

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
ВЫСШЕГО ОБРАЗОВАНИЯ «НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЯДЕРНЫЙ  
УНИВЕРСИТЕТ «МИФИ»»

УДК 539.165.2

ОТЧЕТ  
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ  
РАЗРАБОТКА АЛГОРИТМА ИДЕНТИФИКАЦИИ  
ФОТОНОВ В ЭЛЕКТРОМАГНИТНОМ  
КАЛОРИМЕТРЕ ЭКСПЕРИМЕНТА SPD

Научный руководитель  
к.ф.-м.н., в.н.с.

\_\_\_\_\_ Е. Ю. Солдатов

Студент

\_\_\_\_\_ Г. Е. Петров

Москва 2025

# СОДЕРЖАНИЕ

<b>Введение</b>	<b>3</b>
<b>1 Теоретическая часть</b>	<b>4</b>
1.1 Цель эксперимента SPD . . . . .	4
1.2 Первая стадия эксперимента . . . . .	4
1.3 Вторая стадия эксперимента . . . . .	5
1.4 Назначение и конструкция электромагнитного калориметра SPD . . . . .	6
<b>2 Исследование архитектуры и функционала среды моделирования SPDR00T и текущих алгоритмов реконструкции фотонов в электромагнитном калориметре SPD</b>	<b>8</b>
2.1 Моделирование взаимодействия частиц в среде SPDR00T и тестирование модели электромагнитного калориметра . . . . .	8
2.2 Изучение текущих алгоритмов кластеризации и идентификации частиц в электромагнитном калориметре SPD . . . . .	10
<b>3 Модификация алгоритмов идентификации частиц в электромагнитном калориметре SPD</b>	<b>11</b>
3.1 Анализ параметров из эксперимента ATLAS на применимость к идентификации частиц в электромагнитном калориметре SPD . . . . .	11
3.2 Сравнение эффективности различных алгоритмов классификации кластеров на простых выборках . . . . .	17
3.2.1 Описание подходов к классификации кластеров . . . . .	17
3.2.2 Обучение и тестирование BDT-классификатора с использованием адаптированных параметров кластеров . . . . .	18
3.2.3 Классификация методом фиксированных отборов с использованием адаптированных параметров кластеров . . . . .	23
3.2.4 Использование всей совокупности параметров из ATLAS и текущего алгоритма классификации в SPDR00T для классификации кластеров с помощью BDT . . . . .	25
3.3 Ранжирование переменных полного набора параметров . . . . .	31
3.4 Тестирование классификаторов на более реалистичных выборках . . . . .	33
3.4.1 Уточнение постановки задачи . . . . .	33
3.4.2 Качество классификаторов по мере приближения выборки к более реалистичному случаю . . . . .	34
3.4.3 Сравнение сценариев разделения . . . . .	37
<b>4 Заключение</b>	<b>38</b>

# ВВЕДЕНИЕ

Современные исследования в области физики высоких энергий направлены на изучение фундаментальных свойств материи и взаимодействий на субатомном уровне. Одной из наиболее актуальных задач является исследование структуры нуклона, включая вклад глюонов и кварков в его спин и массу. Для выполнения этой задачи разрабатываются новые экспериментальные установки, такие как Spin Physics Detector (SPD), предназначенный для работы на коллайдере NICA в Объединённом институте ядерных исследований (Дубна).

SPD представляет собой универсальный детектор, созданный для изучения структуры протонов и дейтронов, а также связанных с ними спиновых явлений. Уникальные возможности этой установки позволяют исследовать поляризованные столкновения протонов и дейтронов в ранее недоступных энергетических диапазонах [1], что делает SPD важным дополнением к текущим и будущим экспериментам в таких лабораториях, как RHIC (BNL), LHC (CERN) и EIC (BNL). Исследования в рамках SPD нацелены на решение ключевых вопросов, связанных с содержанием глюонов в нуклонах и динамикой их взаимодействий.

Важную роль в экспериментальной программе SPD играет электромагнитный калориметр, предназначенный для регистрации и измерения энергии фотонов и электронов. Точность реконструкции фотонов в калориметре напрямую влияет на качество физических измерений, включая изучение процессов рождения чармония, открытого чарма и прямых фотонов, которые являются чувствительными зондами глюонной структуры нуклона. В связи с этим особое внимание уделяется разработке и оптимизации алгоритмов реконструкции и идентификации фотонов, а также их адаптации к геометрии и характеристикам детектора SPD.

Настоящая работа посвящена разработке и оптимизации алгоритмов идентификации фотонных кластеров в эксперименте SPD с использованием программной среды моделирования SPDROOT. В рамках работы анализируются существующие подходы к реконструкции и классификации фотонных кластеров. Предлагаются и тестируются новые методы, направленные на повышение точности идентификации фотонов и снижение фоновых вкладов от нейтральных мезонов, таких как  $\pi^0$ . целью работы является улучшение качества идентификации фотонов в электромагнитном калориметре SPD, что позволит повысить точность физических измерений и расширить возможности эксперимента в изучении структуры нуклонов и связанных с ними спиновых явлений.

## ЗАДАЧИ РАБОТЫ

- Исследование существующей архитектуры и функционала модели электромагнитного калориметра SPD и среды моделирования SPDROOT.
- Изучение существующих алгоритмов реконструкции фотонов в электромагнитном калориметре SPD.
- Разработка и внедрение модифицированных алгоритмов идентификации с целью улучшения точности идентификации фотонов.
- Тестирование и сравнение точности модифицированных алгоритмов с исходными.

# 1. ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

## 1.1 Цель эксперимента SPD

Одна из основных целей эксперимента SPD - получение доступа к глюонным функциям распределения, зависящим от поперечного импульса (далее будет использоваться аббревиатура TMD PDFs - Transverse Momentum Dependent Parton Distribution Functions) в протоне и дейтроне [2]. Для получения же доступа к TMD PDFs в самом эксперименте будут анализироваться такие реакции, как рождение чармония, открытого чарма и прямых фотонов [3].

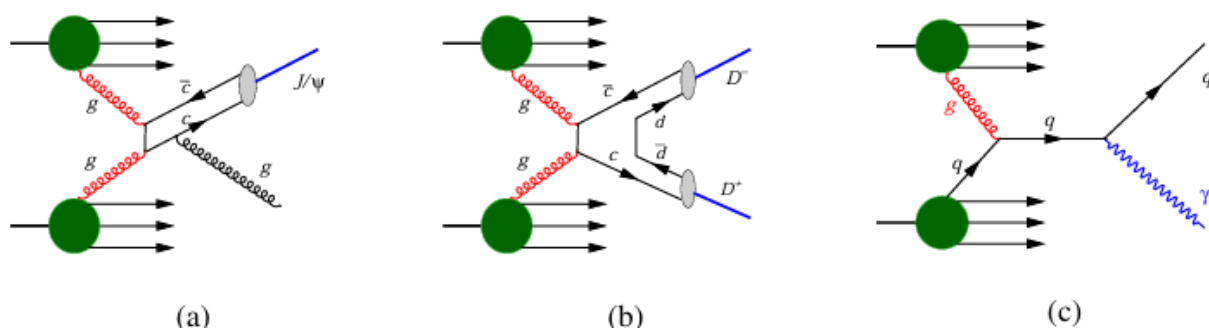


Рисунок 1 — Диаграммы иллюстрирующие реакции: рождение чармония (а), открытого чарма (b) и прямых фотонов (с)

## 1.2 Первая стадия эксперимента

### Цели:

- Подготовка к изучению TMD PDFs глюонов и кварков.
- Сбор данных для последующей модернизации системы детекторов.

### Детекторы, необходимые на первой стадии [4]:

- **Straw-Based Tracking System (ST):** Реконструкция треков заряженных частиц.
- **Micromegas-Based Central Tracker (MCT):** Центральный трекер для улучшения точности треков.
- **Muon Range System (RS):** Идентификация мюонов.

- **Beam-Beam Counters (BBC):** Контроль светимости и поляриметрия.
- **Zero Degree Calorimeters (ZDC):** Контроль светимости и измерение параметров нейтральных частиц.

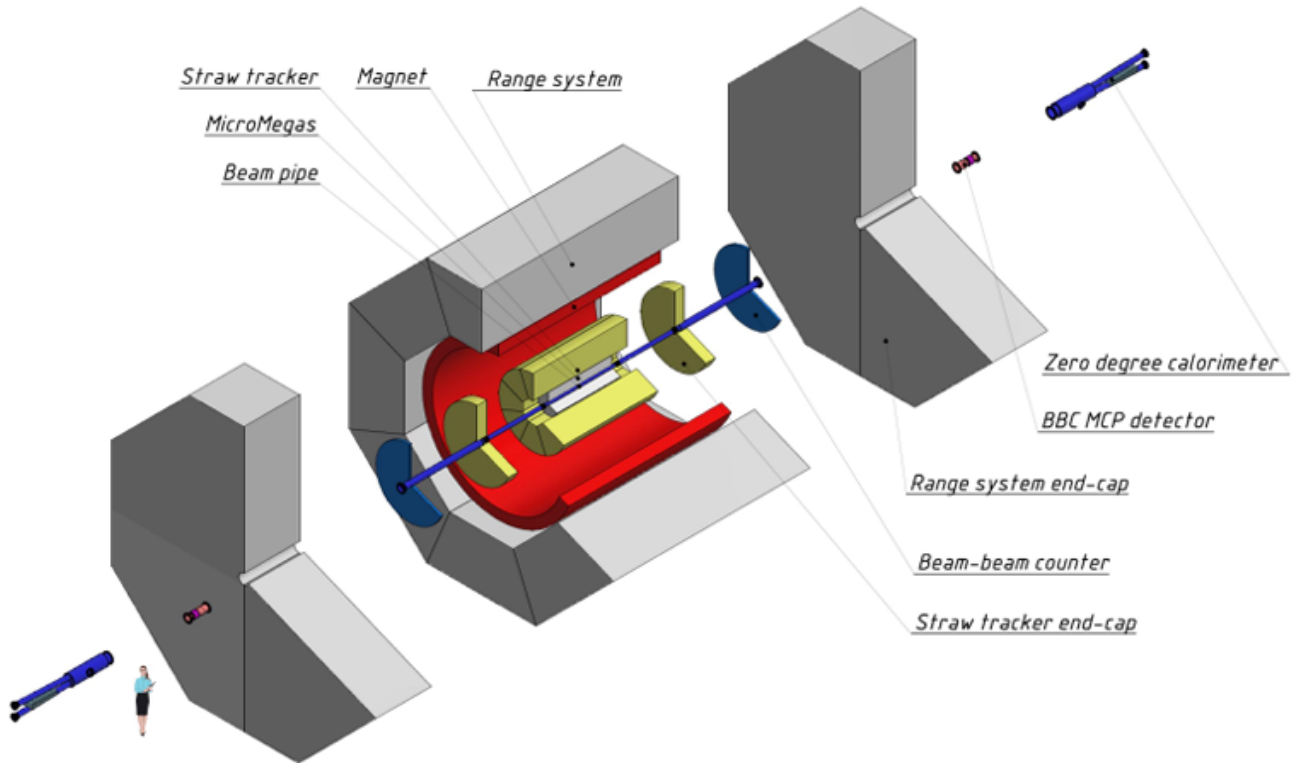


Рисунок 2 — Детекторы первой стадии эксперимента

### 1.3 Вторая стадия эксперимента

Цель:

- Изучение TMD PDFs глюонов и кварков.

Улучшение детекторов:

- **Electromagnetic Calorimeter (ECal):** Детектирование фотонов.
- **Silicon Vertex Detector (SVD):** Система, заменяющая MCT, для реконструкции вершин распада короткоживущих частиц, что важно для точного анализа  $D$ -мезонов.
- **Time-of-Flight (TOF) System:** Введение системы для идентификации частиц по времени пролета (в частности для разрешения  $\pi/K$   $K/p$ )

- **FARICH (Aerogel-Based Detector или AEG)**: Улучшение идентификации частиц (в частности протонов и каонов).

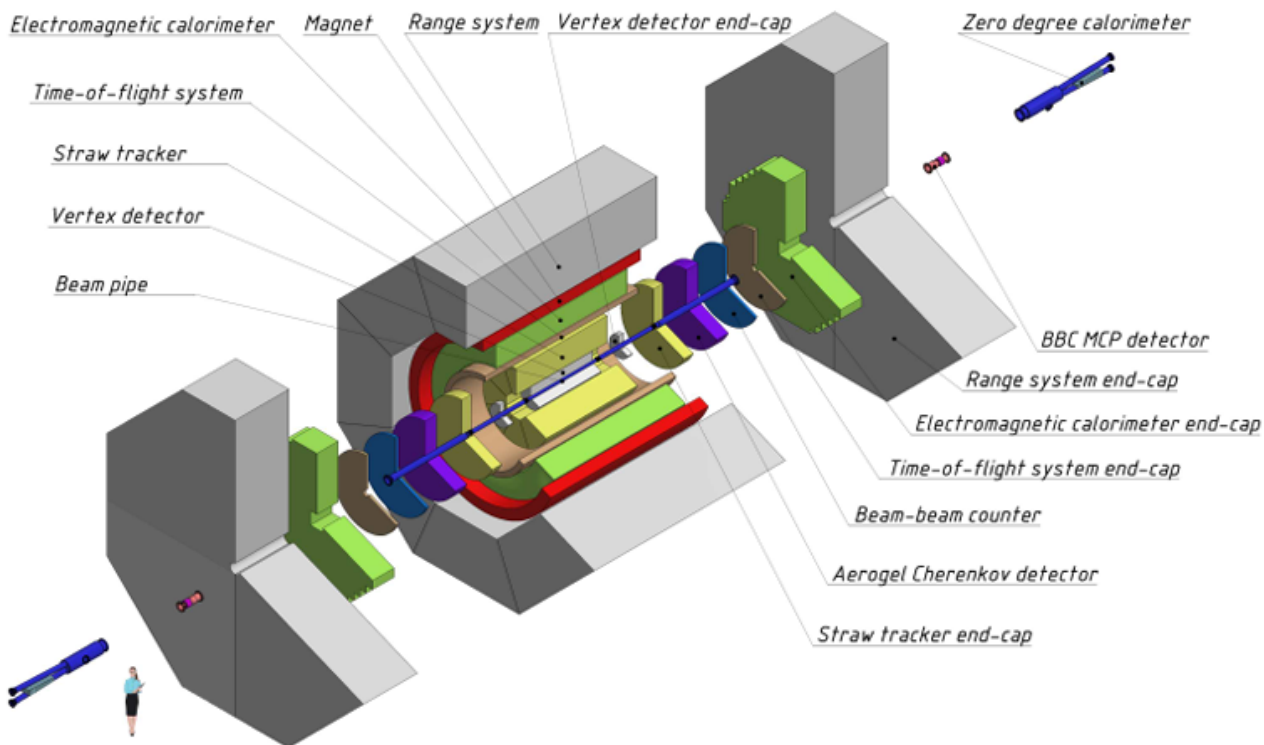


Рисунок 3 — Детекторы второй стадии эксперимента

## 1.4 Назначение и конструкция электромагнитного калориметра SPD

Электромагнитный калориметр (ECal) является одним из ключевых компонентов детектора SPD и предназначен для точного измерения энергии, координат и времени прихода фотонов и электронов, возникающих в результате столкновений частиц. Его значимость обусловлена рядом задач, решаемых в рамках эксперимента SPD, таких как реконструкция нейтральных мезонов ( $\pi^0$ ,  $\eta$ ), подавление фоновых событий при регистрации прямых фотонов [4].

Конструктивно калориметр представляет собой сэмплирующий детектор, состоящий из чередующихся слоёв свинца (поглотитель) и пластикового сцинтиллятора. Каждый модуль включает 190 двойных слоев из 1.5 мм сцинтиллятора и 0.5 мм свинца, что обеспечивает суммарную толщину активной части порядка 380 мм. Свет от сцинтилляторов собирается с помощью волокон со сдвигом длины волны (wavelength-shifting fibers, or WLS) и регистрируется многоэлементными фотодиодами (multi-pixel photon counter, or MPPC) [4]. Общая толщина модуля с учётом конструктивных элементов составляет около 490 мм.

