

МЕТОДОЛОГИЯ ПРИМЕНЕНИЯ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ В ПОИСКЕ “НОВОЙ ФИЗИКИ” НА КОЛЛАЙДЕРАХ И СТАТИСТИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ ОЖИДАЕМЫХ РЕЗУЛЬТАТОВ

© 2022 г. Э. Э. Абасов^{1)*}, М. И. Белоброва¹⁾, П. В. Волков¹⁾, Г. А. Воротников¹⁾,
Л. В. Дудко¹⁾, А. Д. Заборенко¹⁾, М. А. Перфилов¹⁾, Е. С. Сивакова¹⁾

Поступила в редакцию 01.07.2022 г.; после доработки 01.07.2022 г.; принята к публикации 02.07.2022 г.

В настоящей статье описываются методы применения нейронных сетей к поиску “новой физики” в коллайдерных экспериментах. Рассмотрены механизмы отбора наблюдаемых, оптимизации гиперпараметров, построения каскадов глубоких нейронных сетей. Приведены методы статистического анализа дискриминантов нейронных сетей в пакетах *theta* и *CombinedLimit* с учетом различных неопределенностей.

DOI: 10.31857/S0044002722060022

1. ФОРМУЛИРОВКА ЗАДАЧИ

Одними из важнейших задач современной физики элементарных частиц являются проверка и анализ предсказаний Стандартной модели (СМ). Создаваемые и существующие коллайдеры ТэВных энергий призваны дать ответ о границах применимости СМ и о том, какая “новая физика” может лежать за ее пределами. Современные статистические методы позволяют существенно повысить точность анализа данных, получаемых на коллайдерах. В данной статье приведен краткий обзор методологии применения глубоких нейронных сетей при поиске “новой физики” в коллайдерных экспериментах. Отклонения от предсказаний СМ могут проявляться в виде рождения новых частиц, резонансов, в рождении или распаде частиц СМ, или в виде модификации характеристик взаимодействий частиц СМ. В первом случае применение многомерных методов анализа обычно сводится к подавлению процессов СМ и выделению новых процессов с рождением новых резонансов. Статистический анализ распределения характерной наблюдаемой, например, инвариантной массы продуктов распада нового резонанса дает возможность количественно определить возможный вклад “новой физики”. Второй случай предполагает аккуратный анализ характеристик определенных

процессов и на основании отклонений в сечении, состояний поляризации или других наблюдаемых свойств сделать заключение о возможном отклонении от предсказаний СМ. Существуют три подхода при поиске “новой физики”. Первый подход предполагает точное моделирование нужного процесса в рамках конкретной теории и прямой поиск такого процесса. Второй подход основывается на модельно-независимом описании целого класса теоретических моделей, параметризуемых в рамках эффективной теории поля (EFT), или введении конкретных дополнительных членов в лагранжиан взаимодействия. Третий подход формулируется в терминах поиска аномалий в данных и является наименее точным и эффективным, но активно обсуждается в настоящее время в контексте развития методов глубокого обучения и повышения их эффективности. В данной статье мы сконцентрируемся на первых двух подходах и в качестве примера рассмотрим задачу поиска нейтральных токов, меняющих аромат кварков (Flavour Changing Neutral Currents — FCNC), во взаимодействии топ-кварка с *u*- или *s*-кварком и глюоном (*tqg* FCNC). Модельно-независимое теоретическое описание таких процессов может быть найдено в работе [1]. В рамках СМ такие процессы сильно подавлены и увеличение сечения этих процессов будет прямо свидетельствовать о проявлении “новой физики”, параметризуемой в виде одного аномального параметра для каждого из взаимодействий *tug* и *tcg*. Моделирование таких процессов проведено в пакете *CompHEP* [2, 3]. Процессы СМ, приводящие к аналогичным сиг-

¹⁾Московский государственный университет имени М.В. Ломоносова, Научно-исследовательский институт ядерной физики имени Д.В. Скобельцына, Москва, Россия.

*E-mail: abasov.ee19@physics.msu.ru

натурам в детекторе, смоделированы стандартным образом в пакетах MadGraph [4], POWHEG [5] и Pythia [6]. Моделирование отклика детектора проведено в пакете DELPHES [7] с параметрами детектора CMS коллайдера LHC.

В качестве результата необходимо получить ожидаемые ограничения на параметры аномальных взаимодействий в интерпретации верхних ограничений на брэнчинги (вероятности) редких распадов топ-кварка $t \rightarrow cg$ и $t \rightarrow ug$. Вероятности таких редких распадов настолько малы, что прямой поиск таких распадов не имеет смысла, но в процессах рождения одиночного топ-кварка tqg FCNC взаимодействия могут быть обнаружены. Процессы с tqg FCNC очень близки по характеристикам к t -канальному процессу рождения одиночного топ-кварка в рамках СМ. Поэтому одиночное рождение топ-кварка в рамках СМ можно использовать в качестве контрольного канала для проверки методов анализа. Исследование данной задачи в эксперименте CMS приведено в работе [8].

2. ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ

В коллайдерной физике с помощью нейронных сетей решается широкий круг задач, например, классификация адронных струй (jets), реконструкция треков частиц, поиск аномальных событий, ускорение моделирования событий [9]. В данной статье будет описана методология применения нейронных сетей в анализе высокого уровня для поиска новой физики на примере FCNC-взаимодействий.

Как было отмечено выше, в современной физике высоких энергий существует несколько подходов к поиску событий, не предсказываемых Стандартной моделью: модельно-зависимый, модельно-независимый и поиск неизвестных аномалий. В модельно-независимом анализе и поиске аномалий ведется поиск общих отклонений от предсказаний Стандартной модели в большом количестве физических каналов. Преимущество таких методов заключается в покрытии больших областей фазового пространства и пространства параметров, а не в чувствительности [10].

Одним из методов поиска аномальных событий новой физики является использование модели автокодировщика [11]. Автокодировщик — это модель машинного обучения, которая обучается реконструировать входные данные с минимальной ошибкой. Для обучения такой модели не нужно знать класс события, поэтому обучение автокодировщика — пример обучения без учителя, т.е.

без тренировки сети на заранее смоделированных сигнальных событиях. Для того чтобы модель не выучила функцию $f(x) = x$, в нейросети-автокодировщике часто используют принцип “бутылочного горлышка” — понижения числа нейронов в скрытых слоях. Таким образом, автокодировщик понижает размерность данных, “фокусируясь” на наиболее информативных для реконструкции особенностях данных. Поскольку такая модель “выучивает” распределение нормальных данных, аномальные данные, не попадающие в это распределение, будут иметь более высокую ошибку реконструкции — метрику разницы между входным вектором данных и реконструированным. Чем больше такая метрика для определенного аномального события, тем сильнее оно отличается от нормальных, предсказанных в СМ.

В данной статье рассматривается пример задачи, относящейся к классу модельно-независимых, покрывающей некоторый класс теоретических моделей, но требующих моделирования сигнальных событий с некоторой зависимостью от дополнительных аномальных параметров. Определенная предполагаемая модель новой физики используется для генерации событий методом Монте-Карло и в программах реконструкции, после этого на этих событиях тренируется классификатор, чувствительный именно к этому процессу. После этого дискриминатор классификатора используется для проверки гипотез и статистического анализа (пример дискриминатора — рис. 1). Все модели в данном анализе были созданы на базе фреймворка Tensorflow [12].

