

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

Научные руководители: Проклова Н. В.
Солдатов Е. Ю.

Студент: Казакова К. К.



Кафедра физики
элементарных
частиц

Москва 2021

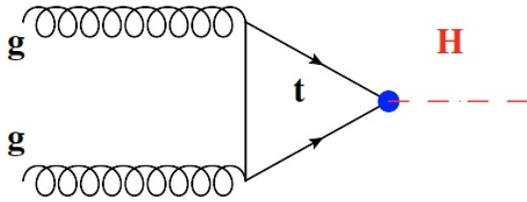


Мотивация

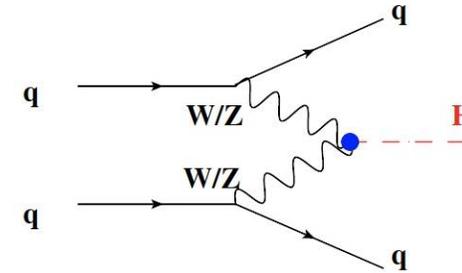
- **увеличить чувствительность анализа путем создания категоризации**
- **проверить возможные способы оптимизации анализа с использованием различных методов машинного обучения из библиотеки TMVA**
- **на основе категорий произвести фитирование смоделированного сигнала**

Механизм Хиггса

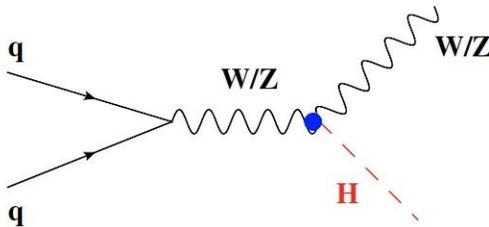
Четыре основных процесса образования бозона Хиггса из протон-протонных столкновений:



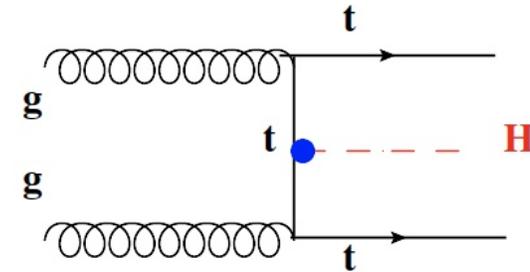
Глюон-глюонное слияние (ggF)



Слияние векторных бозонов (VBF)



Ассоциированное рождение вместе с W или Z бозонами (WH или ZH)



Ассоциированное рождение вместе с топ-кварками (ttH)

Категоризация



Чувствительность $\xi = \frac{N_S}{\sqrt{N_S + N_B}}$

N_S – число сигнальных событий (VBF, ggH, VH, ttH)
 N_B – число фоновых событий (Z + jets, Z γ , Z γ jj)

Из-за особенностей распада бозона Хиггса все события удобно **разделять по количеству адронных струй**

- Для категории N jet = 0 машинное обучение не применялось из-за низкой чувствительности и сложной формы фона
- Разделение для N jet = 0 производилось по переменной pTt. Максимум чувствительности достигается при 40 GeV

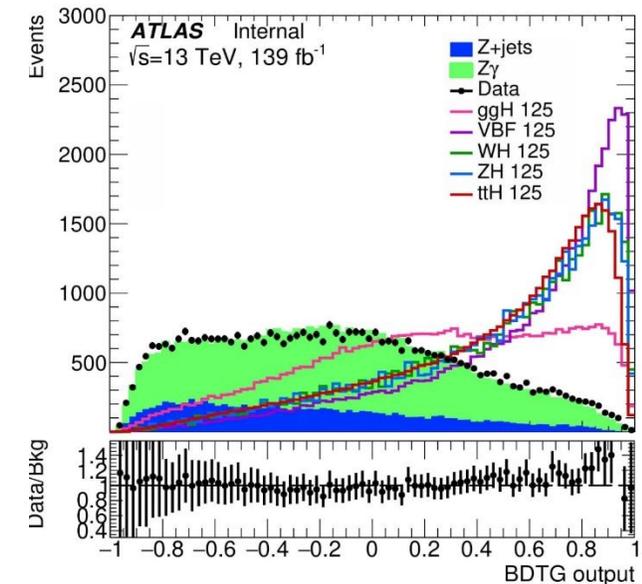
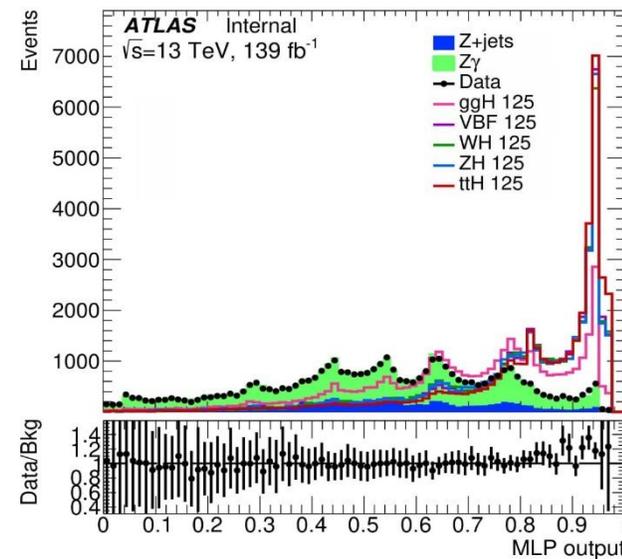
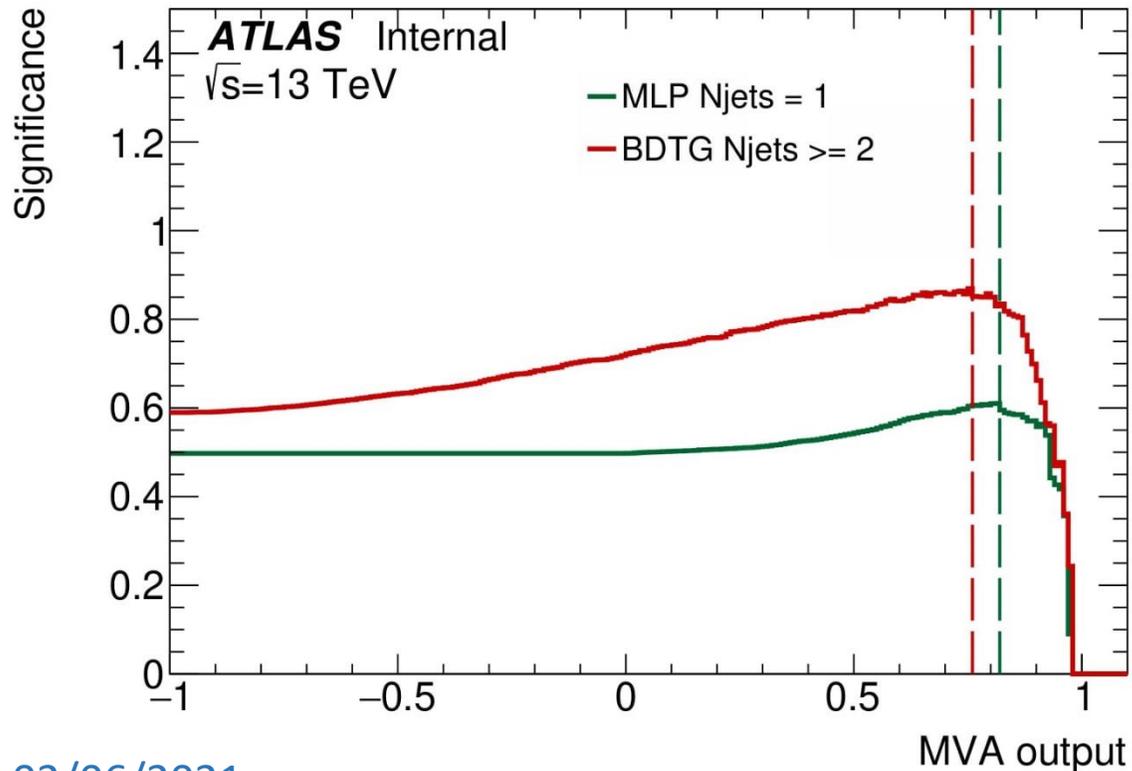
$$p_{Tt} = 2 \cdot \frac{|p_x^Z p_y^\gamma - p_x^\gamma p_y^Z|}{p_T^{Z-\gamma}}$$

