

Инженерная школа информационных технологий и робототехники
 Направление подготовки 09.03.02 «Информационные системы и технологии»
 Отделение информационных технологий

БАКАЛАВРСКАЯ РАБОТА

Тема работы
Разработка модуля поддержки принятия решений в процессе диагностики функционального состояния системы гемостаза

УДК 681.518.22:519.81:616-005.1-08(571.16)

Студент

Группа	ФИО	Подпись	Дата
8И5Б	Чурсина Елена Андреевна		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Кудинов Антон Викторович	к.т.н.		

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Шулинина Юлия Игоревна			

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Немцова Ольга Александровна			

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Цапко Ирина Валериевна	к.т.н.		

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ООП

Код результатов	Результат обучения (выпускник должен быть готов)
Профессиональные и общепрофессиональные компетенции	
P1	Применять базовые и специальные естественнонаучные и математические знания для комплексной инженерной деятельности по созданию, внедрению и эксплуатации геоинформационных систем и технологий, а также информационных систем и технологий в бизнесе.
P2	Применять базовые и специальные знания в области современных информационных технологий для решения инженерных задач.
P3	Ставить и решать задачи комплексного анализа, связанные с созданием геоинформационных систем и технологий, информационных систем в бизнесе, с использованием базовых и специальных знаний, современных аналитических методов и моделей.
P4	Выполнять комплексные инженерные проекты по созданию информационных систем и технологий, а также средств их реализации (информационных, методических, математических, алгоритмических, технических и программных).
P5	Проводить теоретические и экспериментальные исследования, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, проведение эксперимента, анализ и интерпретация полученных данных, в области создания геоинформационных систем и технологий, а также информационных систем и технологий в бизнесе.
P6	Внедрять, эксплуатировать и обслуживать современные геоинформационные системы и технологии, информационные системы и технологии в бизнесе, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья, безопасность труда, выполнять требования по защите окружающей среды.
Универсальные (общекультурные) компетенции	
P7	Использовать базовые и специальные знания в области проектного менеджмента для ведения комплексной инженерной деятельности.
P8	Осуществлять коммуникации в профессиональной среде и в обществе в целом. Владеть иностранным языком (углублённый английский язык), позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты комплексной инженерной деятельности.
P9	Эффективно работать индивидуально и в качестве члена команды, состоящей из специалистов различных направлений и квалификаций,
P10	Демонстрировать личную ответственность за результаты работы и готовность следовать профессиональной этике и нормам ведения комплексной инженерной деятельности.
P11	Демонстрировать знания правовых, социальных, экологических и культурных аспектов комплексной инженерной деятельности, а также готовность к достижению должного уровня физической подготовленности для обеспечения полноценной социальной и профессиональной деятельности.

Инженерная школа информационных технологий и робототехники
Направление подготовки 09.03.02 «Информационные системы и технологии»
Отделение информационных технологий

УТВЕРЖДАЮ:
Руководитель ООП

(Подпись) _____ (Дата) Цапко И.В.
(Ф.И.О.)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

В форме:

Бакалаврской работы

Студенту:

Группа	ФИО
8И5Б	Чурсина Елена Андреевна

Тема работы:

Разработка модуля поддержки принятия решений в процессе диагностики функционального состояния системы гемостаза	
Утверждена приказом директора (дата, номер)	18.02.2019 № 1240/с

Срок сдачи студентом выполненной работы:	
------------------------------------------	--

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе	Работа направлена на разработку модуля поддержки принятия решений в процессе диагностики функционального состояния системы гемостаза.
Перечень подлежащих исследованию, проектированию и разработке вопросов	<ul style="list-style-type: none"> – Анализ предметной области. – Разработка алгоритма классификации. – Проектирование модуля на основе разработанного алгоритма. – Разработка модуля на основе разработанного алгоритма. – Внедрение модуля в программное обеспечение тромбоэластографа. – Финансовый менеджмент. – Социальная ответственность.
Перечень графического материала	
Консультанты по разделам выпускной квалификационной работы	
Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Шулинина Юлия Игоревна
Социальная ответственность	Немцова Ольга Александровна

