

**Федеральное государственное автономное образовательное учреждение
высшего образования Национальный исследовательский университет
«Высшая школа экономики»**

Факультет компьютерных наук

Магистерская программа «Анализ данных в биологии и медицине»

КУРСОВАЯ РАБОТА

на тему “Прогнозирование исхода онкологических заболеваний при
проведения CAR-T клеточной терапии с помощью топологического
анализа данных”.

Выполнил

студент образовательной группы МАДБМ22

«Анализ данных в
биологии и медицине»

А.В. Пунько

Научный руководитель

Доцент, к.ф.-м.н.

В.Л. Чернышев

Москва 2023

Реферат

Изучение особенностей проведения Т-клеточной терапии химерным антигенным рецептором, а также прогнозирование исхода онкологических заболеваний при проведении клеточной терапии могут улучшить эффективность и безопасность данного вида лечения. Применение алгоритмов топологического анализа данных предоставляет возможность для решения таких задач.

В данной работе были проанализированы 2 набора данных: 1) данные доз CAR Т-клеток, которые получили пациенты с различными злокачественными гематологическими заболеваниями, 2) цитокиновый профиль пациентов с острым лимфобластным лейкозом, у которых после проведения CAR Т-терапии возник синдром высвобождения цитокинов, и цитокиновый профиль здоровых доноров. Применены три алгоритма топологического анализа данных: устойчивые гомологии, Mapper и RTD-AE.

В результате применения метода устойчивых гомологий построены диаграммы устойчивости, кривые Бетти, а также были сопоставлены метрики модели классификации RandomForest с использованием топологических данных и стандартного алгоритма. Для отображения основных особенностей данных и для кластеризации данных был получен топологический граф с помощью алгоритма Mapper. В результате применения алгоритма RTD-AE построен график R-Cross-Barcode, который измеряет различия топологии двух облаков точек, а также определено значение RTD.

При исследовании доз CAR Т-клеток, которые получили пациенты с гематологическими опухолями, топологические структуры не были обнаружены, однако при сравнении метрик модели классификации было показано, что точность классификации с использованием топологических данных составляет 57.1%, а при использовании стандартного алгоритма точность падает до 42.9%. При исследовании цитокинового профиля пациентов с CRS и здоровых доноров было обнаружено, что данные классифицируются на группы с точностью 100% даже стандартными методами.

Содержание

Реферат	2
Содержание	3
1. Введение	4
2. Топологический анализ данных	8
3. Описание данных	14
4. Описание эксперимента.....	15
5. Применение метода устойчивых гомологий	16
6. Применение метода Mapper.....	17
7. Применение метода RTD-AE	18
8. Результаты	19
Заключение	28
Список источников	29

1. Введение

Рак считается ведущей причиной смерти и важным препятствием на пути увеличения продолжительности жизни во всех странах мира. По оценкам, в 2020 году во всем мире было зарегистрировано 19,3 миллиона новых случаев и 10 миллионов смертей от рака. Прогнозируется, что в 2040 году во всем мире будет зарегистрировано около 28,4 миллиона новых случаев рака, что на 47% больше, чем в 2020 году, при условии, что показатели, оцененные в 2020 году, останутся постоянными [1, 2].

Для лечения рака используют хирургию, облучение, лекарства и другие методы уменьшения размера опухоли или остановки прогрессирования рака. В зависимости от конкретной ситуации, пациент может получать одно лечение или комбинацию процедур. Варианты лечения зависят от типа и стадии рака, общего состояния здоровья пациента. Целью первичного лечения является полное удаление рака из организма или уничтожение всех раковых клеток. Наиболее распространенным первичным лечением является хирургическое вмешательство. Целью адъювантной терапии является уничтожение любых раковых клеток, которые могут остаться после основного лечения, чтобы уменьшить вероятность рецидива рака. Общие адъювантные методы лечения включают химиотерапию, лучевую терапию и гормональную терапию. Неоадъювантная терапия аналогична, но лечение используется до основного лечения, чтобы сделать основное лечение более легким или более эффективным. Паллиативное лечение может помочь облегчить побочные эффекты лечения или признаки и симптомы, вызванные самим раком. Для облегчения симптомов можно использовать хирургию, лучевую терапию, химиотерапию и гормональную терапию. Другие лекарства могут облегчить такие симптомы, как боль и одышка [3].

В последнее время большое внимание уделяется иммунной системе пациентов и ее активации с помощью биологической терапии. Биологические методы лечения, включая иммунотерапию и терапию онколитического вируса, часто более физиологичны и хорошо переносятся.

Иммунная система помогает организму бороться с инфекциями и другими заболеваниями. Иммунотерапия — это лечение, при котором используются определенные части иммунной системы человека для борьбы с такими заболеваниями как рак. В последние несколько десятилетий иммунотерапия стала важной частью лечения некоторых видов рака. Новые методы иммунотерапии тестируются, одобряются и очень быстрыми темпами открываются новые способы работы с иммунной системой [4].

Для того, чтобы иммунная система находила и атаковала раковые клетки необходимо стимулирование или усиление естественной защиты иммунной системы, чтобы она работала усерднее или умнее. Также есть вариант создать в лаборатории вещества, похожие на компоненты иммунной системы, и использовать их для восстановления или улучшения работы иммунной системы для поиска и атаки раковых клеток.

Существует несколько основных типов иммунотерапии, используемых для лечения рака [5, 6]:

1) Ингибиторы контрольных точек (препараты, которые снимают «тормоза» с иммунной системы, что помогает ей распознавать и атаковать раковые клетки);

2) Т-клеточная терапия химерным антигенным рецептором (CAR) (терапия, при которой берут некоторое количество Т-клеток из крови пациента, смешивают их со специальным вирусом, который заставляет Т-клетки учиться прикрепляться к опухолевым клеткам, а затем возвращает клетки пациента, чтобы они могли найти, прикрепиться и убить рак);

3) Цитокины (стимулируют иммунные клетки к атаке рака);

4) Иммуномодуляторы (укрепляют часть иммунной системы для лечения определенных видов рака);

5) Противораковые вакцины (запускают иммунный ответ против определенных видов рака);

6) Моноклональные антитела (искусственные версии белков иммунной системы, они могут быть очень полезны при лечении рака, поскольку могут быть разработаны для воздействия на очень специфическую часть раковой клетки);

7) Онколитические вирусы (используются вирусы, которые были модифицированы в лаборатории для заражения и уничтожения определенных опухолевых клеток).

Иммунотерапевтические препараты одобрены для лечения многих видов рака. Однако иммунотерапия еще не так широко используется, как хирургия, химиотерапия или лучевая терапия. Иммунотерапия может вызывать побочные эффекты, многие из которых возникают, когда иммунная система, усиленная для борьбы с раком, также действует против здоровых клеток и тканей в организме [7, 8].

Значительный интерес у исследователей и онкологов вызвала клеточная генная терапия, а именно Т-клеточная терапия химерным антигенным рецептором (CAR), в результате которой Т-клетки изменяют таким образом, что они обретают способность находить и уничтожать раковые клетки [9].

Для осуществления CAR-T терапии в первую очередь проводится лейкоферез (извлечение лейкоцитов (включая Т-клетки) из крови пациента). После этого отделяют Т-клетки и изменяют их путем добавления гена специфического химерного антигенного рецептора (CAR). Затем эти клетки растут и делятся в лабораторных условиях. Для производства большого количества CAR-T клеток, необходимого для этой терапии, может потребоваться несколько недель. После этого CAR-T клетки можно вливать пациенту (Рисунок 1). CAR-T клетки могут уничтожить все раковые клетки и могут оставаться в организме через несколько месяцев после завершения инфузии. Такая терапия может привести к длительной ремиссии, а также может быть полезна, когда другие методы лечения уже не работают. В некоторых исследованиях до 90 процентов детей и взрослых с В-ОЛЛ, у которых заболевание либо неоднократно рецидивировало, либо не отвечало на стандартные методы лечения, достигли ремиссии после проведения терапии CAR-T клетками [10].

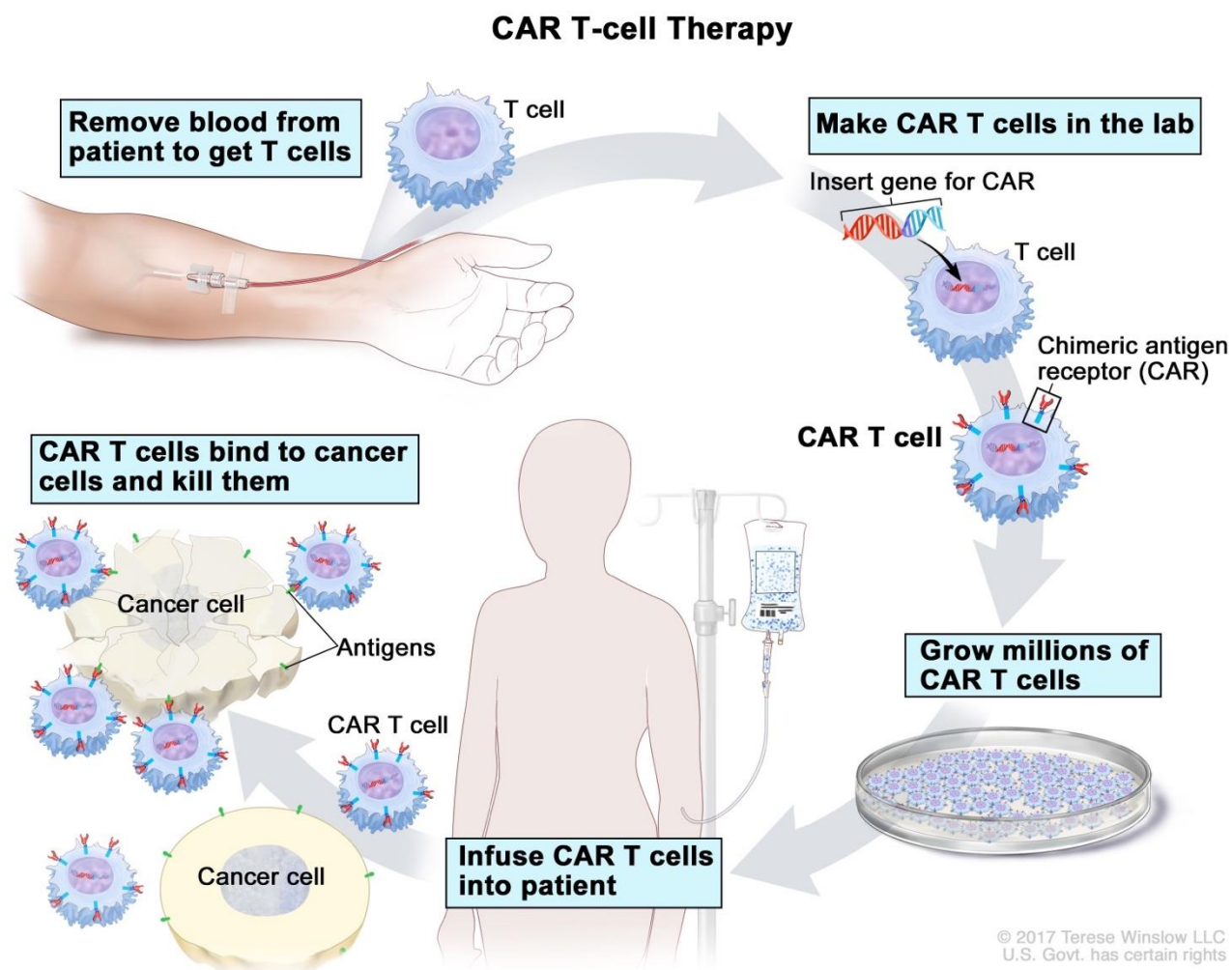


Рисунок 1 – Схема терапии CAR-T клетками [11].