Категоризация

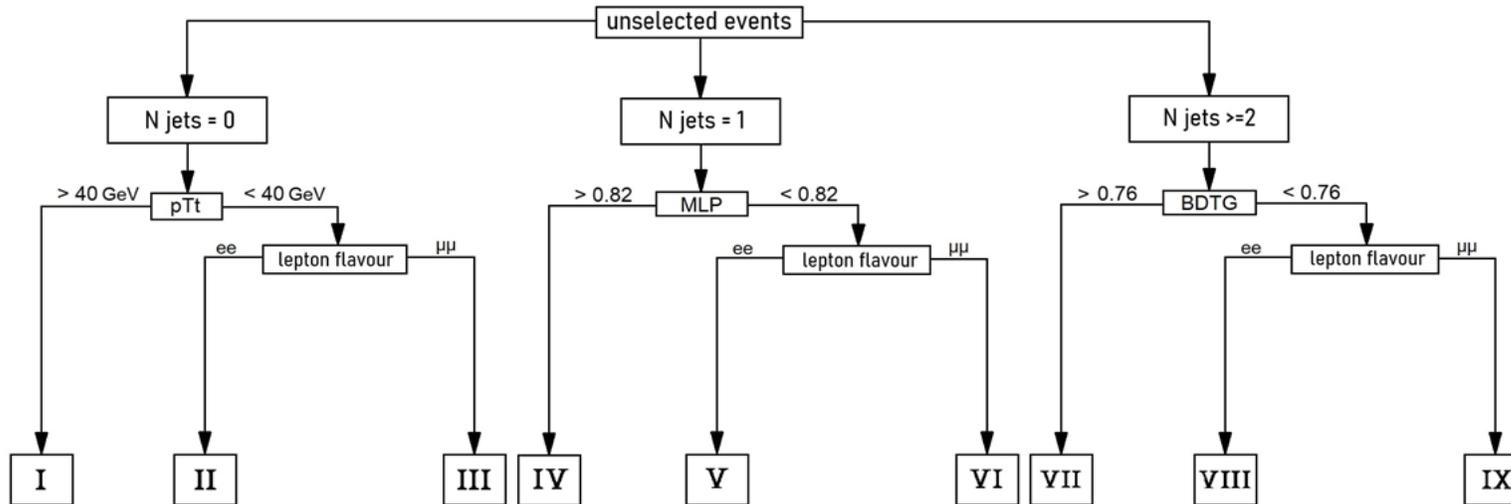
- Для категории N jet = 1 применялся метод MLP, так как он давал наибольшее значение чувствительности. Переменные: $p_{Tt}^{ll\gamma}$, p_T , $\log ME_{kDV_{al}}$, MET_{Jets} , $\Delta\eta^{Z,\gamma}$, $\cos^{H,Z\gamma}\theta$, $\Delta\phi^{Z,\gamma}$, $\log ME_{ggH}$
- Для категории N jet ≥ 2 применялся метод BDTG. При тренировке использовались переменные: $\Delta\phi^{Z,\gamma}$, $\Delta\phi_{Z\gamma,jj}$, $\Delta\eta^{Z,\gamma}$, $\log ME_{kDV_{al}}$, $\log ME_{ggH}$, $\eta^{Z_{epp}}$, $p_{Tt}^{ll\gamma}$, p_T^γ

$$\xi_{MLP}(0.82) = 0.61 \pm 0.01$$

$$\xi_{BDTG}(0.76) = 0.87 \pm 0.02$$



Категоризация



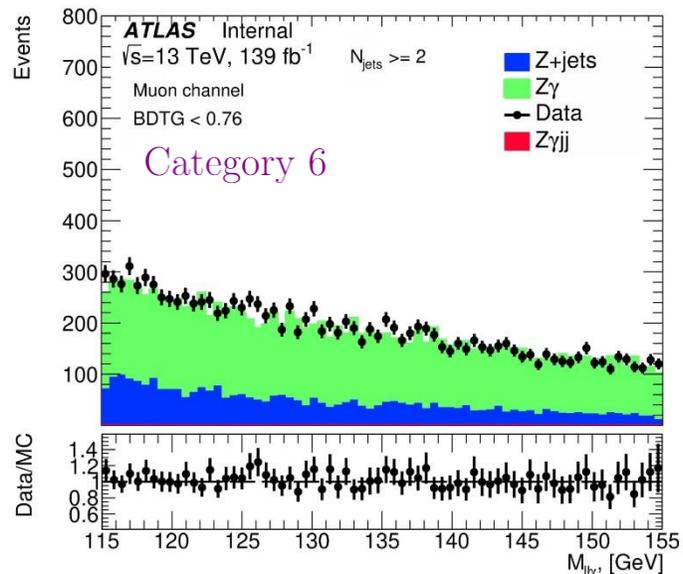
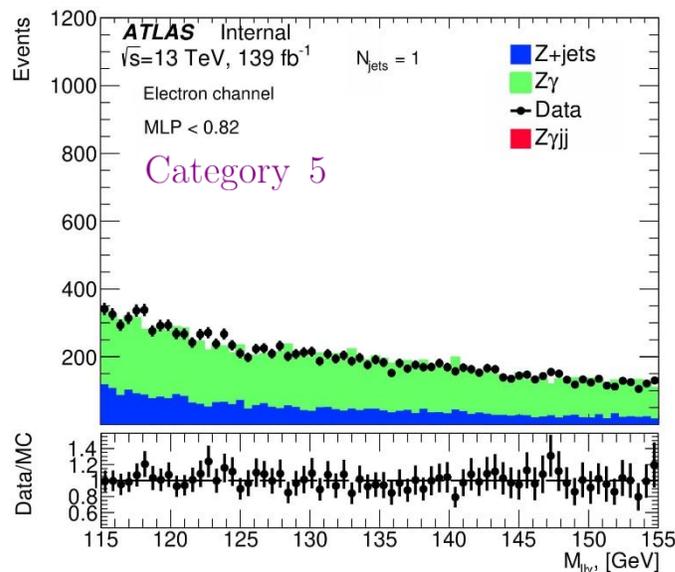
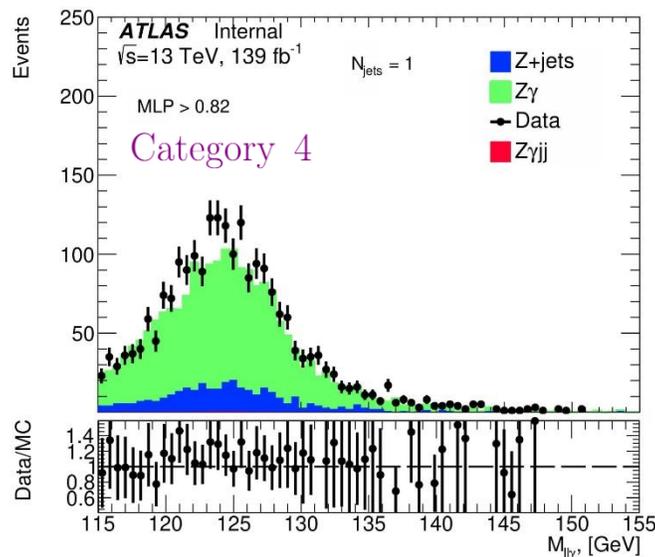
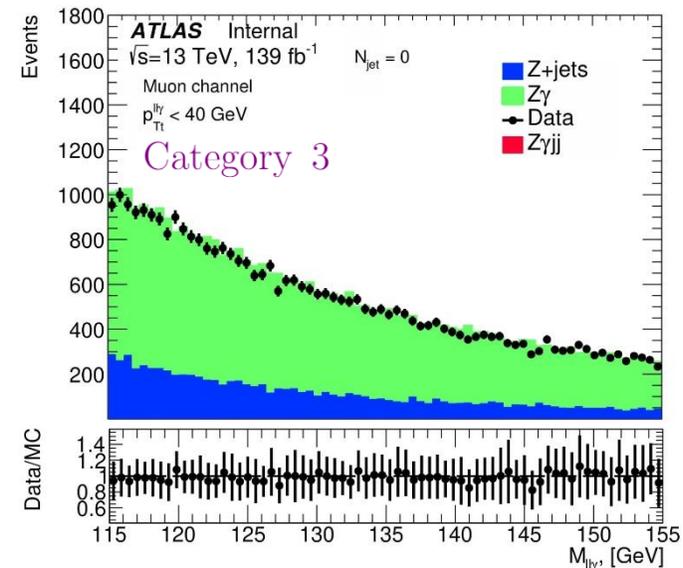
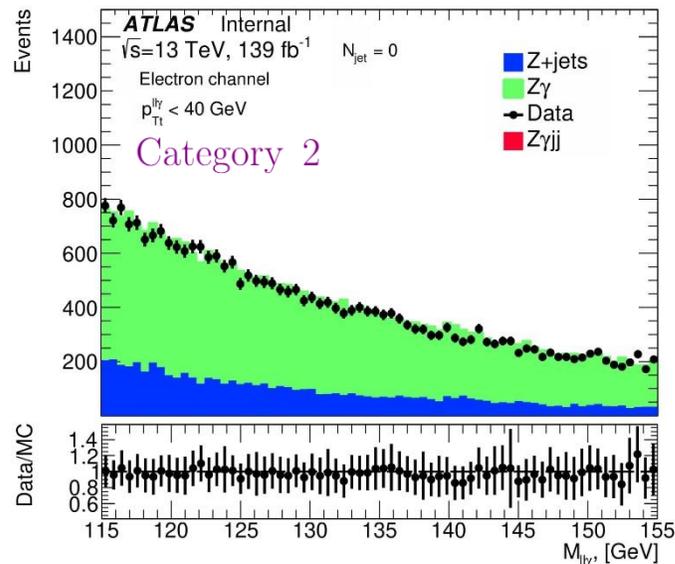
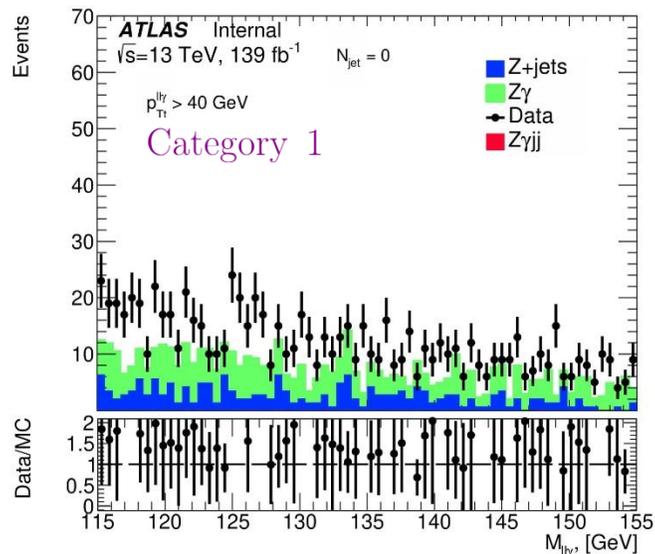
- Так как в анализе рассматривается канал распада $Z \rightarrow \ell\ell$, категории **разделяются по электронному и мюонному каналам**