Важной частью анализа данных является подготовка входных данных для предиктивной модели. Группой ранее был представлен универсальный метод отбора переменных для анализа в физике высоких энергий, основываясь на фейнмановских диаграммах исследуемых процессов [13]. После этого шага рекомендуется удалить переменные, имеющие сильную линейную корреляцию. Для некоторых моделей это уменьшит размер и увеличит скорость тренировки, а для других — напрямую повлияет на качество классификации. Для оптимальной тренировки нейронных сетей необходимо нормализовать вектор переменных — трансформировать его таким образом, чтобы среднее распределения было равно нулю, а стандартное отклонение — единице. Другие многомерные методы анализа, например, деревья решений, не требуют такой трансформации.

Смоделированные методом Монте-Карло события жестких процессов в коллайдерных экспериментах имеют дополнительные весовые коэффициенты, которые необходимо учитывать при тренировке нейронной сети. Эти веса используются в функции ошибки для увеличения влияния

отдельных событий с большим весом на суммарную ошибку классификации набора смоделированных событий (датасета). При минимизации этой функции в процессе обучения сети модель будет в первую очередь “стремиться” правильно классифицировать именно эти события. При объединении нескольких физических процессов в один набор фоновых событий необходимо учитывать их сечение и нормировать общий вес событий таким образом, чтобы интегральный вес событий определенного физического процесса был пропорционален его сечению. Использование весов для событий позволяет в некоторой степени решить проблему несбалансированной классификации — сильного неравенства числа событий в фоновом и сигнальном датасетах. Например, если сигнальные события составляют только 10% всего датасета, то нейросетевая модель, для всех событий предсказывающая фон, будет иметь 90%-ную точность. Уравнивание интеграла весов сигнальных и фоновых событий позволяет вернуть проблему несбалансированной классификации к сбалансированной. Тем не менее, малое количество событий может существенно повлиять на работу классификатора, внося статистический шум.

Важным этапом оценки работы нейросетевой модели является измерение степени влияния отдельных переменных на итоговый результат классификации. Некоторые модели машинного обучения, в частности, деревья решений, имеют встроенные методы оценки важности переменных. Для оценки важности переменных независимо от типа используемой модели машинного обучения применяются общие методы, например, *permutation feature importance* (метод перестановки). В рамках этого метода каждая из входных переменных модели зашумляется, производится предсказание с зашумленной переменной, и результат этого предсказания сравнивается с эталонным предсказанием на “чистых” переменных (пример результата такого анализа — рис. 2). Чем сильнее отличается полученное предсказание от эталонного, тем важнее для классификации эта переменная. Метод оценки важности переменных позволяет оценить, какие переменные имеют больший вес для модели, и внести некоторую ясность в причины принятия решений конкретной модели машинного обучения. Также этот инструмент помогает понять, в каких областях данных следует в первую очередь искать ошибки при отладке нейросетевой модели. Этот анализ можно провести с помощью инструментов библиотеки *sklearn* [14].

Архитектура нейронной сети может напрямую влиять на качество ее предсказаний. Архитектуру простого многослойного перцептрона можно менять, варьируя число скрытых слоев и нейронов в

них, добавляя дропаут и меняя функции активации. Также на качество финальной модели может влиять оптимизатор модели с его собственными параметрами: *learning rate*, *weight decay*, *gradient clipping*, *momentum* и другие. Все архитектурные и оптимизационные параметры задаются пользователем и не меняются в процессе обучения нейронной сети. Эти параметры называются гиперпараметрами, и их необходимо оптимизировать для повышения эффективности нейронной сети.

Оптимизация гиперпараметров должна проводиться по всему числу параметров вместе, так как один гиперпараметр может влиять на другой. Совокупность гиперпараметров и их возможных значений называется пространством гиперпараметров, и в этом пространстве необходимо провести оптимизацию, найдя оптимальное сочетание гиперпараметров для каждой конкретной задачи. Руководствоваться в этой задаче нужно специальной функцией оценки модели, например, площадью под ROC-кривой для бинарной классификации (AUC). Модель тренируется с определенным набором гиперпараметров, затем выбранная метрика оценивается либо на отдельной тестовой выборке (для этого исходный датасет необходимо разбить на три части: тренировочную, тестовую и валидационную), либо с использованием кросс-валидации. Проводить оптимизацию в пространстве гиперпараметров можно разными способами. Можно перебрать все сочетания гиперпараметров, полностью исследовав их пространство. Этот метод позволяет найти глобальный минимум, но требует существенных вычислительных затрат, если пространство гиперпараметров имеет большой размер и тренировка единичной модели требует заметных вычислительных ресурсов. Существуют алгоритмы, с помощью которых можно найти решение, близкое к оптимальному, за существенно меньшее время, например, *Bayesian Optimization* или *Random Search*.

Нашей группой была проведена оптимизация архитектурных гиперпараметров нейронной сети для типичной задачи физики высоких энергий. Их визуализация представлена на рис. 3. В пространстве гиперпараметров существуют довольно широкие области стабильности, дающие одинаково хороший результат. Для типичной задачи физики высоких энергий примерная область оптимальных гиперпараметров следующая: скрытые слои — 1–2 слоя, число нейронов в скрытых слоях — $(4-6)n$, где n — число входных переменных, значение дропаута — 0.2–0.3. Исследование пространства гиперпараметров было проведено с помощью пакета *Keras Tuner* [15].

Следующий возможный этап оптимизации анализа — это объединение моделей в ансамбль. В целом есть два пути формирования ансамбля: варьи-

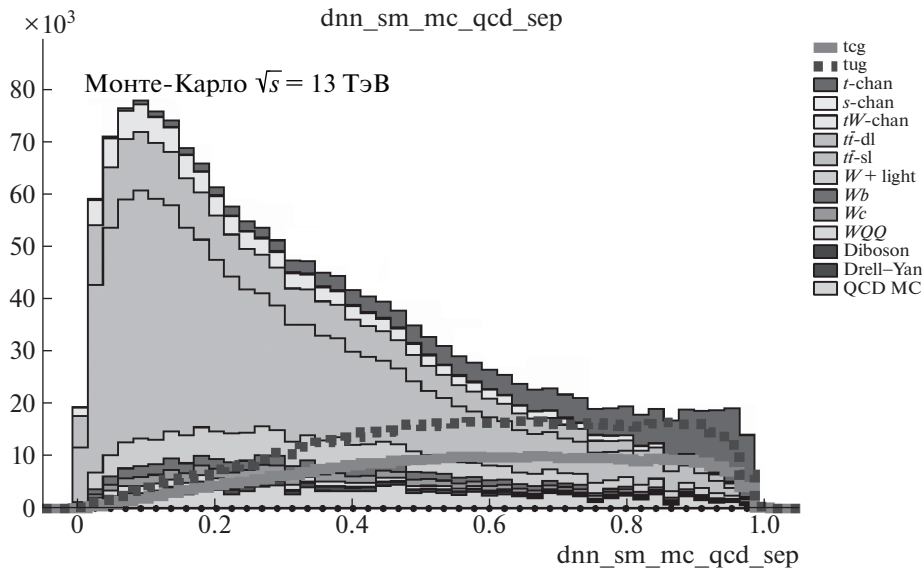


Рис. 1. Пример дискриминатора нейронной сети. Сигнальные процессы (t -chan) сгруппированы ближе к единице, фоновые события расположены ближе к нулю.

рование данных и тренировка модели на разных версиях датасета (Bagging), либо использование различных моделей, обученных на одном датасете (Stacking). В данной статье будет описан первый метод.

