

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский ядерный университет «МИФИ»
(НИЯУ МИФИ)

УДК 539.120.71

ОТЧЁТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

Повышение чувствительности анализа в канале
распада бозона Хиггса в Z бозон и фотон с
применением машинного обучения

Научный руководитель,
к.ф.-м.н., доцент

_____ Е. Ю. Солдатов

Научный руководитель

_____ Н. В. Проклова

Студент

_____ К. К. Казакова

Москва 2021

Содержание

Введение	3
Цель работы	3
1 Устройство детектора ATLAS	4
1.1 Эксперимент ATLAS	4
1.1.1 Система координат детектора ATLAS	4
1.1.2 Внутренний детектор	5
1.1.3 Система калориметров детектора ATLAS	5
1.1.4 Мюонный спектрометр	6
1.1.5 Триггерная система	6
2 Механизм Хиггса	7
2.1 Механизмы образования бозона Хиггса	7
2.2 Канал распада $H \rightarrow Z\gamma$	7
3 Используемые наборы данных	9
3.1 Используемые данные и Монте-Карло данные	9
3.2 Критерии отбора данных	10
4 Категоризация	12
4.1 Результаты тренировки MLP	13
4.2 Результаты тренировки BDTG	16
4.3 Основная категоризация	19
4.4 Сравнение данных и фонового распределения	22
4.5 Другие способы категоризации	24
5 Моделирование сигнала	26
Заключение	30
Список используемых источников	32

Введение

Стандартная модель (СМ) - наиболее тщательно проработанная теория в физике элементарных частиц. То, почему частицы обладают массой, в СМ объясняется через механизм Хиггса[1]. Однако СМ не описывает гравитацию, Тёмную Энергию, Тёмную Материю и многие другие явления. Можно предположить, что СМ лишь часть более универсальной теории. Поэтому так важно наиболее обширно изучить все свойства бозона Хиггса – относительно недавно открытой частицы, и с максимальной точностью их описать для подтверждения СМ или же для открытия «новой физики». В данной работе исследуется редкий канал распада $H \rightarrow Z\gamma$, так как этот канал распада чувствителен к любым отклонениям от СМ из-за особенностей процесса распада. Так как в последствии рассматривается распад Z бозона только в два лептона (в пару электронов либо мюонов), данный распад бозона Хиггса представляет особый интерес из-за чёткого сигнала и относительно небольшого фона от протон-протонных столкновений.

Цель работы

Целью данной работы является увеличение чувствительности анализа путем создания категорий, в которых можно локально описывать сигнальные и фоновые процессы. Далее, на основе категорий необходимо провести фитирование смоделированного сигнала для проверки их применимости. Полученные в ходе работы результаты будут необходимы для анализа при переходе к описанию реальных данных с ускорителя на Большом Адронном Коллайдере (БАК), так как качество набранной статистики непосредственно влияет на вероятность открытия в данных. Также целью работы является проверить возможные способы оптимизации анализа с использованием различных методов машинного обучения из библиотеки TMVA (Toolkit for Multivariate Data Analysis)[2].

1 Устройство детектора ATLAS

1.1 Эксперимент ATLAS

Эксперимент ATLAS[3],[4] - это многоцелевой детектор (рис.1), предназначенный для исследования протон-протонных столкновений и столкновений тяжелых ионов, полученных с БАКа. Программа ATLAS предназначена для поиска бозона Хиггса и «новой физики», а так же для проверки КХД. Детектор радиально симметричен и состоит из разных подчастей, наложенных друг на друга концентрическими слоями. Детектор состоит из внутренней трековой системы, которая окружена сверхпроводящим соле-ноидом, адронного и электромагнитного калориметров, а также мюонного спектрометра.

1.1.1 Система координат детектора ATLAS

В детекторе используется несколько основных систем отчёты, одна из них - прямоугольная система координат. Начало отчёта выбирается в точке взаимодействия, оси расположены так, что ось x направлена к центру БАКа, ось z направлена вдоль движения пучка, а ось y направлена вверх. В цилиндрической системе координат вводятся полярный угол θ - отсчитывается от положительного направления оси z , и азимутальный угол ϕ - определяется в плоскости $0xy$ вокруг оси пучка. Псевдобыстрота задается формулой (1.1).

$$\eta = -\ln \left(\operatorname{tg} \left(\frac{\theta}{2} \right) \right) \quad (1.1)$$

Угловое расстояние между частицами определяется формулой (1.2).

$$\Delta R = \sqrt{(\Delta\eta^2 + \Delta\phi^2)} \quad (1.2)$$

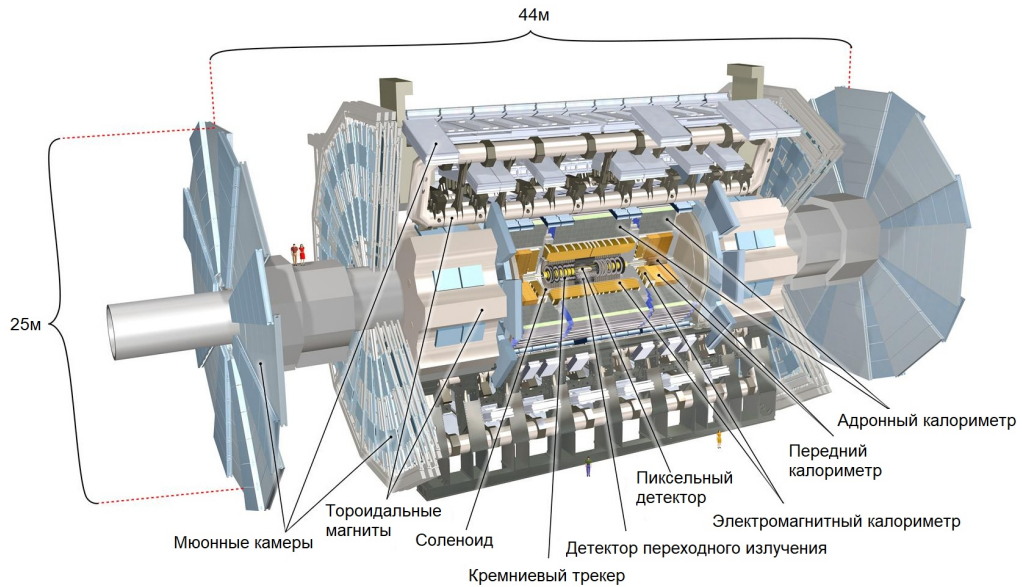


Рисунок 1 — Схема детектора ATLAS и его подсистем

1.1.2 Внутренний детектор

Внутренний детектор - это первая часть детектора ATLAS, регистрирующая продукты распада. Он относительно компактный и очень устойчив к радиационным воздействиям. Детектор состоит из трех частей: пиксельного детектора, детектора переходного излучения и кремниевого трекера, изображённые на рисунке (1), которые охватывают диапазон псевдобыстрот $|\eta| < 2.5$. Внутренний детектор выполняет функции трекинга заряженных частиц. Благодаря магнитному полю от соленоида детектор фиксирует треки частиц, а также восстанавливает их импульсы.

1.1.3 Система калориметров детектора ATLAS

Система калориметров охватывает диапазон псевдобыстрот $|\eta| < 4.9$ и состоит из двух компонентов: электромагнитного и адронного калориметров. Основная задача калориметров - это измерение энергии и положения пришедших в него частиц. Также система калориметров позволяет вычислить потерянную энергию E_T^{miss} . Калориметры сконструированы так, чтобы обеспечивать надежную защиту мюонной системы от проник-

новения электромагнитных и адронных ливней. Электромагнитный калориметр, окружающий внутренний детектор, оптимизирован для измерения энергии фотонов и электронов. Адронный калориметр окружает электромагнитный калориметр и позволяет измерить энергию адронов.

1.1.4 Мюонный спектрометр

Мюоны проходят через описанную в разделе 1.1.3 систему калориметров почти без потерь, поэтому главной задачей мюонного спектрометра является измерение импульсов и идентификация мюонов высоких энергий. Он разработан для обнаружения заряженных частиц и измерения их импульса в пределах значений псевдобыстрот $|\eta| < 2.7$. Мюонный спектрометр состоит из камеры прецессионного слежения и триггерных камер. Камеры слежения измеряют и восстанавливают импульс мюона по виду кривизны трека, изгибаемой магнитным полем тороидных магнитов.

1.1.5 Триггерная система

Триггерная система ATLAS имеет несколько различных подсистем: триггер первого уровня и триггер высокого уровня. Главная задача триггерной системы ATLAS заключается в фильтрации интересных событий от всех остальных фоновых событий. Также триггерная система должна уменьшать частоту событий, примерно 40 кГц, до частоты, которые могут использоваться для обработки данных и для хранения, которая примерно равна 200 Гц.

2 Механизм Хиггса

2.1 Механизмы образования бозона Хиггса

Существует несколько способов образования бозона Хиггса. Поскольку БАК является протон-протонным коллайдером, основной процесс, способствующий возникновению бозона Хиггса, должен содержать глюоны, происходящие от сталкивающихся протонов. Рисунок (4) демонстрирует диаграммы Фейнмана четырех основных процессов[5] образования бозона Хиггса из протон-протонных столкновений:

- Глюон-глюонное слияние (ggF) - наиболее распространенный процесс образования бозона Хиггса, требующий слияния двух глюонов, порождающих бозон Хиггса через петлю тяжелого кварка, обычно топ-кварка из-за его большой массы.
- Слияние векторных бозонов (VBF) - процесс образования бозона Хиггса в результате слияния двух слабых векторных бозонов (W^\pm или Z^0), испускаемых кварками.
- Ассоциированное рождение вместе с W^\pm или Z^0 бозонами (WH или ZH) - процесс, в котором бозон Хиггса образовывается совместно с векторным бозоном W^\pm или Z^0 , образованных из слияния двух кварков.
- Ассоциированное рождение вместе с топ-кварками (ttH) - процесс, в котором бозон Хиггса рождается вместе с парами топ-кварков из расщепления двух глюонов.

