Часть 2. Описание алгоритмов

# Общий алгоритм

Общий алгоритм работы модуля «Диагностика» представлен на рисунке 1 и реализован следующим образом (вариант использования для конечного пользователя):

1. Пользователь выбирает пациента и его результаты ЭКГ;
2. Чтение данных ЭКГ подмодулем «Сегментация ЭКГ»;
3. Сегментация сигнала ЭКГ;
4. Сохранение результатов сегментации, посредством интерфейса модуля «Кардиобаза»;
5. Чтение данных, необходимых для формирования признаков из модуля «Кардиобаза»;
6. Формирование признаков;
7. Сохранение признаков в модуль «Кардиобаза»;
8. Чтение необходимых данных (сигнал ЭКГ, результаты сегментации, выделенные признаки, информация о пациенте) в подмодуль «Прямые методы диагностики»
9. Выполнение прямых алгоритмов диагностики;
10. Сохранение промежуточных результатов диагностики и дополнительных признаков для машинного обучения.
11. Чтение имеющихся признаков для задачи машинного обучения;
12. Методы машинного обучения для диагностики;
13. Интеграция результатов диагностики и сохранение результата в модуль «Кардиобаза».
14. Предоставление результатов диагностики пользователю.

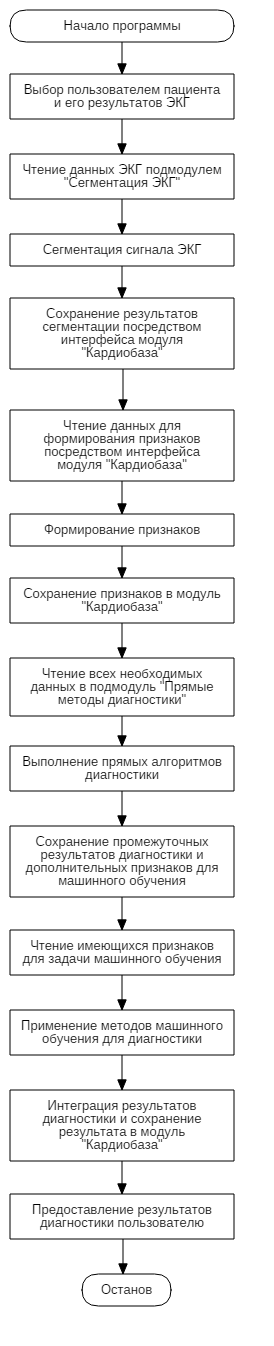


Рисунок 1. Алгоритм работы модуля «Диагностика»

# Алгоритм сегментации

Алгоритм сегментации сигнала ЭКГ представлен на рисунке 2. Работа модуля выполняется следующим образом:

1. Загрузка данных о пациенте и сигнала ЭКГ посредством интерфейса модуля "Кардиобаза".
2. В цикле по числу доступных отведений ЭКГ выполняется:
   1. Фильтрация сигнала ЭКГ.
   2. Выполнение дискретного вейвлет-преобразования и получение детализирующих коэффициентов различных шкал сигнала ЭКГ.
   3. Выделение QRS комплексов сигнала ЭКГ (рисунок 3):
3. Определение шкалы детализирующих коэффициентов.
4. Получение всех пересечений нуля детализирующими коэффициентами соответствующей шкалы.
5. Выбор пересечений нуля, соответствующих R-пикам.
6. Установление отметки об M-образности R-пика, если R-пик имеет M-образную морфологию.
7. Определение точек начала каждого QRS-комплекса.
8. Определение точек окончания каждого QRS-комплекса.
9. Выполнение уточнения расположения R-пика.
   1. Выделение P волн сигнала ЭКГ (рисунок 4):
10. Определение шкалы детализирующих коэффициентов.
11. В цикле по числу найденных QRS-комплексов выполняется:
12. Определение окна поиска волны P.
13. Получение всех пересечений нуля детализирующих коэффициентов выбранной шкалы в найденном окне.
14. Определение пика волны P.
15. Проверка полученного пика волны P на M-образность.
16. Проверка полученного пика волны P на бифазность.
17. Определение начала волны P.
18. Определение окончания волны P.
19. Выполнение уточнения расположения всех найденных пиков волн P.
    1. Выделение T волн сигнала ЭКГ (рисунок 5):
       1. Определение шкалы детализирующих коэффициентов.
       2. В цикле по числу найденных QRS-комплексов выполняется:
20. Определение окна поиска волны T.
21. Получение всех пересечений нуля детализирующих коэффициентов выбранной шкалы в найденном окне.
22. Определение пика волны T.
23. Проверка полученного пика волны T на M-образность.
24. Проверка полученного пика волны T на бифазность.
25. Определение начала волны T.
26. Определение окончания волны T.
    * 1. Выполнение уточнения расположения всех найденных пиков волн T.
27. Сохранение файлов сегментации посредством интерфейса модуля "Кардиобаза".