Для реализации метода бэггинга выбранная модель тренируется не на исходном датасете, а на модифицированном. Для получения такого датасета события из исходного выбираются с заменой: таким образом, какие-то события повторятся несколько раз, а каких-то не будет вовсе. Создав несколько таких датасетов, можно обучить на них несколько моделей и объединить их в единый ансамбль. Так как каждая индивидуальная модель работала с несколько измененным датасетом, их объединение позволяет уменьшить разброс (variance) на тестовых данных, уменьшая тестовую ошибку и повышая обобщающую способность всего предиктора, сравнение эффективности сетей в терминах AUC приведено на рис. 4.

Для событий с весами необходимо несколько модифицировать процедуру семплирования с заменой: необходимо учитывать веса каждого события при их случайной выборке. Вероятность выбора каждого события должна быть пропорциональна его весу. Таким образом, распределение по переменным в датасетах, полученных семплированием с заменой, будет соответствовать распределению в исходном (пример распределений с примененными весами приведен на рис. 5).

Последним возможным этапом оптимизации анализа стало использование каскадов нейронных сетей. Суть метода заключается в том, что тренируется не одна сеть, разделяющая весь фон

и сигнал, а несколько сетей, чувствительных к разным процессам из фона, если их свойства сильно различаются. Например, в данном анализе существуют кинематически похожие процессы, с рождением топ-кварков (процессы tt , tW , $t + X$) и более мягкие процессы без рождения топ-кварков ($W + \text{jets}$, WZ , WW , ZZ). Сети первого уровня тренируются отдельно выделять процессы с рождением топ-кварка от сигнала, вторая сеть тренируется выделять процессы ($W + \text{jets}$, ...) от сигнала, а затем выходы этих сетей соединяются в одномерный дискриминант с помощью третьей нейронной сети (суперсети), имеющей на входе дискриминаторы сетей первого уровня. Такой подход позволяет объединить две более чувствительные к отдельным процессам сети и изменить форму итогового дискриминатора (рис. 6).

3. СТАТИСТИЧЕСКИЙ АНАЛИЗ

Следующим этапом является статистический анализ полученного дискриминанта глубокой нейронной сети (DNN). Данный дискриминант можно представить в виде гистограммы распределения событий по значениям дискриминатора созданной нейронной сети (ансамбля или суперсети). Задача статистического анализа состоит в фитировании данного распределения с учетом данных и моделирования, оценке значений случайных параметров (например, сечений, ошибок реконструкции и т.д.) и их вероятностных распределений. На основе статистического анализа проводятся измерения параметров, представляющих интерес в задаче, или находятся ограничения на возможные значения

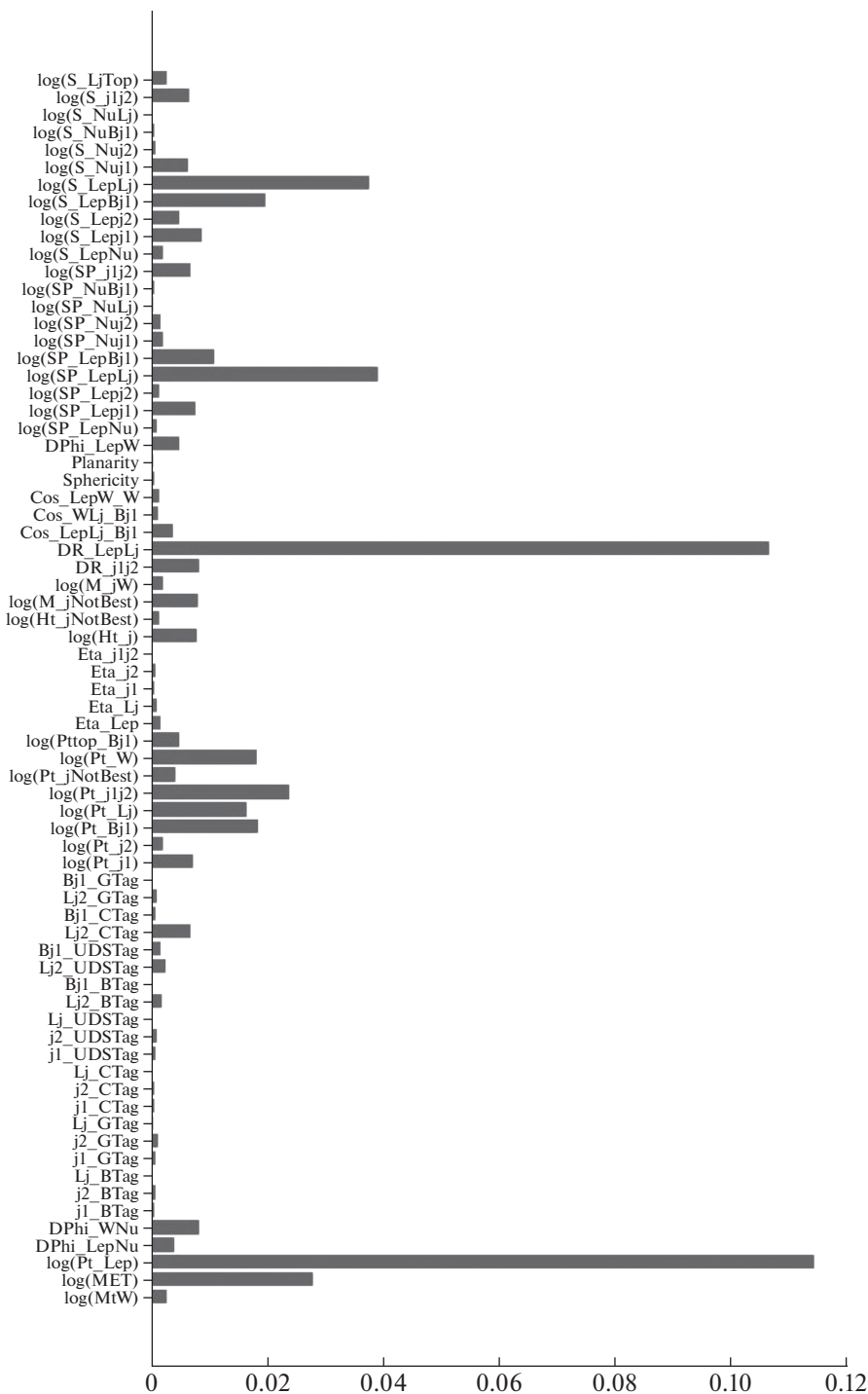


Рис. 2. Пример ранжирования переменных для модели нейронной сети по степени важности. Чем больше значение по абсциссе (условные единицы), тем больше значимость переменной для модели.

этих параметров. В нашей задаче искомыми случайными параметрами являются константы связи, характеризующие возможный вклад аномальных FCNC-взаимодействий, выраженные в терминах брэнчингов редких распадов топ-кварка $t \rightarrow cg$ и $t \rightarrow ug$.