2.2 Канал распада $H \rightarrow Z\gamma$

Бозон Хиггса распадается на фермионы или бозоны, и только при измерении продуктов распада возможно реконструировать информацию о наличии бозона Хиггса в детекторе. Вероятность того, какой именно будет распад пропорционален массе продуктов распада. Это означает, что

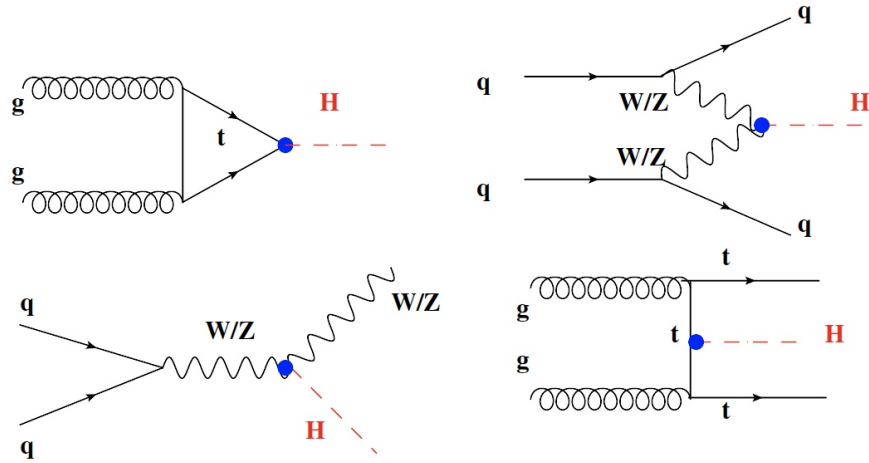


Рисунок 2 — Основные способы образования бозона Хиггса из столкновений адронов. Глюон-глюонное слияние(слева вверху), слияние векторных бозон(справа вверху), ассоциированное рождение вместе с W^\pm или Z^0 (слева внизу), ассоциированное рождение вместе с топ-кварками(справа внизу).

распады с тяжелыми частицами наблюдаются чаще. Из этого следует, что такой распад как распад в Z бозон и фотон ($H \rightarrow Z\gamma$) наблюдается крайне редко, вероятность распада составляет порядка $1.541 \cdot 10^{-3}$ [6]. Распад бозона Хиггса в Z бозон и фотон проходит через бозонные или фермионные петли, обычно W бозоны наиболее вовлечены в петлю, так как их связь с бозоном Хиггса наиболее сильная[7]. На рисунке (3) приведены диаграммы Фейнмана данных процессов.

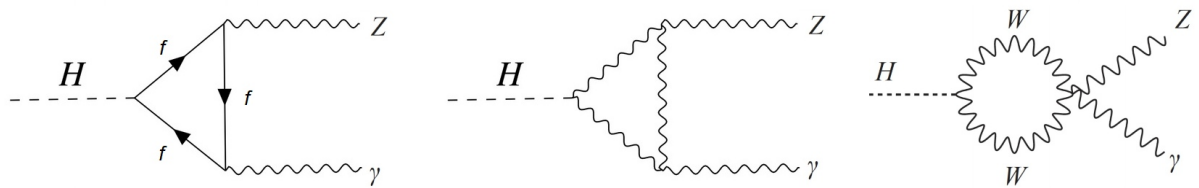


Рисунок 3 — Диаграммы Фейнмана для канала распада $H \rightarrow Z\gamma$

3 Используемые наборы данных

3.1 Используемые данные и Монте-Карло данные

В данной работе использовались наборы Монте-Карло (МК) данных, прошедшие полное моделирование и реконструкцию для геометрии детектора ATLAS. Также в анализе использовались реальные наборы данных с БАКа, полученные в результате протон-протонных столкновений с энергией в системе центра масс 13 TeV и интегральной светимостью 139 fb^{-1} , набранные с 2015 по 2018 гг. К набранным данным была применена фильтрация с условием существования как минимум одного фотона, электрона и мюона в событии. В качестве сигнальных событий использовались сгенерированные с помощью POWHEG+PYTHIA8[8] МК данные для пяти основных каналов образования бозона Хиггса: ggH, VBF, WH, ZH и ttH. Значение массы бозона Хиггса для всех смоделированных образцов было выбрано равным $m_H = 125 \text{ ГэВ}$. Информация о сигнальных МК данных приведена в таблице (1). В качестве фоновых событий использо-

Таблица 1 — Сигнальные МК данные для поиска бозона Хиггса

Канал образования	m_H , ГэВ	Кол-во событий	p-tag	DSID
ggH, Z(ll)gam	125	1.436M	p4062(m16a/d)	345316
VBF, Z(ll)gam	125	1.155M	p4062(m16a/d)	345833
WmH, Z(ll)gam	125	60k	p4062(m16a/d)	345320
WpH, Z(ll)gam	125	60k	p4062(m16a/d)	345321
ttH, Z(ll)gam	125	5.398M	p4062(m16a/d)	346198

вались данные распада $Z + jet$, полученные путем применения обратной фотонной идентификации и изоляции к реальным данным с ускорителя, МК данные распадов $Z + \gamma$ и $Z\gamma + 2jets$, сгенерированные с помощью Shepra[9] и MadGraph[10] генераторов. Информация о МК фоновых событиях приведена в таблицах 2 и 3.

Таблица 2 — Информация об используемых МК данных $Z + \gamma$ в качестве фоновых событий, сгенерируемых с помощью Sherpa

Канал	Количество событий	p-tag	DSID
mc16a $Z + \gamma ee$	998k, 3.996M, 498k, 249k, 249k	p-3705	366140-366144
mc16a $Z + \gamma \mu\mu$	998k, 3.995M, 499k, 250k, 250k	p-3705	366145-366149
mc16d $Z + \gamma ee$	1.247M, 4.956M, 622k, 319k, 318k	p-3705	366140-366144
mc16d $Z + \gamma \mu\mu$	1.246M, 4.985M, 624k, 319k, 320k	p-3705	366145-366149
mc16e $Z + \gamma ee$	1.596M, 6.634M, 835k, 420k, 419k	p-3705	366140-366144
mc16d $Z + \gamma \mu\mu$	1.67M, 6.461M, 834k, 418k, 250k	p-3705	366145-366149

Таблица 3 — Информация об используемых МК данных $Z\gamma + 2jets$ в качестве фоновых событий, сгенерированных с помощью MadGraph

Канал	Количество событий	Идентификатор набора данных	DSID
mc16a ee	0.44M	MGaMcAtNloPy8EG_EWKVgamma	363267
mc16d ee	0.74M	MGaMcAtNloPy8EG_EWKVgamma	363267
mc16e ee	0.77M	MGaMcAtNloPy8EG_EWKVgamma	363267
mc16a $\mu\mu$	0.43M	MGaMcAtNloPy8EG_EWKVgamma	363268
mc16d $\mu\mu$	0.75M	MGaMcAtNloPy8EG_EWKVgamma	363268
mc16e $\mu\mu$	0.77M	MGaMcAtNloPy8EG_EWKVgamma	363268

3.2 Критерии отбора данных

Для анализа использовались МК данные и данные с ускорителя с предварительным отбором лептонов и фотонов. Информация о предварительных отборах приведена в таблице 4. Кандидаты в Z – бозон реконструируются

Таблица 4 — Информация о предварительном отборе лептонов и фотонов, используемых в качестве входных данных

Канал	Электроны	Мюоны	Фотоны
p_T	> 10 ГэВ	> 10 ГэВ	> 10 ГэВ
$ \eta $	$ \eta < 2.47$ кроме $1.37 < \eta < 1.52$	$ \eta < 2.7$ -	$ \eta < 2.37$ кроме $1.37 < \eta < 1.52$
$ d_o /\sigma_{d_o}$	< 5	< 3	-
$z_0 \sin \theta$	< 0.5 mm	< 0.5 mm	-
Идентификация	Мягкая	Средняя	Жёсткая
Изоляция	Мягкий	Мягкая	Мягкая

ируются из пары лептонов противоположных знаков с ограничением по массе $81.2 \text{ GeV} < m_{ll} < 101.2 \text{ GeV}$. Кандидаты в бозон Хиггса восстанавливаются в совокупности с наиболее близкой инвариантной массой Z – бозона к табличному значению и фотона с наибольшей поперечной энергией. Накладываемые ограничения по массе бозона Хиггса $123.59 \text{ GeV} < m_{ll\gamma} < 126.59 \text{ GeV}$.

4 Категоризация

Для повышения чувствительности анализа наборы данных разделяются на категории, поэтому в этой главе будут определяться категории, используемые при поиске бозона Хиггса. Для оптимизации поиска бозона Хиггса чувствительность в различных категориях оценивается по формуле (4.1):

$$\xi = \frac{N_S}{\sqrt{N_S + N_B}} \quad (4.1)$$

где N_S - количество сигнальных событий, N_B - количество фоновых событий $Z + \gamma$, $Z + jet$ и $Z\gamma + 2jet$ процессов, прошедших отбор, описанный в разделе 3.2. Фон нормирован на данные со светимостью $139.0 fb^{-1}$ в диапазоне масс 115 - 170 GeV.

В качестве сигнала использовались смоделированные наборы данных с инвариантной массой 125 GeV, включающие в себя все способы образования бозона Хиггса: ggH , VBF , WH , ZH и ttH . Разделение сигнала и фона в анализе производилось с использованием различных методов машинного обучения из библиотеки TMVA, а именно методами MLP (Multilayer perceptron) и BDTG (Gradient-Boosted Decision Trees).

Ниже приведены несколько самых оптимальных стратегий категоризации, основанных на использовании кинематических переменных и на особенностях различных способов образования бозона Хиггса. Основная стратегия - разделить все события по количеству адронных струй в процессах. Такое разделение оправданно применять, так как процесс VBF характеризуется наличием двух отдельных струй, процесс ggH либо не имеет струй, либо имеет одну дополнительную струю, исходящую от глюонного слияния. Кроме того, события с двумя дополнительными адронными струями также могут происходить из ассоциированного рождения вместе с W или Z бозонами (процессы WH и ZH), где W (jj) или Z (jj) характеризуются двумя дополнительными струями с инвариантной массой равной инвариантной массе W или Z - бозонов. Процессы ZH и WH также можно охарактеризовать отсутствием каких-либо струй при лептонных распадах.

Именно на этих особенностях будет проводиться дальнейшая категоризация.

