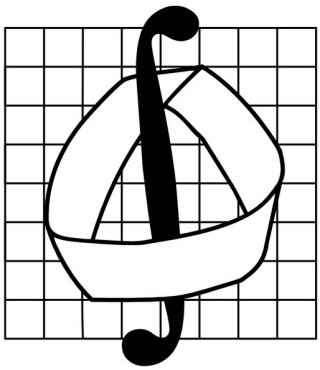
Московский государственный университет имени М. В. Ломоносова
механико-математический факультет
кафедра математической теории интелектуальных систем



Курсовая работа
студента 331 группы
Циберевой Елизаветы Кирилловны

# Построение тест-систем для распознавания глиобоастомы

# Construction of gene signatures for glioblastoma recognition

Научный руководитель:
доцент, к.ф.-м.н.
Галатенко Алексей Владимирович

Москва, 2024

# 1 Введение

Глиобластома (GBM) представляет собой одну из наиболее агрессивных форм злокачественных опухолей головного мозга, характеризующуюся быстрым ростом и высокой степенью инвазивности. Средняя продолжительность жизни пациентов с глиобластомой остается крайне низкой, несмотря на значительные достижения в области медицинской диагностики и лечения. В связи с этим актуальность разработки методов прогнозирования исхода заболевания, основанных на анализе экспрессии генов, неоспорима.

Целью данной работы является разработка тест-систем для распознавания глиобластомы на основе анализа экспрессии генов. Такая система позволит не только улучшить диагностику заболевания на ранних стадиях, но и предоставить важные данные для прогноза течения болезни и выбора оптимальной стратегии лечения.

На сегодняшний день проведено множество исследований, направленных на изучение молекулярных механизмов глиобластомы и поиск биомаркеров, позволяющих диагностировать это заболевание. Например, исследования, посвященные анализу мутаций и изменений экспрессии генов EGFR, PTEN, TP53, MDM2 и CDK4, показали их значимость в патогенезе глиобластомы. Тем не менее, разработка комплексных тест-систем, основанных на совокупности данных о генетических и молекулярных изменениях, остается очень востребованной задачей, так как несмотря на глубокие познания о молекулярном строении глиобластомы, смертность от нее все еще очень значительна. Об актуальности данной проблемы можно прочитать по ссылке: [1].

В рамках данной работы были проведены следующие исследования:

- Анализ данных экспрессии генов из базы данных TCGA для глиобластомы.
- Отбор наиболее значимых генов для диагностики глиобластомы.
- Разработка тест-системы на основе полученных данных, включающей алгоритмы для анализа и интерпретации результатов, а также построение графика выживаемости Каплана-Мейера.

Работа устроена следующим образом: в разделе 2 содержатся предварительные сведения, которые потребуются в ходе изучения данных и построения классификатора, содержащихся в разделе 3. Описасние результатов исследования содержится в разделе 4. Содержательное обсуждение результатов и направления дальнейших исследований содержатся в разделе 5. Заключительные замечания сделаны в разделе 6.

# 2 Предварительные сведения

**Определение 2.1.** Секвенирование РНК (RNA-seq) — это высокопроизводительный метод, используемый для анализа транскриптома в образцах клеток или тканей. Этот метод позволяет исследовать наличие и количество РНК в биологическом образце в

данный момент времени, что дает возможность понять экспрессию генов и ее изменения при различных физиологических или патологических состояниях.

**Определение 2.2.** Экспрессия генов — это процесс, посредством которого генетическая информация, закодированная в ДНК, используется для синтеза функциональных продуктов, таких как белки или РНК.

На протяжении всей курсовой работы будем использовать секвенирование РНК для анализа экспрессии генов, которая в свою очередь обеспечивает реализацию генетической информации в клетках.

**Определение 2.3.** Транскриптом — это полный набор РНК-молекул, которые экспрессируются в клетке или организме в определенный момент времени.

На протяжении всей курсовой работы будем использовать формализацию транскриптома в виде матрицы экспрессий генов.

Важной задачей является анализ данных секвенирования РНК (RNA-seq) с целью нахождения генов, которые экспрессируются по-разному в разных группах образцов. Небольшое количество реплик, дискретность, большой динамический диапазон и наличие выбросов требуют подходящего статистического подхода. Чтобы улучшить стабильность и интерпретируемость оценок, применим DESeq2 - это метод для дифференциального анализа данных подсчета, использующий усреднение для дисперсий и изменений кратности. Подробно изучить метод DESeq2 можно по ссылке: [2].

# 3 Материалы и методы

## 3.1 Входные данные

Будем брать данные об экспресси генов и выживаемости пациентов с сайта `https://xenabrowser.net/datapages/`. А именно, нас интетеруют данные TCGA для глиобластомы: `https://goo.su/AGLUMvt`.

Также данные о норме и отклонении от нее экспресии генов можно найти на сайте [3].

После скачивания по ссылке `https://goo.su/ZthTMcn` отнормированного для удобства массива экспрессии генов в формате .csv (AffyU133a) и данных о выживаемости по ссылке `https://goo.su/bQMe3b` в формате .txt (Curated survival data) приступим к их обработке.

