

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

ФИЗИЧЕСКИЙ ФАКУЛЬТЕТ

КАФЕДРА ОБЩЕЙ ЯДЕРНОЙ ФИЗИКИ

МАГИСТЕРСКАЯ РАБОТА

**«МАШИННОЕ ОБУЧЕНИЕ В ЗАДАЧЕ ПРЕДСКАЗАНИЯ
ДИФФЕРЕНЦИАЛЬНЫХ СЕЧЕНИЙ И СТРУКТУРНЫХ ФУНКЦИЙ
ЭЛЕКТРОРОЖДЕНИЯ ОДИНОЧНОГО ПИОНА НА ПРОТОНЕ
В РЕЗОНАНСНОЙ ОБЛАСТИ»**

Выполнила студентка
213М группы
Чистякова Влада Витальевна

Научный руководитель:
к.ф.-м.н., с.н.с. Исупов Евгений Леонидович

Допущена к защите
Зав. кафедрой _____

МОСКВА

2025

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	3
Глава 1. Нуклонные резонансы как инструмент для изучения непертурбативной КХД.....	6
1.1. Возбуждённые состояния нуклонов	6
1.2. Электровозбуждение нуклонных резонансов	7
1.3. Изучение эксклюзивной реакции $e^-p \rightarrow e^-p \pi^0$	9
Глава 2. Описание экспериментальной установки.....	13
2.1. Ускоритель CEBAF	13
2.2. Детектор CLAS	15
2.2.1. Тороидальный магнит	16
2.2.2. Дрейфовые камеры.....	16
2.2.3. Черенковские счётчики	18
2.2.4. Электромагнитные калориметры	18
2.2.5. Система времени пролёта	19
Глава 3. Сбор и хранение экспериментальных данных, полученных на детекторе CLAS.....	20
3.1. Идентификация частиц в детекторе CLAS	20
3.1.1. Идентификация нейтральных частиц.....	20
3.1.2. Идентификация заряженных частиц	21
3.2. Результаты коллаборации CLAS из данных реакций $e^-p \rightarrow e^-p \pi^0$ и $e^-p \rightarrow e^-n \pi^+$	21
3.3. База данных CLAS Physics Database.....	23
3.4. База данных реакции $e^-p \rightarrow e^-p \pi^0$	23
Глава 4. Обзор методов машинного обучения, использованных в работе	25
4.1. Методы классического машинного обучения.....	26
4.2. Полносвязная нейронная сеть.....	32
Глава 5. Результаты.....	36
ЗАКЛЮЧЕНИЕ	49
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	50

ВВЕДЕНИЕ

Существенным достижением фундаментальной физики второй половины XX века стало создание теории сильных взаимодействий – квантовой хромодинамики (КХД), описывающей формирование структуры адронов и атомных ядер [1]. Экспериментальное обнаружение трёхструйных событий в процессе e^-e^+ -аннигиляции, открытие чармония и топ-кварка убедительно подтвердили корректность теории.

Заметным отличием КХД от квантовой электродинамики (КЭД) является существование взаимодействия между переносчиками сильного взаимодействия – глюонами, поскольку они, подобно кваркам, обладают цветовым зарядом, выполняющим роль заряда сильного взаимодействия. Данное явление определяет характер зависимости бегущей константы сильного взаимодействия от расстояния. Если константа электромагнитного взаимодействия возрастает с уменьшением расстояния по причине экранирования электрического заряда виртуально образующимися e^-e^+ -парами, то процессы антиэкранирования цветового заряда превалируют над процессами экранирования цвета кварк-антикварковыми парами, и бегущая константа сильного взаимодействия растёт с увеличением расстояния (рисунок 1).

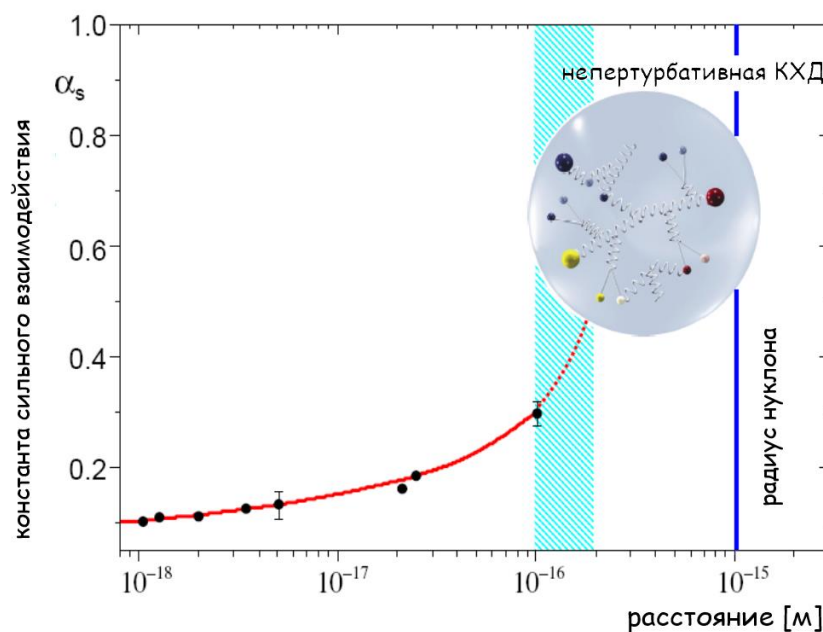


Рисунок 1 – Зависимость константы сильного взаимодействия от расстояния.

Итак, область малых расстояний ($r < 0.1$ фм) соответствует пертурбативной КХД с малой величиной безразмерного параметра связи. Эксперименты при высоких энергиях показали, что адрон состоит из точечных центров рассеяния – партонов [2], впоследствии отождествлённых с кварками и глюонами. При увеличении r константа α_s становится почти равной единице, кварки и глюоны окружаются облаками виртуальных кварк-антикварковых пар, и пертурбативный подход перестаёт быть применимым. Считается, что процесс такого «одевания» точечных кварков и глюонов приводит к формированию примерно 98% массы адрона [3]. К непертурбативным явлениям можно также отнести конфайнмент кварков, имеющий место при $r > 1$ фм. Детальное описание процессов сильного взаимодействия в непертурбативном режиме пока остаётся нерешённой проблемой физики. Вместе с тем переходная область от асимптотической свободы к конфайнменту ответственна за возникновение у нуклонов возбуждённых состояний. Изучение нуклонных резонансов, по-видимому, приблизит нас к пониманию эволюции динамики сильного взаимодействия в зависимости от расстояния.

Информация о спектре и структуре возбуждённых состояний нуклонов главным образом извлекается из реакций эксклюзивного электророждения мезонов на нуклонах. Неоспоримая заслуга в накоплении доминирующей части мировых данных об этих реакциях принадлежит коллаборации CLAS, в состав которой входит НИИ ядерной физики им. Д.В. Скобельцына МГУ.

Важную роль в изучении реакций электророждения мезонов в области возбуждения нуклонных резонансов играют каналы рождения одного и двух пионов. Если масса резонанса не превышает 1.6 ГэВ, то его распад осуществляется преимущественно с испусканием одиночного пи-мезона, исследование однопионного канала необходимо для независимого изучения электромагнитных параметров нуклонных резонансов. В настоящее время экспериментальная информация об однопионном канале доступна только в ограниченной области фазового пространства. Возникает потребность в создании инструмента для интерполяции/экстраполяции имеющихся данных в

области, где экспериментальные данные отсутствуют. Для решения этой задачи можно обратиться к методам машинного обучения, которые успешно применяются в ядерной физике для идентификации частиц, реконструкции их треков, отделения фоновых событий [4-6]. Значимость методов машинного обучения для научных исследований была подтверждена присуждением Нобелевских премий по физике Дж. Хопфилду и Дж. Хинтону и химии Д. Бейкеру, Д. Хассабису и Дж. Джемперу в 2024 г.

Цель работы

Цель моей магистерской работы заключается в построении алгоритма машинного обучения для предсказания дифференциальных сечений реакции электророждения одиночного нейтрального пи-мезона на протоне, способного осуществлять надёжную интерполяцию данных в пятимерном фазовом пространстве.

Глава 1. Нуклонные резонансы как инструмент для изучения непертурбативной КХД

1.1. Возбуждённые состояния нуклонов

Нуклон обладает спектром возбуждённых состояний, что является явным указанием на его сложную внутреннюю структуру. Изучение нуклонных резонансов началось в середине XX в. с рассеяния пи-мезонов на водородных мишенях. Так, в 1952 г. Э. Ферми и Г.Л. Андерсоном был открыт Δ -резонанс [7] при кинетической энергии пиона $T_\pi = 195$ МэВ, что соответствует массе системы пион-протон $1232 \text{ МэВ}/c^2$. Время жизни резонанса составляет $10^{-24} - 10^{-22}$ с и может быть оценено по величине ширины распада Γ с использованием соотношения неопределённостей: $\tau \sim \hbar/\Gamma$. Спин данного Δ -резонанса, равный $3/2$, был найден из анализа углового распределения пи-мезонов от его распада. Δ -резонансы существуют в 4 зарядовых состояниях, следовательно изоспин частицы $I = 3/2$. Позже были обнаружены нуклонные резонансы с изоспином $I = 1/2$, они обозначаются N^* .

Нуклонные резонансы записываются $L_{2I2J}(M)$, где I – изоспин, J – спин, L – орбитальный момент в системе пион-нуклон, M – масса. На рисунке 2 показаны некоторые открытые в реакциях упругого рассеяния пионов на протоне нуклонные резонансы. Видно, что $\Delta(1232)P_{33}$ — единственный изолированный максимум в спектре, он соответствует состоянию нуклона с энергией возбуждения 300 МэВ. Другие максимумы представляют собой суперпозицию нескольких резонансов.

Нуклонные резонансы в целом хорошо согласуются с систематикой кварковой модели. На 2024 г. есть достоверные сведения о 14 N^* и 8 Δ -резонансах с различными массами [8], но кварковая модель допускает наличие большего числа резонансов. Поиск недостающих резонансов или ответ на вопрос, почему некоторые резонансы не проявляются экспериментально, — одна из задач современной физики адронов.

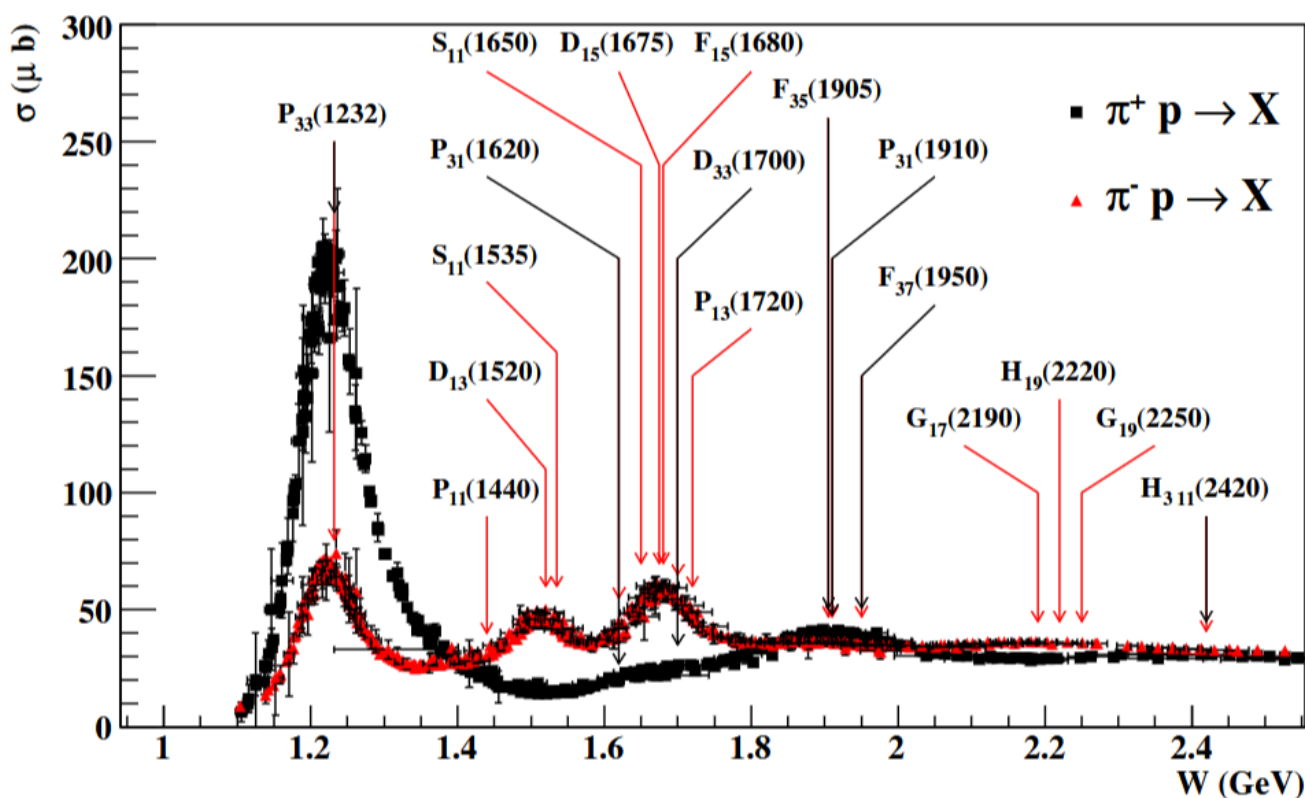


Рисунок 2 – Зависимость полных поперечных сечений реакций π^+p и π^-p от массы системы пион-протон [9].

Низколежащие нуклонные резонансы распадаются на нуклон и пи-мезон. Не все резонансы могут быть обнаружены в реакциях пион-нуклонного рассеяния, поскольку более массивные резонансы распадаются преимущественно с испусканием мезонов, отличных от пионов, или с испусканием пары $\pi^+\pi^-$. Поэтому в современных экспериментах (например, в JLab) нуклонные резонансы изучаются в реакциях электровозбуждения [10-12]. При этом для получения более детальной информации о дифференциальных сечениях таких реакций исследуются эксклюзивные каналы распада резонансов, то есть измеряются кинематические характеристики всех частиц в конечном состоянии.

1.2. Электровозбуждение нуклонных резонансов

На рисунке 3 представлен схематический процесс электромагнитного возбуждения нуклонного резонанса на нуклоне в приближении однофотонного

обмена. Система «виртуальный фотон – протон» имеет 6 возможных спиральных состояний (спиральность фотона: $\lambda_{\gamma_v} = 0, \pm 1$, спиральность протона: $\lambda_p = \pm 1/2$), закон сохранения чётности сокращает их количество вдвое. Следовательно, вершина взаимодействия виртуального фотона с протоном $\gamma_v p N^*$, приводящая к формированию резонанса, может быть полностью охарактеризована тремя амплитудами электровозбуждения $A_{1/2}$, $A_{3/2}$, $S_{1/2}$, зависящими только от квадрата переданного четырёхимпульса фотона (Q^2) [12-13]. Амплитуды $A_{1/2}$, $A_{3/2}$ соответствуют взаимодействию поперечно поляризованного фотона с нуклоном, амплитуда $S_{1/2}$ описывает возбуждение резонансов продольно поляризованными виртуальными фотонами. Если фотон в реакции реальный, то $S_{1/2} = 0$. Индексы у амплитуд показывают проекцию суммарного спина системы начальных частиц на направление импульса фотона в системе центра масс.

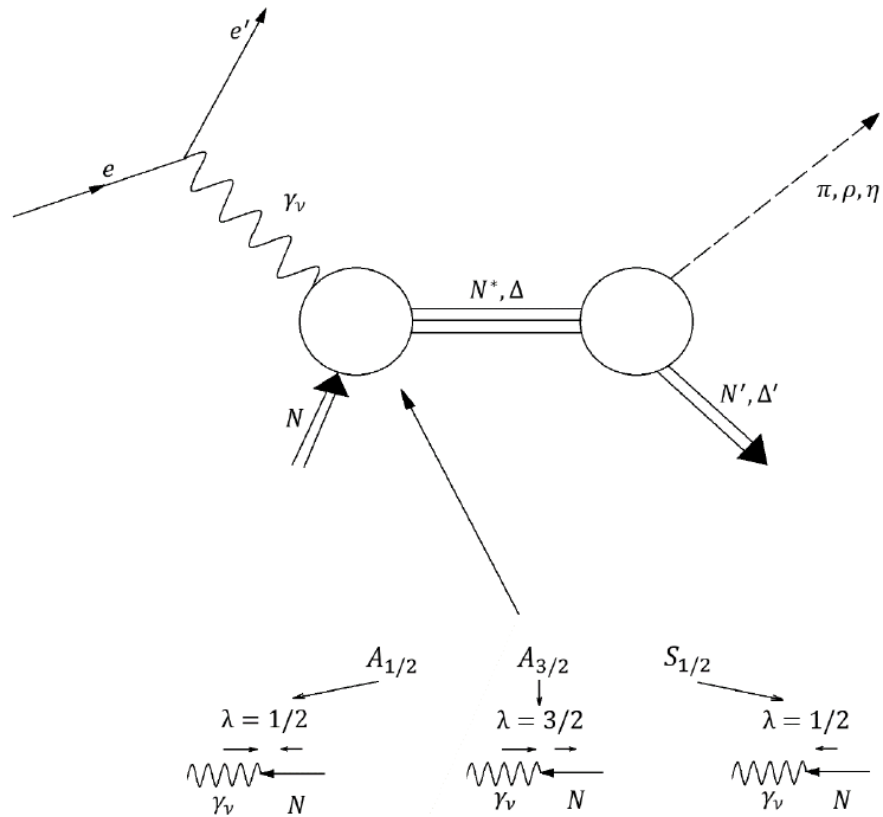


Рисунок 3 – Схема электромагнитного возбуждения резонанса на протоне.

