

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ
Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра динамического моделирования и биомедицинской инженерии

«Сопоставление границ применимости методов статистического и
корреляционного анализа сигнала R-R интервалограмм при
диагностике артериальной гипертензии»

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

Студента 4 курса 4081 группы
направления 12.03.04 «Биотехнические системы и технологии»
профиль подготовки «Методы и устройства обработки биосигналов»
институт физики
Гузенин Даниил Дмитриевич

Научный руководитель:
доцент, к.ф.-м.н.


14.06.2024
подпись, дата

Ю.М. Ишбулатов

Зав. кафедрой динамического
моделирования и
биомедицинской инженерии,
д.ф.-м.н., профессор


14.06.2024
подпись, дата

А.С. Караваяев

Саратов 2024

Введение. Артериальная гипертензия — это синдром, при котором наблюдается стабильное повышенное артериальное давление (АД), систолическое и/или диастолическое.

Вариабельность ритма сердца (ВРС) - это физиологическое явление, которое заключается в изменении интервалов между последовательными синусовыми сердечными сокращениями.

Одной из областей применения спектрального, статистического и корреляционного анализа ВРС является диагностика состояния системы автономной регуляции системы кровообращения. Результирующие индексы характеризуют активность симпатических и парасимпатических отделов системы автономной регуляции и характер их взаимодействия

Актуальность данной работы обусловлена тем, что артериальная гипертензия является распространённым социально-значимым заболеванием. При этом большая часть больных артериальной гипертензией не знают о наличии у себя заболевания. Спектральный, фазовый и корреляционный анализы сигнала R-R-интервалограмм являются перспективными методами ранней неинвазивной диагностики. Так же, регистраторы RR-интервалограмм доступны в виде коммерческих носимых устройств, что обеспечивает доступность применения методов.

Целью данной работы является сопоставление границ применимости методов статистического, спектрального и корреляционного анализов при диагностике артериальной гипертензии по экспериментальным временным рядам R-R интервалограмм – последовательностей временных интервалов между R-пиками электрокардиограммы (ЭКГ). В ходе работы были поставлены и решены следующие задачи:

- Выделение RR-интервалограмм из электрокардиограмм
- Применение методов статистического, спектрального и корреляционного к RR-интервалограммам здоровых и пациентов с артериальной гипертензией

- ROC-анализ результатов
- Сопоставление точности методов при диагностике артериальной гипертензии

Объектом исследования являются электрокардиограммы людей, обратившихся в больницу, длительностью 5 минут каждая, частотой дискретизации 250 Гц. Общее количество обратившихся 60 человек, среди которых 20 здоровых и 40 с подтвержденной гипертонией. Данные о диагнозе подтверждены специалистами.

Вклад автора заключается в: выделении RR-интервалограмм из ЭКГ; статистическом, спектральном и корреляционном анализе RR-интервалограмм; ROC-анализе результатов

Объем и структура работы: выпускная квалификационная работа состоит из введения; двух глав, первая глава содержит 10 подпунктов, вторая – 4 подпункта; заключения; списка литературы. Общий объем ВКР 42 страницы. Список литературы содержит 5 наименований.

Основное содержание работы. Во введении обоснована актуальность темы исследования, сформулированы цели и задачи работы. Изложены основные положения, выносимые на защиту.

Первая глава, именуемая «Материалы и методы», содержит теоретическую и статистическую информацию об артериальной гипертензии, методах анализа ВРС, ROC-анализе, машинном обучении, метриках оценки результатов.

Артериальная гипертензия — это синдром, при котором наблюдается стабильное повышенное артериальное давление (АД), систолическое и/или диастолическое. ВОЗ (Всемирная Организация Здравоохранения) считает показатель артериального давления 140/90 мм рт. ст. (миллиметров ртутного столба) пороговым значением для диагностики артериальной гипертонии, поскольку эффективность лечения возрастает у пациентов, у которых

показатель артериального давления превышает это значение. Различают первую (140-159 САД/90-99 ДАД), вторую (160-179 САД/100-109 ДАД) и третью (180 САД и выше/110 ДАД и выше) степени АГ, а также ИСАГ (Изолированная систолическая артериальная гипертензия). Артериальная гипертензия часто протекает бессимптомно и имеет ряд опасных осложнений, таких как: гипертонический криз, сердечная и почечная недостаточности, инфаркт, инсульт, поражения глаз, аневризма аорты. Диагностика гипертензии происходит посредством измерения артериального давления (АД), анализа крови, ЭКГ.