## 3.2 Отбор генов для построения выборки

Теперь среди большого количества генов в массиве нужно выбрать именно те, которые влияют на глиобластому, для составления выборки, подходящей для прогнозирования смерти человека.

Интересными для нашего исследования являются следующие гены:

- EGFR (Epidermal Growth Factor Receptor)

EGFR часто мутирует или амплифицируется в клетках глиобластомы, что приводит к неконтролируемому росту клеток.

- PTEN (Phosphatase and Tensin Homolog)

PTEN является опухолевым супрессором, и его потеря или мутация часто наблюдаются в глиобластоме. PTEN регулирует сигнальный путь PI3K/AKT, и его потеря приводит к повышенной клеточной пролиферации и выживанию.

- TP53 (Tumor Protein p53)

TP53 — это ключевой опухолевый супрессор, который регулирует клеточный цикл и апоптоз. Мутации в TP53 часто встречаются в глиобластоме и приводят к потере контроля над клеточным циклом и предотвращению апоптоза.

- MDM2 (Mouse Double Minute 2 Homolog)

MDM2 является негативным регулятором TP53 и его амплификация или повышенная экспрессия могут привести к инактивации TP53. Это способствует выживанию опухолевых клеток и их устойчивости к апоптозу.

- CDK4 (Cyclin-Dependent Kinase 4)

CDK4 играет важную роль в регуляции клеточного цикла, особенно в переходе от фазы G1 к фазе S. Амплификация или повышенная активность CDK4 способствует неконтролируемому клеточному делению.

Таким образом, все вышеперечисленные гены участвуют в ключевых сигнальных путях, которые регулируют клеточную пролиферацию, апоптоз и выживание. Их мутации или нарушения экспрессии играют важную роль в развитии и прогрессировании глиобластомы, что делает их важными биомаркерами для прогнозирования течения заболевания.

## 3.3 Экспрессия генов

Средний уровень экспрессии **EGFR** в нормальных тканях мозга составляет примерно 5.0-6.0 в единицах log2(affy RMA). В исследованиях рака часто упоминается, что экспрессия EGFR в нормальных тканях значительно ниже по сравнению с опухолевыми клетками.

**PTEN** обычно экспрессируется на уровне 6.5-7.5 в нормальных тканях. Этот ген играет важную роль как опухолевый супрессор и часто имеет стабильный уровень экспрессии в здоровых клетках. Экспрессия PTEN в глиобластоме значительно снижена или отсутствует, часто уровень составляет около 3.0-4.0. Потеря PTEN связана с повышенной клеточной пролиферацией и снижением апоптоза.

**TP53** экспрессируется на уровне около 4.5-5.5 в нормальных условиях. TP53 сохраняет базовую экспрессию, чтобы обеспечивать клеточный цикл и апоптоз в ответ на повреждение ДНК. В глиобластоме экспрессия TP53 может быть изменена, однако точные значения могут варьироваться.

В нормальных тканях уровень экспрессии **MDM2** составляет примерно 4.0-5.0. Повышенная экспрессия MDM2 подавляет активность TP53, что способствует выживанию опухолевых клеток.

**CDK4** экспрессируется на уровне около 5.5-6.5 в нормальных тканях. Экспрессия CDK4 в глиобластоме значительно выше, чем в нормальных тканях, и составляет около 8.0-9.0. Это связано с усиленной регуляцией клеточного цикла и пролиферацией опухолевых клеток.

Для наглядности средние уровни экспрессии генов здорового человека и человека, имеющего глиобластому, представлены в Таблице 1.

| **Ген** | **Норма экспрессии** | **Экспрессия при глиобластоме** |
|---|---|---|
| EGFR | 5.0-6.0 | 9.5-10.5 |
| PTEN | 6.5-7.5 | 3.0-4.0 |
| TP53 | 4.5-5.5 | 4.0-5.0 |
| MDM2 | 4.0-5.0 | 6.0-7.0 |
| CDK4 | 5.5-6.5 | 8.0-9.0 |

Таблица 1: Сравнение нормальной экспрессии генов и экспрессии генов при глиобластоме в единицах log2(affy RMA)

## 3.4 Построение выборки

В пункте 3.2 мы объяснили выбор пяти генов для дальнейшего исследования. Отфильтруем наш массив и помещаем необходимые гены в отдельный.

Также для построения классификатора нам понадобится таблица Curated survival data, в которой содержится информация об общей выживаемости(OS) и времени жизни с момента постановки диагноза(OS.time).

Проверим, совпадают ли пациенты в новом массиве экспрессии генов и данной таблице с помощью set.intersection. Оказалось, что совпадения есть у 518 из 539 пациентов. Далее проводим фильтрацию данных по общим пациентам.

Для корректного объединения двух массивов необходимо транспонировать один из них. Проделаем это с массивом экспресии генов.

Таким образом, объединив массив экспрессии генов и таблицу с данными выживаемости, а также удалив столбцы, не нужные для исследования, получим массив, который будем использовать для построения модели, предсказывающей смерть пациента.