Амплитуды $A_{1/2}$, $A_{3/2}$, $S_{1/2}$ могут быть получены из таких физических наблюдаемых, как сечение в системе центра масс, асимметрия фотонного пучка, поляризация ядра отдачи, асимметрия мишени. Зависимость массы одетого кварка от его импульса проявляется в Q^2 -эволюции амплитуд, их извлечение из эксклюзивных реакций электровозбуждения резонансов важно для понимания динамики сильного взаимодействия.

Амплитуды любой эксклюзивной реакции представляют собой суперпозицию вкладов нуклонных резонансов и совокупности нерезонансных механизмов (рисунок 4) [12, 14-15]. Нерезонансные процессы в различных эксклюзивных каналах различаются, но амплитуды $A_{1/2}$, $A_{3/2}$, $S_{1/2}$ для всех каналов одинаковы. Для надёжного определения амплитуд электровозбуждения необходим большой объём экспериментальных данных по разнообразным эксклюзивным каналам.

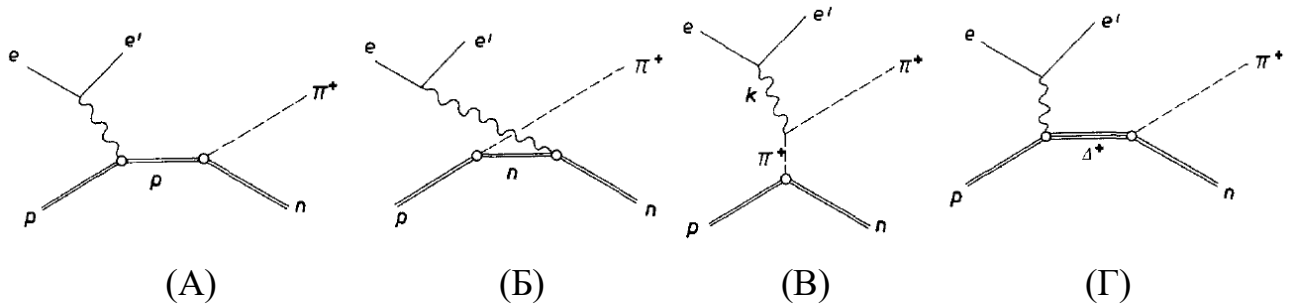


Рисунок 4 – (А), (Б), (В) нерезонансные вклады, (Г) – резонансный вклад реакции $e^- p \rightarrow e^- n \pi^+$.

1.3. Изучение эксклюзивной реакции $e^- p \rightarrow e^- p \pi^0$

В настоящей работе рассматривается эксклюзивная реакция электророждения нейтрального пиона на протоне. В физике резонансов часто пользуются следующими переменными: Q^2 – виртуальность фотона, равная величине квадрата переданного четырёхимпульса фотона q^2 с противоположным знаком, и W – инвариантная масса системы конечных адронов. Виртуальность фотона можно выразить через энергии начального E_e и

конечного электронов E'_e (в предположении равенства единице скорости света $c = 1$):

$$Q^2 = -q^2 = \mathbf{q}^2 - (E_e - E'_e) = 2(E_e E'_e - |\mathbf{p}_e| |\mathbf{p}'_e| \cos \theta_e) - 2m_e^2 \approx$$

$$\approx 4E_e E'_e \sin^2 \left(\frac{\theta_e}{2} \right),$$

где $\theta_e = \arccos \frac{p'_z}{E'_e}$ – угол рассеяния электрона в лабораторной системе (p'_z – проекция импульса рассеянного электрона на ось z), m_e – масса электрона, $\mathbf{q} = \mathbf{p}_e - \mathbf{p}'_e$ – переданный импульс фотона, $\mathbf{p}_e, \mathbf{p}'_e$ – импульсы налетающего и рассеянного электронов. Инвариантная масса конечных адронов W :

$$W = (P_p + q)^2 = M_p^2 + 2M_p(E_e - E'_e) - Q^2,$$

где P_p и M_p – четырёхимпульс протона и масса протона соответственно.

Дифференциальное сечение под действием виртуальных фотонов на неполяризованном нуклоне в случае неполяризованного электронного пучка и образования двух адронов в конечном состоянии описывается выражением [14-16]:

$$\frac{d\sigma_{\gamma v}}{d\Omega_\pi} = \frac{d\sigma_u}{d\Omega_\pi} + \varepsilon \frac{d\sigma_{tt}}{d\Omega_\pi} \cdot \cos 2\varphi_\pi + \sqrt{2\varepsilon(1+\varepsilon)} \frac{d\sigma_{lt}}{d\Omega_\pi} \cdot \cos \varphi_\pi \equiv$$

$$\equiv A + B \cos 2\varphi_\pi + C \cos \varphi_\pi,$$

$$\frac{d\sigma_u}{d\Omega_\pi} = \frac{d\sigma_t}{d\Omega_\pi} + \varepsilon \frac{d\sigma_l}{d\Omega_\pi},$$

где $\frac{d\sigma_u}{d\Omega_\pi}, \frac{d\sigma_t}{d\Omega_\pi}, \frac{d\sigma_l}{d\Omega_\pi}, \frac{d\sigma_{lt}}{d\Omega_\pi}, \frac{d\sigma_{tt}}{d\Omega_\pi}$ – неполяризованная, поперечная, продольная, продольно-поперечная, поперечно-поперечная структурные функции соответственно, ε – поляризация виртуального фотона:

$$\varepsilon = \left(1 + 2 \left(1 + \frac{v^2}{Q^2} \right) \tan^2 \frac{\theta_e}{2} \right)^{-1},$$

$v = E_e - E'_e$ – переданная электроном энергия, или энергия фотона в лабораторной системе. φ_π – это угол между плоскостью рассеяния электрона и плоскостью реакции (плоскостью, образованной начальным нуклоном и

конечными частицами), θ_π – полярный угол вылета пиона в системе центра масс «виртуальный фотон – протон», они указаны на рисунке 5.

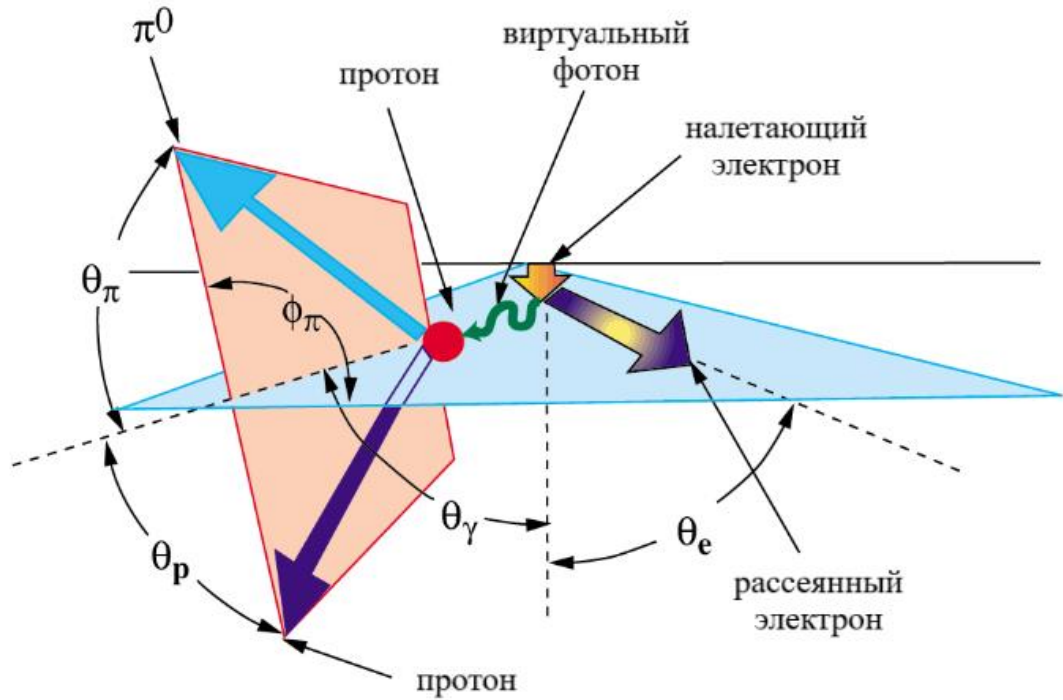


Рисунок 5 – Схематическое изображение углов θ_e , φ_π , θ_π реакции $e^- p \rightarrow e^- p \pi^0$.

При $Q^2 \rightarrow 0$ структурные функции $\frac{d\sigma_l}{d\Omega_\pi}$ и $\frac{d\sigma_{lt}}{d\Omega_\pi}$ также стремятся к нулю, и сечение эксклюзивных реакций электророждения мезонов на неполяризованном нуклоне под действием неполяризованных электронов переходит в сечение реакций под действием реальных неполяризованных фотонов на неполяризованном нуклоне.

Структурные функции не зависят от φ_π и ε , они могут быть получены из дифференциального сечения:

$$A(E_e, Q^2, W, \cos \theta_\pi) = \frac{1}{2\pi} \int_0^{2\pi} \frac{d\sigma_{\gamma v}(\varphi_\pi)}{d\Omega_\pi} d\varphi_\pi,$$

$$B(E_e, Q^2, W, \cos \theta_\pi) = \frac{1}{\pi} \int_0^{2\pi} \frac{d\sigma_{\gamma v}(\varphi_\pi)}{d\Omega_\pi} \cos 2\varphi_\pi d\varphi_\pi,$$

$$C(E_e, Q^2, W, \cos \theta_\pi) = \frac{1}{\pi} \int_0^{2\pi} \frac{d\sigma_{\gamma v}(\varphi_\pi)}{d\Omega_\pi} \cos \varphi_\pi d\varphi_\pi.$$

В настоящее время существует два подхода для извлечения амплитуд электровозбуждения из экспериментальных данных эксклюзивных реакций $e^-p \rightarrow e^-p \pi^0$ и $e^-p \rightarrow e^-n \pi^+$: модель дисперсионных соотношений (Dispersion relations, DR) и унитарная изобарная модель (Unitary isobar model, UIM) [17]. На рисунке 6 представлена зависимость вычисленных амплитуд $A_{1/2}$, $S_{1/2}$ на основе двух подходов для Ропер резонанса от Q^2 .

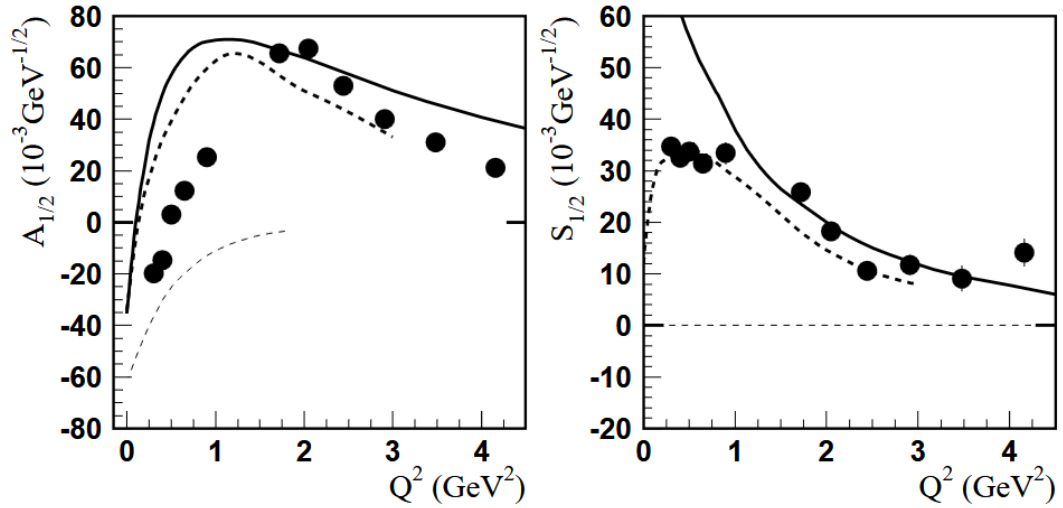


Рисунок 6 – Амплитуды электровозбуждения $A_{1/2}$, $S_{1/2}$ для перехода $\gamma_v p \rightarrow N(1440)P_{11}$. Кружки – усреднённые результаты подходов DR и UIM для данных коллаборации CLAS. Жирные сплошная и пунктирная линии – теоретические расчёты амплитуд в предположении, что Ропер резонанс – радиальное возбуждение трёхкваркового состояния, тонкая пунктирная линия – в предположении, что резонанс – возбуждение гибридного бариона ($q^3 g$).

Таким образом, анализ дифференциальных сечений и других наблюдаемых эксклюзивных каналов электророждения мезонов позволяет определить структурные функции, с помощью которых можно выразить амплитуды электровозбуждения резонансов. Эти данные, в свою очередь, служат источником информации о динамике сильного взаимодействия.

Глава 2. Описание экспериментальной установки

Эксклюзивные реакции под действием реальных и виртуальных фотонов служат наиболее удобным инструментом для изучения возбуждённых состояний нуклонов. Для успешного исследования таких реакций требуется использование электронных ускорителей с высоким фактором заполнения пучка, а также детекторов, охватывающих почти полный телесный угол эмиссии конечных частиц. Данным характеристикам удовлетворяют электронный ускоритель CEBAF [18] и детектор CLAS [19] Национальной Лаборатории им. Томаса Джефферсона (JLab). Основанная в 1984 г. в городе Ньюпорт-Ньюс, США, она является одним из важнейших мировых центров по изучению внутренней структуры нуклона. Программа исследований Лаборатории охватывает также поиск недостающих резонансов и экзотических частиц, изучение конфайнмента кварков.

2.1. Ускоритель CEBAF

Электронный ускоритель непрерывного действия CEBAF (the continuous electron beam accelerator facility) состоит из двух параллельных линейных ускорителей длиной 1.4 км, соединённых пятью арками (рисунок 7).

Для создания электронного пучка в инжекторе на арсенидогаллиевый катод направляют луч лазера, и за счёт фотоэффекта из него вылетают электроны с энергией 67 МэВ. Свободные электроны группируются электромагнитами в последовательность сгустков по миллиону частиц в каждом, временной интервал между двумя соседними сгустками составляет 0.67 нс. Временное разрешение детектора ~ 1 нс, поэтому электронный пучок можно считать непрерывным. Степень поляризации электронов в пучке достигает 90%.

Электроны ускоряются в северном и южном линейных ускорителях со средним темпом ускорения 7.5 МэВ/м. Применяется технология сверхпроводящих ускоряющих резонаторов (superconducting radiofrequency, SRF). Ниобиевые резонаторы (рисунок 8) представляют собой полости (камеры), внутри которых поддерживаются стоячие электромагнитные волны с частотой

1497 МГц. Резонаторы с помощью жидкого гелия охлаждаются до температуры 2 К, ниобий при таких условиях становится сверхпроводником.

Арки поворотных магнитов обеспечивают пятикратный проход электронного пучка через оба ускорителя. Максимальная энергия электронного пучка – 6 ГэВ. Ускоренный пучок электронов направляется в один из трёх экспериментальных залов, детектор CLAS расположен в зале В.

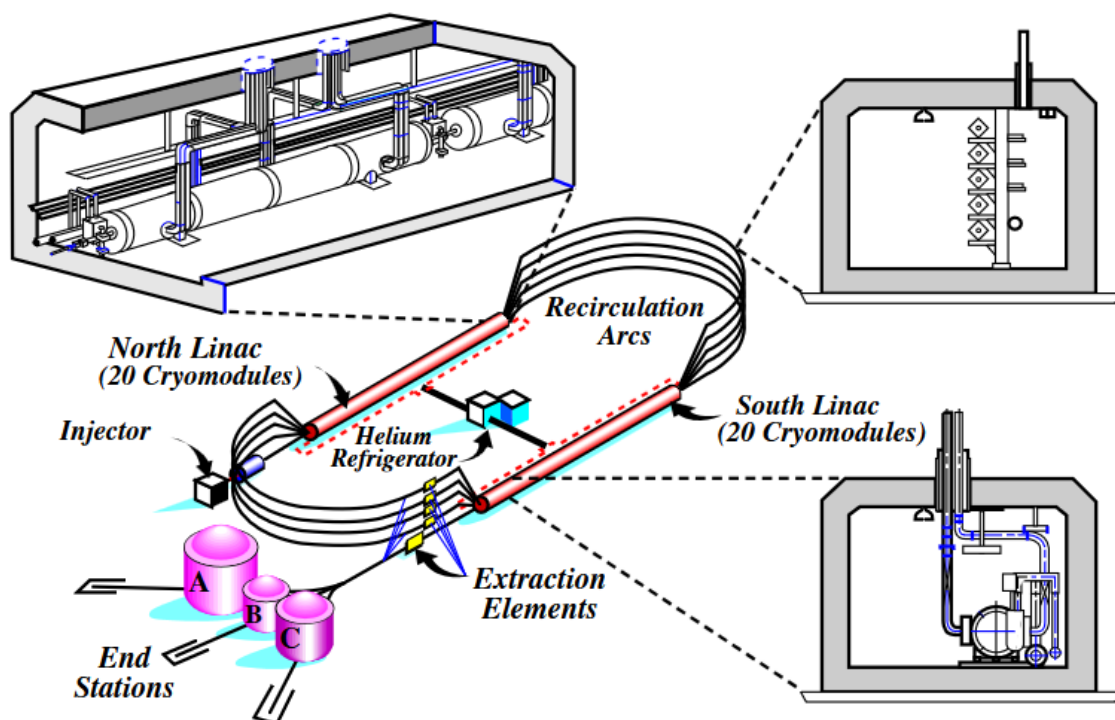


Рисунок 7 – Схематическое устройство ускорителя CEBAF.



