

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВПО «МГИУ»)

КАФЕДРА «ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ»

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

по специальности «Математическое обеспечение и администрирование
информационных систем»

студента Николаевой Маргариты Владимировны
на тему «РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ
СТАТИСТИЧЕСКОЙ ОБРАБОТКИ РАСПРЕДЕЛЕНИЯ ПЛОТНОСТИ
ТОКОВ ПОЛУЧЕННЫХ В РЕЗУЛЬТАТЕ ПРОВЕДЕНИЯ
МАГНИТОКАРДИОГРАФИИ»

Руководитель работы:
доцент, к.ф.-м.н. Белова Ирина Михайловна

Студент-дипломник
Руководитель работы,
доцент, к.ф. –м.н.

Николаева М.В.

Белова И.М.

ДОПУСКАЕТСЯ К ЗАЩИТЕ

Зав. кафедрой 36,
доцент, к.ф. –м.н.

Роганов Е.А

МОСКВА 2012

Аннотация

В выпускной квалификационной работе представлена разработка программы, позволяющая обработать распределения плотности токов полученных в результате проведения магнитокардиографии. В работе был использован метод главных компонент, одноклассовый классификатор на основе главных компонент и одноклассовый классификатор SVDD.

Выпускная квалификационная работа содержит 45 страницы, 3 таблицы, 15 иллюстраций и 9 элементов в списке используемой литературы.

Ключевые слова: МЕТОД ГЛАВНЫХ КОМПОНЕНТ, СОКРАЩЕНИЕ РАЗМЕРНОСТИ СИГНАЛА, КЛАССИФИКАТОР НА ОСНОВЕ АНАЛИЗА ГЛАВНЫХ КОМПОНЕНТ, ОДНОКЛАССОВЫЙ КЛАССИФИКАТОР SVDD.

Оглавление

Аннотация	2
Введение.....	4
Глава 1. Описание магнитографического комплекса	6
1.1. Общая структура программного обеспечения	6
1.2. Пакет начальной обработки МКГ	8
1.3. Предобработка МКГ-сигналов	11
Глава 2. Постановка задачи.	18
Глава 3. Сокращение размерности сигнала.....	20
3.1. Метод главных компонент	20
3.2. Структура анализа главных компонент.....	20
3.3. Основные представления данных	25
3.4. Сокращение размерности	26
Глава 4. Методы классификации сокращенных сигналов многоканальных ЭКГ	29
4.1. Классификатор на основе анализа главных компонент	29
4.2. Статистический одноклассовый классификатор SVDD	29
Глава 5. Программная реализация и обсуждение результатов.	35
Выводы	42
Литература.....	43

Введение

В последнее десятилетие магнитокардиография (МКГ) становится одним из новых инструментальных средств для диагностики заболеваний сердца.

Этот метод интенсивно развивается и предполагает исследование величин параметров магнитного поля сердца, зарегистрированных в воздухе над грудной клеткой пациента. Измерения МКГ выполняются бесконтактно, а сама магнитометрическая система не оказывает никакого воздействия на электрофизиологические процессы в сердце, так как при регистрации диагностической информации она не излучает никакой энергии.

В электрофизиологии сердца, установлено, что причина практически любой его патологии связана с механизмом изменения плотности тока ионов через мембрану клетки. Изменение этой плотности по отношению к норме – это и есть первые функциональные нарушения, которые приводят к патологии.

Таким образом, МКГ, чувствительная, главным образом, к внутри - и внеклеточным ионным токам сердца, дает более раннюю диагностическую информацию, чем известные потенциальные методы: электрокардиография; к артирование поверхностных потенциалов тела (КППТ).

Более того, в данных потенциальных измерений (ЭКГ) нет информации о форме элементов тока источника поля, потенциалы которого регистрируют на поверхности тела, в то время как магнитометрические данные содержат информацию и о пространственной конфигурации ионных токов в сердце.

Следовательно, разработка математических моделей и программного обеспечения для статистической обработки данных, полученных в результате проведения магнитокардиографии является актуальной задачей.

Основной целью работы является статистическая обработка распределения плотности токов.

Главной задачей работы является обработка данных распределения токов и попытка классифицировать больных на основе анализа главных компонент.

Глава 1. Описание магнитографического комплекса

Программное обеспечение для определенности ориентировано на МКГ измерительную систему с девятью измерительными каналами в виде осесимметричных градиентометров второго порядка и референтным векторным (три ортогональных приемника) магнитометром. Алгоритм исследования предполагает последовательные измерения магнитокардиосигнала в четырех пространственных позициях в одной плоскости над грудной клеткой пациента таким образом, чтобы точки измерений составляли регулярную квадратную сетку 66 точек с шагом 40 мм. В каждой позиции синхронно с МКГ измеряется ЭКГ по второму стандартному отведению. Схема измерений и положение измерительных позиций по отношению к анатомическому ориентиру показана на рис.1

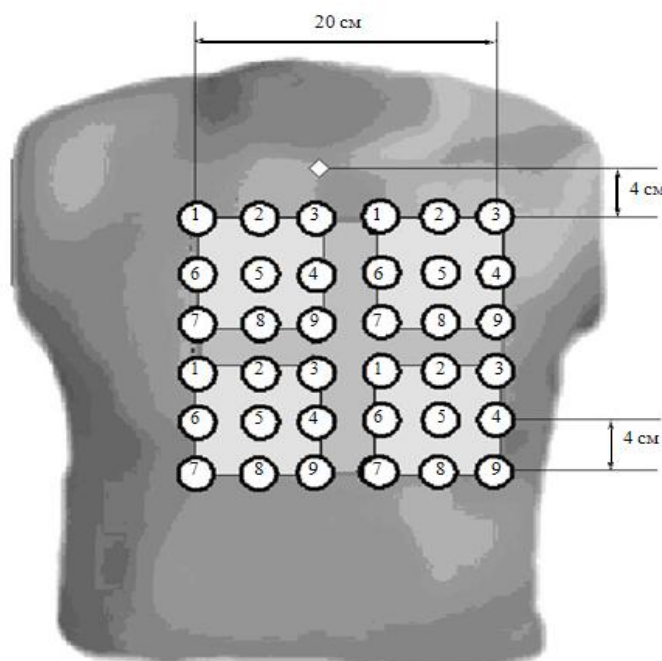


Рис. 1. Схема расположения точек измерения и позиций измерителя при регистрации МКГ.

1.1. Общая структура программного обеспечения

В ходе МКГ исследования можно выделить две последовательные

относительно автономные стадии:

1. ввод магнитокардиосигналов, сохранение результатов измерений в базе данных и цифровая предобработка сигналов;
2. выделение интервала кардиоцикла для исследования, пространственно-временной анализ и определение характеристик МКГ, решение обратной задачи и пространственно-временной анализ распределения источников сигналов.

Основной задачей первого этапа, кроме непосредственного ввода и сохранения данных измерений, является выделение МКГ-кардиоциклов и очистка сигналов от помех. На втором этапе, кроме расчета параметров и характеристик магнитного поля и его источников, выполняется также визуализация пространственных распределений источников магнитного сигнала и анализ временной динамики их изменений.

Каждая из двух стадий МКГ-исследования реализована в виде отдельного программного пакета. Принципиальная схема структуры программного обеспечения магнитокардиографа показана на рис. 2.

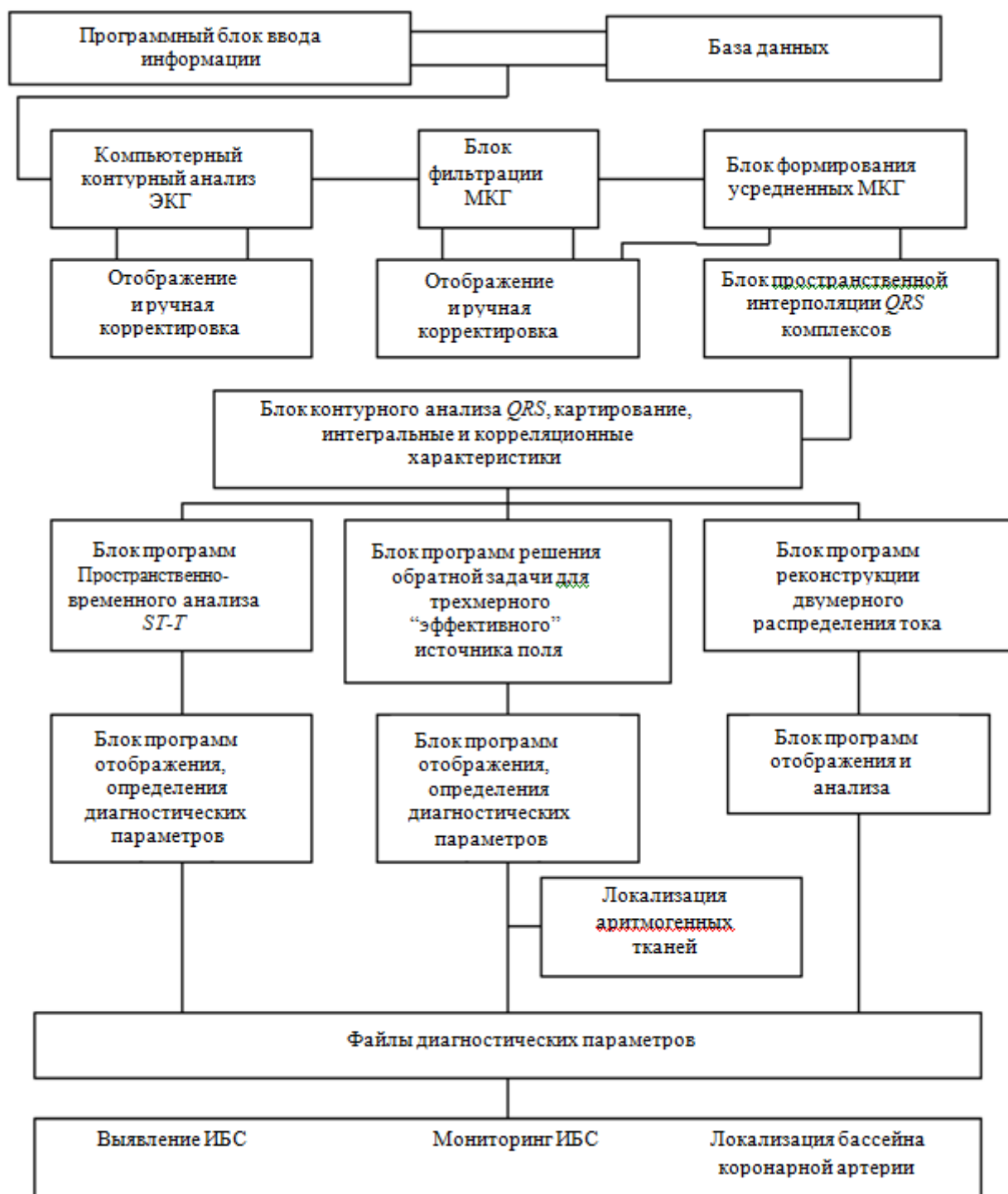


Рис. 2. Принципиальная блок-схема программного обеспечения магнитокардиографа.