Терапия CAR T-клетками может быть очень эффективной против некоторых видов трудно поддающихся лечению видов рака, но иногда она также может вызывать серьезные или даже опасные для жизни побочные эффекты [12].

Синдром высвобождения цитокинов: по мере размножения CAR T-клеток они могут выделять в кровь большое количество химических веществ, называемых цитокинами, которые могут усиливать иммунную систему. Серьезные побочные эффекты цитокинового шторма могут включать: высокая температура и озноб, затрудненное дыхание, сильная тошнота, рвота, головокружения, сильное чувство усталости, боль в мышцах, суставах, низкое кровяное давление/тахикардия, остановка сердца, сердечные аритмии, сердечная недостаточность, гипоксия, почечная недостаточность, плохая оксигенация легких, полиорганная недостаточность, неврологические симптомы.

Синдром активации макрофагов: этот побочный эффект тесно связан с тяжелым Синдромом высвобождения цитокинов. Данное состояние вызвано чрезмерной активацией и размножением T-клеток и макрофагов.

Нейротоксичность: иногда лечение может иметь серьезные последствия для нервной системы, что может привести к таким симптомам, как: головные боли, изменения в сознании, путаница или возбуждение, судороги, тремор, проблемы с речью и пониманием, потеря равновесия.

Синдром лизиса опухоли: это группа метаболических осложнений, которые могут возникать из-за распада умирающих клеток, обычно в начале токсического лечения рака. Однако Синдром лизиса опухоли может быть отсроченным и возникнуть через месяц или более после терапии CAR-T клетками. Синдром лизиса опухоли может вызывать повреждение органов и может быть опасным для жизни.

Анафилаксия: у пациента, получающего терапию CAR T-клетками, существует потенциальная вероятность подавляющего иммунного ответа против самого CAR, называемого «анафилаксией». Симптомы, связанные с анафилаксией, включают крапивницу, отек лица, низкое кровяное давление и респираторный дистресс.

Целевая, внеопухолевая токсичность: важным фактором безопасного и успешного использования CAR T-клеток является выбор правильного антигена, ассоциированного с опухолью, в качестве мишени. Многие опухолевые антигены также экспрессируются на здоровых клетках тканей. Повреждение нормальной ткани CAR-T-клетками может представлять опасный для жизни риск, особенно когда клетки основных тканей, таких как сердце, легкие или печень, экспрессируют антиген-мишень.

В-клеточная аплазия: терапия CAR-T клетками, нацеленная на антигены, обнаруженные на поверхности В-клеток, уничтожает не только раковые В-клетки, но и нормальные В-клетки. Следовательно, аплазия В-клеток является ожидаемым результатом успешного лечения CD19-специфическими CAR-T клетками и служит полезным индикатором продолжающейся активности CAR-T клеток. Однако этот эффект приводит к меньшей способности вырабатывать антитела, защищающие от инфекции [8, 13].

Для более эффективного проведения CAR-T терапии на различных стадиях применяются методы математического моделирования, а также аналитические методы, с помощью которых можно оценить и спрогнозировать безопасность и/или эффективность терапии, а также дополнить лабораторные исследования для понимания основных механизмов заболевания. Благодаря современным аналитическим методам исследования появляется возможность сопоставить клинические результаты с факторами, вызывающие проблемы с безопасностью и эффективностью терапии (включая такие проблемы как наработка CAR-T клеток, побочные явления, высокая стоимость терапии) [14].

2. Топологический анализ данных

Топологический анализ данных (Topological Data Analysis – TDA) предоставляет ряд потенциальных возможностей для решения проблем, возникающих при работе со сложными данными. TDA включает в себя набор инструментов для визуализации, исследования и анализа данных, основанных на топологии (область математики, изучающая абстрактные понятия формы и связности). Сосредоточив внимание на «форме» данных, TDA использует принципиально иную точку зрения, чем другие подходы. TDA – это недавнее направление, возникшее из различных работ в прикладной (алгебраической) топологии и вычислительной геометрии в первом десятилетии нашего столетия. TDA находится на пересечении анализа данных, алгебраической топологии, вычислительной геометрии, информатики, статистики и других смежных областей [15].

TDA наиболее полезен для изучения систем, которые учитывают преимущества топологической точки зрения, таких как системы с непрерывной или основанной на форме структурой, которая возникает в нескольких масштабах. Инструменты TDA способны сжато обобщить эту структуру таким образом, что часто обнаруживаются особенности, пропущенные другими методами. TDA также можно использовать как инструмент уменьшения размерности, который сохраняет основные топологические отношения или в сочетании с моделями машинного обучения для повышения их точности и способности обобщать новые данные. Кроме того, TDA может быть полезен для определенных типов данных, которые создают проблемы для стандартных инструментов, поскольку он более устойчив к определенным проблемам качества данных, таким как разреженность и шум, особенно для данных высокой размерности. Кроме того, TDA можно использовать для меньших наборов данных, чем было бы достаточно для обучения моделей машинного обучения или получения достаточной мощности для статистики.

Главной мотивацией TDA является идея о том, что топология и геометрия предоставляют мощный подход для вывода качественной, а иногда и количественной информации о структуре данных [16]. Цель TDA – предоставить обоснованные математические, статистические и алгоритмические методы для вывода, анализа и использования сложных топологических и геометрических структур, лежащих в основе данных, которые часто представлены в виде облаков точек в евклидовом или более общем метрическом пространстве. За последние несколько лет были проделаны значительные усилия по созданию надежных и эффективных структур данных и алгоритмов для TDA, которые теперь реализованы и доступны и легко используются через стандартные библиотеки. Хотя TDA все еще быстро развивается, он уже предоставляет набор зрелых и эффективных инструментов, которые могут использоваться в сочетании или в дополнение к другим инструментам науки о данных [17].

Для внедрения этого современного метода есть несколько веских причин. Во-первых, при количественном анализе упускается важная информация, зашифрованная в данных. Во-вторых, во многих системах неясно, какое значение имеют фактические расстояния между данными, поэтому метрики не всегда оправданы. TDA измеряет расстояния и кластеры, чтобы получить представление данных в топологических пространствах. TDA сочетает алгебраическую топологию с другими инструментами из чистой математики, что позволяет изучать форму данных. Высокая размерность наборов данных сильно ограничивает возможности по их визуализации. Именно поэтому TDA может улучшить способность иллюстрировать и анализировать эти данные. Наиболее широко обсуждаемые топологии данных включают связанные компоненты, туннели, пустоты и т.д.

Начиная с метрического пространства, то есть данных, в которых известно расстояние каждой точки до каждой другой, TDA позволяет измерить форму и структуру данных. Данные с расстоянием могут быть точками, пикселями; атомами, аминокислотами или биомолекулами; ядрами, клетками или организмами; генетическими и корреляционными расстояниями. TDA можно использовать не только для извлечения данных из формы, но и по своей сути TDA предполагает, что данные имеют форму. Концепция формы ограничена только природой лежащих в основе данных, и, если рассматривать ее в абстрактном смысле, способность измерять форму становится мощным инструментом анализа данных, который может быть применен практически к любому набору данных [18].

Методы TDA позволили получить множество новых знаний во все более разнообразных областях, включая изучение белков, 3-мерную структуру ДНК, развитие клеток, робототехнику, сигналы в изображениях, периодичность во временных рядах, исследования в онкологии, филогенетику, природные изображения, распространение инфекций, финансовые сети, различные приложения в нейронауке, сети сотрудничества, анализ данных мобильных телефонов, коллективное поведение в биологии, вывод временных рядов динамических систем, анализ естественного языка и многое другое [19].

Таким образом TDA является мощным инструментом, который способен дать критически важную информацию о прогнозе лечения, диагностике опухолей, классификации заболеваний и решить множество других задач в области онкологии [20, 21].

Метод устойчивых гомологий

Для изучения качественных характеристик данных с помощью TDA необходимы точные определения качественных характеристик, инструменты для их вычисления на практике и некоторая гарантия устойчивости этих характеристик. Одним из способов решения всех трех задач является метод устойчивых гомологий. Этот метод основан на алгебраической топологии, которая дает хорошо понятную теоретическую основу для изучения качественных характеристик данных со сложной структурой, вычислим с помощью линейной алгебры и устойчив к небольшим возмущениям во входных данных [22]. Диаграмма устойчивости может показать большое количество информации о данном облаке точек, например, кластеризацию без параметра связности, который обычно необходим. Она также может описывать более сложные структуры, такие как петли и пустоты, которые не видны при использовании других методов. Метод устойчивых гомологий относится к классу методов, использующих симплициальные комплексы для изучения топологии. Он позволяет выводить топологическую информацию из геометрических свойств объектов.