Рисунок 8 – Ниобиевые ускоряющие пятикамерные структуры, длина каждой – 0.5 м.

2.2. Детектор CLAS

Детектор CLAS (the CEBAF large acceptance spectrometer) предназначен для работы с пучками электронов и фотонов до 6 ГэВ со 100% фактором заполнения. Благодаря аксептансу 4π детектор способен надёжно регистрировать до 6 частиц в каждом событии и определять их четырёхимпульсы и углы рассеяния при светимости $10^{34} \text{ см}^{-2}\text{с}^{-1}$. В среднем обрабатывается около 2 тысяч событий в секунду. В качестве мишени чаще всего используется жидкий водород, реже – жидкий дейтерий и гелий, а также ядра углерода, железа и свинца.

Детектор имеет почти сферическую форму с диаметром ~ 9 м, на рисунке [9](#) приведено его схематическое устройство.

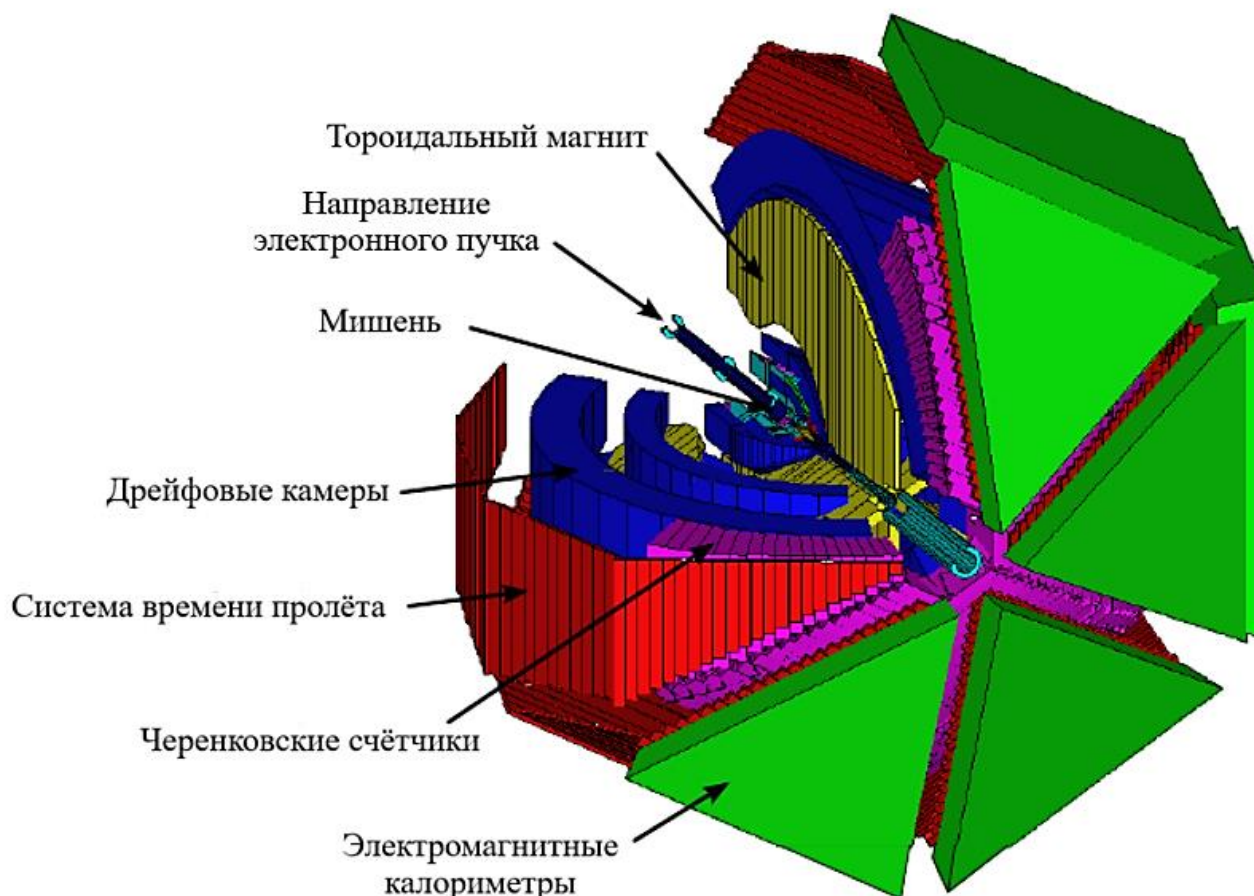


Рисунок 9 – Схематическое устройство детектора CLAS.

2.2.1. Тороидальный магнит

Магнитное поле в детекторе CLAS обеспечивается шестью сверхпроводящими сегментами тороидального магнита (рисунок 10 А), аксиально симметрично расположенными вокруг линии электронного пучка. В отличие от соленоида тороидальная форма позволяет получить свободную от магнитного поля внутреннюю область, что важно при использовании поляризованных мишеней. Почковидная форма катушек формирует высокий интеграл поля для частиц, движущихся под малыми углами по отношению к оси первичного пучка (обычно с высоким импульсом) и более низкий интеграл поля для частиц, выпускаемых под большими углами (рисунок 10 Б).

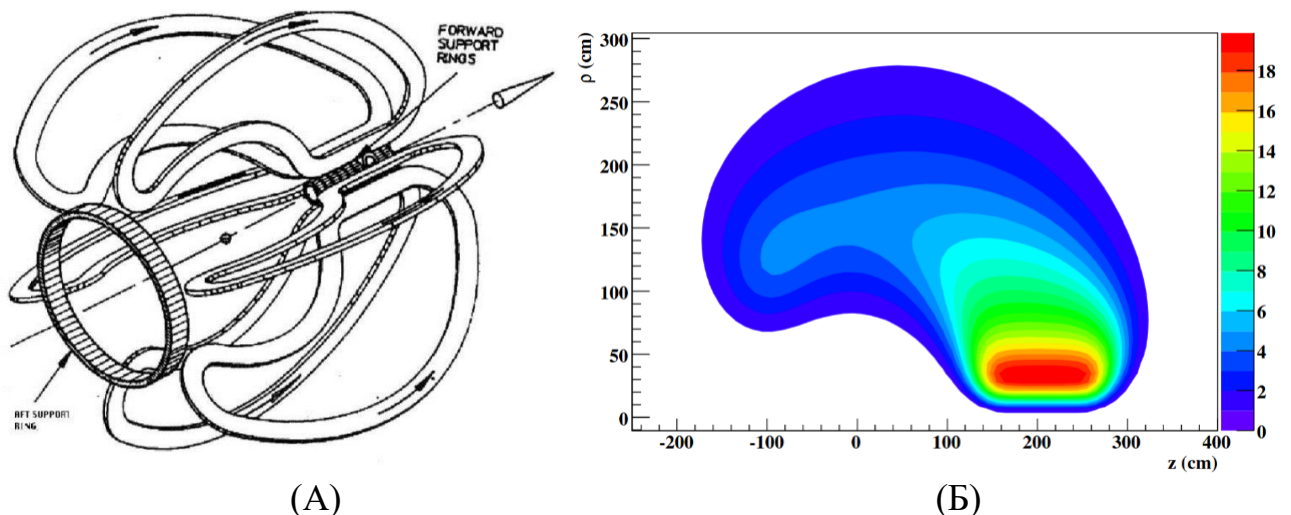


Рисунок 10 – (А) Конфигурация тороидального магнита. (Б) Индукция магнитного поля в плоскости между двумя катушками, единица измерения у цветовой панели – кГс, z – координата в направлении электронного пучка (мишень располагается в $z = -10$ см), ρ – радиальная координата до электронного пучка.

2.2.2. Дрейфовые камеры

Детектор CLAS способен эффективно регистрировать заряженные частицы с импульсами более 200 МэВ/с в диапазоне полярных углов $\theta \in [8, 142]^\circ$. Тороидальный магнит отклоняет заряженные частицы от направления первичного электронного пучка, а 18 дрейфовых камер,

сгруппированных в трёх зонах (Region 1, 2, 3), служат для определения их импульса и траектории (рисунок 11 А). Внутреннее строение дрейфовых камер представляет собой шестиугольные ячейки, в вершинах которых находятся полевые провода при отрицательном потенциале, а в центре каждой ячейки – чувствительные провода при положительном потенциале (рисунок 11 Б). При взаимодействии с проводами заряженная частица индуцирует ток, траекторию её движения можно установить по собранным сигналам от определённых проводов.

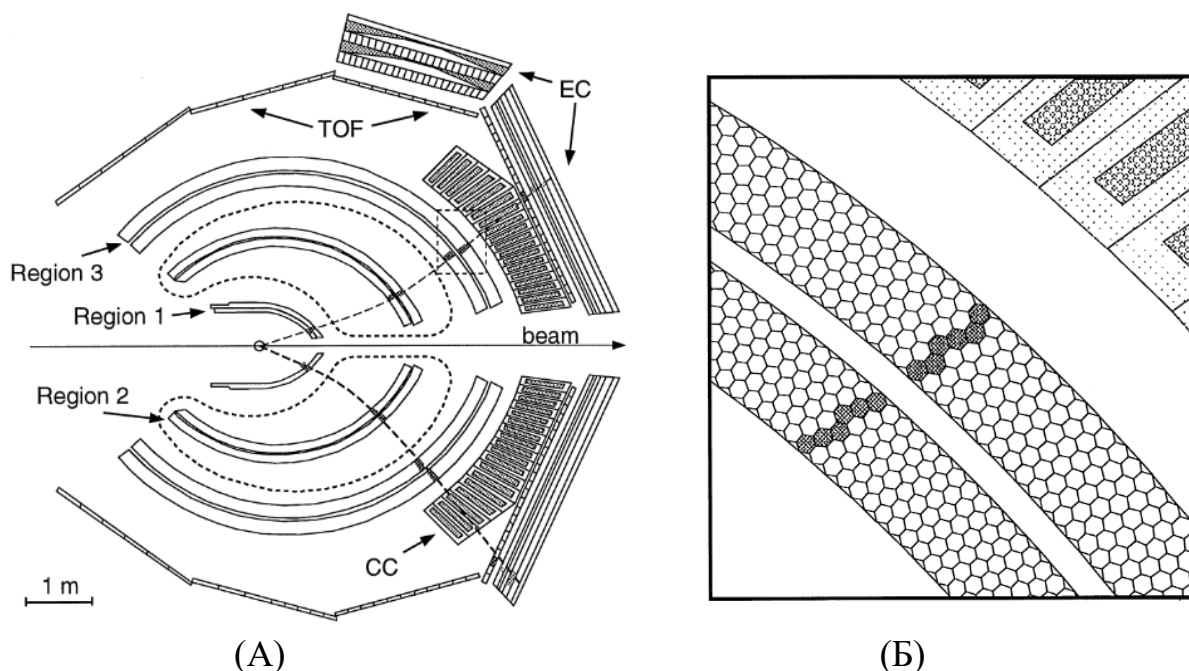


Рисунок 11 – (А) Две заряженные частицы пересекают дрейфовые камеры (Region 1, 2, 3), черенковские счётчики (СС) и электромагнитные калориметры (ЕС) в двух полусферах детектора. (Б) Выделенный прямоугольник на рисунке (А) в крупном масштабе, представлена ячеистая структура дрейфовых камер: в вершинах шестиугольников находятся полевые провода, в центре – чувствительные провода, периметры ячеек нарисованы для удобства восприятия. Закрашенные шестиугольники показывают движение заряженной частицы. В правом верхнем углу изображена часть черенковского счётчика.

2.2.3. Черенковские счётчики

Излучение Вавилова-Черенкова – это излучение атомов среды при движении в ней заряженной частицы со скоростью $v > c/n$, где c – скорость света, n – показатель преломления среды. Основная задача черенковских счётчиков детектора CLAS – разделить сигналы от электронов и от отрицательно заряженных пи-мезонов. Счётчики наполнены газом C_4F_{10} с показателем преломления 1.00153, минимальная величина энергии частицы для возникновения излучения $E_{min} = \frac{m}{\sqrt{1 - \frac{1}{n^2}}}$, где m – масса частицы, что составляет 2.5 ГэВ для пионов и 9.2 МэВ для электронов. Черенковское излучение собирается системой зеркал и фокусируется в фотоэлектронный умножитель (рисунок 12).

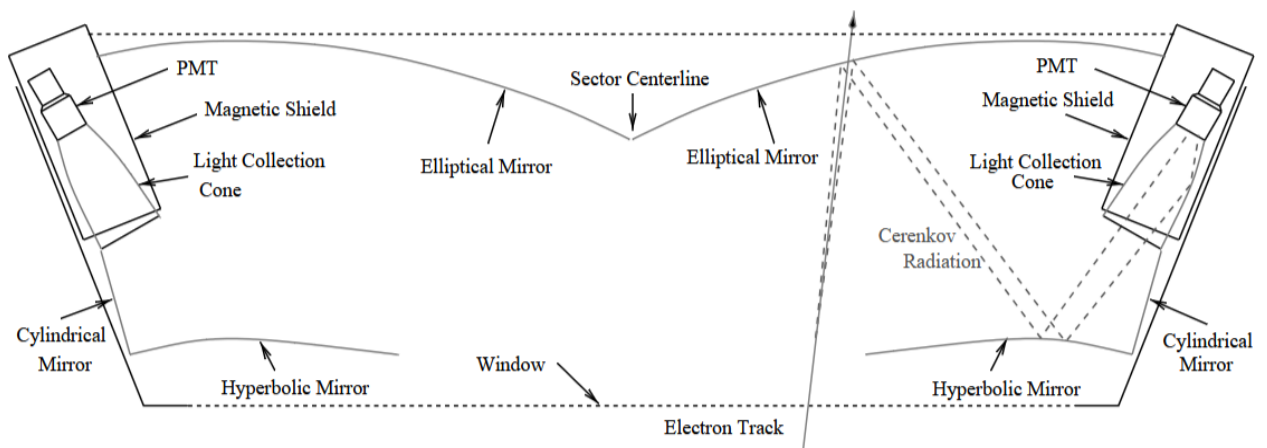


Рисунок 12 – Схематическое устройство модуля черенковского счётчика детектора CLAS. Фотоэлектронные умножители (photomultiplier tubes, PMT) располагаются в области под катушками тороидального магнита.

2.2.4. Электромагнитные калориметры

Система электромагнитных калориметров детектора CLAS представлена шестью передними калориметрами, охватывающими углы θ до 45° , и двумя калориметрами больших углов с перекрытием по θ от 45° до 75° . Каждый калориметр составлен из чередующихся слоёв свинца и сцинтилляторов (рисунок 13). Калориметры предназначены для регистрации электронов с

энергиями выше 0.5 ГэВ, фотонов от распадов π^0 и η с энергиями выше 0.2 ГэВ и обнаружения нейтронов.

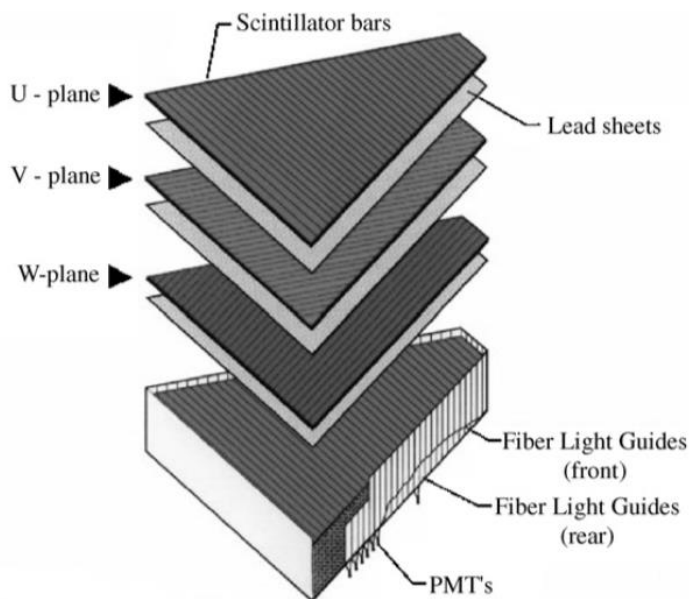


Рисунок 13 – Вид модуля электромагнитного калориметра CLAS.

2.2.5. Система времени пролёта

Сцинтилляторы системы времени пролёта, располагающиеся между черенковскими счётчиками и электромагнитными калориметрами, определяют амплитуду импульса частиц и время их регистрации, успешно отделяют сигналы от пионов и каонов с энергиями до 2 ГэВ друг от друга. Система имеет высокое временное разрешение: $\sigma = 120$ пс для частиц, образующихся под малыми углами (обычно высокоэнергичные) и $\sigma = 250$ пс для частиц под углами $\theta > 90^\circ$.