1.2. Пакет начальной обработки МКГ

Данный пакет включает в себя три логически раздела программных блока:

- блок ввода данных из измерительной системы в компьютер;
- база данных МКГ- исследований;
- предобработка МКГ-сигналов.

Первые два блока реализованы в виде независимых программных *exe*-модулей. Третий разделяется на последовательные этапы, которые также реализованы как отдельный *exe*-модуль. Связь между блоками, выбор режима работы с базами данных (БД) и режима предобработки, а также интерфейс взаимодействия с пользователем обеспечивает отдельная управляющая программа пакета.

1.2.1. Программа ввода и регистрации МКГ

Программа 9MCG предназначена для ввода и накопления данных МКГ-измерений в компьютере, контроля качества полезного МКГ- сигнала

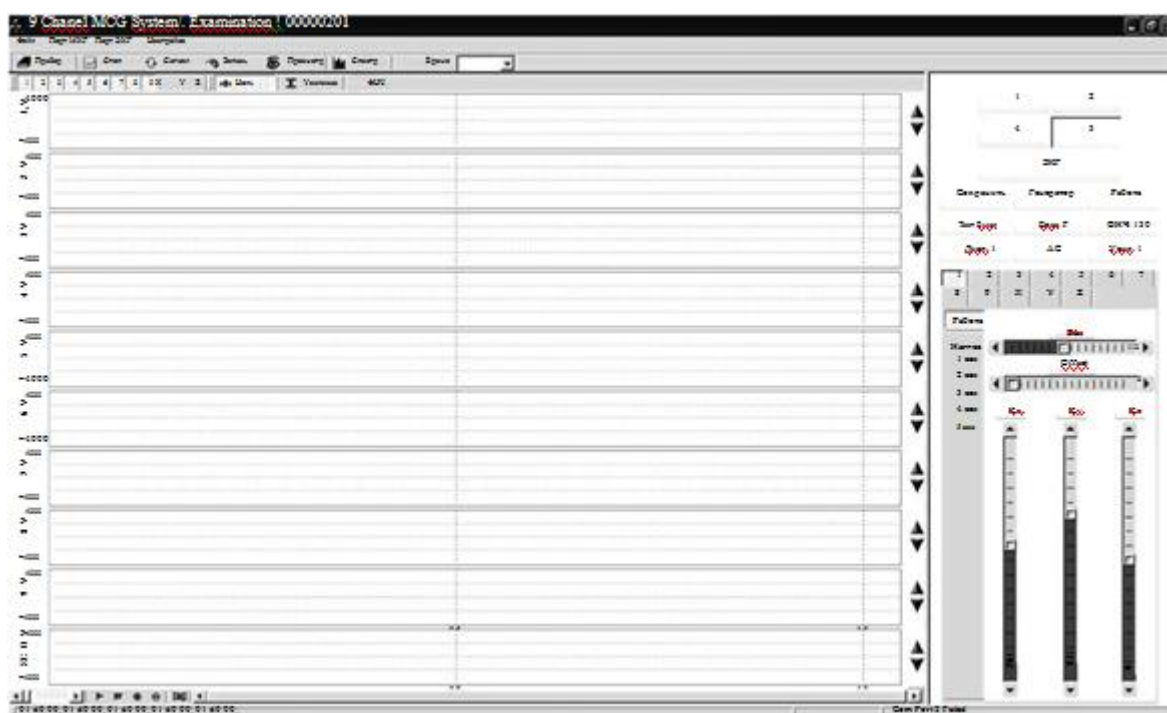


Рис. 3. Отображение МКГ данных пациента на экране компьютера в одном из режимов работы программы ввода.

на фоне помех в режиме реального времени, а также настройки характеристик технических элементов магнитометрической системы. Основные функциональные возможности программы ввода данных: синхронный ввод и

отображение текущих данных по 13 каналам в режиме “онлайн”; отображение текущих данных в режиме “оффлайн” (“замороженное” отображение); запись данных в файлы на жесткий диск компьютера; управление параметрами записи и отображением данных во всех режимах работы программы; цифровая фильтрация отображаемых данных, расчет и отображение спектров вводимых сигналов; просмотр и анализ качества записанных данных в текущем сеансе измерений в статическом режиме; настройка характеристик технических элементов магнитометрического модуля МКГ-системы. На рис. 3 представлено основное окно программы 9MSG в процессе записи магнитометрических данных.

1.2.2. База данных МКГ-исследований

Для хранения информации о пациентах, МКГ-измерениях и результатах исследований разработана база данных МКГ-исследований, которая включает картотеку пациентов с возможностями быстрого поиска, отбора и сортировки данных.

Для каждого пациента динамически подгружается таблица с МКГ-измерениями. Предусмотрен ввод обширных комментариев по каждому исследованию и результату. У пользователя имеется возможность откорректировать или удалить ранее введенную в БД информацию (рис. 4).

В рамках БД реализовано запоминание, хранение, обновление и извлечение исходных данных МКГ-измерений в виде плоской структуры сжатых бинарных файлов. Каждый такой компрессированный файл соответствует одному полному МКГ-исследованию и содержит файлы с исходными записями МКГ, основными промежуточными и конечными результатами. Предусмотрено три режима работы БД МКГ-исследований:

Ввод карточки нового пациента и/или нового исследования, инициализация работы внешнего программного модуля для ввода данных, запоминание введенных данных в БД.

1. Извлечение из БД ранее накопленных данных МКГ для последующей обработки, анализа и интерпретации.
2. Запоминание в БД результатов анализа конкретного МКГ-исследования.
3. В качестве дополнительных возможностей реализовано создание новой пустой БД, сохранение текущей БД в другом месте (например, для ведения архива) и открытие другой БД.

1.3. Предобработка МКГ-сигналов

Цель предобработки состоит в повышении качества введенных в компьютер сигналов с измерительных каналов магнитокардиографа. Качество МКГ сигналов при измерениях в неэкранированных помещениях в присутствии сильных внешних помех (главным образом сигналов помех на промышленной частоте 50 Гц и ее гармониках) в большой степени определяется отношением сигнал/помеха в записанных в компьютер данных измерений. Как показывает опыт, в исходных данных даже в моменты времени с максимально сильным сигналом и в тех пространственных точках, где сигнал магнитного поля сердца имеет максимальную амплитуду, отношение сигнал/помеха не превышает 1. Алгоритмы цифровой обработки МКГ-сигналов позволяют подавить шум до уровня, когда в усредненном кардиоцикле отношение сигнал/помеха становится больше 100.

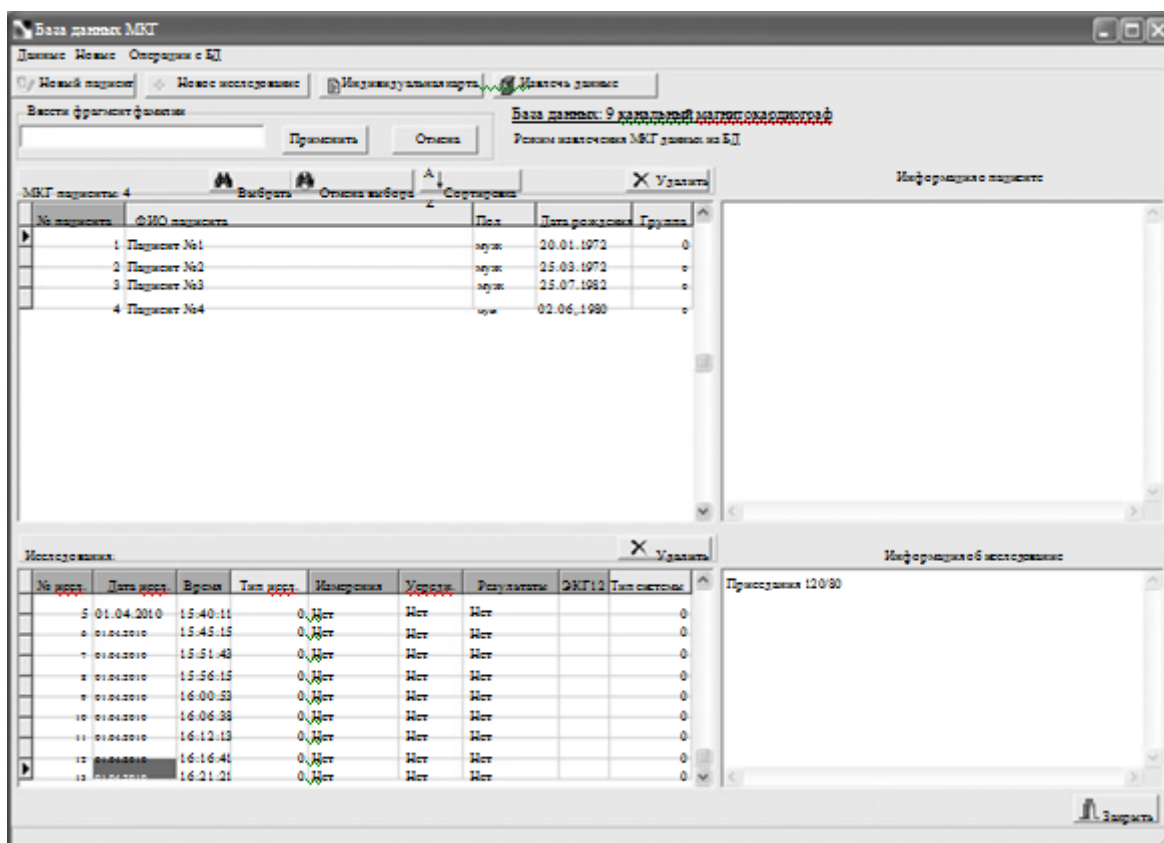


Рис. 4. Изображение основного окна базы данных МКГ-исследований.