Анализ ВСР. Вариабельность ритма сердца (ВРС) - это физиологическое явление, которое заключается в изменении интервалов между последовательными синусовыми сердечными сокращениями. Одной из областей применения спектрального, статистического и корреляционного анализа ВРС является диагностика состояния системы автономной регуляции системы кровообращения. Результирующие индексы характеризуют активность симпатических и парасимпатических отделов системы автономной регуляции и характер их взаимодействия.

Симпатическая регуляция кровообращения активизируется в ситуациях снижения артериального давления в аорте и каротидных синусах, активация данной системы приводит к увеличению частоты и силы сердечных сокращений и увеличению тонуса артериальных сосудов, что ведёт к нормализации артериального давления (АД). Активность симпатической регуляции проявляется в низкочастотном (0.04-0.15 Гц, LF) диапазоне сигнала RR-интервалограммы.

Парасимпатическая регуляция активизируется в ситуациях повышения АД в аорте и каротидных синусах, активация данной системы приводит к уменьшению частоты и силы сердечных сокращений. Активность парасимпатической регуляции проявляется в высокочастотном (0.15-0.4 Гц, HF) диапазоне сигнала RR-интервалограммы.

Спектральные методы анализа ВСР позволяют получить информацию о

распределении мощности колебаний по частотам, оценить различные частотные составляющие сердечного ритма и выявить активность регуляторных элементов. При анализе важно учитывать размер выборки. В работе используются компоненты HF, LF, VLF и ULF, представляющие собой частотные диапазоны дыхательных волн: высокие, низкие, очень низкие и ультранизкие соответственно. Важно учитывать физиологическую и клиническую интерпретацию данных при анализе вариабельности ритма сердца. Необходимо сравнивать полученные результаты с нормативными показателями.

Показатель HF отвечает за степень торможения активности автономного контура регуляции, за который ответствен парасимпатический отдел. Снижение этого показателя указывает на смещение вегетативного баланса в сторону преобладания симпатического отдела, то есть увеличение симпатической активности. Показатель LF отвечает за состояние системы регуляции сосудистого тонуса. Увеличение частоты в диапазоне этих волн говорит о снижении активности вазомоторного центра или замедлении барорефлекторной регуляции. Показатель VLF характеризует влияние высших вегетативных центров на сердечно-сосудистый подкорковый центр, отражает состояние нейрогуморального и метаболического уровней регуляции

Статистические методы анализа ВРС.

Существует три формата представления данных для математического анализа ВРС:

1) динамический ряд NN интервалов - NN_i , $i = 1, 2, \dots, n$. Термин «NN» используется вместо RR, чтобы подчеркнуть тот факт, что обработанные доли являются «нормальными» частями, то есть прошедшими «очистку» от артефактов и процедуру «нормализации»;

2) данные, вычисленные на основе разницы между NN-интервалами;

3) новый ряд дискретных значений x_i , $i = 1, 2, \dots, N$. Построение нового ряда основано на положении, что ВРС задается непрерывной функцией от

времени – $x(t)$, определенной на множестве элементарных событий – моментах появления R – зубцов. Значения функции в эти моменты равны величинам соответствующих NN-интервалов. Значения функции в промежутках времени между моментами появления R зубцов рассчитываются методом сплайновой кубической интерполяции. Ряд строится квантованием функции $x(t)$ с шагом 250 мс .