№	N_{jet}	Category	N_S	N_B	$S/\sqrt{(S+B)}$	$\xi_{Possion}$
1	0	$p_{Tl}^{\ell\gamma} > 40$	1.54	49.55	0.22	0.22
2	0	ee $p_{Tl}^{\ell\gamma} < 40$	15.31	2782.33	0.29	0.29
3	0	$\mu\mu p_{Tl}^{\ell\gamma} < 40$	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

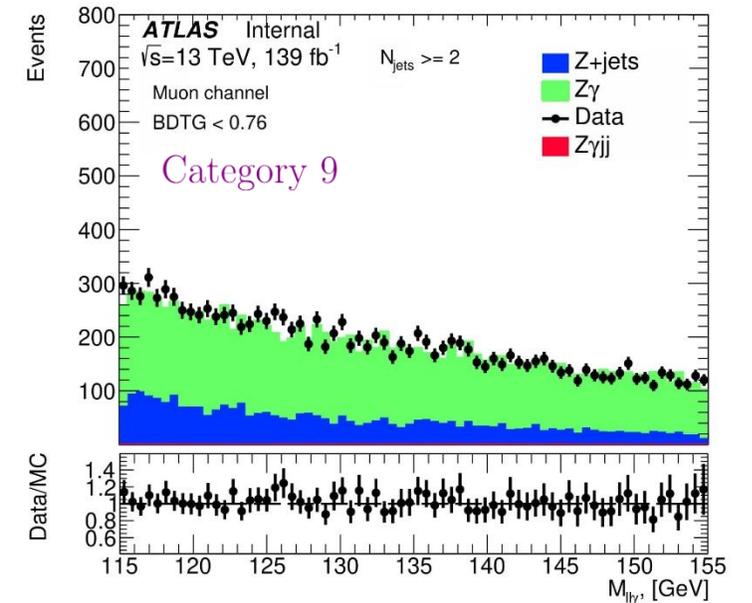
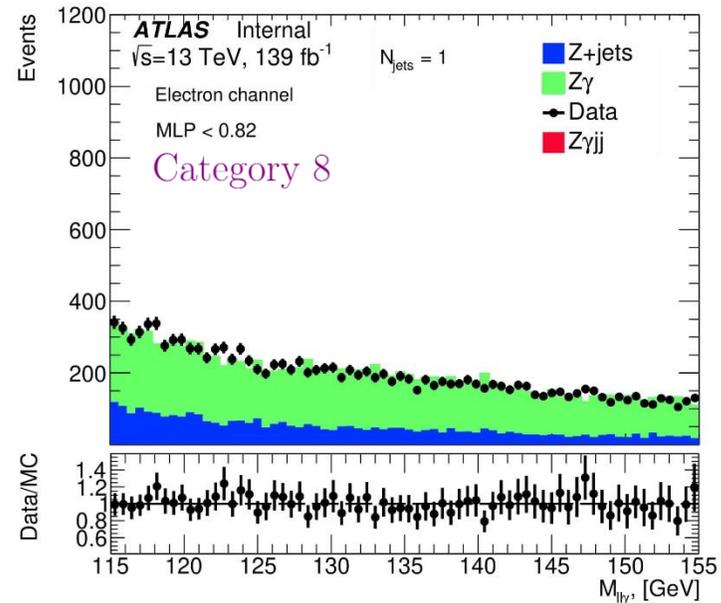
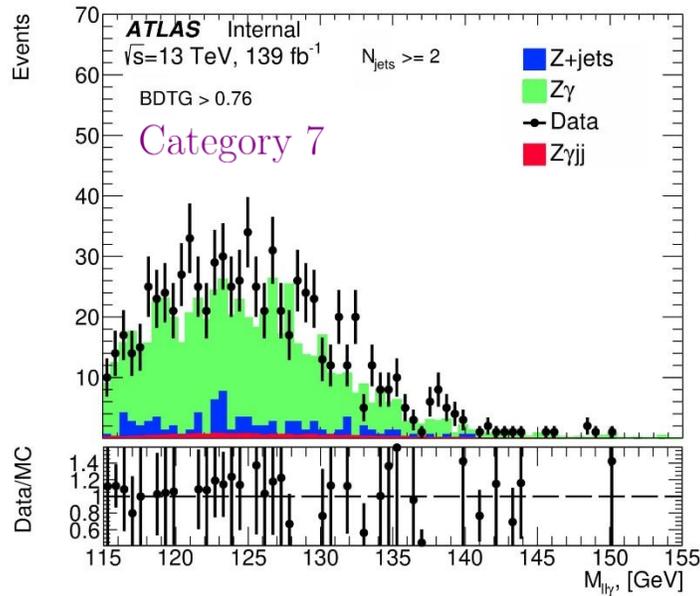
Суммарная чувствительность для всех категорий $\xi = 1.28$