|  |  |
| --- | --- |
|  |  |
| Рисунок 2. Алгоритм работы подмодуля «Сегментация ЭКГ» | Рисунок 3. Алгоритм выделения QRS-комплекса |

|  |  |
| --- | --- |
|  |  |
| Рисунок 4. Алгоритм выделения волны P | Рисунок 5. Алгоритм выделения волны T |

# Алгоритм прямого метода

Алгоритм подмодуля «Прямые методы диагностики» представлен на рисунке ???. Работа подмодуля выполняется следующим образом:

1. Загрузка данных о пациенте и признаков посредством интерфейса модуля "Кардиобаза".
2. Определение, является ли ритм синусовым:
   1. Проверка, обнаружен ли зубец P в отведениях V1 и II.
   2. Если зубец P обнаружен и ЧСС находится между 30 и 180, ритм синусовый, иначе ритм не синусовый.
3. Проверка регулярности ритма.
4. Диагностика тахикардии и брадикардии.
5. Если в одном из отведений обнаружен зубец P и ЧСС > 120, поставить диагноз синусовая тахикардия.
6. Если в одном из отведений обнаружен зубец P и ЧСС < 30, поставить диагноз синусовая брадикардия.
7. Диагностика на основе длительности и морфологии зубца P.
8. Диагностика на основе интервала PQ.
9. Диагностика на основе комплекса QRS.
10. Диагностика на основе интервала QT.
11. Диагностика на основе сегмента ST.
12. Сохранение файлов с диагнозом и кардиологическим заключением и дополнительными признаками посредством интерфейса модуля "Кардиобаза".



# Алгоритм графового подхода

Алгоритм автоматической диагностики заболеваний сердца на основе графового подхода (parenclitic networks) представлен на рисунке 2. Работа метода выполняется следующим образом:

Построение паренклитической сети:

* 1. Выбор контрольной группы (например, 15 человек) из группы здоровых.
  2. Для всех пар признаков mi и mj, где
     1. Построение эталонной модели на контрольной группе: + \* , где и - коэффициенты эталонной модели.
     2. Нахождение стандартного отклонения от объектов до эталонной модели для контрольной группы.
     3. Добавление узлов mi и mj и связи между ними, которая имеет вес в сеть пациента.

Подготовка и описание сетей с помощью топологических индексов.

* 1. Применение порога. К взвешенным сетям пациентов применяется порог. Удаляются все связи между узлами со значениями выше порогового. Сеть становится невзвешенной.
  2. Формирование нового набора данных. Сеть каждого пациента характеризуется определенными значениями – центральностями (centrality) и другими характеристиками:
     1. Degree centrality.
     2. Closeness centrality.
     3. Betweenness centrality.
     4. Eigenvector centrality.
     5. Katz centrality.
     6. Edge betweenness centrality.
     7. Current flow closeness centrality.
     8. Current flow betweenness centrality.
     9. Communicability centrality.
     10. Load centrality.
  3. Каждая характеристика - это вектор, для которого находится среднее и максимальные значение. В конечном итоге, эти значения становятся признаками пациента для нового набора данных.

Применение методов классификации к новым данным, используя кросс-валидацию по 10-блокам для оценки качества классификации и вывод результатов классификации.

|  |
| --- |
|  |
|  |

|  |
| --- |
| ../Desktop/Parenclitic%20networks%20algorithm.png |
| Рисунок 2. Алгоритм автоматической диагностики заболеваний сердца на основе графового подхода (parenclitic networks) |

# Алгоритм машинного обучения и диагностики

#### Обучение:

1. Получение данных через модуль “Кардиобаза”. Тип считываемых признаков заранее известен и для каждой записи ЭКГ представляется числовым вектором заданной длины с, возможно, некоторыми неизвестными значениями компонент.
2. Всё множество векторов разделяется на классы по болезням, выбираются только представительные классы
3. Ставиться k независимых задач бинарной классификации (k - число классов болезней), в выборку попадают здоровые и люди с какой-то конкретной болезнью. Для каждой такой задачи:
   1. Выполняется этап предобработки данных: обнаружение выбросов, вставка пропущенных значений, нормализация, уменьшение размерности пространства признаков.
   2. Решение задачи различными методами машинного обучения, выбор из множества решений лучшего по некоторому критерию качества
4. На выходе этапа обучения получается k классификаторов, способных давать прогноз по выбранным заболеваниям.

#### Диагностика:

1. На вход программе прогнозирования подаётся числовой вектор, характеризующий подлежащую диагностированию запись ЭКГ. Интерпретация компонент вектора такая же, как на этапе обучения.
2. Этап предобработки вектора (такой же как при обучении)
3. По результату вычислений программа выдает вероятность наличия каждой из k-болезней.

Схема программы обучения алгоритмов машинного обучения и программы диагностики:

