Всероссийский конкурс учебно-исследовательских работ старшеклассников по политехническим дисциплинам для учащихся 9-11 классов

**информатика и ИКТ**

***ROC-анализ критериев дифференциации раковых и нормальных клеток молочной железы человека***

|  |  |
| --- | --- |
|  | **Выполнила**:  Лихачева Дарья Федоровна 11 класс, МБОУ «Лицей №1»,  город Пермь  **Научный руководитель**:  канд. физ.-мат. наук, науч. сотр.  Никитюк Александр Сергеевич |

**Пермь. 2022.**

**Abstract**

Researching cancer cells is a relevant aim of oncology as cancer is the main cause of human death.

This research work is devoted to the problem of the effectiveness of early cancer diagnosis.

The aim of the research is to analyze the error curves of differentiation criteria of cancer cells and normal cells, which were received on the basis of the phase-modulation laser interference microscopy data processing.

To accomplish the task it was necessary to study the basic principles of the phase-modulation laser interference microscopy and medical applications of receiver operating characteristic curves (ROC-curves). To calculate error curves an algorithm implemented in programming language Python has been developed. Program execution results are presented in charts. Area under curve (AUC), which is a criterion of ROC-curve, was compared to the expert scale and during the research it was found out which differentiation criterion is more effective for early diagnosis of cancer.

In conclusion it should be noted that received results will allow oncologists to reduce diagnostic errors and get more accurate values during cell examinations.

Оглавление

[Введение 4](#_Toc94117795)

[Глава 1. Лазерная интерференционная микроскопия 7](#_Toc94117796)

[Глава 2. Медицинские приложения ROC-анализа 10](#_Toc94117797)

[Глава 3. Описание критериев дифференциации раковых и нормальных клеток молочной железы человека 14](#_Toc94117798)

[Глава 4. Реализация алгоритмов расчета ROC-кривой. Анализ полученных данных 16](#_Toc94117799)

[Заключение 23](#_Toc94117800)

[Список литературы 24](#_Toc94117801)

[Приложение А 25](#_Toc94117802)

[Приложение Б 27](#_Toc94117803)

[Приложение В 30](#_Toc94117804)

[Приложение Г 33](#_Toc94117805)

# Введение

Злокачественные новообразования являются одной из основных причин смерти в мире. Среди онкологических заболеваний рак молочной железы занимает лидирующую позицию, причем заболеваемость постоянно растет, и за последнее время увеличилась почти в два раза. По данным Всемирной Организации Здравоохранения, за 2020 год рак молочной железы был диагностирован у 2,3 млн. человек. Ввиду роста заболеваемости возникает вопрос об эффективных методах ранней диагностики рака. Разработанные на данный момент методы обладают высокой чувствительностью и низкой специфичностью, что влияет на достоверность результатов обследования и правильную постановку диагноза. Поэтому создание новых эффективных методов ранней диагностики рака является актуальной задачей в онкологии.

Эффективность метода ранней диагностики обычно оценивается по следующим критериям: чувствительность и специфичность исследуемых раковых и нормальных клеток. Чувствительность — это доля действительно болеющих людей, которые по результатам теста выявляются как больные. Специфичность — это доля тех, у которых тест отрицателен, среди всех людей, не имеющих болезни.

Анализ кривой ошибок (ROC-анализ) позволяет сопоставить чувствительность и специфичность диагностического теста и оценить качество классификации. На основе результатов проведенного анализа можно выбрать наиболее эффективный метод диагностики, который дает наименьшее количество ложных результатов. Новые критерии диагностики позволят сделать постановку диагноза более точной.

Анализ кривой ошибок успешно используется в медицине. Например, полиорганная недостаточность у новорождённых с экстренной хирургической абдоминальной патологией была исследована с применением ROC-кривых [1]. Таким образом, оценена вероятность летального исхода и получены значимые критерии, отражающие патологический процесс. В кардиологии также было проведено изучение эффективности диагностирования острого коронарного синдрома с помощью кривой ошибок [2]. Данное исследование основано на результате анализа белка «тропонин». Было получено, что анализ на тропонин исключает заболевание у 40% пациентов.

Лазерная интерференционная микроскопия является частью класса оптических методов, которая называется количественной фазовой микроскопией. Данный вид микроскопии определяется как перспективный подход в исследовании морфометрических показателей клеточных структур, а также их динамики. Современные методы математической обработки интерферограмм дают возможность получить изображение, содержащее количественную информацию о свойствах клетки. Исследования, проводимые с помощью лазерной интерференционной микроскопией, позволяют сравнивать динамику фазовых толщин клеток. Полученные данные отражают развитие онкологических заболеваний. Такой подход изучения может использоваться в качестве нового диагностического инструмента в клинической практике [3].

Цель работы – проанализировать кривые ошибок критериев дифференциации раковых и нормальных клеток молочной железы человека, полученные на основе обработки данных лазерной интерференционной микроскопии.

Задачи работы:

1. Сделать обзор литературы по темам «Дифференциация раковых и нормальных клеток молочной железы человека методом лазерной интерференционной микроскопии» и «Медицинские приложения ROC-анализа».
2. Численно реализовать алгоритмы расчета ROC-кривой.
3. Осуществить поиск и выбор критериев дифференциации раковых и нормальных клеток молочной железы человека, основанный на лазерной интерференционной микроскопии.
4. Рассчитать ROC-кривые критериев дифференциации раковых и нормальных клеток молочной железы человека.
5. Выполнить анализ полученных результатов.

Научная новизна работы заключается в исследовании оригинальных критериев раковых и нормальных клеток при реализации оценки эффективности методов диагностики. Эти критерии были получены с помощью лазерного модуляционного интерференционного микроскопа.

# Глава 1. Лазерная интерференционная микроскопия

Лазерная интерференционная микроскопия относится к одному из классов количественной фазовой микроскопии. Принцип этой микроскопии заключается в математической обработке интерферограмм. Полученное таким образом изображение предоставляет количественную информацию об оптических, геометрических и динамических свойствах клетки.