4.1 Результаты тренировки MLP

При тренировке TMVA методом MLP были использованы следующие настройки нейронной сети: NeuronType = tanh, VarTransform = N, NCycles = 600, HiddenLayers = N+10, N, N, TestRate = 5. Перед обучением и тестированием к сигнальным и фоновым событиям была применена селекция по адронным струям $N_{jet} = 1$. Для этой категории выбран именно метод MLP, так как для определённых переменных он давал наилучшие значения чувствительности. В таблице (5) приведены переменные, используемые при тренировке. Их распределения для сигнала, фона и данных показаны на рисунке 4. Форма сигнальных Монте-Карло данных нормирована на количество событий в данных. Качество отделения сигнала от фона переменной определяется как: $\langle S^2 \rangle = 1/2 * \int \frac{(yS-yB)^2 * dy}{yS+yB}$, где $= 0$ означает, что формы сигнала и фона полностью совпадают. Значимость переменной задается формулой: $I_i = \overline{x_i^2} \sum_{j=1}^{n_h} (\omega_{ij}^{(1)})^2$, $i = 1, \dots, n_{var}$. Переменные выбирались так, чтобы иметь наилучшие характеристики отделения сигнала от фона среди всех кинематически распределений, и также, чтобы можно было описать форму фона аналитическими функциями. Результаты тренировки методом MLP приведены на рисунке 5.

Таблица 5 — Используемые переменные при тренировке методом MLP и их определения

Переменная	Определение	$\langle S^2 \rangle$	I_m	Ранг
$p_{Tt}^{ll\gamma}$	p_T частицы, ортогональный вектору разности между импульсами Z-бозона и фотона	0.084	1370	1
p_T^γ	Поперечный импульс фотона	0.030	1184	2
$\log ME_{kDVal}$	$2 \cdot [\log(ME_{ggH}) - \log(ME_{bkg})]$	0.119	475	3
ME_{Jets}	Потерянный поперечный импульс системы jj	0.017	147	4
$\Delta\eta^{Z,\gamma}$	Разница псевдобыстрот между Z-бозоном и γ	0.033	44	5
$\cos^{H,Z\gamma} \theta$	Косинус угла между $H \rightarrow Z\gamma$ и $Z\gamma$	0.031	42	6
$\Delta\phi^{Z,\gamma}$	Азимутальный угол между Z-бозоном и фотоном	0.011	36	7
$\log ME_{ggH}$	Логарифм матричного элемента ggH	0.118	17	8

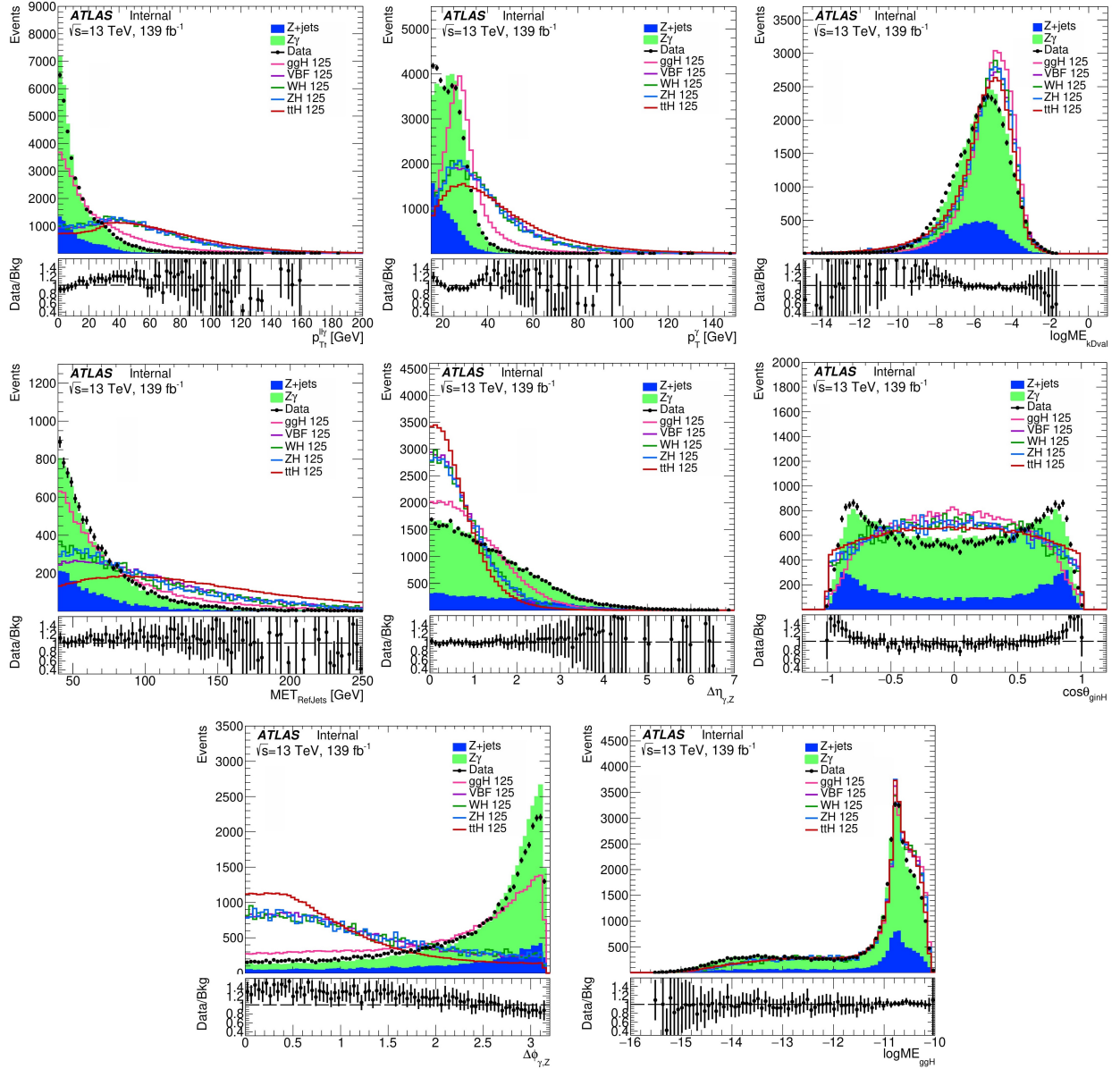


Рисунок 4 — Распределение переменных, используемых при тренировке методом MLP. Монте-Карло данные нормированы на количество событий в данных

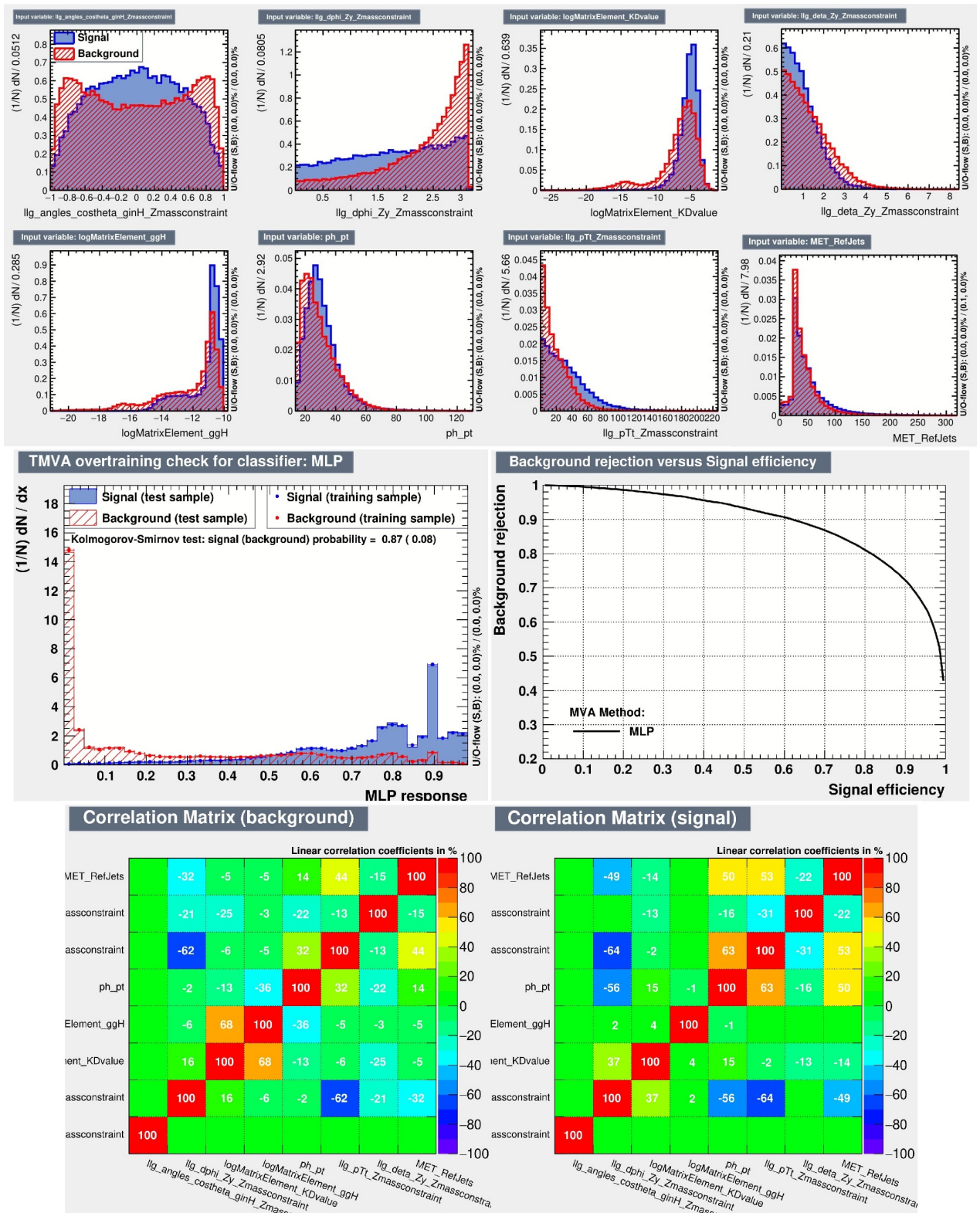


Рисунок 5 — Используемые переменные при тренировке MLP методом, тест на перетренировку модели, ROC - кривая, корреляционные матрицы используемых переменных

После тренировки было построено распределение чувствительности для всех событий от отклика MVA, показанное на рисунке 6 (слева). На основе графика максимальное значение чувствительности находится в точке $MVA > 0.82$ и составляет $\xi_{MLP} = (0.61 \pm 0.01)$. Распределение отклика MLP представлено на рисунке 6 (справа).