Расчет основных параметров variability должен включать в себя следующие показатели:

- ЧСС (HR) - определяется как количество NN-интервалов в записи, деленное на продолжительность их записи;
- X_{mean} - среднее значение;
- D – дисперсия;
- SDNN - стандартное отклонения полного массива кардиоинтервалов;
- RMSSD - квадратный корень из суммы разностей последовательного ряда кардиоинтервалов;
- pNN50 –число пар кардиоинтервалов с разностью более 50 мс. в % к общему числу кардиоинтервалов в массиве;
- CV - Коэффициент вариации полного массива кардиоинтервалов

Автокорреляционный анализ включает в себя расчет индексов CC1 и CC0. CC1 - коэффициент корреляции после первого сдвига, физиологической интерпретацией которого является степень активности автономного контура регуляции: $CC1 = r_{0,1}$, где $r_{0,1}$ - коэффициент корреляции, который рассчитывается путем вычисления автокорреляционной функции при величине сдвига – 1 секунда. Автокорреляционная функция строится по значениям ряда коэффициентов корреляции между исходным динамическим рядом и новыми рядами, полученными при последовательных его смещениях на одно значение.

CC0 – число сдвигов автокорреляционной функции до первого

нулевого значения коэффициента корреляции, физиологической интерпретацией которого является степень активности центрального контура регуляции.

ROC-кривая (англ. receiver operating characteristic) - кривая рабочих характеристик приемника представляет собой график, иллюстрирующий качество модели двоичного классификатора (также может использоваться для классификации нескольких классов) при различных пороговых значениях. ROC-кривая представляет собой график зависимости истинно положительного показателя (TPR) от частоты ложноположительных результатов (FPR) при каждом пороговом значении. TPR определяется как доля правильно классифицированных положительных результатов относительно всех положительных результатов в данных. FPR определяет долю ошибочно классифицированных отрицательных результатов относительно всех отрицательных результатов.

Кривая ROC представляет собой графическое представление компромисса между чувствительностью и специфичностью при различных порогах классификации. Идеальная модель классификации будет стремиться к точке в верхнем левом углу графика, где TPR равно 1, а FPR равно 0.

AUC (Area Under Curve) - площадь под кривой - равна вероятности того, что классификатор оценит случайно выбранный положительный объект выше, чем случайно выбранный отрицательный (предполагая, что «положительный» ранжируется выше, чем «отрицательный»). Другими словами, при наличии одного случайно выбранного положительного объекта и одного случайно выбранного отрицательного объекта, AUC — это вероятность того, что классификатор сможет сказать, какой из них есть какой.

Существуют различные методы машинного обучения, самые распространенные среди которых это обучение с учителем, обучение без учителя, обучение с подкреплением, глубокое обучение. В данной работе, при обучении с учителем машине выдается обучающая выборка с

размеченными заранее данными, в которых указаны правильные ответы. И на основе этих данных машина уже сама определяет правильные ответы в схожих с примером задачах. В рамках данной работы используется классификация на основе сходства, включающая в себя следующие методы: метод ближайших соседей, метод парзеновского окна, метод потенциальных функций, метод радиальных базисных функций, отбор эталонных объектов.

Алгоритм работы метода ближайших соседей выглядит следующим образом:

- сначала вычисляется расстояние между тестовым и всеми обучающими образцами;
- далее из них выбирается k -ближайших образцов (соседей), где число k задаётся заранее;
- итоговым прогнозом среди выбранных k -ближайших образцов будет мода в случае классификации и среднее арифметическое в случае регрессии;
- предыдущие шаги повторяются для всех тестовых образцов.

Матрица ошибок - это таблица показывающая эффективность алгоритма классификации путем сравнения прогнозируемого значения целевой переменной с ее фактическим значением. Матрица классифицируется по следующим категориям: Positive – наблюдение, относящееся к целевой популяции, Negative – наблюдение, не относящееся к целевой популяции, True/False – верное или неверное соответствие. Матрица отражает число ложноположительных, ложноотрицательных, истинно положительных, истинно отрицательных целей.

Вторая глава, именуемая «Результаты», содержит практическую часть работы, а именно: выделение RR-интервалограмм из ЭКГ; применение методов статистического, спектрального и корреляционного к RR-интервалограммам здоровых и пациентов с артериальной гипертензией; ROC-анализ результатов.