Глава 3. Сбор и хранение экспериментальных данных, полученных на детекторе CLAS

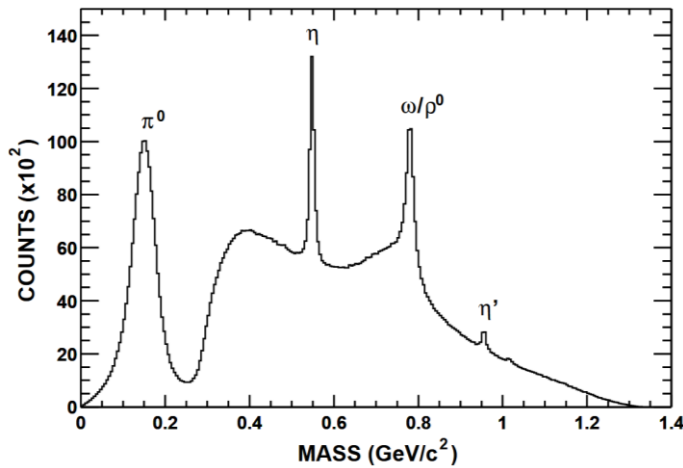
3.1. Идентификация частиц в детекторе CLAS

3.1.1. Идентификация нейтральных частиц

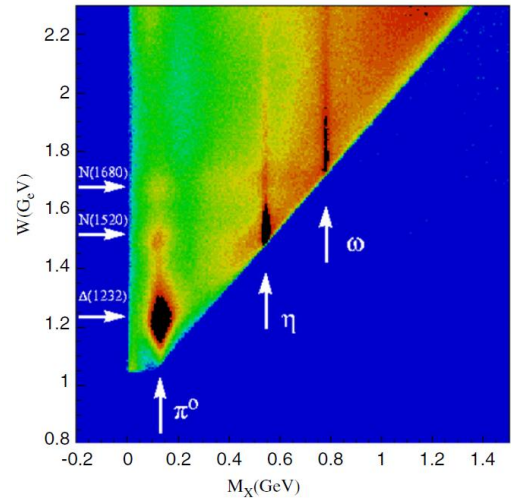
π^0 -мезон может быть восстановлен из его распада на 2 фотона. Однако Монте-Карло моделирование предсказывает 50% эффективность регистрации в электромагнитном калориметре фотонов с энергией ≈ 100 МэВ [19]. Поэтому часто применяется метод недостающей массы: в реакции $e^-p \rightarrow e^-p X$, где X – любая нейтральная частица, недостающая масса равна

$$MM^2(epX) = (P_e + P_p - P_{e'} - P_{p'})^2,$$

$P_e, P_{e'}, P_p, P_{p'}$ – четырёхвекторы начального и конечного электронов, начального и конечного протонов соответственно. На рисунке 14 А изображён спектр недостающей массы реакции $e^-p \rightarrow e^-p X$, на рисунке 14 Б приведена зависимость инвариантной массы от недостающей массы реакции $e^-p \rightarrow e^-p X$ при энергии электронного пучка 4 ГэВ.



(А)



(Б)

Рисунок 14 – (А) Спектр недостающей массы реакции $e^-p \rightarrow e^-p X$. Отчётливо видны пики, обусловленные рождением $\pi^0, \eta, \omega/\rho^0, \eta'$. (Б) Зависимость инвариантной массы от недостающей массы. Заметна корреляция между восстановленными π^0, η, ω и конкретными резонансными состояниями.

3.1.2. Идентификация заряженных частиц

Для регистрации заряженного адрона необходимо получить информацию о длине его пути L и времени прохождения t от мишени до одного из сцинтилляторов системы времени пролёта [19-20]. Тогда скорость частицы равна $\beta = \frac{v}{c} = \frac{L}{t \cdot c}$, а её масса $m = \frac{p}{\beta \gamma}$, где p – импульс, $\gamma = (1 - \beta^2)^{-1/2}$. Детектор CLAS может различить пионы, каоны и протоны в области $p < 2$ ГэВ/ c^2 (рисунок 15).

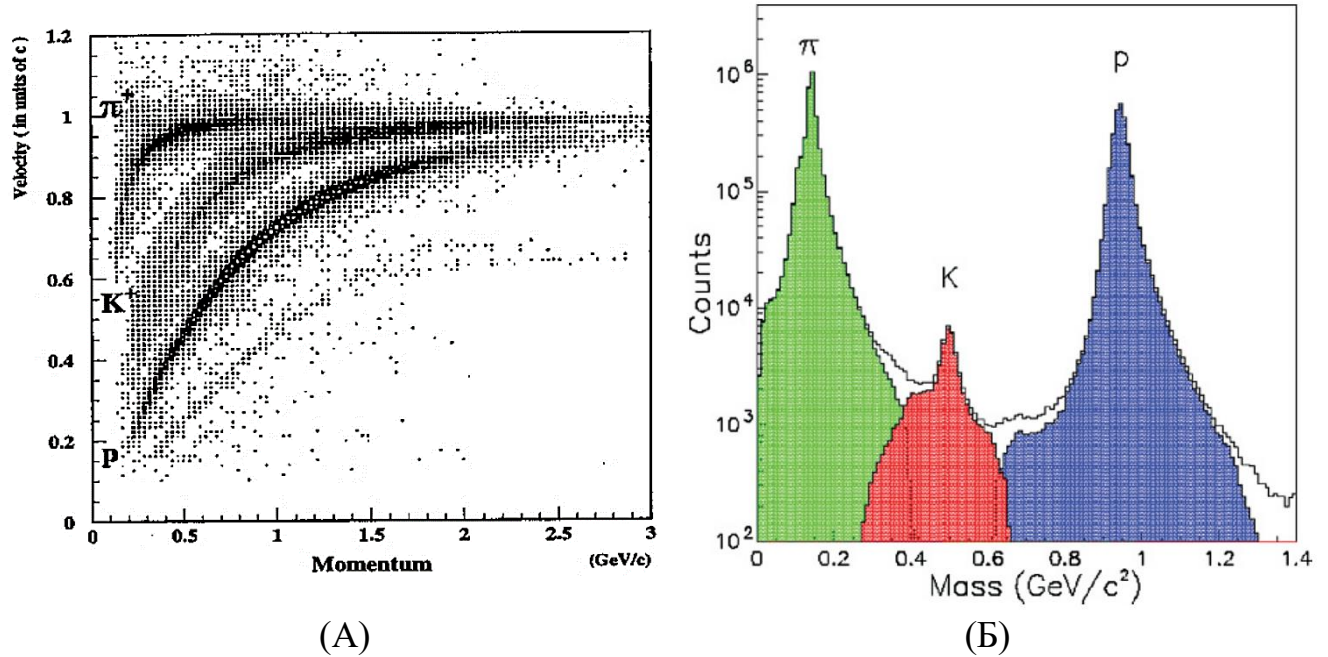


Рисунок 15 – (А) Двумерное распределение скорости и импульса для заряженных адронов. (Б) Распределение масс заряженных адронов.

Идентификация электронов – более трудоёмкая задача: сначала по совпадениям сигналов от черенковского счётчика и электромагнитного калориметра устанавливается наличие трека отрицательно заряженной частицы, а затем по величине поглощённой калориметром энергии частицы происходит отделение электронов от π^- -мезонов.

3.2. Результаты коллаборации CLAS из данных реакций $e^-p \rightarrow e^-p \pi^0$ и $e^-p \rightarrow e^-n \pi^+$

За время работы детектора CLAS была получена большая часть мировых данных эксклюзивных реакций электророждения мезонов на протоне в

кинематической области Q^2 до 5 ГэВ² и W до 3 ГэВ. В таблицах [1](#), [2](#) содержится информация об измеренных наблюдаемых и резонансах, для которых найдены амплитуды электровозбуждения в эксклюзивных реакциях электророжения одиночного пиона [\[21\]](#).

Конечное адронное состояние	Диапазон по W , ГэВ	Диапазон по Q^2 , ГэВ ²	Измеренные наблюдаемые
$p \pi^0$	1.10-1.38	0.16-0.36	$\frac{d\sigma}{d\Omega}$
	1.1125-1.7875	0.45-0.95	$\frac{d\sigma}{d\Omega}$
	1.10-1.68	0.40-1.15	$\frac{d\sigma}{d\Omega}, A_{LT'}, A_t, A_{et}$
	1.08-1.16	2.00-4.50	$\frac{d^2\sigma}{d \cos \theta_\pi d\varphi_\pi}$
	1.10-1.39	3.00-6.00	$\frac{d\sigma}{d\Omega}$
$n \pi^+$	1.10-1.38	0.16-0.36	$\frac{d\sigma}{d\Omega}$
	1.10-1.55	0.30-0.60	$\frac{d\sigma}{d\Omega}$
	1.10-1.70	1.70-4.50	$\frac{d\sigma}{d\Omega}, A_{LT'}$
	1.60-2.00	1.80-4.50	$\frac{d\sigma}{d\Omega}$

Таблица 1 – Измеренные наблюдаемые эксклюзивных реакций $e^-p \rightarrow e^-p \pi^0$ и $e^-p \rightarrow e^-n \pi^+$ на детекторе CLAS в соответствующих областях по W и Q^2 .

$\frac{d\sigma}{d\Omega}$ – дифференциальные сечения в системе центра масс, $A_{LT'}, A_t, A_{et}$ – асимметрии продольно-поляризованного пучка, мишени и пучка-мишени соответственно.

Эксклю- зивные каналы	Возбуждённые состояния протона	Диапазон по Q^2 в реакции $\gamma_p p N^*$, ГэВ ²
$p \pi^0, n \pi^+$	$\Delta(1232) 3/2^+$	0.16-6.00
	$N(1440) 1/2^+, N(1520) 3/2^-, N(1535) 1/2^-$	0.30-4.16
$n \pi^+$	$N(1675) 5/2^-, N(1680) 5/2^+, N(1710) 1/2^+$	1.60-4.50

Таблица 2 – Нуклонные резонансы, для которых получены амплитуды электровозбуждения из анализа данных CLAS по эксклюзивным реакциям

$$e^-p \rightarrow e^-p \pi^0 \text{ и } e^-p \rightarrow e^-n \pi^+.$$

3.3. База данных CLAS Physics Database

В сотрудничестве Национальной Лаборатории им. Томаса Джефферсона (JLab) с ОЭПБАЯ НИИ ядерной физики им. Д.В. Скобелыцына МГУ была создана база данных CLAS Physics Database [22-23]. Она содержит в себе значительную часть уникальных экспериментальных результатов (дифференциальных сечений, асимметрий пучка и мишени, структурных функций) по инклюзивным, полуинклюзивным и эксклюзивным реакциям на протонах и ядрах, полученных коллаборацией CLAS в период эксплуатации детектора в 1996-2012 гг. Хранящиеся в базе данных наблюдаемые не зависят от специфических характеристик экспериментальной установки, таких как эффективность регистрации событий или аксептанс.

3.4. База данных реакции $e^-p \rightarrow e^-p \pi^0$

База данных реакции $e^-p \rightarrow e^-p \pi^0$, используемая в работе, содержит информацию о дифференциальных сечениях $\frac{d\sigma_{\gamma p}}{d\Omega_\pi}$, определённых с ошибками, в зависимости от виртуальности фотона Q^2 , инвариантной массы системы конечных адронов W , углов θ_π и φ_π для четырёх значений энергии начального электронного пучка: $E_e \in [1.64, 2.04, 2.44, 5.75]$ ГэВ. На рисунке 16 приведены

некоторые события из этой базы данных. На рисунке [17](#) показана доступная в эксперименте кинематическая область для этих значений энергии.

	Ebeam	W	Q2	cos_theta	phi	dsigma_d0mega	error
0	1.64	1.1000	0.40	-0.9	0.261799	0.912	0.982353
1	1.64	1.1000	0.40	-0.9	0.785398	0.175	0.356108
2	1.64	1.1000	0.40	-0.9	1.308997	1.650	2.769879
3	1.64	1.1000	0.40	-0.9	1.832596	0.560	0.979330
4	1.64	1.1000	0.40	-0.9	2.356194	1.130	0.909689
...	...	...	...	...	...	...	...
81073	2.04	1.7875	0.95	0.9	5.105088	0.000	0.000000
81074	2.04	1.7875	0.95	0.9	5.366887	0.000	0.000000
81075	2.04	1.7875	0.95	0.9	5.628687	0.000	0.000000
81076	2.04	1.7875	0.95	0.9	5.890486	0.000	0.000000
81077	2.04	1.7875	0.95	0.9	6.152286	0.000	0.000000

81078 rows × 7 columns

Рисунок 16 – База данных реакции $e^-p \rightarrow e^-p \pi^0$, используемая в работе.

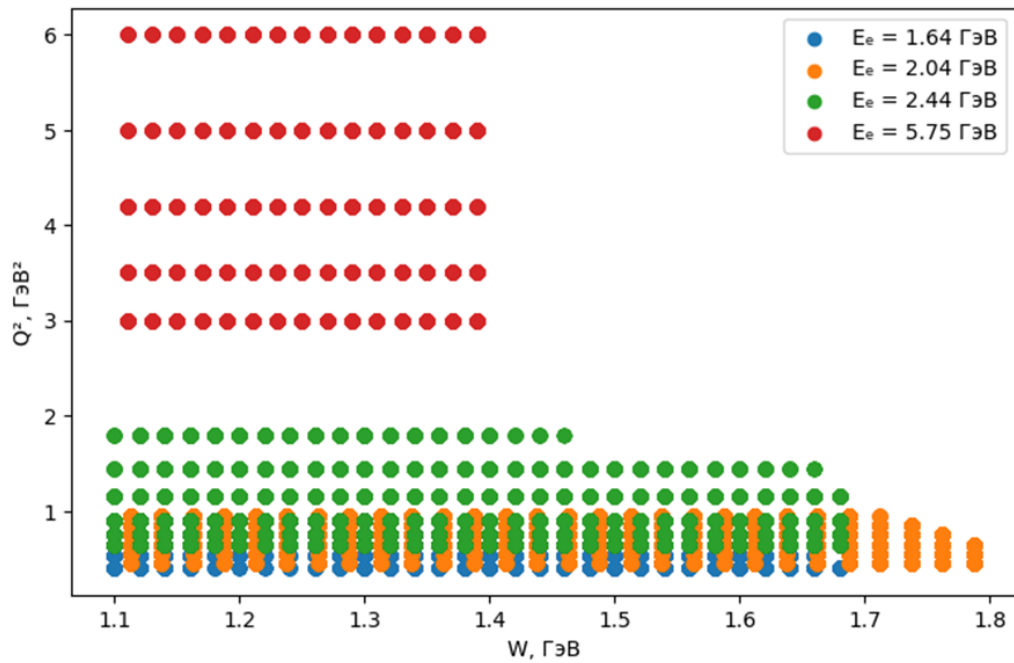


Рисунок 17 – Доступная в эксперименте кинематическая область для значений энергии начального электрона $E_e \in [1.64, 2.04, 2.44, 5.75]$ ГэВ. По оси абсцисс отложена инвариантная масса W , по оси ординат – Q^2 .

Глава 4. Обзор методов машинного обучения, использованных в работе

Машинное обучение (machine learning) – это сфера знаний в области искусственного интеллекта, которая направлена на разработку моделей, способных выявлять зависимости и закономерности в данных без явных инструкций со стороны человека. В качестве данных могут выступать таблицы, пространственно-структурированная информация (видео, изображения), последовательности (текст, звук). Объект (object) в задаче машинного обучения – это сущность, которая характеризуется совокупностью признаков (features). Наиболее распространённые виды признаков: численные, бинарные, категориальные и упорядоченные.

Чтобы модель могла находить различия и корреляции в подаваемых ей данных, она должна обучиться на некотором их наборе, то есть настроить значения своих параметров. Обычно выделяют следующие способы обучения:

1) обучение с учителем (supervised learning): алгоритму заранее известен правильный ответ (целевая переменная, target) для каждого объекта. Этот тип обучения включает задачи классификации – отнесение образца к одному из нескольких множеств (классов) и задачи регрессии – предсказание для объекта некоторого непрерывного численного значения;

2) обучение без учителя (unsupervised learning): в данных целевые переменные неизвестны. Примером обучения без учителя является задача кластеризации – разбиение множества объектов на группы (кластеры) таким образом, чтобы объекты одного кластера были более похожи друг на друга, чем на объекты других кластеров по какому-либо критерию;

3) обучение с подкреплением (reinforcement learning): обучение алгоритма (агента) осуществляется посредством его взаимодействия со средой (источником данных, не обязательно стационарным) и получения вознаграждения/штрафа за свои действия.

Кроме того, существуют комбинации описанных выше способов, например, обучение с частичным привлечением учителя (semi-supervised learning) и др.

Многочисленные методы машинного обучения можно разделить на две группы: классическое машинное обучение и глубокое обучение (deep learning). К первой группе относятся линейная и логистическая регрессии, дерево решений, метод опорных векторов (SVM), методы понижения размерности (например, метод главных компонент, PCA) и др. Глубокое обучение рассматривается как совокупность методов машинного обучения, основанных на искусственных нейронных сетях. Эта группа методов автоматически распознаёт необходимые признаки в данных, в отличие от алгоритмов классического машинного обучения, для которых признаки создают люди вручную. В зависимости от типа обрабатываемых данных и характера поставленной задачи были внедрены различные топологии нейронных сетей, например, полносвязные (FNN), свёрточные (CNN), рекуррентные (RNN) нейронные сети, а также трансформеры, автоэнкодеры и др.

Далее будет приведён краткий обзор алгоритмов, которые применялись в работе для предсказания дифференциальных сечений реакции электророждения одиночного нейтрального пиона на протоне.

4.1. Методы классического машинного обучения

Несмотря на широкое распространение нейронных сетей, методы классического машинного обучения не теряют своей актуальности, поскольку они часто позволяют получить разумное качество предсказания за короткое время обучения. Более того, их часто применяют в случае табличных данных.

