Отчет о проекте «Оценка эффективности диагностического теста»

**Введение в проблему и актуальность ее решения**

Существует множество ситуаций, подходы к которым будут различаться, соответственно необходимо уметь правильно оценивать эффективность метода диагностики, причем сделать это нужно на объективной основе.

Диагностическим тестом является процедура или инструмент, который помогает выявить людей, находящихся в различных состояниях. В медицинской практике бывают случаи, когда сложно поставить диагноз, имея только данные результатов анализов. Поэтому возникает вопрос отнести пациента в группу «здоровых» или в группу «больных». В таком случае необходим диагностический тест, который бы позволил определить пациентов в одну из групп. При этом тест должен быть эффективным. Оценка диагностической эффективности теста основана на взвешивании возможных диагностических ошибок. Это взвешивание можно провести с помощью операционных характеристик теста:

Специфичности – доля, здоровых людей, у которых получен отрицательный результат теста.

Чувствительности - доля больных, у которых получен положительный результат теста.

Прогностичности положительного результата – доля больных среди лиц с положительным результатом теста.

Прогностичности отрицательного результата – доля здоровых среди лиц с отрицательным результатом теста.

Все операционные характеристики связаны между собой.

Чувствительность и специфичность – вероятность в определенной группе (условная вероятность).

Прогностичность – вероятность состояния при известном результате исследования (апостериорная вероятность).

Показатели признака у больных и здоровых людей отличаются. Типичной является ситуация с частичным перекрытием между двумя выборками, что приводит к проблеме отнесения пациента в группу больных или здоровых. Поэтому возникает потребность в поиске оптимальной точки разделения, в которой число ошибок минимально, и оценке эффективности данного диагностического теста. Эффективность оценивается по площади под ROC-кривой. Полученное значение площади под кривой позволит нам оценить насколько эффективен тест, чем больше значение стремится к 1, тем лучше тест.

Одной из проблем диагностического теста может оказаться его высокая чувствительность, чувствительный тест редко относит больных к категории здоровых и применяется для предварительной диагностики. Другой проблемой может быть его высокая специфичность, в таком случае тест редко относит здоровых к категории больных и применяется для подтверждения наличия заболеваний. В идеальном тесте специфичность и чувствительность будут близки к 1. Тест редко бывает одновременно высокочувствительным и высокоспецифичным, поэтому его оценка эффективности перед использованием очень важна.

**Техническое задание на программное средство**

**Наименование программы**

«Программное приложение для оценки операционных характеристик и диагностической эффективности исследования пациента»

**1.Введение**

Техническое задание направлено на создание ПО, используемого для определения диагностической эффективности теста.

Программное обеспечение предназначено для использования в области медицины в целях научного исследования.

Назначение разработки ПО

ПО создается для нахождения оптимальной точки разделения - СOP-точки, которая приводит к минимальной сумме количества ложноположительных и ложноотрицательных ошибок в двух выборках или к максимальному индексу диагностической эффективности.

Построения ROC-кривой, которая показывает зависимость количества верно классифицированных положительных объектов (истинно положительное множество) от количества неверно классифицированных отрицательных объектов (ложно отрицательное множество). На основе этих данных будет определена диагностическая эффективность теста.

**2.Требования к ПО**

2.1 Организация входных и выходных данных: ПО должно учитывать возможность импорта данных из других программ (формат csv)

2.2 Программное обеспечение должно иметь дружественный интерфейс, рассчитанный на пользователя

2.3 ПО должно обладать инструментарием для графического отображения полученных результатов

2.4 Программное обеспечение должно работать платформе Windows 7 и на всех платформах, новее Windows 7

2.5Программное обеспечение должно занимать небольшой объем памяти

**3.Стадии и этапы разработки:**

3.1 Генерация двух частично перекрывающихся случайных модельных выборок по количественному признаку. Импорт входных данных в массив csv

3.2 Определение интервала перекрытия выборок. Нахождение максимума интервала здоровых (max H). Нахождение минимума интервала больных (min D)

3.3 Вычисление шага сдвига

Интервал [min D, max H] делится пополам , а затем делится на N – количество объектов в данном интервале

3.4. Определение оптимальной точки разделения диагностического теста и расчет операционных характеристик в этой точке.