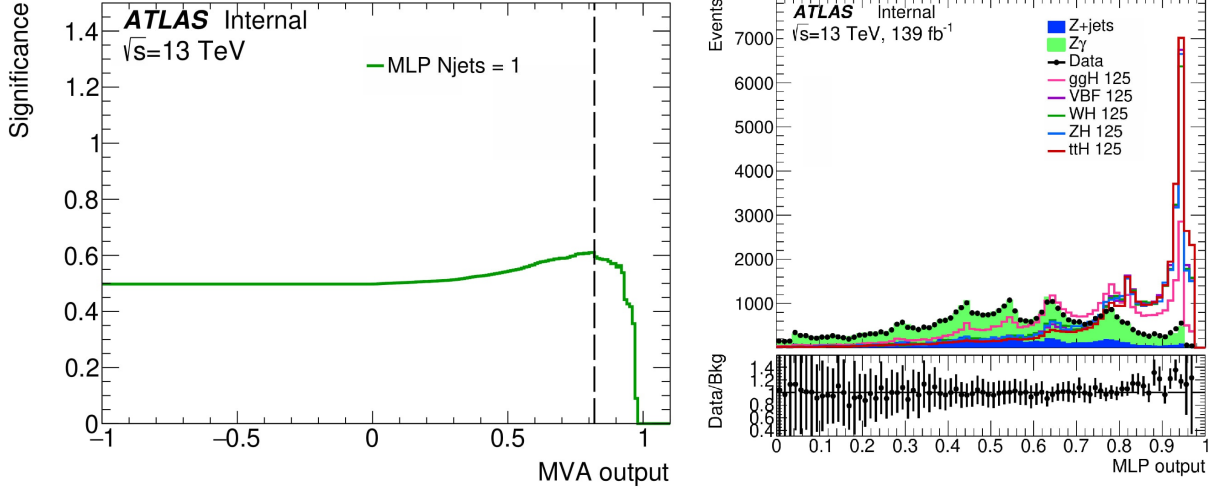


Рисунок 6 — Распределение чувствительности MLP (слева), распределение отклика MLP в результате тренировки MVA (справа)

4.2 Результаты тренировки BDTG

При тренировке TMVA методом BDTG были использованы следующие настройки модели: NTrees = 2000, MinNodeSize = 2.5%, BoostType = Grad, Shrinkage = 0.10, UseBaggedBoost, BaggedSampleFraction = 0.5, nCuts = 20, MaxDepth = 3. Сигнальные и фоновые события прошли селекцию по числу адронных струй $N_{jet} \geq 2$, которая применяется перед обучением и тестированием. В таблице (6) приведены переменные, используемые в тренировке, а их распределения для сигнала, фона и данных показаны на рисунке 7. Форма сигнальных Монте-Карло данных нормирована на количество событий в данных. Результаты тренировки методом BDTG приведены на рисунке 8.

Значимость переменной при тренировке задается формулой: $I_m = |a_m| \sqrt{s_m(1.0 - s_m)}$. После тренировки построено распределение чувствительности для всех событий от отклика MVA, показанное на рисунке 9 (слева). Максимальное значение чувствительности составляет $\xi_{BDTG} = (0.87 \pm 0.02)$

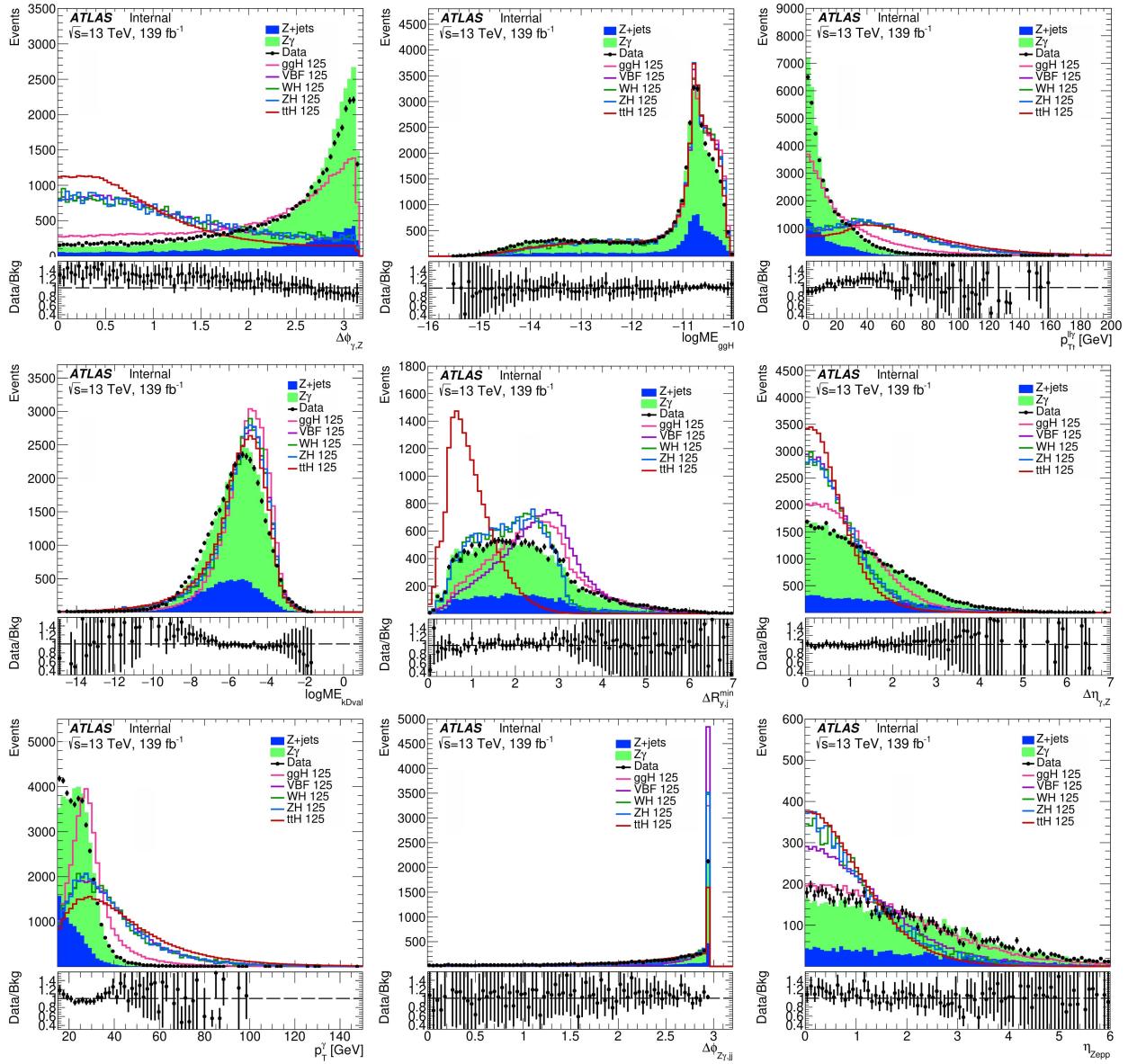


Рисунок 7 — Распределение переменных, используемых при тренировке методом BDTG. Монте-Карло данные нормированы на количество событий в данных

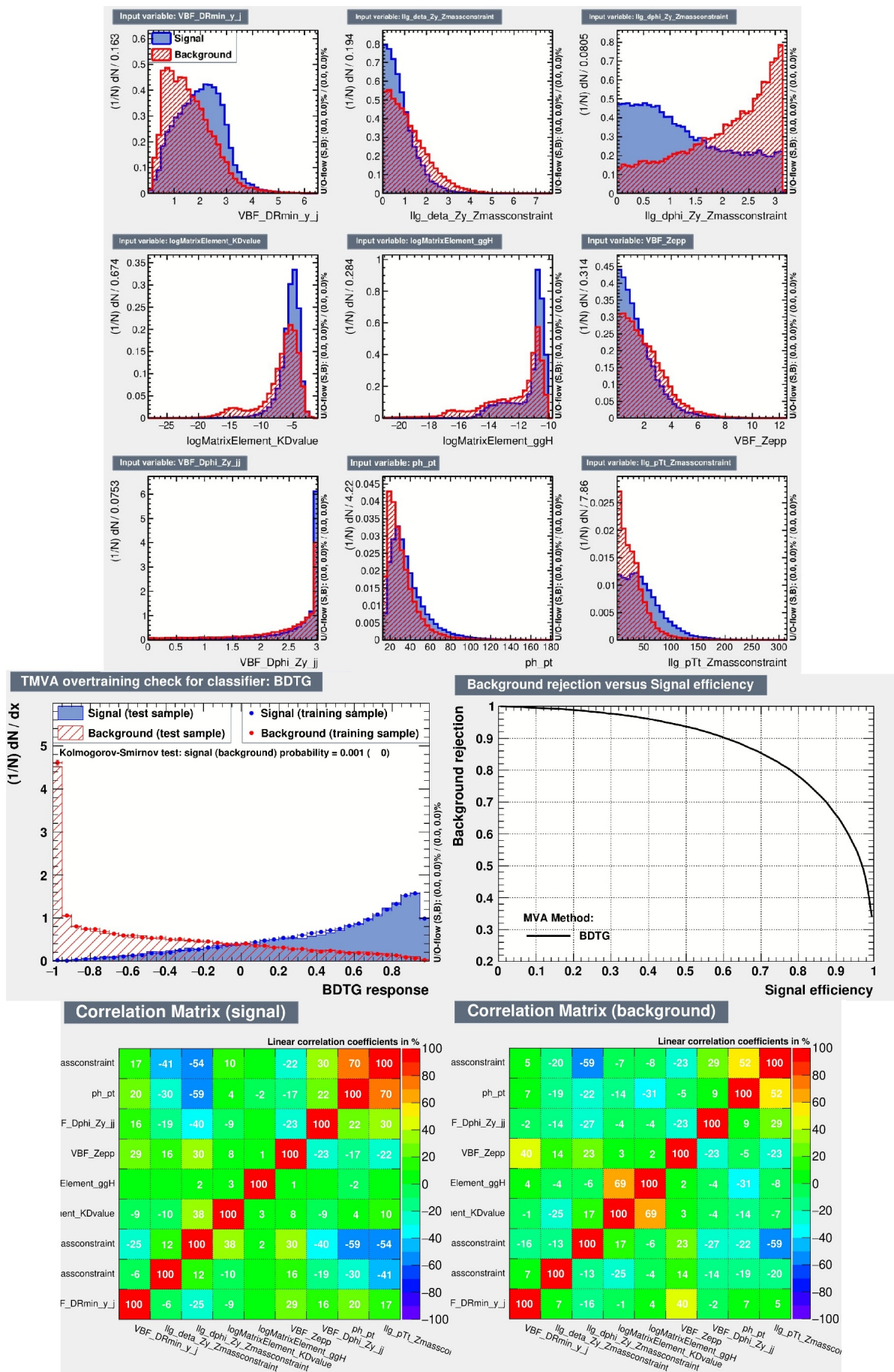


Рисунок 8 — Используемые переменные при тренировке BDTG методом, тест на перетренировку модели, ROC - кривая, корреляционные матрицы используемых переменных