Оптическая схема лазерного микроскопа МИМ-340 (рисунок 1) представлена на рисунке 2. В таблице 1 указаны технические характеристики микроскопа [4]. Измерение оптической толщины клеточного материала осуществляется с помощью луча красного лазера, который делится на два при прохождении через светоделитель. Первый луч фокусируется на исследуемом объекте, размещенном на предметном стекле, а второй – на зеркальной поверхности. После отражения оба луча попадают в камеру измерительного канала, где возникает интерференционное изображение.

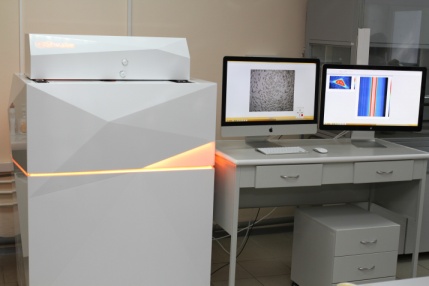


Рисунок 1 – Лазерный микроскоп МИМ-340

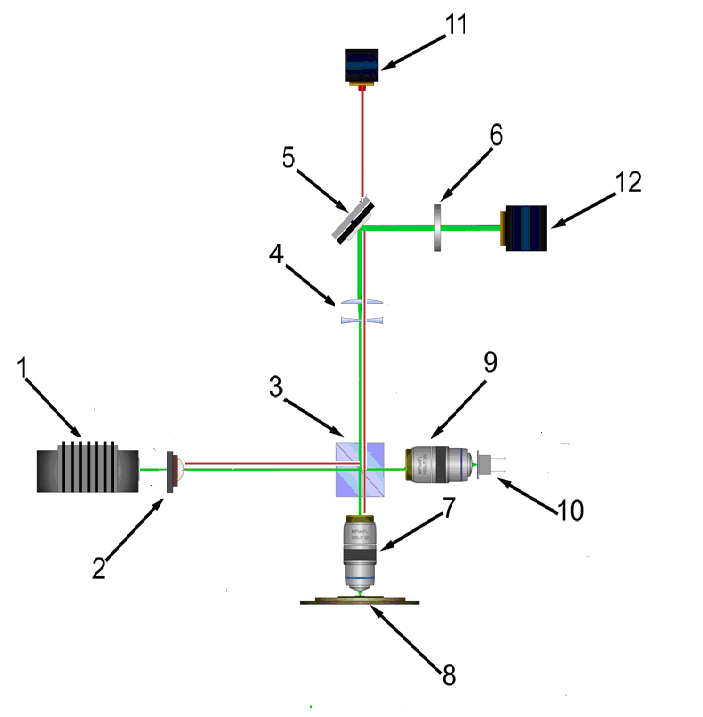


Рисунок 2 – Принципиальная оптическая схема лазерного микроскопа МИМ-340, где 1 – лазер, 2 – светодиод (белый свет), 3 – светоделитель, 4 – тубусная линза, 5 – поворотное зеркало, 6 – проектив, 7 – микрообъектив предметного плеча, 8 – исследуемый объект, 9 – микрообъектив опорного плеча, 10 – фазовый модулятор, 11 – цифровая камера навигационного канала, 12 – цифровая камера измерительного канала.

Таблица №1 – Технические характеристики лазерного микроскопа МИМ-340

|  |  |
| --- | --- |
| Характеристики микроскопа | |
| Оптическое увеличение | 10х |
| Поле зрения, мкм | 51,4х32,8 |
| Разрешение по вертикали, нм | 0,3 |
| Разрешение в плоскости XY, нм | 10-100 |
| Размер кадра, пикс. | 752х480 |
| Скорость съемки, кадр/сек | 1,3-100 |
| Задержка между кадрами, мс | 12 |
| Источник света | Лазер 655 нм |
| Характеристики предметного стола | |
| Длина хода (X - Y - Z), мм | 300х300х100 |
| Результирующая точность, нм | 150 |
| Разрешение обратной связи, нм | 100 |
| Неплоскостность хода, нм | 80 |
| Непрямолинейность хода, нм | 40 |
| Скорость перемещения, мм/с | До 100 |
| Грузоподъемность, кг | 100 |
| Вес, кг | 700 |

С помощью лазерного интерференционного микроскопа измеряются флуктуации оптической толщины клеток путем анализа разностного кадра, полученного при вычитании первого фазового изображения из второго изображения, сделанного через 60 секунд.

# Глава 2. Медицинские приложения ROC-анализа

ROC-анализ (Receiver Operator Characteristic) – анализ классификаций с применением ROC-кривых [5]. При оценке качества бинарной классификации применяется соотношение между верно классифицированными положительными примерами и неверно классифицированными отрицательными примерами.

Чувствительность (истинно положительная пропорция) отражает долю положительных результатов, которые верно определены. Показывает вероятность того, что больной пациент будет классифицирован именно как больной.

Специфичность (истинно отрицательная пропорция) показывает долю отрицательных результатов, которые правильно идентифицированы, то есть вероятность того, что здоровые пациенты будут классифицированы именно как не больные.

Задача классификации состоит в том, чтобы относить ранее неизвестные объекты к тому или иному классу: заболел пациент (положительный результат) или нет (отрицательный результат). Тогда в результате классификации может наблюдаться четыре различных ситуации [6]:

* Истинно положительный результат (true-positive, TP) — пациент болен, диагноз положительный;
* Ложно положительный результат (false-positive, FP) — пациент здоров, диагноз положительный (ошибка II рода – ложное обнаружение);
* Истинно отрицательный результат (true-negative, TN) — пациент здоров, диагноз отрицательный;
* Ложно отрицательный результат (false-negative, FN) — пациент болен, диагноз отрицательный (ошибка I рода – ложный пропуск).