## 3.5 Построение классификатора

**Логистическая регрессия** — это статистический метод для анализа набора данных, в котором одна или несколько независимых переменных определяют исход. Исход, или зависимая переменная, кодируется бинарно (0 или 1).

Мы выбрали логистическую регрессию для предсказания выживаемости пациентов на основе экспрессии генов по следующим причинам:

- Логистическая регрессия используется для моделирования зависимой переменной, которая является бинарной. В нашем случае зависимая переменная — это выживаемость пациента, которая кодируется как 1 (не выжил) или 0 (выжил).
- Логистическая регрессия относительно проста для реализации и интерпретации. Она хорошо работает с наборами данных, в которых есть несколько независимых переменных.
- Логистическая регрессия может быть использована в контексте анализа выживаемости для моделирования вероятности выживания пациента. Это полезно в медицинских исследованиях для оценки факторов риска и прогнозирования исходов лечения.
- Существует множество инструментов для оценки качества моделей логистической регрессии, таких как ROC-кривые и AUC (площадь под кривой), которые позволяют легко интерпретировать и сравнивать результаты.

**Алгоритм построения логистической регрессии:**

1. Из библиотек sklearn.model_selection и sklearn.linear_model загружаем только функции для построения тренировочной и тестовой выборки (train_test_split) и для построения логистической регрессии (LogisticRegression) соответственно. Из библиотеки sklearn.metrics загружаем функции, отвечающие за метрики качества (classification_report, accuracy_score).
2. Подготавливаем данные, а именно – отбираем только необходимые стлбцы для построения логистической регрессии (столбцы с генами и столбец OS, содержащий бинарные данные о выживаемости пациентов)
3. Разделяем данные на обучающую и тестовую выборки с помощью train_test_split для обучения и оценки модели.
4. Используем модель логистической регрессии для обучения на данных об экспрессии генов и информации о выживаемости пациентов с помощью logreg.fit.
5. Прогнозируем выживаемость пациентов на тестовых данных с помощью logreg.predict и проверяем качество модели с использованием различных метрик.

При оценке качества модели логистической регрессии используем следующие метрики: **точность (accuracy)** и **отчёт о классификации (classification report)**.

**Точность (accuracy)** — это доля правильно предсказанных наблюдений от общего числа наблюдений. Ее мы вычисляем с помощью функции accuracy_score.

Отчёт о классификации включает несколько метрик, таких как точность (precision), полнота (recall), f1-score и поддержка (support) для каждого класса. Рассмотрим каждую из них подробнее:

1. **Точность (precision)** — это доля правильно предсказанных положительных наблюдений от всех наблюдений, которые модель предсказала как положительные.
2. **Полнота (recall)** — это доля правильно предсказанных положительных наблюдений от всех наблюдений, которые на самом деле являются положительными.
3. **f1-score** — это средневзвешенная метрика точности и полноты, которая учитывает как ложноположительные, так и ложноотрицательные ошибки.
4. **Поддержка (support)** — это количество истинных наблюдений для каждого класса в наборе данных.

С помощью функции classification_report создаем отчёт о классификации, включающий precision, recall, f1-score и support для каждого класса.

**Алгоритм построения ROC-кривой:**

**ROC (Receiver Operating Characteristic) кривая** используется для оценки качества бинарного классификатора, такого как логистическая регрессия. ROC-кривая строится на основе истинных положительных и ложных положительных значений при различных порогах классификации. Основные аспекты ROC-кривой включают TPR (True Positive Rate) и FPR (False Positive Rate).

1. Из библиотеки sklearn.metrics загружаем функции roc_curve, roc_auc_score, отвечающие за построение ROC-кривой и подсчет площади (AUC) под ее графиком.
2. Получаем вероятности предсказания с помощью метода predict_proba, который возвращает вероятности принадлежности к каждому из классов для каждого наблюдения в тестовом наборе данных. Мы берем вероятности принадлежности к классу 1.
3. Переходим к построению ROC-кривой с помощью функции roc_curve, которая возвращает значения ложноположительной доли (False Positive Rate, FPR) и истинноположительной доли (True Positive Rate, TPR) для различных порогов классификации. Функция roc_auc_score вычисляет площадь под ROC-кривой (AUC).
4. С помощью функций из библиотеки matplotlib.pyplot строим график ROC-кривой.

Благодаря ROC-кривой мы можем визуально оценить качество построенной модели. Чем ближе кривая к левому верхнему углу, тем производительнее наша модель. Площадь под кривой (AUC) – это метрика качества модели. Чем ближе площадь к 1, тем лучше модель.

Результаты построения логистической регрессии и ROC-кривой будут описаны в разделе 4.

## 3.6 Кривые выживаемости Каплана-Мейера

**Кривые выживаемости Каплана-Мейера** являются важным инструментом в анализе времени до наступления события. Они используются для оценки функции выживаемости и отображения вероятности выживания по времени. Эти кривые особенно полезны в медицинских исследованиях для анализа данных о времени до наступления таких событий, как смерть или рецидив заболевания. Также кривые Каплана-Мейера позволяют визуально сравнивать выживаемость между различными группами субъектов, что будет проделано ниже.

