min} , используемые для обучения алгоритмов классификации. Видно, что типы частиц не могут быть определены исключительно из их энергий.

2.4. Выбор тренировочной и тестирующей выборки

Чтобы получить объективное значение точности классификации, необходимо разделить данные моделирования на два набора: по одному будет проводится тренировка классификатора, по другому будет оцениваться его производительность.

Наборы для тренировки и тестирования классификатора выбираются в соответствии с 3-я условиями:

1. Наборы данных для тренировки и тестирования должны быть одинакового размера (для достижения баланса между качеством классификации и точностью оценки производительности);
2. Количество событий в каждом наборе должно быть как можно большим, но не более 2^{15} (ограничение вводится из-за высокой вычислительной сложности и высоких требований по количеству памяти алгоритмов классификации);
3. События для наборов выбираются случайным образом в количестве, вычисленном в соответствии с предыдущими двумя пунктами.

Для наборов данных, изучаемых в данной работе, остаются следующие количества событий в зависимости от минимального значения энергии E_{\min} :

1. $E_{\min} = 10$ ЭэВ. Набор состоит из 29351 событий.
2. $E_{\min} = 31.6$ ЭэВ. Набор состоит из 3992 событий.
3. $E_{\min} = 100$ ЭэВ. Набор состоит из 459 событий.

2.5. Алгоритмы классификации

Для обучения классификатора, разделяющего фотоны и протоны, используются следующие алгоритмы машинного обучения:

- **Логистическая регрессия** [9, р. 128]. Для предсказания типа частиц используется фит логистической функцией $F(t) = \frac{e^t}{e^t + 1}$, где t – функция значений параметров: $t = \beta_0 + \sum_{i=1}^N \beta_i x_i$, где x_i – значения параметров, а β_i – параметры фита.
- **Наивный байесовский классификатор** [10]. Для классификации событий предполагается, что параметры независимы и используется теорема Байеса из теории вероятностей.
- **Метод k ближайших соседей** [11]. В данном методе используются классы k соседей, ближайших в пространстве параметров, и выбирается класс, встречающийся среди соседей наиболее часто. k – параметр метода, выбираемый для максимизации качества классификации.
- **Искусственная нейронная сеть** [12]. Алгоритм, имитирующий центральную нервную систему животных. Классификация происходит с помощью нескольких слоев функций (нейронов), каждая из которых вычисляет свое значение как линейную комбинацию значений нейронов предыдущего слоя. Коэффициенты линейных функций, используемых каждым нейроном, являются параметрами классификации.
- **Random forest** [13]. Классификация работает путем конструирования нескольких деревьев принятия решений, которые опре-

деляют тип частицы исходя из ответов на множество вопросов о параметрах события.

- **Метод опорных векторов** [14]. В данном методе пространство параметров делится на две области, в каждой из которых находятся события соответствующие разным типам частиц, таким образом, что точки соответствующие разным частицам как можно сильнее разделены. В базовом методе области различных классов разделены прямой, однако можно использовать ядерный метод приводящий к полиномиальной границе между областями.

2.6. Выбор наилучшей границы классификации

Результат работы классификатора для каждого события – вероятность того, что данная частица является фотоном или протоном. Обычно, для классификатора между фотонами и протонами, событие классифицируется как фотон, если вероятность этого превышает 50%. Однако, для установления верхней оценки на поток фотонов, оптимальнее использовать другую границу классификации, например, классифицировать частицу как фотон если вероятность этого более 95%. Для того, чтобы понять, какую границу необходимо выбрать, нужно провести вычисление погрешности потока фотонов как функции от точностей классификации фотонов и протонов.

Пусть $\epsilon_{p\gamma}$ – часть протонов, ошибочно классифицирующихся как фотоны, а $\epsilon_{\gamma p}$ – часть фотонов, классифицирующихся в качестве протонов. Кроме того, пусть γ и p – истинные количества фотонов и протонов. Наконец, пусть γ_{obs} и p_{obs} – количества событий, классифици-

рованных в качестве фотонов и протонов соответственно.