Существует два основных подхода к статистиче-

скому анализу: частотный и байесовский. Частотные методы позволяют достичь высокой скорости обработки данных (например, асимптотические методы [16]), но, в отличие от байесовских методов, где на выходе статистического анализа вычисляются апостериорные распределения по каждому параметру, в частотных методах можно получить

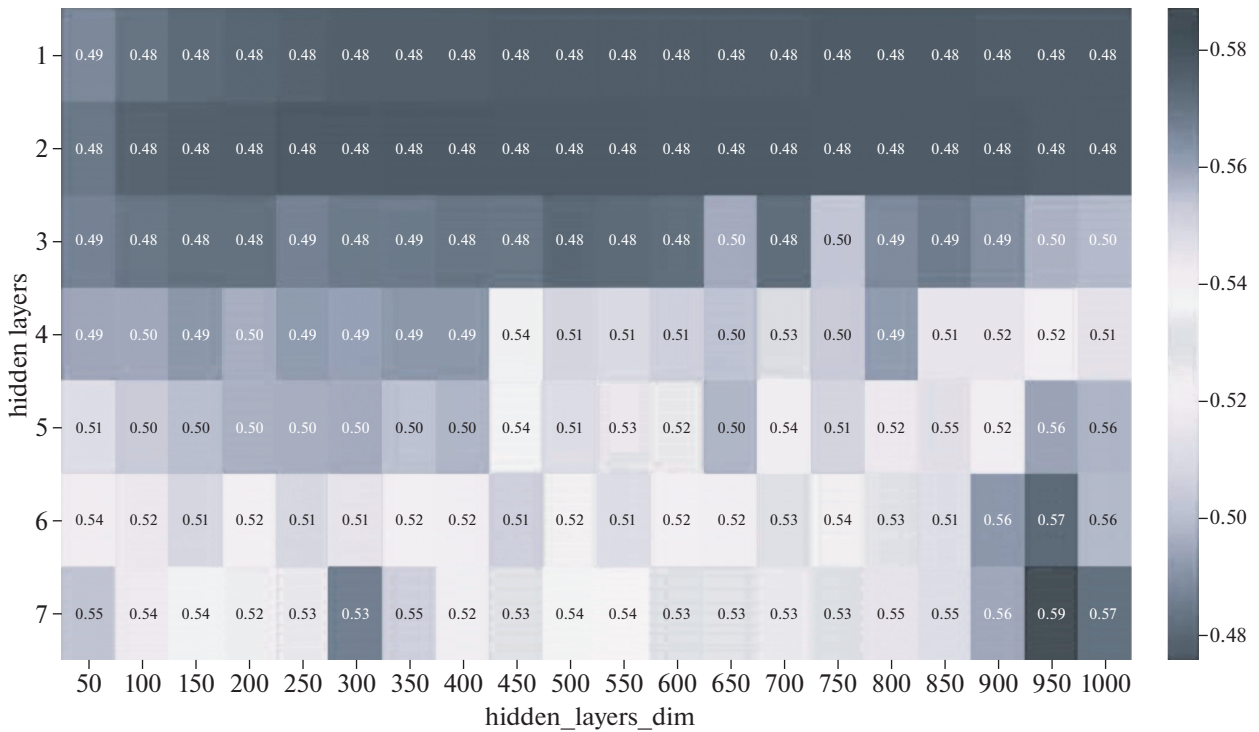


Рис. 3. Пример визуализации пространства гиперпараметров: показана зависимость предиктивной способности модели (функции бинарной кросс-энтропии: чем меньше, тем лучше) от числа нейронов в скрытых слоях (абсцисса) и числа скрытых слоев (ордината). По остальным гиперпараметрам было проведено усреднение. Заметен большой регион стабильности: для 1–2 слоев и 200–1000 нейронов сеть показывает почти одинаково хороший результат классификации.

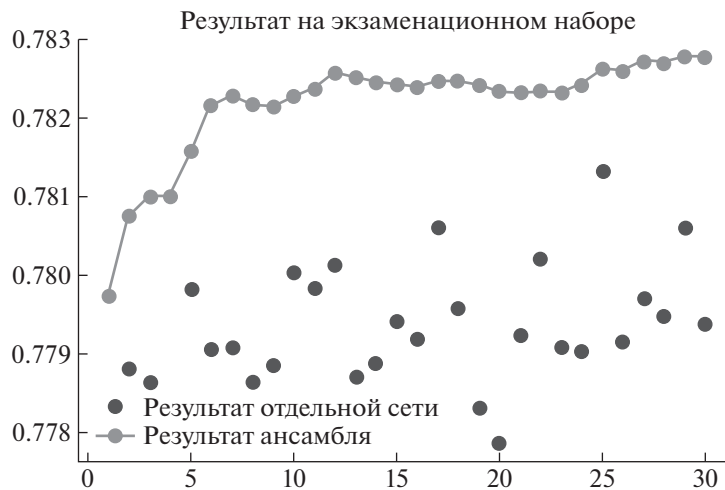


Рис. 4. Зависимость качества классификации AUC от числа моделей в ансамбле. Отдельными точками показаны метрики индивидуальных моделей. Обобщающая способность и AUC ансамбля превосходит любую индивидуальную модель.

только конечные ограничения и оценки. Для проведения статистического анализа используется пакет *theta* [17] и *CombinedLimit* [18, 19]. Далее все этапы анализа описаны на примере их реализации в пакете *theta*.

3.1. Построение статистической модели

В самом общем случае статистическая модель в *theta* представляет собой набор шаблонов для

наблюдаемых o_1, o_2, \dots, o_N . Каждая статистическая модель является функцией, зависящей от случайных параметров \mathbf{p} . Если исходные данные описываются гистограммой (в нашем случае дискриминантом DNN), то необходимо задать шаблоны для каждого физического процесса с ожидаемым количеством событий этого процесса в каждом

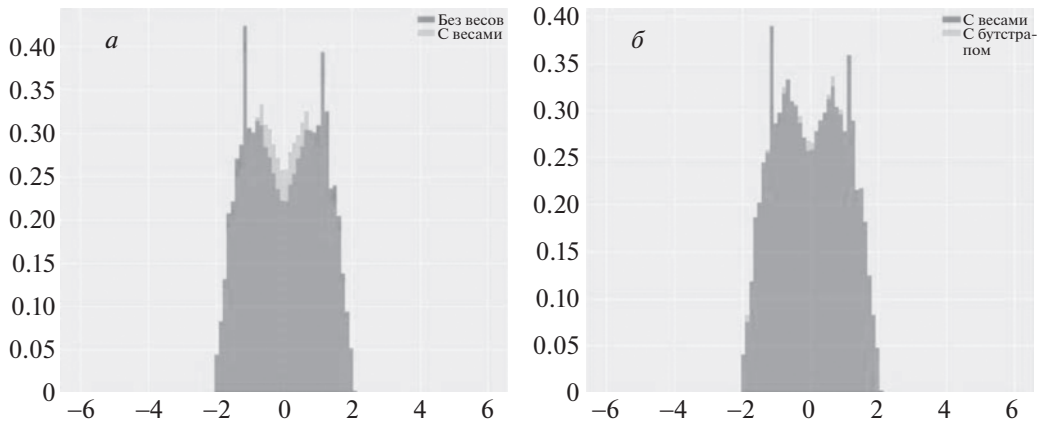


Рис. 5. Пример применения весов к бутстрап-семплированию событий: сравнение распределения одной наблюдаемой с примененными весами и без них. Рисунок 5б демонстрирует хорошее согласование распределений после семплирования.