Без категоризации чувствительность $\xi = 0.83$

Сравнение данных и фона

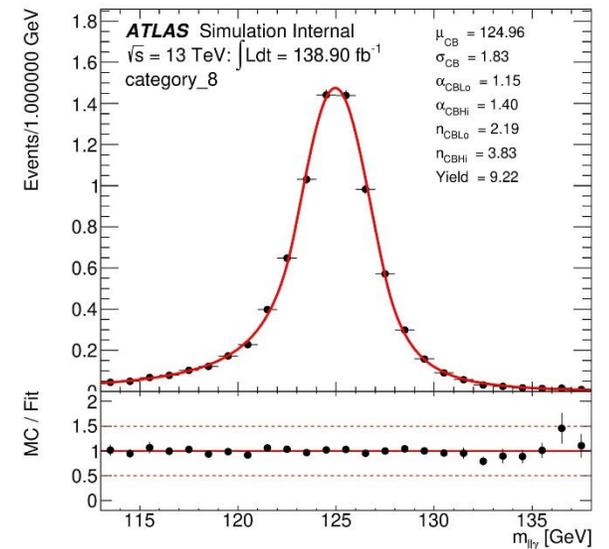
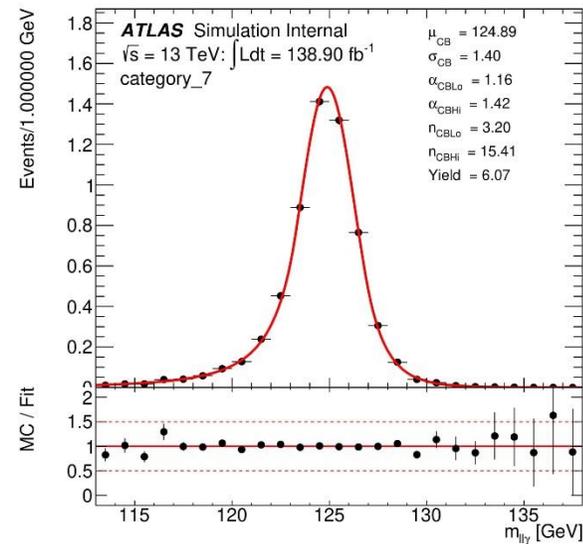
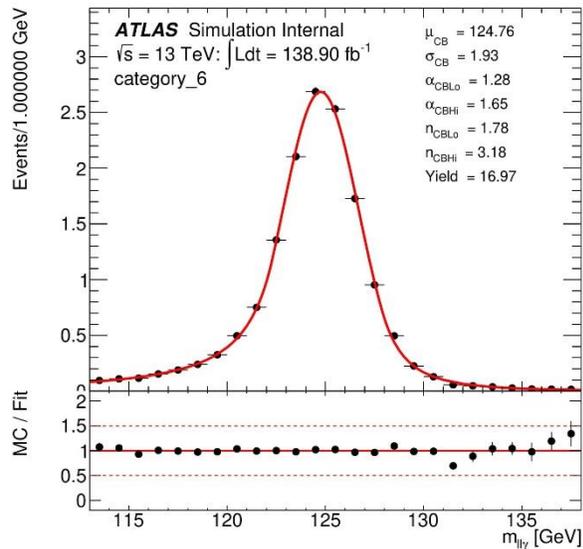
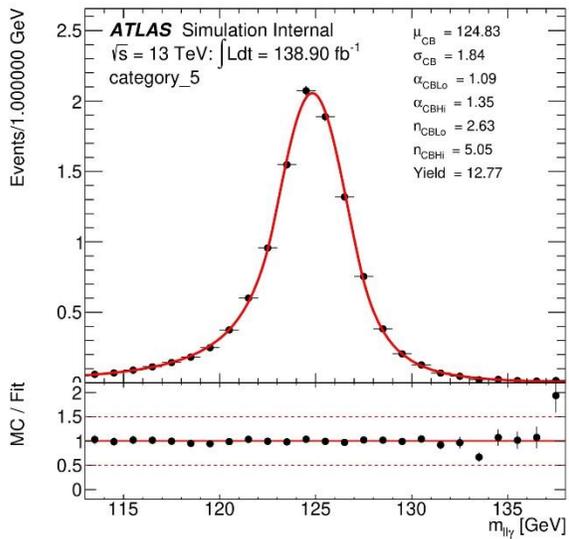
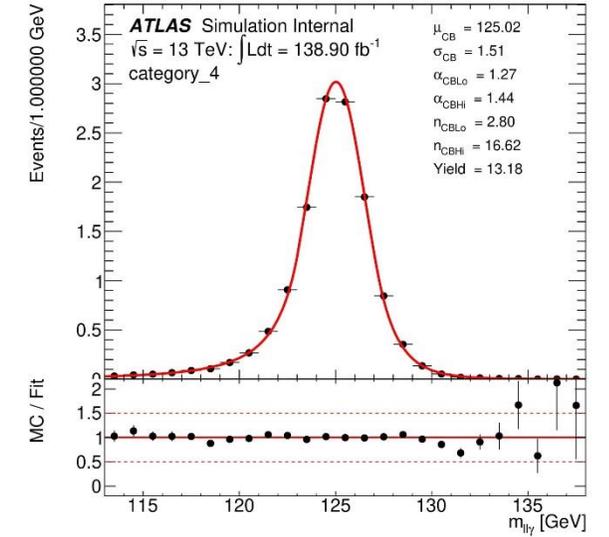
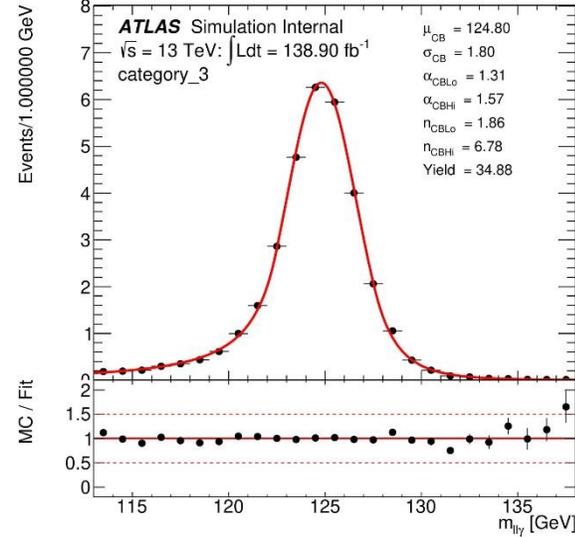
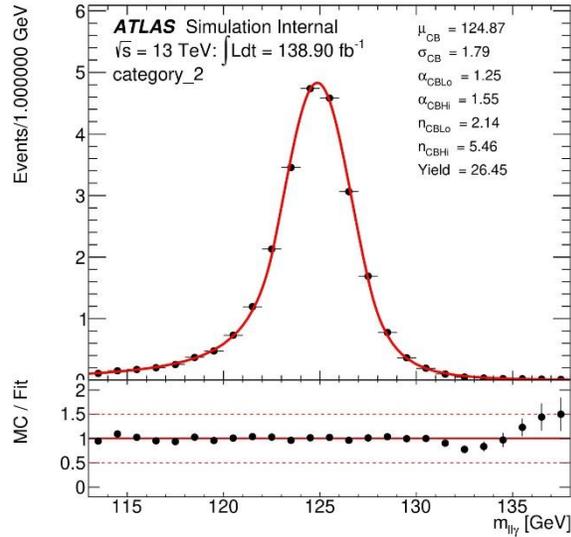
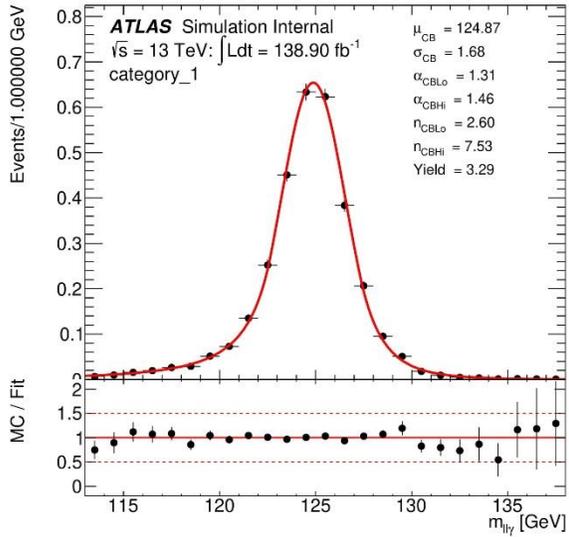


Сравнение данных и фона

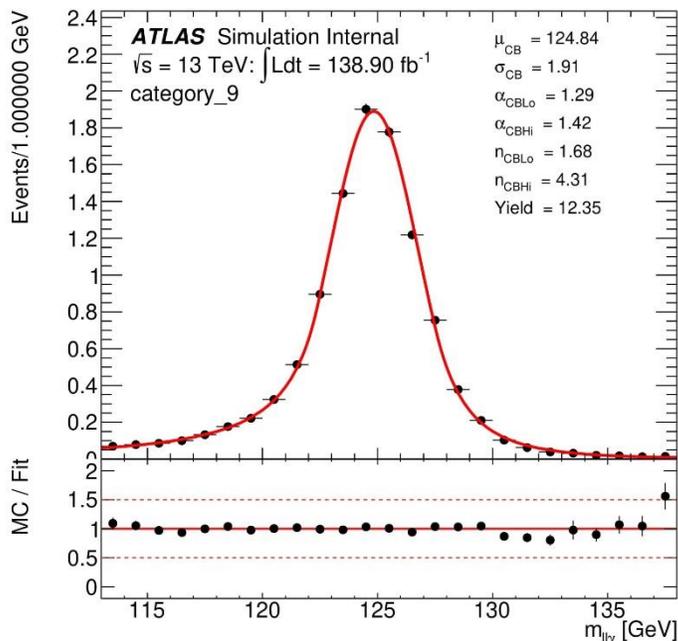


Распределение инвариантной массы для фона простое и **поддается описанию аналитическими функциями** в каждой категории, что говорит о применимости выбранной категоризации

Фитирование сигнала



Фитирование сигнала



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
$N_{jet} = 1, \text{MLP} > 0.82$	125.02 ± 0.65	1.51 ± 0.90	1.84
$N_{jet} = 1, ee, \text{MLP} < 0.82$	124.83 ± 0.94	1.84 ± 1.21	1.38
$N_{jet} = 1, \mu\mu, \text{MLP} < 0.82$	124.76 ± 0.86	1.93 ± 1.02	2.10
$N_{jet} \geq 2, \text{BDTG} > 0.76$	124.89 ± 0.92	1.40 ± 0.95	0.89
$N_{jet} \geq 2, ee, \text{BDTG} < 0.76$	124.96 ± 1.11	1.83 ± 1.13	1.28
$N_{jet} \geq 2, \mu\mu, \text{BDTG} < 0.76$	124.84 ± 0.92	1.91 ± 0.87	1.67

- сигнал в каждой категории **хорошо описывается функцией DSCB** с достаточно небольшой погрешностью
- большая погрешность для значений σ_{CB} обусловлена малым количеством статистики в каждой из категорий

Заключение

- Была **создана категоризация** с использованием оптимальных настроек различных методов машинного обучения. **Суммарная чувствительность всех категорий составляет 1.28**, что повышает общую чувствительность анализа. Для каждой из полученных категорий форма распределения фона поддается описанию аналитическими функциями
- Проведено **фитирование смоделированного сигнального распределения инвариантной массы для каждой категории** функцией DSCB с наиболее оптимальными параметрами, и получено, что форма сигнала описывается аналитическими функциями

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

СПАСИБО ЗА ВНИМАНИЕ!