Опишем, что такое симплициальные комплексы. Данные представляют собой набор отдельных точек, а для того, чтобы хранить взаимосвязи между точками данных, используются графы. В некотором смысле, графы кодируют одномерный скелет данных. То есть вершины можно рассматривать как 0-мерные объекты, а ребра – как 1-мерные объекты. Но есть и более высокоразмерные отношения, которые теряются, когда мы видим только скелет. Симплициальные комплексы обобщают понятие графов, позволяя использовать 2-, 3- и более размерные строительные блоки, называемые симплексами. Симплициальный комплекс строится путем объединения симплексов различных размерностей. Например, две вершины могут быть объединены, чтобы создать ребро, три вершины – чтобы создать треугольник и т.д. Все симплексы в комплексе должны быть связаны друг с другом через общие грани [23]. Симплициальные комплексы – это геометрические объекты, которые возникают по мере увеличения некоторого радиуса вокруг точек в пространстве при их соединении. Интуитивно понятно, что структура симплициального комплекса на пространстве является выражением пространства как объединения точек, интервалов, треугольников и других аналогов более высокой размерности. Процесс объединения, при котором возникает последовательность вложенных симплициальных комплексов, называется фильтрацией. Фильтрация будет управлять тем, как точки данных преобразуются в симплициальные комплексы, чтобы преобразовать наборы точек в топологическое пространство. Симплициальные комплексы могут использоваться для представления формы данных в более высоких размерностях, что позволяет более точный анализ и распознавание шаблонов. Они также могут быть использованы для построения графов и сетей, которые помогают визуализировать данные и выявлять связи между ними [19].

Комплекс Вьеториса–Рипса. Следующей задачей является построение простого комплекса, представляющего структуру данных и использующего исходные данные в качестве набора вершин. Предположим, что мы определились с метрикой для точек данных и выбрали для начала число $t \geq 0$. Комплекс Вьеториса–Рипса для параметра t строится следующим образом. Набор вершин задается самим набором данных. Для каждой пары точек x, y в наборе данных мы включаем ребро xy , если расстояние между ними не больше t : $d(x, y) \leq t$. Для симплекса более высокой размерности, заданного вершинами x_0, \dots, x_d , мы включаем симплекс, если комплекс имеет все возможные ребра; в явном виде это означает, что каждая вершина x_0, \dots, x_d находится на расстоянии t от каждой другой вершины симплекса [23].

Идея устойчивых гомологий состоит из двух частей. Гомология – это инструмент из классической алгебраической топологии, который может измерять определенные структуры простого комплекса. "Устойчивая" часть происходит от рассмотрения всех возможных значений t , чтобы увидеть, где структура появляется и исчезает; то есть, коллекция значений t , для которых структура сохраняется. Это позволяет нам не выбирать конкретное значение t , а вместо этого посмотреть на континуум возможных значений t в поисках диапазонов, которые, как кажется, представляют что-то интересное с точки зрения структуры.

Гомологии делятся на различные измерения, представляющие размерность измеряемой структуры (Рисунок 2). Здесь 0-мерные гомологии измеряют связные компоненты, 1-мерные гомологии измеряют петли, а 2-мерные гомологии измеряют пустоты (как бы пузырьки воздуха). Также существуют определения k -мерных гомологий для более высоких k . Мы теряем способность интуитивно визуализировать структуры, которые мы фиксируем для более высоких k , поэтому они еще не нашли широкого применения в приложениях. На рисунке 2 приведено так называемое число Бетти β_k для различных примеров пространств. Число Бетти – это ранг k -мерной гомологической группы; в частности, оно подсчитывает количество структур, наблюдаемых в этом измерении. Числа Бетти – это формализация подсчета количества независимых k -мерных поверхностей в пространстве (β_0 означает количество связных компонент, β_1 – количество одномерных или "круглых" отверстий, а β_2 – количество двумерных "пустот" или "полостей"). Числа Бетти – это вектор или сигнатура целых чисел $(\beta_0, \beta_1, \beta_2)$, описывающих пространство. Этот вектор будем называть топологической сигнатурой данных.






	β_0	β_1	β_2	β_3
	1	•	•	•
	1	1	•	•
	1	•	1	•
	1	2	1	•
	1	2	1	•

Рисунок 2 – Числа Бетти β_k для различных топологических пространств [23].

В то время как гомологии измеряют структуру одного, неизменного пространства, устойчивые гомологии следят за тем, как эта структура меняется по мере изменения пространства. Таким образом, можно выбрать представителя одномерного гомологического класса при одном t и посмотреть, будет ли он по-прежнему представлять класс при большем t . Диаметр t , при котором класс появляется впервые, называется диаметром рождения, а диаметр t , при котором класс больше не отличается от ранее встречавшихся классов, называется диаметром смерти. Для каждого класса у нас есть пара чисел (a, b) для диаметра рождения и диаметра смерти. Если класс умирает очень скоро после своего рождения (то есть, если у него короткий срок жизни), то он будет представлен точкой, расположенной близко к диагонали. Если же класс имеет долгий срок жизни, то он будет представлен точкой, расположенной далеко от диагонали. Во многих приложениях существование нескольких точек, удаленных от диагонали относительно остальных точек на диаграмме устойчивости, может означать, что эти классы являются результатом присущей структуре данных, в то время как остальные точки являются артефактами шума [24, 25].

Метод Mapper

Еще одним методом, возникшим на основе TDA, является Mapper [26]. Идея заключается в том, чтобы представить одномерную структуру набора данных с помощью графа. В отличие от диаграммы устойчивости, точки данных имеют ассоциированное расположение с точками в графе, который строит метод Mapper. Поскольку эти графы гораздо легче визуализировать, чем, возможно, высокоразмерные данные, используемые для их построения, Mapper является интересным инструментом для исследования и визуализации структуры набора данных. Визуализация означает качественное представление, поэтому Mapper не является собственно алгоритмом анализа данных. Он может быть использован для создания топологического графа, который отражает основные характеристики данных. Преимуществами Mapper являются его нечувствительность к метрикам, понимание чувствительности к изменениям параметров и многомасштабный подход. Он широко используется для анализа данных, особенно в области биологии и здравоохранения [27 – 31].

Как и в случае с диаграммами устойчивости, мы начинаем с облака точек и выбора расстояния между ними. Однако, в отличие от диаграммы постоянства, нужно сделать еще несколько выборов. Для построения графа отображения необходима функция фильтрации. Функция фильтрации представляет собой назначение вещественного числа для каждой точки данных и используется для распределения данных. Затем выбирается покрытие значений функции фильтрации. Покрытие для интервала $[a, b]$ – это совокупность перекрывающихся множеств, таких, что каждое число между a и b входит хотя бы в одно из них. Наконец, мы рассматриваем подмножество точек со значениями функции в одном множестве из покрытия, а затем кластеризуем точки. Каждый кластер становится узлом в графе картографа. Грани включаются на основе информации о пересечениях из пересекающихся наборов в покрытии [23]. Важно отметить, что каждый узел графа отображения представляет собой подмножество точек данных. Это делает данный метод особенно полезным для исследования данных, поскольку он может разделять точки данных с различными свойствами, даже если стандартная кластеризация не может их различить.

3. Описание данных

В рамках курсовой работы исследовались данные пациентов со злокачественными гематологическими заболеваниями, которые получили лечение с помощью химерных антигенных рецепторов (CAR) Т-клеток [32 – 35]. Было проанализировано 54 пациента (31 мужчина и 23 женщины). Медиана возраста составила 26 лет (диапазон 5 – 77 лет). Иммуносупрессивные препараты, включающие системные кортикостероиды выше физиологической дозировки, не допускались в течение 4 недель до инфузии Т-клеток CAR19. Химиотерапия и терапия антителами должны были быть прекращены за 2 недели до инфузии Т-клеток CAR19. Пациенты с признаками реакции “трансплантат против хозяина” после проведения аллогенной трансплантации гемопоэтических стволовых клеток были исключены. Оценка ответа проводилась в определенные протоколом временные точки и была основана на Международных критериях ответа на лечение [36]. У пациентов наблюдались следующие виды ответа на терапию: полный ответ (complete response – CR), стабильная болезнь (stable disease – SD), прогрессивное заболевание (progressive disease – PD), частичная ремиссия (partial remission – PR), отсутствие ответа (no response – NR).

Также в работе исследовались данные пациентов с острым лимфобластным лейкозом, у которых после проведения CAR Т-терапии возник синдром высвобождения цитокинов (CRS), а также данные здоровых доноров, для которых был определен цитокиновый профиль [37]. Для прогнозирования возникновения синдрома высвобождения цитокинов были проанализированы следующие цитокины: IL2, IL4, IL5, IL6, IL7, IL8, IL10, IL12, IL13, IL15, IL17, sIL2R α , MCP1, EGF, IL1RA, sIL1RI, sIL1RII, sIL4R, sIL6R, sgp130, TNF α , IFN α , IFN γ , MIP1 α , MIP1 β , MIG, GCSF, GM-CSF, IP10. Было проанализировано 100 пациентов (50 пациентов с CRS и 50 доноров). Начало CRS определялось как день первой лихорадки у пациентов, у которых лихорадка появилась после инфузии CAR19. Согласно протоколу, лихорадка определялась как 38.0°C или выше. Прекращение CRS определялось как первый день отсутствия лихорадки. Все пациенты с лихорадкой были обследованы на наличие сопутствующих инфекций. Система оценки CRS [38] была следующей: Grade 1 – легкая реакция, которая лечится только поддерживающей терапией. Grade 2 – умеренная реакция, требующая внутривенной терапии или парентерального питания. Легкие признаки нарушения функции органов. Grade 3 – более тяжелые реакции. Умеренные признаки нарушения функции органов. Гипотония, лечимая внутривенными жидкостями или вазоактивными препаратами в малых дозах. Гипоксемия, требующая оксигенации. Grade 4 – угрожающие жизни осложнения, включая гипотонию, требующую применения высоких доз вазоактивных препаратов, или гипоксемию, требующую механической вентиляции легких. Grade 5 – смерть.

4. Описание эксперимента

На первом этапе для прогнозирования ответа были использованы данные о вводимой дозе CAR Т-клеток, индикация, возраст и пол пациентов со злокачественными гематологическими заболеваниями. На втором этапе были использованы данные пациентов с острым лимфобластным лейкозом, у которых после проведения CAR Т-терапии возник синдром высвобождения цитокинов, а также данные здоровых доноров, для которых был определен цитокиновый профиль.