В рамках подсистемы предобработки МКГ разработаны формы представления входной, промежуточной и выходной информации в виде бинарных файлов на различных стадиях обработки. Разработана схема движения информации и интерфейсы между отдельными программами, входящими в подсистему, а также между базой данных и программой анализа усредненных комплексов. Предобработка может вестись в двух режимах: 1) автоматическом, когда все этапы обработки выполняются последовательно без участия пользователя; 2) пошаговом, когда на каждом этапе пользователь может проанализировать результаты обработки, скорректировать параметры и повторить этап. Работа программного обеспечения протестирована на нескольких сотнях МКГ-исследований.

Алгоритмы предобработки включают: 1) алгоритмы автоматического контурного анализа ЭКГ; 2) алгоритмы фильтрации МКГ-записей; 3) алгоритмы усреднения МКГ-кардиокомплексов (*P-QRS-T*-комплексов). Ясно, что качество обработки данных на каждом этапе существенно влияет на конечный результат.

Рассмотрим особенности реализации трех основных разделов общего алгоритма.

1.3.1. Контурный анализ ЭКГ

При использовании МКГ - систем с небольшим количеством измерительных каналов (в нашем случае девять каналов) в каждом исследовании для покрытия стандартной области 20×20 см измерения проводятся последовательно в четырех пространственных позициях системы относительно пациента. Для синхронизации данных МКГ, соответствующих разным пространственным позициям, используются записи ЭКГ по второму стандартному отведению.



Рис. 5 Отображение на экране компьютера результатов контурного анализа ЭКГ в одном из исследований.

Электрокардиограммы записываются синхронно с МКГ в каждой позиции. На этапе усреднения МКГ QRS-комплексов усредняются также и

QRS-комплексы в ЭКГ. Усредненные комплексы ЭКГ для всех позиций совмещаются относительно одинаковой базисной точки в комплексе. С теми же смещениями, что и для ЭКГ, совмещаются и МКГ QRS-комплексы для всех пространственных позиций. Кроме синхронизации данных для разных пространственных позиций, записи ЭКГ используются и при усреднении QRS-комплексов в каждом канале МКГ (рис.5).

В задачи контурного анализа ЭКГ входит обнаружение и выделение QRS-комплексов, распознавание и разделение по типам (кластеризация), определение положений базисных точек и границ комплексов. Разработан оригинальный алгоритм контурного анализа ЭКГ, который реализован в виде отдельной программы. Перед началом анализа производится цифровая фильтрация высоких (50 Гц и выше) и низких (ниже 0.5...0.7 Гц) частот в записях ЭКГ. Далее на основе анализа амплитуд, длительности и крутизны фронтов производной ЭКГ по времени выделяются QRS-комплексы и находится положение левой и правой границы каждого из них.

Завершается контурный анализ ЭКГ оценкой типов комплексов и формированием кластеров – последовательностей комплексов одного типа. Формирование кластеров выполняется итерационно с уточнением усредненного или наиболее представительного комплекса каждого типа. Таким образом, в ЭКГ автоматически распознаются и разделяются комплексы основного синусового типа, экстрасистолы различных типов, а также комплексы основного типа, предшествующие или последующие экстрасистолам. Анализ типов комплексов позволяет избежать случайного усреднения комплексов из разных кластеров. В программе реализована возможность для пользователя вручную скорректировать результаты компьютерного анализа: изменить тип комплекса, удалить или добавить комплекс. При завершении работы программы определяется средняя, максимальная и минимальная частота сердечных сокращений (ЧСС), а также среднеквадратичное отклонение ЧСС для основного типа комплексов в каждой ЭКГ.

Работа программы контурного анализа ЭКГ проверена и отлажена на нескольких тысячах записей ЭКГ для пациентов с разнообразными патологиями и нарушениями ритма. В результате необходимость ручной корректировки сведена к минимуму.

1.3.2. Фильтрация МКГ-записей

Помехи в каналах МКГ имеют две составляющие: некоррелированный случайный шум, обусловленный электронным трактом (усилители, фильтры и т.д.) и мощные помехи от внешних источников магнитного поля различной природы. Этап обработки, связанный с фильтрацией МКГ-записей, направлен главным образом на подавление внешних коррелированных помех. Специально для борьбы с этими помехами разработанная МКГ-система включает в свой состав и использует дополнительный трехканальный референтный векторный магнитометр (РВМ). Адаптивное вычитание данных РВМ из данных измерительных МКГ-каналов позволяет существенно подавить влияние внешних помех в рабочем частотном диапазоне полезного МКГ-сигнала. Алгоритм комплексного подавления помех в данных МКГ-измерений содержит следующие этапы:

1. Цифровая фильтрация низких частот с частотой среза 50 Гц. Фильтруются параллельно данные МКГ- и РВМ -каналов. Для большинства задач в МКГ исследованиях, в частности, при исследовании S-T и P-Q интервалов, высокочастотной составляющей (выше 50 Гц) в магнитном поле сердца можно пренебречь. В случае необходимости учета высоких частот, цифровая фильтрация нижних частот может быть заменена узкополосным пробочным фильтром с центральной частотой, равной частоте электрической сети. Данный этап алгоритма предназначен в первую очередь для фильтрации сетевой гармоники- наиболее мощной, как правило, внешней помехи. Однако амплитуда и фаза сетевой гармоники внешнего поля могут изменяться на протяжении записи, в том числе, и скачкообразно. После фильтрации эти изменения переходят в

тренд и скачки базовой нулевой линии МКГ записей и когерентных базовых линий записей РВМ.

2. Цифровая фильтрация высоких частот с частотой среза 0.5 Гц. Фильтруются также параллельно данные МКГ- и РВМ-каналов. Такой ВЧ-фильтр позволяет уменьшить влияние низкочастотных колебаний базовой линии за счет, например дыхания пациента.

3. Адаптивное вычитание отфильтрованных данных РВМ из отфильтрованных данных МКГ. Разработана модель, методика и алгоритм определения оптимальных коэффициентов для адаптивного смешивания сигналов. Предложенный алгоритм адаптивной фильтрации позволяет подавить помехи в диапазоне рабочих частот полезного МКГ- сигнала, не искажая при этом сам полезный сигнал. При настройке алгоритма фильтрации для конкретной помеховой обстановки предусмотрена возможность выбора различных типов ФНЧ и ФБЧ, добавление любого количества узкополосных фильтров для МКГ- и РВМ-данных. Изменение набора и параметров фильтров можно выполнить как в специальной таблице, так и непосредственно в процессе анализа спектра. Весь алгоритм цифровой фильтрации МКГ реализован в виде отдельной программы, работа которой была опробована и оттестирована на многочисленных МКГ-записях. Пользователь имеет возможность визуально сравнивать результаты фильтрации на отдельных этапах, в том числе и в спектральной форме.

1.3.3. Усреднение МКГ QRS- комплексов.

Процедура усреднения кардиоциклов необходима для уменьшения уровня помех, случайных и слабо коррелированных с МКГ-сигналами. Усреднение основано на результатах контурного анализа ЭКГ и выполняется отдельно для каждого из выделенных ЭКГ типов комплексов, если они присутствуют для всех измеренных позиции системы. Предусмотрена возможность автоматического определения наиболее зашумленных QRS-комплексов в МКГ-данных и удаления их из усредняемого набора. Кроме того, пользователь может вручную на графике предобработанной МКГ удалить из

усреднения тот или иной комплекс. Реализовано графическое представление результатов усреднения и остаточных после вычитания усредненных комплексов МКГ-данных. Конечным результатом всей предобработки является получение файла с усредненными синхронизированными кардиоциклами для всех пространственных точек измерений МКГ.

Глава 2. Постановка задачи.

Задача исследования в кардиографии ставится как задача классификации - получения диагностических параметров, которые могут быть использованы для выработки диагностического заключения. Задача классификации состоит из отнесения некоторого индивидуума (пациента) к одной из k популяций W_1, W_2, \dots, W_k на основе измерения p параметров X_1, X_2, \dots, X_p .

В нашем случае измеренными параметрами X_1, X_2, \dots, X_p являются результаты определения пространственно-временного распределения источников кардиомагнитного сигнала в сердце на выделенном интервале кардиоцикла в виде найденной в дискретном наборе точек ($31 \times 31 = 961$) в плоскости, параллельной плоскости измерений.

При постановке задачи предполагается, что магнитное поле сердца в произвольной точке плоскости измерений описывается вектором магнитной индукции \mathbf{B} и пространственными производными этого вектора первого D_1 ($d\mathbf{B}_k/dr_i; k, i = x, y, z$) второго D_2 порядка ($d^2\mathbf{B}_k/dr_i^2; k, i, j = x, y, z$). Поскольку точки измерения расположены в немагнитной, непроводящей, однородной среде (в воздухе), то вектор магнитной индукции удовлетворяет однородным уравнениям Максвелла.

Предполагаем, что магнитное поле в точках наблюдения (измерения) описывается уравнением

$$\rho_i = \sum_{k \neq i} [1 + (r_{xi} - r_{xk})^2 + (r_{yi} - r_{yk})^2 + (r_{zi} - r_{zk})^2]^{-1} \quad (2.1)$$

и источник поля расположен в плоскости, параллельной плоскости измерений, требуется найти пространственную конфигурацию двумерной системы токов, магнитное поле которой соответствует измеренному.

С помощью диагностического комплекса “МАГ-СКАН” получаем пространственную конфигурацию двумерной системы токов, которую мы должны исследовать.