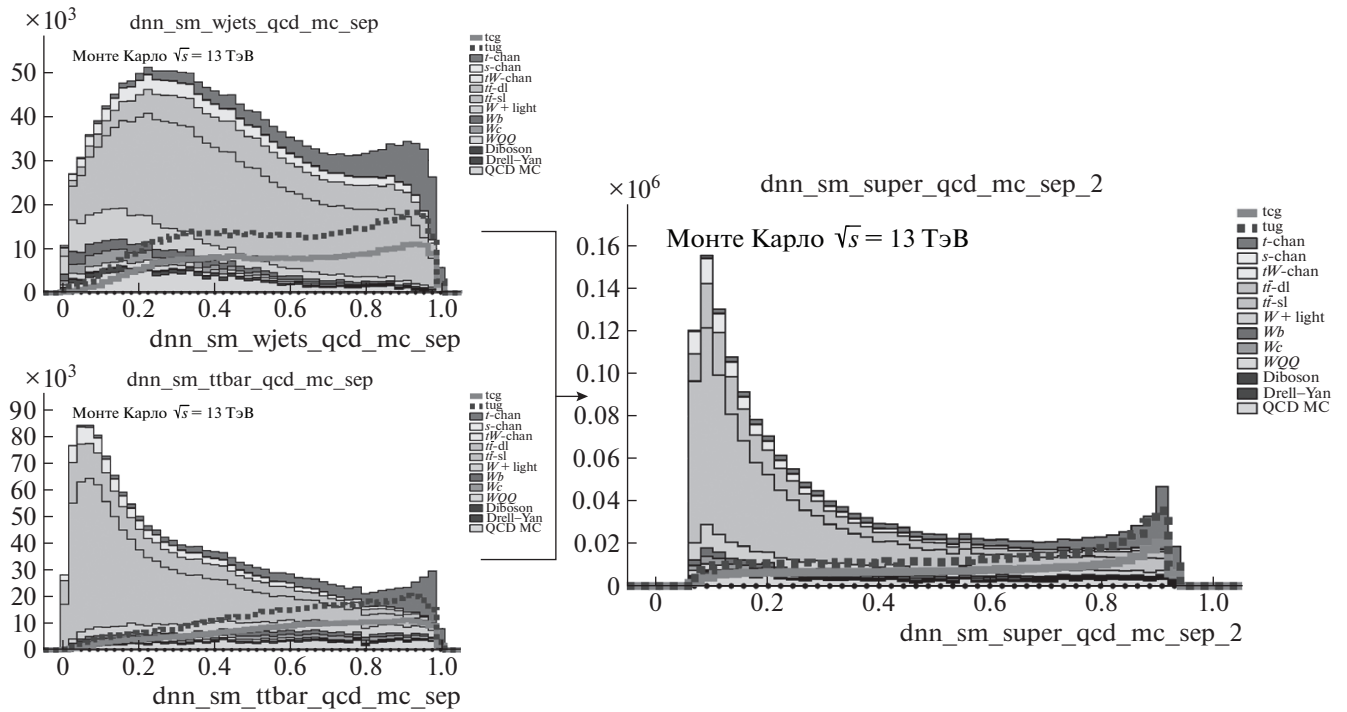


Рис. 6. Использование ансамбля сетей: сети первого уровня чувствительны к событиям ($t\bar{t}$, tW , $t + X$) и ($W + jets$), сеть второго уровня объединяет их выходы.

бине гистограммы, а затем просуммировать:

$$m_i(\mathbf{p}) = \sum_{k=1}^{M_i} c_{i,k}(\mathbf{p}) t_{i,k}(\mathbf{p}), \quad (1)$$

где $c_{i,k}$ — вещественные коэффициенты, а $t_{i,k}$ — шаблоны каждого из M_i процессов в модели. Принимается, что для каждого бина $t_{i,k}$ — строго положительная величина.

Тогда вероятность получить количество измененных данных d при условии параметров \mathbf{p} можно

записать как

$$p_m(d|\mathbf{p}) = \prod_{i=1}^N \prod_{l=1}^{b_i} \text{Poisson}(d_{i,l}|m_{i,l}(\mathbf{p})), \quad (2)$$

где индекс i пробегает по всем наблюдаемым, а индекс l — по бинам (b_i — количество бинов для наблюдаемой o_i). Здесь

$$\text{Poisson}(n|\lambda) = \frac{\lambda^n e^{-\lambda}}{n!} \quad (3)$$

— распределение Пуассона со средним λ .

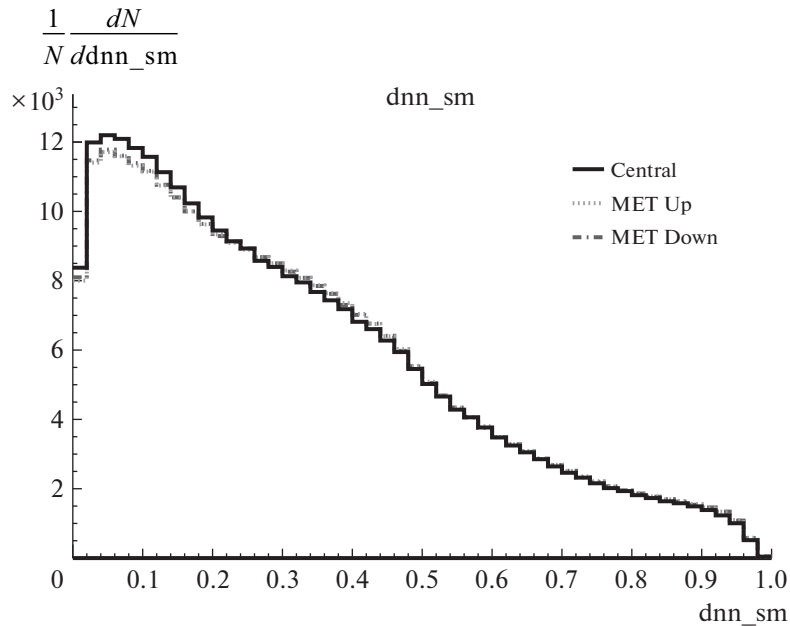


Рис. 7. Демонстрация влияния неопределенности на форму распределения. Приведено распределение смоделированных событий фона по величине дискриминатора нейронной сети. Показано распределение для центрального значения систематической ошибки определения недостающего поперечного импульса в событии (MET) и распределения дискриминанта после сдвига систематической неопределенности вычисления MET на $\pm 1\sigma$.

В нашем примере наблюдаемая единственная — дискриминант выхода нейронной сети; число бинов выбирается при проведении анализа вручную. В связи с неопределенностью выбора количества бинов должен быть проведен отдельный анализ влияния количества бинов на результат и достигнута стабильность результата относительно количества бинов.

3.2. Добавление систематических неопределенностей

Систематические неопределенности включают в модель как дополнительный вектор \mathbf{q} , который параметризует влияние различных неопределенностей на предсказанные распределения. В пакете theta эти параметры обрабатываются так же, как и \mathbf{p} , которые были описаны выше.

В самом простом случае систематическая неопределенность может являться лишь неопределенностью нормировки конкретного процесса. Этот случай обсуждается в разд. 3.2.1. В общем случае неопределенность может влиять также и на форму распределений (неопределенности формы). Этот случай описан в разд. 3.2.2.

3.2.1. Неопределенности нормировки. Для данного типа неопределенности добавление ее в модель заключается лишь в добавлении нового

нормировочного множителя τ , который будет варьироваться отдельно в соответствии с распределением возможных значений данной неопределенности. Случай $\tau = 1$ отвечает отсутствию изменений по данному типу неопределенности. В качестве примера из ранее описанной модели в разделе 3.1 неопределенность нормировки добавлена к первому процессу.

$$m_i(\mathbf{p}) = \tau c_{i,1}(\mathbf{p}) t_{i,1}(\mathbf{p}) + \sum_{k=2}^{M_i} c_{i,k}(\mathbf{p}) t_{i,k}(\mathbf{p}). \quad (4)$$

В нашем примере основными неопределенностями нормировки являются неопределенности теоретических сечений отдельных процессов и ошибка интегральной светимости, полученной в эксперименте. В качестве априорных распределений для сечений фоновых процессов берутся логнормальные распределения, а для сигнального процесса — плоское распределение в положительной области [19]. Выбор логнормальных распределений обусловлен тем, что полные сечения процессов всегда неотрицательны. Плоское априорное распределение для сечения сигнального процесса (или другого параметра интереса) обусловлено равной априорной вероятностью всех возможных значений этого измеряемого случайного параметра.