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	21.01.2019
-------------------------------------------------------------------------------------------------	------------

Задание выдал руководитель / консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Кудинов Антон Викторович	к.т.н		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8И5Б	Чурсина Елена Андреевна		

Инженерная школа информационных технологий и робототехники
 Направление подготовки 09.03.02 «Информационные системы и технологии»
 Уровень образования – бакалавриат
 Отделение информационных технологий
 Период выполнения осенний / весенний 2018/2019 учебного года

Форма представления работы:

Бакалаврская работа

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
21.01.2018	Анализ предметной области	10
01.02.2018	Аналитический обзор	15
25.02.2019	Разработка классификатора	20
13.04.2019	Разработка модуля поддержки принятия решений	20
15.05.2019	Внедрение модуля	15
28.05.2019	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	10
03.06.2019	Социальная ответственность	10

СОСТАВИЛ:

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Кудинов Антон Викторович	к.т.н		

СОГЛАСОВАНО:

Руководитель ООП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Цапко Ирина Валериевна	к.т.н		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И
РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

Группа		ФИО	
8И5Б		Чурсиной Елене Андреевне	
Школа	ИШИТР	Отделение школы (НОЦ)	ОИТ
Уровень образования	Бакалавриат	Направление/специальность	09.03.02 Информационные системы и технологии

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. <i>Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	Оклад ассистента – 21760 руб. Оклад руководителя – 33664 руб.
2. <i>Нормы и нормативы расходования ресурсов</i>	Премияльный коэффициент 30%. Коэффициент доплат и надбавок 20%. Районный коэффициент 30%. Коэффициент дополнительной заработной платы 12%. Накладные расходы 16%.
3. <i>Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	Коэффициент отчислений на уплату во внебюджетные фонды 30%.

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. <i>Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения</i>	– Сравнение конкурентных технических решений. – Многоугольник конкурентоспособности.
2. <i>Планирование и формирование бюджета научных исследований</i>	Формирование плана и графика разработки: – определение структуры работ; – определение трудоемкости работ; – разработка графика Ганта. Формирование бюджета затрат на научное исследование: – материальные затраты; – заработная плата (основная и дополнительная); – отчисления на социальные цели; – накладные расходы.
3. <i>Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования</i>	– Определение потенциального эффекта исследования.

Перечень графического материала (с точным указанием обязательных чертежей):

1. <i>Оценочная карта для сравнения конкурентных технических решений</i> 2. <i>Многоугольник конкурентоспособности</i> 3. <i>Матрица SWOT</i> 4. <i>График Ганта</i> 5. <i>Расчет бюджета затрат</i>

Дата выдачи задания для раздела по линейному графику

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент ОСГН ШБИП	Шулинина Ю.И.			

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8И5Б	Чурсина Е.А.		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»**

Студенту:

Группа	ФИО
8И5Б	Чурсиной Елене Андреевне

Школа	ИШИТР	Отделение школы (НОЦ)	Информационных технологий
Уровень образования	Бакалавриат	Направление/специальность	09.03.02 Информационные системы и технологии

Исходные данные к разделу «Социальная ответственность»:	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Объект исследования – модуль поддержки принятия решений в процессе оценки состояния системы гемостаза. Разработка и эксплуатация ПО, в состав которого входит модуль, осуществляется на ПЭВМ.
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. Правовые и организационные вопросы обеспечения безопасности: 1.1 специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства. 1.2 организационные мероприятия при компоновке рабочей зоны.	Основные проводимые правовые и организационные мероприятия по обеспечению безопасности трудящихся в аудиториях. Анализ правильного расположения и организации рабочего места, а также режима работы.
2. Производственная безопасность: 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	Выделение и анализ следующих вредных и опасных факторов: – отклонение показателей микроклимата; – превышение уровня шума на рабочем месте; – недостаточная освещенность рабочей зоны; – нарушение предельно допустимых значений напряжений прикосновения и токов.
3. Экологическая безопасность:	Анализ влияния процесса разработки на атмосферу, гидросферу и литосферу.
4. Безопасность в чрезвычайных ситуациях:	Составление и анализ перечня возможных чрезвычайных ситуаций. Определение мер по превентивному противодействию и реагированию при пожаре.