**Построение кривых Каплана-Мейера:**

1. Из библиотеки lifelines загружаем функцию KaplanMeierFitter для построения кривых выживаемости.
2. Инициализируем объект KaplanMeierFitter и подгоняем модель для общей кривой выживаемости, используя время до наступления смерти (столбец OS.time) и результат события (столбец OS).
3. Строим общую кривую выживаемости с помощью функций из библиотеки matplotlib.pyplot.
4. Теперь для каждого отобранного гена посчитаем его среднюю экспрессию и разделим данные о каждом гене на две группы: в одной из них будет содержаться информация об экспрессии, которая выше средней, а во второй – гены с более низкой экспрессией.
5. Создаем отдельные объекты KaplanMeierFitter для групп с высокой и низкой экспрессией каждого гена.
6. Подгоняем модели выживаемости для каждой из групп.
7. Строим кривые выживаемости для каждой из групп. Для наглядности построим кривые Каплана-Мейера для каждого гена на отдельном графике, также на каждом из них построим общую кривую выживаемости для сравнения результатов, которые будут описаны в разделе 4.

Анализ кривых Каплана-Мейера помогает понять, как различные факторы (например, экспрессия генов) влияют на выживаемость пациентов, что может быть полезно для принятия клинических решений и разработки стратегий лечения.

# 4 Результаты

## 4.1 Результаты построения логистической регресии и ROC-кривой

Логистическая регрессия была обучена для предсказания выживаемости пациентов на основе экспрессии пяти ключевых генов: EGFR, PTEN, TP53, MDM2, CDK4.

В первой версии построения логистической регрессии и ROC-кривой мы получили следующие результаты:

1. После обучения модели и прогноза на тестовых данных были рассчитаны метрики качества, которые отображены на Рисунке 1.

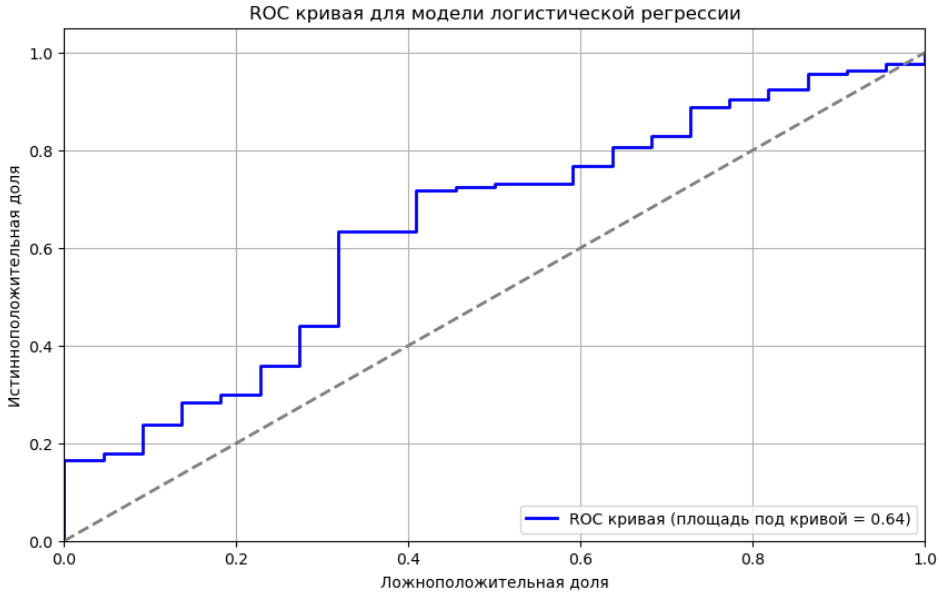
Рис. 1: Метрики качества работы первой логистической регрессии

Accuracy: 0.8589743589743589
Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 22 |
| 1 | 0.86 | 1.00 | 0.92 | 134 |
| accuracy | | | 0.86 | 156 |
| macro avg | 0.43 | 0.50 | 0.46 | 156 |
| weighted avg | 0.74 | 0.86 | 0.79 | 156 |

2. График ROC-кривой для данной модели представлен на Рисунке 2

Рис. 2: ROC-кривая для первой логистической регрессии



Заметим, что несмотря на высокую точность (accuracy = 0.86), в отчете о классификации наблюдаются проблемы с предсказыванием живых пациентов в связи с недостаточным набором данных о них. Также площадь под кривой AUC = 0.64 – не очень высокий показатель, что говорит о плохой производительности модели.

Попробуем это исправить балансировкой классов с помощью функции SMOTE и поиском оптимального порога классификации.

В этом случае мы получили следующие результаты:

1. После обучения модели и прогноза на тестовых данных были рассчитаны метрики качества, которые отображены на Рисунке 3.