Категориальные признаки были преобразованы с помощью Dummy-кодирования. Для применения методов топологического анализа использовался высокопроизводительный топологический инструментальный машинного обучения на Python, построенный поверх scikit-learn, – Giotto-tda [39]. Преимущество использования библиотеки Giotto-tda заключается в ее прямой интеграции с scikit-learn и поддержке табличных данных, временных рядов, графиков или изображений. Оба эти преимущества упрощают построение конвейеров машинного обучения. Метод устойчивых гомологий [40] использовался для количественной оценки топологических особенностей форм и функций в наборе данных, чтобы получить как качественную, так и количественную информацию. Алгоритм Mapper [41] использовался для упрощения и визуализации наборов данных. Также был применен новый метод снижения размерности с учетом топологии: автоэнкодер, усиленный дифференцируемыми потерями RTD: RTD-AE [42].

В данном исследовании анализируются два пространства размерностью 54x4 и 100x29, которые можно рассматривать как облако точек. Для первичной визуализации данных выполнялось снижение размерности до двух или трех компонентов с помощью библиотеки Uniform Manifold Approximation and Projection (UMAP) [43]. UMAP строит высокоразмерное графовое представление данных и оптимизирует низкоразмерный граф, чтобы он был как можно более структурно похож на данные. Для того чтобы построить исходный высокоразмерный граф, UMAP строит нечто, называемое нечетким симплициальным комплексом (представление взвешенного графа, с весами ребер, представляющими вероятность того, что две точки связаны). UMAP, как симплициальный комплекс, расширяет радиус от каждой точки, соединяя точки, когда эти радиусы пересекаются. Затем он "размывает" граф, уменьшая вероятность взаимодействия по мере увеличения радиуса. Данный метод гарантирует, что локальная структура поддерживается в равновесии с глобальной структурой.

5. Применение метода устойчивых гомологий

При использовании метода устойчивых гомологий в первую очередь необходимо определить облако точек. Затем используется метод фильтрации для создания простых комплексов и, наконец, определяются топологические сигнатуры данных (например, числа Бетти).

Первым шагом является установка трансформатора VietorisRipsPersistence [44] и вычисление так называемых диаграмм устойчивости для данной коллекции облаков точек. Комплекс Vietoris-Rips – это метод фильтрации. VietorisRipsPersistence – это устойчивая симплициальная гомология метрического пространства. В результате получаем диаграммы устойчивости. Далее вычисляем кривые Бетти и энтропию устойчивости из диаграмм устойчивости, чтобы найти топологические сигнатуры.

Вычисление топологических сигнатур позволяет охарактеризовать данные. С помощью этого обоснования также можно получить количественную информацию, которую можно использовать для статистической классификации. Giotto-tda с модулем gtdata.pipeline [45] расширяет модуль scikit-learn, определяя конвейеры для выполнения таких вычислений. Был построен базовый конвейер, позволяющий использовать алгоритмы классификации sklearn для получения статистики.

Сначала была посчитана диаграмма устойчивости и ее энтропия устойчивости. Затем был осуществлен поиск статистической модели в sklearn.ensemble, чтобы получить задачу классификации с соответствующим показателем точности классификации. Для выполнения классификации был выбран RandomForestClassifier [46].

Также была выполнена обычная классификация с помощью той же модели RandomForestClassifier без использования топологической информации данных, чтобы сравнить результаты.

С помощью конвейера был классифицирован первый набор данных между пациентами, которые полностью ответили на терапию CAR T-клетками и у которых возник рецидив. А также был классифицирован второй набор данных между пациентами, которые имели цитокиновый профиль характерный для синдрома высвобождения цитокинов.

6. Применение метода Mapper

Для визуализации данных с помощью алгоритма Mapper необходимо выбрать функцию фильтрации, определить покрытие и вычислить кластеризацию.

Обсудим выбор функции фильтрации. Библиотека Giotto-tda предлагает три альтернативы: проекция, энтропия и *eccentricity* [47]. Для облака точек в исследуемых датасетах была выбрана функция *eccentricity*, так как она обеспечивает лучшее разрешение фильтрации. Метрика функции расстояния по умолчанию для *eccentricity* – евклидово расстояние.

Покрытие. Существует два варианта: *CubicalCovering* для многомерных данных и *OneDimensionCover* для одномерных данных [48]. *CubicalCovering* имеет важные опции: количество интервалов и доли перекрытия. Большее количество интервалов означает большее количество точек для кластеризации и большее количество узлов в графе. Доля перекрытия устанавливает границу между двумя интервалами. В данном случае было установлено тридцать интервалов, чтобы иметь богатое представление о данных. Чтобы сохранить скрытые локальные структуры в данных, доля перекрытия выбрана между 30%.

Кластер. Алгоритм кластеризации по умолчанию – *scikit-learn DBSCAN* [49]. Самый важный выбор, который предстоит сделать, – это параметр *epsilon*, то есть максимальное расстояние между двумя выборками для того, чтобы одна из них считалась находящейся по соседству с другой. Результаты работы алгоритма Mapper полностью зависят от этого параметра. Ввиду его важности, будет использоваться метрика для проверки расстояния разделения между полученными кластерами DBSCAN (*Silhouette Coefficient*) [50].

Алгоритм Mapper строит пайплайн с параметрами, описанными выше. Для построения графа можно выбрать между статическим и интерактивным Mapper. Интерактивный картограф позволяет взаимодействовать с параметрами конвейера. В данном случае использовался *plot_static_mapper_graph* [51]. Размер узла – это метрика для представления количества точек в узле. Цвет узла представляет собой среднее значение точек, входящих в него.

7. Применение метода RTD-AE

Representation Topology Divergence (RTD) – метод снижения размерности с учетом топологии. RTD измеряет топологическое расхождение между двумя облаками точек при соответствии облаков один к одному. Основным препятствием для включения RTD в глубокое обучение является его дифференциация. Существуют подходы к дифференцированию бар-кодов, общих функций на основе бар-кодов относительно деформаций фильтрации [52] и к дифференцированию TDA в особых случаях [53, 54].

RTD-AE – новый метод снижения размерности с учетом топологии: автоэнкодер, усиленный дифференцируемыми потерями. Минимизация потерь RTD между реальным и латентным пространствами обеспечивает близость топологических особенностей и их локализацию с сильными теоретическими гарантиями. RTD-AE превосходит современные методы снижения размерности в плане сохранения глобальной структуры и топологии многообразия данных [42]. В некоторых случаях предложенный RTD-AE производит более точные и визуально привлекательные низкоразмерные вкрапления, чем современные алгоритмы.

8. Результаты

Прогнозирование ответа на иммунотерапию CAR T-клетками у пациентов с гематологическими заболеваниями.

Результаты систематического обзора литературы [55] свидетельствуют о том, что в терапии клеток CAR-T может быть оптимальная доза эффективности, при которой достигается максимальный клинический эффект и при которой не возникает необходимость в дополнительной противоопухолевой терапии. Однако увеличение дозы сверх оптимальной может привести к более высокой частоте и/или тяжести неблагоприятных событий. Подбор дозы является критически важным фактором, определяющим успех любой терапии рака, включая клеточную терапию. Для такого вида терапии, как CAR-T клетки, подбор дозы может быть более сложным, так как иммунотерапия не может быть описана типичными принципами клинической фармакологии.

В настоящее время рекомендации по начальной дозе даются на основе доклинических моделей и эмпирических данных предыдущих соответствующих исследований с аналогичными конструкциями при том же типе рака. Однако вопрос о возможном повышении эффективности при увеличении дозы продолжает оставаться предметом обсуждения в рамках клинической разработки, поскольку существуют противоречивые данные о дозовом ответе CAR-T клеток. В некоторых исследованиях сообщалось о положительной корреляции между увеличением ответа и повышением дозы [56, 57], тогда как в других исследованиях корреляции не наблюдалось, а эффективность была одинаковой при всех уровнях дозы [58].

В связи с этим были проанализированы данные пациентов (вводимая доза CAR T-клеток, индикация, возраст и пол) для прогнозирования ответа на проведение иммунотерапии. Оценка ответа проводилась в определенные временные точки и была основана на Международных критериях ответа на лечение [36]. У пациентов наблюдались следующие виды ответа на терапию: полный ответ (complete response – CR), стабильная болезнь (stable disease – SD), прогрессивное заболевание (progressive disease – PD), частичная ремиссия (partial remission – PR), отсутствие ответа (no response – NR).

Визуализации данных с помощью библиотеки UMAP.

На рисунке 3 показан результат визуализации набора данных с помощью UMAP. Точки окрашены значениями ответа пациентов на терапию, так что красные точки представляют пациентов, которые достигли полной ремиссии после проведения терапии, синие точки представляют пациентов, у которых наблюдался рецидив. Четкие границы между красными и синими точками в облаке точек отсутствуют.