Целью настоящей работы является разработка программного обеспечения для статистической обработки распределения плотности токов,

полученных в результате проведения магнитокардиографии.

Для уменьшения размерности данных, полученных в результате проведения магнитокардиографии применяем метод главных компонент.

Метод главных компонент позволяет уменьшить размерность полученных данных без повреждений и потерь данных.

По меньшему количеству данных можно классифицировать исследуемого - болен он или здоров.

Глава 3. Сокращение размерности сигнала

3.1. Метод главных компонент

Метод главных компонент (Principal components analysis, PCA) — один из основных способов уменьшить размерность данных, потеряв наименьшее количество информации. Данный метод является математической процедурой, целью которой является преобразование некоторого числа коррелируемых данных в меньшее число некоррелируемых, называемых главные компоненты. Первые главные компоненты содержат в себе наибольшее количество информации об исходных данных.

Метод главных компонент — это простейший метод многомерного анализа, основанный на использовании собственных значений и собственных векторов. Обычно данные методы используются для того, чтобы понять внутреннюю структуру данных. Например, если мы имеем набор многомерных данных, которые можно представить как координаты вектора (или векторов) в пространстве с большой размерности (один компонент данных вдоль каждой оси), то наш метод позволит исследователю визуализировать эти данные в 2-х и 3-х мерном пространстве, используя первые главные компоненты, и получив тем самым «тень» объекта в самом информативном ракурсе.

3.2. Структура анализа главных компонент

Под выделением признаков понимается процесс, в котором пространство данных преобразуется в пространство признаков, теоретически имеющее ту же размерность, что и исходное пространство. Обычно преобразование выполняется таким образом, чтобы пространство данных могло быть представлено сокращенным количеством «эффективных» признаков. Таким образом, остается только существенная часть информации, содержащейся в данных. Предположим, что существует некоторый вектор \mathbf{x} размерности m , который мы хотим передать с помощью l чисел, где $l < m$. Если мы просто обрежем вектор \mathbf{x} , это приведет к тому, что среднеквадратичная ошибка будет равна сумме дисперсий элементов, вырезанных из вектора \mathbf{x} . Поэтому возни-

кает вопрос: существует ли такое обратимое линейное преобразование \mathbf{T} , для которого обрезание вектора $\mathbf{T}\mathbf{x}$ будет оптимальным в смысле квадратичной ошибки? Естественно при этом преобразование \mathbf{T} должно иметь свойство маленькой дисперсии своих отдельных компонентов. Традиционный метод главных компонент максимизирует скорость уменьшения дисперсий и, таким образом, вероятность правильного выбора.

Пусть \mathbf{x} – m -мерный случайный вектор, имеющий нулевое среднее значение:

$$E(\mathbf{X}) = 0.$$

Если \mathbf{X} имеет ненулевое среднее, можно вычесть это значение еще до начала анализа. Пусть \mathbf{q} – единичный вектор размерности m , на который проектируется вектор \mathbf{X} . Эта проекция определяется как скалярное произведение векторов \mathbf{X} и \mathbf{q} :

$$A = \mathbf{X}^T \mathbf{q} = \mathbf{q}^T \mathbf{X}, \quad (3.1)$$

при ограничении

$$\|\mathbf{q}\| = (\mathbf{q}^T \mathbf{q})^{1/2} = 1. \quad (3.2)$$

Проекция A представляет собой случайную переменную со средним значением и с дисперсией, связанными со статистикой случайного вектора \mathbf{X} . В предположении, что случайный вектор \mathbf{X} имеет нулевой среднее значение, среднее значение его проекции A также будет нулевым.

$$E[A] = \mathbf{q}^T E[\mathbf{X}] = 0.$$

Таким образом, дисперсия A равна:

$$\sigma^2 = E[A^2] = E[(\mathbf{q}^T \mathbf{X})(\mathbf{X}^T \mathbf{q})] = \mathbf{q}^T E[\mathbf{X}\mathbf{X}^T] \mathbf{q} = \mathbf{q}^T \mathbf{R} \mathbf{q}. \quad (3.3)$$

Матрица \mathbf{R} является матрицей корреляции случайного вектора \mathbf{X} .

$$\mathbf{R} = E[\mathbf{X}\mathbf{X}^T], \quad \mathbf{R}^T = \mathbf{R} \quad (3.4)$$

Из выражения (3.3) видно, что дисперсия σ^2 проекции A является функцией единичного вектора \mathbf{q} . Таким образом, можно записать:

$$\psi(\mathbf{q}) = \sigma^2 = \mathbf{q}^T \mathbf{R} \mathbf{q}, \quad (3.5)$$

на основании чего $\psi(\mathbf{q})$ можно представить как дисперсионный зонд.

Следующим вопросом, подлежащим рассмотрению является поиск тех единичных векторов \mathbf{q} , для которых функция $\psi(\mathbf{q})$ имеет экстремальные или стационарные значения (локальные минимумы и максимумы) при ограниченной Евклидовой норме вектора \mathbf{q} . Решение этой задачи лежит в собственной структуре матрицы корреляции \mathbf{R} . Если \mathbf{q} – единичный вектор, такой, что дисперсионный зонд $\psi(\mathbf{q})$ имеет экстремальное значение, то для любого возмущения $\delta\mathbf{q}$ единичного вектора \mathbf{q} выполняется:

$$\psi(\mathbf{q} + \delta\mathbf{q}) = \psi(\mathbf{q}). \quad (3.6)$$

Из определения дисперсионного зонда можно вывести следующее соотношение:

$$\psi(\mathbf{q} + \delta\mathbf{q}) = (\mathbf{q} + \delta\mathbf{q})^T \mathbf{R} (\mathbf{q} + \delta\mathbf{q}) = \mathbf{q}^T \mathbf{R} \mathbf{q} + 2(\delta\mathbf{q})^T \mathbf{R} \mathbf{q} + (\delta\mathbf{q})^T \mathbf{R} \delta\mathbf{q}.$$

Игнорируя слагаемое второго порядка $(\delta\mathbf{q})^T \mathbf{R} \delta\mathbf{q}$ и используя определение (3.5), можно записать следующее:

$$\psi(\mathbf{q} + \delta\mathbf{q}) = \mathbf{q}^T \mathbf{R} \mathbf{q} + 2(\delta\mathbf{q})^T \mathbf{R} \mathbf{q} = \psi(\mathbf{q}) + 2(\delta\mathbf{q})^T \mathbf{R} \mathbf{q}. \quad (3.7)$$

Отсюда, подставляя (3.6) в (3.7) получим:

$$2(\delta\mathbf{q})^T \mathbf{R} \mathbf{q} = 0. \quad (3.8)$$

Естественно, любые возмущения $\delta\mathbf{q}$ вектора \mathbf{q} нежелательны, ограничим их только теми возмущениями, для которых норма возмущенного вектора $\mathbf{q} + \delta\mathbf{q}$ остается равной единице:

$$\|\mathbf{q} + \delta\mathbf{q}\| = 1.$$

Или, что эквивалентно,

$$(\mathbf{q} + \delta\mathbf{q})^T (\mathbf{q} + \delta\mathbf{q}) = 1.$$

Исходя из этого, в свете равенства (3.2) требуется, чтобы для возмущения первого порядка $\delta\mathbf{q}$ выполнялось соотношение

$$(\delta\mathbf{q})^T \mathbf{q} = 0. \quad (3.9)$$

Это значит, что возмущения $\delta\mathbf{q}$ должны быть ортогональны вектору \mathbf{q} и, таким образом, допускаются только изменения в направлении вектора \mathbf{q} .

Элементы единичного вектора \mathbf{q} являются безразмерными в физическом смысле. Таким образом, можно скомбинировать (3.8) и (3.9), ввести дополнительный масштабируемый множитель λ в последнее равенство с той же

размерностью, что и вхождение в матрицу корреляции **R**. После этого можно записать следующее:

$$(\delta \mathbf{q})^T \mathbf{R} \mathbf{q} - \lambda (\delta \mathbf{q})^T \mathbf{q} = 0,$$

или, эквивалентно,

$$(\delta \mathbf{q})^T (\mathbf{R} \mathbf{q} - \lambda \mathbf{q}) = 0. \quad (3.10)$$

Для того чтобы выполнялось условие (3.10), необходимо и достаточно, чтобы

$$\mathbf{R} \mathbf{q} = \lambda \mathbf{q}. \quad (3.11)$$

(3.11) – это уравнение определения таких единичных векторов **q**, для которых дисперсионный зонд $\psi(\mathbf{q})$ принимает экстремальные значения.

В уравнении (3.11) можно легко узнать задачу определения собственных значений из области линейной алгебры. Эта задача имеет нетривиальные решения только для некоторых значений λ , которые и называются собственными значениями матрицы корреляции **R**. При этом соответствующие векторы **q** называют собственными векторами. Матрица корреляции характеризуется действительными, неотрицательными собственными значениями. Соответствующие собственные векторы являются единичными (если все собственные значения различны). Обозначим собственные значения матрицы **R** размерности $m \times m$ как $\lambda_1, \lambda_2, \dots, \lambda_m$, а соответствующие им собственные векторы – $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_m$ соответственно. Тогда можно записать следующее:

$$\mathbf{R} \mathbf{q}_j = \lambda_j \mathbf{q}_j, j = 1, 2, \dots, m. \quad (3.12)$$

Пусть соответствующие собственные значения упорядочены следующим образом:

$$\lambda_1 > \lambda_2 > \dots > \lambda_j > \dots > \lambda_m. \quad (3.13)$$

При этом $\lambda_1 = \lambda_{\max}$. Пусть из соответствующих собственных векторов построена следующая матрица размерности $m \times m$:

$$\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_j, \dots, \mathbf{q}_m]. \quad (3.14)$$

Тогда систему m уравнений (3.12) можно объединить в одно большое матричное уравнение:

$$\mathbf{RQ} = \mathbf{Q}\mathbf{\Lambda}, \quad (3.15)$$

где $\mathbf{\Lambda}$ – диагональная матрица, состоящая из собственных значений матрицы корреляции \mathbf{R} .