Таблица 6 — Используемые переменные при тренировке методом BDTG и их определения

Переменная	Определение	$\langle S^2 \rangle$	I_m	Ранг
$\Delta\phi^{Z,\gamma}$	Азимутальный угол между Z-бозоном и фотоном	0.157	0.140	1
$\Delta\phi_{Z\gamma,jj}$	Азимутальный угол между $Z\gamma$ и системой jj	0.030	0.036	2
$\Delta\eta^{Z,\gamma}$	Разница псевдобыстрот между Z-бозоном и γ	0.052	0.111	3
$\log ME_{kDV al}$	$2 \cdot [\log(ME_{ggH}) - \log(ME_{bkg})]$	0.092	0.110	4
$\Delta R_{\gamma,j}^{min}$	Минимальное расстояние между $Z\gamma$ и системой jj	0.072	0.110	5
$\log ME_{ggH}$	Логарифм матричного элемента ggH	0.124	0.110	6
η^{Zep}	$ \eta_{Z\gamma} - 0.5 \cdot (\eta_{j_1} + \eta_{j_2}) $	0.026	0.100	7
$p_{Tt}^{ll\gamma}$	p_T частицы, ортогональный вектору разности между импульсами Z-бозона и фотона	0.110	0.096	8
p_T^γ	Поперечный импульс фотона	0.048	0.095	9

в точке BDTG > 0.76 . Распределение отклика BDTG представлено на рисунке 9 (справа).

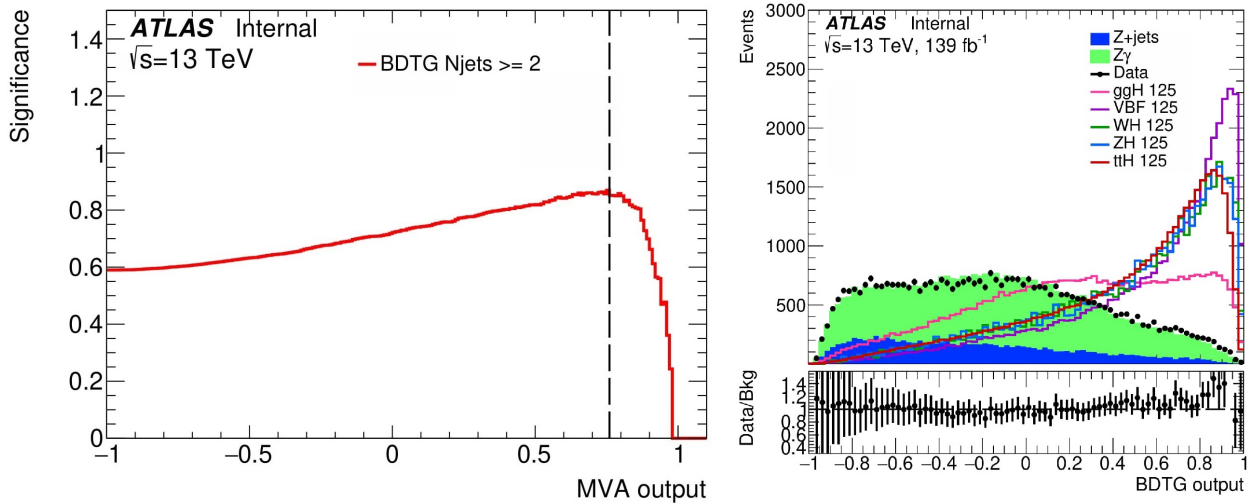


Рисунок 9 — Распределение чувствительности BDTG (слева), распределение отклика BDTG в результате тренировки MVA (справа)

4.3 Основная категоризация

При поиске распада $Z\gamma$ бозона Хиггса все события могут быть разделены на категории, определяющиеся на основе различных кинематических свойств распада бозона Хиггса в Z – бозон и фотон. Одно из таких свойств – это рождение H с 0, 1, 2 и больше адронными струями. Для событий с $N_{jet} = 1$ и $N_{jet} \geq 2$ было применено машинное обучение, а для событий

$N_{jet} = 0$ применение каких либо методов MVA оказалось неэффективным, так как в результате распределение фона не описывалось аналитическими функциями в регионе инвариантной массы бозона Хиггса. Таким образом можно определить первые три категории событий в зависимости от количества струй:

- Категория, где струи отсутствуют: $N_{jet} = 0$
- Категория с одной струей: $N_{jet} = 1$
- Категория с двумя и больше струями: $N_{jet} \geq 2$

Для разделения событий без адронных струй была найдена наиболее чувствительная к разделению сигнала и фона переменная p_{Tt} , которая определяется по формуле (4.2). Так как в анализе исследуется канал распада $H \rightarrow Z\gamma$, в котором Z – бозон распадается только в два лептона (в пару электронов либо мюонов), все события можно разделить на электронный и мюонный каналы. То есть найдены две переменные, по которым можно производить дальнейшую категоризацию:

- компонента поперечного импульса p_T частицы-кандидата на бозон Хиггса, ортогональная вектору разности между импульсами Z -бозона и фотона
- переменная, определяющая лептонный канал: ee канал или $\mu\mu$ канал

$$p_{Tt} = 2 \cdot \frac{|p_x^Z p_y^\gamma - p_x^\gamma p_y^Z|}{p_T^{Z-\gamma}} \quad (4.2)$$

Максимум чувствительности по переменной p_{Tt} определяется по формуле (4.1) и достигается при $p_{Tt} > 40$ GeV для категории $N_{jet} = 0$. На основе разделов 4.1 и 4.2 для категорий $N_{jet} = 1$ и $N_{jet} \geq 2$ события можно разделить по откликам для каждой из категорий, а также по лептонному каналу в регионах меньше максимумов чувствительности. Тогда исходный набор данных можно разделить на 9 категорий, суммарная чувствительность которых составляет 1.27 :

- «Категория 1», для $p_{Tt} > 40$ GeV
- «Категория 2», для ee и $p_{Tt} < 40$ GeV
- «Категория 3», для $\mu\mu$ и $p_{Tt} < 40$ GeV
- «Категория 4», для MLP отклика > 0.82
- «Категория 5», для ee канала и MLP отклика < 0.82
- «Категория 6», для $\mu\mu$ канала и MLP отклика < 0.82
- «Категория 7», для BDTG отклика > 0.76
- «Категория 8», для ee канала и BDTG отклика < 0.76
- «Категория 9», для $\mu\mu$ и BDTG отклика < 0.76

В таблице 7 приведены количество сигнальных и фоновых событий для каждой категории и значения чувствительностей, определяемых в диапазоне масс 123.59 - 126.59 GeV. На рисунке 11 приведена схема полученных категорий. Если произвести более точное вычисление по формуле Пуассона (4.3), то суммарная чувствительность составит 1.28.

Таблица 7 — Количество сигнальных и фоновых событий, значения чувствительности для каждой категории в диапазоне масс 123.59-126.59 GeV

№	N_{jet}	Category	N_S	N_B	$S/\sqrt{(S+B)}$	$\xi_{Poisson}$
1	0	$p_{Tt}^{l\gamma} > 40$	1.54	49.55	0.22	0.22
2	0	ee $p_{Tt}^{l\gamma} < 40$	15.31	2782.33	0.29	0.29
3	0	$\mu\mu$ $p_{Tt}^{l\gamma} < 40$	18.94	3734.02	0.31	0.31
total			35.79	6565.90	0.48	0.48
4	1	MLP > 0.82	14.01	504.47	0.62	0.62
5	1	ee MLP < 0.82	6.23	1126.03	0.19	0.19
6	1	$\mu\mu$ MLP < 0.82	8.21	1584.65	0.21	0.21
total			28.45	3215.15	0.67	0.68
7	≥ 2	BDTG > 0.76	9.30	105.22	0.87	0.89
8	≥ 2	ee BDTG < 0.76	7.42	821.55	0.26	0.26
9	≥ 2	$\mu\mu$ BDTG < 0.76	10.10	1111.13	0.30	0.30
total			26.81	2037.92	0.96	0.97
combined			91.05	11818.97	1.27	1.28

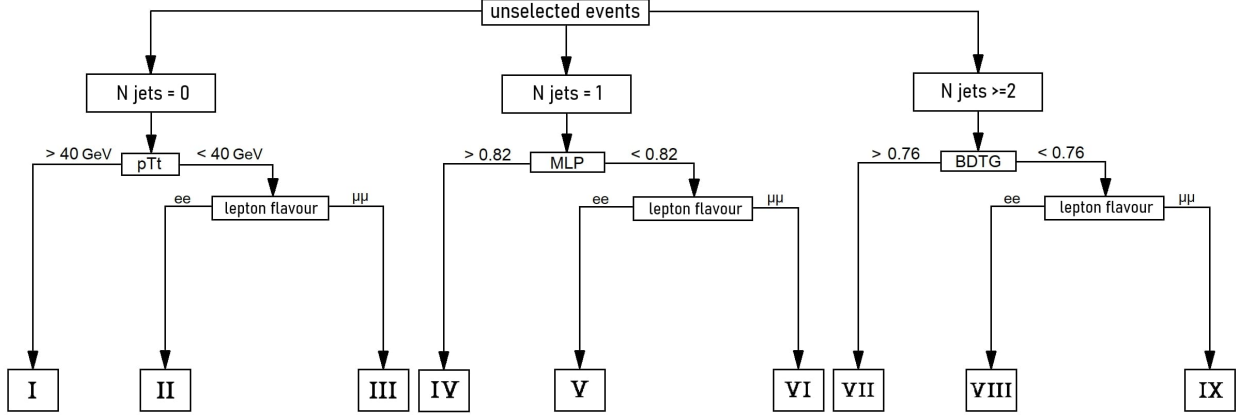


Рисунок 10 — Схема категоризации

$$\xi_{Possion} = \sqrt{2 \cdot \left[(S + B) \times \ln \left(1 + \frac{S}{B} \right) - S \right]} \quad (4.3)$$

4.4 Сравнение данных и фонового распределения

Доминирующий фон в рассматриваемом процессе распада бозона Хиггса $H \rightarrow Z\gamma$ является процесс $Z\gamma$, в котором Z бозон образован посредством КХД рождения, также вклад в фоновые процессы вносит $Z + jets$, где адронная струя неверно идентифицируется как фотон. В анализе рассматривается еще один фоновый процесс $Z\gamma jj$, являющийся результатом электрослабого образования, преимущественно вносящий вклад в VBF процесс, и составляет порядка 2.4% от всего смоделированного Монте-Карло фона.