Четыре возможных исхода могут быть оформлены в виде [таблицы сопряжённости](https://ru.wikipedia.org/wiki/%D0%A2%D0%B0%D0%B1%D0%BB%D0%B8%D1%86%D0%B0_%D1%81%D0%BE%D0%BF%D1%80%D1%8F%D0%B6%D1%91%D0%BD%D0%BD%D0%BE%D1%81%D1%82%D0%B8) (см. таблицу 2)

Таблица №2 – Таблица сопряжённости

|  |  |  |
| --- | --- | --- |
| Модель | Фактически положительно | Фактически отрицательно |
| Положительно | TP | FP |
| Отрицательно | FN | TN |

Для расчета чувствительности алгоритма применяется следующее соотношение:

.

Специфичность алгоритма определяется таким образом:

. (1.2)

При анализе используют две величины:

1. Доля истинно положительных классификаций (True Positives Rate):

*TP R = Sen;* (1.3)

1. Доля ложно положительных классификаций (False Positives Rate):

*FPR = 1 – Sp.* (1.4)

ROC-кривая получается следующим образом: для каждого значения порога отсечения рассчитываются значения чувствительности Sen и специфичности Sp. Затем строится график зависимости: по оси Y откладывается TPR, по оси X — FPR.

Количественную интерпретацию ROC даёт показатель *AUC* (area under ROC curve) — площадь, ограниченная ROC-кривой и осью доли ложных положительных классификаций.

Рассмотрим пример расчета кривой ошибок на тестовых данных приведенных в таблице №3. Кривая ошибок рассчитывалась на основе алгоритма, реализованного на языке программирования Python (см. приложение А) в виде отдельной программы. Данная программа также позволяет получить показатель *AUC*.

Таблица №3 – Тестовые данные

|  |  |
| --- | --- |
| Оценка | Класс |
| 0,5 | 0 |
| 0,1 | 0 |
| 0,2 | 0 |
| 0,6 | 1 |
| 0,2 | 1 |
| 0,3 | 1 |
| 0,0 | 0 |

На рисунке 3 представлена кривая ошибок для тестовых данных, приведенных в таблице №3. По оси абсцисс отложена доля ложных положительных классификаций, по оси ординат – доля истинных положительных классификаций.

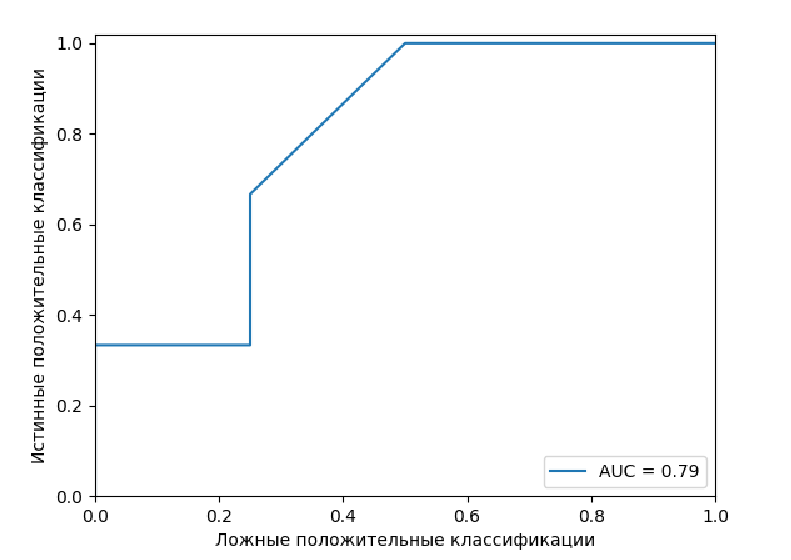


Рисунок 3 – Кривая ошибок для тестовых данных, приведенных в таблице №3

По результатам расчётов было получено, что площадь участка под ROC-кривой равна 0,79. Можно сделать вывод, что доля объектов, упорядоченных алгоритмом верно, от числа всех объектов тестовых данных составляет 0,79.

# Глава 3. Описание критериев дифференциации раковых и нормальных клеток молочной железы человека

Для анализа критериев дифференциации раковых и нормальных клеток молочной железы человека были использованы морфометрические показатели фазовых изображений раковых культуральных клеток (MCF7) и нормальных клеток (MCF10a) молочной железы человека. К морфометрическим показателям относятся следующие геометрические характеристики: высота, минимальный и максимальный диаметры, периметр, площадь и объем.

При вычислении высоты фазового изображения (*h*) рассчитывается разница между минимальным (*∆ϕmin*) и максимальным (*∆ϕmax*) значениями оптической толщины клетки внутри контура, границы которого определены согласно методу Кэнни

. (1.5)

Затем из все возможных диаметров клетки внутри выделенного контура определяются минимальный и максимальный (*dmin*, *dmax*) диаметры. Периметр, площадь и объем фазового изображения клетки рассчитывались согласно следующим соотношениям:

, (1.6)

, (1.7)

, (1.8)

где *P* – это периметр фазового изображения клетки, *l* – размер пикселя, *n*Ω – количество точек, принадлежащих множеству точек контура Ω, *nω* – количество точек внутри контура Ω*, V* – объем фазового изображения клетки, ∆*ϕ*(*x*,*y*) – значение фазовой толщины клетки в точке (*x*,*y*).

На основе результатов статистического и корреляционного анализа для корректного сравнения морфометрических параметров клеток MCF-7 и MCF-10A предложено использовать следующие безразмерные комплексы *h/M(r)* и *V/(0,5 λS0)*, где

*M(r) = (dmax + dmin) / 4* (1.9)

и

*S0 = 4/3πr2*. (1.10)

Эпителиальная клетка молочной железы человека считается нормальной, если значение *h/M(r)* не превышает 2,2 x 104 и значение *V/(0,5λS0)* не превышает 0,02, в противном случае исследуемая клетка – раковая [7].