Осуществляется цикл по интервалу с найденным шагом сдвига, в каждой точке разделения находится количество ложноположительных и ложноотрицательных ошибок. Далее производится занесение в массив найденных значений (4 столбца: PH, ND, NH, PD; строки – точки COP) Производится суммирование ошибок в каждой точке СОР и находится точка с минимальной суммой – оптимальная точка разделения

5. Расчет операционных характеристик: Se (чувствительности) и 1-Sp (неспецифичности) в каждой точке COP

6 Оценка диагностической эффективности теста по форме характеристической кривой и S под кривой:

- построение ROC-кривой по парам операционных характеристик

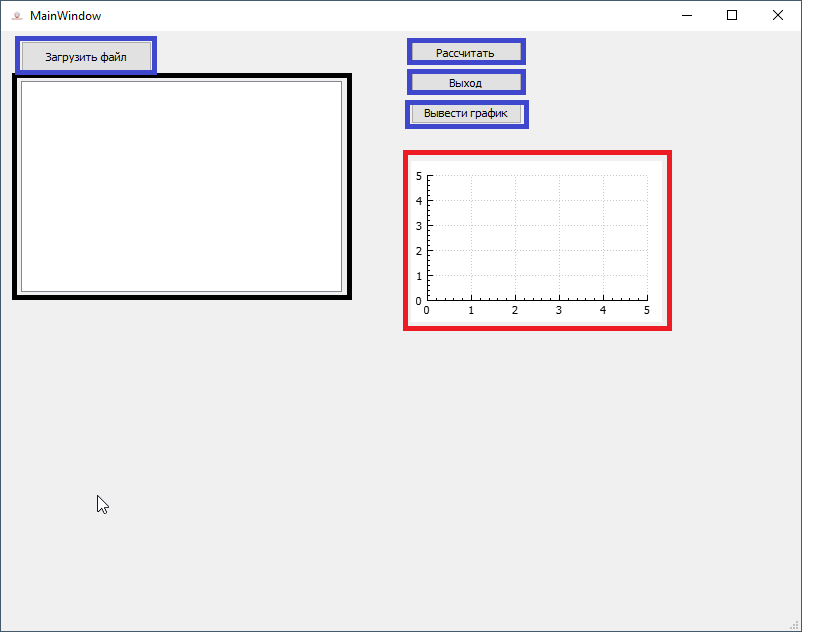
- нахождение площади под ROC-кривой: выделение трапеций по полученным точкам, определение площади каждой трапеции (произведение полусуммы оснований на высоту), суммирование всех площадей.

7.Контроль и приёмка ПО

Испытание проекта и контроль качества его работы провести путем изменения входных данных. Прием проекта будет осуществлен, если он соответствует техническому заданию к ПО.

**Описание программного приложения в виде совокупности модулей с указанием их функций и иллюстрацией экранными формами**

В папке с релизом приложения открыта форма c расширением «.exe».



На форме чёрным цветом выделен объект tableView, в который выводятся исходные данные выборок.

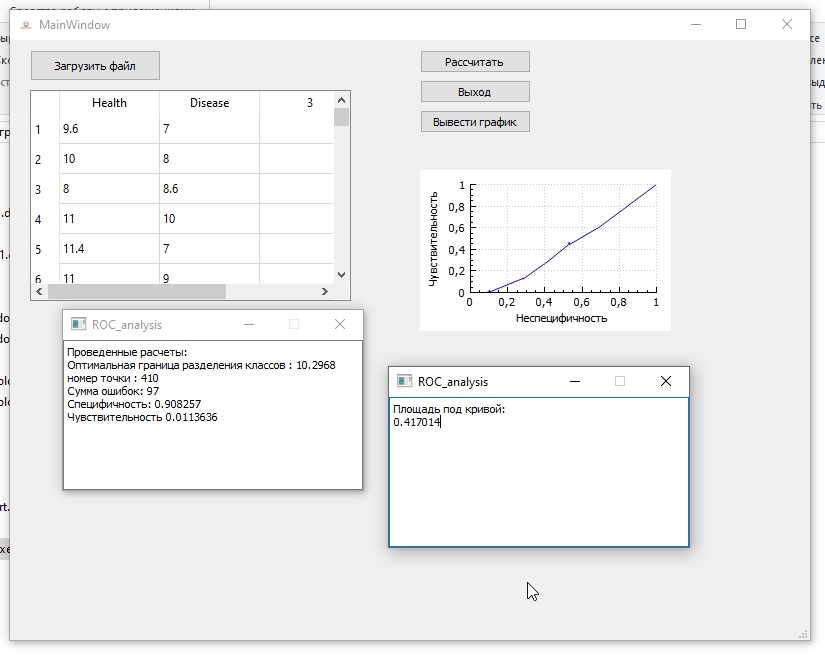
Красным цветом выделен объект widget, специально подключенный к QCustomPlot, этот элемент выводит ROC-кривую на экран.

Синим цветом выделены объекты pushButton, с помощью которых пользователь загружает исходные данные, проводит вычисления, рисует график и выходит из программы.

**Программа испытаний и результаты тестирования программного приложения**

Для проведения тестирования программы использовались выборки с разным перекрытием.