Калориметр разделён на цилиндрическую (баррельную) часть и два торцевых эндкапа, что позволяет покрывать почти полный  $4\pi$  телесный угол: по  $\phi$  от 0 до  $2\pi$ , по  $\eta$  приблизительно от  $-3.2$  до  $3.2$ .

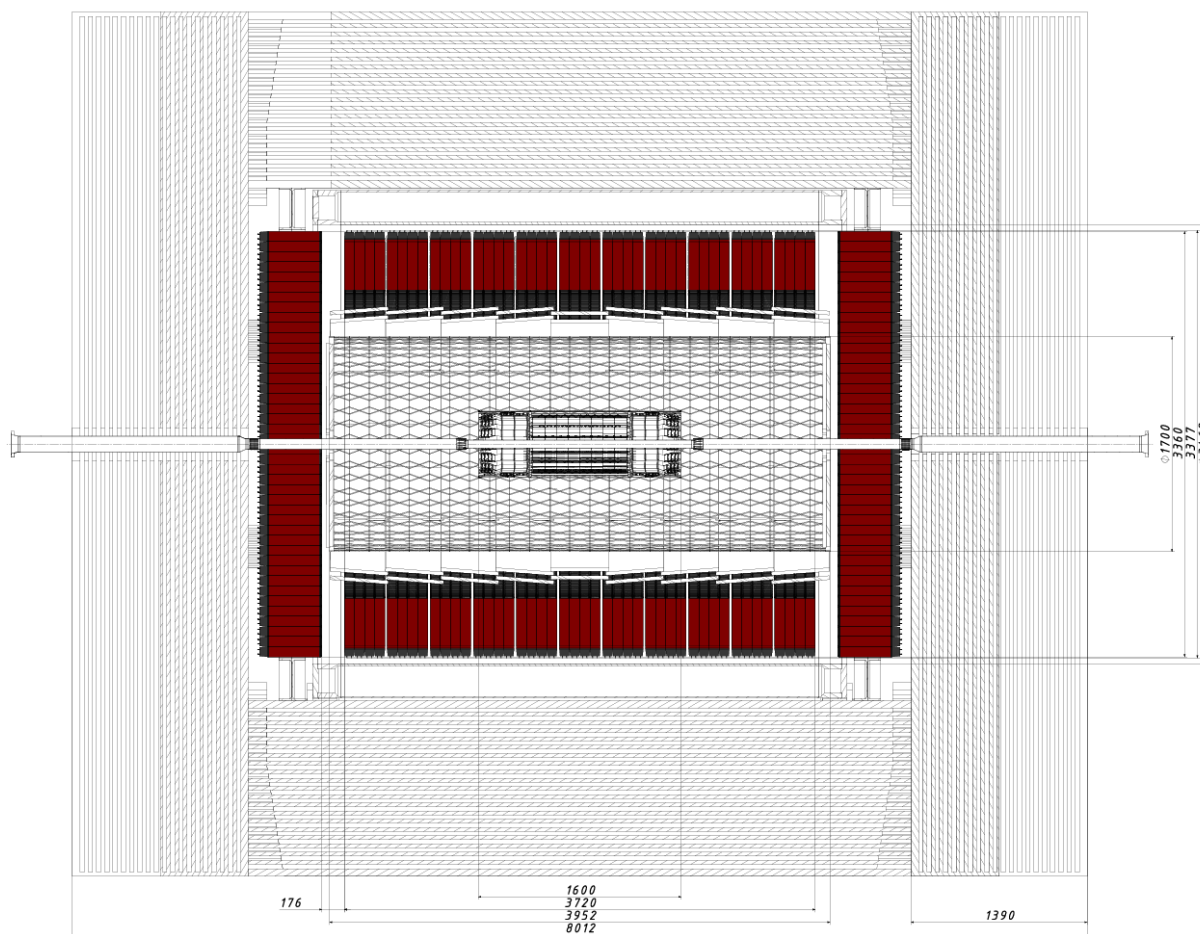


Рисунок 4 — Калориметр, вид сбоку, красным изображены эндкапы и баррельная часть калориметра

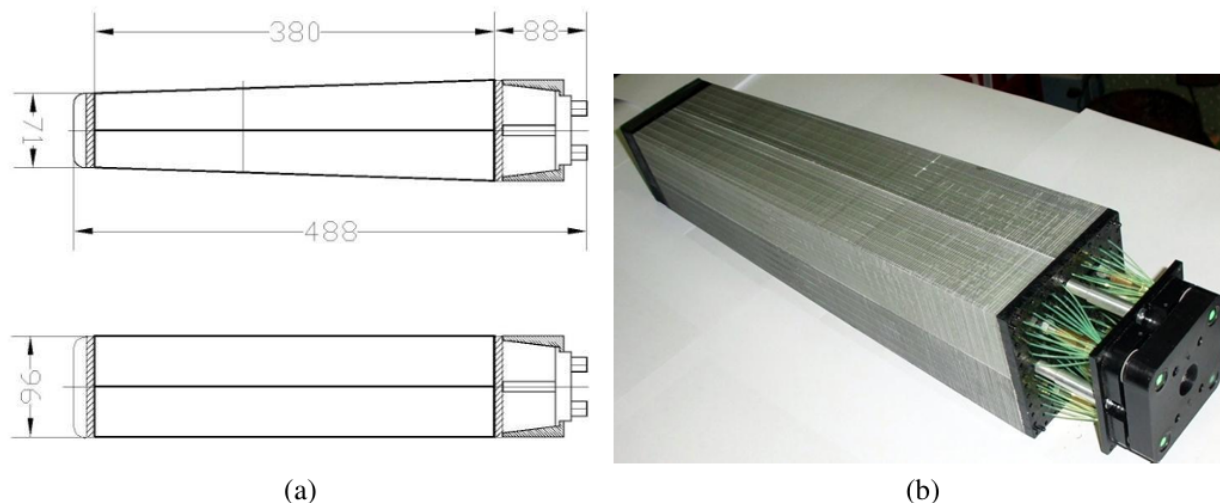


Рисунок 5 — Чертеж отдельного модуля калориметра (a) и фото модуля без внешнего корпуса (b)

## 2. ИССЛЕДОВАНИЕ АРХИТЕКТУРЫ И ФУНКЦИОНАЛА СРЕДЫ МОДЕЛИРОВАНИЯ SPDROOT И ТЕКУЩИХ АЛГОРИТМОВ РЕКОНСТРУКЦИИ ФОТОНОВ В ЭЛЕКТРОМАГНИТНОМ КАЛОРИМЕТРЕ SPD

### 2.1 Моделирование взаимодействия частиц в среде SPDROOT и тестирование модели электромагнитного калориметра

SPDROOT является специализированной средой моделирования и анализа данных, разработанной для эксперимента Spin Physics Detector (SPD) на коллайдере NICA. Эта среда основана на фреймворке ROOT, который широко используется в физике высоких энергий для обработки и анализа больших объемов данных. SPDROOT предоставляет инструменты для моделирования детекторов, генерации событий, реконструкции частиц и анализа данных. Она включает в себя модули для симуляции взаимодействий частиц с детекторами, что позволяет исследователям оценивать эффективность различных алгоритмов реконструкции и оптимизировать параметры детекторов. В данной работе для модификации алгоритмов реконструкции фотонов необходимо было изучить способы запуска отдельных частиц и удостовериться, что модель электромагнитного калориметра дает адекватные

результаты при реконструкции фотонов. Для этого были написаны макросы на языке C++, которые позволяют запускать частицы с заданными параметрами и анализировать результаты их взаимодействия с калориметром. В среде SPDR00T были проведены тестовые запуски фотонов 1 GeV 10000 событий перпендикулярно баррелю калориметра. В результате была построена гистограмма распределения энергии, зарегистрированной калориметром, суммарно по всем реконструированным частицам.

## Gen vs Reco Energy for Gamma

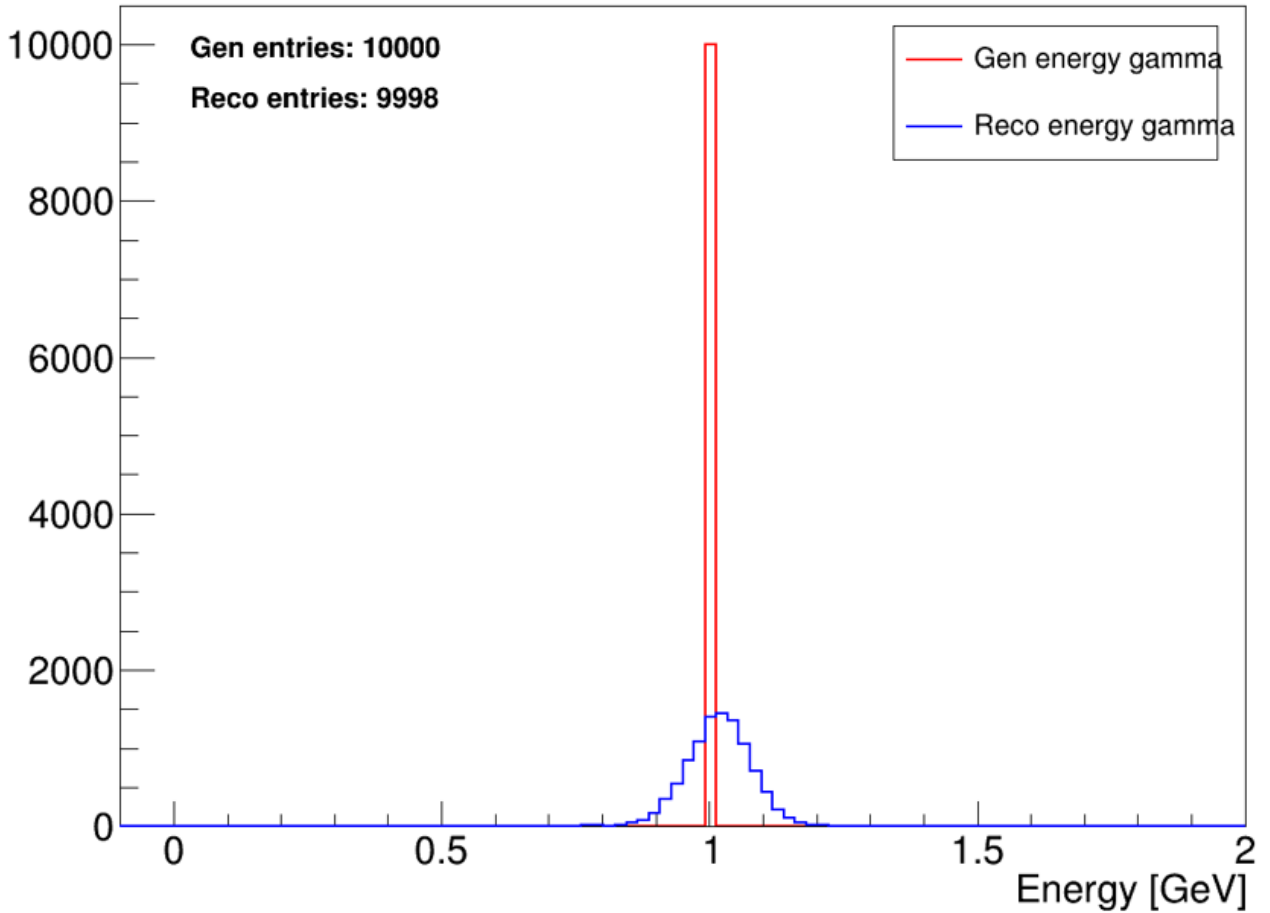


Рисунок 6 — Распределение энергии, зарегистрированной калориметром при запуске фотонов 1 GeV перпендикулярно баррелю калориметра

В результате была получена гистограмма с пиком около 1 GeV, что свидетельствует о корректной работе модели электромагнитного калориметра в среде SPDR00T и адекватной реконструкции фотонов. Процент фотонов, зарегистрированных калориметром (которые оставили в модели хотя бы какую-то энергию), составил около 99.98%, что является хорошим показателем эффективности детектора.

## 2.2 Изучение текущих алгоритмов кластеризации и идентификации частиц в электромагнитном калориметре SPD

На момент написания данной работы в среде SPDR00T реализованы базовые алгоритмы кластеризации и проведены первые исследования по разделению кластеров, соответствующих фотонам, от кластеров, соответствующих  $\pi^0$  [5].

Кластер формируется следующим образом: берётся любой хит (срабатывание ячейки - минимального детектирующего элемента электромагнитного калориметра) с энергией выше порога, объявляется началом кластера, после чего в этот кластер добавляются все остальные хиты, которые находятся достаточно близко хотя бы к одному уже добавленному хиту. Процесс повторяется, пока добавлять больше нечего.

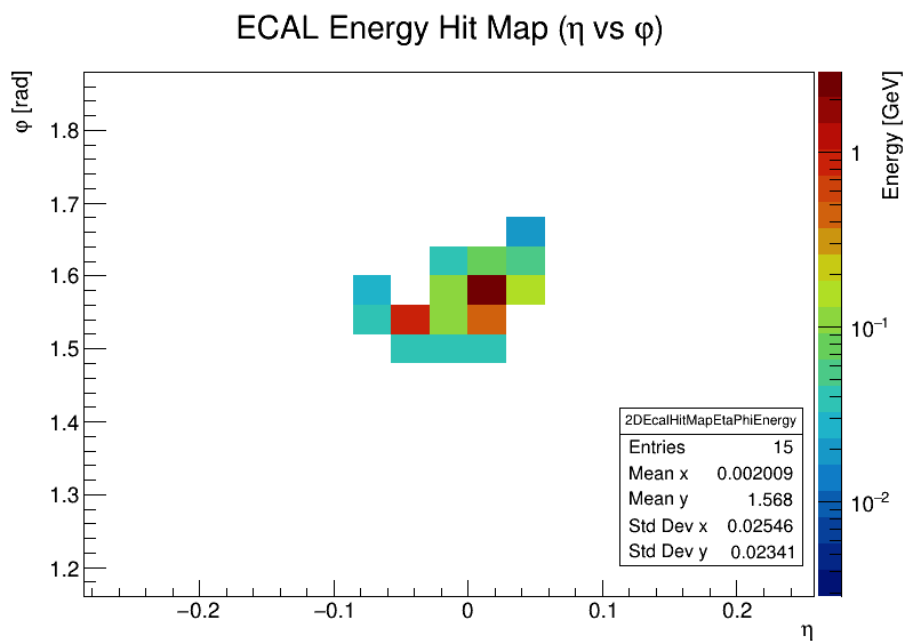


Рисунок 7 — Пример кластера от  $\pi^0$  при энергии 5 GeV

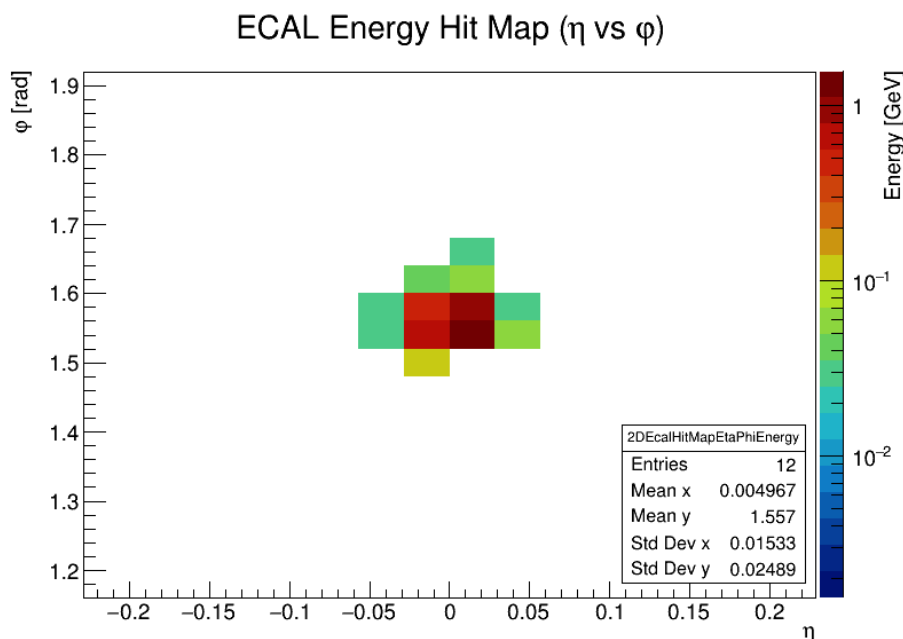


Рисунок 8 — Пример кластера  $\gamma$  при энергии 5 GeV

Разделение кластеров фотонов и  $\pi^0$  проводится на основе параметров зависящих от формы распределения энергии внутри кластера. [5]

Далее используется MLP нейронная сеть для классификации кластеров на фотонные и  $\pi^0$ -кластеры. В текущей версии алгоритмов разделения были достигнуты следующие результаты: около 90%  $\pi^0$  режекции при 80%  $\gamma$  эффективности [6].