Рис. 3: Метрики качества работы второй логистической регрессии

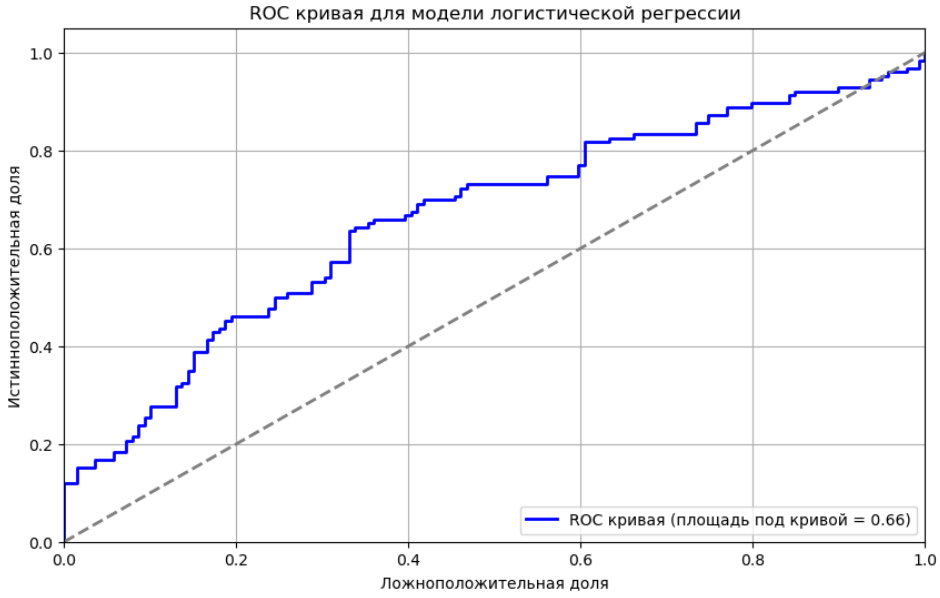
```
Accuracy: 0.6528301886792452
Classification Report:
              precision    recall  f1-score   support

           0       0.67      0.66      0.67       139
           1       0.63      0.64      0.64       126

    accuracy                           0.65       265
   macro avg       0.65      0.65      0.65       265
weighted avg       0.65      0.65      0.65       265
```

2. График ROC-кривой для данной модели представлен на Рисунке 4.

Рис. 4: ROC-кривая для второй логистической регрессии



Ключевые признаки, использованные в модели (экспрессия генов EGFR, PTEN, TP53, MDM2, CDK4), имеют важное биологическое значение, влияющее на выживаемость пациентов, о котором говорилось в разделе 3.2

Суть полученных результатов будет подробно раскрыта в разделе 5.

## 4.2 Результаты построения кривых выживаемости Каплана-Мейера

Ниже представлено пять графиков кривых выживаемости. Каждый из них содержит кривую Каплана-Мейера общей выживаемости, а также кривые выживаемости пациентов с высокой и низкой экспрессией генов EGFR, PTEN, TP53, MDM2, CDK4 соответственно.