# Глава 4. Реализация алгоритмов расчета ROC-кривой. Анализ полученных данных

Для выполнения построения ROC-кривой были использованы критерии раковых и нормальных клеток молочной железы человека (mcf7 и mcf10a соответственно). Предоставленные данные находились в CSV-файлах и были обработаны с помощью программы и функций для расчета и построения ROC-кривых. Морфометрические показатели фазовых изображений клеток, находящиеся в этих файлах, а именно высота, минимальный и максимальный диаметр, периметр, площадь и объем, были преобразованы по формуле *h / (( d\_max + d\_min) / 4)*. Полученные значения были классифицированы как «оценка», т.е. значения по оси ординат на графике зависимости TPR(FPR). «Класс» данных был определен по умолчанию: раковым клеткам присвоено значение – 1, здоровым клеткам – 0. В результате выполнения алгоритмов программы (Приложение Б) была построена ROC-кривая, предоставленная на графике (см. рисунок 4).

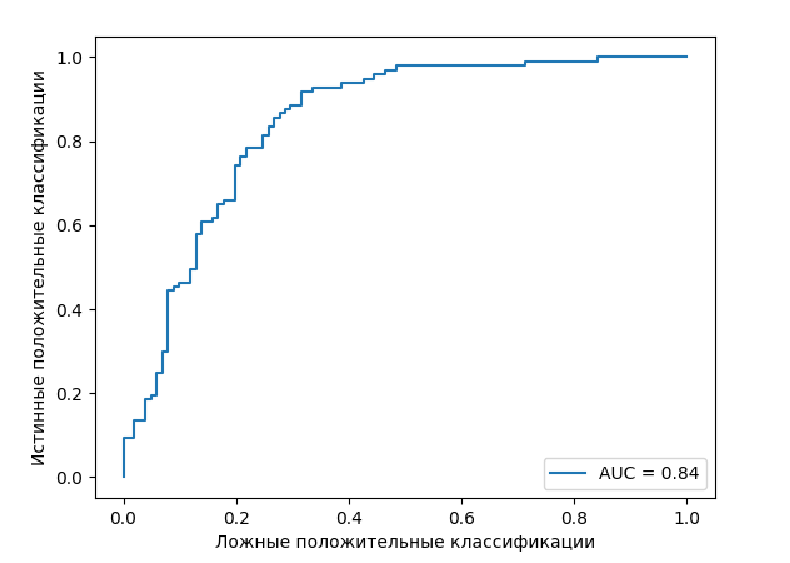


Рисунок 4 – Кривая ошибок, построенная для комплексов *h/M(r)*

На графике указано, что площадь под ROC-кривой равна 0,84. Существует экспертная шкала [6], позволяющая судить о качестве модели. Её критерии указаны в таблице 4.

Таблица №4 – Оценка качества модели.

|  |  |
| --- | --- |
| Интервал AUC | Качество модели |
| 0,5 – 0,6 | Неудовлетворительное |
| 0,6 – 0,7 | Среднее |
| 0,7 – 0,8 | Хорошее |
| 0,8 – 0,9 | Очень хорошее |
| 0,9 – 1,0 | Отличное |

При соотнесении полученного результата с величинами из приведенной таблицы видно, что данная модель исследования определена по этой классификации как «очень хорошая». Стоит отметить, что на практике невозможно добиться идеальной модели с результатом равным 1.

При построении кривой ошибок для второго безразмерного комплекса, который определяется по формуле *V/(0,5λS0)*, так же были использованы критерии раковых и нормальных клеток молочной железы человека (mcf7 и mcf10a соответственно). Были сформированы списки, содержащие величины объёма каждой клетки и её средней площади, которая зависит от величины радиуса клетки. Величина радиуса определяется как среднее значение максимального и минимального диаметра данной клетки. Раковым клеткам соответствует класс 1, а нормальным клеткам – 0. Величина λ была принята равной 650 нм. Результаты выполнения алгоритма (Приложение В) для построения кривой ошибок представлены на рисунке 5.

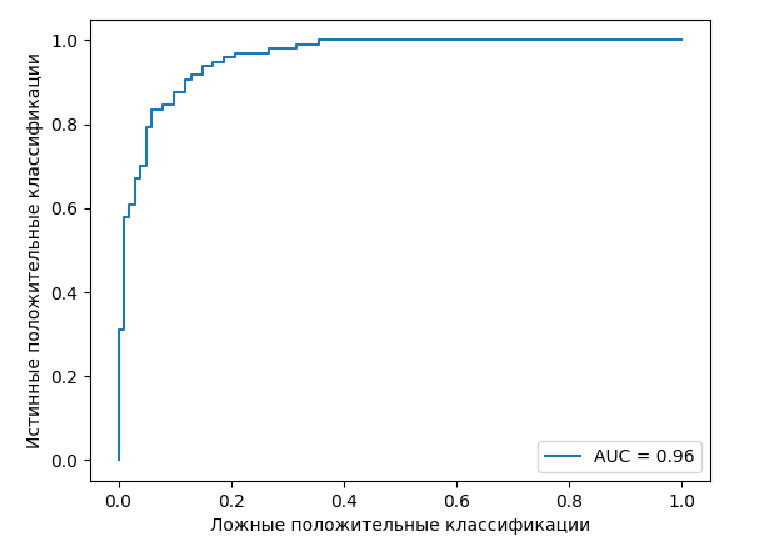


Рисунок 5 – Кривая ошибок, рассчитанная для комплексов *V/(0,5λS0)*

По построенной кривой ошибок для второго комплекса, определённого по формуле *V/(0,5λS0)*, видно, что площадь под графиком равна 0,96. Данный результат очень близок к 1. Если сопоставить полученную величину с экспертной шкалой, можно сделать вывод, что качество этой модели классифицируется как «отличное»

Аналогичным образом были построены кривые для каждого морфометрического показателя фазовых изображений раковых и нормальных клеток по отдельности. В приложении Г представлена реализация алгоритма. Значения критериев для анализа (высота, минимальный диаметр, максимальный диаметр, периметр, площадь, объём) были помещены в список «оценка» из файлов, содержащих результаты лазерной интерференционной микроскопии. «Класс» раковых клеток соответствует 1, нормальных клеток – 0. Результаты выполнения программы представлены на рисунках 6-11.