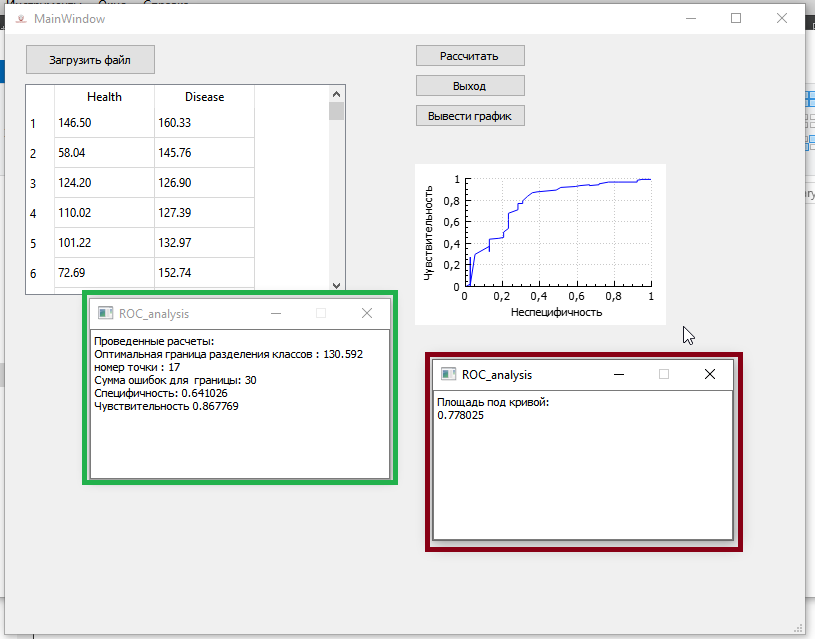
1.Выборки с полным перекрытием. Количество пациентов в каждой группе 110. Значения меняются от 7 до 11. Площадь под ROC-кривой равна 0,417014, мы можем сделать вывод что диагностический тест плохой.



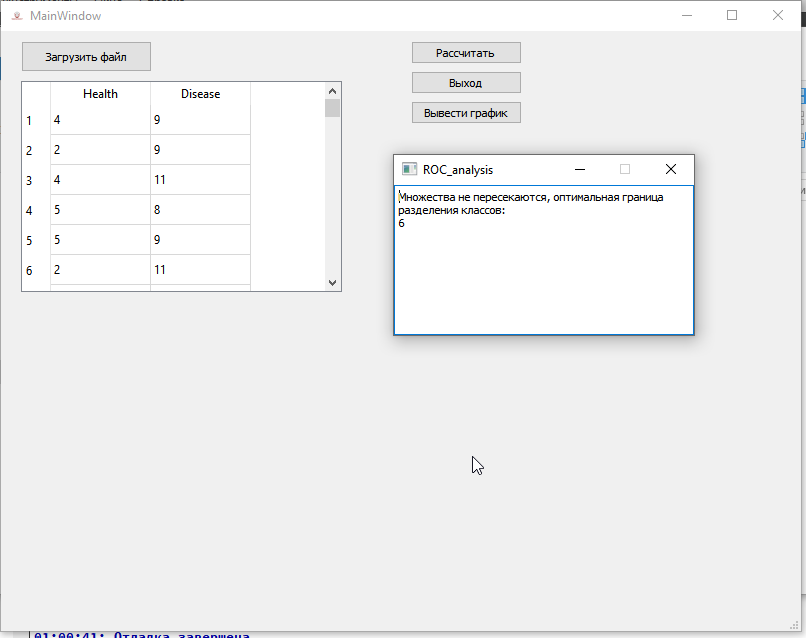
2.Выборки с частичным перекрытием в 30 процентов. Количество пациентов в каждой группе 121. Значения «здоровых» меняется от 22 до 160, значения больных меняется от 115 до 176.

Зеленым цветом выделен объект TextEdit, в котором происходит вывод оптимальной границы разделения классов, номера точки COP, сумма ошибок для оптимальной границы, специфичность и чувствительность для полученной границы. Специфичность и чувствительность равны 0.641026 и 0.867769 соответственно, это неплохие характеристики текста.

Площадь под ROC- кривой равна 0.778025 , что говорит о хорошем диагностическом тесте.



3.Выборки без перекрытия. Значения «здоровых» меняется от 1 до 5, значения «больных» от 7 до 11. Максимум здоровых – 5, Минимум больных – 7, Оптимальная точка ((max+min)/2) = 6



**Область применения и конечные пользователи программного приложения**

Данная программа может применяться в образовательных целях в медицине.

Пользователями могут быть:

1.Научные сотрудники

2.Врачи

3.Студенты

4.Преподаватели

**Руководство пользователя программного приложения**

1. Запустить программу Diagnosis\_test.exe
2. Нажать кнопку «Загрузить файл».
3. Выбрать нужный файл в формате \*.csv.
4. Нажать кнопку «Рассчитать».
5. Нажать кнопку «Вывести график».
6. Для выхода из программы нажать кнопку «Выход».

**Дополнительно**

скачать релиз вместе с отчетом, кодом программы и другими вложениями можно по ссылке: <https://github.com/Aryunaa/ROC_analysis>