3.2.2. Неопределенности формы. В общем случае неопределенностей нормировки недостаточно для полного описания влияния ошибок на

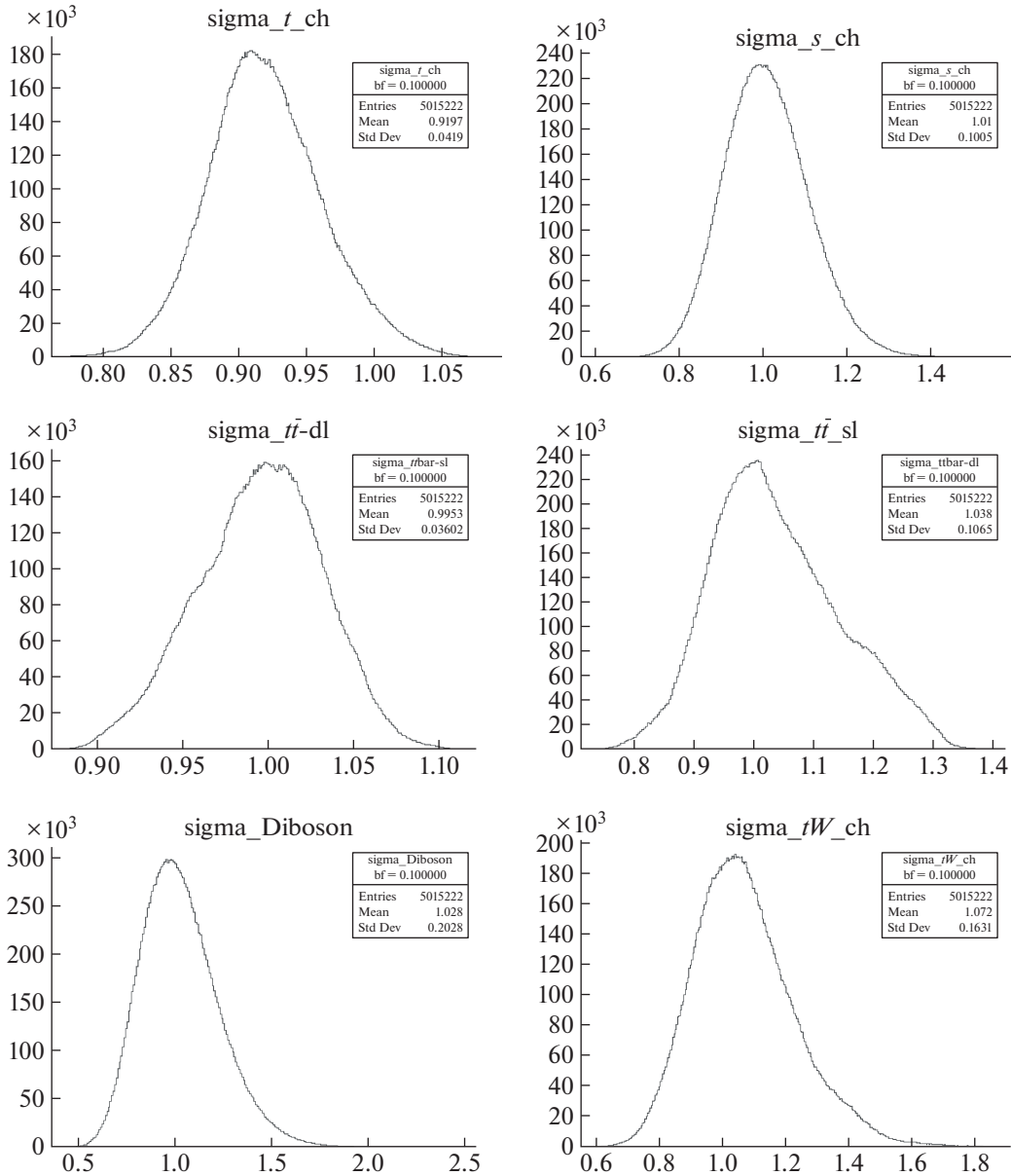


Рис. 8. Демонстрация получаемых в байесовском подходе плотностей вероятности для некоторых случайных параметров, не являющихся искомыми в задаче. Приведены апостериорные плотности вероятности значений сечений некоторых фоновых процессов, нормированных на теоретические сечения, полученные при анализе в пакете theta.

распределения наблюдаемых. Необходимо учитывать также влияние некоторых неопределенностей на форму распределений, т.е. могут получаться различные нормировочные коэффициенты для каждого бина диаграммы, а не одинаковый коэффициент для изменения нормировки конкретного процесса во всех бинах. В качестве примера приведено изменение распределения дискриминанта нейронной сети для фоновых процессов при варьировании недостающей поперечной энергии (MET) в пределах ошибки на $\pm 1\sigma$ (рис. 7).

Для включения таких неопределенностей в модель вводится параметр δ , который используется

для интерполяции между исходным шаблоном (без влияния систематики) и сдвинутыми шаблонами, проводится вариация соответствующей неопределенности на одно стандартное отклонение σ вверх и вниз в каждом бине. В theta используется функция интерполяции

$$t(\delta) = t \left(\frac{t^{\text{sign}(\delta)}}{t} \right)^{|\delta|}. \quad (5)$$

Здесь t — несмещенная (центральная) гистограмма, а t_{\pm} — смещенные гистограммы. Данный выбор функции не единственен, необходимо лишь соблюдение некоторых условий:

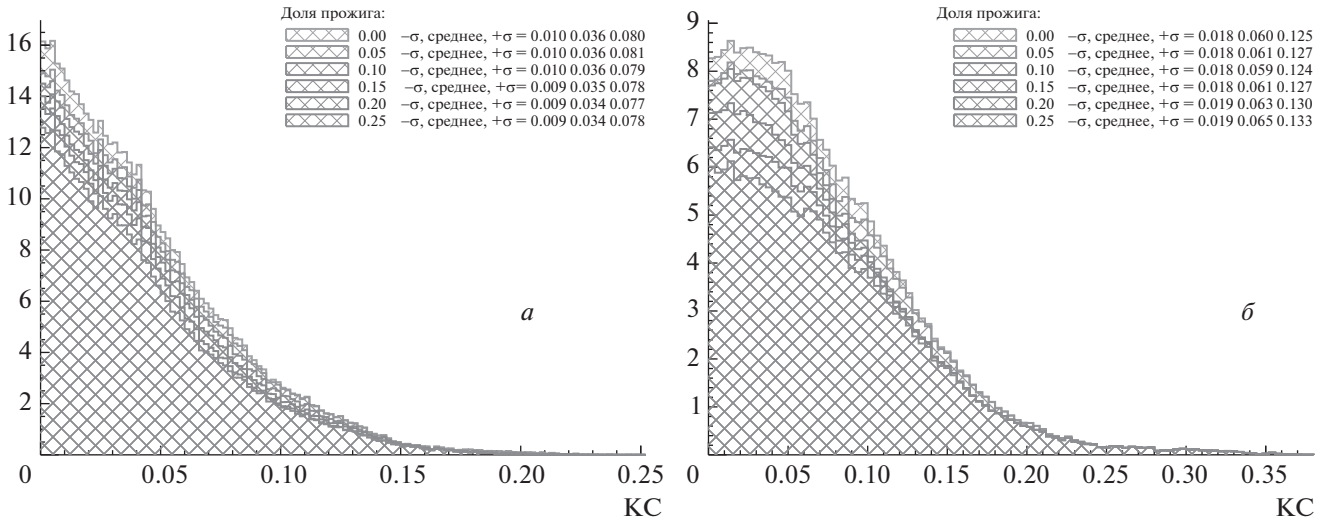


Рис. 9. Демонстрация получаемых в байесовском подходе плотностей вероятности искомого случайного параметра. Приведены апостериорные ожидаемые плотности вероятности параметра, характеризующего вклад $\text{tg} \text{ FCNC}$, полученные одной марковской цепью, рис. 9а и 9б демонстрируют распределения из разных марковских цепей. Приведены распределения для разного количества исключаемых первых итераций прожига марковской цепи, соответствующих исключению первых 25%, 20%–0% итераций.