Тогда, по определениям вышеприведенных величин,

$$\gamma_{\text{obs}} = \gamma (1 - \epsilon_{\gamma p}) + p \epsilon_{p\gamma}, \quad (1)$$

$$p_{\text{obs}} = p (1 - \epsilon_{p\gamma}) + \gamma \epsilon_{\gamma p}. \quad (2)$$

Решая данные уравнения для γ , получаем:

$$\gamma = \frac{\gamma_{\text{obs}} - p_{\text{obs}} \epsilon_{p\gamma} - \gamma_{\text{obs}} \epsilon_{p\gamma}}{1 - \epsilon_{p\gamma} - \epsilon_{\gamma p}} \quad (3)$$

Однако, так как $\gamma \ll p$, в данной работе вычисляется только верхнее ограничение потока фотонов, а значит, необходимо вычислить погрешность количества фотонов γ :

$$\Delta_\gamma = \sqrt{\left(\frac{\partial}{\partial \epsilon_{p\gamma}} \left(\frac{\gamma_{\text{obs}} - p_{\text{obs}} \epsilon_{p\gamma} - \gamma_{\text{obs}} \epsilon_{p\gamma}}{1 - \epsilon_{p\gamma} - \epsilon_{\gamma p}} \right) \Delta_{\epsilon_{p\gamma}} \right)^2 + \left(\frac{\partial}{\partial \epsilon_{\gamma p}} \left(\frac{\gamma_{\text{obs}} - p_{\text{obs}} \epsilon_{p\gamma} - \gamma_{\text{obs}} \epsilon_{p\gamma}}{1 - \epsilon_{p\gamma} - \epsilon_{\gamma p}} \right) \Delta_{\epsilon_{\gamma p}} \right)^2} \quad (4)$$

Подставляя значения для γ_{obs} и p_{obs} из уравнений 1 и 2, получаем:

$$\Delta_\gamma = \sqrt{\frac{\Delta_{\epsilon_{p\gamma}}^2 p^2 + \Delta_{\epsilon_{\gamma p}}^2 \gamma^2}{(1 - \epsilon_{p\gamma} - \epsilon_{\gamma p})^2}} \quad (5)$$

Предполагая, что $\gamma \rightarrow 0$, получаем:

$$\Delta_\gamma = \frac{\Delta_{\epsilon_{p\gamma}} p}{1 - \epsilon_{p\gamma} - \epsilon_{\gamma p}} \quad (6)$$

Значение $\Delta_{\epsilon_{p\gamma}}$ можно оценить используя биномиальное распределение, зная количество событий в наборе и само значение $\epsilon_{p\gamma}$. Для этого необходимо выполнить два шага:

1. Найти такое биномиальное распределение путем подбора n , чтобы вероятность получить в нем значение случайной величины совпадающее со значением $\epsilon_{p\gamma}$, умноженным на количество событий, была не слишком маленькой (в данной работе берется значение 5%).
2. Вычислить аргумент функции распределения, полученного в предыдущем пункте, таким, чтобы ее значение было равно 95%. Полученный аргумент является максимально возможным количеством частиц, которые могут быть неправильно классифицированы как фотоны в наборе из данного количества частиц.

Считая в первом приближении, что $\Delta_{\epsilon_{p\gamma}} \approx \epsilon_{p\gamma}$, получаем, что необходимо сделать $\epsilon_{p\gamma}$ существенно меньшим, чем $\epsilon_{\gamma p}$, для того, чтобы минимизировать Δ_γ . Поэтому, можно пренебречь $\epsilon_{p\gamma}$ в знаменателе, и получить:

$$\Delta_\gamma \approx \frac{\Delta_{\epsilon_{p\gamma}} p}{1 - \epsilon_{\gamma p}}, \quad (7)$$

где $\Delta_{\epsilon_{p\gamma}}$ вычисляется в соответствии с алгоритмом, приведенным выше.

Таким образом, задача состоит в том, чтобы выбрать границу вероятностей разделения между фотонами и протонами так, чтобы минимизировать Δ_γ из предыдущего уравнения.