В данной же работе планируется исследовать другие подходы к разделению кластеров с целью улучшения текущих показателей.

### 3. МОДИФИКАЦИЯ АЛГОРИТМОВ ИДЕНТИФИКАЦИИ ЧАСТИЦ В ЭЛЕКТРОМАГНИТНОМ КАЛОРИМЕТРЕ SPD

#### 3.1 Анализ параметров из эксперимента ATLAS на применимость к идентификации частиц в электромагнитном калориметре SPD

Задача фотонной идентификации уже решалась в эксперименте ATLAS, в котором были использованы следующие параметры [7]:

**Покрывтие детектора.** Кандидаты в фотоны должны удовлетворять условию

$$|\eta| < 2.37,$$

за исключением переходной области калориметра

$$1.37 < |\eta| < 1.52.$$

**Адронная утечка.** Доля поперечной энергии, утекающей в адронный калориметр, используется для подавления адронного фона. В зависимости от псевдобыстроты определены две переменные:

- $R_{\text{had1}}$ : отношение поперечной энергии, выделенной в первом слое адронного калориметра, к поперечной энергии электромагнитного кластера; используется в диапазонах

$$|\eta| < 0.8 \quad \text{и} \quad |\eta| > 1.37;$$

- $R_{\text{had}}$ : отношение поперечной энергии, выделенной во всём адронном калориметре, к поперечной энергии электромагнитного кластера; используется в диапазоне

$$0.8 < |\eta| < 1.37.$$

**Переменные второго слоя калориметра.** Переменные, характеризующие поперечное развитие электромагнитного ливня во втором слое калориметра:

- $R_\eta$ : отношение энергии, выделенной в окне  $3 \times 7$  ячеек, к энергии в окне  $7 \times 7$  ячеек в направлении  $\eta$ ;
- $w_2$ : поперечная ширина электромагнитного ливня во втором слое калориметра;
- $R_\phi$ : отношение энергии, выделенной в окне  $3 \times 3$  ячеек, к энергии в окне  $3 \times 7$  ячеек в направлении  $\phi$ .

**Переменные strip-слоя калориметра.** Переменные, использующие тонкую сегментацию первого (strip) слоя калориметра и особенно чувствительные к распадам  $\pi^0 \rightarrow \gamma\gamma$ :

- $w_{s3}$ : ширина ливня, вычисленная по трём стрипам, центрированным на стрипе с максимальной энергией;
- $w_{\text{stot}}$ : полная поперечная ширина ливня в strip-слое;
- $F_{\text{side}}$ : доля энергии, выделенной вне ядра трёх центральных стрипов, но внутри окна из семи стрипов;
- $\Delta E$ : разность между энергией второго локального максимума в strip-слое и минимальной энергией, реконструированной в стрипе между первым и вторым максимумами;
- $E_{\text{ratio}}$ : 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 strip-слое.

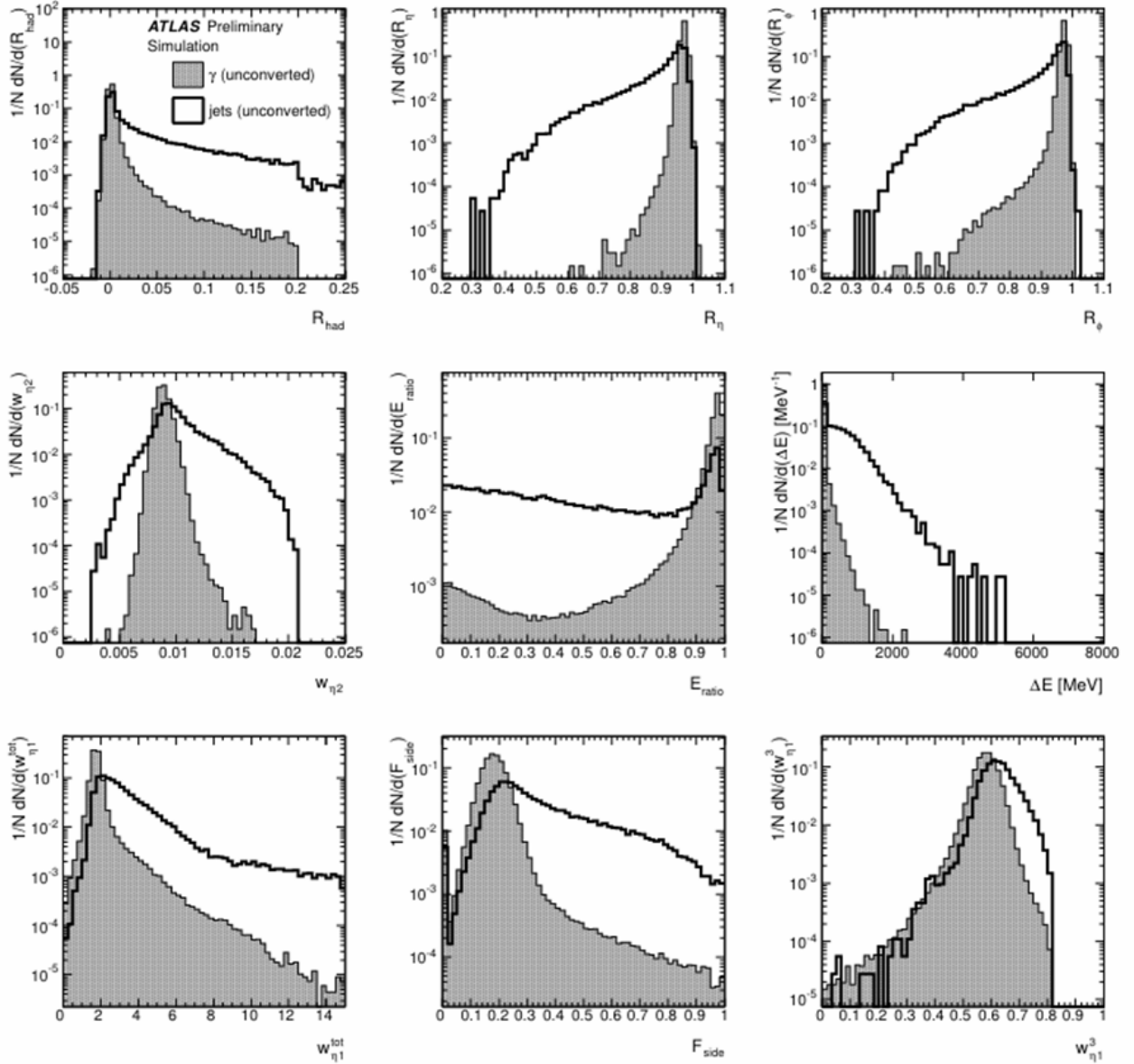


Рисунок 9 — Распределения параметров кластеров в эксперименте ATLAS

Поскольку электромагнитный калориметр SPD имеет отличную от ATLAS конструкцию и геометрию, не все перечисленные параметры могут быть применены напрямую. Среди данных параметров были выбраны те, которые можно адаптировать к модели электромагнитного калориметра SPD. Были реализованы функции для вычисления следующих адаптированных параметров кластеров:

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

- $w_{\eta 2}$ : 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.

Для оценки эффективности новых параметров были проведены тестовые запуски фотонов и  $\pi^0$  в среде SPDR00T с последующим вычислением адаптированных параметров для каждого кластера. Были построены гистограммы распределений новых параметров для кластеров, соответствующих фотонам и  $\pi^0$ .

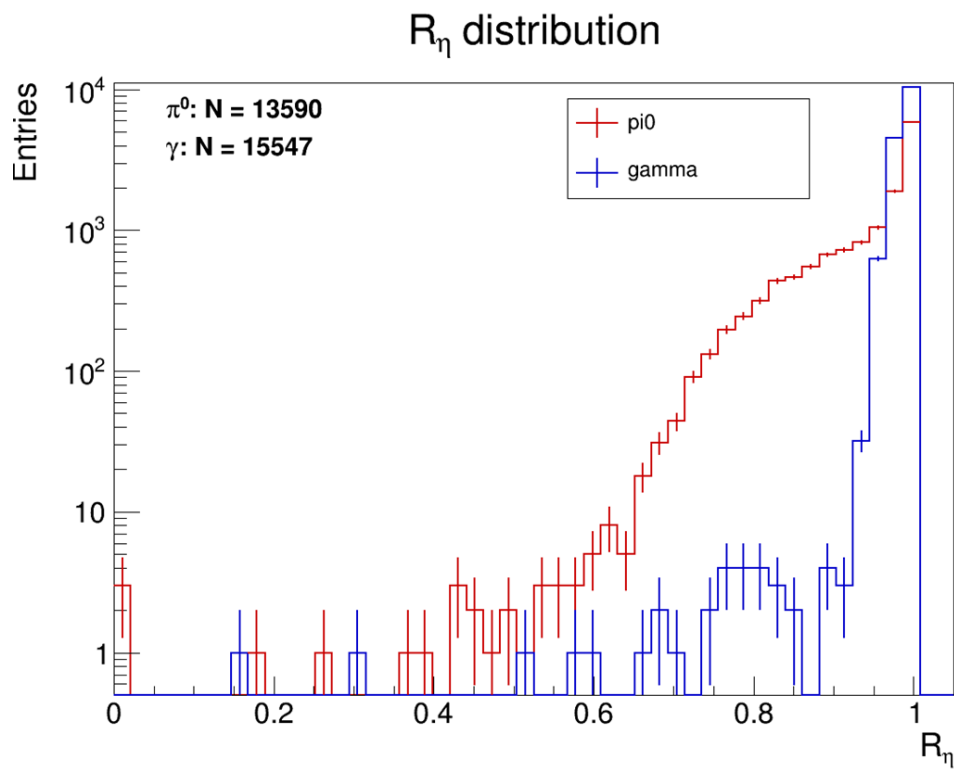


Рисунок 10 — Распределение адаптированного параметра  $R_\eta$ , 10000 событий  $\gamma$  (синим) и  $\pi^0$  (красным) при энергии 4 GeV

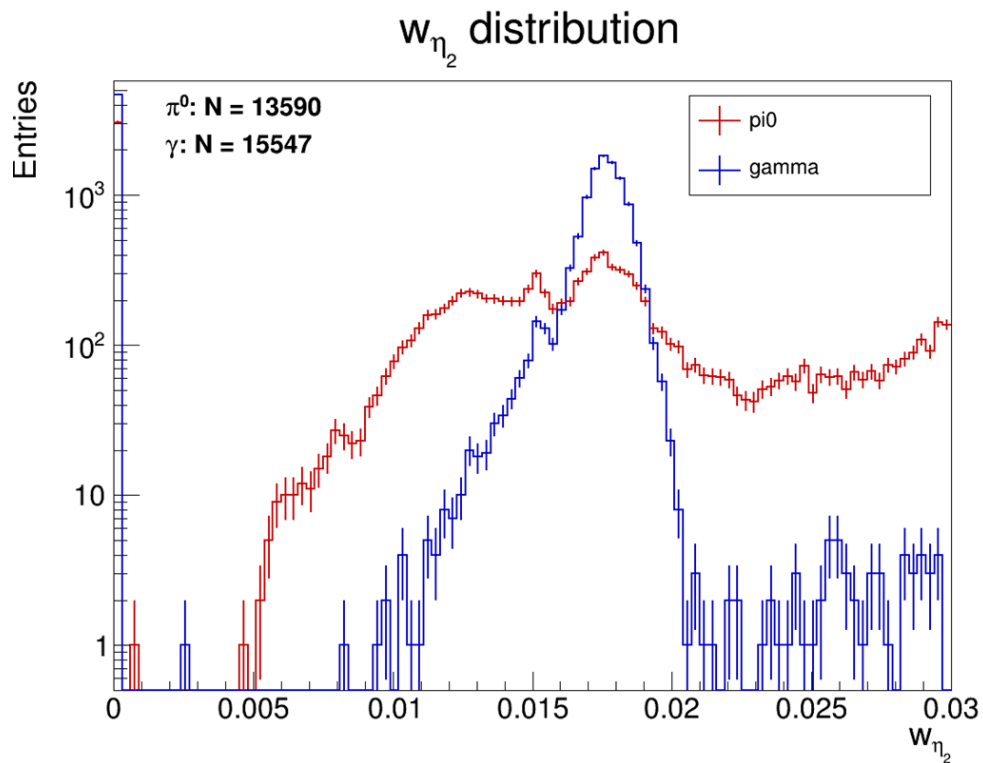


Рисунок 11 — Распределение адаптированного параметра  $w_{\eta_2}$ , 10000 событий  $\gamma$  (синим) и  $\pi^0$  (красным) при энергии 4 GeV

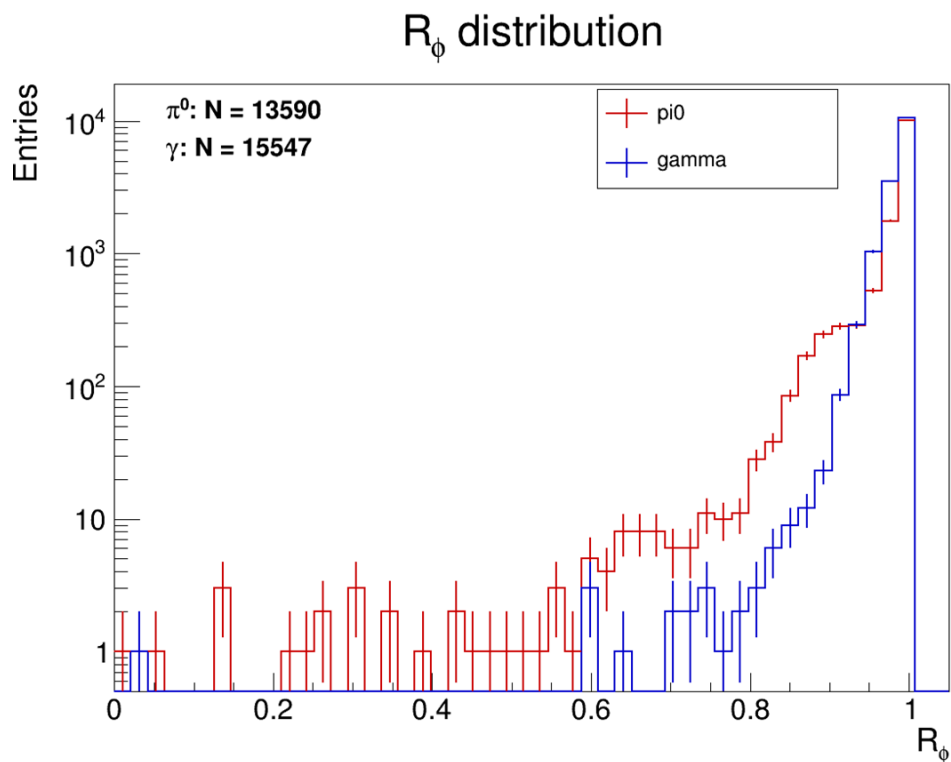


Рисунок 12 — Распределение адаптированного параметра  $R_\phi$ , 10000 событий  $\gamma$  (синим) и  $\pi^0$  (красным) при энергии 4 GeV