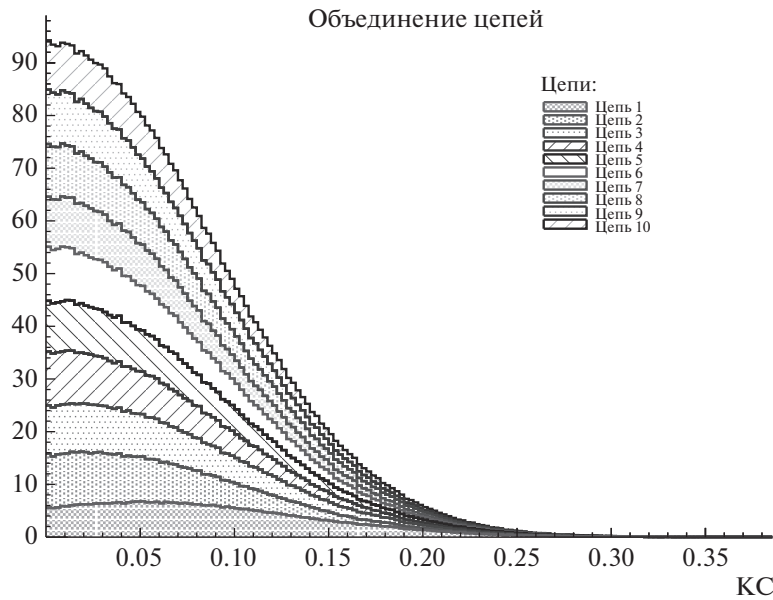


Рис. 10. Демонстрация объединения распределений из десяти марковских цепей. Приведена апостериорная ожидаемая плотность вероятности параметра, характеризующего вклад $\text{tg} \text{ FCNC}$ по результатам объединения десяти марковских цепей.

- Для любых значений δ все бины гистограммы должны быть положительны.
- Функция $t(\delta)$ должна иметь непрерывную первую производную.

3.3. Байесовский подход в статистическом анализе

В нашем примере используется байесовский подход к статистическому анализу. Такой метод,

в отличие от частотного, позволяет получать апостериорные распределения параметров, полностью описывающих все случайные параметры, а не только значения квантилей. В байесовской статистике апостериорное распределение параметра может считаться результатом измерения и основано на априорной информации о распределении данного параметра.

Рассмотрим простую модель с двумя парамет-

рами: сечением сигнала μ_s и сечением фона μ_b . Априорные распределения для параметров зададим как $\pi(\mu_s)$, $\pi(\mu_b)$. Тогда по теореме Байеса получим

$$p(\mu_s, \mu_b|d) = p(d|\mu_s, \mu_b) \frac{\pi(\mu_s) \pi(\mu_b)}{\pi(d)}, \quad (6)$$

$p(d|\mu_s, \mu_b)$ получено при моделировании сигнала и фона, $\pi(d)$ вычисляется из условия нормировки.

Для выделения из полученного многомерного распределения искомого параметра интереса (в данном примере это μ_s) производится интегрирование по всем остальным случайным параметрам:

$$p(\mu_s|d) = \int p(\mu_s, \mu_b|d) d\mu_b. \quad (7)$$

Интегрирование в пакете theta реализовано при помощи алгоритма Markov Chain Monte Carlo. Апостериорные распределения некоторых вспомогательных случайных параметров статистической модели приведены на рис. 8. Показаны плотности вероятности значений сечений, нормированных на теоретически предсказанное сечение, для процессов одиночного t -, s - и tW -каналов одиночного рождения топ-кварка, полулептонного и дилептонного каналов парного рождения топ-кварков и рождения двух массивных калибровочных бозонов.

Из полученного апостериорного распределения стандартным образом извлекаются среднее и квантили уровней 0.16 и 0.84 (отклонения на $-\sigma$ и $+\sigma$ соответственно). В табл. 1 приведены вышеописанные квантили для сечений основных процессов и светимости.

3.4. Многомерный статистический анализ

В качестве альтернативы анализу дискриминанта одной сети, в нашем примере натренированной на разделение t -канального одиночного рождения t -кварка (t -ch) от всех остальных процессов, можно провести анализ дискриминантов двух сетей, обученных на разделение сигнала, и по отдельности двух фоновых процессов, сильно отличающихся между собой по кинематическим характеристикам: $t\bar{t}$ (парное рождение топ-кварков) и W jets (рождение W -бозона и адронных струй). Такой анализ практически не отличается от одномерного: исходное двумерное распределение проходит процедуру побинной развертки, после чего анализируется так же, как и выход одной сети. Это возможно в силу независимости анализа от порядка бинов. Полученные квантили для 2D-анализа приведены в табл. 2.

Таблица 1. Квантили для ожидаемых сечений основных процессов, нормированных на теоретически предсказываемые сечения, и светимости, полученные из анализа в пакете theta

Параметр	$-\sigma$	Среднее	$+\sigma$
sigma_t_ch	0.866	0.913	0.958
sigma_s_ch	0.911	1.01	1.11
sigma_tW_ch	0.906	1.06	1.23
sigma_t \bar{t} -sl	0.903	0.959	1.01
sigma_t \bar{t} -dl	0.96	1.08	1.22
sigma_Diboson	0.832	1.01	1.23
sigma_DY	0.805	0.975	1.19
sigma_WQQ	0.877	1.17	1.54
sigma_Wc	0.828	1.08	1.41
sigma_Wb	1.19	1.61	2.06
sigma_Wother	0.904	1.12	1.56
sigma_Wlight	0.775	1.02	1.34
sigma_QCD	0.491	0.644	0.799
светимость	0.989	1.01	1.03

Таблица 2. Квантили для ожидаемых сечений основных процессов, нормированных на теоретически предсказываемые сечения, и светимости, полученные из анализа при 2D анализе в пакете theta

Параметр	$-\sigma$	Среднее	$+\sigma$
sigma_t_ch	0.912	0.946	0.982
sigma_s_ch	0.936	1.04	1.15
sigma_tW_ch	1.02	1.16	1.34
sigma_t \bar{t} -sl	0.936	0.969	1
sigma_t \bar{t} -dl	1.08	1.15	1.2
sigma_Diboson	0.854	1.04	1.28
sigma_DY	0.839	1.01	1.22
sigma_WQQ	0.926	1.21	1.53
sigma_Wc	1.1	1.34	1.58
sigma_Wb	1.57	1.82	2.09
sigma_Wother	1.1	1.28	1.49
sigma_Wlight	0.631	0.802	1.03
sigma_QCD	0.635	0.698	0.76
светимость	0.98	0.992	1

Таблица 3. Квантили для ожидаемых сечений основных процессов, нормированных на теоретически предсказываемые сечения, и светимости, полученные из анализа в пакете CombinedLimit

Параметр	$-\sigma$	Среднее	$+\sigma$
sigma_t_ch	0.852	0.888	0.936
sigma_s_ch	0.915	1	1.11
sigma_tW_ch	0.942	1.05	1.15
sigma_tt-sl	0.942	0.968	0.981
sigma_tt-dl	0.992	1.05	1.11
sigma_Diboson	0.845	1.01	1.21
sigma_DY	0.839	1.01	1.22
sigma_WQQ	0.943	1.2	1.55
sigma_Wc	0.916	1.17	1.45
sigma_Wb	1.1	1.34	1.75
sigma_Wother	1	1.21	1.38
sigma_Wlight	0.775	0.959	1.21
sigma_QCD	0.534	0.71	0.855
светимость	0.986	1	1.02

4. ПРОВЕРКА РЕЗУЛЬТАТОВ СТАТИСТИЧЕСКОГО АНАЛИЗА

На примере результатов поиска FCNC в данном разделе описываются различные методы проверки и контроля статистического анализа. Данные методы позволяют убедиться в непредвзятости анализа в зависимости от различных параметров. Для каждого полученных результатов проводится проверка независимости от числа бинов гистограммы, а также любых других желаемых параметров анализа путем нескольких запусков при разных значениях параметра и дальнейшего построения зависимости ограничений от выбранного параметра.