Матрица \mathbf{Q} является ортогональной в том смысле, что векторы столбцы удовлетворяют условию ортогональности:

$$\mathbf{q}_i^T \mathbf{q}_j = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases} \quad (3.16)$$

Выражение (3.16) предполагает, что собственные значения различны. Эквивалентно, можно записать:

$$\mathbf{Q}^T \mathbf{Q} = \mathbf{I},$$

из чего можно заключить, что обращение матрицы \mathbf{Q} эквивалентно ее транспонированию:

$$\mathbf{Q}^T = \mathbf{Q}^{-1}. \quad (3.17)$$

Это значит, что выражение (3.15) можно переписать в форме, называемой ортогональным преобразованием подобия:

$$\mathbf{Q}^T \mathbf{R} \mathbf{Q} = \mathbf{\Lambda}, \quad (3.18)$$

или в расширенной форме:

$$\mathbf{q}_j^T \mathbf{R} \mathbf{q}_k = \begin{cases} \lambda_j, & k = j, \\ 0, & k \neq j. \end{cases} \quad (3.19)$$

Ортогональное преобразование подобия (3.18) трансформирует матрицу корреляции \mathbf{R} в диагональную матрицу, состоящую из собственных значений. Сама матрица корреляции может быть выражена в терминах своих собственных векторов и собственных значений следующим образом:

$$\mathbf{R} = \sum_{i=1}^m \lambda_i \mathbf{q}_i \mathbf{q}_i^T. \quad (3.20)$$

Это выражение называют спектральной теоремой. Произведение векторов $\mathbf{q}_i \mathbf{q}_i^T$ имеет ранг 1 для всех i .

Уравнения (3.18) и (3.20) являются двумя эквивалентными представлениями разложения по собственным векторам матрицы корреляции \mathbf{R} .

Анализ главных компонент и разложение по собственным векторам матрицы \mathbf{R} являются в сущности одним и тем же – различается только под-

ход к задаче. Эта эквивалентность следует из уравнений (3.5) и (3.20), из которых ясно видно равенство собственных значений и дисперсионного зонда, т.е.

$$\psi(\mathbf{q}_j) = \lambda_j, j = 1, 2, \dots, m. \quad (3.21)$$

Таким образом можно сделать выводы, касающиеся метода главных компонент.

- Собственные векторы матрицы корреляции \mathbf{R} принадлежат случайному вектору \mathbf{X} с нулевым средним и определяют единичные векторы \mathbf{q}_j , представляющие основные направления, вдоль которых дисперсионный зонд $\psi(\mathbf{q})$ принимает экстремальные значения.
- Соответствующие собственные значения определяют экстремальные значения дисперсионного зонда.

3.3. Основные представления данных

Пусть вектор данных \mathbf{x} является реализацией случайного вектора \mathbf{X} .

При наличии m возможных значений единичного вектора \mathbf{q} следует рассмотреть m возможных проекций вектора данных \mathbf{x} . В частности, согласно формуле (3.1)

$$a_j = \mathbf{q}_j^T \mathbf{x} = \mathbf{x}^T \mathbf{q}_j, \quad j = 1, 2, \dots, m,$$

где a_j – проекция вектора \mathbf{x} на основные направления, представленные единичными векторами \mathbf{q}_j . Эти проекции называют главными компонентами. Их количество соответствует размерности вектора данных \mathbf{x} .

Для того чтобы восстановить вектор исходных данных \mathbf{x} непосредственно из проекций a_j , нужно выполнить следующее. Прежде всего, объединим множество проекций $\{a_j | j = 1, 2, \dots, m\}$ в единый вектор:

$$\mathbf{a} = [a_1, a_2, \dots, a_m]^T = [\mathbf{x}^T \mathbf{q}_1, \mathbf{x}^T \mathbf{q}_2, \dots, \mathbf{x}^T \mathbf{q}_m]^T = \mathbf{Q}^T \mathbf{x}. \quad (3.22)$$

Перемножим обе части уравнения (3.22) на матрицу \mathbf{Q} , после чего используем соотношение (3.17). В результате исходный вектор данных \mathbf{x} будет

реконструирован в следующем виде:

$$\mathbf{x} = \mathbf{Q}\mathbf{a} = \sum_{j=1}^m a_j \mathbf{q}_j, \quad (3.23)$$

который можно рассмотреть как формулу синтеза. В этом контексте единичные векторы \mathbf{q}_j будут представлять собой пространства данных.

3.4. Сокращение размерности

С точки зрения задачи статистического распознавания практическое значение метода главных компонент состоит в том, что он обеспечивает эффективный способ сокращения размерности. В частности, можно сократить количество признаков, необходимых для эффективного представления данных, устраняя те линейные комбинации (3.23), которые имеют малые дисперсии, и оставляя те, дисперсии которых велики. Пусть $\lambda_1, \lambda_2, \dots, \lambda_l$ – наибольшие l собственных значений матрицы корреляции \mathbf{R} . Тогда вектор данных \mathbf{x} можно аппроксимировать, отсекая члены разложения (3.23) после l -го слагаемого:

$$\hat{\mathbf{x}} = \sum_{j=1}^l a_j \mathbf{q}_j = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_l] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_l \end{bmatrix}, l \leq m. \quad (2.24)$$

Имея исходный вектор \mathbf{x} , можно вычислить главные компоненты следующим образом:

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_l \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \\ \vdots \\ \mathbf{q}_l^T \end{bmatrix} \mathbf{x}, l \leq m. \quad (3.25)$$

Обратите внимание, что доминирующие собственные значения $\lambda_1, \lambda_2, \dots, \lambda_l$ не участвуют в вычислениях (3.24) и (3.25). Они просто определяют количество главных компонент используемых для преобразования сигнала.

Вектор ошибки аппроксимации \mathbf{e} равен разности между вектором исходных данных \mathbf{x} и вектором приближенных данных $\hat{\mathbf{x}}$:

$$\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}. \quad (3.26)$$

Подставляя (3.23) и (3.24) в (3.26), получим:

$$\mathbf{e} = \sum_{j=l+1}^m a_j \mathbf{q}_j^T \quad (3.27)$$

Вектор ошибки \mathbf{e} является ортогональным вектору приближенных данных $\hat{\mathbf{x}}$. Другими словами, скалярное произведение векторов \mathbf{e} и $\hat{\mathbf{x}}$ равно нулю. Используя (3.24) и (3.27), это свойство можно доказать следующим образом:

$$\mathbf{e}^T \hat{\mathbf{x}} = \sum_{i=l+1}^m a_i \mathbf{q}_i^T \sum_{j=1}^l a_j \mathbf{q}_j = \sum_{i=l+1}^m \sum_{j=1}^l a_i a_j \mathbf{q}_i^T \mathbf{q}_j = 0, \quad (3.28)$$

где учитывается второе условие выражения (3.16).

Общая дисперсия m компонент вектора данных \mathbf{x} составляет:

$$\sum_{j=1}^m \sigma_j^2 = \sum_{j=1}^m \lambda_j, \quad (3.29)$$

где σ_j^2 - дисперсия j -го главного компонента a_j . Общая дисперсия l элементов приближенного вектора $\hat{\mathbf{x}}$ равна:

$$\sum_{j=1}^l \sigma_j^2 = \sum_{j=1}^l \lambda_j. \quad (3.30)$$

Таким образом, общая дисперсия $(m-l)$ элементов вектора ошибки аппроксимации $\mathbf{x} - \hat{\mathbf{x}}$ равна:

$$\sum_{j=l+1}^m \sigma_j^2 = \sum_{j=l+1}^m \lambda_j. \quad (3.31)$$

Числа $\lambda_{l+1}, \dots, \lambda_m$ являются наименьшими $(m-l)$ собственными значениями матрицы корреляции \mathbf{R} . Они соответствуют слагаемым, исключенным из разложения (3.24), используемого для построения приближенного вектора $\hat{\mathbf{x}}$. Чем ближе эти собственные значения к нулю, тем более эффективным будет сокращение размерности (как результат применения метода главных компонент вектора данных \mathbf{x}) в представлении информации исходных данных.

Таким образом, для того чтобы обеспечить сокращение размерности входных данных, нужно вычислить собственные значения и векторы матрицы корреляции векторов входных данных, а затем ортогонально проецировать эти данные на подпространство, задаваемое собственными векторами, соответствующими доминирующим собственным значениям этой матрицы.

Глава 4. Методы классификации сокращенных сигналов многоканальных ЭКГ.

4.1. Классификатор на основе анализа главных компонент

Структура классификатора

После вычисления главных осей тренировочного набора, вектор признаков тестового объекта проецируется на подпространство, образованное главными осями. Вычисляются две характеристики: Distance in Feature Space (DIFS) и Distance From Feature Space (DFFS). Исходя из этих характеристик выносится решение о принадлежности тестового объекта классу, образованному тренировочным набором.

Первая характеристика - DIFS – Distance in Feature Space позволяет найти расстояние от проекции тестового вектора до среднего вектора тренировочного набора. Это расстояние называется дистанцией Махаланобиса в подпространстве.

$$DIFS(x) = \sum_i \frac{y_i^2}{\lambda_i} \quad (4.1)$$

DIFS определяет в подпространстве F концентрические эллипсы, точки которых равноудалены от математического ожидания (в определении дистанции Махаланобиса). Найдя максимальный параметр DIFS у векторов исходной выборки мы определим границу нашего класса. В него входя элементы с меньшим параметром DIFS.

4.2. Статистический одноклассовый классификатор SVDD

Рассмотрим задачу классификации данных по признаку принадлежности, или не принадлежности определенному классу, который в свою очередь определяется определенным набором данных (тренировочным набором).

Обычно задача одно классовой классификации сложнее, чем задача об двух классовой классификации, так как в данном случае граница определяется “с двух сторон”, путем оценки принадлежности тестируемого объекта к обоим классам. В нашем случае нам доступен лишь один класс объектов, поэтому возникает вопрос, насколько плотно граница допуска объектов к заданному классу должна прилегать к тренировочным данным по каждому измерению.