UMAP projection 2D and 3D

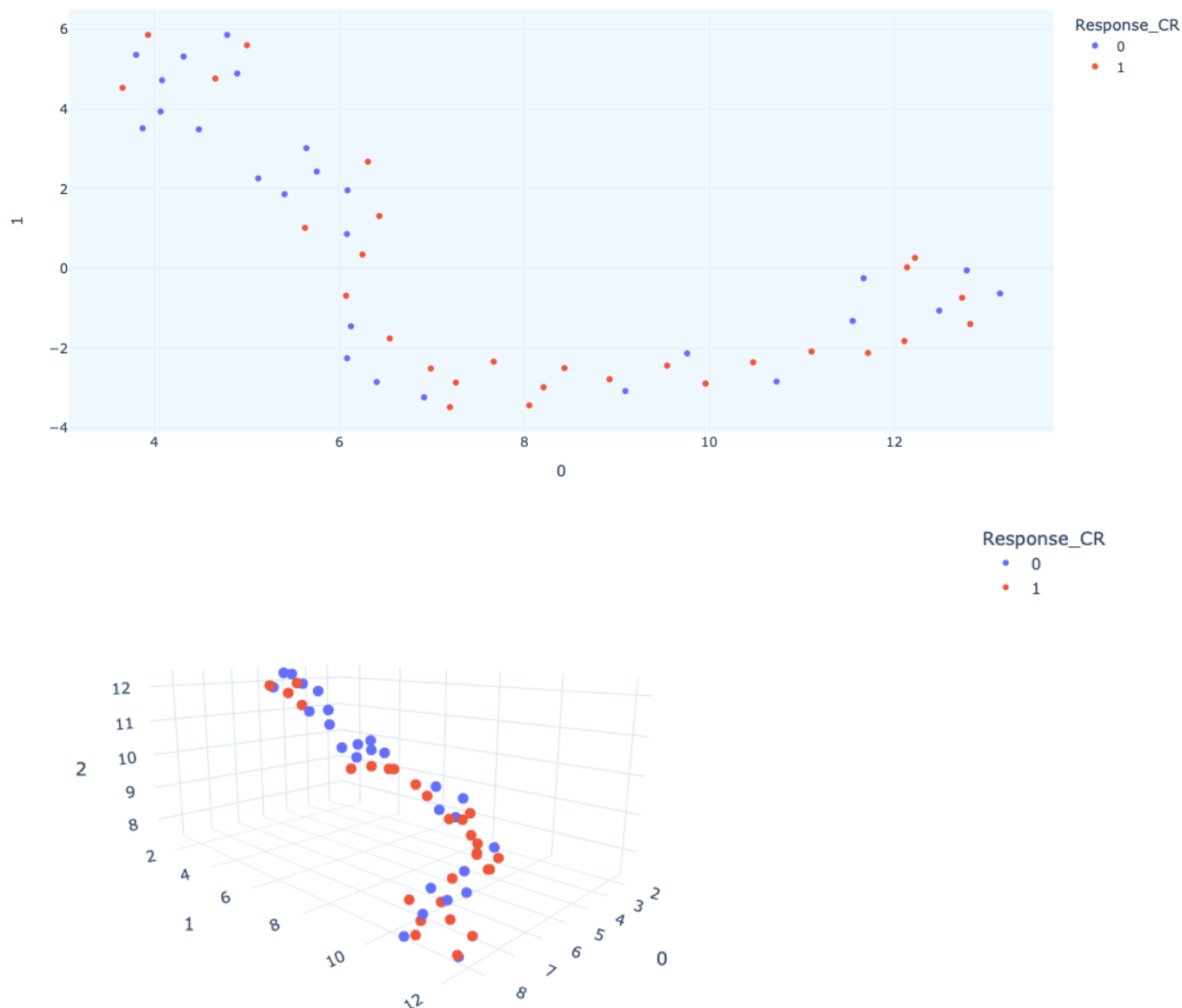


Рисунок 3 – Визуализация набора данных методом UMAP в 2D и 3D.

Применение метода устойчивых гомологий.

В результате применения метода устойчивых гомологий была построена диаграмма устойчивости (Рисунок 4). Оси показывают "рождение" и "смерть" компонент. H_0 (или β_0) обозначает количество связных компонент (информация о связности нулевого уровня), H_1 (или β_1) – количество одномерных отверстий (информация о связности первого уровня). Точки на пунктирной линии рассматриваются как шум. Для β_0 (оранжевым цветом) видно, что все компоненты начинаются с 0. Самая высокая точка соответствует пределу облака точек. Все оранжевые точки очень близки, кроме последних. Зеленые точки соответствуют β_1 . За пределами уровня шума точки не наблюдаются.

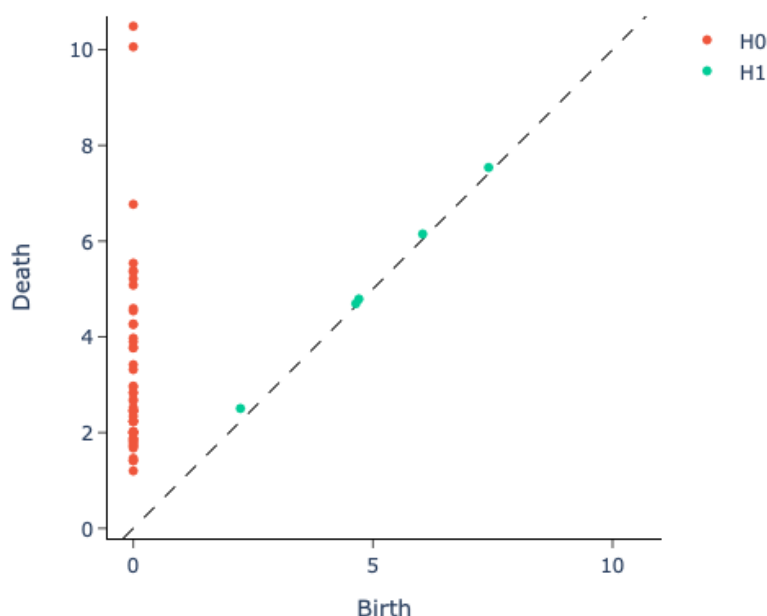


Рисунок 4 – Диаграмма устойчивости для параметра ответа на терапию.

Кривые Бетти показаны на рисунке 5. Параметр фильтрации (или значение радиуса роста, используемое для фильтрации) показан на оси x, а кривая Бетти – на оси y. Эта кривая подходит для деталей связности нулевого уровня, уровня один и уровня два, поэтому на рисунке вектор из трех чисел Бетти (β_0 , β_1 , β_2) для каждого параметра фильтрации.

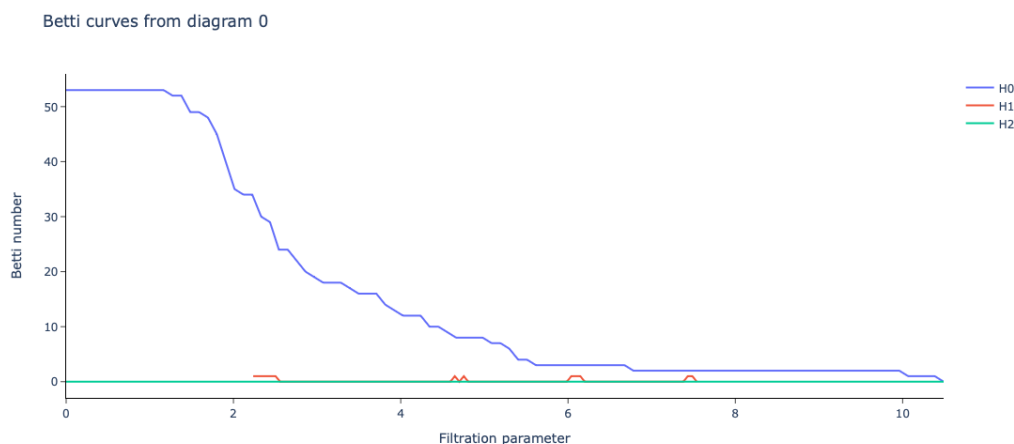


Рисунок 5 – Кривые Бетти для параметра ответа на терапию.

По всей видимости в нашем облаке точек отсутствуют какие-либо топологические структуры, однако после сравнения метрик модели классификации RandomForest оказалось, что точность классификации с использованием топологических данных составляет 57.1%, а при использовании стандартного алгоритма sklearn.ensemble точность падает до 42.9%.

Визуализации данных с помощью алгоритма Mapper.

В результате применения метода Mapper был получен топологический граф (Рисунок 6), который отражает основные особенности данных. Размер узла показывает количество точек в узле, а цвет узла представляет собой среднее значение точек, входящих в него. На графе видно 7 связных компонент и 10 несвязных узлов. С помощью данного алгоритма не было обнаружено никаких структур, топологических особенностей (например, петли, отверстия) в исследуемых данных. А в связи с тем, что такие структуры в дальнейшем используются для выявления кластеров, в данном случае кластеризовать данные не представляется возможным.

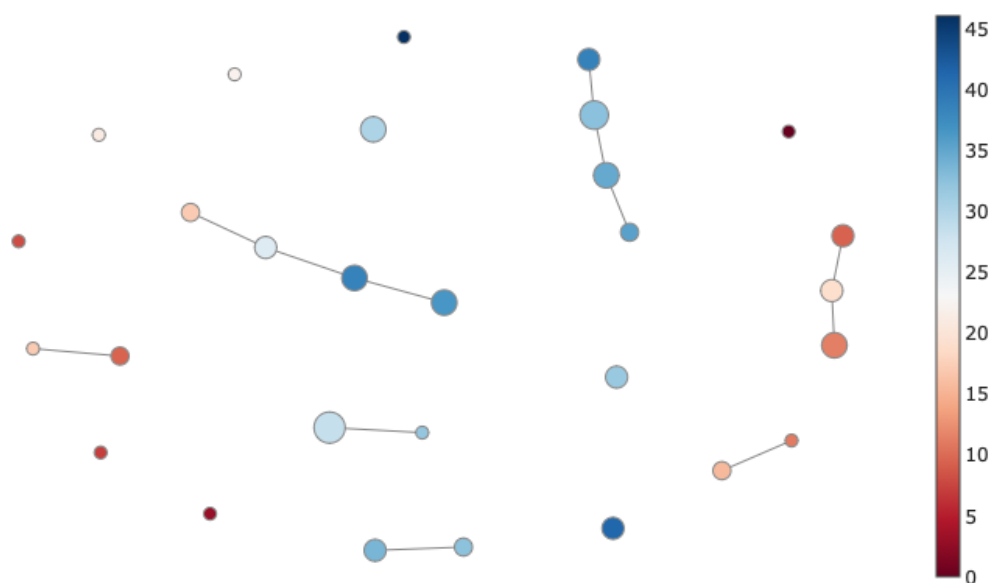


Рисунок 6 – Граф, построенный методом Mapper, для кластеризации данных по виду ответа на терапию.

Снижение размерности данных с учетом топологии.

В результате применения метода RTD-AE было измерено различие в топологии между двумя облаками точек (пациенты, у которых наблюдался полный ответ после терапии, и пациенты, у которых болезнь прогрессировала). Был построен график R-Cross-Barcode (Рисунок 7), который измеряет различия в многомасштабной топологии двух облаков точек с односторонним соответствием между точками. На основе R-Cross-Barcode было определено значение RTD – коэффициент, измеряющий многомасштабное топологическое различие между двумя наборами данных ($RTD = 1.46$).