Так как при интегрировании используются марковские цепи, то проводится визуальный анализ апостериорных распределений от величины прожига (burn-in) — числа начальных точек цепи, исключаемых из анализа в силу их зависимости от стартового значения. Данные распределения приведены на рис. 9. На рисунке показаны апостериорные гистограммы для проверки burn-in для различной доли исходных данных, удаленных при прожиге: от 0 до 0.25 с шагом 0.05.

4.1. Объединение марковских цепей

В силу большого числа учитываемых систематических неопределенностей в анализе и большой размерности пространства случайных параметров

стала наблюдаться нестабильность в апостериорных распределениях. В качестве одного из решений данной проблемы анализ проводится несколько раз при одинаковых условиях, запускается несколько марковских цепей, а затем апостериорные распределения суммируются. Распределения для некоторых отдельных итераций приведены на рис. 9, объединенное распределение нескольких марковских цепей приведено на рис. 10. В отличие от простого усреднения квантилей каждой итерации, такой метод статистически объективнее, так как сохраняет и форму распределения в процессе суммирования. Объединение цепей позволяет избежать предвзятости при выборе конкретной итерации анализа.

4.2. Сравнение пакетов

Для дальнейшей проверки результатов проводится сравнение анализа, проведенного при помощи пакета theta (табл. 1), с полностью аналогичным анализом в CombinedLimit (табл. 3). В CombinedLimit был применен метод MarkovChainMC. Такое сравнение позволяет убедиться в независимости результатов от выбранного пакета для анализа и повышает достоверность полученных результатов.

5. ЗАКЛЮЧЕНИЕ

В статье приведены основные этапы применения методов глубоких нейронных сетей в задаче поиска “новой физики” в анализе данных коллайдерных экспериментов. Рассмотрена задача поиска нейтральных токов, меняющих аромат кварков, во взаимодействии топ-кварка с u - или c -кварком и глюоном, и описаны методы оптимизации применения глубоких нейронных сетей и их статистической интерпретации для измерения ожидаемых ограничений на параметры, характеризующие возможный вклад таких отклонений от СМ. Сформулированные рецепты могут быть напрямую использованы в других задачах поиска “новой физики” в коллайдерных экспериментах.

Результаты применения разрабатываемых методов в задаче поиска tqg FCNC получены при поддержке гранта Российского научного фонда (проект № 22-12-00152). Исследования по оптимизации построения нейронных сетей выполнены в рамках научной программы Национального центра физики и математики, проект “Физика частиц и космология”.

СПИСОК ЛИТЕРАТУРЫ

1. E. E. Boos, V. E. Bunichev, L. V. Dudko, M. A. Perfilov, and G. A. Vorotnikov, *Phys. At. Nucl.* **83**, 984 (2020); arXiv: 2004.14498.
2. E. Boos, V. Bunichev, M. Dubinin, L. Dudko, V. Ilyin, A. Kryukov, V. Edneral, V. Savrin, A. Semenov, and A. Sherstnev (CompHEP), *Nucl. Instrum. Methods A* **534**, 250 (2004); hep-ph/0403113.
3. A. Pukhov, E. Boos, M. Dubinin, V. Edneral, V. Ilyin, D. Kovalenko, A. Kryukov, V. Savrin, S. Shichanin, and A. Semenov, hep-ph/9908288.
4. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *JHEP* **1407**, 079 (2014); arXiv: 1405.0301.
5. S. Alioli, P. Nason, C. Oleari, and E. Re, *JHEP* **1006**, 043 (2010); arXiv: 1002.2581.
6. T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *Comput. Phys. Commun.* **191**, 159 (2015); arXiv: 1410.3012.
7. J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaitre, A. Mertens, and M. Selvaggi (DELPHES 3), *JHEP* **1402**, 057 (2014); arXiv: 1307.6346.
8. CMS Collab. (V. Khachatryan *et al.*), *JHEP* **1702**, 028 (2017); arXiv: 1610.03545.
9. D. Guest, K. Cranmer, and D. Whiteson, *Ann. Rev. Nucl. Part. Sci.* **68**, 161 (2018); arXiv: 1806.11484.
10. R. M. Bianchi, PhD Thesis, Freiburg U. (2014), <https://cds.cern.ch/record/1757501>
11. T. Finke, M. Krämer, A. Morandini, A. Mück, and I. Oleksiyuk, *JHEP* **2106**, 161 (2021); arXiv: 2104.09051.
12. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, *et al.*, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems* (2015), software available from <https://www.tensorflow.org/>
13. L. Dudko, G. Vorotnikov, P. Volkov, M. Perfilov, A. Chernoded, D. Ovchinnikov, and A. Shporin, *Int. J. Mod. Phys. A* **35**, 2050119 (2020); arXiv: 2002.09350.
14. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, *J. Machine Lear. Research* **12**, 2825 (2011).
15. T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, L. Invernizzi, *et al.*, *Keras Tuner (2019)*, <https://github.com/keras-team/keras-tuner>
16. G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Eur. Phys. J. C* **71**, 1554 (2011); *Eur. Phys. J. C* **73**, 2501 (Erratum) (2013); arXiv: 1007.1727.
17. Th. Müller, J. Ott, and J. Wagner-Kuhr, Tech. Rep. (2010), CMS-IN-2010/017.
18. CMS-NOTE-2011-005; ATL-PHYS-PUB-2011-11; Tech. Rep., CERN, Geneva (2011), <https://cds.cern.ch/record/1379837>; <https://github.com/cms-analysis/HiggsAnalysis-CombinedLimit>
19. G. Cowan, *Statistics*, Chap. 39 in Particle Data Group, <http://pdg.lbl.gov/2016/reviews/rpp2016-rev-statistics.pdf>; C. Patrignani *et al.* (Particle Data Group), *Chin. Phys. C* **40**, 100001 (2016).

METHODOLOGY FOR THE APPLICATION OF DEEP NEURAL NETWORKS IN THE SEARCH FOR “NEW PHYSICS” AT COLLIDERS AND STATISTICAL INTERPRETATION OF THE EXPECTED RESULTS

E. Abasov¹⁾, M. Belobrova¹⁾, P. Volkov¹⁾, G. Vorotnikov¹⁾, L. Dudko¹⁾, A. Zaborenko¹⁾,
M. Perfilov¹⁾, E. Sivakova¹⁾

¹⁾ Skobeltsyn Institute of Nuclear Physics of Lomonosov Moscow State University, Russia

This article focuses on methods of application of neural networks to search for “New Physics” in collider experiments. Mechanisms of observable selection, hyperparameter optimization, and constructing cascades of deep neural networks are considered. Methods for statistical analysis of neural network discriminants in packages theta and CombinedLimit are presented, taking into account various uncertainties.