ВАСКУР СЛАЙДЫ

Используемые наборы данных

Сигнальные МК данные

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

МК данные фонового процесса Z + gamma

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

МК данные фонового процесса Zgammajj

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

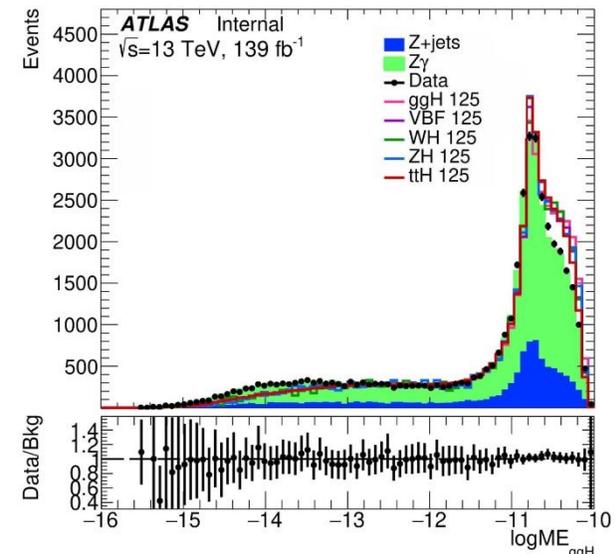
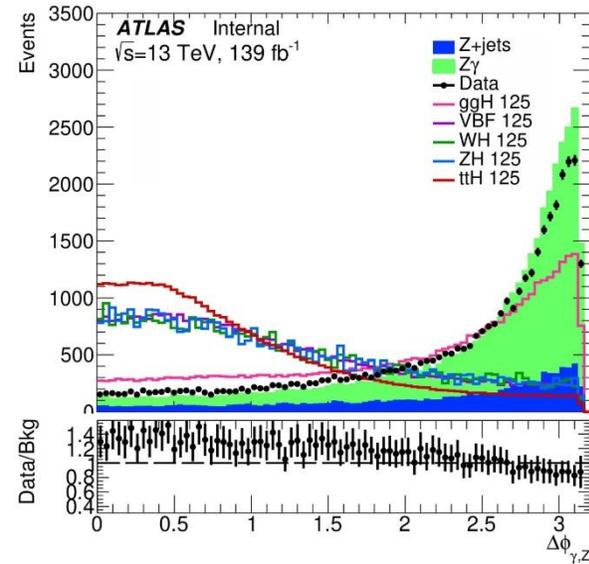
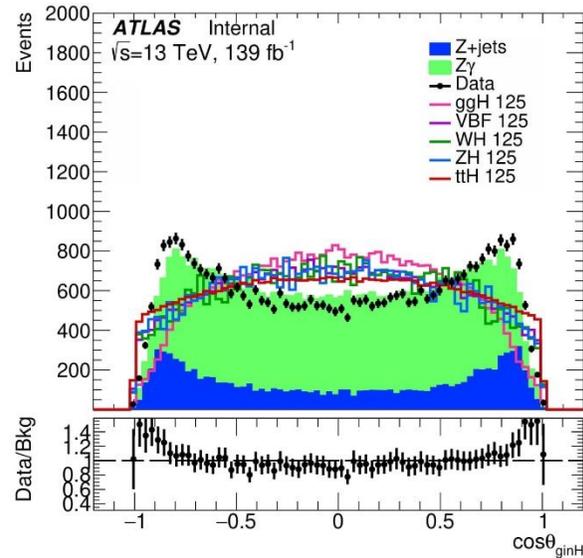
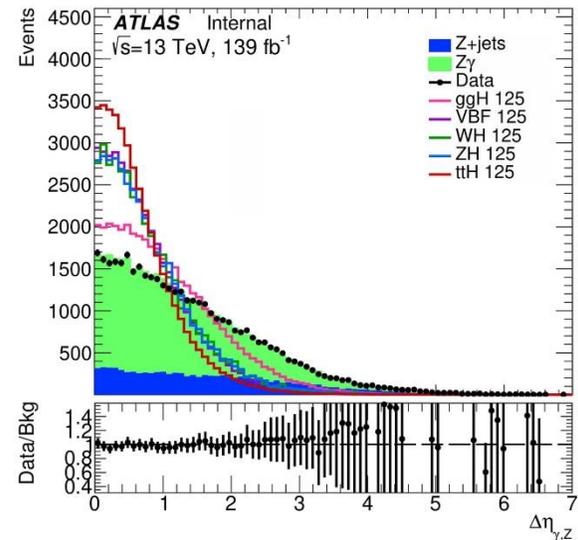
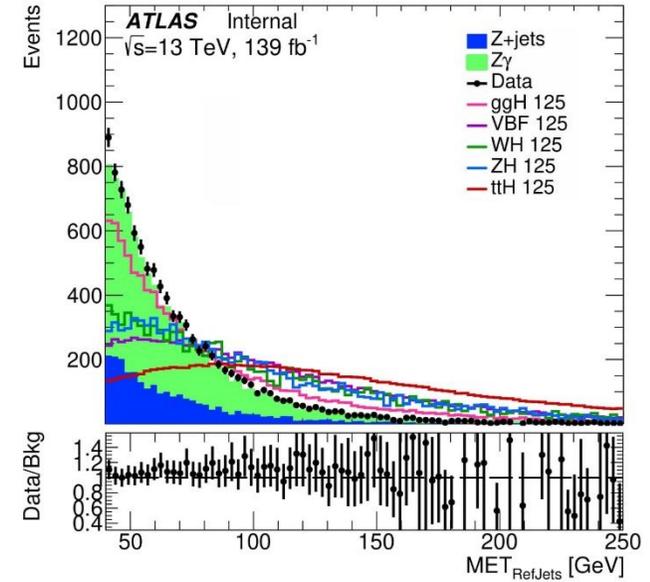
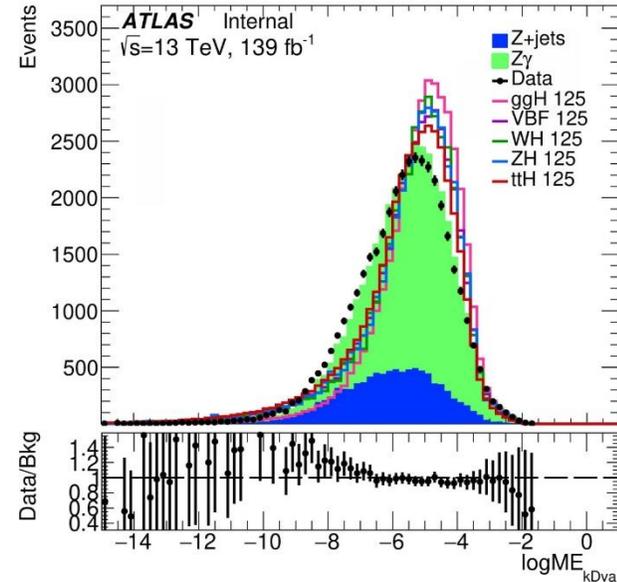
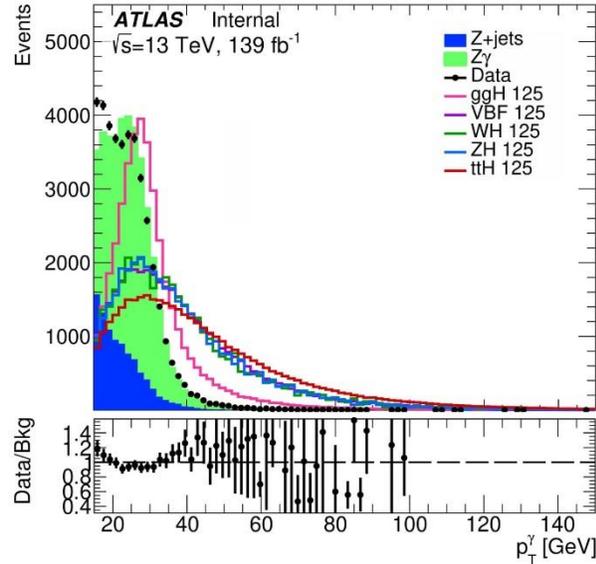
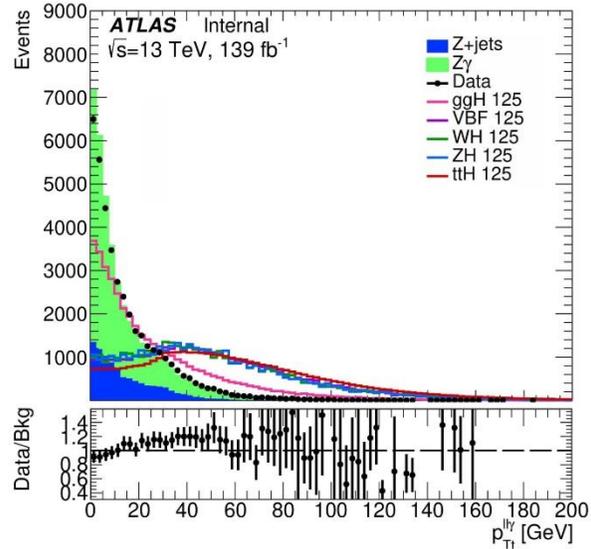
Отбор лептонов и фотонов