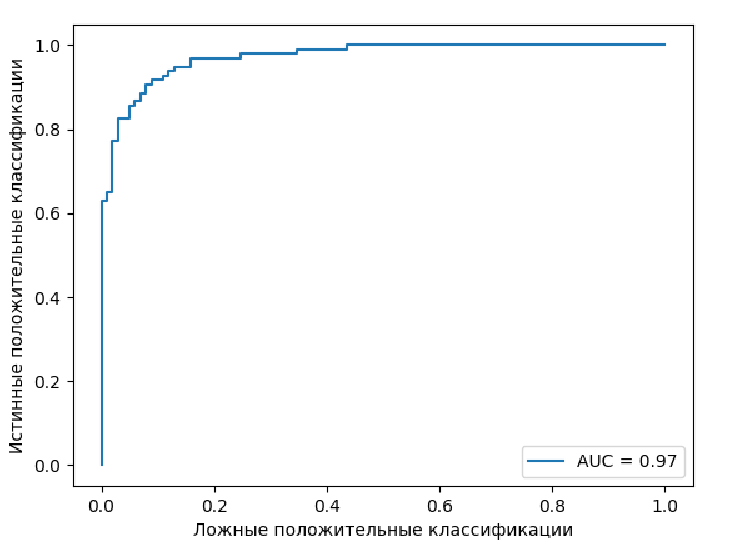


Рисунок 6 – Кривая ошибок для критерия «высота фазового изображения клетки»

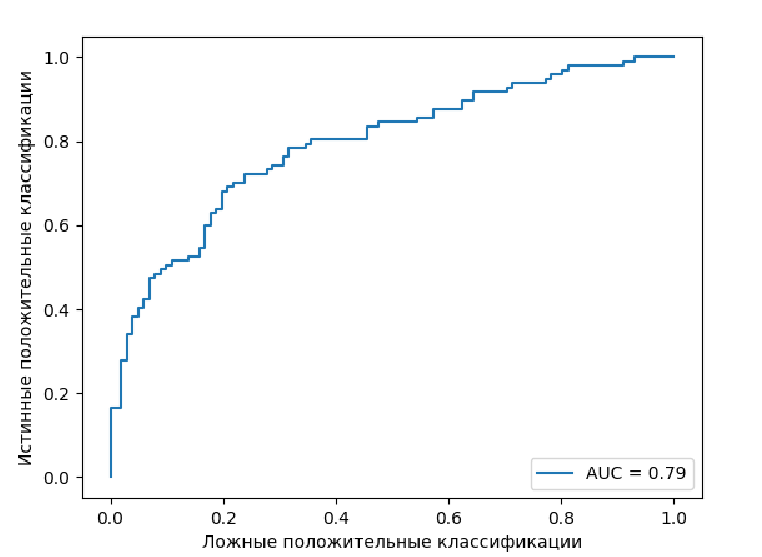


Рисунок 7 – Кривая ошибок для критерия «минимальный диаметр фазового изображения клетки»

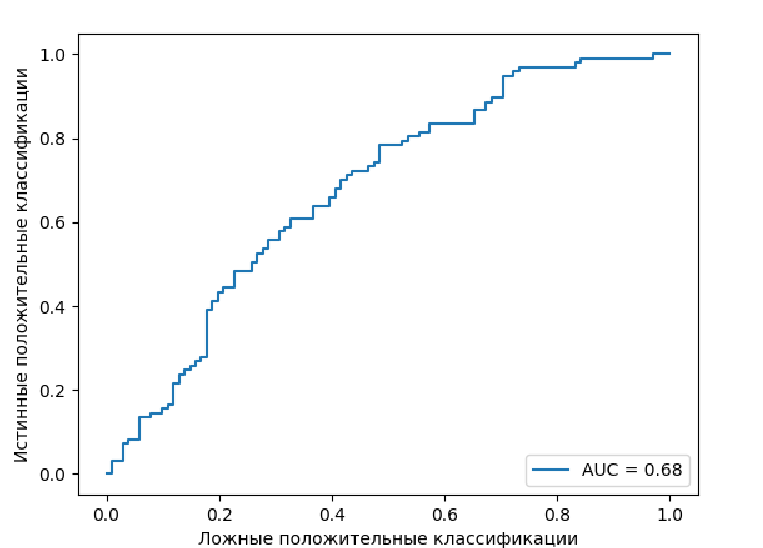


Рисунок 8 – Кривая ошибок для критерия «максимальный диаметр фазового изображения клетки»

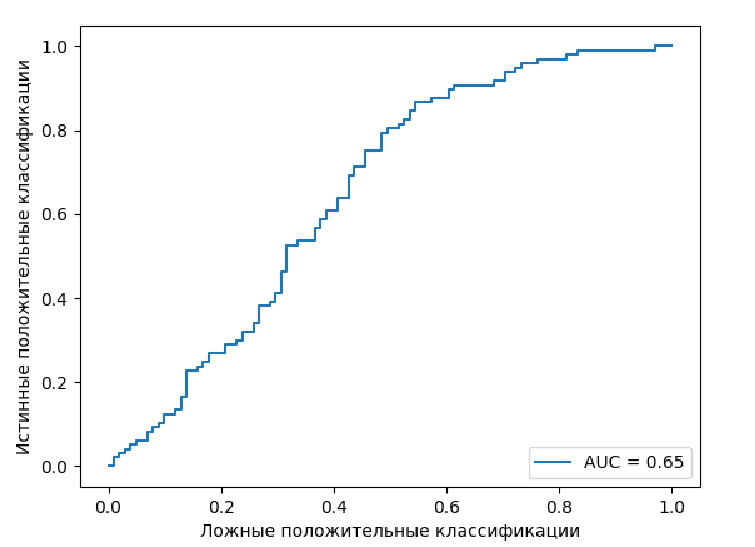


Рисунок 9 – Кривая ошибок для критерия «периметр фазового изображения клетки»