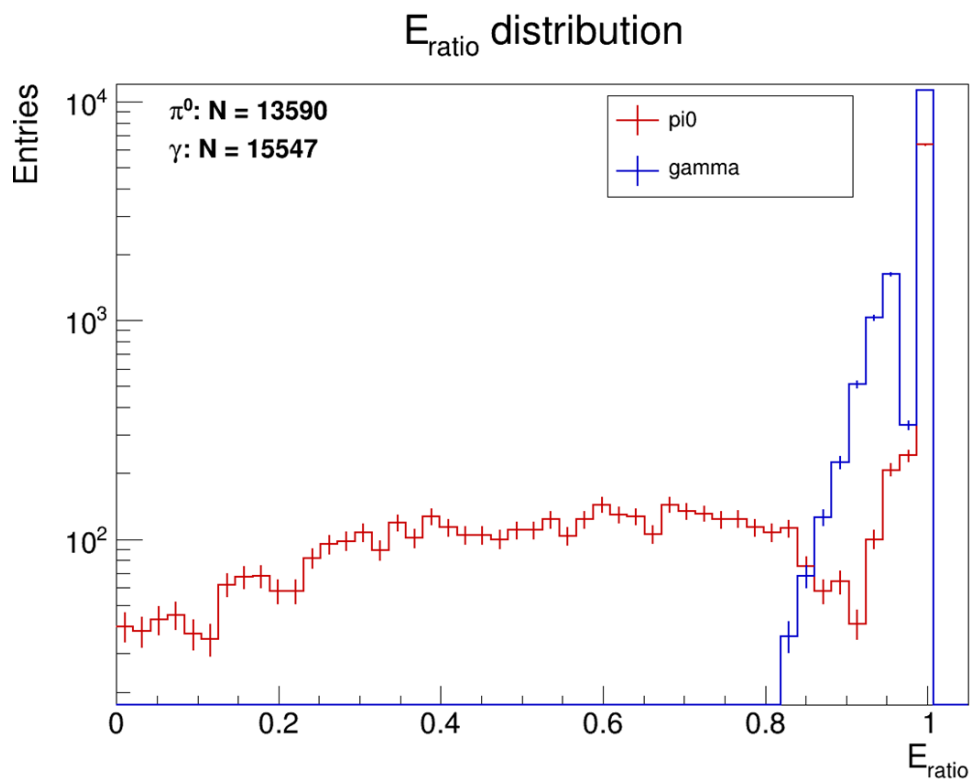


Рисунок 13 — Распределение адаптированного параметра  $E_{\text{ratio}}$ , 10000 событий  $\gamma$  (синим) и  $\pi^0$  (красным) при энергии 4 GeV

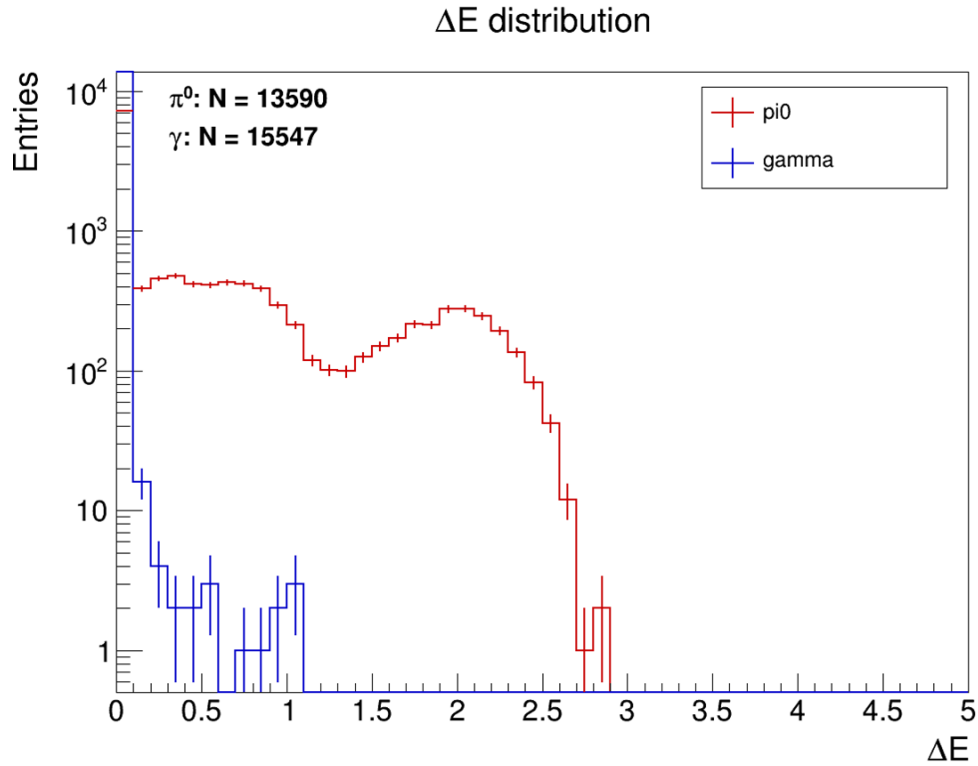


Рисунок 14 — Распределение адаптированного параметра  $\Delta E$ , 10000 событий  $\gamma$  (синим) и  $\pi^0$  (красным) при энергии 4 GeV

Анализ полученных распределений показал, что новые параметры демонстрируют различия между фотонными и  $\pi^0$ -кластерами, что свидетельствует об их потенциальной полезности для задачи разделения.

## 3.2 Сравнение эффективности различных алгоритмов классификации кластеров на простых выборках

### 3.2.1 Описание подходов к классификации кластеров

В данной работе были реализованы и протестированы несколько подходов к классификации кластеров в электромагнитном калориметре SPD с целью разделения фотонных и  $\pi^0$ -кластеров. Сравнивались следующие методы:

- **Метод фиксированных отборов (Rectangular Cuts).** Данный метод основан на применении прямоугольных пороговых отсеков по набору физических переменных кластера. Оптимальные значения порогов подбираются автоматически в TMVA путём перебора допустимых диапазонов параметров с целью максимизации режекции фона при заданной эффективности. Метод является интерпретируемым, но не учитывает корреляции между переменными. Для выбора наилучшего набора порогов используется метрика  $EffSel$  — эффективность селекции сигнала на тренировочной

выборке, которая максимизирует выделение сигнала при заданном уровне подавления фона. [8]

- **MLP (Multi-Layer Perceptron)**. Многослойный перцептрон представляет собой нейросетевой классификатор, способный моделировать нелинейные зависимости между входными переменными. Сеть состоит из входного слоя, одного или нескольких скрытых слоев с нелинейными активационными функциями и выходного слоя, который выдаёт вероятность принадлежности к классу. (данный метод уже был реализован в текущей версии алгоритмов идентификации кластеров в SPDROOT и показал хорошие результаты [6]).
- **BDT (Boosted Decision Trees)**. Метод основан на ансамбле решающих деревьев, обучаемых с использованием бустинга (AdaBoost). Каждое дерево выполняет последовательные бинарные разбиения пространства признаков, а итоговое решение формируется как взвешенная сумма ответов отдельных деревьев. BDT эффективно учитывает корреляции между переменными, устойчив к выбросам и, как правило, демонстрирует высокую разделяющую способность при анализе многомерных данных. Для подбора оптимального разбиения в каждом узле используется *Gini Index*:

$$\text{Gini} = 2 \cdot p_s \cdot (1 - p_s),$$

где  $p_s$  — доля сигналов в узле. Деревья строятся так, чтобы суммарно максимизировать разделение сигнал/фон по всей тренировочной выборке. [8]

### 3.2.2 Обучение и тестирование BDT-классификатора с использованием адаптированных параметров кластеров

Были построены корреляционные матрицы для адаптированных параметров кластеров, которые показали наличие корреляций между некоторыми из них. Но в целом корреляции были не слишком сильными (в частности некоторые параметры которые коррелировали для одного типа кластеров не коррелировали для другого), что позволило использовать все параметры в качестве входных для BDT-классификатора.

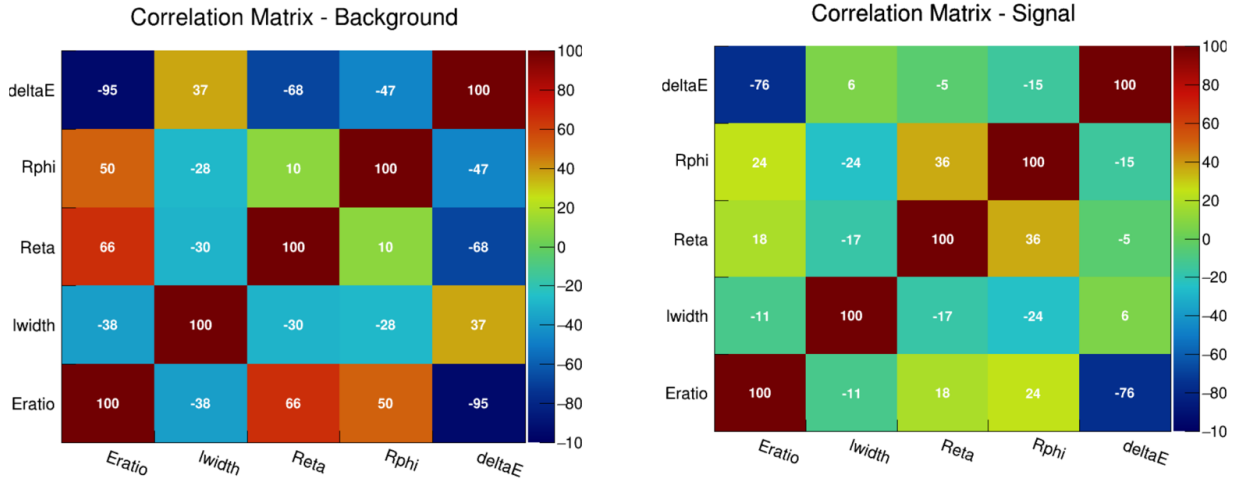


Рисунок 15 — Корреляционные матрицы адаптированных параметров кластеров

Для обучения BDT-классификатора была подготовлена выборка из 10000 событий  $\gamma$  и 10000 событий  $\pi^0$  при энергии 2, 4, 8 GeV. 70% выборки было использовано для обучения, 30% — для тестирования. Был проведен анализ гиперпараметров BDT с целью выбора оптимальных настроек модели. Была проведена сеточная оптимизация по следующим гиперпараметрам:

- Количество деревьев в ансамбле (NTrees) (от 100 до 1600 с шагом 200);
- Максимальная глубина каждого дерева (MaxDepth) (от 2 до 5 с шагом 1);
- Скорость обучения (Beta) (0.3, 0.5, 0.8)
- Количество разбиений по каждому признаку при построении дерева (nCuts) (20, 40).

Подбор лучших гиперпараметров и дальнейший анализ проводился на основе метрики ROC AUC на тестовой выборке. Данная метрика показывает площадь под ROC-кривой и является стандартным показателем качества бинарной классификации.

На основе проведённого анализа были выбраны оптимальные гиперпараметры для каждого энергетического диапазона. Лучшие ROC-кривые и BDT scores для каждого энергетического диапазона приведены ниже.

Из ROC-кривых видно, что BDT-классификатор тем хуже разделяет фотонные и  $\pi^0$ -кластеры, чем выше энергия частиц. Данные результаты согласуются с ожиданиями и предыдущими исследованиями в данной области. [5] Это вероятно связано с тем, что при увеличении энергии расстояние между фотонами от распада  $\pi^0$  уменьшается, и максимумы начинают сливаться, что затрудняет их разделение.

Сравнение эффективности BDT-классификатора с данными параметрами с текущим MLP-классификатором, реализованным в SPDR00T, показало, что BDT демонстрирует сопоставимые результаты на 2GeV, но на 4GeV и 8GeV эффективность BDT-классификатора сильно снижается.

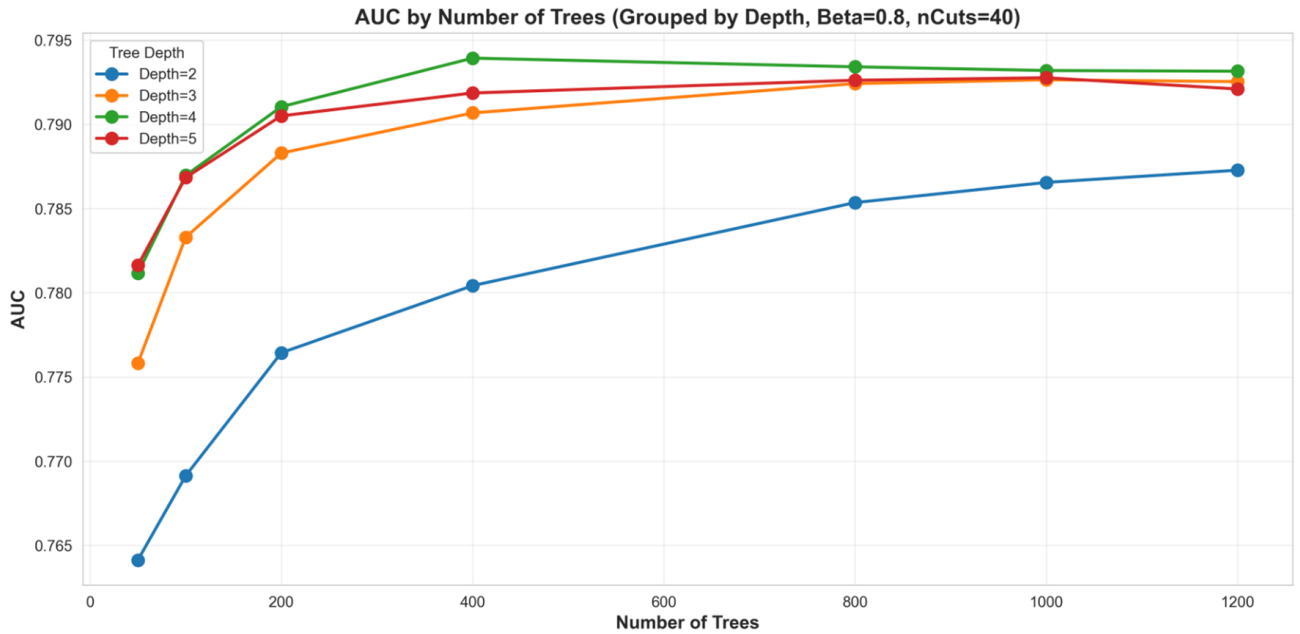


Рисунок 16 — Пример анализа гиперпараметров (переменные из эксперимента ATLAS, 8 GeV)

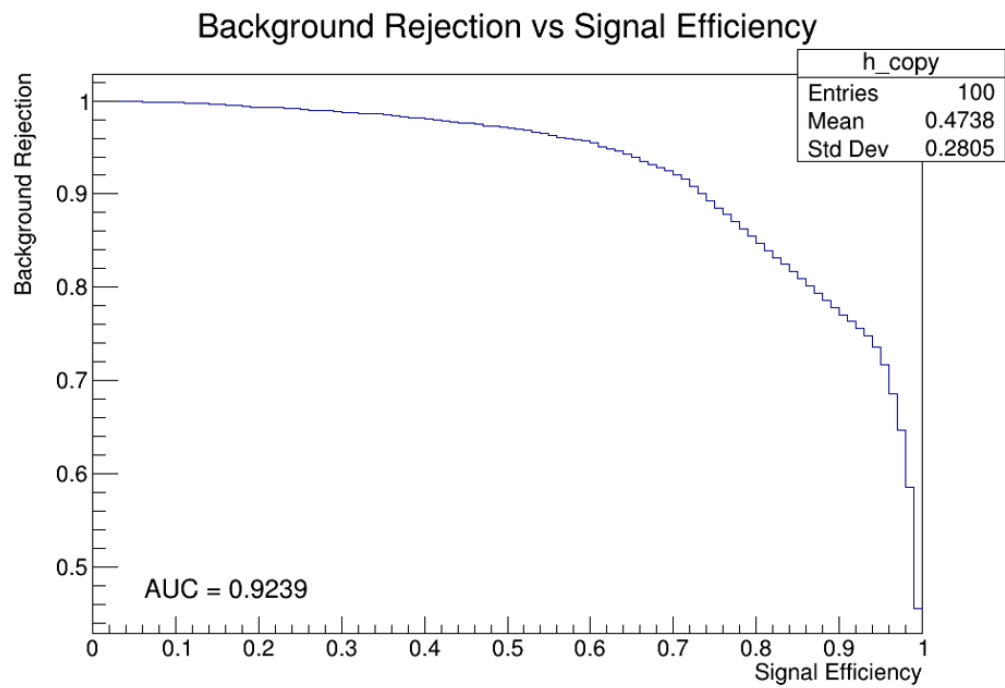


Рисунок 17 — Лучшая ROC-кривая (переменные из эксперимента ATLAS, для 2 GeV)