Рис. 5: Кривые Каплана-Мейера для гена EGFR

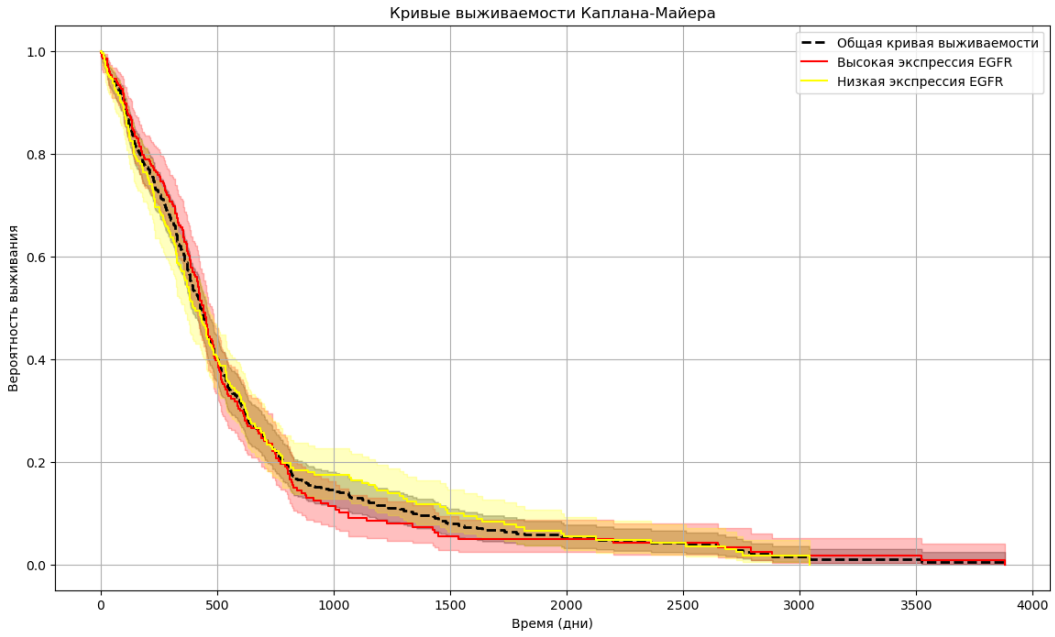


Рис. 6: Кривые Каплана-Мейера для гена PTEN

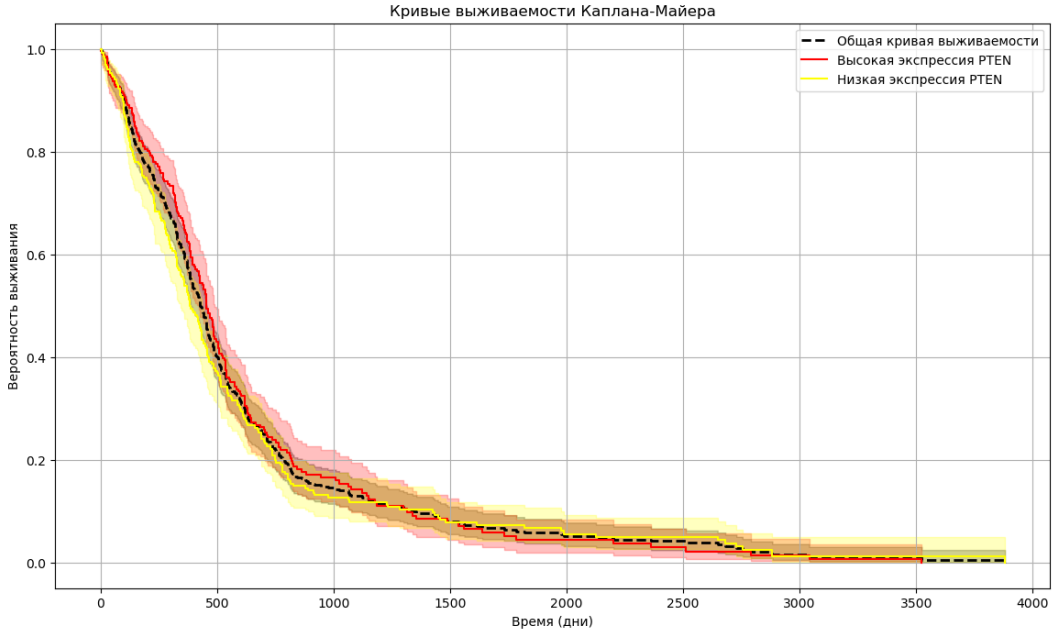



Рис. 7: Кривые Каплана-Мейера для гена TP53

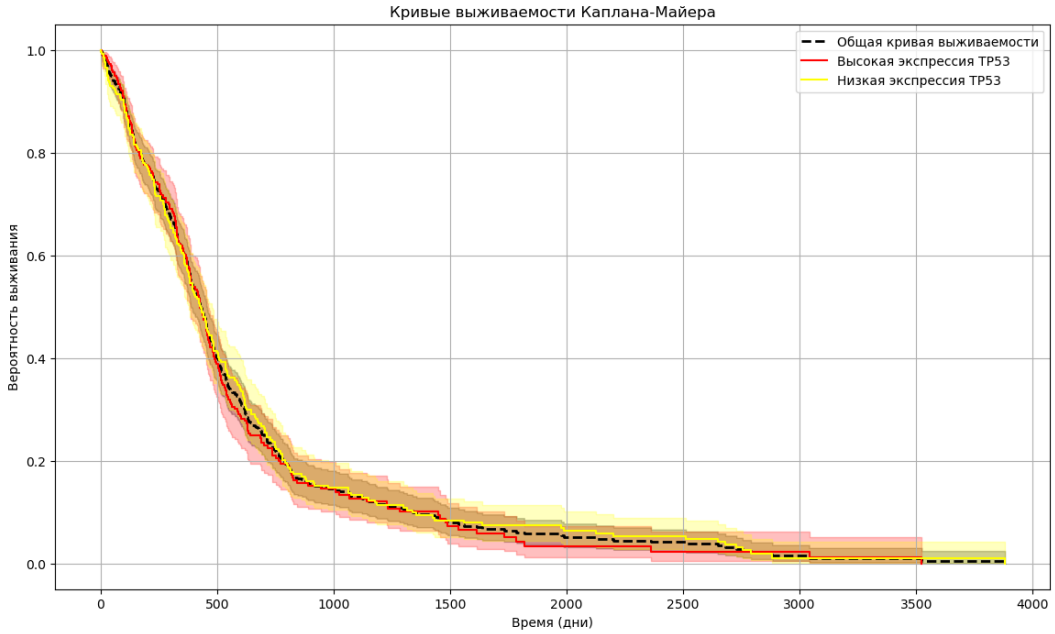


Рис. 8: Кривые Каплана-Мейера для гена MDM2

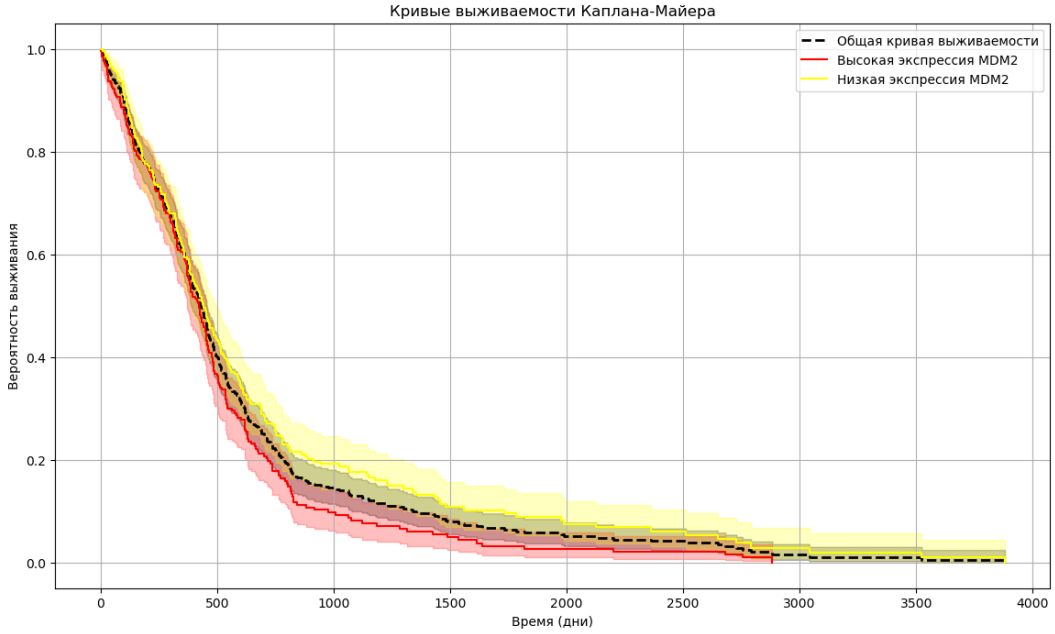
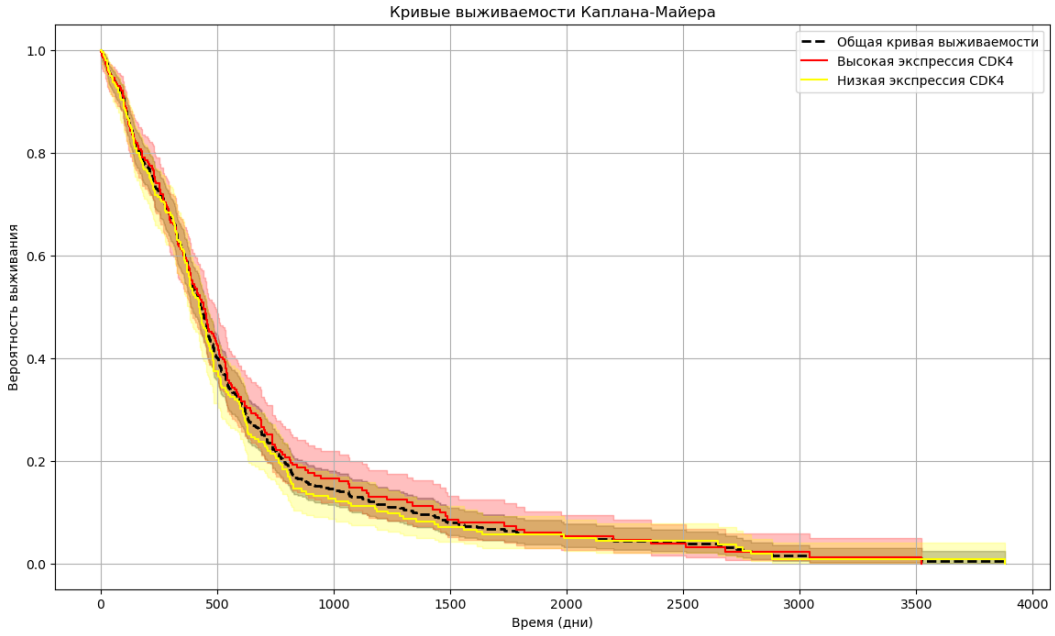



Рис. 9: Кривые Каплана-Мейера для гена CDK4




Интерпретация полученных результатов будет произведена в разделе 5.

С полным кодом можно ознакомиться на GitHub по ссылке:

https://github.com/Elizaveta47/COURSEWORK/tree/main

# 5 Обсуждение

С учетом результатов, полученных в разделе 4.1, можно сделать следующие выводы:

## 5.1 Анализ первой модели

- Модель очень хорошо классифицирует класс 1 (невыжившие) с высокой точностью (precision) 0.86 и полнотой (recall) 1.00. Это приводит к высокому значению f1-score 0.92.
- Модель не справляется с классификацией класса 0 (выжившие) — все метрики (precision, recall, f1-score) равны 0.00. Это указывает на сильный дисбаланс в данных и неспособность модели выявлять выживших.
- Площадь под графиком ROC-кривой равна 0.64, что говорит нам о почти случайности построенной модели. Для того, чтобы модель была идеальной, площадь под графиком ROC-кривой должна стремиться к 1.

В связи с этим были выполнены балансировка классов и поиск оптимального порога классификации с целью увеличить количество живых пациентов и улучшить метрики качества модели.