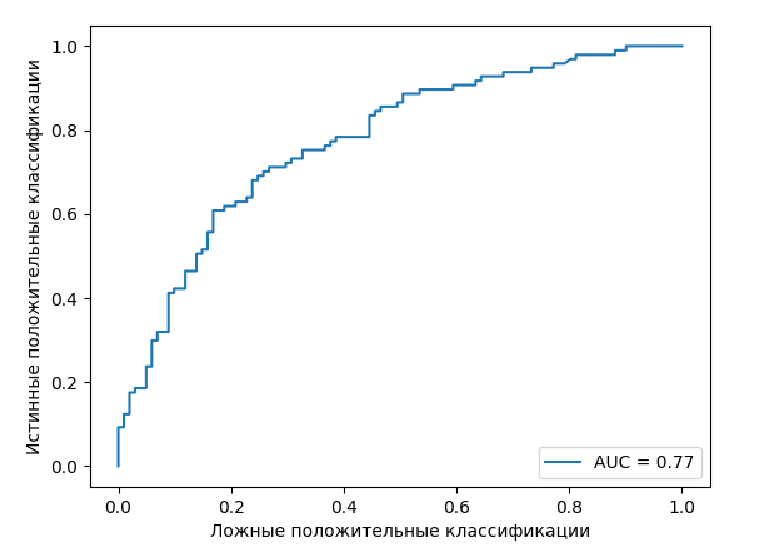


Рисунок 10 – Кривая ошибок для критерия «площадь фазового изображения клетки»

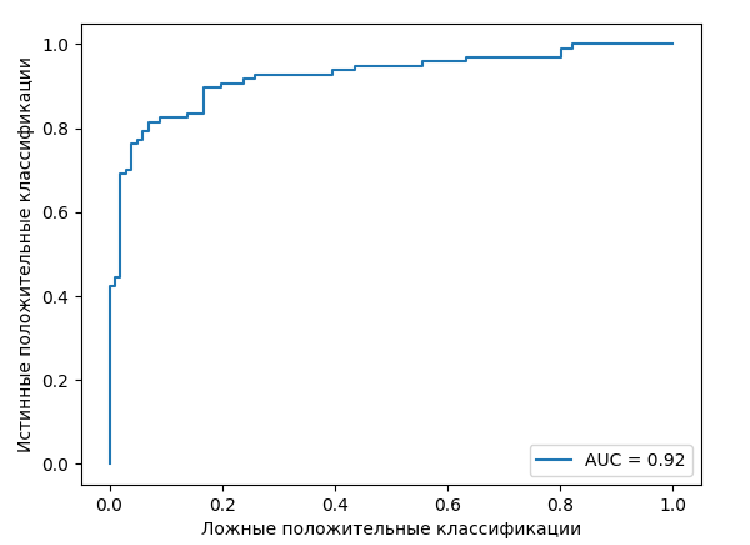


Рисунок 11 – Кривая ошибок для критерия «объём фазового изображения клетки»

По выполнении расчётов площади под графиком кривых для морфометрических показателей фазовых изображений клеток были получены следующие результаты: для критерия «высота» площадь равна 0,97, для «минимального диаметра» - 0,79, для «максимального диаметра» - 0,68, для «периметра» - 0,65, для «площади» и для «объёма» - 0,77 и 0,92 соответственно. В соответствии с экспертной шкалой, указанной в таблице №4, данные модели исследования можно разделить на несколько классификационных групп. Наименее эффективными оказались критерии периметра и максимального диаметра фазового изображения клетки. Значения их AUC находятся в интервале 0,6-0,7, и качество этих моделей определяется «как среднее». К моделям «хорошего» качества можно отнести критерии площади и минимального диаметра клеток. Результаты вычисления их площади под графиком кривых принадлежит интервалу от 0,7 до 0,8. Наиболее эффективный результат был обнаружен при построении ROC-кривой для критериев высоты и объёма. Эти модели исследования имеют площадь под графиком близкую к 1, что соответствует «отличному» качеству.

# Заключение

В ходе учебно-исследовательской работы была рассмотрена проблема эффективности критериев дифференциации раковых и нормальных клеток молочной железы человека для ранней диагностики. При решении этой задачи были изучены принципы построения кривых ошибок, на основании которых можно получить зависимости, позволяющие судить об эффективности выбранных критериев дифференциации клеток. При разработке алгоритмов для построения графиков использовались результаты исследования клеток двух культур: раковых и нормальных клеток молочной железы человека. Эти данные были получены на основе лазерной модуляционной интерференционной микроскопии, рассмотрение принципа работы которой также являлось одной из задач исследования. Построенные кривые ошибок отразили успешность дифференциации клеток по выбранным критериям. Анализ полученных зависимостей и численных расчётов позволили выбрать наиболее эффективную модель для дифференциации раковых и нормальных клеток молочной железы человека.

# Список литературы

1. Бударова К.В., Шмаков А.Н., Сирота С.И. Возможности ROC-анализа в интенсивной терапии новорожденных // Журнал научных статей «Здоровье и образование в XXI веке». – 2017. – № 6. – С. 88–91.
2. Richard Body, Edward Carlton, Matthew Sperrin, Philip S Lewis, Gillian Burrows, Simon Carley, Garry McDowell, Iain Buchan, Kim Greaves, Kevin Mackway-Jones Troponin-only Manchester Acute Coronary Syndromes (T-MACS) decision aid: single biomarker re-derivation and external validation in three cohorts // Emergency Medicine Journal. – 2017. – P. 349–356
3. Наймарк О.Б., Гришко В.В., Баяндин Ю.В., Никитюк А.С. Механобиологическое исследование динамики и морфологии клеточных структур методом лазерной микроскопии и приложения в онкологии // Вестник ПФИЦ. – 2020. – № 1. – С. 61–71.
4. Лазерный интерференционный микроскоп МИМ-340. URL: <http://permsc.ru/ru/main/activity/tsentry-kollektivnogo-polzovaniya/73-issledovaniya-materialov-i-veshchestva/329-lazernyj-interferentsionnyj-mikroskop-mim-340> (дата обращения: 17.06.2021).
5. ROC-кривая. URL: [https://ru.wikipedia.org/wiki/ROC-кривая](https://ru.wikipedia.org/wiki/ROC-%D0%BA%D1%80%D0%B8%D0%B2%D0%B0%D1%8F) (дата обращения: 15.06.2021)
6. Логическая регрессия и ROC-анализ – математический аппарат. URL: <https://loginom.ru/blog/logistic-regression-roc-auc> (дата обращения: 15.06.2021).
7. Никитюк А.С. Математическая модель нелинейной кинетики молекулы ДНК и ее применение для анализа клеточной динамики. // Российская национальная библиотека. – 2020. – С. 59–88.