В данном разделе будет рассмотрен одно классовый классификатор SVDD(Support Vector Data Description). Данный метод позволяет выделить набор объектов из тренировочного набора данных, которые называются SV (Support Vectors) количество данных векторов определяет, насколько сложной будет граница, определяющая принадлежность объектов к заданному классу. Минимальное количество векторов (SV) образует сферическую границу, по мере увеличения их числа границ принимает более сложные формы.

Алгоритм SVDD

Метод SVDD был разработан Дэвидом Таксом и Робертом Дуином.

В данном методе мы размещаем данные внутри гиперсферы с минимальным объемом. Минимизируя радиус сферы, мы пытаемся минимизировать шанс ошибочного допуска тестируемого объекта к классу, определяемому тренировочным набором. Введем понятие структурной ошибки:

$$\varepsilon_{struct}(R, a) = R^2 \quad (4.2)$$

Где a – это центр сферы. Далее это выражение (1,4) нужно минимизировать с ограничением:

$$\|x_i - a\|^2 \leq R^2, \forall i \quad (4.3)$$

Ослабим условие (4.4), введя набор переменных ξ :

$$\varepsilon_{struct}(R, a, \xi) = R^2 + C \sum_i \xi_i \quad (4.4)$$

где выбором константы C мы сможем найти компромисс между объемом сферы и полученной ошибкой.

Таким образом, получим ограничение, при котором почти все объекты будут лежать внутри сферы:

$$\|x_i - a\|^2 \leq R^2 + \xi, \quad \xi \leq 0 \quad \forall i \quad (4.5)$$

Путем введения множителей Лагранжа α, γ и построения Лагранжиана из равенства (4.6) с учетом ограничений (4.7) получим:

$$\begin{aligned} L(R, a, \xi, \alpha, \gamma) = \\ = R^2 + C \sum_i \xi_i - \sum_i \alpha_i (R^2 + \xi_i - (\|x_i\|^2 - 2ax_i + \|a\|^2)) - \sum_i \gamma_i \xi_i \end{aligned} \quad (4.6)$$

Приравняв частные производные выражения (4.5) по R, a и ξ к 0, получим следующие ограничения:

$$\sum_i \alpha_i = 1 \quad (4.7)$$

$$a = \sum_i \alpha_i x_i \quad (4.8)$$

$$C - \alpha_i - \gamma_i = 0 \quad (4.9)$$

Применяя выражения (4.7 – 4.9) к выражению (4.8) получим выражение для лагранжиана L :

$$L = \sum_i \alpha_i (x_i \cdot x_i) - \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) \quad (4.10)$$

Максимизация (4.12) с заданными ограничениями (4.7) и (4.9) даст нам набор α_i . Объекты x_i , для которых $\alpha_i > 0$ называются опорными векторами. Опорные вектора лежат на границе (если $0 < \alpha_i < C$) или за границей (если $\alpha_i = C$). Выражение (4.10) показывает, что центр сферы является линейной комбинацией опорных векторов. Задача максимизации является задачей квадратичного программирования с линейными ограничениями. Существует набор стандартных алгоритмов для её решения, например алгоритм Франка-Вульфа.

Процесс классификации

Определив набор $\{\alpha_i\}$ мы сможем определить функцию, задающую границу для задачи одно классовой классификации, и тем самым определяя принадлежность нового объекта исследуемого классу:

$$f_{SVDD}(z, \alpha, R) = I(|z - a|^2 \leq R^2) =$$

$$= I\left((z \cdot z) - 2 \sum_i \alpha_i (z \cdot x_i) + \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) \leq R^2\right) \quad (4.11)$$

Где z - тестируемый объект, а I – функция, определенная как:

$$I(A) = 1 \Leftrightarrow A \text{ верно}$$

$$I(A) = 0, \text{ иначе} \quad (4.12)$$

Support Vectors – это те элементы тренировочного набора данных, для которых $\alpha=0$. Таким образом, только эти элементы дадут ненулевые слагаемые в суммах.

Функция ядра.

Гиперсфера – очень жесткая граница вокруг данных, и обычно не дает

правильное представление об их структуре. Идея метода SVDD – нелинейное отображение данных из тренировочного набора в пространство с большой размерностью, и построение разделяющей гиперплоскости в этом пространстве. Тем самым мы получим нелинейные границы в исходном пространстве. С использованием функции ядра мы сможем вычислить разделяющую гиперплоскость без конкретного отображения данных в пространство большой размерности. Введем функцию ядра:

$$K(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j)) \quad (4.13)$$

Функцией ядра может быть любая функция, удовлетворяющая теореме Мерсера. В данной работе мы будем использовать ядро Гаусса:

$$K(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|^2}{s^2}\right) \quad (4.14)$$

Заменяя скалярные произведения в выражении (9) получим новое выражение для Лангранжиана

$$L = -\sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) \quad (4.15)$$

Выражение (4.11) так же примет другой вид:

$$f_{SVDD}(z, \alpha, R) = I\left(\sum_i \alpha_i \exp\left(-\frac{|z - x_i|^2}{s^2}\right) > \frac{1}{2}(B - R^2)\right) \quad (4.16)$$

где

$$B = 1 + \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j). \quad (4.17)$$

Выражение в левой части неравенства в выражении (4.24) по сути является взвешенной суммой Гауссианов $K(z, x_i)$. Таким образом, граница строго зависит от параметра s . При маленьких значениях s обычно все объекты представляют собой тренировочный набор. Для больших значений s мы получим границу в виде гиперсферы.

С увеличением значения s , число векторов с ненулевыми α_i уменьшается. Когда тренировочный набор представляет собой верное распределение данных, количество тренировочных векторов позволяет дать оценку ошибки при проверке очередного вектора. С увеличением s увеличивается число векторов, и, соответственно, уменьшается ошибка. Однако, когда граница допуска новых объектов шире, то увеличивается риск принять ложный объект.

Глава 5. Программная реализация и обсуждение результатов.

Была разработано программное обеспечение, которое обрабатывает распределение плотности токов полученных в результате проведения магнитокардиографии.

С помощью метода главных компонент уменьшаем размерность полученных данных, далее вычисляя характеристику - DIFS , классифицируем пациентов на больных и здоровых.

Метод главных компонент был реализован с помощью библиотеки `gs1` на языке программирования `c++`.

В методе `debug_show_data` находим матрицу корреляции. Далее в методе `debug_show_cov` находим собственные значения матрицы, далее по критерию снижения размерности отбираем их и расставляем их в порядке убывания.

В методе `calculate_eigen_data()` находим собственные вектора.

Далее в методе `calculate()` получаем на выход матрицу с главными компонентами по диагонали, а во всех остальных ячейках нули.

Одноклассовый классификатор SVDD был реализован с помощью библиотеки `matlab` на языке программирования `c++`.

В библиотеке `matlab` реализован метод франка-вульфа с помощью него находим набор $\{\alpha_i\}$.

Далее в методе `I_function` вычисляем формулу (4.16) и в методе `function_1` вычисляем формулу (4.11).

Возьмем данных о пациентах из двух групп: больные (группа «больные») и здоровые (группа «норма»).

Исходные данные: распределения плотности токов полученные в результате магнитокардиографии выглядят следующим образом:

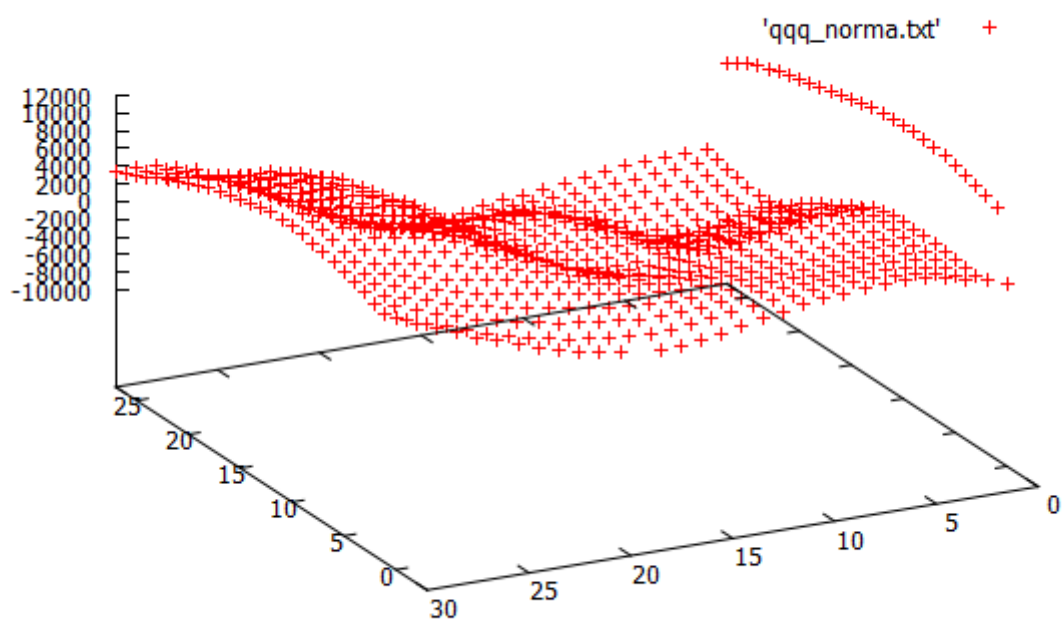


Рис. 6 Данные из группы «норма».