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

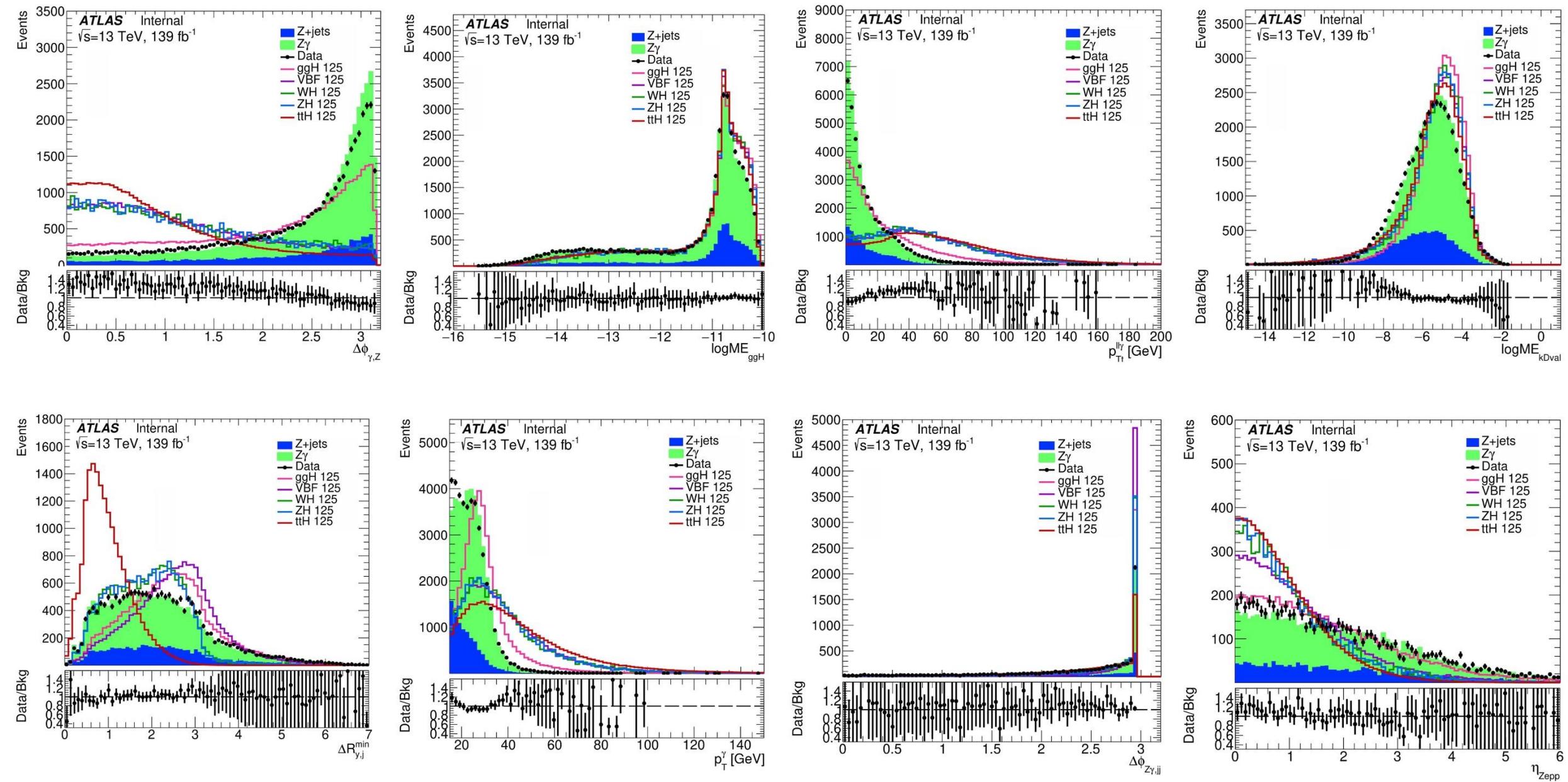
Отбор Z бозона по массе $81.2 \text{ GeV} < m_{ll} < 101.2 \text{ GeV}$

Отбор H бозона по массе $123.59 \text{ GeV} < m_{ll\gamma} < 126.59 \text{ GeV}$

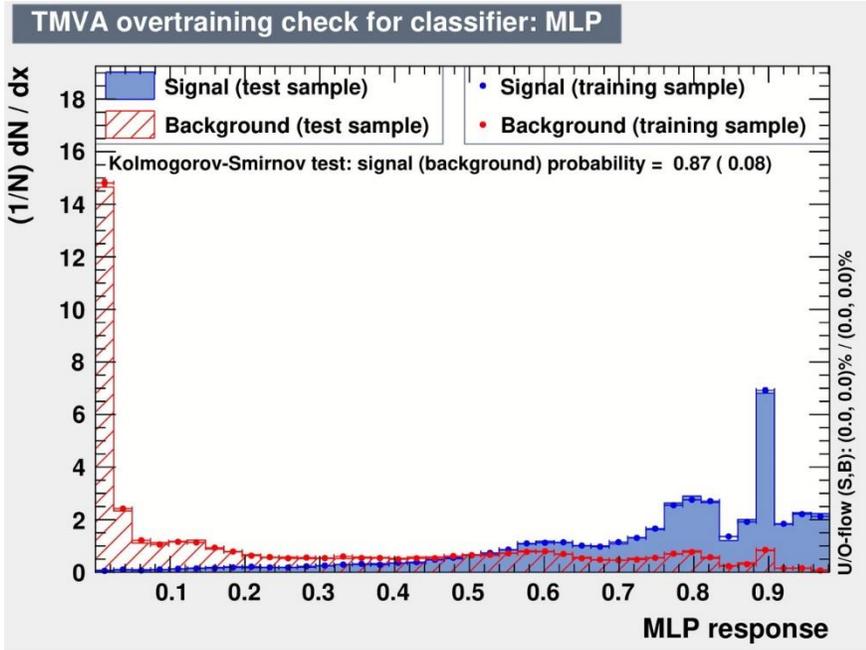
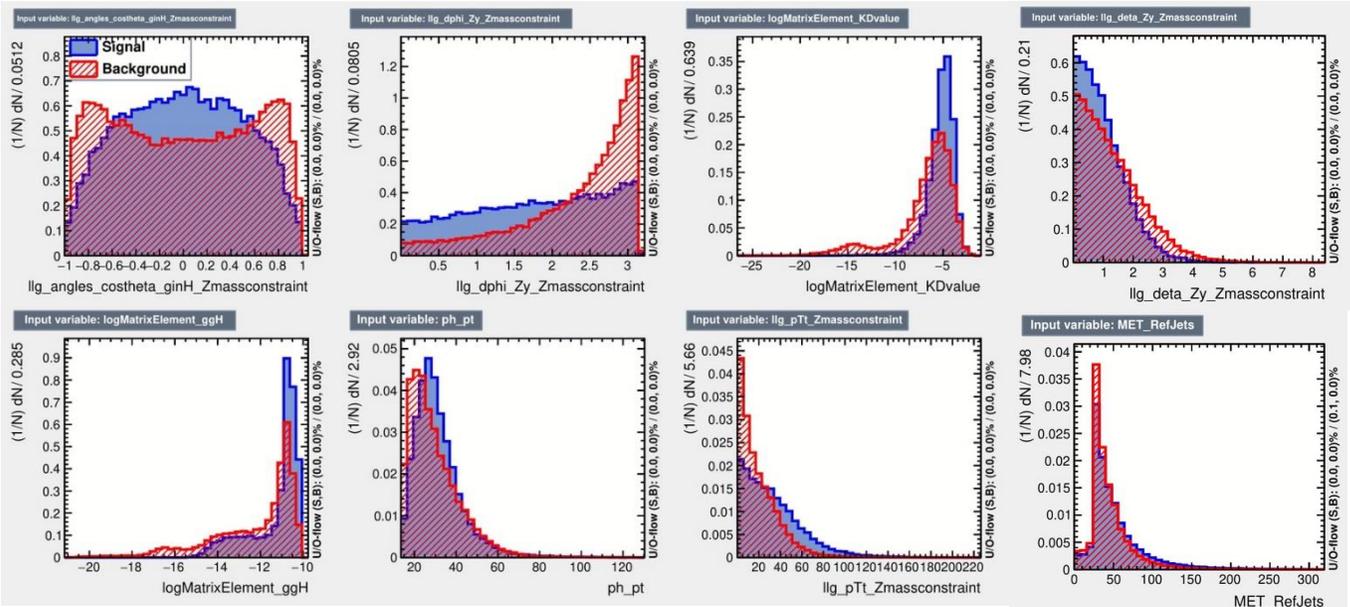
Переменные при тренировке методом MLP



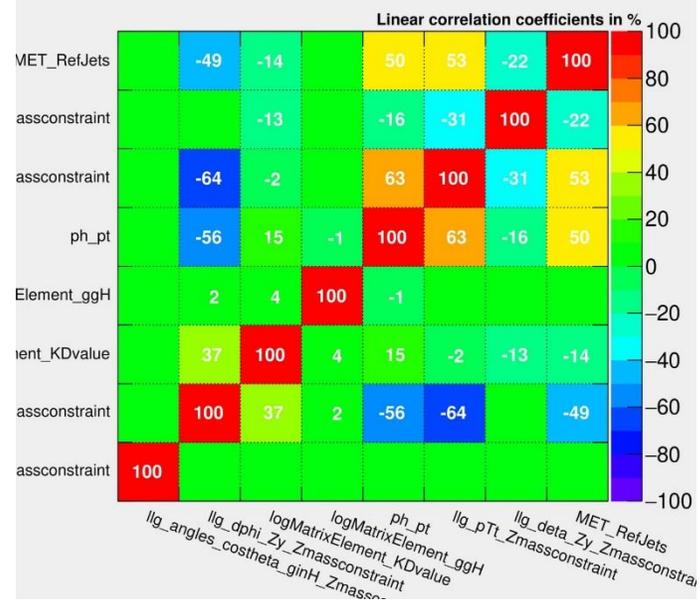
Переменные при тренировке методом BDTG



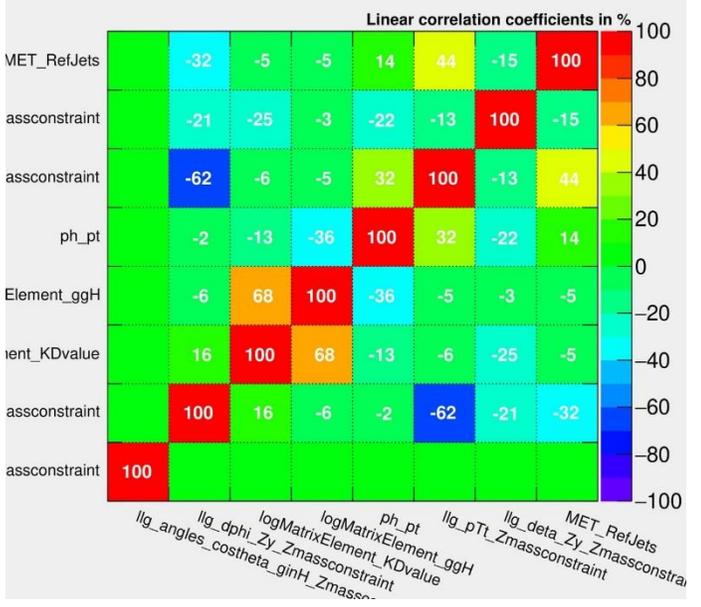
Результаты тренировки методом MLP



Correlation Matrix (signal)