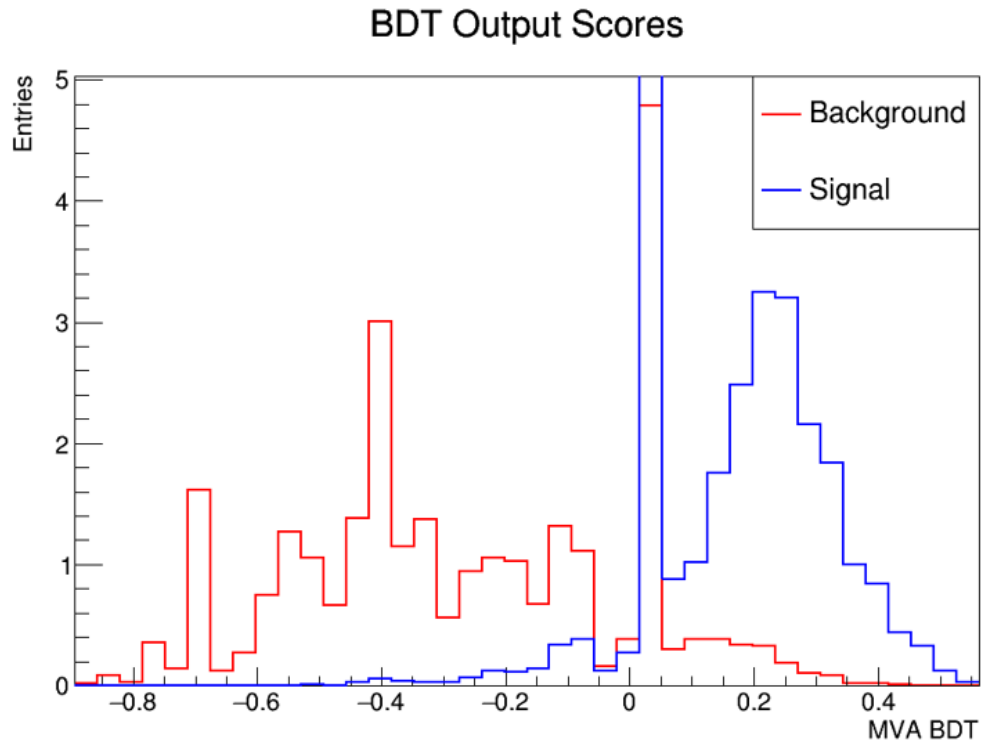


Рисунок 18 — BDT scores для лучшей ROC-кривой (переменные из эксперимента ATLAS, для 2 GeV)

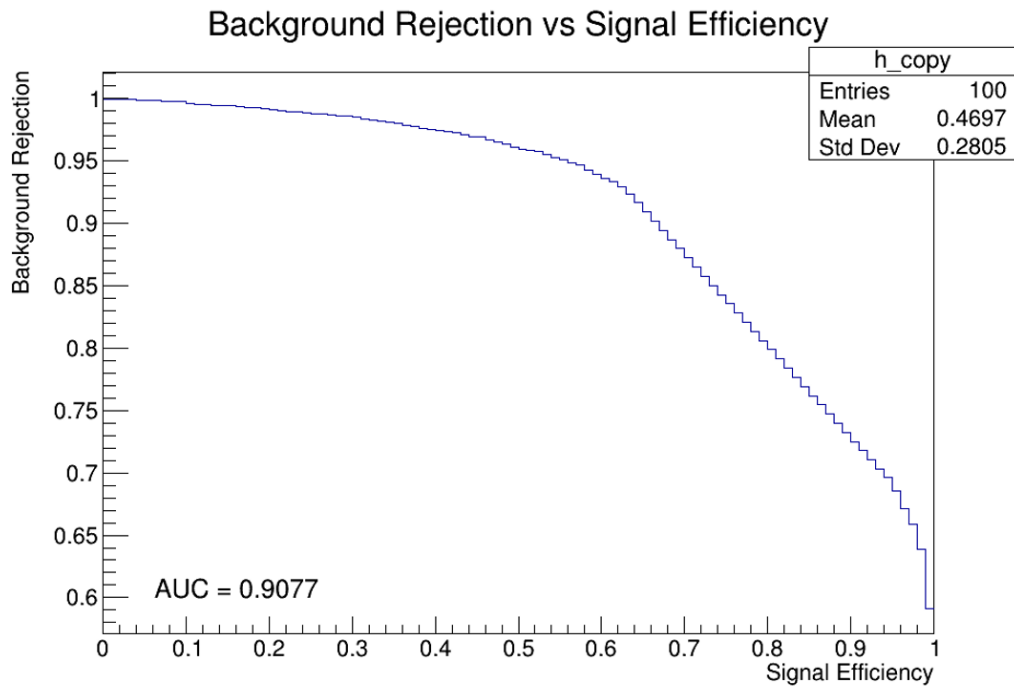


Рисунок 19 — Лучшая ROC-кривая (переменные из эксперимента ATLAS, для 4 GeV)

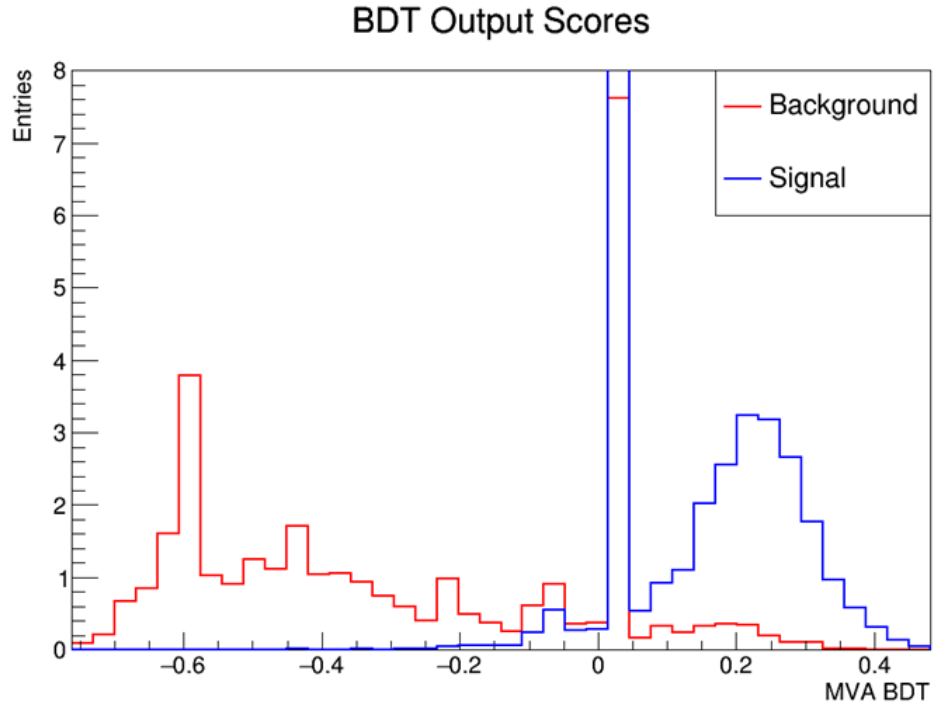


Рисунок 20 — BDT scores для лучшей ROC-кривой (переменные из эксперимента ATLAS, для 4 GeV)

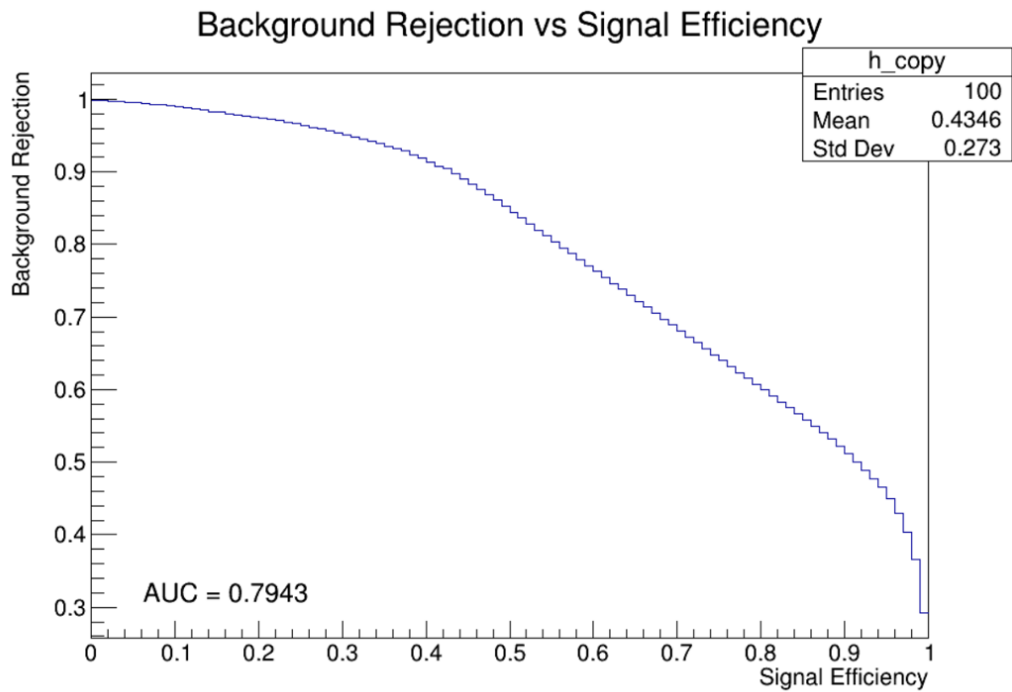


Рисунок 21 — Лучшая ROC-кривая (переменные из эксперимента ATLAS, для 8 GeV)

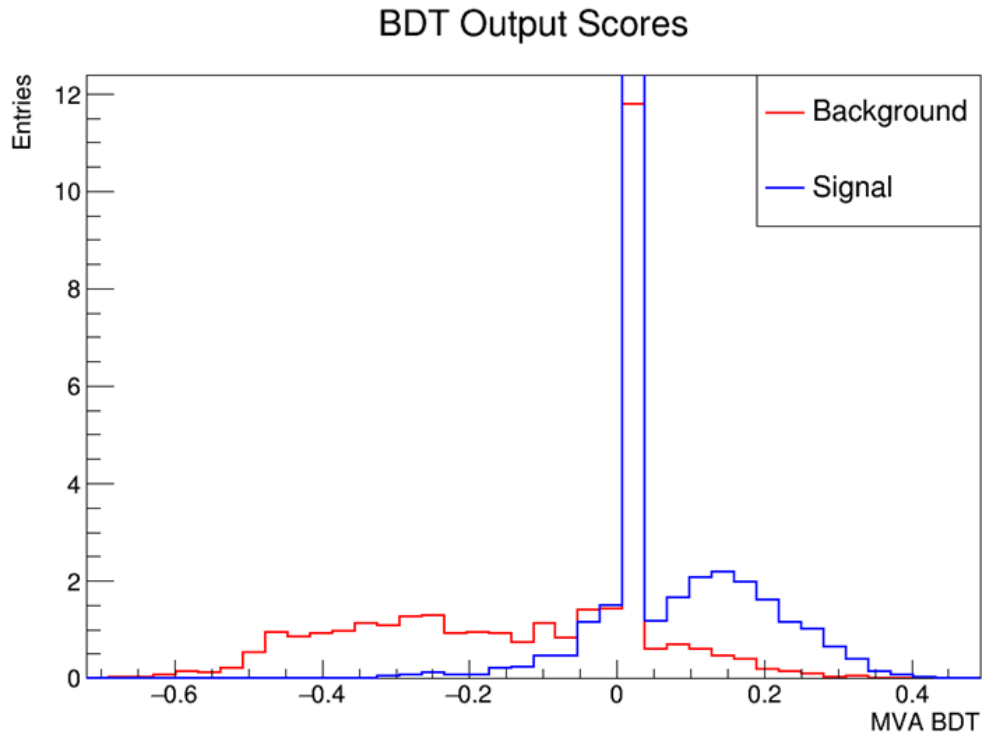
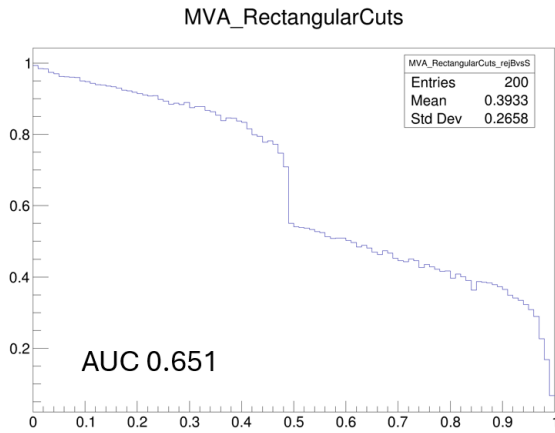


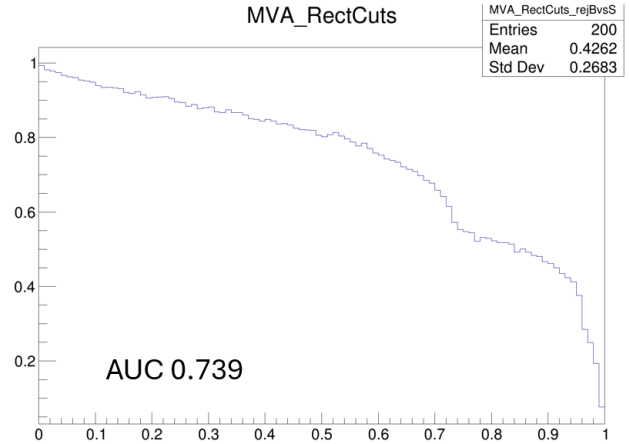
Рисунок 22 — BDT scores для лучшей ROC-кривой (переменные из эксперимента ATLAS, для 8 GeV)

### 3.2.3 Классификация методом фиксированных отборов с использованием адаптированных параметров кластеров

Для сравнения с BDT-классификатором был реализован классификатор, основанный на фиксированных отборах, с использованием тех же адаптированных параметров кластеров и той же выборки. С помощью TMVA были подобраны оптимальные пороговые значения для каждого параметра с целью максимизации режекции фона при заданной эффективности. В результате была построена ROC-кривая для классификатора, основанного на фиксированных отборах.



До удаления кластеров из 1 ячейки



После удаления кластеров состоящих из 1 ячейки

Рисунок 23 — ROC-кривые классификатора, основанного на фиксированных отборах (переменные из эксперимента ATLAS, для 8 GeV)

Анализ ROC-кривой показал присутствие резкого порога в режекции фона при достижении определённой эффективности сигнала (рисунок слева). Была высказана гипотеза, что это связано с кластерами, состоящими из одного хита (ячейки калориметра). Такие кластеры имеют специфические значения параметров, что может приводить к резкому изменению эффективности при прохождении определённых порогов. Для проверки этой гипотезы была проведена дополнительная классификация с исключением кластеров из одного хита (рисунок справа). При исключении таких кластеров резкий порог в ROC-кривой исчезает, что подтверждает гипотезу. Сложно представить себе модель разделяющую такие кластеры, поскольку они не несут информации о форме распределения энергии. Данные кластеры должны рассматриваться либо как шум, либо как особый класс событий, требующий отдельного анализа. При этом при исключении таких кластеров общая эффективность классификации значительно возрастает. Поэтому в дальнейшем такие кластеры будут исключаться из анализа. В последствии так же определено, что пики на BDT scores на рисунках 18, 20, 22 связаны именно с кластерами из одного хита.