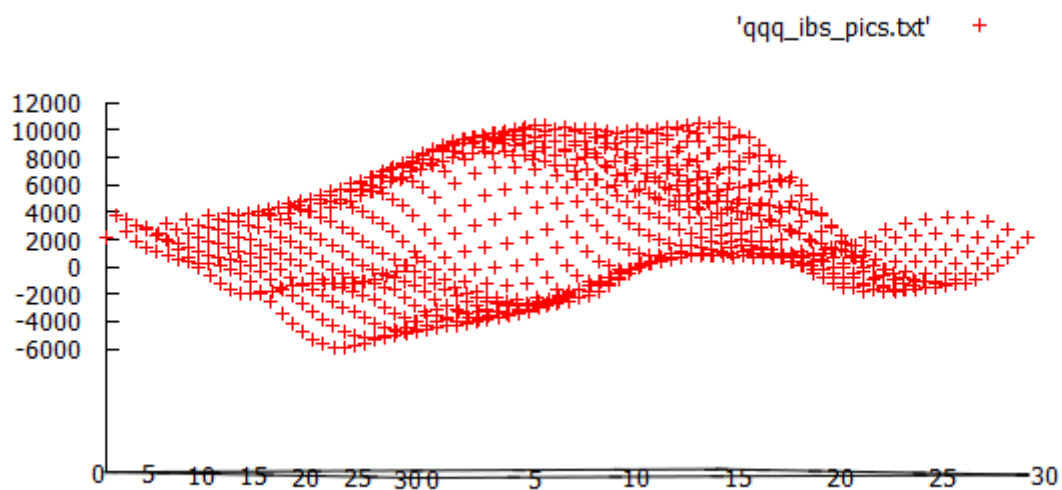


Рис.7 Данные из группы «больные».

Результаты обработки данных получены у 20 пациентов в таблице ниже приведены данные о 3-х здоровых пациентов(группа «норма») и трех пациентов (группа «больные»), приведены в таблице:

Таблица 5.1

Значения главных компонент

номер компонент	Значения компонент для групп пациентов					
	группа "норма"			группа "больные"		
	norma1	norma2	norma3	ibs-pics-1	ibs-pics-2	ibs-pics-3
1	1	1	1	1	1	1
2	0,212637017	0,4934312	0,461367	0,432741938	0,754322	0,376987515
3	0,075982041	0,12117989	0,043754	0,337942701	0,089835	0,241095802
4	0,016088592	0,00432047	0,027271	0,053110443	0,014727	0,049979462
5	0,009327216	0,0017046	0,003212	0,002267434	0,000802	0,006042402
6	0,000493023	0,00060245	0,002517	0,001488774	0,000395	0,001070545
7	0,000084	0,000156	0,001354	0,000208	0,000137	0,000122
8	0,000024		0,000916	0,000054	0,000028	0,000013
9			0,000111	0,000005	0,000003	
10			0,000003			

Ниже приведены графики для каждого пациента из группы «норма» и группы «больные»

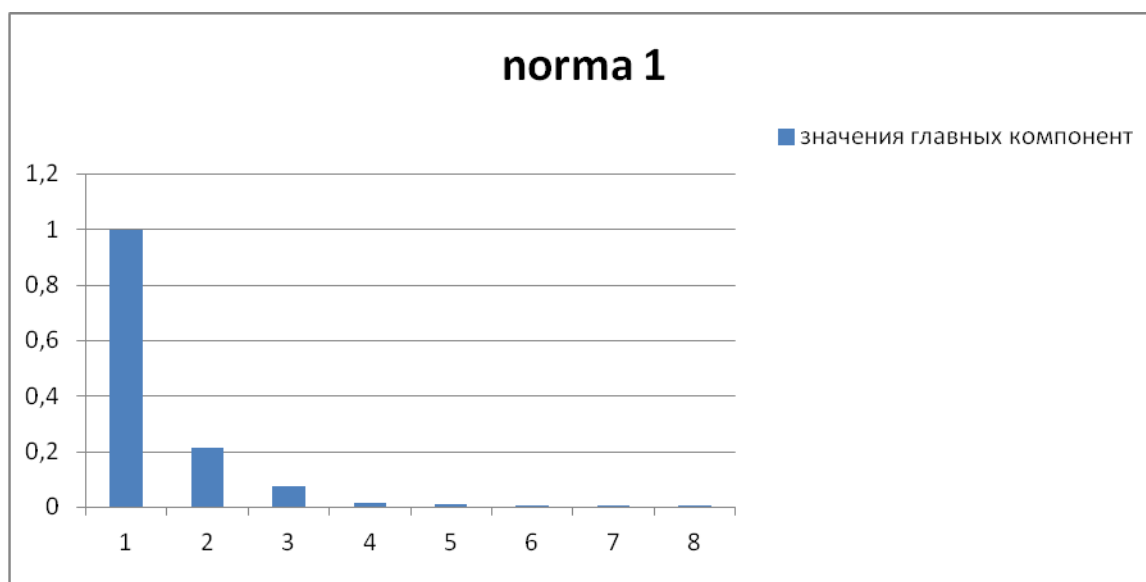


Рис.8 Значения главных компонент для пациента1 (из группы «норма»).

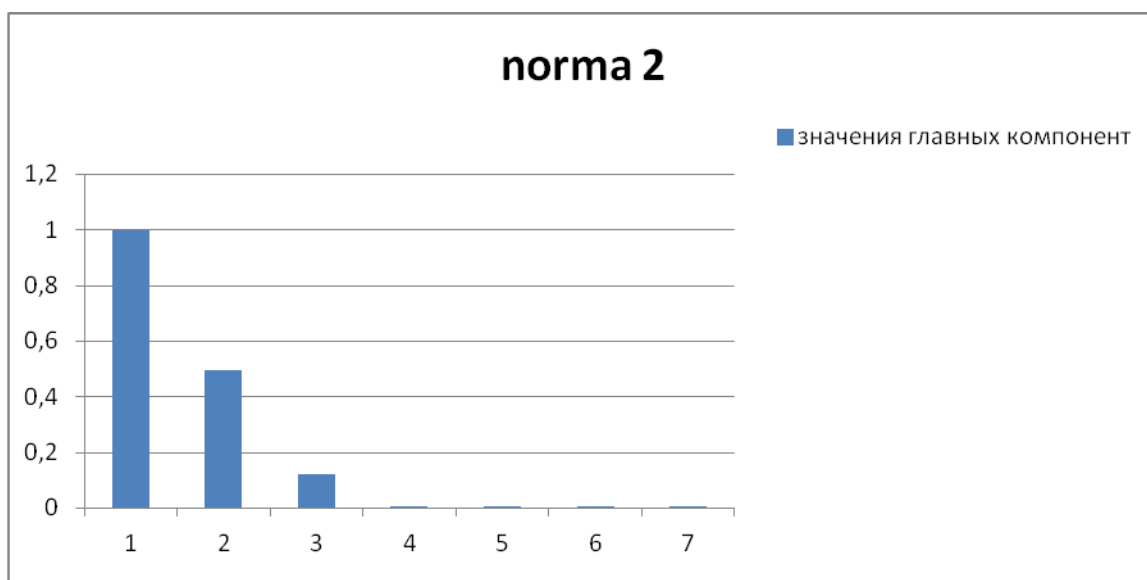


Рис.9 Значения главных компонент для пациента2 (из группы «норма»).

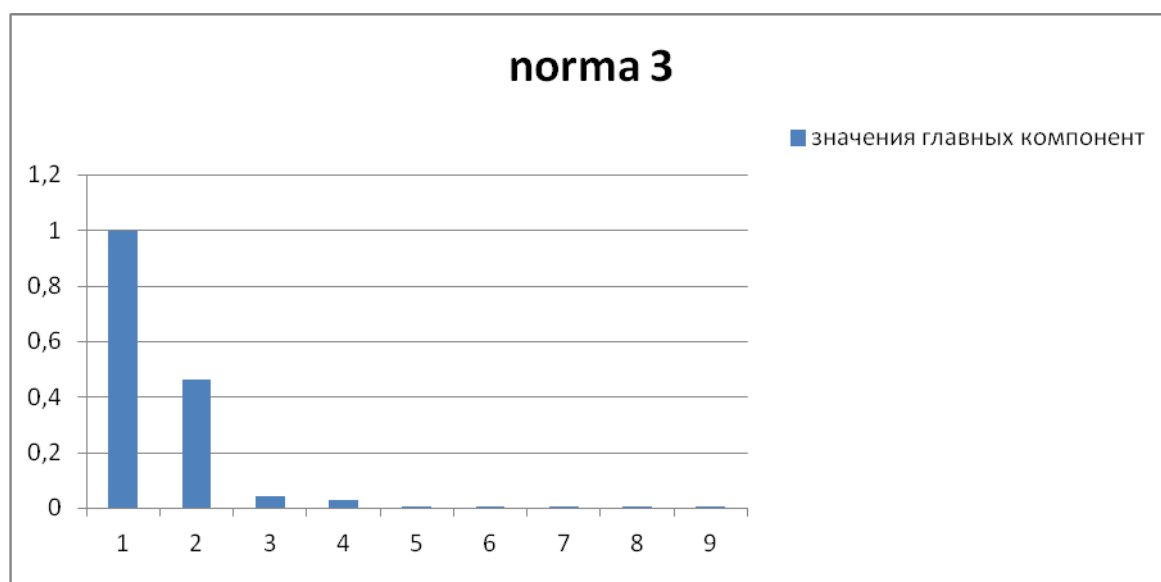


Рис.10 Значения главных компонент для пациента3 (из группы «норма»).

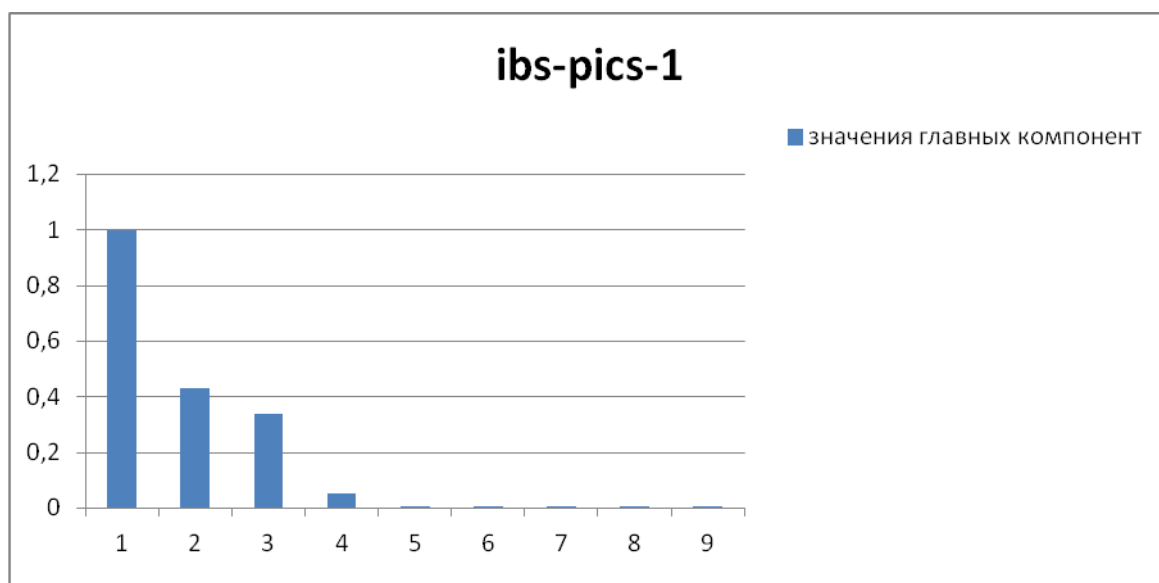


Рис.11 Значения главных компонент для пациента4 (из группы «больные»).