2.7. Вычисление фотонного потока

Учитывая результат предыдущей подсекции, задача состоит в том, чтобы вычислить значения формулы 7, используя разные алгоритмы

Граница классификации	Точность классификации фотонов	Точность классификации протонов	Верхняя граница потока
0.5	0.89	$1 - 8.1 \times 10^{-2}$	0.096
0.1	0.70	$1 - 1.5 \times 10^{-2}$	0.025
0.01	0.37	$1 - 2.4 \times 10^{-3}$	0.0092
0.001	0.14	1	0.0014
0.0002	0.068	1	0.0030
0.0001	0.048	1	0.0043

Таблица 1: Производительность классификатора на основе искусственных нейронных сетей для набора событий с минимальной энергией $E_{\min} = 10$ ЭэВ в зависимости от вероятностной границы разделения типов частиц. Верхняя граница потока равна $\frac{\Delta_\gamma}{p}$

классификации и разные границы вероятностей разделения, и выбрать минимальное значение.

Лучшая производительность для всех трех диапазонов энергий достигается с помощью алгоритма искусственных нейронных сетей. Точность в зависимости от границы разделения приведена в таблицах 1, 2 и 3. На Рис. 2 можно увидеть распределение фотонности – степени уверенности нейронной сети в том, что данное событие является фотоном.

Граница классификации	Точность классификации фотонов	Точность классификации протонов	Верхняя граница потока
0.5	0.92	$1 - 6.2 \times 10^{-2}$	0.082
0.1	0.78	$1 - 1.7 \times 10^{-2}$	0.031
0.01	0.48	$1 - 5.0 \times 10^{-3}$	0.019
0.001	0.19	1	0.0077
0.0002	0.088	1	0.017

Таблица 2: Производительность классификатора на основе искусственных нейронных сетей для набора событий с минимальной энергией $E_{\min} = 31.6$ ЭэВ в зависимости от вероятностной границы разделения типов частиц. Верхняя граница потока равна $\frac{\Delta_\gamma}{p}$

Граница классификации	Точность классификации фотонов	Точность классификации протонов	Верхняя граница потока
0.5	0.93	$1 - 5.2 \times 10^{-2}$	0.097
0.1	0.73	$1 - 1.7 \times 10^{-2}$	0.057
0.01	0.35	1	0.037
0.001	0.057	1	0.23

Таблица 3: Производительность классификатора на основе искусственных нейронных сетей для набора событий с минимальной энергией $E_{\min} = 100$ ЭэВ в зависимости от вероятностной границы разделения типов частиц. Верхняя граница потока равна $\frac{\Delta_\gamma}{p}$