В сравнении с BDT-классификатором классификатор, основанный на фиксированных отборах, показал результаты значительно хуже: так на 8 GeV при эффективности сигнала 80% режекция фона составила всего около 40%, в то время как у BDT он был около 60% (сравнение до исключения кластеров из одного хита).

### 3.2.4 Использование всей совокупности параметров из ATLAS и текущего алгоритма классификации в SPDROOT для классификации кластеров с помощью BDT

Для дальнейшего улучшения классификации кластеров был проведён анализ с использованием всей совокупности параметров из ATLAS и текущего алгоритма классификации в SPDROOT (MLP). Были использованы все параметры, представленные в SPDROOT. При этом были внесены небольшие незначительные изменения, упрощающие их вычисление с использованием уже реализованных функций. Подробное описание адаптации переменных  $((x, y)$  соответствуют  $(\eta, \varphi)$ ):

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

- **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):**

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

где

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

Были построены корреляционные матрицы для полного набора параметров кластеров. Анализ показал наличие значительных корреляций между некоторыми параметрами, что может повлиять на эффективность классификации. В дальнейших исследованиях будет рассмотрено влияние этих корреляций на работу BDT-классификатора.

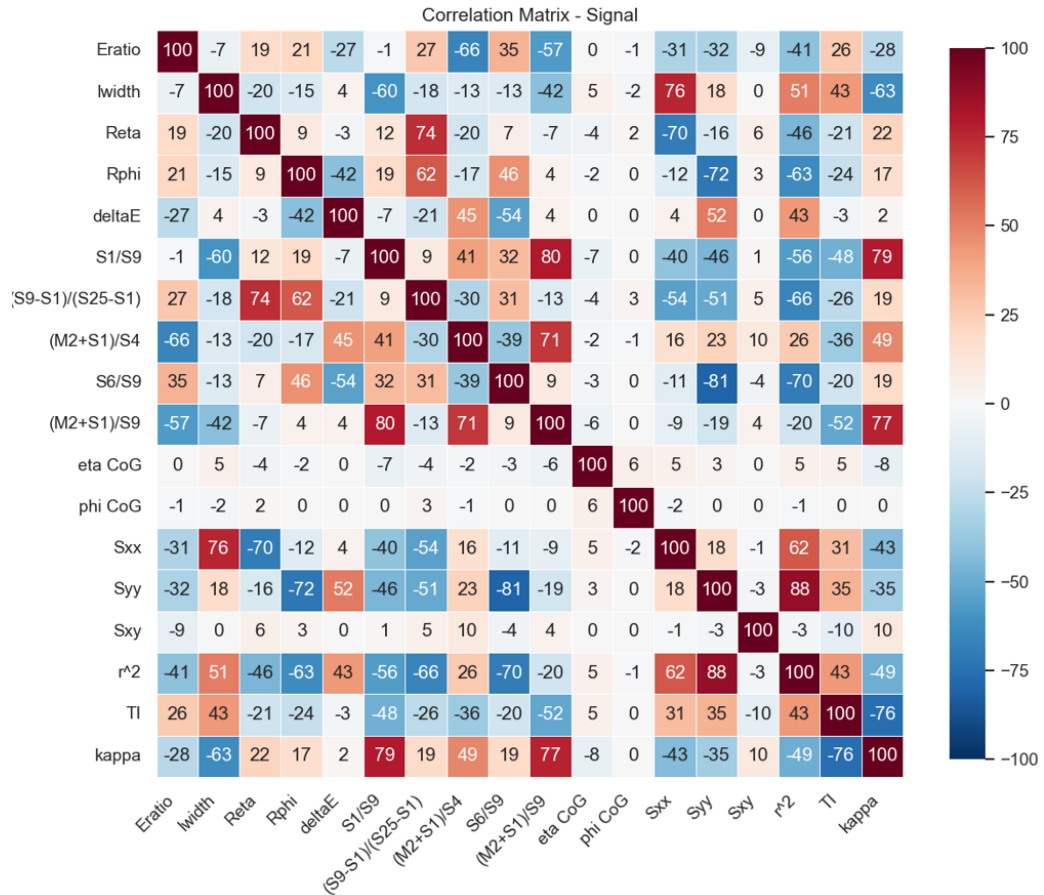


Рисунок 24 — Корреляционная матрица полного набора параметров кластеров для сигналов (фотоны)

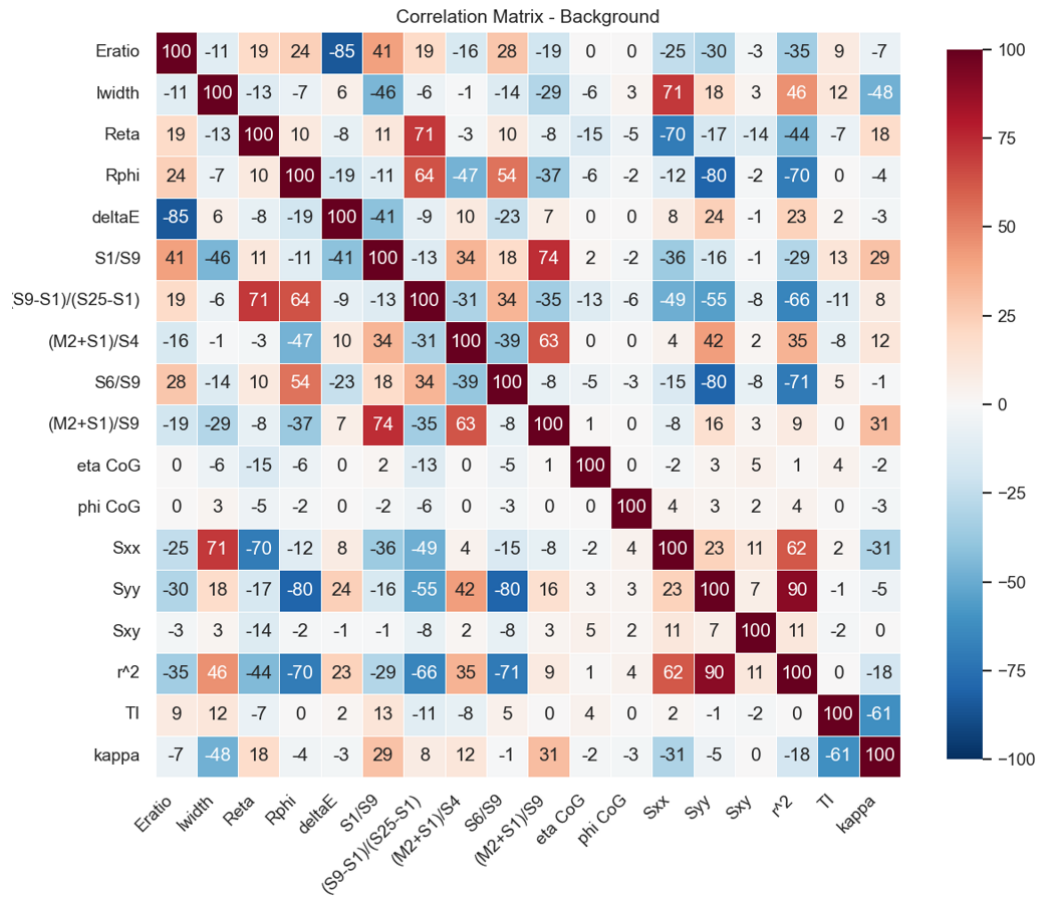


Рисунок 25 — Корреляционная матрица полного набора параметров кластеров для фона ( $\pi^0$ )

Далее был обучен BDT-классификатор с использованием полного набора параметров кластеров на той же выборке для 2, 4 и 8 GeV. Были проведены оптимизации гиперпараметров аналогично предыдущему разделу. Лучшие ROC-кривые и BDT scores для каждого энергетического диапазона приведены ниже.

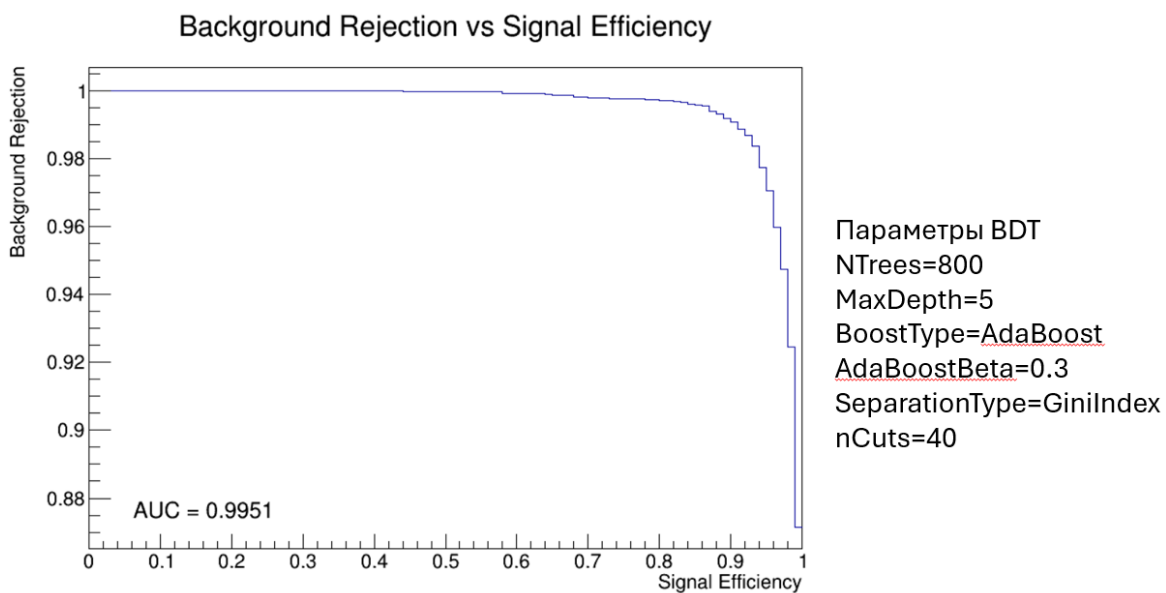


Рисунок 26 — Лучшая ROC-кривая (все параметры, для 2 GeV)

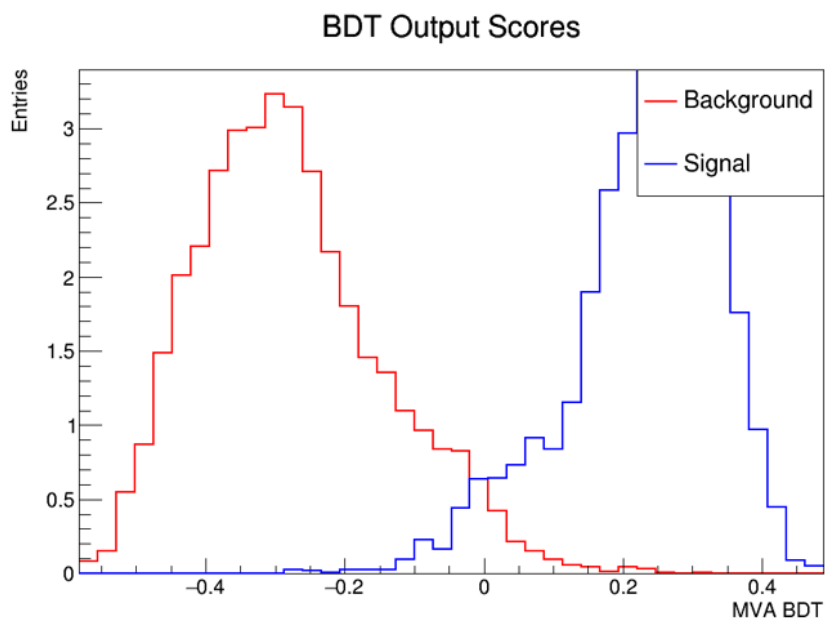


Рисунок 27 — BDT scores для лучшей ROC-кривой (все параметры, для 2 GeV)

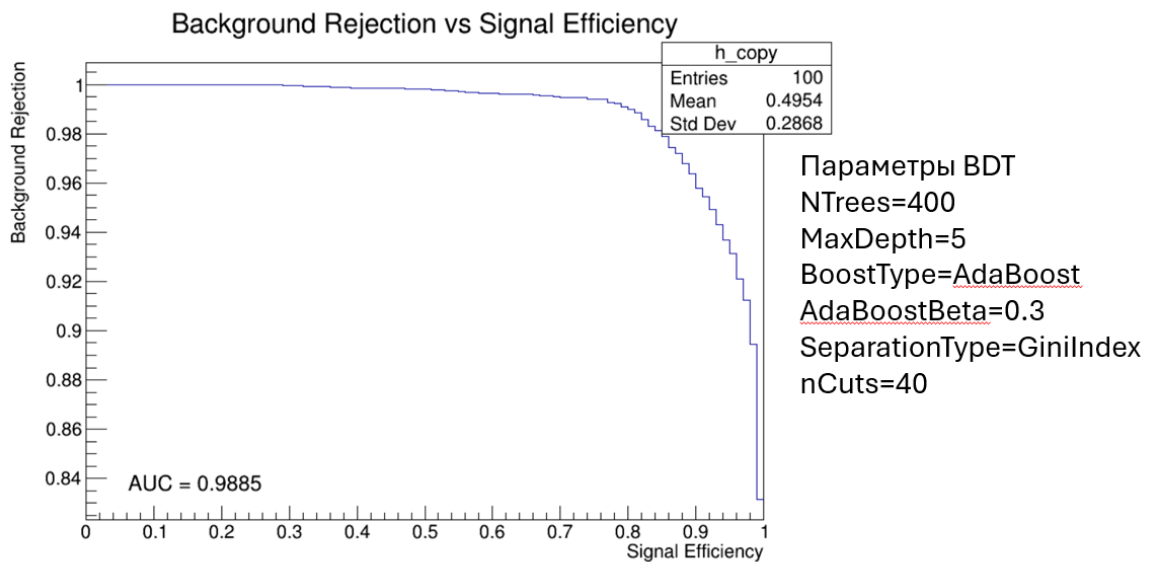


Рисунок 28 — Лучшая ROC-кривая (все параметры, для 4 GeV)

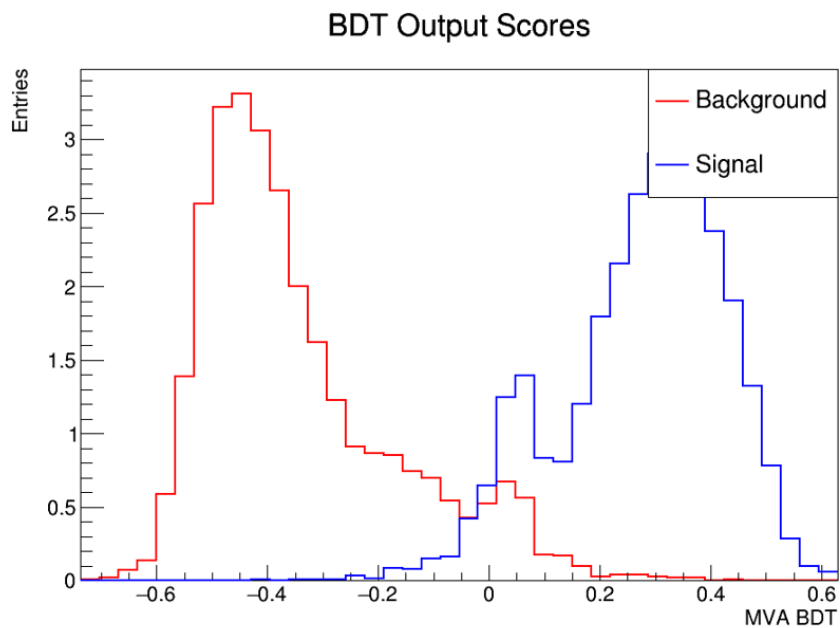


Рисунок 29 — BDT scores для лучшей ROC-кривой (все параметры, для 4 GeV)