Для каждой категории необходимо оценить форму смоделированного суммарного фона на возможность описать распределение аналитическими функциями, а также сопоставить реальные данные с коллайдера фоновым процессам. На рисунках 11 и 12 показаны сравнения формы распределения инвариантной массы бозона Хиггса между данными и суммой фоновых процессов ($Z\gamma$, $Z + jets$, $Z\gamma jj$).

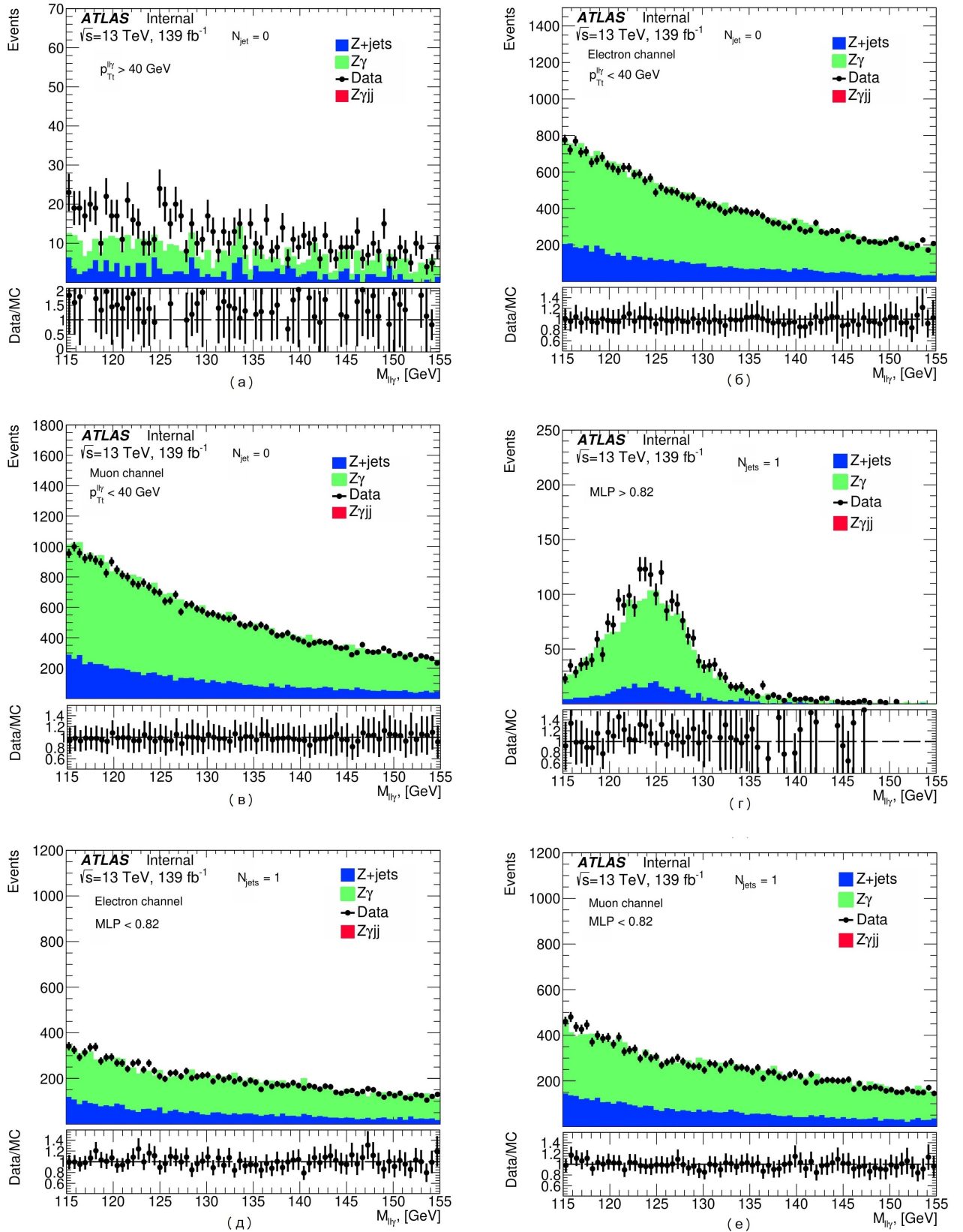


Рисунок 11 — Сравнение формы фона для распределения массы с данными: (а) «категория 1», (б) «категория 2», (в) «категория 3», (г) «категория 4», (д) «категория 5», (е) «категория 6», распределение инвариантной массы для данных и фона ($Z\gamma$, $Z + jets$, $Z\gamma\text{jj}$)

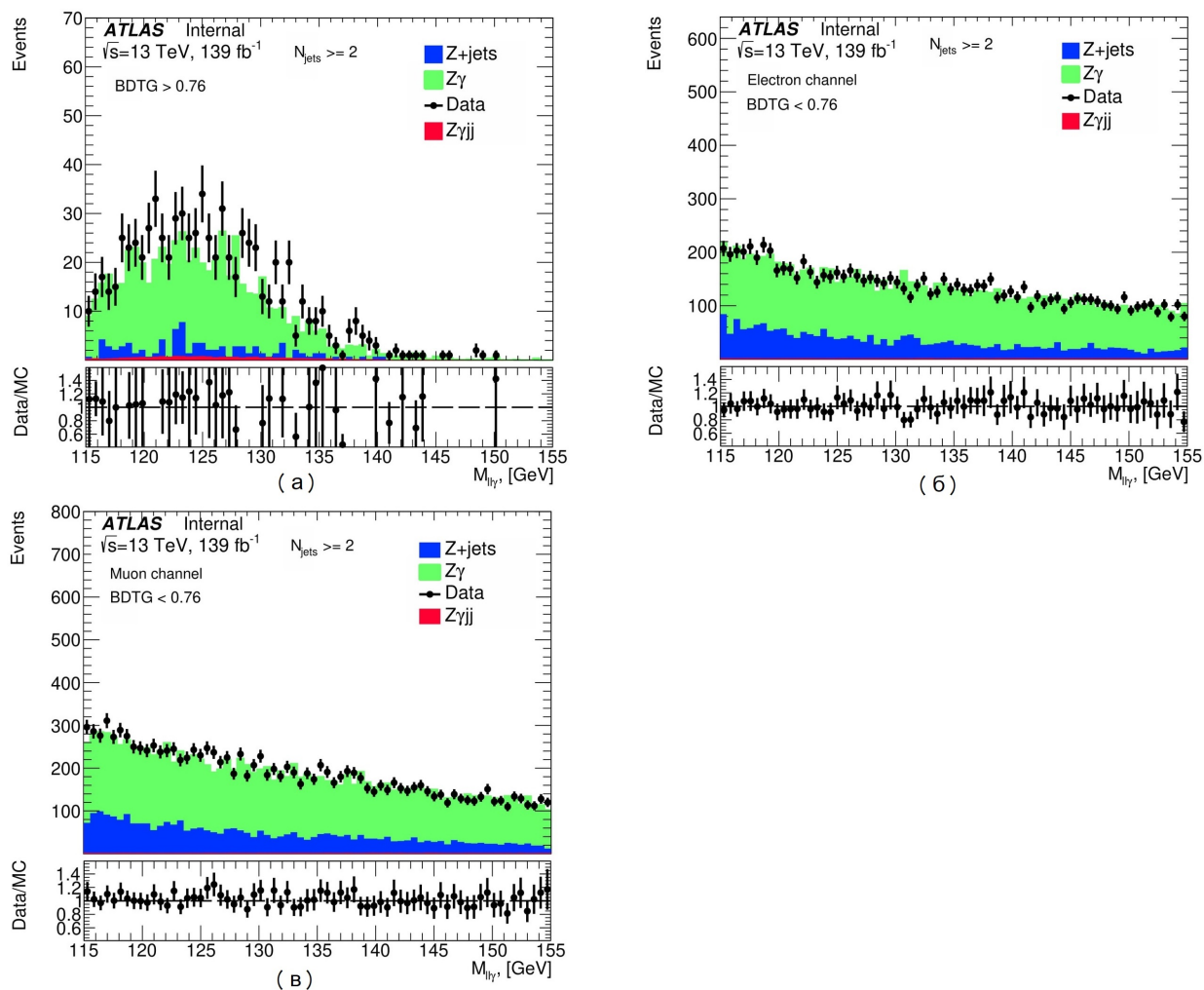


Рисунок 12 — Сравнение формы фона для распределения массы с данными: (а) «категория 7», (б) «категория 8», (в) «категория 9», распределение инвариантной массы для данных и фона ($Z\gamma$, $Z + jets$, $Z\gamma jj$)

4.5 Другие способы категоризации

В данном разделе показаны другие способы категоризации, которые имеют высокую чувствительность анализа, но полученную форму фона будет сложно описать аналитическими функциями. Это приведет к значительному увеличению систематической погрешности на данное описание, из-за чего такой выбор категорий не рассматривается в окончательном анализе реальных данных. Рассматривались следующие категории:

- $N_{jet} = 0$, $N_{jet} = 1$ и $N_{jet} \geq 2$ с другими переменными при тренировке и настройками методов MLP и BDTG (суммарная $\xi = 1.30$)
- $N_{jet} \geq 2$ и $N_{jet} < 2$ (суммарная $\xi = 1.23$)

На рисунках 13 и 14 приведено сравнение формы фона для распределения массы с данными. Из картинок видно, что в регионе $m_{ll\gamma} = 125$ GeV фон имеет резкие скачки.