Одной из первых и наиболее интуитивно понятных моделей классического машинного обучения является дерево решений (decision tree) [24]. Рисунок 18 демонстрирует принцип его работы. Пусть есть датасет, объектами которого являются точки, и у каждой из них есть 2 вещественных признака. Для всех точек в узле t_1 необходимо создать вопрос: признак x_1 больше 0.7 или меньше? Если больше 0.7, то это красная точка. Если меньше 0.7, то попадаем во второй узел t_2 и спрашиваем: признак x_2 меньше 0.5 или больше? Если больше 0.5, то точка будет синяя, в противном случае — красная. Листья дерева — это

последние узлы, в которых нет разбиения объектов. Дерево решений делит пространство признаков с помощью плоскостей на области, и в каждой из этих областей предсказывается константная величина.

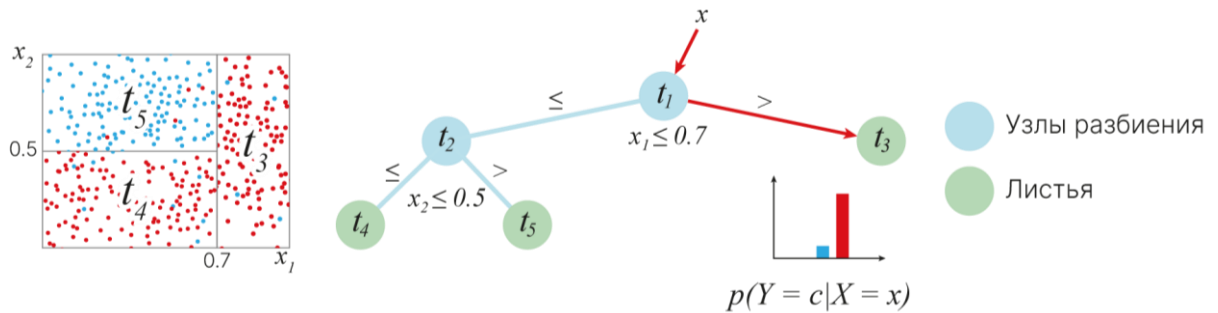


Рисунок 18 – Принцип работы дерева решений.

Возникает вопрос: как построить такое дерево, то есть по каким признакам должны сравниваться объекты из датасета в узлах разбиения? Пусть теперь перед нами ответственная задача – определить, есть ли у человека инфаркт по 3 бинарным признакам. Разбиение всех объектов представлено на рисунке [19](#). Наилучшее разбиение создаёт узлы, в которых сосредоточены объекты преимущественно одного класса. В качестве критерия разбиения можно использовать метрику Джини:

$$Gini = 1 - \sum_i p_i^2,$$

где p_i – вероятность класса i . Чем ниже метрика, тем лучше получившийся узел. Далее можно оценить, насколько улучшился результат в зависимости от выбранного признака:

$$Impurity\ decrease = Gini_0 - \left(\frac{n_1}{n_1 + n_2} Gini_1 + \frac{n_2}{n_1 + n_2} Gini_2 \right),$$

где $Gini_0$ – чистота исходного узла, n_1, n_2 – количество объектов в двух листьях (рисунок [20](#)). К наибольшему impurity decrease приводит признак «боль в груди», а значит он и является оптимальным для первого узла разбиения.



Рисунок 19 – Разбиение объектов по бинарным признакам.

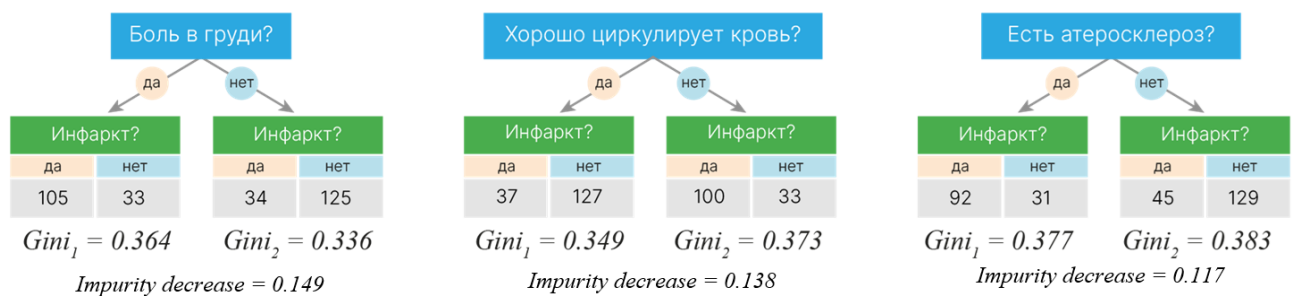


Рисунок 20 – Вычисленные метрики Джини и Impurity decrease.

В случае вещественного признака необходимо сперва отсортировать все значения, после чего выбрать какой-либо порог для разбиения. Impurity decrease считается для каждого такого порога, и на её основе определяется наиболее удачная величина порога (рисунок 21).



Рисунок 21 – Разбиение объекта в случае вещественного признака.

Для задачи регрессии (предсказания не метки класса, а вещественной переменной) можно предсказывать наиболее часто встречающееся значение, медиану или среднее значение (рисунок 22). Часто выбирают среднее.

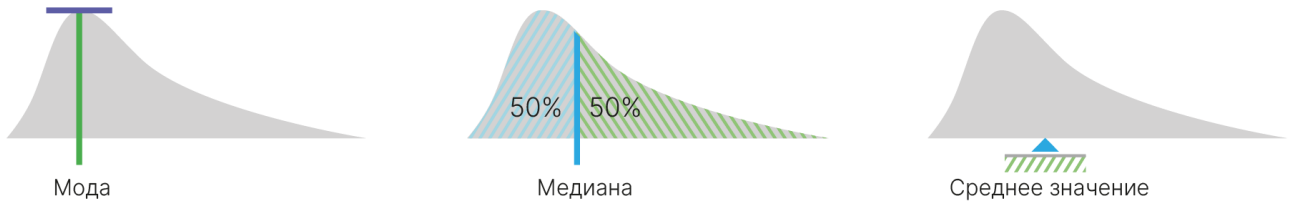


Рисунок 22 – Способы формирования предсказания в листе для регрессии.

Критерий разбиения для регрессионного дерева состоит в том, чтобы как можно сильнее уменьшить отличия вещественных значений в узле, то есть минимизировать дисперсию $D(Y) = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$, где n – количество объектов в узле, $\bar{y} = \frac{\sum_{i=1}^n (y_i)}{n}$ – среднее значение. Для этого вычисляется средняя квадратичная ошибка (MSE):

$$MSE = \frac{1}{N} \sum_{l=1}^L \sum_{i=1}^{n_l} (y_{li} - \bar{y}_l)^2 = \frac{1}{N} \sum_{l=1}^L \sum_{i=1}^{n_l} \frac{n_l - 1}{n_l} D(Y_l),$$

где N – количество объектов в тренировочной выборке, L – общее число листьев, n_l – число объектов в листе l . Уменьшение MSE означает уменьшение дисперсии. Дополнительно необходимо взвешивать дисперсию на размер объектов в узлах: $\frac{D_1 n_1 + D_2 n_2}{n_1 + n_2}$, эта величина должна получиться меньше дисперсии объектов в первичном узле разбиения.

Дерево можно строить до тех пор, пока в каждом листе не окажется по одному объекту, но это приведёт к сильному переобучению (невозможности модели обобщать закономерности на новых данных). Для того чтобы этого не допустить, можно задать минимально возможное количество объектов в листе, ограничить глубину дерева или ввести критерий остановки (например, при достижении определённой процентной доли правильно распознанных объектов).

Деревья могут обрабатывать большой объём данных, работать с категориальными признаками, они устойчивы к выбросам. Однако деревья в чистом виде почти не используются, так как они крайне чувствительны к небольшим изменениям во входных данных, что также может привести к переобучению. Поэтому прибегают к ансамблированию деревьев, например, с помощью алгоритма случайного леса [25]. Основная его суть заключается в том, чтобы построить много базовых некоррелированных деревьев, дающих по отдельности невысокое качество предсказания, и усреднить результаты всех деревьев для получения более точного прогноза. Чтобы построить независимые деревья, можно провести процедуру бэггинга (bagging, bootstrap aggregating): из исходного датасета создать N подвыборок с повторением, на каждой из которых обучить по одному дереву. Предсказание для нового объекта будет строиться путём усреднения предсказаний N обученных моделей (рисунок 23).

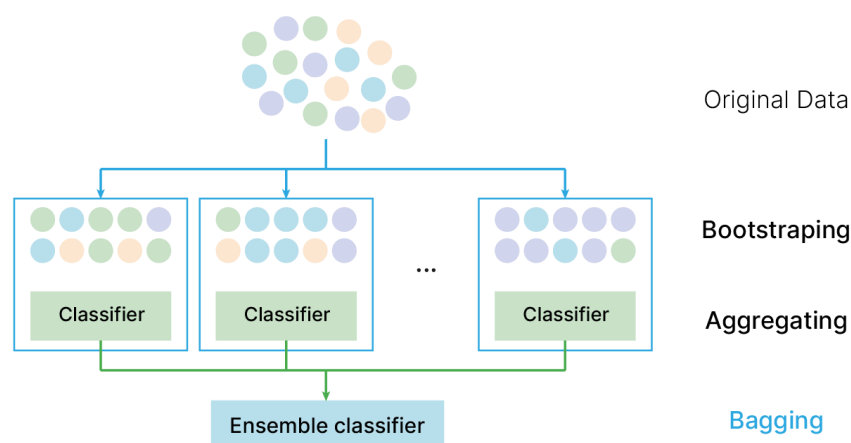


Рисунок 23 – Бэггинг.

Ещё одним способом получения некоррелированных деревьев является метод случайных подпространств (random subspace method, RSM): модели обучаются на полном наборе объектов из тренировочной выборки, но в каждом узле просматриваются не все признаки, а некоторое их количество (рисунок 24).

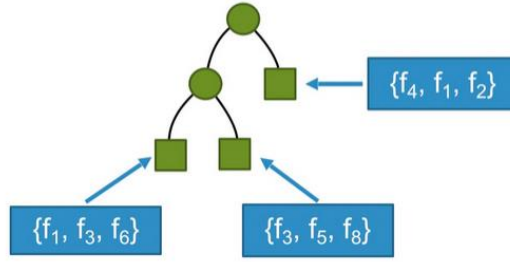


Рисунок 24 – Метод случайных подпространств.

Вторым вариантом ансамблирования деревьев служит градиентный бустинг [25], в котором производится итеративный процесс построения моделей; каждая новая модель, учитывая ошибки предыдущих, должна стремиться улучшить предсказание всего ансамбля (другими словами, минимизировать функцию потерь, то есть функцию, оценивающую, насколько предсказания модели близки к истинным). В бустинге рассматриваются «слабые» деревья: качество их работы должно быть чуть лучше случайного угадывания. Минимизация ошибки осуществляется методом градиентного спуска (рисунок 25). Кратко его можно изложить так: инициализируется начальное предсказание (можно любым числом) $\hat{f}(x) = \hat{f}_0$, вычисляется антиградиент функции потерь $r_{it} = -\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}$, строится новое дерево $h_t(x)$ для уменьшения текущей ошибки предсказания ансамбля $\arg \min_{\theta} \sum_{i=1}^n L(h(x_i, \theta), r_{it})$, вычисляются коэффициент $\rho_t = \arg \min_{\rho} \sum_{i=1}^n L(y_i, \hat{f}(x_i) + \rho \cdot h_t(x_i, \theta))$ и приближение (λ – коэффициент скорости обучения, learning rate) $\hat{f}_t(x) = \lambda \cdot \rho_t \cdot h_t(x)$ и, наконец, обновляется приближение $\hat{f}(x) \leftarrow \hat{f}(x) + \hat{f}_t(x) = \sum_{i=0}^t \hat{f}_i(x)$. Итоговое предсказание – это линейная комбинация предсказаний всех деревьев в ансамбле $\hat{f}(x) = \sum_{i=0}^M \hat{f}_i(x)$.

Одной из модификаций градиентного бустинга является алгоритм XGBoost [26]. Характерной его особенностью является наличие штрафа в функции потерь, учитывающего как количество листьев, так и значения весов в листьях.



Рисунок 25 – Градиентный спуск.

4.2. Полносвязная нейронная сеть

Фактически искусственная нейронная сеть – это сложная математическая функция. Согласно теореме Цыбенко полносвязная нейронная сеть прямого распространения с одним скрытым слоем и сигмоидальными функциями активации может поточечно аппроксимировать любую непрерывную функцию многих переменных с любой наперёд заданной точностью, а значит сеть можно рассматривать в качестве универсального аппроксиматора.

Полносвязная нейронная сеть (или персептрон) [27] – простейший вид нейронных сетей, в которых все нейроны соседних слоёв связаны друг с другом (рисунок 26). Нейрон – это базовый структурный элемент сети, на вход он получает значения от нейронов предыдущего слоя, выполняет взвешенное суммирование, добавляет к получившейся сумме коэффициент смещения (bias), применяет к результату нелинейную функцию, которую называют функцией активации, а затем передаёт информацию нейронам следующего слоя (рисунок 27). Нейроны входного слоя не производят подобной обработки, они служат для передачи информации на первый скрытый слой.

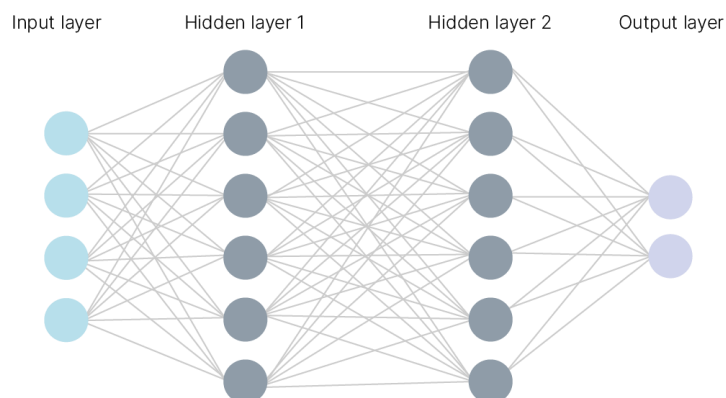


Рисунок 26 – Пример полносвязной сети с двумя скрытыми слоями.

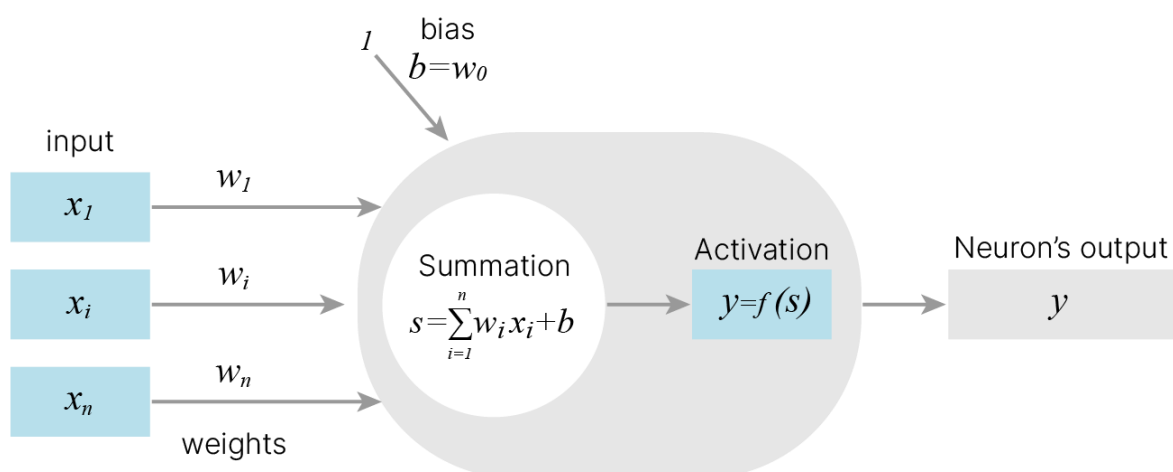


Рисунок 27 – Пример работы нейрона.

Функции активации вводят нелинейности в сети, они должны обладать свойствами дифференцируемости и непрерывности, на рисунке [28](#) представлены некоторые их разновидности. Без функций активации все слои сети эквивалентны одному слою, так как линейная комбинация линейных отображений есть линейное отображение.

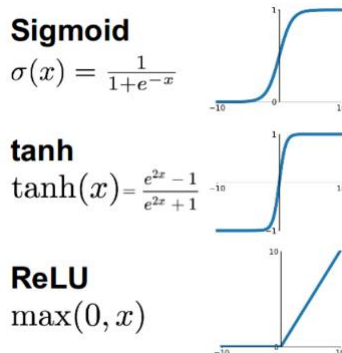


Рисунок 28 – Примеры функций активации.

Веса и смещения нейронов представляют собой обучаемые параметры нейронной сети, при создании сети они инициализируются случайными величинами. Для того чтобы сеть могла аппроксимировать неизвестную функцию, необходимо её обучить, то есть найти подходящие значения весов и смещений для каждого нейрона. Обучение нейронной сети осуществляется методом градиентного спуска и заключается в том, чтобы минимизировать функцию потерь. В ходе прямого распространения (forward pass) модель предсказывает значение искомой величины (метки класса, вероятности принадлежности классу или вещественной переменной) y_{pred} , которое зависит как от входных данных x , так и от параметров модели $\mathbf{W}$: $y_{pred} = model(x, \mathbf{W})$. Функция потерь необходима для сравнения y_{pred} с истинным значением y_{true} , она тоже зависит от $\mathbf{W}$: $L = Loss(y_{true}, y_{pred}) = f(y_{true}, x, \mathbf{W})$. На этапе обратного распространения (backward pass) рассчитывается градиент функции потерь по обучаемым параметрам $\nabla_{\mathbf{W}} L$ методом обратного распространения ошибки (backpropagation) и обновляются веса и смещения (рисунок 29).