## 5.2 Анализ второй модели

- Модель после балансировки классов имеет более сбалансированные метрики для обоих классов.
- Для класса 0 (выжившие) и класса 1 (невыжившие), метрики precision, recall и F1-score примерно равны (около 0.65), что указывает на улучшенную способность модели различать оба класса по сравнению с первой моделью.
- Площадь под графиком ROC-кривой равна 0.66, что не сильно лучше предыдущего результата.

Первая модель показала высокую точность (0.86), но за счет очень плохой классификации выживших пациентов. Это объясняется сильным дисбалансом в данных, где большинство образцов принадлежит классу невыживших.

Вторая модель с применением SMOTE для балансировки классов показала более уравновешенные метрики (точность 0.65). Метрики precision, recall и f1-score для обоих классов стали более сбалансированными, что указывает на улучшенную способность модели классифицировать оба класса.

Несмотря на снижение общей точности, балансировка классов улучшила способность модели правильно классифицировать выживших пациентов, что является критически важным для медицинских исследований.

Таким образом, в дальнейшем планируется улучшение полученных результатов в следующих направлениях:

- Исследование других методов балансировки данных, таких как ADASYN или комбинации методов, чтобы улучшить производительность модели.
- Включение дополнительных генов и клинических данных. Это может помочь улучшить точность модели и понять влияние различных факторов на выживаемость пациентов.
- Применение более сложных моделей, таких как random forest или gradient boosting, что может улучшить производительность модели.

## 5.3 Анализ кривых выживаемости Каплана-Мейера

Кривые Каплана-Мейера являются мощным инструментом для анализа данных выживаемости и позволяют делать важные выводы о влиянии различных факторов на выживаемость пациентов. Они позволяют оценить вероятность выживаемости пациентов с течением времени. В данном случае мы построили общую кривую выживаемости, а также кривые для групп пациентов с высокой и низкой экспрессией генов EGFR, PTEN, TP53, MDM2, CDK4.

Изучив графики, полученные в разделе 4.2, можно сделать следующие выводы:

- На начальном этапе вероятность выживания высока, но с течением времени она значительно уменьшается. Уже спустя год выживаемость в среднем составляет 50%.
- В среднем, выживаемость с низкой и высокой экспрессией выбранных генов не сильно отличается от общей выживаемости. Наибольшие расхождения с общей выживаемостью имеют гены EGFR и MDM2.
- Вероятность выживания для пациентов с высокой экспрессией EGFR и MDM2 в среднем ниже по сравнению с пациентами с низкой экспрессией. Кривые падают быстрее, что указывает на более высокую смертность в этих группах.

Таким образом, нам не удалось выявить явные расхождения вероятности выживания в общем случае с пациентами с высокой или низкой экспрессией выбранных генов. Поэтому дальнейший ход исследования будет заключаться в следующем:

- Исследование влияния других генов на выживаемость пациентов.
- Изучение комбинаций генов и клинических признаков для улучшения прогностических моделей.

# 6 Заключение

В данной курсовой работе был произведен анализ выживаемости пациентов с глиобластомой на основе экспрессии ключевых генов, в том числе прогнозирование смерти. Основное внимание было уделено генам EGFR, PTEN, TP53, MDM2 и CDK4, которые играют значительную роль в патогенезе глиобластомы.

Для достижения поставленных целей были выполнены следующие этапы:

1. Сбор и предобработка данных.
2. Обучение моделей логистической регрессии: результаты показали, что первая модель обладает высокой точностью (0.86), но плохо справляется с классификацией выживших пациентов. Вторая модель, несмотря на снижение общей точности до 0.65, продемонстрировала более сбалансированные метрики для обоих классов, что свидетельствует о лучшей способности модели различать выживших и невыживших пациентов.
3. Построение ROC-кривых для обеих моделей, расчет площади под кривой (AUC). Первая модель показала AUC = 0.64, а вторая модель - AUC = 0.66, что подтверждает небольшое улучшение качества модели после балансировки классов.
4. Были построены кривые Каплана-Мейера для общей выживаемости пациентов, а также для групп с высокой и низкой экспрессией отобранных генов. Анализ показал, что высокая экспрессия EGFR и MDM2 ассоциируется с более низкой вероятностью выживания, что подчеркивает важность этих генов в прогнозировании исходов при глиобластоме.

Данная курсовая работа продемонстрировала важность использования методов машинного обучения и статистических методов для анализа медицинских данных. Полученные результаты подчеркивают значимость генов EGFR, PTEN, TP53, MDM2 и CDK4 в контексте выживаемости пациентов с глиобластомой и предлагают направления для дальнейших исследований и улучшений.

# Список литературы


[1] X. Z. Zhong Lan, Xin Li, Glioblastoma: An update in pathology, molecular mechanisms and biomarkers, Int. J. Mol. Sci. 25 (2024) 1–5.

[2] W. H. Simon Anders, Differential expression analysis for sequence count data, Genome Biolog 11 (2010).

[3] M. J.-A. María González-Tablas, Daniel Arandia, Heterogeneous egfr, cdk4, mdm4, and pdgfra gene expression profiles in primary gbm, Cancers 12 (2020) 1–3.