# Приложение А

**Листинг программы для построения кривой ошибок по тестовым данным**

import numpy as np

import matplotlib.pyplot as plt

from sklearn.metrics import roc\_curve, plot\_roc\_curve, roc\_auc\_score, auc, RocCurveDisplay

import csv

def main():

# Загрузка данных для расчета ROC-кривой и характеристики AUC

# Список mark содержит значения "оценки"

# Список classification содержит значения "класса"

mark = []

classification = []

with open("test\_data.txt") as file:

reader = csv.reader(file)

for row in reader:

mark.append(float(row[0]))

classification.append(int(row[1]))

# Расчет ROC-кривой

x = np.array(classification)

y = np.array(mark)

fpr, tpr, thresholds = roc\_curve(x, y, pos\_label = 1)

# Расчет характеристики AUC

roc\_auc = auc(fpr, tpr)

# Визуализация ROC-кривой

viz = RocCurveDisplay(fpr = fpr, tpr = tpr, roc\_auc = roc\_auc)

viz.plot()

plt.xlim([0.0, 1.0])

plt.ylim([0.0, 1.02])

plt.xlabel('Ложные положительные классификации')

plt.ylabel('Истинные положительные классификации')

plt.savefig("viz.png")

if \_\_name\_\_ == "\_\_main\_\_":

main()

# Приложение Б

**Листинг программы для построения кривой ошибок комплексов *h/M(r)***

import numpy as np

import matplotlib.pyplot as plt

from sklearn.metrics import roc\_curve, plot\_roc\_curve, roc\_auc\_score, auc, RocCurveDisplay

import csv

import sys

def main():

# Проверка количества аргументов командной строки. Указание необходимой последовательности файлов при вводе

if len(sys.argv) != 3:

sys.exit("Usage: python roc.py morphometric\_parameters\_mcf7.txt morphometric\_parameters\_mcf10a.txt")

# Загрузка морфометрических показателей (высота, минимальный диаметр, максимальный диаметр) раковых клеток

h\_mcf7 = []

d\_min\_mcf7 = []

d\_max\_mcf7 = []

with open(sys.argv[1]) as file:

reader = csv.reader(file)

for row in reader:

h\_mcf7.append(float(row[0]))

d\_min\_mcf7.append(float(row[1]))

d\_max\_mcf7.append(float(row[2]))

# Загрузка морфометрических показателей (высота, минимальный диаметр, максимальный диаметр) нормальных клеток

h\_mcf10a = []

d\_min\_mcf10a = []

d\_max\_mcf10a = []

with open(sys.argv[2]) as file:

reader = csv.reader(file)

for row in reader:

h\_mcf10a.append(float(row[0]))

d\_min\_mcf10a.append(float(row[1]))

d\_max\_mcf10a.append(float(row[2]))

# Расчет значений по формуле для раковых клеток и определение класса 1 по умолчанию

mcf7 = []

cl\_mcf7 = []

for i in range(len(h\_mcf7)):

tmp = h\_mcf7[i] / ((d\_max\_mcf7[i] + d\_min\_mcf7[i]) / 4)

mcf7.append(float(tmp))

cl\_mcf7.append(1)

# Расчет значений по формуле для нормальных клеток и определение класса 0 по умолчанию

mcf10a = []

cl\_mcf10a = []

for i in range(len(h\_mcf10a)):

tmp = h\_mcf10a[i] / ((d\_max\_mcf10a[i] + d\_min\_mcf10a[i]) / 4)

mcf10a.append(float(tmp))

cl\_mcf10a.append(0)

# Расчет ROC-кривой

# Объединение списков

classification = cl\_mcf7 + cl\_mcf10a

mark = mcf7 + mcf10a

x = np.array(classification)

y = np.array(mark)

fpr, tpr, thresholds = roc\_curve(x, y, pos\_label = 1)

# Расчет характеристики AUC

roc\_auc = auc(fpr, tpr)

# Визуализация ROC-кривой

roc = RocCurveDisplay(fpr = fpr, tpr = tpr, roc\_auc = roc\_auc)

roc.plot()

plt.xlabel('Ложные положительные классификации')

plt.ylabel('Истинные положительные классификации')

plt.savefig("roc.png")

if \_\_name\_\_ == "\_\_main\_\_":

main()

# Приложение В

**Листинг программы для построения кривой ошибок комплексов *V/(0,5λS0)***

import numpy as np

import matplotlib.pyplot as plt

from sklearn.metrics import roc\_curve, plot\_roc\_curve, roc\_auc\_score, auc, confusion\_matrix, RocCurveDisplay

import csv

import sys

def main():

# Проверить количество аргументов командной строки

if len(sys.argv) != 3:

sys.exit("Usage: python roc2.py morphometric\_parameters\_mcf7.txt morphometric\_parameters\_mcf10a.txt")

# Загрузка морфометрических показателей раковых клеток

V\_mcf7 = []

d\_min\_mcf7 = []

d\_max\_mcf7 = []

with open(sys.argv[1]) as file:

reader = csv.reader(file)

for row in reader:

d\_min\_mcf7.append(float(row[1]))

d\_max\_mcf7.append(float(row[2]))

V\_mcf7.append(float(row[5]))