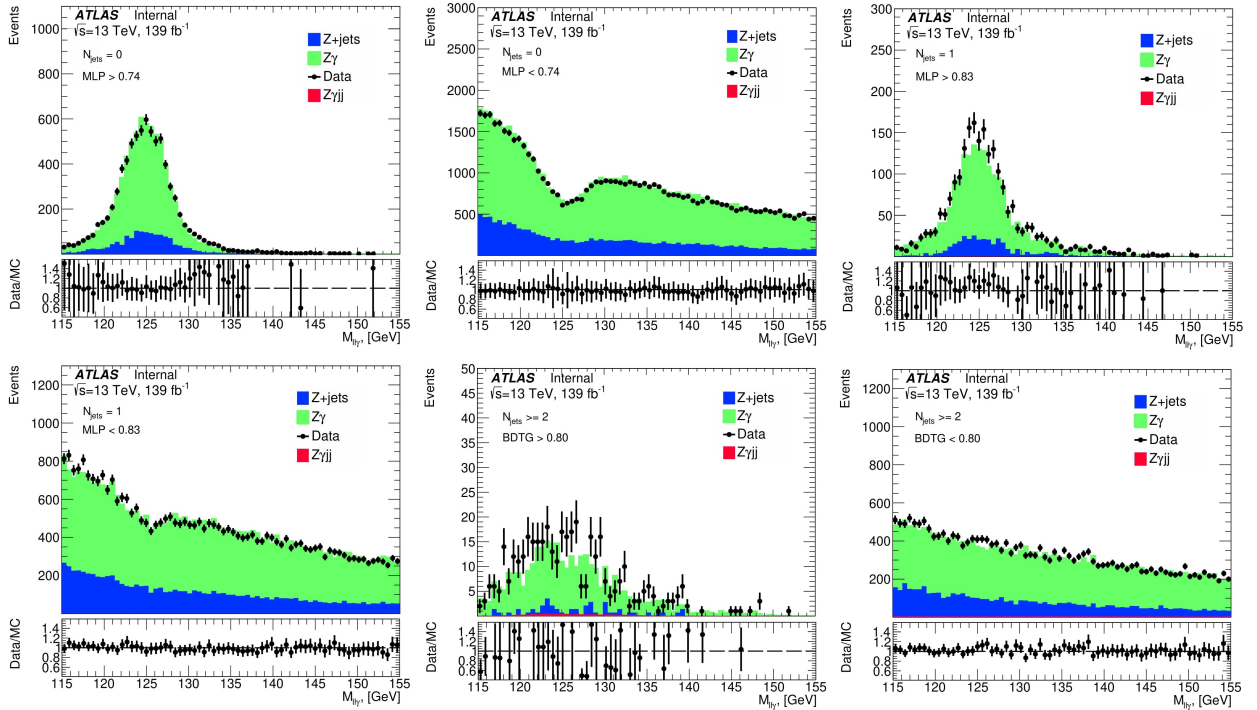


Рисунок 13 — Непременимые категории, разделенные по N_{jet} 0, 1, ≥ 2 с другими переменными и настройками методов машинного обучения

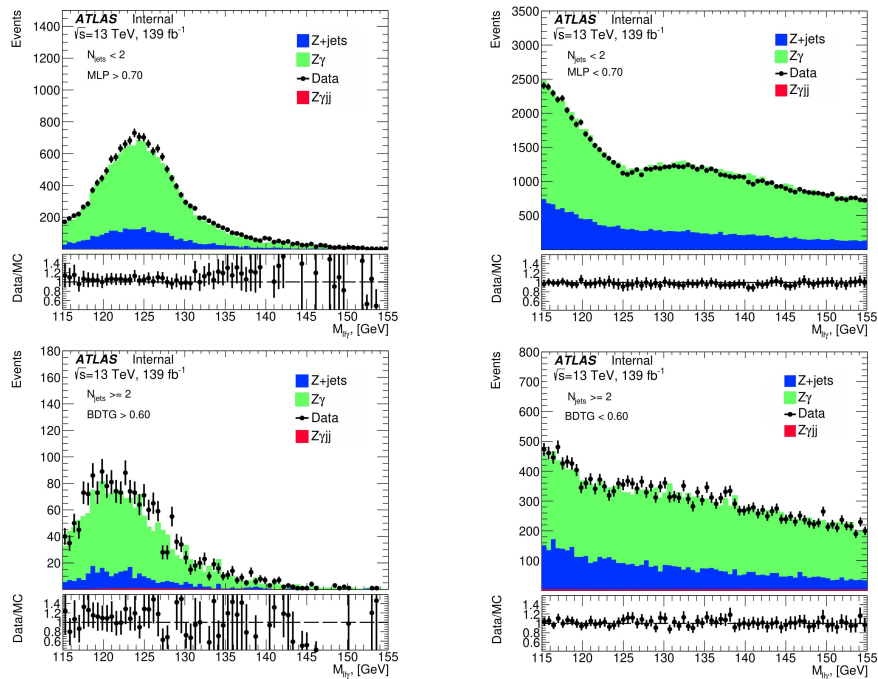


Рисунок 14 — Непременимые категории, разделенные по $N_{jet} \leq 2$ и ≥ 2 с другими переменными и настройками методов машинного обучения

5 Моделирование сигнала

Выбор инвариантной массы $m_{ll\gamma}$ позволяет эффективно отделить форму сигнального распределения от фонового. Распределения сигнала и фона параметризуются аналитическими функциями. В данной главе была проведена параметризация только сигнального распределения. Основные параметры, которые входят в финальное фитирование процесса $H \rightarrow Z\gamma$:

- Пик распределения инвариантной массы бозона Хиггса
- Значение σ_{CB} для оценки справедливости полученных результатов
- Отношение фитирования к моделированию сигнала

Среди всех смоделированных данных для двух способов образования бозона Хиггса с учетом нормировки на их поперечное сечение (ggH и VBF, так как они вносят наибольший вклад в сигнальные события) распределение по инвариантной массе трёх частиц для резонансной массы m_H , как было получено в предыдущем семестре, хорошо описывается функцией Double-Sided Crystal Ball (DSCB).

Центральная часть функции DSCB описывается функцией Гаусса, а оба хвоста задаются степенным законом. Функция задается как (5.1):

$$N \cdot \begin{cases} e^{t^2/2}, & -\alpha_{Lo} \leq t \leq \alpha_{Lo} \\ \frac{e^{-0.5\alpha_{Lo}^2}}{\left[\frac{\alpha_{Lo}}{n_{Lo}} \left(\frac{n_{Lo}}{\alpha_{Lo}} - \alpha_{Lo} - t\right)\right]^{n_{Lo}}}, & t < -\alpha_{Lo} \\ \frac{e^{-0.5\alpha_{Hi}^2}}{\left[\frac{\alpha_{Hi}}{n_{Hi}} \left(\frac{n_{Hi}}{\alpha_{Hi}} - \alpha_{Hi} + t\right)\right]^{n_{Hi}}}, & t > \alpha_{Hi} \end{cases} \quad (5.1)$$

где $t = \Delta m_H / \sigma_{CB}$, $\Delta m_H = M_{Z\gamma} - \mu_{CB}$, N - нормировочный параметр, μ_{CB} - пик распределения Гаусса, σ_{CB} - ширина распределения Гаусса, α_{Lo} (α_{Hi}) - точки, где функция Гаусса переходит в степенную функцию на левом (правом) участке, n_{Lo} (n_{Hi}) - показатель степенной функции.