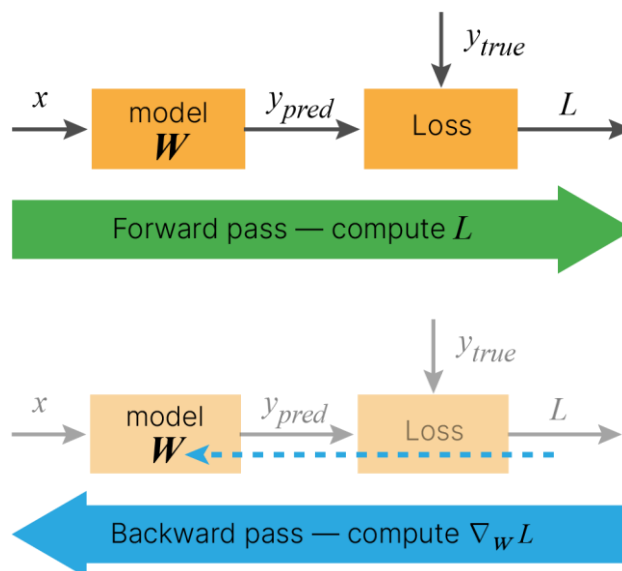


Рисунок 29 – Прямое распространение и вычисление ошибки (сверху), обратное распространение и вычисление $\nabla_{\mathbf{W}} L$ (снизу).

Перед обучением имеющиеся данные обычно делят на тренировочную, валидационную и тестовую выборки. Тренировочный набор необходим для

обновления весов и смещений, валидационный – для контроля качества работы сети и отслеживания переобучения, тестовый – для оценки качества предсказания сети уже после её обучения. При обучении данные проходят по сети не целиком, а небольшими порциями (разделяются на батчи одинакового объёма). Эпохой обучения называют прохождение тренировочной выборки в прямом и обратном направлениях один раз и прохождение валидационной выборки в прямом направлении.

Глава 5. Результаты

Сначала для предсказания дифференциальных сечений реакции $\gamma_\nu p \rightarrow p \pi^0$ были рассмотрены алгоритмы случайного леса и XGBoost. Подбор гиперпараметров осуществлялся с использованием библиотеки Optuna [28], в таблицах 3 и 4 содержится информация о найденных оптимальных значениях. Данные делились для обучения и тестирования в соотношении 8:2. На рисунке 30 представлены метрики качества $MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$ и $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$ (где N – количество реакций, y_i – целевое значение дифференциального сечения для i -й реакции, $\hat{y}_i$ – предсказанное алгоритмом значение) данных моделей.

Гиперпараметр	Поисковый диапазон значений	Оптимальное значение
Количество деревьев (n_estimators)	[20; 200]	160
Максимальная глубина дерева (max_depth)	[2; 30]	26
Минимальное количество объектов, необходимое для разделения узла (min_samples_split)	[2; 10]	5
Минимальное количество объектов в листе (min_samples_leaf)	[1; 10]	2
Доля признаков, отбираемая для построения дерева (max_features)	[0.6, 0.8, 1.0]	0.8
Доля объектов из тренировочной выборки, отбираемая для построения дерева (max_samples)	[0.5; 1.0]	0.9

Таблица 3 – Оптимальные значения гиперпараметров для случайного леса.

Гиперпараметр	Поисковый диапазон значений	Оптимальное значение
Количество деревьев (n_estimators)	[200; 4000]	3010
Максимальная глубина дерева (max_depth)	[2; 20]	17
Максимальное количество листьев (max_leaves)	[20; 120]	99
Минимальное изменение значения функции потерь для разделения узла (gamma)	[0; 10]	0.007
Скорость обучения (eta)	[1e-5; 0.3]	0.03
Доля объектов из тренировочной выборки, отбираемая для построения дерева (subsample)	[0.5; 1.0]	0.9
Доля признаков, отбираемая для построения дерева (colsample_bytree)	[0.6, 0.8, 1.0]	1.0
Параметр L2-регуляризации (reg_lambda)	[0; 10]	0.686
Параметр L1-регуляризации (reg_alpha)	[0; 10]	8.563

Таблица 4 – Оптимальные значения гиперпараметров для XGBoost.

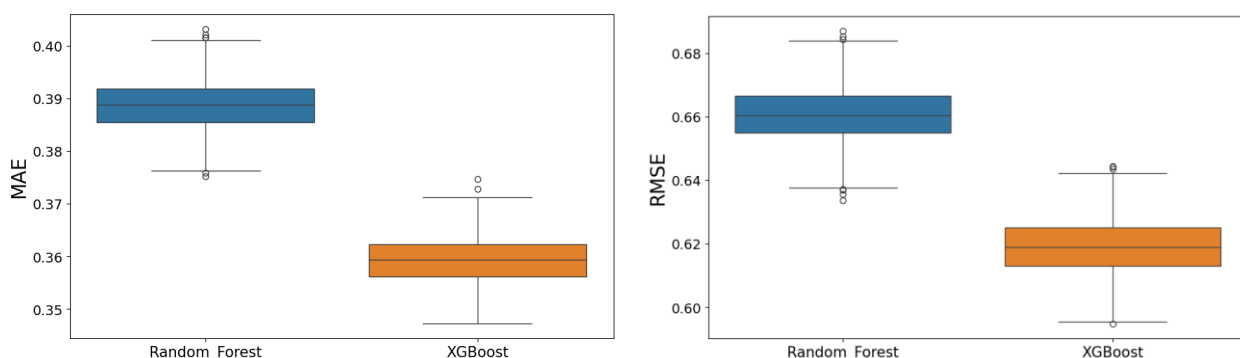


Рисунок 30 – Ошибки MAE (слева) и $RMSE$ (справа), полученные на бутстрэп-наборах из тестовой выборки, для алгоритмов случайного леса (Random_Forest) и XGBoost.

Чтобы проверить, насколько успешно обученные алгоритмы могут предсказывать дифференциальные сечения $\frac{d\sigma_{\gamma\nu}}{d\Omega_\pi}$ на новых данных, создавалась

сетка по E_e , Q^2 , W , $\cos \theta_\pi$, φ_π с небольшим шагом от минимальных до максимальных значений из исходной базы данных. На рисунках [31](#), [32](#) показаны зависимости предсказанных сечений $\frac{d\sigma_{\gamma\nu}}{d\Omega_\pi}$ (зелёные кривые) от угла φ_π моделей случайного леса и XGBoost соответственно для 4 кинематических областей. Красные точки – это экспериментальные данные из исходной базы. Фитирование (синие кривые) проводилось с помощью функции $\frac{d\sigma_{\gamma\nu}}{d\Omega_\pi} = A + B\cos 2\varphi_\pi + C\cos \varphi_\pi$. Видно, что алгоритмы классического машинного обучения демонстрируют недостаточную точность и полученные распределения нельзя назвать непрерывными. Поэтому следующим шагом в работе стало применение полносвязной нейронной сети.

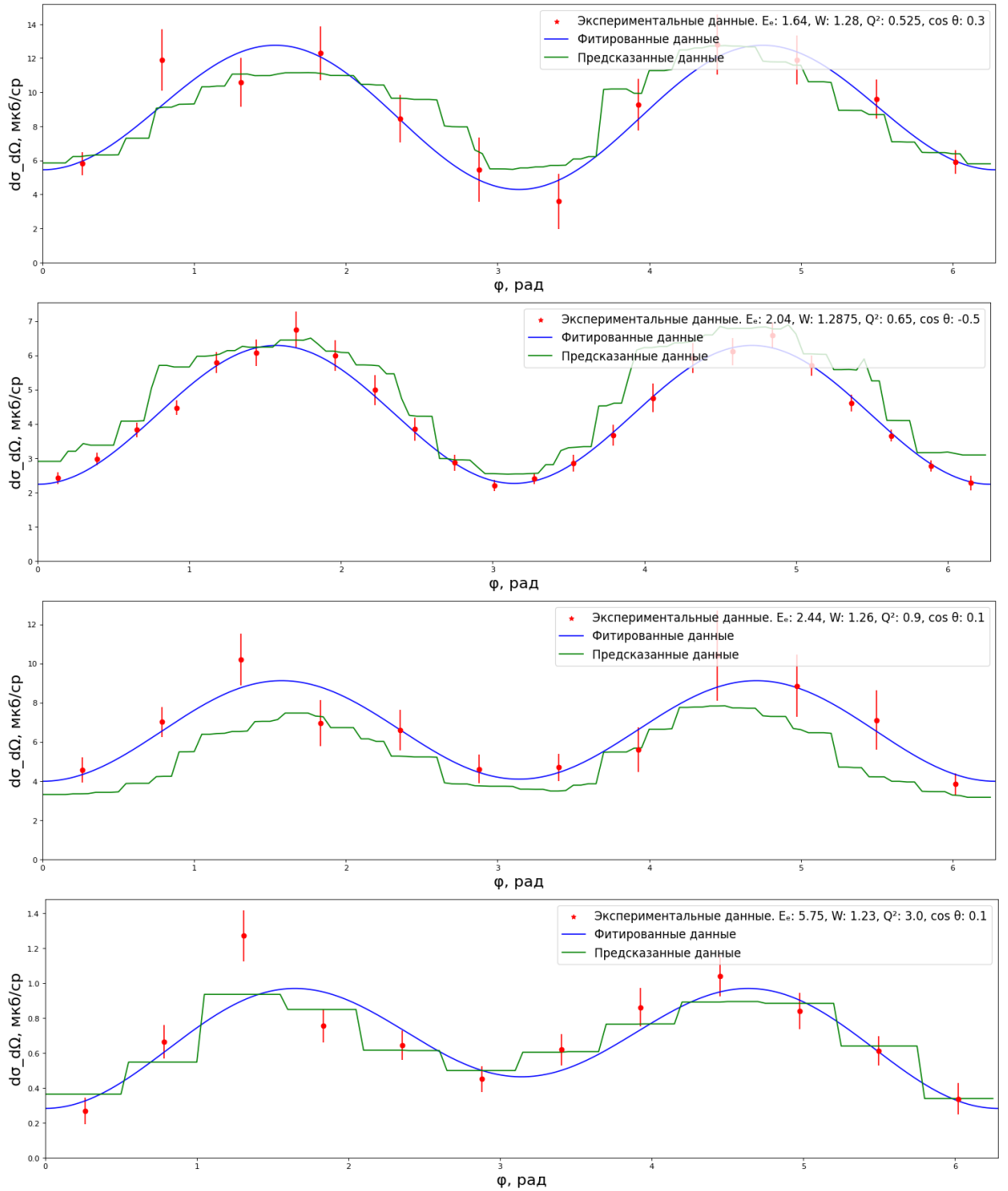


Рисунок 31 – Зависимость экспериментальных данных (красные точки), фитирования (синяя кривая) и предсказаний алгоритма случайного леса (зелёная кривая) дифференциальных сечений $\frac{d\sigma_{\gamma\nu}}{d\Omega_\pi}$ реакции $\gamma_\nu p \rightarrow p \pi^0$ от угла φ_π для областей фазового пространства (сверху вниз): $E_e = 1.64$ ГэВ, $W = 1.28$ ГэВ, $Q^2 = 0.525$ ГэВ², $\cos \theta_\pi = 0.3$; $E_e = 2.04$ ГэВ, $W = 1.2875$ ГэВ, $Q^2 = 0.65$ ГэВ², $\cos \theta_\pi = -0.5$; $E_e = 2.44$ ГэВ, $W = 1.26$ ГэВ, $Q^2 = 0.9$ ГэВ², $\cos \theta_\pi = 0.1$; $E_e = 5.75$ ГэВ, $W = 1.23$ ГэВ, $Q^2 = 3.0$ ГэВ², $\cos \theta_\pi = 0.1$.

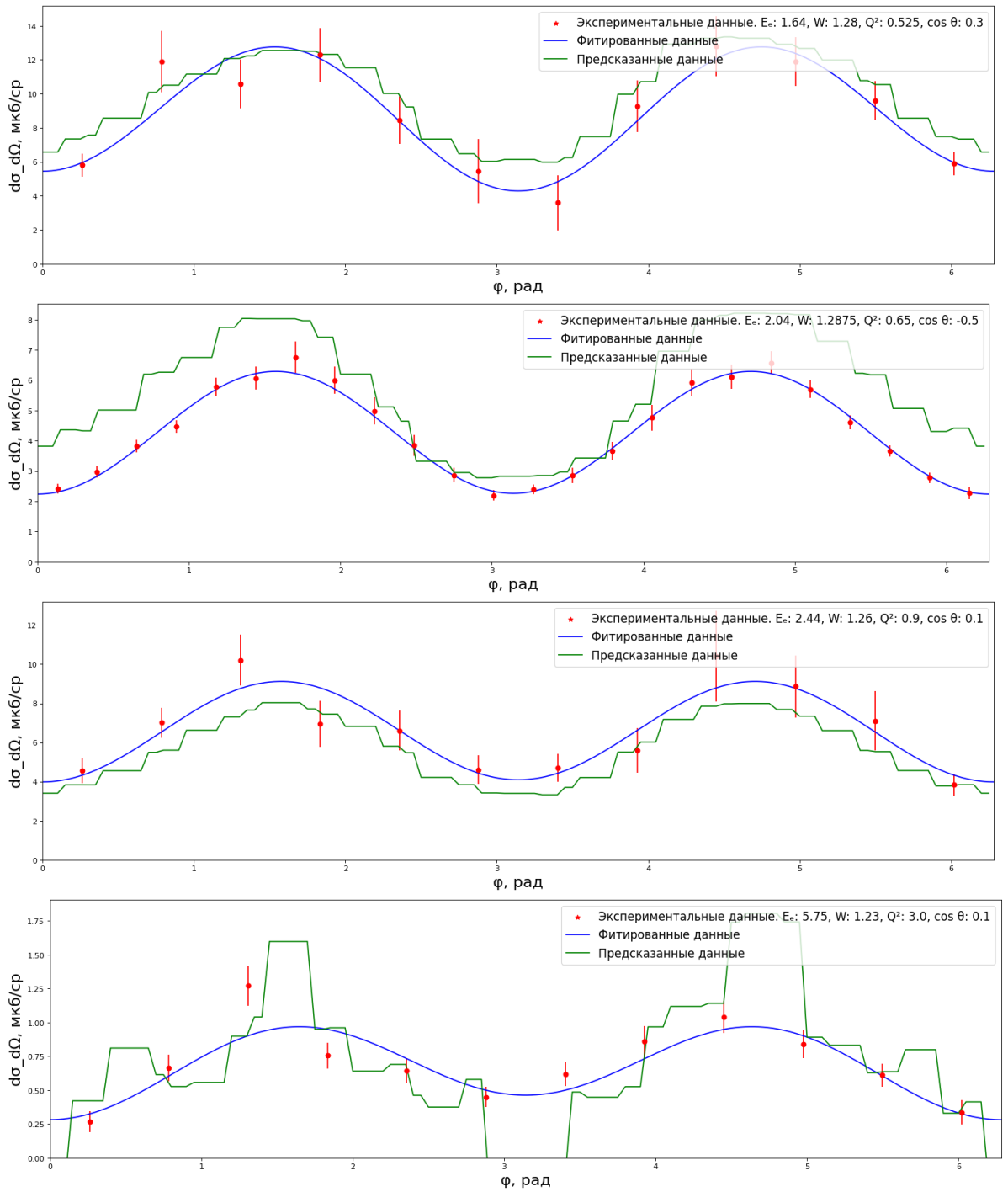


Рисунок 32 – Зависимость экспериментальных данных (красные точки), фитирования (синяя кривая) и предсказаний алгоритма XGBoost (зелёная кривая) дифференциальных сечений $\frac{d\sigma_{\gamma\nu}}{d\Omega_\pi}$ реакции $\gamma\nu p \rightarrow p\pi^0$ от угла φ_π для областей фазового пространства (сверху вниз): $E_e = 1.64$ ГэВ, $W = 1.28$ ГэВ, $Q^2 = 0.525$ ГэВ², $\cos\theta_\pi = 0.3$; $E_e = 2.04$ ГэВ, $W = 1.2875$ ГэВ, $Q^2 = 0.65$ ГэВ², $\cos\theta_\pi = -0.5$; $E_e = 2.44$ ГэВ, $W = 1.26$ ГэВ, $Q^2 = 0.9$ ГэВ², $\cos\theta_\pi = 0.1$; $E_e = 5.75$ ГэВ, $W = 1.23$ ГэВ, $Q^2 = 3.0$ ГэВ², $\cos\theta_\pi = 0.1$.

Архитектура полносвязной нейронной сети для предсказания дифференциальных сечений реакции $\gamma_\nu p \rightarrow n \pi^+$ и реализация обучения в библиотеке PyTorch Lightning [29] были разработаны аспирантом кафедры общей ядерной физики физического факультета МГУ Голда А.В. Количество скрытых слоёв нейронной сети – 19, общее количество обучаемых параметров составляет 5 миллионов. Алгоритм был адаптирован для работы с реакцией $\gamma_\nu p \rightarrow p \pi^0$, а также произведён анализ влияния различных гиперпараметров на качество предсказания сети.