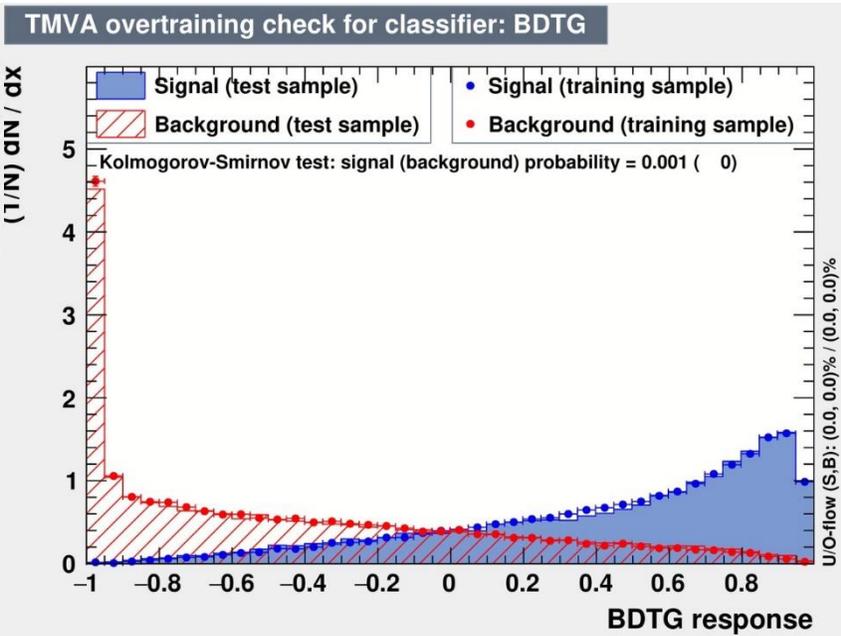
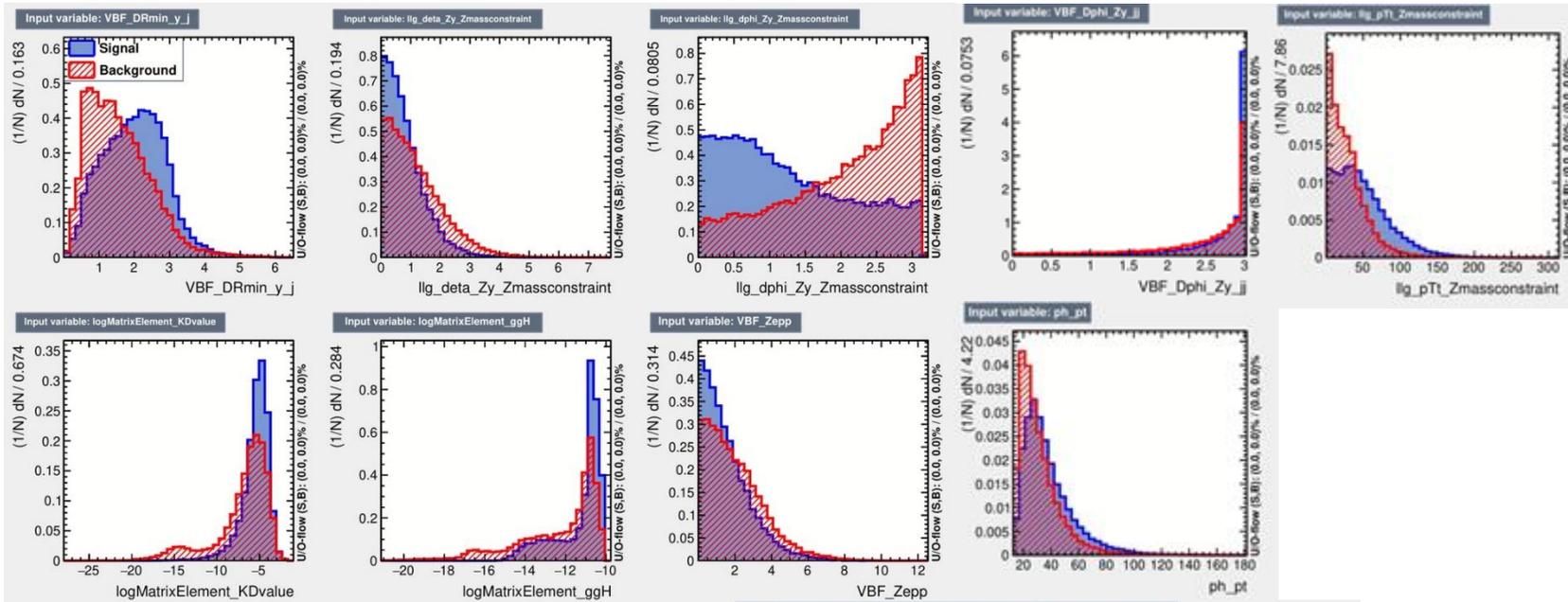
Correlation Matrix (background)



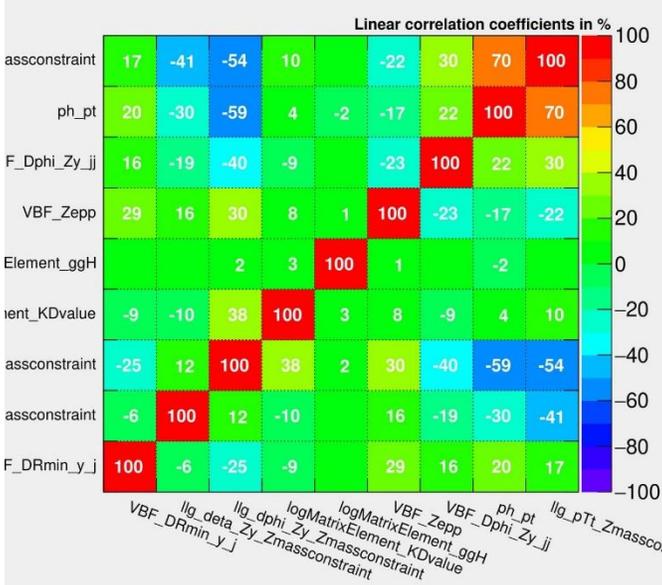
ROC integral 0.890

Настройки метода MLP:
 NeuronType = tanh, VarTransform = N,
 NCycles = 600, HiddenLayers = N+10, N, N,
 TestRate = 5

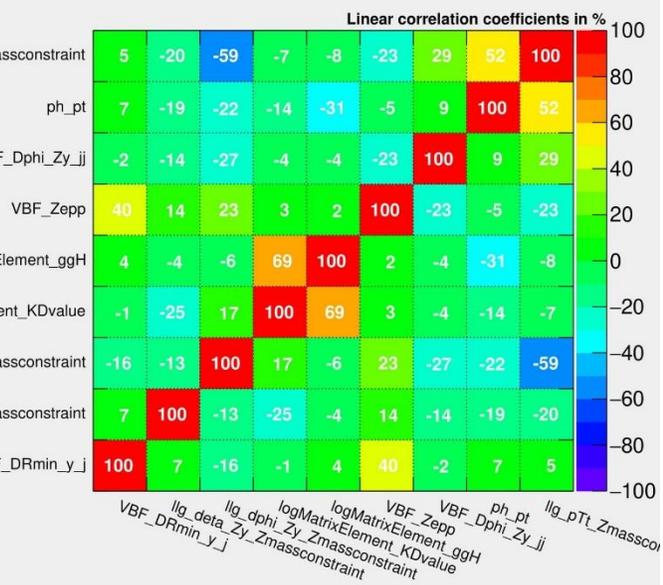
Результаты тренировки методом BDTG



Correlation Matrix (signal)



Correlation Matrix (background)



ROC integral 0.915

Настройки метода BDTG:
 NTrees = 2000, MinNodeSize = 2.5%,
 BoostType = Grad, Shrinkage = 0.10,
 UseBaggedBoost, BaggedSampleFraction = 0.5, nCuts = 20, MaxDepth = 3