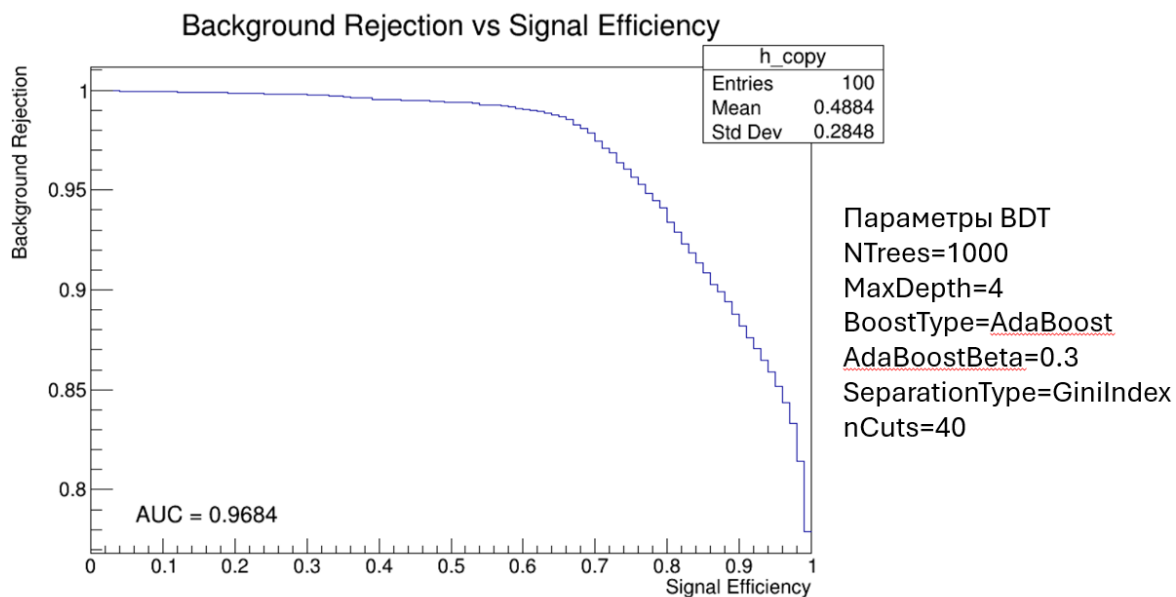


Рисунок 30 — Лучшая ROC-кривая (все параметры, для 8 GeV)

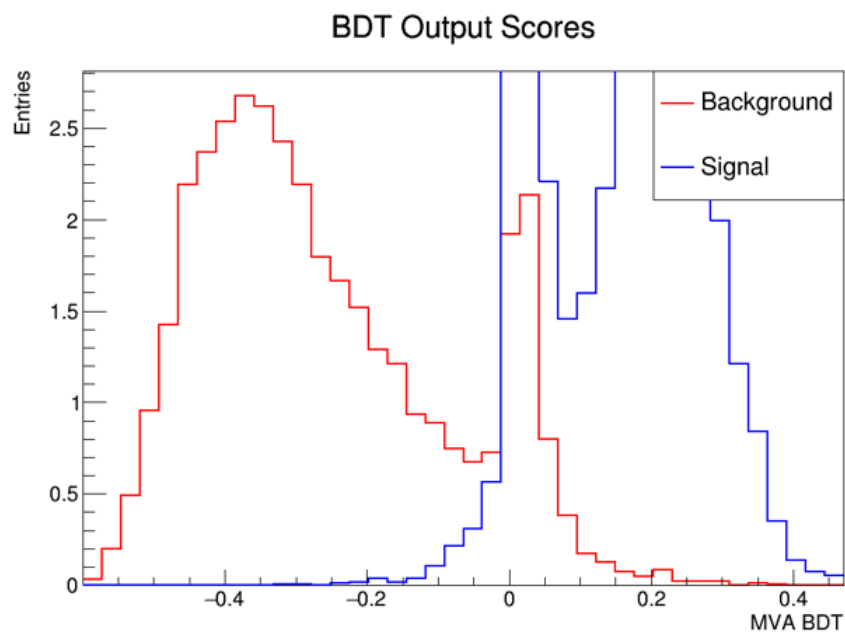


Рисунок 31 — BDT scores для лучшей ROC-кривой (все параметры, для 8 GeV)

Из анализа ROC-кривых можно заключить, что данный BDT-классификатор показывает улучшенные результаты по сравнению с предыдущими версиями, особенно на высоких энергиях (4 и 8 GeV). Для эффективности сигнала около 80% режекция фона достигает примерно 95% на 8 GeV и 99.5% на 2 GeV, что значительно лучше предыдущих результатов. Данный результат значительно превосходит MLP-классификатор, реализованный в SPDR00T, и данный подход может быть положен в основу дальнейших улучшений алгоритмов идентификации кластеров в электромагнитном калориметре SPD.

### 3.3 Ранжирование переменных полного набора параметров

Изначально было проведёно первичное ранжирование переменных для полного набора параметров по метрикам separation и variable importance [8]. Результаты приведены ниже.

```
-----  
: Rank : Variable : Separation  
-----  
: 1 : Sxy 5x5 : 2.868e-01  
: 2 : S1/S9 : 2.267e-01  
: 3 : Shape variable kappa : 2.241e-01  
: 4 : (M2+S1)/S9 : 2.224e-01  
: 5 : Sxx 5x5 : 2.146e-01  
: 6 : lwidth : 1.792e-01  
: 7 : Syy 5x5 : 1.748e-01  
: 8 : S6/S9 : 1.560e-01  
: 9 : (M2+S1)/S4 : 1.332e-01  
: 10 : Reta : 1.262e-01  
: 11 : Tails importance r2/r4 : 1.042e-01  
: 12 : (S9-S1)/(S25-S1) : 9.598e-02  
: 13 : Rphi : 8.876e-02  
: 14 : Eratio : 5.640e-02  
: 15 : r^2 : 5.439e-02  
: 16 : phi CoG 5x5 : 3.566e-02  
: 17 : deltaE : 2.884e-02  
: 18 : eta CoG 5x5 : 2.320e-02  
-----
```

Рисунок 32 — Ранжирование переменных полного набора параметров (для 8 GeV) по метрике separation [8]

Rank	Variable	Variable Importance
1	kappa	8.121e-02
2	Sxy25	7.127e-02
3	s1_divided_s9	7.047e-02
4	Sxx25	6.329e-02
5	phiCog25	6.120e-02
6	m2_plus_s1_over_s9	6.096e-02
7	tailsImportance	6.076e-02
8	Reta	6.001e-02
9	s6_over_s9	5.938e-02
10	Syy25	5.858e-02
11	m2_plus_s1_over_s4	5.845e-02
12	lwidth	5.557e-02
13	Rphi	5.134e-02
14	r2	5.102e-02
15	s9_minus_s1_over_s25_minus_s1	5.063e-02
16	eratio	4.299e-02
17	etaCog25	3.509e-02
18	deltaE	7.788e-03

Рисунок 33 — Ранжирование переменных полного набора параметров (для 8 GeV) по метрике variable importance [8]

Как видно из анализа важности переменных, значительный вклад в классификацию вносят как адаптированные параметры из ATLAS, так и текущие параметры из SPDR00T.

Далее был проведен более подробный анализ важности переменных, основанный на методе N-1: проводилось сравнение метрики ROC AUC для моделей обученных на N-1 переменных с перебором по всем возможным комбинациям. Наихудшая переменная на каждой итерации исключалась. Ниже представлено влияние исключения каждой переменной по сравнению с предыдущей итерацией.

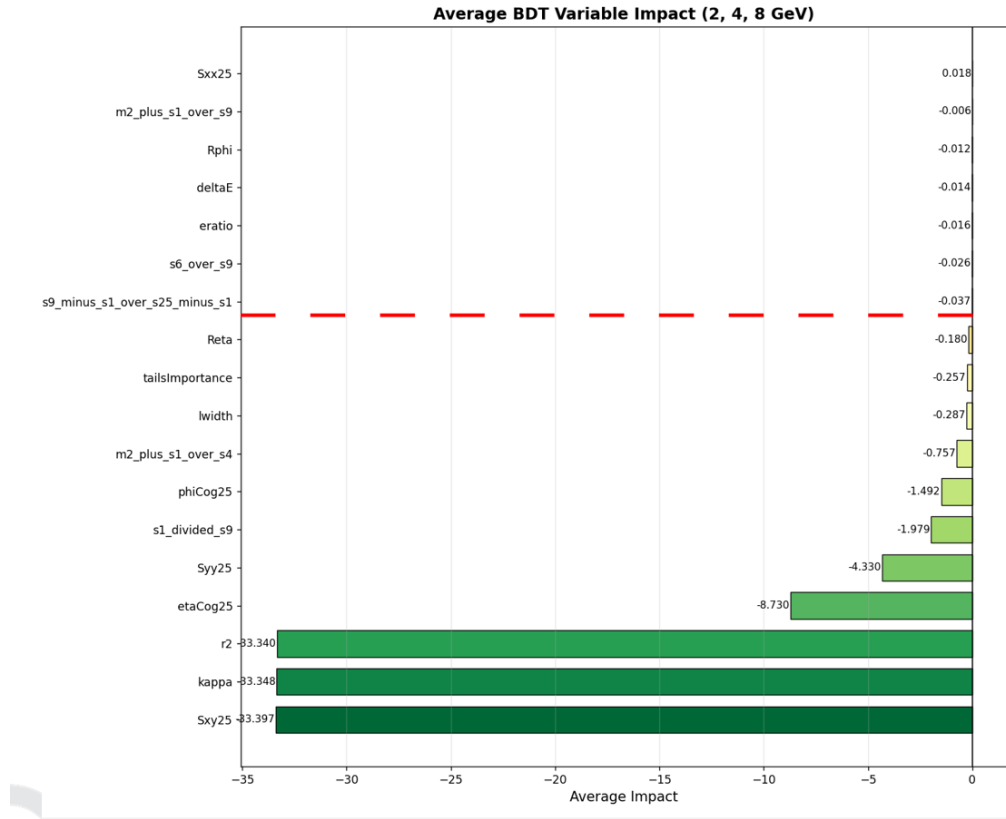


Рисунок 34 — Влияние исключения каждой переменной по сравнению с предыдущей итерацией

В итоге на основе анализа был проведён отбор переменных для каждого из классификаторов:

- Для классификатора фиксированных отборов отобраны переменные:  $\kappa$ ,  $R_\phi$ ,  $R_\eta$ ,  $w_{\eta 2}$ ,  $\Delta E$ ,  $S_6/S_9$ .
- Для BDT-классификатора, напротив, были отброшены переменные:  $E_{\text{ratio}}$ ,  $\Delta E$ ,  $R_\phi$ ,  $(M_2 + S_1)/S_4$ ,  $S_{XX}$ ,  $S_6/S_9$ ,  $(S_9 - S_1)/(S_{25} - S_1)$ .

### 3.4 Тестирование классификаторов на более реалистичных выборках

#### 3.4.1 Уточнение постановки задачи

При работе с более реалистичными выборками кластеры разных подтипов от  $\pi^0$  и  $\gamma$  могут давать совершенно различные распределения параметров, что негативно влияет на качество классификаторов. В связи с этим возникает необходимость уточнить постановку задачи.

Для кластеров от  $\pi^0$  существует 2 принципиальных случая (другие типы кластеров маловероятны либо будут идентифицироваться другими детекторами):

- Фотоны от  $\pi^0$  попали в один кластер (условное обозначение  $[\pi^0, \gamma, \gamma]$ ): тогда такой кластер имеет специфическую форму
- Фотоны от  $\pi^0$  попали в разные кластеры (условное обозначение  $[\pi^0, \gamma]$ ): такие кластеры необходимо идентифицировать из кинематики

В дальнейшем в качестве фона рассмотрены кластеры 1го типа, именно такие кластеры отделялись от «фотонных кандидатов»: кластеров 2го типа и кластеров от прямых фотонов (для прямых фотонов брались только кластеры, в которых был один фотон).

### 3.4.2 Качество классификаторов по мере приближения выборки к более реалистичному случаю

Тестирование проводилось в 3 этапа:

- запуск одиночных  $\gamma/\pi^0$  перпендикулярно баррелю с равномерным распределением по Pt от 0 до 10 GeV и отбором на энергию кластера  $>1$  GeV; выборка (train) сигнала: 13867 кластеров; выборка (train) фона: 7457 кластеров (далее этап 1)
- запуск одиночных  $\gamma/\pi^0$  с равномерным распределением по телесному углу и равномерным распределением по Pt от 0 до 10 GeV и отбором на энергию кластера  $>1$  GeV; выборка (train) сигнала: 10139 кластеров; выборка (train) фона: 7149 кластеров (далее этап 2)
- minbias pp столкновения 27 GeV и с отбором на энергию кластера  $>1$  GeV; выборка (train) сигнала: 29746 кластеров; выборка (train) фона: 119319 кластеров (далее этап 3)

Результаты тестирования классификаторов на более реалистичных выборках приведены на рисунках ниже:

Background Rejection vs Signal Efficiency

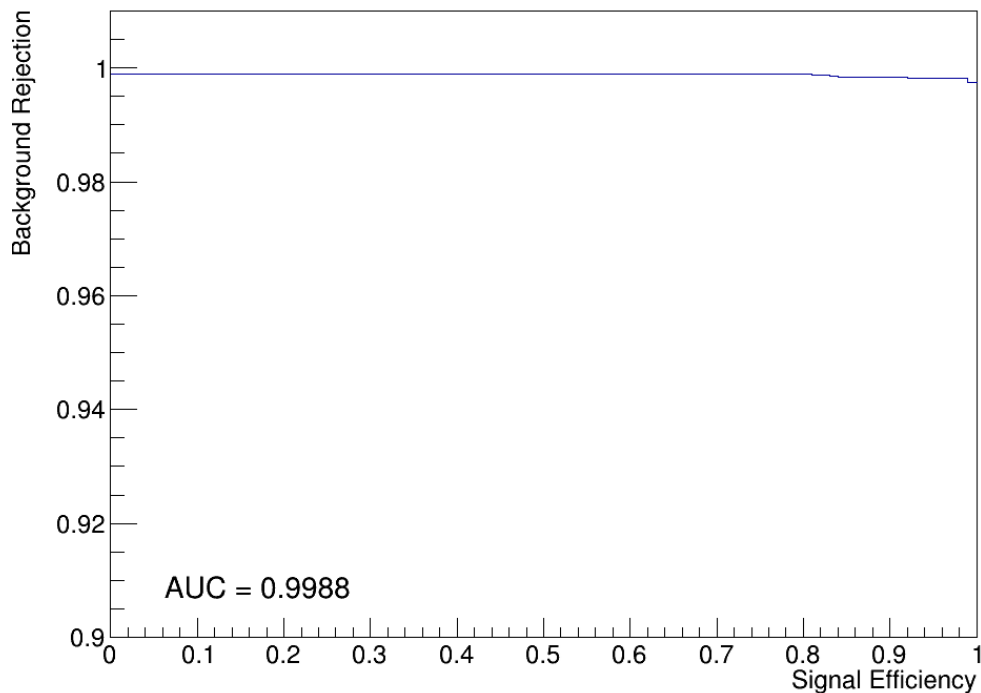


Рисунок 35 — ROC-кривая BDT-классификатора (с параметрами BoostType = grad, nTrees = 800, MaxDepth = 5, nCuts = 40) для этапа 1