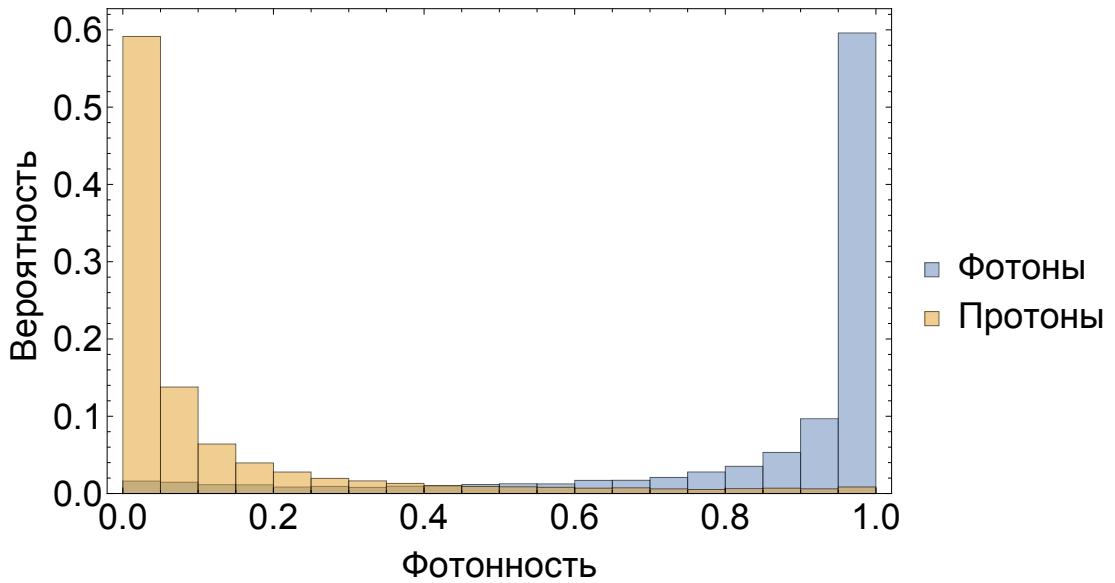


Рис. 2: Распределение нового параметра – результата работы классификатора – фотонности. Видно, что фотонность для фотонов и протонов существенно различается.

3. Результаты

В данной работе разработан классификатор между различными типами частиц для данных обсерватории Telescope Array, а также разработан алгоритм, позволяющий оценить верхнюю границу потока фотонов. Для используемых результатов Монте-Карло моделирования, в зависимости от минимального значения энергии, верхние оценки потока, определяемого данным алгоритмом, составляют:

1. $E_{\min} = 10$ ЭэВ. Верхняя граница на поток фотонов $\frac{\Delta\gamma}{p} = 0.0014$.
2. $E_{\min} = 31.6$ ЭэВ. Верхняя граница на поток фотонов $\frac{\Delta\gamma}{p} = 0.0077$.
3. $E_{\min} = 100$ ЭэВ. Верхняя граница на поток фотонов $\frac{\Delta\gamma}{p} = 0.037$.

После дополнительного тестирования метод будет представлен коллаборации Telescope Array с целью получения разрешения на использование данных установки.

Еще одним направлением развития данной работы может быть использование исходных, не реконструированных, показаний детекторов в качестве входных данных классификатора. Данное изменение существенно повысит вычислительную сложность алгоритмов машинного обучения, однако может улучшить точность классификации, а значит, уменьшить верхнюю границу потока.

Список литературы

- [1] M. Ackermann *et al.*, “Detection of the Characteristic Pion-Decay Signature in Supernova Remnants,” *Science*, vol. 339, p. 807, 2013, 1302.3307.
- [2] I. Anatoly, “The Yakutsk array experiment: Main results and future directions,” *EPJ Web Conf.*, vol. 53, p. 04003, 2013.
- [3] J. Abraham *et al.*, “Properties and performance of the prototype instrument for the Pierre Auger Observatory,” *Nucl.Instrum.Meth.*, vol. A523, pp. 50–95, 2004.
- [4] H. Kawai *et al.*, “Telescope array experiment,” *Nucl.Phys.Proc.Suppl.*, vol. 175-176, pp. 221–226, 2008.
- [5] P. Abreu *et al.*, “The Pierre Auger Observatory II: Studies of Cosmic Ray Composition and Hadronic Interaction models,” 2011, 1107.4804.

- [6] “CORSIKA Simulation of the Telescope Array Surface Detector,” 2014, 1403.0644.
- [7] T. Abu-Zayyad *et al.*, “Upper limit on the flux of photons with energies above 10^{19} eV using the Telescope Array surface detector,” *Phys.Rev.*, vol. D88, no. 11, p. 112005, 2013, 1304.5614.
- [8] G. Ros, A. Supanitsky, G. Medina-Tanco, L. del Peral, J. D’Olivo, *et al.*, “A new composition-sensitive parameter for Ultra-High Energy Cosmic Rays,” *Astropart.Phys.*, vol. 35, pp. 140–151, 2011, 1104.3399.
- [9] D. A. Freedman, *Statistical Models: Theory and Practice*. Cambridge University Press, 2009.
- [10] D. J. Hand and K. Yu, “Idiot’s bayes—not so stupid after all?,” *International statistical review*, vol. 69, no. 3, pp. 385–398, 2001.
- [11] N. S. Altman, “An introduction to kernel and nearest-neighbor nonparametric regression,” *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.
- [12] C. M. Bishop *et al.*, “Neural networks for pattern recognition,” 1995.
- [13] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [14] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.