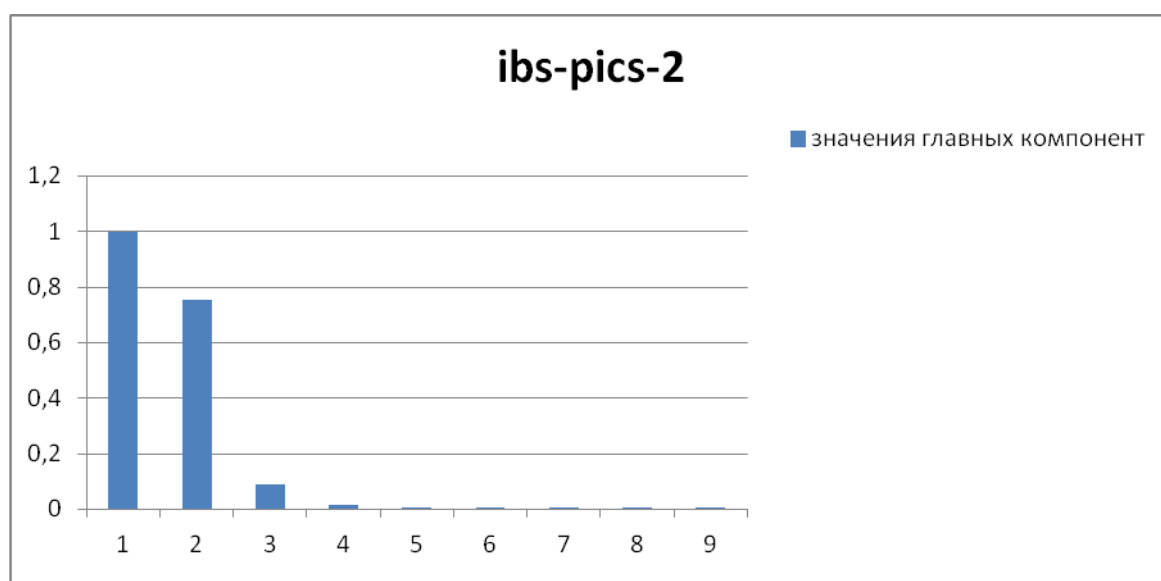


Рис.12 Значения главных компонент для пациента5 (из группы «больные»).

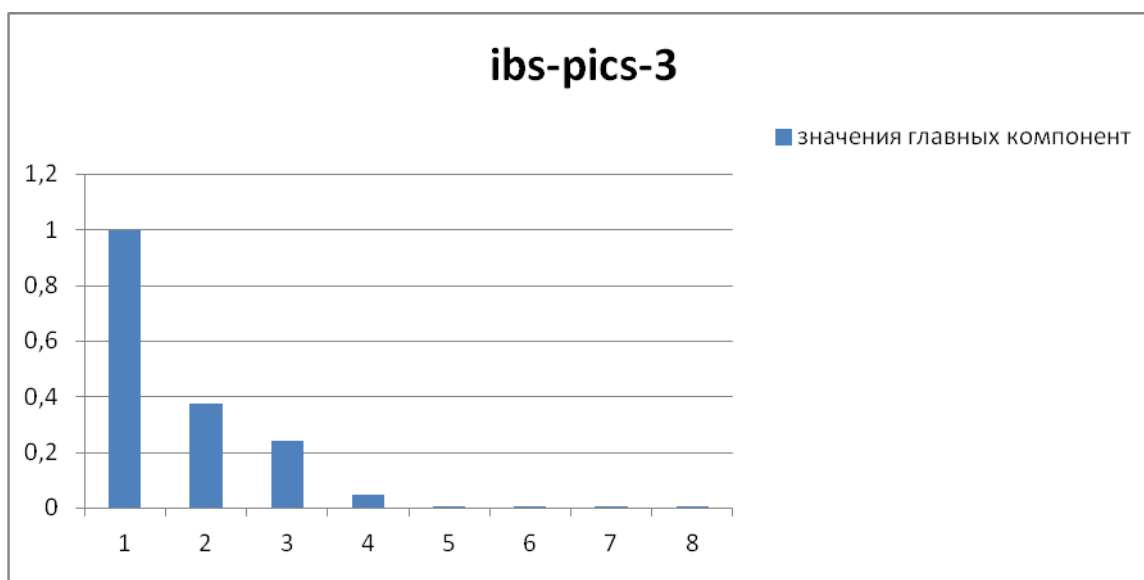


Рис.13 Значения главных компонент для пациента5 (из группы «больные»).

После того, как уменьшили размерность исходных данных, получаем 7-10 компонент, по которым можно классифицировать пациент относится к группе «норма» или к группе «больные». Для этого надо вычислить характеристику-DIFS. По ее значениям отчетливо понятно, какой пациент к какой группе относится.

Таблица 5.2
Значения DIFS

№ п/п	difs	
	norma	ips-pics
1	639	-4 790
2	1 750	-1 830
3	5 600	-1 608
4	5 150	-4 300
5	1 020	-2 170

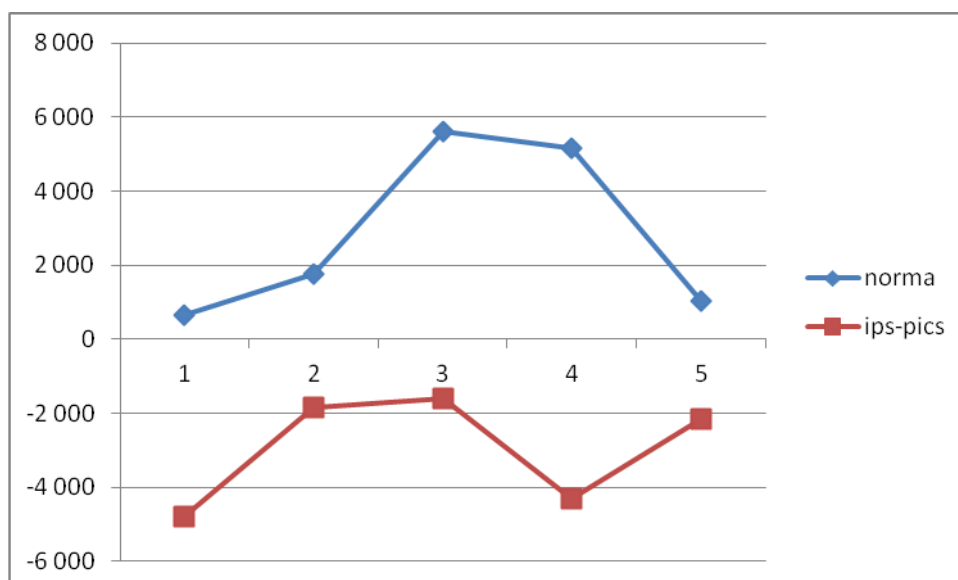


Рис.14 Значения DIFS.

Также можно классифицировать пациентов, используя одноклассовый классификатор SVDD. После проведения классификации также видно, какой пациент к какой группе относится.

Таблица 5.3
Значения SVDD

	SVDD		
	c=2,s=0,6	c=2,s=0,8	c=2,s=1,2
Обработано пациентов из группы "норма"	7	10	11
Обработано пациентов из группы "больные"	8	8	5

Выводы

Была разработано программное обеспечение, которое обрабатывает распределение плотности токов полученных в результате проведения магнитокардиографии и классифицирует их.

Для уменьшения размерности данных, полученных в результате проведения магнитокардиографии был применен метод главных компонент и проведена классификация больных.

Показано, что из большого количества данных (961 точка) можно получить 7-10 компонент, по которым можно проводить классификацию.

Разработана модель одноклассового статистического классификатора на основе главных компонент и одноклассовый классификатор SVDD для анализа многоканальной ЭКГ, и проведена оценка его эффективности для контрольных групп.

Литература

- 1) М. А. Примин, И.В. Недавода, Ю.В. Масленников, Ю.В. Гуляев Магнитокардиографический комплекс для раннего выявления и мониторинга заболеваний сердца: Программное обеспечение/Радиотехника и электроника, 2010, том 55, №10, с.1250-1269
- 2) Гантмахер Ф. Р. Теория матриц // М.: Наука, 1966.
- 3) Претт У. Цифровая обработка изображений. Том 2. // М.: Мир, 1982
- 4) Menser B., Muller F. Face detection in color images using principal component analysis // Image Processing and Its Applications, 1999, vol. 1, p. 620-624
- 5) Tax D., R. Duin, Support vector data description / Machine Learning, 2004, Vol. 54, p. 45-66.
- 6) Turk M., Pentland A. Eigenspaces for recognition.// Journal of cognitive neuroscience, 1991, vol 3, 71-86.
- 7) Mahalanobis, P C. On the generalised distance in statistics // Proceedings of the National Institute of Sciences of India **2** (1), 1936, 49–55.
- 8) Tax D., Duin R. Support vector data description. // Pattern recognition letters, 1999, vol. 20, 1191-1199.
- 9) Vilaplama V., F. Marques. 13th European Signal Processing Conference, September, 4-8, 2005. Support vector data description based on PCA Features for face detection. Potential Maps. 2005.