Выделение RR-интервалограмм из ЭКГ представляет собой преобразование временного ряда RR- интервалов в соответствующую интервалограмму. Для этого нужно обнаружить R-пики и выделить интервалы между ними. Необходимо провести расчет и построение производной временного ряда ЭКГ со сглаживанием полинома. Рассчитанная производная, есть скорость изменения данного сигнала. Прделанные действия позволяют установить необходимый порог для поиска R зубцов уже на графике производной. Теперь, когда был задан порог амплитуды производной R-пики, необходимо провести поиск вершин этих самых R-пики. Следующим этапом является генерация ряда RR-интервалов – конечная цель обработки электрокардиограммы. Итак, были выделены RR-интервалограммы из реализации ЭКГ и очищены от артефактов.

Далее происходит расчет статистических и спектральных характеристик. С помощью специальной программы проводится анализ RR-интревалограммы, на основе анализа выделяются спектральные характеристики HF, LF, VLF, ULF и статистические и корреляционные индексы HR, Xmean, D, SDNN, CV, RMSSD, PNN50, CC1 и CC0.

В подпункте «Применение спектральных характеристик в методах машинного обучения» была разработана программа, используемая для диагностики АГ по показателям вышеперечисленных индексов. Для создания программы использовался Jupiter Notebook. Итоговый файл содержит программный код, разбитый на смысловые блоки:

1. Подключение библиотек и чтение данных.
2. Преобразование данных в вид, пригодный для подачи на вход классификатора.
3. Разделение данных на тестовую и тренировочную выборку.
4. Выбор наилучшей модели.
5. Предсказания на тестовой выборке.

Результатами работы программы являются матрица спутанности,

значение точности диагностики, чувствительности и специфичности. Матрица имеет размерность 2×2 , поскольку осуществляется разделение данных на два класса — 0 и 1. При выводе реальных значений матрицы видно, что большая часть предсказанных значений оказались верными. Так из 94 тренировочных объектов, которые принадлежали классу 0, для 12 был предсказан класс 1; а из 88 объектов, принадлежащих классу 1, для 21 был выбран класс 0. Полученное значение точности равно 0.74. Показатель специфичности равен 0.87. Показатель чувствительности равен 0.76.

Подпункт «Построение ROC – кривых в Python» содержит программу, используемую для построения ROC-кривых по данным различных коэффициентов и характеристик (необходим лишь текстовый формат загружаемого файла) и вычисления AUC. Кривые строятся по данным из двух текстовых файлов для двух осей соответственно. В одном файле содержатся показатели здоровых людей, в другом файле показатели гипертоников. Программа имеет следующую структуру:

1. Импорт библиотек и чтение данных.
2. Расчёт TPR и FPR.
3. Расчёт AUC и вывод результатов.

В ходе работы были получены несколько ROC-кривых и значений AUC для различных показателей методов статистического, спектрального и корреляционного анализа, наименьшую точность из которых имеет спектральная характеристика ULF при $AUC = 0.64$, а наибольшую точность имеет корреляционный показатель CC1 при $AUC = 0.88$.

Заключение. В заключении подводятся итоги выпускной квалификационной работы, излагаются его основные результаты:

В ходе работы сопоставлялись границы применимости статистического, спектрального и корреляционного анализов при диагностике артериальной гипертензии по экспериментальным временным рядам R-R интервалограмм. Количественной мерой диагностической силы

выступала площадь по ROC-кривой.

Среди методов автокорреляционного анализа наибольшую диагностическую силу продемонстрировал первый коэффициент автокорреляционной функции. Площадь под кривой составила 0.88, при специфичности - 0.93, и чувствительности - 0.83.

Среди методов статистического анализа наибольшую диагностическую силу продемонстрировал коэффициент вариации. Площадь под кривой составила 0.87, при специфичности - 0.93, и чувствительности - 0.72.

Среди методов спектрального анализа наибольшую диагностическую силу продемонстрировал коэффициент вариации. Площадь под кривой составила 0.78, при специфичности - 0.86, и чувствительности - 0.58. При использовании спектральных индексов в качестве входных данных для алгоритма машинного обучения (к-ближайших соседей) была достигнута точность диагностики артериальной гипертензии 0.80, при специфичности равной 0.87 и чувствительности равной 0.76.

Согласно полученным результатам, наибольшей точностью диагностики артериальной гипертензии обладают методы автокорреляционного анализа, в частности, значение первого коэффициента автокорреляционной функции, который трактуют как индекс степени активности автономного контура регуляции.

14.06.2024  Тугенци Д. Д.