Общее количество событий реакции $e^-p \rightarrow e^-p \pi^0$ составляет около 81 тыс. (рисунок 16), в процессе предобработки были исключены полностью идентичные события, реакции, у которых значение дифференциального сечения меньше величины ошибки, а также реакции с $\frac{d\sigma_{\gamma_\nu}}{d\Omega_\pi} = 0$. Оставшиеся данные (72 тыс.) разделялись на тренировочную и валидационную подвыборки в соотношении 9:1. Для тестового набора создавалась сетка по E_e , Q^2 , W , θ_π , φ_π с небольшим шагом от минимальных до максимальных значений из исходной базы данных.

На вход нейронной сети подавались следующие переменные: E_e , Q^2 , W , $\cos \theta_\pi$, $\cos \varphi_\pi$, $\sin \varphi_\pi$. Ввиду того, что углы φ_π и $\varphi_\pi \pm 2\pi \cdot n$ должны рассматриваться сетью как равные, значения φ_π были преобразованы в $\cos \varphi_\pi$ и $\sin \varphi_\pi$. Для того чтобы избежать влияния возможных выбросов в данных, искажённого восприятия важности одного признака над другим вследствие большего числового значения, необходимо было привести признаки и целевые переменные к единому масштабу. С этой целью входные данные стандартизировались:

$$x' = \frac{x - \mu}{\sigma},$$

где x – начальное значение признака, $\mu = \frac{1}{N} \sum_{i=1}^N x_i$ – среднее, $\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$ – стандартное отклонение признака из данных тренировочной выборки. К целевым значениям применялся MinMaxScaler [30]:

$$y' = \frac{y - y_{min}}{y_{max} - y_{min}},$$

где y – исходное значение дифференциального сечения, y_{min} – наименьшее его значение в тренировочном наборе, а y_{max} – наибольшее. Эмпирически было проверено, что конструирование новых признаков – линейных комбинаций величин E_e , Q^2 , W , $\cos \theta_\pi$, $\cos \varphi_\pi$, $\sin \varphi_\pi$, возведённых в различные показатели степени – не приводит к улучшению предсказания нейронной сети.

В качестве функции активации была выбрана ReLU [31, рисунок 28]. Добавление в архитектуру сети батч-нормализации (Batch Normalization) и регуляризации Dropout не повысило качество работы.

В процессе обучения нейронной сети минимизировалась функция MAE . Наиболее успешно из всех реализованных в PyTorch оптимизаторов проявил себя *Adam* [32], сочетающий в себе идеи инерции (momentum) в функции потерь и адаптивной скорости обучения (adaptive learning rate). Скорость обучения уменьшалась на плато [33]: если ошибка MAE на валидационной выборке не снижалась в течение 5 эпох подряд, скорость обучения сокращалась в 2 раза, её начальное значение равно 0.001.

На рисунке 33 приведены зависимости ошибок MAE , $RMSE$ и функции потерь для валидационного и тренировочного наборов от эпохи обучения.

Результаты работы нейронной сети для тех же четырёх кинематических областей представлены на рисунке 35. Красные точки – экспериментальные данные, зелёная кривая – предсказания алгоритма. Фитирование проводилось с помощью функции $\frac{d\sigma_{\gamma\nu}}{d\Omega_\pi} = A + B\cos 2\varphi_\pi + C\cos \varphi_\pi$, ему соответствует синяя кривая на графиках. Дополнительная проверка качества работы нейронной сети осуществлялась посредством построения распределений неполяризованной структурной функции A и поляризованных структурных функций B , C в зависимости от инвариантной массы W (рисунок 36). На графиках красные точки соответствуют экспериментальным данным, чёрные – предсказанию алгоритма.

Качество работы нейронной сети заметно превосходит результаты алгоритмов классического машинного обучения.

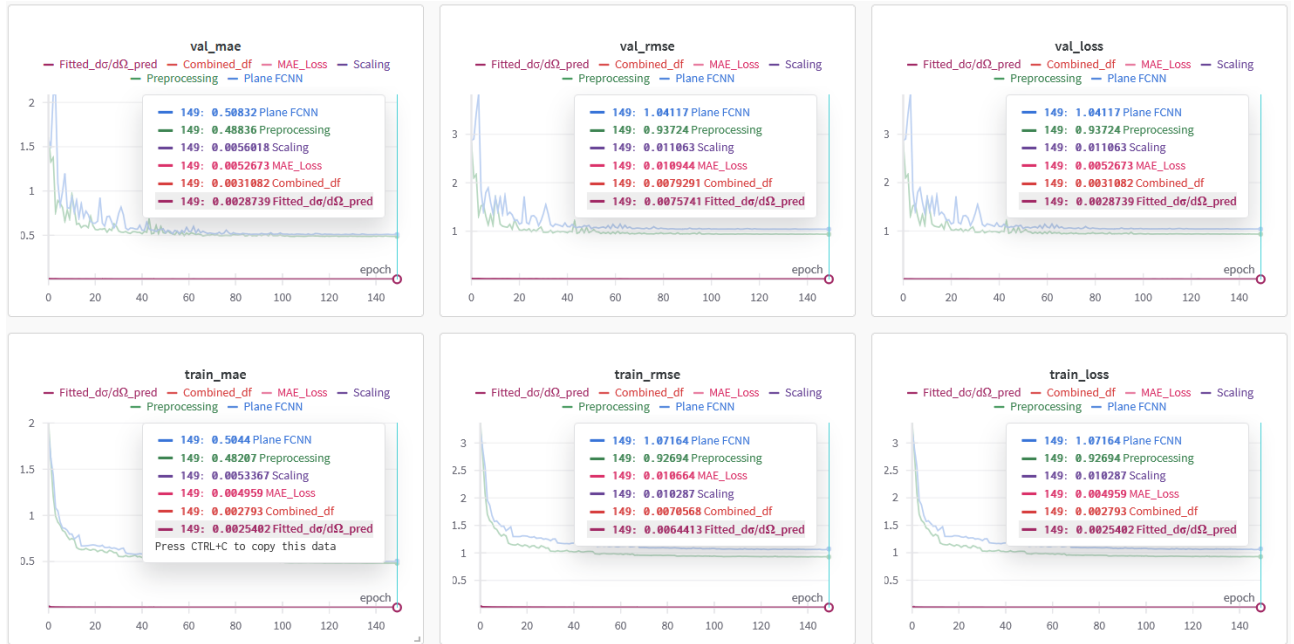


Рисунок 33 – В первом ряду слева направо: зависимость ошибок *MAE*, *RMSE* и функции потерь для валидационной выборки; во втором ряду слева направо: зависимость ошибок *MAE*, *RMSE* и функции потерь для тренировочной выборки от эпохи обучения. Представлены результаты обучения полносвязной сети (Plane FCNN), последовательного введения в процесс обучения предобработки данных (Preprocessing), стандартизации и минимаксного масштабирования (Scaling), замены функции потерь с *RMSE* на *MAE* (MAE_Loss), а также обучения на объединённой базе данных (Combined_df) и предсказания аппроксимированного значения дифференциального сечения (Fitted_dσ/dΩ_pred).

Результаты сохранялись на платформе Weights & Biases [34].

Также нейронная сеть была модифицирована для предсказания дифференциальных сечений обеих реакций электроорождения одиночного пиона: $\gamma_\nu p \rightarrow p \pi^0$ и $\gamma_\nu p \rightarrow n \pi^+$. С этой целью она обучалась на объединённой базе данных двух реакций, дополнительно для каждой реакции вводился бинарный категориальный признак (рисунок 34). Энергия электронного пучка в реакции

$e^-p \rightarrow e^-n\pi^+$ принимает значения $E_e \in [1.515, 5.499, 5.754]$ ГэВ, аналогичные результаты для данной реакции представлены на рисунках [37](#), [38](#).

	Ebeam	W	Q2	cos_theta	phi	dsigma_dOmega	error	cos_phi	sin_phi	reaction_type
0	1.640	1.10	0.4	-0.900	0.261799	0.9120	0.982353	0.965926	0.258819	1
1	1.640	1.10	0.4	-0.900	0.785398	0.1750	0.356108	0.707107	0.707107	1
2	1.640	1.10	0.4	-0.900	1.308997	1.6500	2.769879	0.258819	0.965926	1
3	1.640	1.10	0.4	-0.900	1.832596	0.5600	0.979330	-0.258819	0.965926	1
4	1.640	1.10	0.4	-0.900	2.356194	1.1300	0.909689	-0.707107	0.707107	1
...	...	...	...	...	...	...	...	...	...	...
160831	5.499	2.01	4.0	0.975	3.730641	0.1012	0.043165	-0.831470	-0.555570	0
160832	5.499	2.01	4.0	0.975	3.992441	0.1199	0.076638	-0.659346	-0.751840	0
160833	5.499	2.01	4.0	0.975	4.646939	0.1578	0.095391	-0.065403	-0.997859	0
160834	5.499	2.01	4.0	0.975	4.777839	0.2346	0.158557	0.065403	-0.997859	0
160835	5.499	2.01	4.0	0.975	6.086836	0.1250	0.077753	0.980785	-0.195090	0

160836 rows × 10 columns

Рисунок 34 – Объединённая база данных двух реакций $e^-p \rightarrow e^-p\pi^0$ и $e^-p \rightarrow e^-n\pi^+$.

Дополнительно обобщающая нейронная сеть была обучена предсказывать аппроксимированное значение дифференциального сечения, которое определялось следующим образом: отбирались реакции с одинаковыми величинами E_e , Q^2 , W , $\cos \theta_\pi$, для каждой такой группы рассчитывались коэффициенты А, В, С:

$$A = \frac{1}{2\pi} \int_{\varphi_{min}}^{\varphi_{max}} \frac{d\sigma_{\gamma v}(\varphi_\pi)}{d\Omega_\pi} d\varphi_\pi,$$

$$B = \frac{1}{\pi} \int_{\varphi_{min}}^{\varphi_{max}} \frac{d\sigma_{\gamma v}(\varphi_\pi)}{d\Omega_\pi} \cos 2\varphi_\pi d\varphi_\pi,$$

$$C = \frac{1}{\pi} \int_{\varphi_{min}}^{\varphi_{max}} \frac{d\sigma_{\gamma v}(\varphi_\pi)}{d\Omega_\pi} \cos \varphi_\pi d\varphi_\pi,$$

где φ_{min} , φ_{max} – минимальное и максимальное значение угла φ_π в группе, после чего для каждой реакции вычислялось значение $\frac{d\sigma_{\gamma v}}{d\Omega_\pi} = A + B\cos 2\varphi_\pi + C\cos \varphi_\pi$.

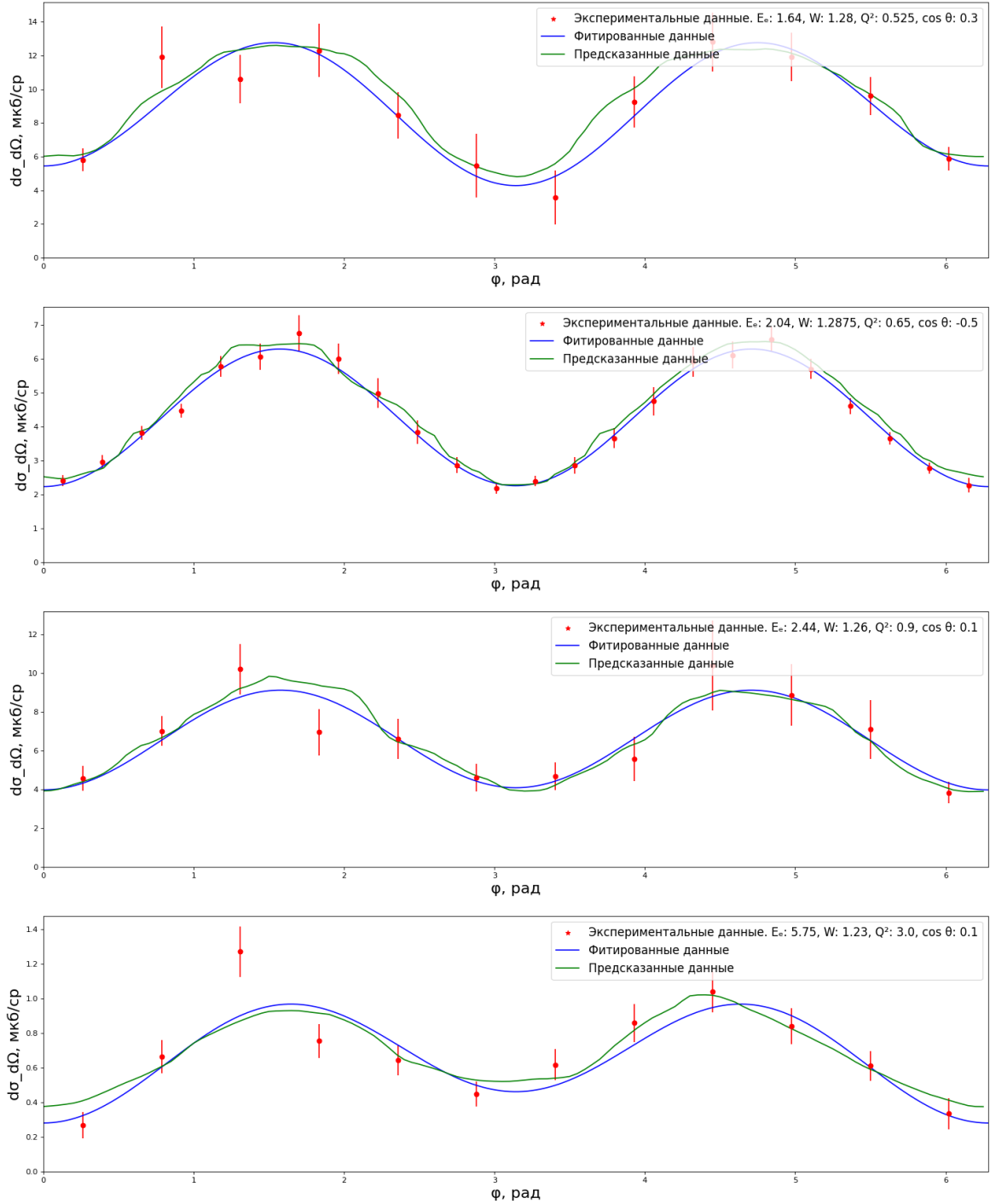


Рисунок 35 – Зависимость экспериментальных данных (красные точки), фитирования (синяя кривая) и предсказаний полносвязной нейронной сети (зелёная кривая) дифференциальных сечений $\frac{d\sigma_{\gamma\nu}}{d\Omega_\pi}$ реакции $\gamma_\nu p \rightarrow p \pi^0$ от угла φ_π для областей фазового пространства (сверху вниз): $E_e = 1.64$ ГэВ, $W = 1.28$ ГэВ, $Q^2 = 0.525$ ГэВ², $\cos \theta_\pi = 0.3$; $E_e = 2.04$ ГэВ, $W = 1.2875$ ГэВ, $Q^2 = 0.65$ ГэВ², $\cos \theta_\pi = -0.5$; $E_e = 2.44$ ГэВ, $W = 1.26$ ГэВ, $Q^2 = 0.9$ ГэВ², $\cos \theta_\pi = 0.1$; $E_e = 5.75$ ГэВ, $W = 1.23$ ГэВ, $Q^2 = 3.0$ ГэВ², $\cos \theta_\pi = 0.1$.

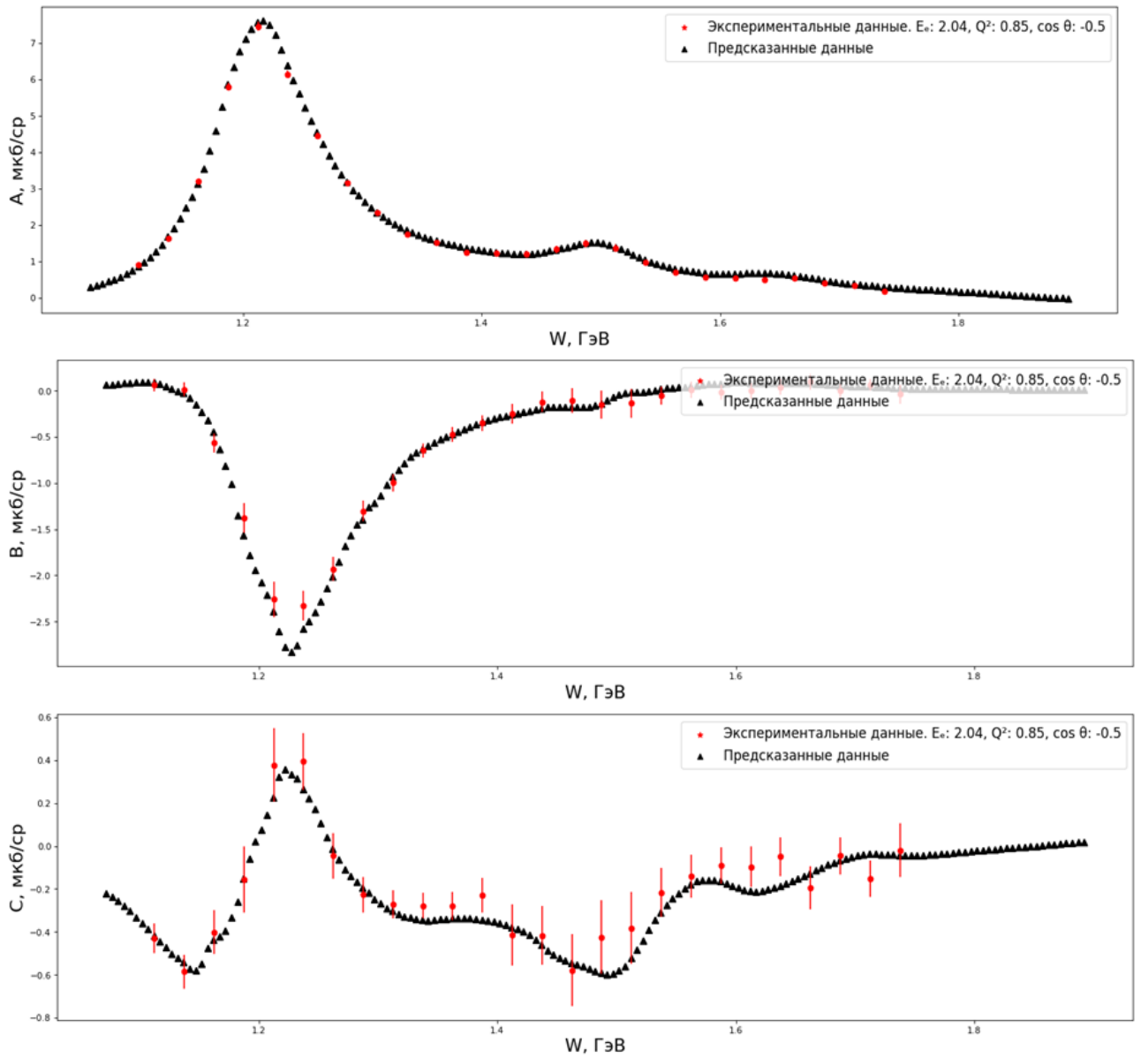


Рисунок 36 – Зависимость неполяризованной структурной функции A и поляризованных структурных функций B и C (сверху вниз) реакции $\gamma_\nu p \rightarrow p \pi^0$ от инвариантной массы W для области фазового пространства $E_e = 2.04$ ГэВ, $Q^2 = 0.85$ ГэВ², $\cos \theta_\pi = -0.5$. Красные точки соответствуют экспериментальным данным, чёрные – предсказанию алгоритма.

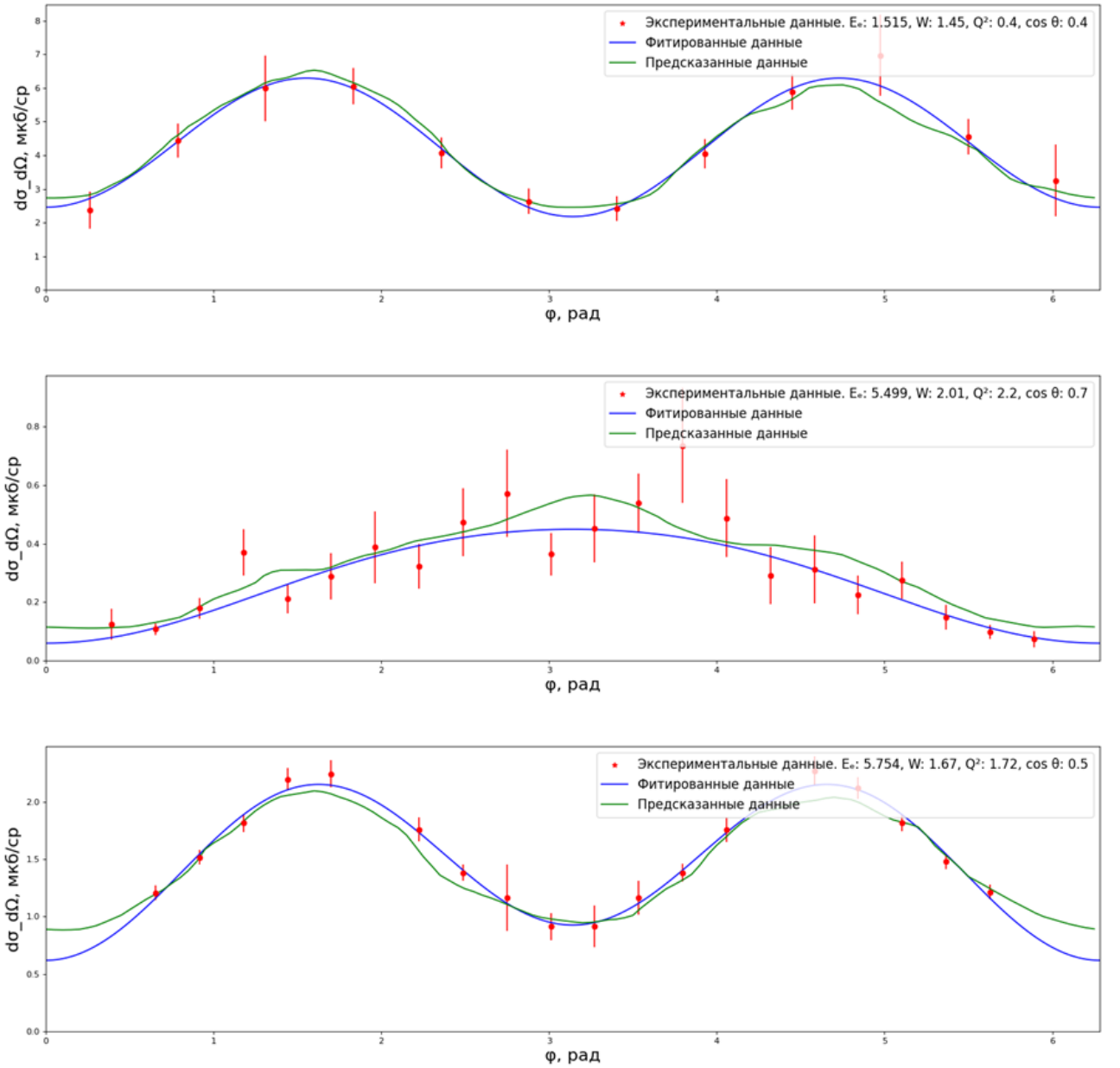


Рисунок 37 – Зависимость экспериментальных данных (красные точки), фитирования (синяя кривая) и предсказаний полносвязной нейронной сети (зелёная кривая) дифференциальных сечений $\frac{d\sigma_{\gamma\nu}}{d\Omega_\pi}$ реакции $\gamma_\nu p \rightarrow n \pi^+$ от угла

φ_π для областей фазового пространства (сверху вниз): $E_e = 1.515$ ГэВ, $W = 1.45$ ГэВ, $Q^2 = 0.4$ ГэВ², $\cos \theta_\pi = 0.4$; $E_e = 5.499$ ГэВ, $W = 2.01$ ГэВ, $Q^2 = 2.2$ ГэВ², $\cos \theta_\pi = 0.7$; $E_e = 5.754$ ГэВ, $W = 1.67$ ГэВ, $Q^2 = 1.72$ ГэВ², $\cos \theta_\pi = 0.5$.

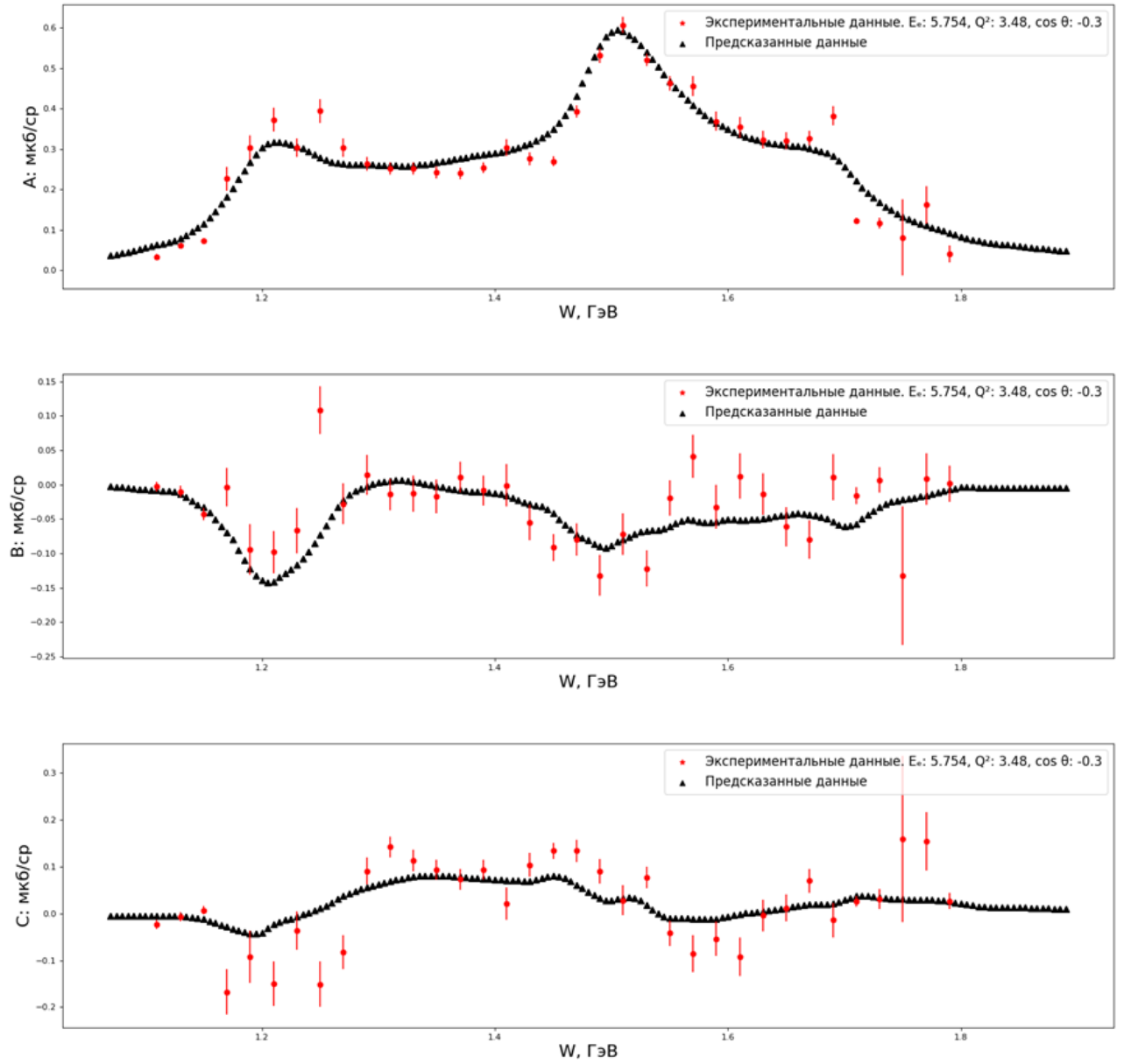


Рисунок 38 – Зависимость неполяризованной структурной функции A и поляризованных структурных функций B и C (сверху вниз) реакции $\gamma_p p \rightarrow n \pi^+$ (сверху вниз) от инвариантной массы W для области фазового пространства $E_e = 5.754$ ГэВ, $Q^2 = 3.48$ ГэВ², $\cos \theta_\pi = -0.3$. Красные точки соответствуют экспериментальным данным, чёрные – предсказанию алгоритма.

ЗАКЛЮЧЕНИЕ

В магистерской работе приведено сравнение результатов по предсказанию дифференциальных сечений реакции электророждения одиночного нейтрального пиона на протоне алгоритмов случайного леса и XGBoost, а также полносвязной нейронной сети. Нейронная сеть продемонстрировала более высокое качество работы по сравнению с алгоритмами классического машинного обучения. К тому же ошибка предсказания нейронной сети оказалась на порядок ниже ошибки в исходной базе данных. Была проведена работа по поиску оптимальных гиперпараметров нейронной сети, была создана обобщающая модель для реакций $\gamma_v p \rightarrow p \pi^0$ и $\gamma_v p \rightarrow n \pi^+$. Так как наблюдается хорошее согласие предсказанных нейронной сетью данных с экспериментальными, то полученная модель может быть использована в качестве Монте-Карло генератора.

По результатам работы была опубликована статья в журнале «Учёные записки физического факультета Московского Университета» [35].

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Gross F., Klempt E., Brodsky S.J. et al.* 50 Years of Quantum Chromodynamics // The European Physical Journal C. 2023. **83**. 1125.
2. *Feynman R.P.* Very High-Energy Collisions of Hadrons // Phys. Rev. Lett. 1969. **23**. 1415.
3. *Carman D.S., Gothe R.W., Mokeev V.I., Roberts C.D.* Nucleon Resonance Electroexcitation Amplitudes and Emergent Hadron Mass // Particles. 2023. **6(1)**. P. 416-439.
4. *Carleo G., Cirac I, Cranmer K. et al.* Machine learning and the physical sciences // Rev. Mod. Phys. 2019. **91**. 045002.
5. *Thomadakis P., Angelopoulos A., Gavalian G., Chrisochoides N.* Using Machine Learning for Particle Track Identification in the CLAS12 Detector // Comput. Phys. Commun. 2023. **287**. 108694.
6. *Radovic A., Williams M., Rousseau D. et al.* Machine learning at the energy and intensity frontiers of particle physics // Nature. 2018. **560**. P. 41–48.
7. *Anderson H. L., Fermi E., Long E. A., Nagle D. E.* Total Cross Sections of Positive Pions in Hydrogen // Phys. Rev. 1952. **85**. 936.
8. Particle Data Group (PDG), <https://pdg.lbl.gov>
9. *Williams M.* Measurement of Differential Cross Sections and Spin Density Matrix Elements along with a Partial Wave Analysis for $\gamma p \rightarrow p\omega$ using CLAS at Jefferson Lab // A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy. 2007.
10. *Roberts C.D.* N^* Structure and Strong QCD // Few Body Syst. 2018. **59**. 72.
11. *Burkert V.D., Roberts C.D.* Colloquium: Roper resonance: Toward a solution to the fifty year puzzle // Rev. Mod. Phys. 2019. **91**. 011003.
12. *Carman D.S., Joo K., Mokeev V.I.* Strong QCD Insights from Excited Nucleon Structure Studies with CLAS and CLAS12 // Few Body Syst. 2020. **61**. 29.
13. *Mokeev V.I., Carman D.S.* Photo- and Electrocouplings of Nucleon Resonances // Few Body Syst. 2022. **63**. 59.

14. *Aznauryan I.G., Burkert V.D.* Electroexcitation of nucleon resonances // Progress in Particle and Nuclear Physics. 2012. **67**. P. 1-54.
15. *Amaldi E., Fubini S., Furlan G.* Pion-Electroproduction. Electroproduction at Low Energy and Hadron Form Factors // Springer Tracts in Modern Physics. 1979. **83**.
16. *Скородумина Ю.А., Буркерт В.Д., Головач Е.Н. и др.* Нуклонные резонансы в эксклюзивных реакциях фото- и электророждения мезонов // ВМУ. Серия 3. ФИЗИКА. АСТРОНОМИЯ. 2015. 6.
17. *Aznauryan I.G., Burkert V.D., the CLAS Collaboration.* Electroexcitation of nucleon resonances from CLAS data on single pion electroproduction // Phys. Rev. 2009. **C 80**. 055203.
18. *Leemann C.W., Douglas D.R., Krafft G.A.* The Continuous Electron Beam Accelerator Facility: CEBAF at the Jefferson Laboratory // Annu. Rev. Nucl. Part. Sci. 2001. **51**. P. 413–450.
19. *Mecking B.A., the CLAS Collaboration.* The CEBAF large acceptance spectrometer (CLAS) // Nucl. Inst. and Meth. in Phys. Res. 2003. **A 503**. P. 513–553.
20. *Dugger M., Ritchie B.G., the CLAS Collaboration.* π^0 photoproduction on the proton for photon energies from 0.675 to 2.875 GeV // Phys. Rev. 2007. **C 76**. 025211.
21. *Moiseev V.I.* Nucleon Resonance Structure from Exclusive Meson Electroproduction with CLAS // Few Body Syst. 2018. **59**. 46.
22. *Чесноков В.В., Голубенко А.А., Ишханов Б.С., Моисеев В.И.* База данных коллаборации CLAS для исследования структуры адронов в электромагнитных процессах // Физика элементарных частиц и ат. ядра. 2022. **53(2)**. С. 181-191.
23. CLAS Physics Database, <https://clas.sinp.msu.ru/cgi-bin/jlab/db.cgi>
24. *Kushwah J.S., Kumar A., Patel S. et al.* Comparative study of regressor and classifier with decision tree using modern tools. // Mat. Today: Proc. 2022. **56(6)**. P. 3571-3576.
25. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. // Springer Ser. in Stat. 2009.

26. *Chen T., Guestrin C.* XGBoost: A Scalable Tree Boosting System. // *Advances in Pure Mathematics.* 2016. **6.** 9.
27. *Hornik K., Stinchcombe M., White H.* Multilayer feedforward networks are universal approximators // *Neural Networks.* 1989. **2.** P. 359-366.
28. *Akiba T., Sano S., Yanase T et al.* Optuna: A Next-generation Hyperparameter Optimization Framework // 2019. [arXiv: 1907.10902].
29. PyTorch Lightning, <https://lightning.ai>
30. MinMaxScaler, <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>
31. ReLU, <https://pytorch.org/docs/stable/generated/torch.nn.ReLU.html>
32. *Kingma D.P., Ba J.* Adam: A Method for Stochastic Optimization // 2017. [arXiv: 1412.6980].
33. ReduceLROnPlateau, https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html
34. Weights & Biases, <https://wandb.ai/site/>
35. *Чистякова В.В., Голда А.В., Русова А.А., Исупов Е.Л.* Машинное обучение в задаче предсказания дифференциальных сечений и структурных функций электророждения пионов в резонансной области. // *Учен. зап. физ. факта Моск. ун-та.* 2025. **2.** 2520213.