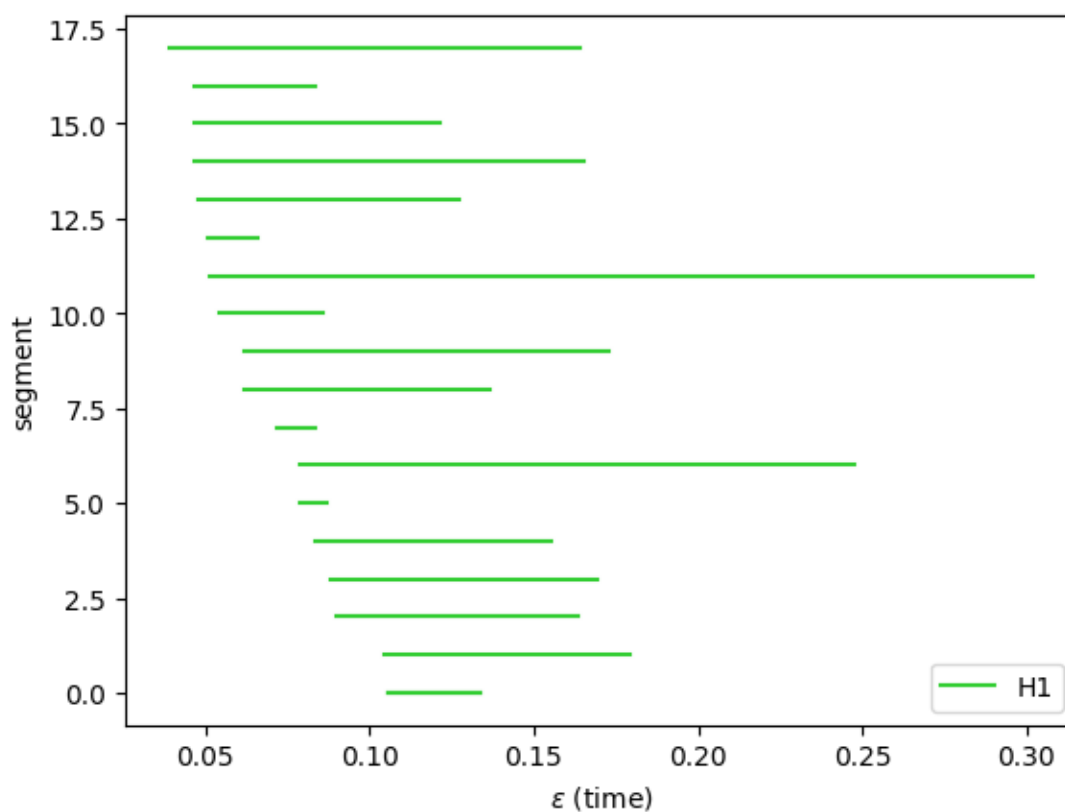


Рисунок 7 – График R-Cross-Barcode.

Прогнозирование синдрома высвобождения цитокинов после проведения иммунотерапии CAR T-клетками у пациентов с острым лимфобластным лейкозом.

Тяжелый синдром высвобождения цитокинов (CRS) – это потенциально опасная для жизни токсичность. Возможность предсказать, у каких пациентов может развиться тяжелая форма CRS до ее развития, может быть полезна для смягчения токсичности, так как цитокинонаправленная терапия может быть начата до того, как пациент станет тяжело больным. Пациенты, у которых прогнозируется развитие тяжелой формы CRS, могут находиться под более тщательным наблюдением для раннего начала агрессивной поддерживающей терапии. Напротив, способность предсказать, у каких пациентов маловероятно развитие тяжелого CRS, может предотвратить ненужную раннюю госпитализацию и/или воздействие ненужной цитокинонаправленной терапии.

В связи с этим были проанализированы данные пациентов (цитокиновый профиль пациентов с острым лимфобластным лейкозом, у которых после проведения CAR T-терапии возник CRS, а также данные здоровых доноров) для прогнозирования синдрома высвобождения цитокинов после проведения иммунотерапии CAR T-клетками.

Визуализации данных с помощью библиотеки UMAP.

На рисунке 8 показан результат визуализации набора данных с помощью UMAP. Синие точки представляют данные здоровых доноров, для которых был определен цитокиновый профиль, а красные точки представляют пациентов с острым лимфобластным лейкозом, у которых после проведения CAR T-терапии возник синдром высвобождения цитокинов. Между красными и синими точками в облаке точек есть четкие границы.

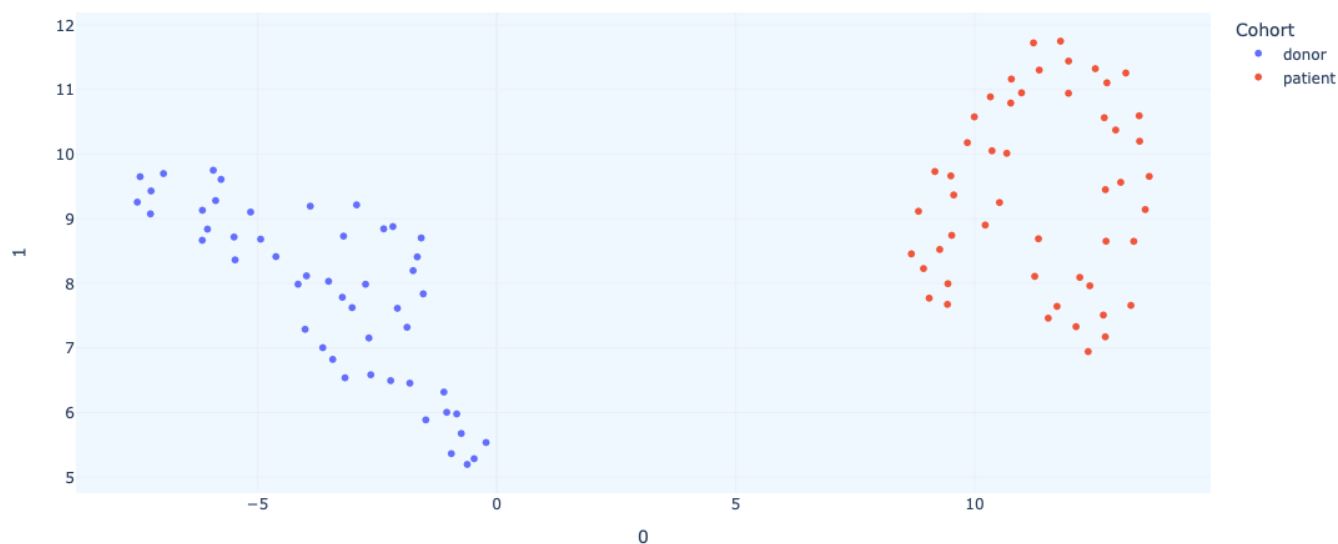


Рисунок 8 – Визуализация набора данных методом UMAP в 2D.

Применение метода устойчивых гомологий.

В результате применения метода устойчивых гомологий была построена диаграмма устойчивости (Рисунок 9). Оси показывают "рождение" и "смерть" компонент. H_0 (или β_0) обозначает количество связных компонент (информация о связности нулевого уровня), H_1 (или β_1) – количество одномерных отверстий (информация о связности первого уровня), H_2 (или β_2) – количество двумерных "пустот" или "полостей" на диаграмме. Точки на пунктирной линии рассматриваются как шум. Для β_0 (оранжевым цветом) видно, что все компоненты начинаются с 0. Самая высокая точка соответствует пределу облака точек. Зеленые точки соответствуют β_1 , фиолетовые точки соответствуют β_2 . За пределами уровня шума точки не наблюдаются.

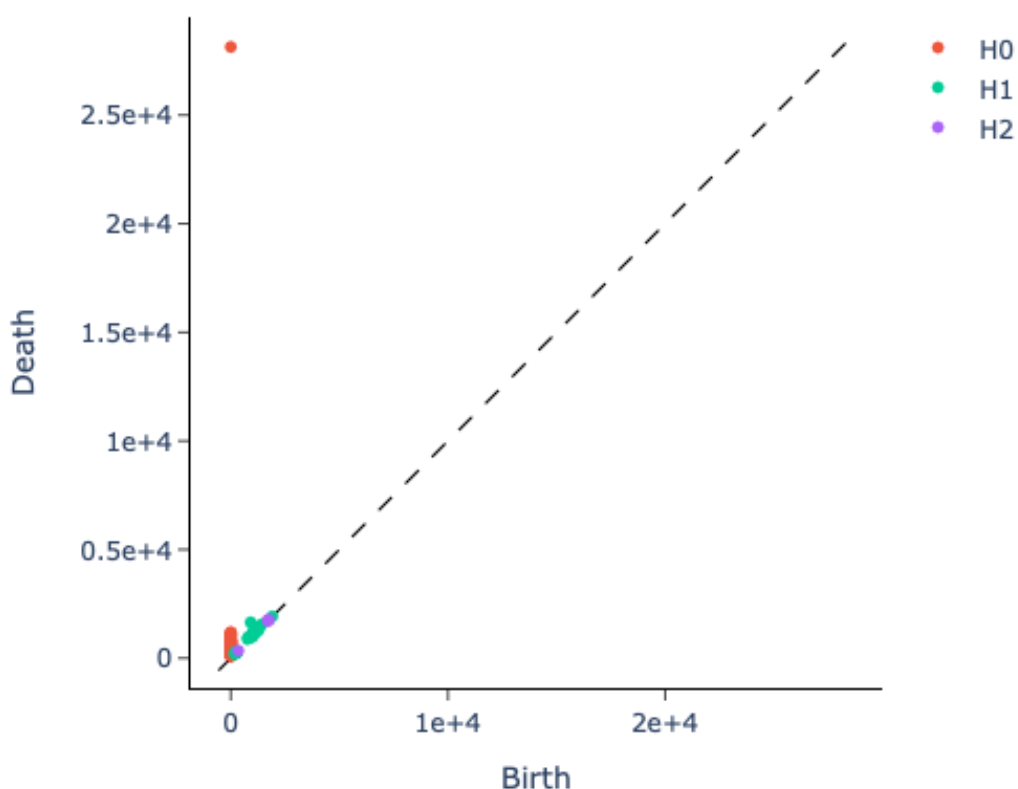


Рисунок 9 – Диаграмма устойчивости для параметра Когорта (пациент/здоровый донор).

Кривые Бетти показаны на рисунке 10. Параметр фильтрации (или значение радиуса роста, используемое для фильтрации) показан на оси x , а кривая Бетти – на оси y . Эта кривая подходит для деталей связности нулевого уровня, уровня один и уровня два, поэтому на рисунке вектор из трех чисел Бетти ($\beta_0, \beta_1, \beta_2$) для каждого параметра фильтрации.