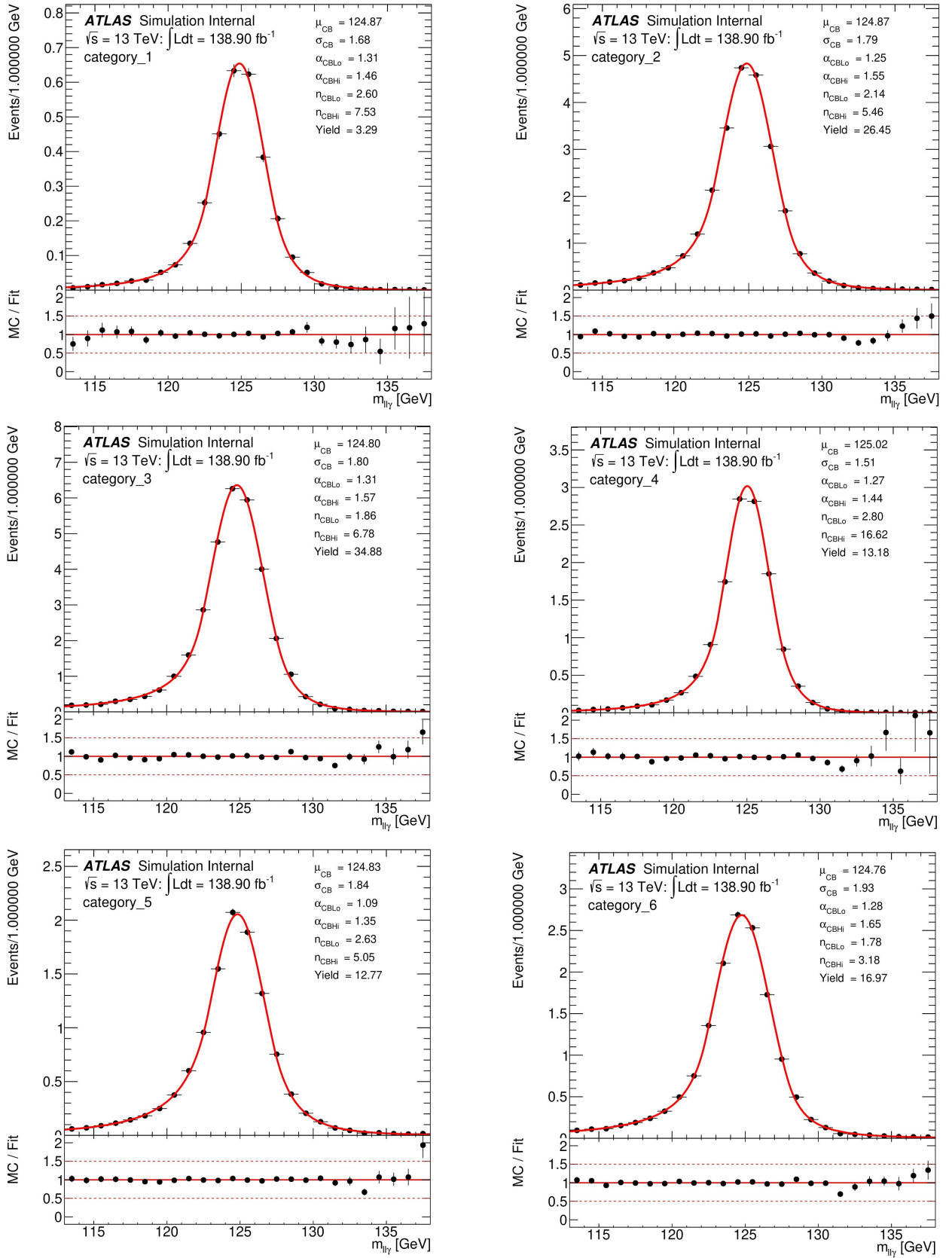


Рисунок 15 — Фитирование смоделированного сигнала для каждой категории

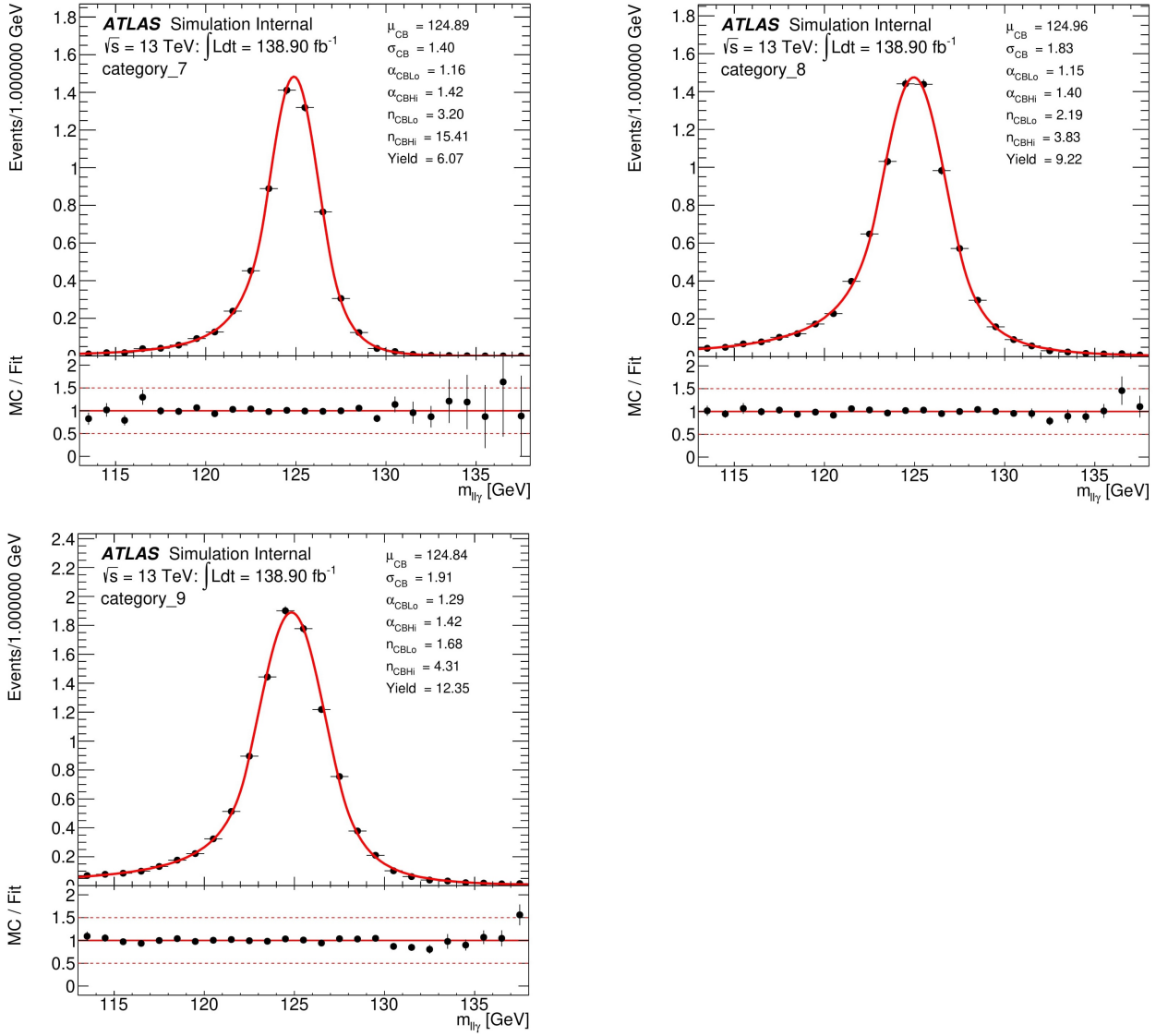


Рисунок 16 — Фитирование смоделированного сигнала для каждой категории

Для каждой категории подобраны наиболее подходящие параметры фитирования. На рисунках 15 и 16 приведены результаты фитирования смоделированного сигнала для каждой категории. Основные параметры функций фитирования приведены в таблице 8.

На основе приведенных фитированных распределений и на данных в таблице можно сказать, что сигнал в каждой из полученной категории хорошо описывается аналитическими функциями с достаточно небольшой погрешностью. Большая погрешность для значений σ_{CB} обусловлена малым количеством статистики в каждой из категорий.

Таблица 8 — Значения инвариантной массы, ширины распределения и коэффициента согласия Пирсона для каждой категории

Категория	Значение массы, ГэВ	σ_{CB} , ГэВ	χ^2/Ndf
$N_{jet} = 0, p_{Tt}^{ll\gamma} > 40$	124.87 ± 1.49	1.68 ± 0.96	0.86
$N_{jet} = 0, ee, p_{Tt}^{ll\gamma} < 40$	124.87 ± 0.59	1.79 ± 0.78	2.12
$N_{jet} = 0, \mu\mu, p_{Tt}^{ll\gamma} < 40$	124.80 ± 0.48	1.80 ± 0.64	3.41
$N_{jet} = 1, MLP > 0.82$	125.02 ± 0.65	1.51 ± 0.90	1.84
$N_{jet} = 1, ee, MLP < 0.82$	124.83 ± 0.94	1.84 ± 1.21	1.38
$N_{jet} = 1, \mu\mu, MLP < 0.82$	124.76 ± 0.86	1.93 ± 1.02	2.10
$N_{jet} \geq 2, BDTG > 0.76$	124.89 ± 0.92	1.40 ± 0.95	0.89
$N_{jet} \geq 2, ee, BDTG < 0.76$	124.96 ± 1.11	1.83 ± 1.13	1.28
$N_{jet} \geq 2, \mu\mu, BDTG < 0.76$	124.84 ± 0.92	1.91 ± 0.87	1.67

Заключение

Главная цель проведенной работы заключалась в увеличении чувствительности анализа канала распада бозона Хиггса в Z бозон и фотон путём создания категоризации, а также в проверке различных методов оптимизации анализа с помощью машинного обучения.

Для достижения поставленных целей в результате проведенной работы были получены следующие результаты:

- на основе кинематических переменных и особенностей различных способов образования бозона Хиггса создана категоризация с использованием оптимальных настроек различных методов машинного обучения из библиотеки TMVA. Суммарная чувствительность всех категорий составляет 1.28, что является высоким показателем и повышает общую чувствительность анализа. Для каждой из полученных категорий форма фонового распределения хорошо описывается аналитическими функциями
- проведено фитирование смоделированного сигнального распределения инвариантной массы для каждой категории функцией DSCB с наиболее оптимальными параметрами, и получено, что форма сигнала описывается математическими функциями с достаточно небольшой погрешностью

В будущем планируется провести моделирование фонового распределения инвариантной массы для подтверждения справедливости категоризации, а так же проверить другие способы оптимизации и методы машинного обучения из других библиотек, таких как TensorFlow 2.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. *Smestad L.* Preparing for an Unbiased Study of the $H \rightarrow \gamma\gamma$ Background with the ATLAS Experiment at LHC. — 2008. — URL: <https://cds.cern.ch/record/1402030> ; Presented 2008.
2. *Speckmayer P.* [et al.]. The toolkit for multivariate data analysis, TMVA 4 // Journal of Physics: Conference Series. — 2010. — Apr. — Vol. 219, no. 3. — P. 032057. — URL: <https://doi.org/10.1088/1742-6596/219/3/032057>.
3. *Collaboration A.* The performance of ATLAS detector. — Heidelberg : Springer, 2011. — URL: <https://cds.cern.ch/record/1485681> ; Reprinted from The European Physical Journal C (articles published between summer 2010 and spring 2011).
4. *ATLAS Collaboration.* The ATLAS Experiment at the CERN Large Hadron Collider // JINST. — 2008. — Vol. 3. — S08003.
5. *Lamberti M.* Measurement of differential cross-sections for Higgs boson production in the $\gamma\gamma$ decay channel at $\sqrt{s} = 13\text{TeV}$ with the ATLAS experiment / Lamberti Mario. — 03/2020. — URL: <https://cds.cern.ch/record/2746556> ; Presented 02 Apr 2020.
6. *Basalaev A.* Search for the Higgs boson in the final state with two leptons and a photon produced in pp collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector. — 2020. — URL: <https://cds.cern.ch/record/2744275>.
7. *ATLAS Collaboration.* Search for the decay of a Higgs boson in the $\ell\ell\gamma$ channel in proton-proton collisions at $\sqrt{s} = 13\text{ TeV}$ // JHEP. — 2018. — June. — Vol. 11, arXiv:1806.05996. CMS-HIG-17-007-003. — 152. 40 p. — URL: <https://cds.cern.ch/record/2624385> ; Submitted to JHEP.

8. *Buckley A., Gupta D. B.* Powheg-Pythia matching scheme effects in NLO simulation of dijet events. — 2017. — arXiv: [1608.03577](https://arxiv.org/abs/1608.03577) [[hep-ph](#)].
9. *Gleisberg T.* [et al.]. Event generation with SHERPA 1.1 // Journal of High Energy Physics. — 2009. — Feb. — Vol. 2009, no. 02. — P. 007–007. — ISSN 1029-8479. — URL: <http://dx.doi.org/10.1088/1126-6708/2009/02/007>.
10. *Alwall J.* [et al.]. MadGraph 5: going beyond // Journal of High Energy Physics. — 2011. — June. — Vol. 2011, no. 6. — ISSN 1029-8479. — URL: [http://dx.doi.org/10.1007/JHEP06\(2011\)128](http://dx.doi.org/10.1007/JHEP06(2011)128).