Background Rejection vs Signal Efficiency

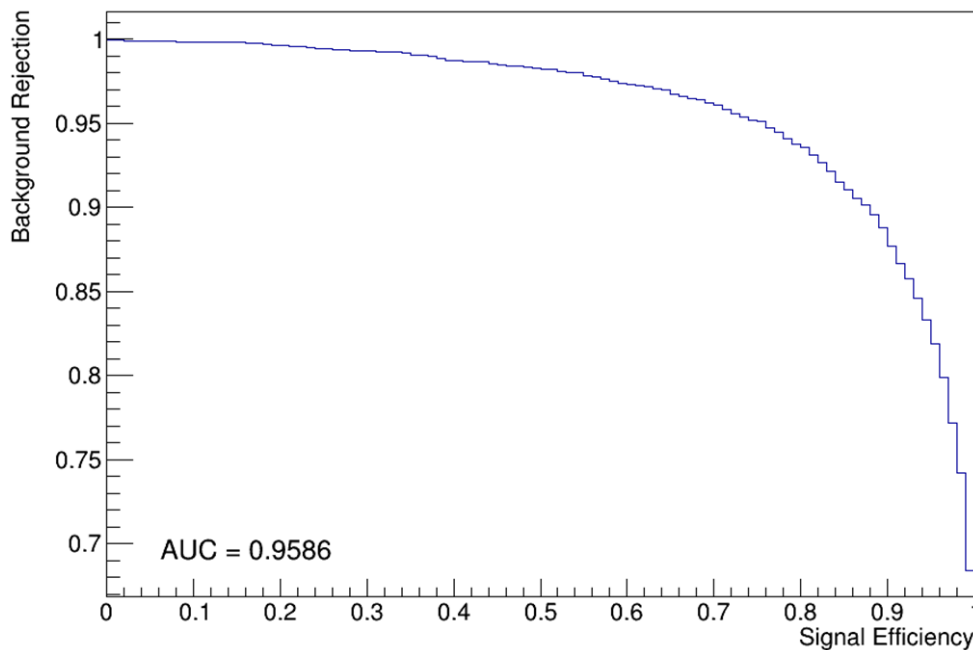


Рисунок 36 — ROC-кривая BDT-классификатора (с параметрами BoostType = grad, nTrees = 800, MaxDepth = 5, nCuts = 40) для этапа 2

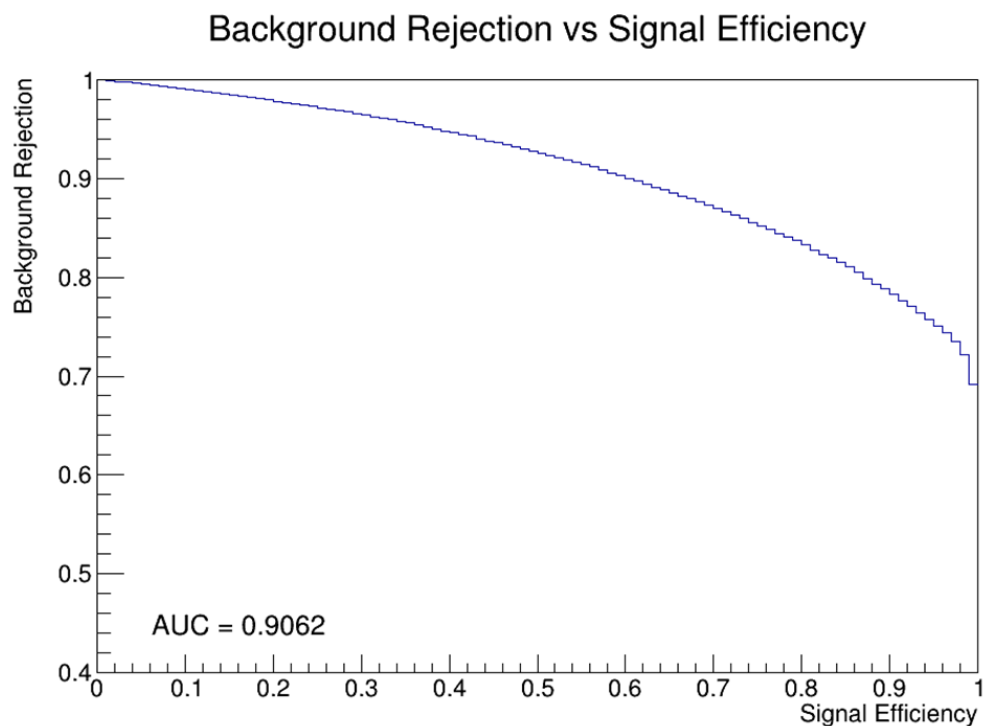


Рисунок 37 — ROC-кривая BDT-классификатора (с параметрами BoostType = grad, nTrees = 800, MaxDepth = 5, nCuts = 40) для этапа 3

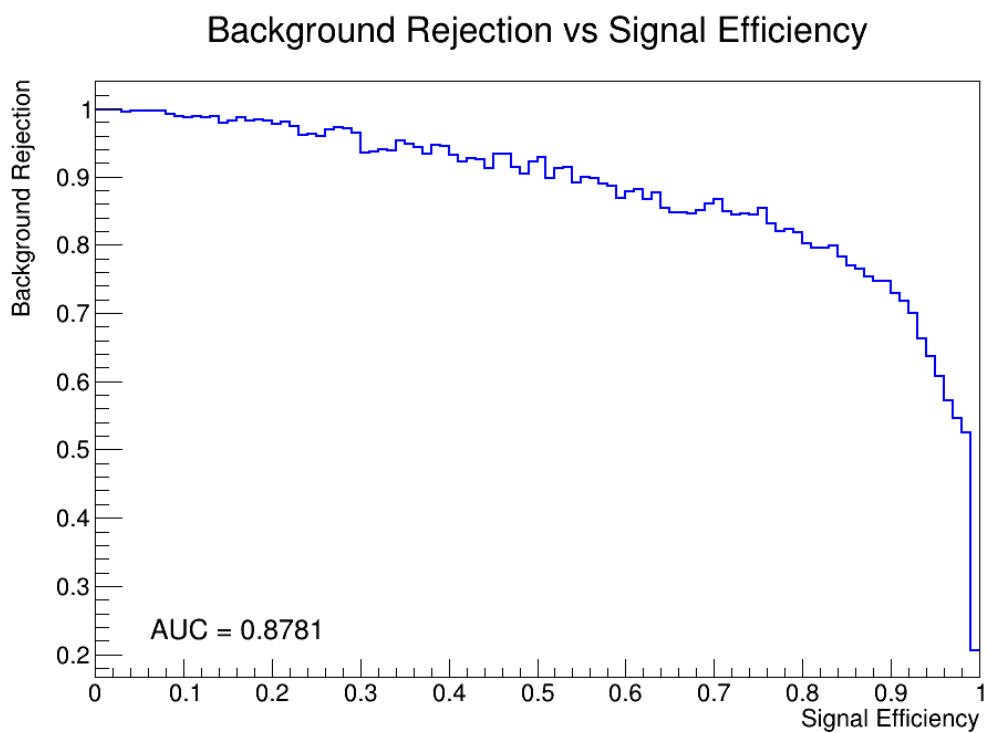


Рисунок 38 — ROC-кривая cut-based-классификатора (genetic algorithm) для этапа 2, отборы на переменные:  $w_{\eta 2} \leq 0.022$ ;  $R_{\eta} \geq 0.473$ ;  $R_{\phi} \geq 0.648$ ;  $\Delta E \leq 0.066$ ;  $S_6/S_9 \geq 0.968$ ;  $\kappa \leq 0.517$

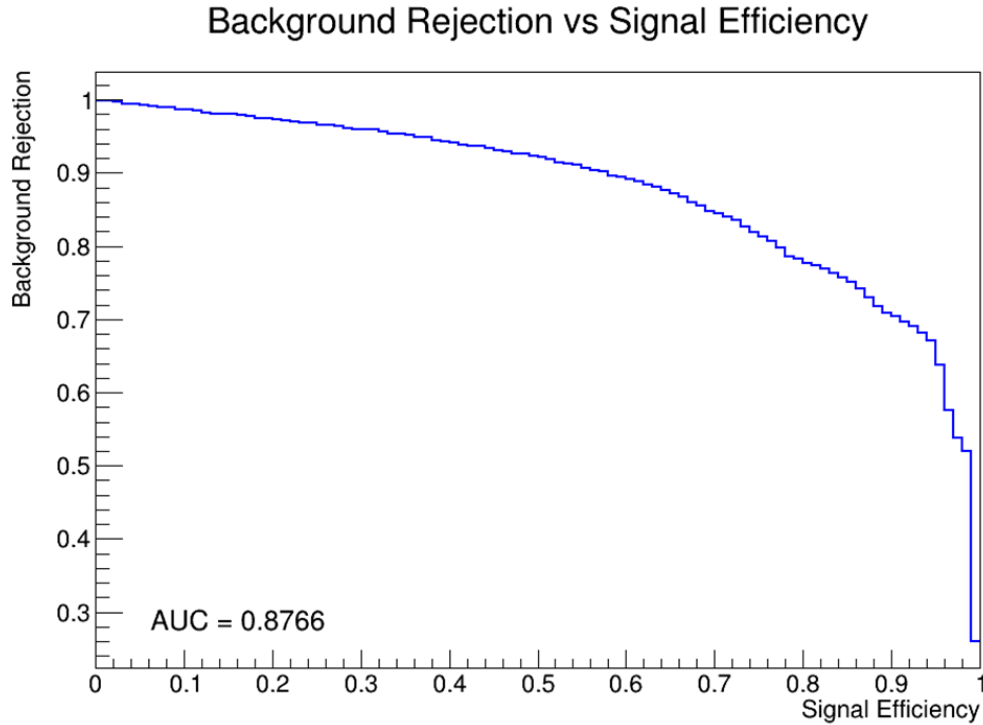


Рисунок 39 — ROC-кривая cut-based-классификатора (genetic algorithm) для этапа 3, отборы на переменные  $w_{\eta 2} \geq -2.049 \times 10^{-4}$ ;  $R_{\eta} \geq 0.137$ ;  $R_{\phi} \geq 0.771$ ;  $\Delta E \leq 0.100$ ;  $S_6/S_9 \geq -3.483 \times 10^{-3}$ ;  $\kappa \leq 1.007$

Итоговое сравнение метрик различных моделей на различных этапах тестирования представлено ниже:

Таблица 1 — Сравнение метрик классификаторов

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

Можно заключить, что в одинаковых условиях BDT классификатор дает лучший результат в сравнении с MLP-классификатором и cut-based-классификатором. Но важно учитывать, что в случае BDT классификатора было использовано больше параметров кластеров, что может негативно влиять на погрешности классификатора.

### 3.4.3 Сравнение сценариев разделения

Также было проведено сравнение эффективности bdt-классификатора для сценариев разделения кластеров типа (кластеры сигнала/кластеры фона):

- $[\gamma], [\pi^0, \gamma]/[\pi^0, \gamma, \gamma]$  (в сигнал добавлены также "фотонные кандидаты" от  $\pi^0$ , которые должны будут исключаться другим алгоритмом);
- $[\gamma]/[\pi^0, \gamma, \gamma]$ .

Для первого сценария была получена ROC-кривая:

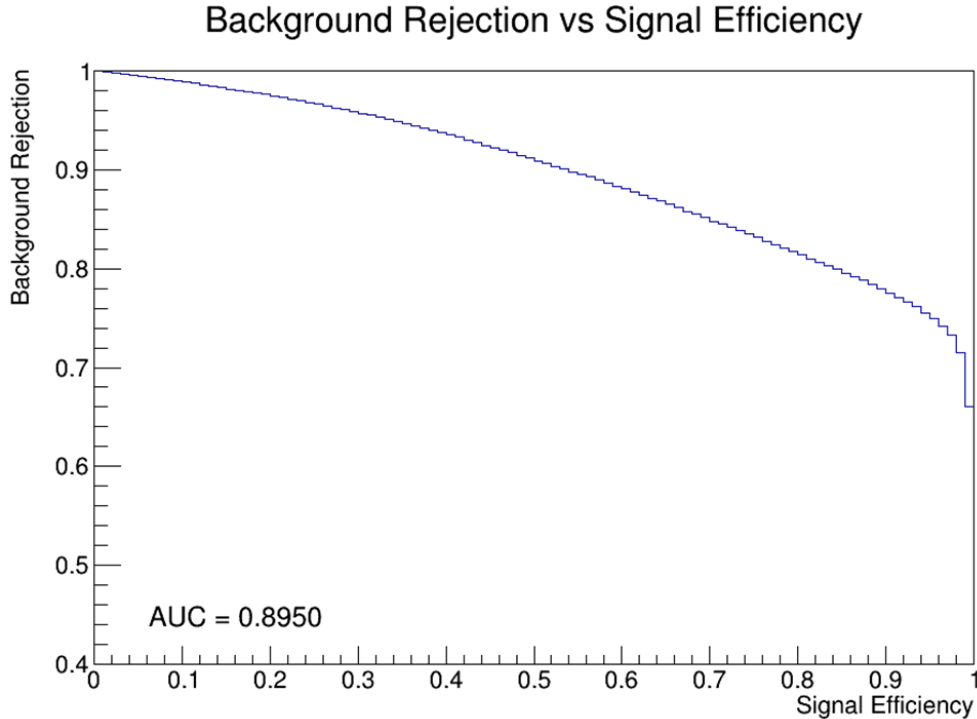


Рисунок 40 — ROC-кривая BDT-классификатора для сценария  $[\gamma], [\pi^0, \gamma]/[\pi^0, \gamma, \gamma]$

Для второго же сценария ROC-кривая была представлена ранее (рисунок 37). Как видно из ROC-кривых, добавление "фотонных кандидатов" практически не влияет на эффективность классификатора (влияние порядка 1%). Что ожидаемо, учитывая схожую форму кластеров от различных типов сигнала.

## 4. ЗАКЛЮЧЕНИЕ

В данной работе разрабатывались и тестировались алгоритмы идентификации фотонов в электромагнитном калориметре SPD.

### Основные результаты:

- Адаптированы параметры фотонной идентификации из эксперимента ATLAS к геометрии калориметра SPD и реализованы в анализе
- Изучены и адаптированы параметры, использованные ранее для решения данной задачи MLP-подходом, для совместного использования с переменными ATLAS

- Произведено ранжирование и отбор лучших параметров кластеров для различных классификаторов
- Выполнено сравнение методов классификации кластеров  $\gamma/\pi^0$ : метод фиксированных отборов, MLP и BDT в различных условиях и сценариях
- Показано, что BDT с расширенным набором параметров (ATLAS + SPDROOT) улучшает разделение при одинаковых условиях
- На minbias достигнуто лучшее значение ROC AUC 0.906, полученное BDT классификатором

**Дальнейшие планы:**

- Разработка алгоритмов идентификации других типов кластеров и других частиц
- Оценка возможности использования множества моделей в различных областях энергии и  $\eta$  для увеличения качества классификации
- Интеграция алгоритмов в основную цепочку реконструкции SPDROOT
- Исследование альтернативных ML-подходов для дальнейшего повышения качества фотонной идентификации

## СПИСОК ЛИТЕРАТУРЫ

- [1] *Spin Physics Detector Project*. URL: <https://spd.jinr.ru/>. (дата обр.: 26.12.2024).
- [2] V. M. Abazov и др. «Conceptual design of the Spin Physics Detector». В: (январь. 2021). arXiv: 2102.00442 [hep-ex].
- [3] A. Guskov и др. «Probing Gluons with the Future Spin Physics Detector». В: *Physics* 5 (2023), с. 672–687.
- [4] V. Abazov и др. «Technical Design Report of the Spin Physics Detector at NICA». В: (апрель. 2024). arXiv: 2404.08317 [hep-ex].
- [5] Andrei Maltsev. *Separation of photon clusters from neutral pion decay in ECAL: exploratory study*. URL: [https://indico.jinr.ru/event/5549/contributions/33072/attachments/23155/40918/SPD\\_PMC\\_20250903\\_Maltsev\\_ECAL\\_separation-1.pdf](https://indico.jinr.ru/event/5549/contributions/33072/attachments/23155/40918/SPD_PMC_20250903_Maltsev_ECAL_separation-1.pdf).
- [6] Andrei Maltsev. *Status of reconstruction in ECAL*. URL: [https://indico.jinr.ru/event/5532/contributions/33735/attachments/23799/41947/Maltsev\\_SPD\\_CM\\_2025.pdf](https://indico.jinr.ru/event/5532/contributions/33735/attachments/23799/41947/Maltsev_SPD_CM_2025.pdf).
- [7] ATLAS collaboration. «Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016». В: *The European Physical Journal C* 79 (2019).
- [8] K. Albertsson и др. *Toolkit for Multivariate Data Analysis with ROOT (TMVA)*. URL: <https://root.cern.ch/download/doc/tmva/TMVAUsersGuide.pdf>.