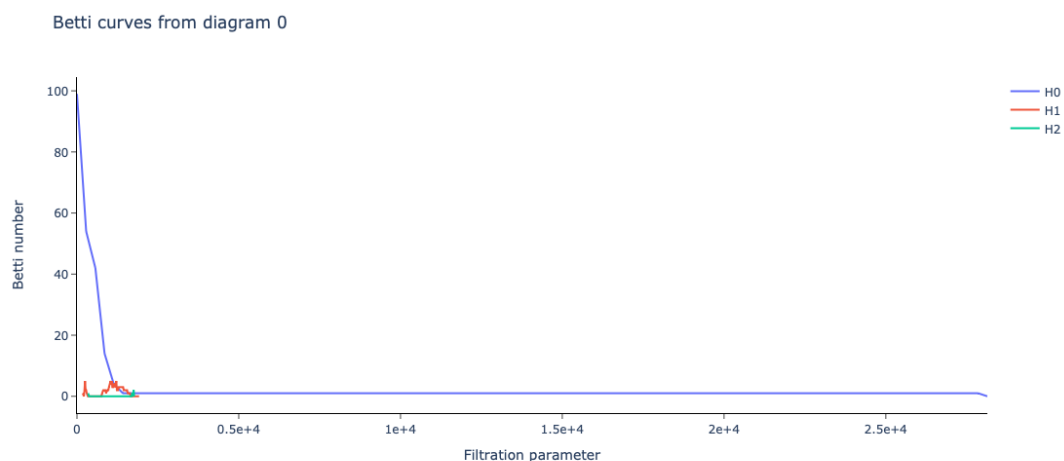


Рисунок 10 – Кривые Бетти для параметра Когорта (пациент/здоровый донор).

Точность классификации с использованием топологических данных и при использовании стандартного алгоритма `sklearn.ensemble` (RandomForest) одинакова и составляет 100%.

Визуализации данных с помощью алгоритма Mapper.

В результате применения метода Mapper был получен топологический граф (Рисунок 11), который отражает основные особенности данных. Размер узла показывает количество точек в узле, а цвет узла представляет собой среднее значение точек, входящих в него. На графе видно 6 связанных компонент и 1 несвязный узел. Есть главный компонент 1, который группирует большинство точек в соответствии с размерами узлов.

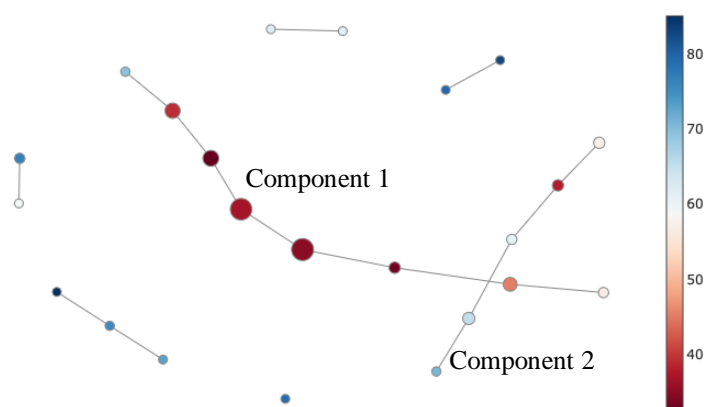


Рисунок 11 – Граф, построенный методом Mapper, для кластеризации данных по цитокиновому профилю пациентов с CRS и здоровых доноров.

Снижение размерности данных с учетом топологии.

В результате применения метода RTD-AE было измерено различие в топологии между двумя облаками точек (цитокиновый профиль пациентов, у которых развился CRS после терапии, и цитокиновый профиль здоровых доноров). Был построен график R-Cross-Barcode (Рисунок 12), который измеряет различия в многомасштабной топологии двух облаков точек с односторонним соответствием между точками. На основе R-Cross-Barcode было определено значение RTD – коэффициент, измеряющий многомасштабное топологическое различие между двумя наборами данных ($RTD = 3.75$).

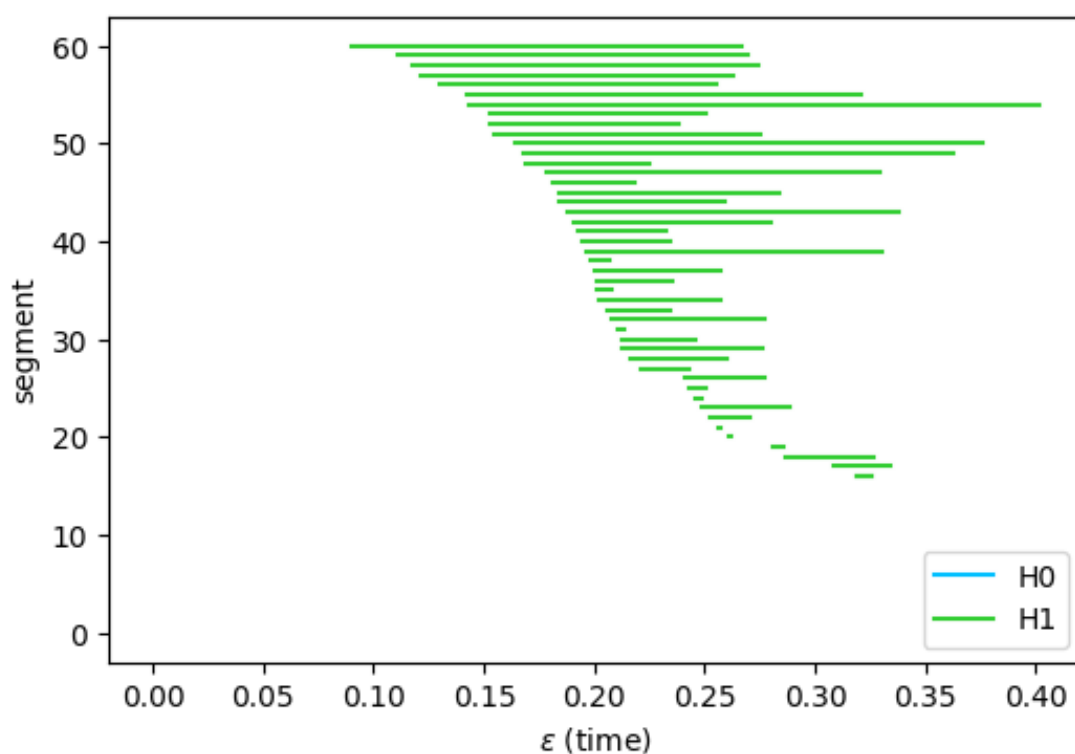


Рисунок 12 – График R-Cross-Barcode.

Заключение

В результате выполнения курсовой работы освоены алгоритмы топологического анализа данных: устойчивые гомологии, Mapper и RTD-AE. Построены диаграммы устойчивости, кривые Бетти, также были сопоставлены метрики модели классификации RandomForest с использованием топологических данных и стандартного алгоритма. Для отображения основных особенностей данных и для кластеризации данных был получен топологический граф с помощью алгоритма Mapper. В результате применения алгоритма RTD-AE построен график R-Cross-Barcode, который измеряет различия топологии двух облаков точек, а также определено значение RTD.

При исследовании доз CAR T-клеток, которые получили пациенты с гематологическими опухолями, топологические структуры не были обнаружены, однако при сравнении метрик модели классификации было показано, что точность классификации с использованием топологических данных составляет 57.1%, а при использовании стандартного алгоритма точность падает до 42.9%. При исследовании цитокинового профиля пациентов с CRS и здоровых доноров было обнаружено, что данные классифицируются на группы с точностью 100% даже стандартными методами.

Список источников

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021 May;71(3):209-249. doi: 10.3322/caac.21660. Epub 2021 Feb 4. PMID: 33538338.
2. Chhikara, B. S., & Parang, K. (2022). Global Cancer Statistics 2022: the trends projection analysis. *Chemical Biology Letters*, 10(1), 451.
3. Schirmacher V. From chemotherapy to biological therapy: A review of novel concepts to reduce the side effects of systemic cancer treatment (Review). *Int J Oncol.* 2019 Feb;54(2):407-419. doi: 10.3892/ijo.2018.4661. Epub 2018 Dec 10. PMID: 30570109; PMCID: PMC6317661.
4. Nicholson LB. The immune system. *Essays Biochem.* 2016 Oct 31;60(3):275-301. doi: 10.1042/EBC20160017. PMID: 27784777; PMCID: PMC5091071.
5. Riley RS, June CH, Langer R, Mitchell MJ. Delivery technologies for cancer immunotherapy. *Nat Rev Drug Discov.* 2019 Mar;18(3):175-196. doi: 10.1038/s41573-018-0006-z. PMID: 30622344; PMCID: PMC6410566.
6. Abbott M, Ustoyev Y. Cancer and the Immune System: The History and Background of Immunotherapy. *Semin Oncol Nurs.* 2019 Oct;35(5):150923. doi: 10.1016/j.soncn.2019.08.002. Epub 2019 Sep 13. PMID: 31526550.
7. Yahya EB, Alqadhi AM. Recent trends in cancer therapy: A review on the current state of gene delivery. *Life Sci.* 2021 Mar 15;269:119087. doi: 10.1016/j.lfs.2021.119087. Epub 2021 Jan 19. PMID: 33476633.
8. Mohanty R, Chowdhury CR, Arega S, et al. CAR T cell therapy: A new era for cancer treatment (Review). *Oncol Rep.* 2019 Dec;42(6):2183-2195. doi: 10.3892/or.2019.7335. Epub 2019 Sep 24. PMID: 31578576.
9. Ahmad A. CAR-T Cell Therapy. *Int J Mol Sci.* 2020 Jun 17;21(12):4303. doi: 10.3390/ijms21124303. PMID: 32560285; PMCID: PMC7352955.
10. Sermer D, Brentjens R. CAR T-cell therapy: Full speed ahead. *Hematol Oncol.* 2019 Jun;37 Suppl 1:95-100. doi: 10.1002/hon.2591. PMID: 31187533.
11. <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/car-t-cell-therapy>
12. Sterner RC, Sterner RM. CAR-T cell therapy: current limitations and potential strategies. *Blood Cancer J.* 2021 Apr 6;11(4):69. doi: 10.1038/s41408-021-00459-7. PMID: 33824268; PMCID: PMC8024391.
13. Mehrabadi AZ, Ranjbar R, Farzanehpour M, et al. Therapeutic potential of CAR T cell in malignancies: A scoping review. *Biomed Pharmacother.* 2022 Feb;