Определения переменных при тренировке

MLP метод

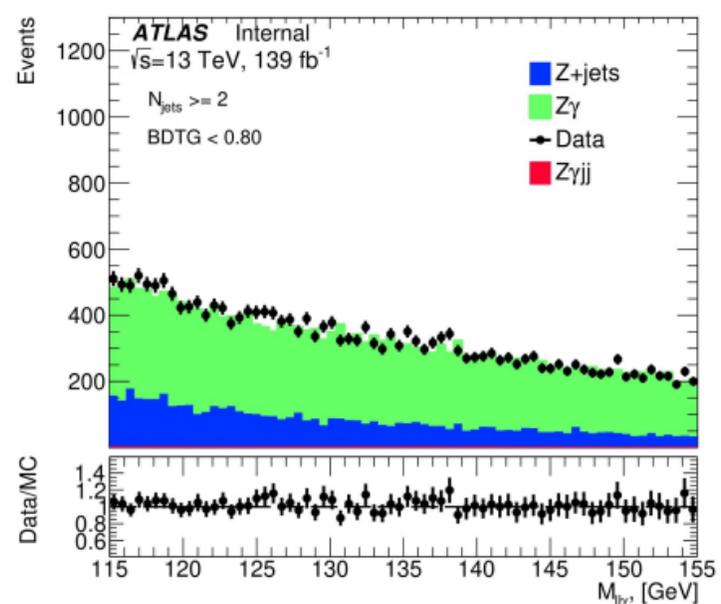
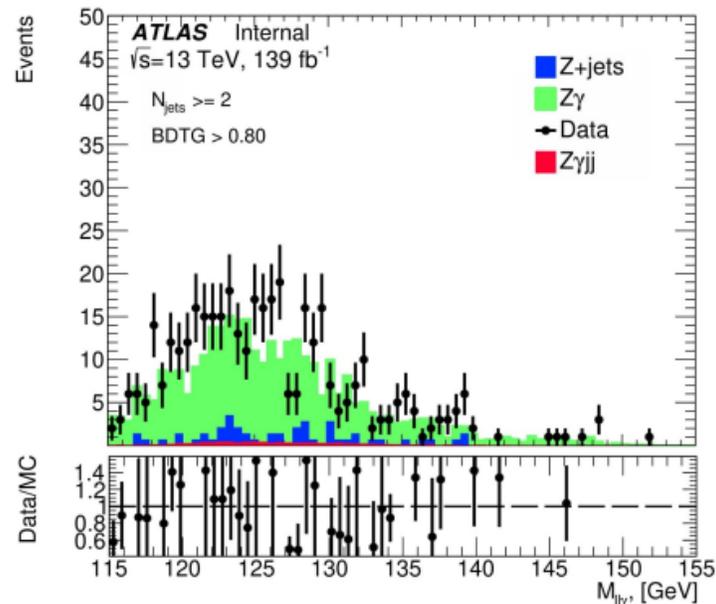
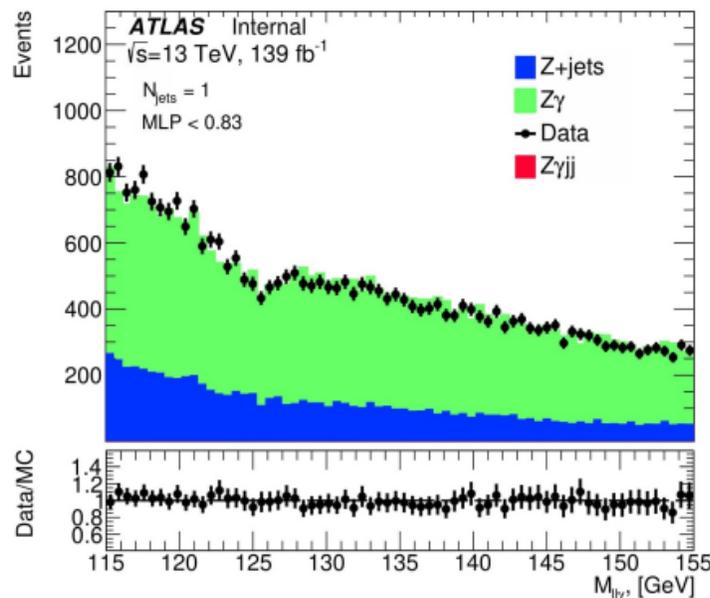
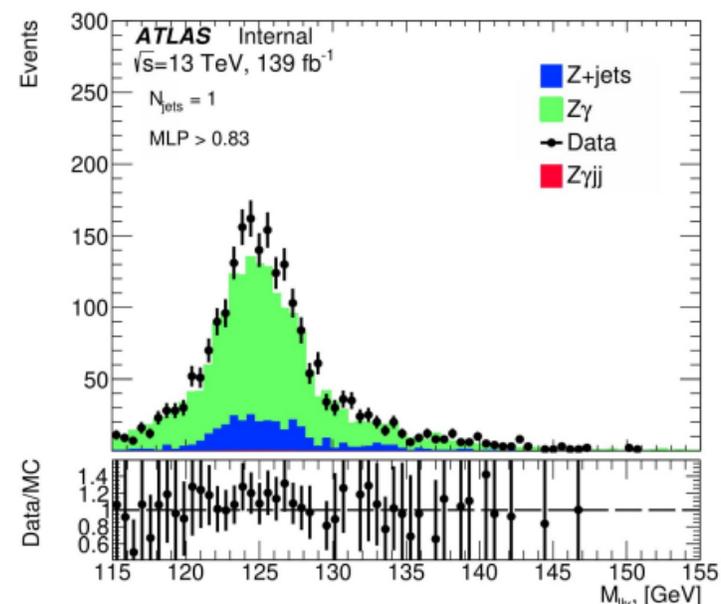
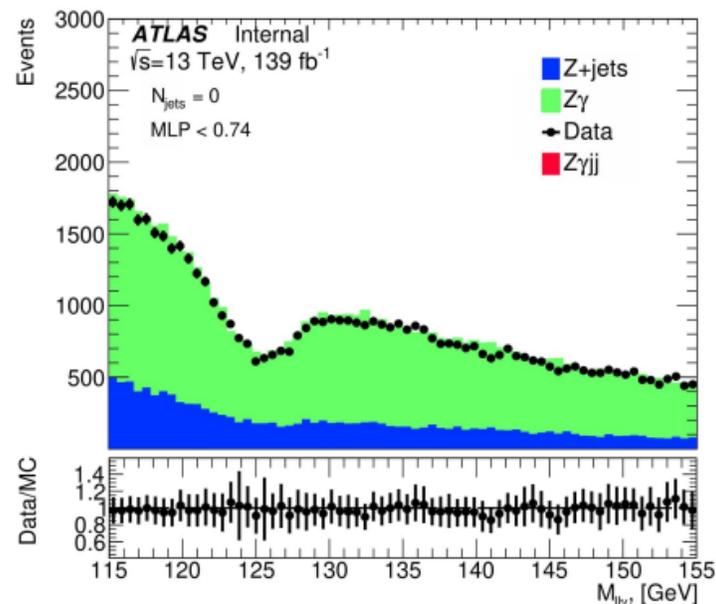
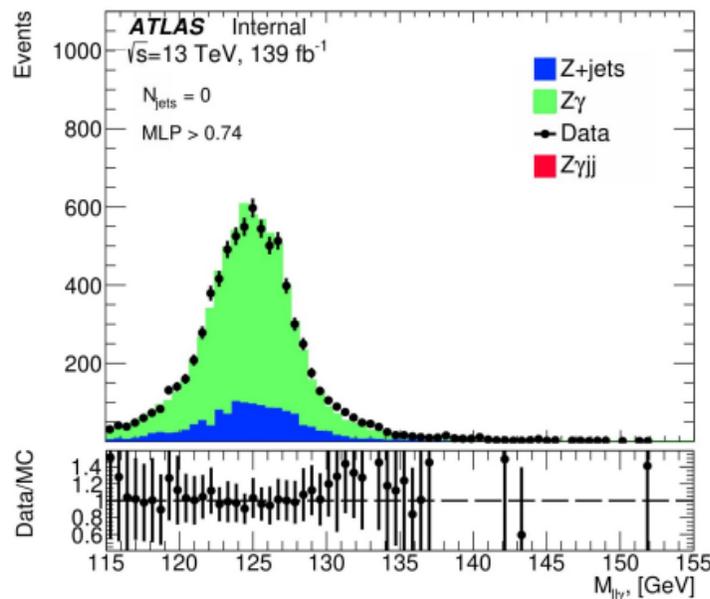
Переменная	Определение	$\langle S^2 \rangle$	I_m	Ранг
$p_{Tt}^{ll\gamma}$	p_T частицы, ортогональный вектору разности между импульсами Z-бозона и фотона	0.084	1370	1
p_T^γ	Поперечный импульс фотона	0.030	1184	2
$\log ME_{kDVal}$	$2 \cdot [\log(ME_{ggH}) - \log(ME_{bkg})]$	0.119	475	3
MET_{Jets}	Потерянный 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

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

BDTG метод

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
$\log ME_{kDVal}$	$2 \cdot [\log(ME_{ggH}) - \log(ME_{bkg})]$	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^{Zerr}	$ \eta_{Z\gamma} - 0.5 \cdot (\eta_{j1} + \eta_{j2}) $	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

Другие способы категоризации



Параметры функции DSCB

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

$$t = \Delta m_H / \sigma_{CB}$$

$$\Delta m_H = M_{Z\gamma} - \mu_{CB}$$

N - нормировочный параметр.

μ_{CB} - пик распределения Гаусса

σ_{CB} - ширина распределения Гаусса