# Загрузка морфометрических показателей нормальных клеток

V\_mcf10a = []

d\_min\_mcf10a = []

d\_max\_mcf10a = []

with open(sys.argv[2]) as file:

reader = csv.reader(file)

for row in reader:

d\_min\_mcf10a.append(float(row[1]))

d\_max\_mcf10a.append(float(row[2]))

V\_mcf10a.append(float(row[5]))

# Расчет значения средней площади для раковых клеток

S\_mcf7 = []

for i in range(len(d\_min\_mcf7)):

tmp = 3.14 \* (((d\_min\_mcf7[i] + d\_max\_mcf7[i]) / 4) \*\* 2)

S\_mcf7.append(float(tmp))

# Расчет значения средней площади для нормальных клеток

S\_mcf10a = []

for i in range(len(d\_min\_mcf10a)):

tmp = 3.14 \* (((d\_min\_mcf10a[i] + d\_max\_mcf10a[i]) / 4) \*\* 2)

S\_mcf10a.append(float(tmp))

# Расчет значений по формуле для раковых клеток и определение класса

mcf7 = []

cl\_mcf7 = []

for i in range(len(S\_mcf7)):

tmp = V\_mcf7[i] / (0.5 \* 650 \* 10 \*\* (-9) \* S\_mcf7[i])

mcf7.append(float(tmp))

cl\_mcf7.append(1)

# Расчет значений по формуле для нормальных клеток и определение класса

mcf10a = []

cl\_mcf10a = []

for i in range(len(S\_mcf10a)):

tmp = V\_mcf10a[i] / (0.5 \* 650 \* 10 \*\* (-9) \* S\_mcf10a[i])

mcf10a.append(float(tmp))

cl\_mcf10a.append(0)

# Расчет ROC-кривой

# Объединение списков

classification = cl\_mcf7 + cl\_mcf10a

mark = mcf7 + mcf10a

x = np.array(classification)

y = np.array(mark)

fpr, tpr, thresholds = roc\_curve(x, y, pos\_label = 1)

# Расчет характеристики AUC

roc\_auc = auc(fpr, tpr)

# Визуализация ROC-кривой

roc = RocCurveDisplay(fpr = fpr, tpr = tpr, roc\_auc = roc\_auc)

roc.plot()

plt.xlabel('Ложные положительные классификации')

plt.ylabel('Истинные положительные классификации')

plt.savefig("roc2.png")

if \_\_name\_\_ == "\_\_main\_\_":

main()

# Приложение Г

**Листинг программы для построения кривой ошибок морфометрических показателей клетки**

import numpy as np

import matplotlib.pyplot as plt

from sklearn.metrics import roc\_curve, plot\_roc\_curve, roc\_auc\_score, auc, confusion\_matrix, RocCurveDisplay

import csv

import sys

def main():

# Проверить количество аргументов командной строки

if len(sys.argv) != 4:

sys.exit("Usage: python roc3.py morphometric\_parameters\_mcf7.txt morphometric\_parameters\_mcf10a.txt argument")

# Загрузка морфометрических показателей раковых и нормальных клеток

# h - высота, d\_min - минимальный диаметр, d\_max - максимальный диаметр, P - периметр, S - площадь, V - объем

mcf7 = []

mcf10a = []

if sys.argv[3] == "h":

with open(sys.argv[1]) as file:

reader = csv.reader(file)

for row in reader:

mcf7.append(float(row[0]))

with open(sys.argv[2]) as file:

reader = csv.reader(file)

for row in reader:

mcf10a.append(float(row[0]))

elif sys.argv[3] == "d\_min":

with open(sys.argv[1]) as file:

reader = csv.reader(file)

for row in reader:

mcf7.append(float(row[1]))

with open(sys.argv[2]) as file:

reader = csv.reader(file)

for row in reader:

mcf10a.append(float(row[1]))

elif sys.argv[3] == "d\_max":

with open(sys.argv[1]) as file:

reader = csv.reader(file)

for row in reader:

mcf7.append(float(row[2]))

with open(sys.argv[2]) as file:

reader = csv.reader(file)

for row in reader:

mcf10a.append(float(row[2]))

elif sys.argv[3] == "P":

with open(sys.argv[1]) as file:

reader = csv.reader(file)

for row in reader:

mcf7.append(float(row[3]))

with open(sys.argv[2]) as file:

reader = csv.reader(file)

for row in reader:

mcf10a.append(float(row[3]))

elif sys.argv[3] == "S":

with open(sys.argv[1]) as file:

reader = csv.reader(file)

for row in reader:

mcf7.append(float(row[4]))

with open(sys.argv[2]) as file:

reader = csv.reader(file)

for row in reader:

mcf10a.append(float(row[4]))

elif sys.argv[3] == "V":

with open(sys.argv[1]) as file:

reader = csv.reader(file)

for row in reader:

mcf7.append(float(row[5]))

with open(sys.argv[2]) as file:

reader = csv.reader(file)

for row in reader:

mcf10a.append(float(row[5]))

# Определение класса для раковых клеток

cl\_mcf7 = []

for i in range(len(mcf7)):

cl\_mcf7.append(1)

# Определение класса для нормальных клеток

cl\_mcf10a = []

for i in range(len(mcf10a)):

cl\_mcf10a.append(0)

# Расчет ROC-кривой

# Объединение списков

classification = cl\_mcf7 + cl\_mcf10a

mark = mcf7 + mcf10a

x = np.array(classification)

y = np.array(mark)

fpr, tpr, thresholds = roc\_curve(x, y, pos\_label = 1)

# Расчет характеристики AUC

roc\_auc = auc(fpr, tpr)

# Визуализация ROC-кривой

roc = RocCurveDisplay(fpr = fpr, tpr = tpr, roc\_auc = roc\_auc)

roc.plot()

plt.xlabel('Ложные положительные классификации')

plt.ylabel('Истинные положительные классификации')

plt.savefig("roc3.png")

if \_\_name\_\_ == "\_\_main\_\_":

main()