- 146:112512. doi: 10.1016/j.biopha.2021.112512. Epub 2021 Dec 9. PMID: 34894519.
14. Nukala U, Rodriguez Messan M, Yogurtcu ON, Wang X, Yang H. A Systematic Review of the Efforts and Hindrances of Modeling and Simulation of CAR T-cell Therapy. *AAPS J.* 2021 Apr 9;23(3):52. doi: 10.1208/s12248-021-00579-9. PMID: 33835308.
 15. Chazal F, Michel B. An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists. *Front Artif Intell.* 2021 Sep 29; 4:667963. doi: 10.3389/frai.2021.667963. PMID: 34661095; PMCID: PMC8511823.
 16. Chazal F, Fasy BT, Lecci F, Michel B, Rinaldo A, Wasserman L (2014) Robust topological inference: distance to a measure and kernel distance. *arXiv:1412.7197*
 17. Boissonnat J-D, Maria C (2014) Computing persistent homology with various coefficient fields in a single pass. In: Schulz AS, Wagner D (eds) *Algorithms - ESA 2014. Lecture notes in computer science*, vol 8737. Springer, Berlin, pp 185-196.
 18. Amézquita EJ, Quigley MY, Ophelders T, Munch E, Chitwood DH. The shape of things to come: Topological data analysis and biology, from molecules to organisms. *Dev Dyn.* 2020 Jul;249(7):816-833. doi: 10.1002/dvdy.175. Epub 2020 Apr 13. PMID: 32246730; PMCID: PMC7383827.
 19. Otter N, Porter MA, Tillmann U, Grindrod P, Harrington HA. A roadmap for the computation of persistent homology. *EPJ Data Sci.* 2017;6(1):17. doi: 10.1140/epjds/s13688-017-0109-5. Epub 2017 Aug 9. PMID: 32025466; PMCID: PMC6979512.
 20. Skaf Y, Laubenbacher R. Topological data analysis in biomedicine: A review. *J Biomed Inform.* 2022 Jun;130:104082. doi: 10.1016/j.jbi.2022.104082. Epub 2022 May 1. PMID: 35508272.
 21. Bukkuri A, Andor N, Darcy IK. Applications of Topological Data Analysis in Oncology. *Front Artif Intell.* 2021 Apr 13;4:659037. doi: 10.3389/frai.2021.659037. PMID: 33928240; PMCID: PMC8076640.
 22. Zomorodian, A., & Carlsson, G. (2004). Computing persistent homology. *Discrete & Computational Geometry*, 33(2), 249–274. <http://dx.doi.org/10.1007/s00454-004-1146-y>
 23. Munch, E. (2017). A User's Guide to Topological Data Analysis. *Journal of Learning Analytics*, 4(2), 47–61. <https://doi.org/10.18608/jla.2017.42.6>
 24. Zomorodian A (2012) Topological data analysis. In: Zomorodian A (ed) *Advances in applied and computational topology. Proceedings of symposia in applied mathematics*, vol 70. Am. Math. Soc., Providence, pp 1-39

25. Adler RJ, Bobrowski O, Borman MS, Subag E, Weinberger S (2010) Persistent homology for random fields and complexes. In: *Borrowing strength: theory powering applications - a festschrift for Lawrence D. Brown*. IMS collections, vol 6. Institute of Mathematical Statistics, Beachwood, pp 124-143
26. Singh, G., Mémoli, F., & Carlsson, G. (2007). Topological methods for the analysis of high dimensional data sets and 3D object recognition. In *Eurographics Symposium on Point-Based Graphics*. Retrieved from <https://research.math.osu.edu/tgda/mapperPBG.pdf>.
27. Nicolau, M., Levine, A. J., & Carlsson, G. (2011). Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences*, 108(17), 7265–7270. <http://dx.doi.org/10.1073/pnas.1102826108/-/DCSupplemental>.
28. Li, L., Cheng, W.-Y., Glicksberg, B. S., Gottesman, O., Tamler, R., Chen, R., Bottinger, E.P., & Dudley, J. T. (2015). Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Science Translational Medicine*, 7(311), 311ra174–311ra174. <http://dx.doi.org/10.1126/scitranslmed.aaa9364>.
29. Torres, B. Y., Oliveira, J. H. M., Thomas Tate, A., Rath, P., Cumnock, K., & Schneider, D. S. (2016). Tracking resilience to infections by mapping disease space. *PLoS Biology*, 14(4), 1–19.
30. Nielson, J. L., Paquette, J., Liu, A. W., et al. (2015). Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. *Nature Communications*, 6. <http://dx.doi.org/10.1038/ncomms9581>
31. Yao, Y., Sun, J., Huang, X., et al. (2009). Topological methods for exploring low-density states in biomolecular folding pathways. *The Journal of Chemical Physics*, 130(14). <http://dx.doi.org/10.1063/1.3103496>
32. Brudno JN, Somerville RP, Shi V, et al. Allogeneic T Cells That Express an Anti-CD19 Chimeric Antigen Receptor Induce Remissions of B-Cell Malignancies That Progress After Allogeneic Hematopoietic Stem-Cell Transplantation Without Causing Graft-Versus-Host Disease. *J Clin Oncol*. 2016 Apr 1;34(10):1112-21. doi: 10.1200/JCO.2015.64.5929. Epub 2016 Jan 25. PMID: 26811520; PMCID: PMC4872017.
33. Kalos M, Levine BL, Porter DL, Katz S, Grupp SA, Bagg A, June CH. T cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia. *Sci Transl Med*. 2011 Aug 10;3(95):95ra73. doi: 10.1126/scitranslmed.3002842. PMID: 21832238; PMCID: PMC3393096.
34. Li S, Zhang J, Wang M, et al. Treatment of acute lymphoblastic leukaemia with the second generation of CD19 CAR-T containing either CD28 or 4-1BB. *Br J*

- Haematol. 2018 May;181(3):360-371. doi: 10.1111/bjh.15195. Epub 2018 Apr 10. PMID: 29637550.
35. Lee DW, Kochenderfer JN, Stetler-Stevenson M, et al. T cells expressing CD19 chimeric antigen receptors for acute lymphoblastic leukaemia in children and young adults: a phase 1 dose-escalation trial. *Lancet*. 2015 Feb 7;385(9967):517-528. doi: 10.1016/S0140-6736(14)61403-3. Epub 2014 Oct 13. PMID: 25319501; PMCID: PMC7065359.
 36. Hallek M, Cheson BD, Catovsky D, et al. International Workshop on Chronic Lymphocytic Leukemia. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood*. 2008 Jun 15;111(12):5446-56. doi: 10.1182/blood-2007-06-093906. Epub 2008 Jan 23. Erratum in: *Blood*. 2008 Dec 15;112(13):5259. PMID: 18216293; PMCID: PMC2972576.
 37. Teachey DT, Lacey SF, Shaw PA, et al. Identification of Predictive Biomarkers for Cytokine Release Syndrome after Chimeric Antigen Receptor T-cell Therapy for Acute Lymphoblastic Leukemia. *Cancer Discov*. 2016 Jun;6(6):664-79. doi: 10.1158/2159-8290.CD-16-0040. Epub 2016 Apr 13. PMID: 27076371; PMCID: PMC5448406.
 38. Porter DL, Hwang WT, Frey NV, et al. Chimeric antigen receptor T cells persist and induce sustained remissions in relapsed refractory chronic lymphocytic leukemia. *Sci Transl Med*. 2015 Sep 2;7(303):303ra139. doi: 10.1126/scitranslmed.aac5415. PMID: 26333935; PMCID: PMC5909068.
 39. giotto-tda: A Topological Data Analysis Toolkit for Machine Learning and Data Exploration, Tauzin et al, arXiv:2004.02551, 2020.
 40. <https://giotto-ai.github.io/gtda-docs/0.2.2/modules/homology.html>
 41. <https://giotto-ai.github.io/gtda-docs/0.2.2/modules/mapper.html>
 42. Trofimov, I., Cherniavskii, D., Tulchinskii, E., Balabin, N., Burnaev, E., & Barannikov, S. (2023). Learning topology-preserving data representations. arXiv preprint arXiv:2302.00136.
 43. McInnes et al., (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, 3(29), 861, <https://doi.org/10.21105/joss.00861>
 44. <https://giotto-ai.github.io/gtda-docs/0.2.2/modules/generated/homology/gtda.homology.VietorisRipsPersistence.html>
 45. <https://giotto-ai.github.io/gtda-docs/0.2.2/modules/pipeline.html>
 46. Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278–282).
 47. <https://giotto-ai.github.io/gtda-docs/0.2.2/modules/mapper.html#filters>

48. <https://giotto-ai.github.io/gtda-docs/0.2.2/modules/mapper.html#covers>
49. Ester, M., Kriegel, H.-P., Sander, J., Xu, X., & others. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd* (Vol. 96, pp. 226–231).
50. Peter J. Rousseeuw (1987). “Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis”. *Computational and Applied Mathematics* 20: 53-65.
51. https://giotto-ai.github.io/gtda-docs/0.2.1/modules/generated/mapper/visualization/gtda.mapper.visualization.plot_static_mapper_graph.html
52. M. Carrière, F. Chazal, M. Glisse, Y. Ike, H. Kannan and Y. Umeda, ‘Optimizing persistent homology based functions’, *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021, pp. 1294–1303.
53. Hofer, C., Kwitt, R., Niethammer, M., and Dixit, M. Connectivity-optimized representation learning via persistent homology. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2751–2760. PMLR, 2019.
54. A. Poulenard, P. Skraba and M. Ovsjanikov, ‘Topological Function Optimization for Continuous Shape Matching’ *Computer Graphics Forum* 37.5, 2018, pp. 13–25.
55. Rotte A, Frigault MJ, Ansari A, Gliner B, Heery C, Shah B. Dose-response correlation for CAR-T cells: a systematic review of clinical studies. *J Immunother Cancer*. 2022 Dec;10(12):e005678. doi: 10.1136/jitc-2022-005678. PMID: 36549782; PMCID: PMC9791395.
56. Neelapu SS, Locke FL, Bartlett NL, et al. Axicabtagene Ciloleucel CAR T-Cell Therapy in Refractory Large B-Cell Lymphoma. *N Engl J Med*. 2017 Dec 28;377(26):2531-2544. doi: 10.1056/NEJMoa1707447. Epub 2017 Dec 10. PMID: 29226797; PMCID: PMC5882485.
57. Munshi NC, Anderson LD Jr, Shah N, et al. Idecabtagene Vicleucel in Relapsed and Refractory Multiple Myeloma. *N Engl J Med*. 2021 Feb 25;384(8):705-716. doi: 10.1056/NEJMoa2024850. PMID: 33626253.
58. Abramson JS, Palomba ML, Gordon LI, et al. Lisocabtagene maraleucel for patients with relapsed or refractory large B-cell lymphomas (TRANSCEND NHL 001): a multicentre seamless design study. *Lancet*. 2020 Sep 19;396(10254):839-852. doi: 10.1016/S0140-6736(20)31366-0. Epub 2020 Sep 1. PMID: 32888407.