Дата выдачи задания для раздела по линейному графику	
-------------------------------------------------------------	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Немцова Ольга Александровна			

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8И5Б	Чурсина Елена Андреевна		

РЕФЕРАТ

Выпускная квалификационная работа: 84 с., 17 рис., 22 табл., 31 источника, 8 прил.

Ключевые слова: гемостаз, тромбоэластограф, низкочастотная пьезотромбоэластография, машинное обучение, классификация, случайный лес, искусственная нейронная сеть.

Объектом исследования являются данные об исследованиях крови человека, проведённые с помощью тромбоэластографа.

Цель работы – разработка модуля поддержки принятия решений в процессе диагностики функционального состояния системы гемостаза, а также внедрение разработанного модуля в программное обеспечение тромбоэластографа.

В результате исследования был разработан классификатор для диагностики гемостазиологических расстройств. Данный классификатор был реализован в модуле поддержки принятия решений в процессе диагностики функционального состояния системы гемостаза. Апробация модуля на данных исследования крови пациентов показала высокие значения метрик, определяющих точность и устойчивость алгоритма.

Основные конструктивные, технологические и технико-эксплуатационные характеристики: модуль, разработанный на языке программирования Python и использующий библиотеки pandas и scikit-learn.

Степень внедрения: внедрено в программное обеспечение клинико-диагностического комплекса АРП-01М «МЕДНОРД».

Область применения: диагностика гемостазиологических расстройств.

Экономическая значимость работы заключается в ускорении и улучшении качества процесса диагностики гемостазиологических расстройств.

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

- ПО** – программное обеспечение;
- НПТЭГ** – низкочастотная пьезотромбоэластография;
- ИНС** – искусственная нейронная сеть;
- ОС** – операционная система;
- ЯП** – язык программирования;
- о.е.** – относительные единицы;
- мин.** – минимум;
- макс.** – максимум.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	14
1 Анализ предметной области	16
1.1 Оценка состояния системы гемостаза человека с помощью тромбоэластографа.....	16
1.1.1 Обзор метода низкочастотной пьезотромбоэластографии.....	16
1.1.2 Диагностика гемостазиологических расстройств на основе результатов НПТЭГ	17
1.2 Обзор аналогов модуля поддержки принятия решений в процессе оценки состояния системы гемостаза	18
1.3 Теоретические основы	18
1.3.1 Постановка задачи классификации	18
1.3.2 Обзор классификаторов.....	19
1.3.2.1 Искусственная нейронная сеть.....	20
1.3.2.2 Случайный лес	20
1.3.2.3 Градиентный бустинг	21
1.3.2.4 Многоэтапные классификаторы.....	21
1.3.3 Настройка классификаторов	21
1.3.3.1 Алгоритм обратного отбора признаков.....	22
1.3.3.2 Анализ матрицы корреляции	22
1.3.3.3 Поиск по решётке.....	22
1.3.4 Оценка качества работы классификатора	23
1.3.4.1 Матрица несоответствий.....	23
1.3.4.2 Точность тестовой выборки.....	23
1.3.4.3 Точность перекрёстной проверки	23

1.3.5	Оценка важности признаков в процессе классификации	24
1.4	Программное обеспечение тромбозластографа.....	24
1.5	Программные инструменты	25
2	Разработка классификатора.....	26
2.1	Анализ данных исследований крови с помощью тромбозластографа ..	26
2.2	Исследование классификаторов	27
2.2.1	Алгоритм настройки классификаторов	28
2.2.2	Выбор метрик	28
2.2.3	Классификатор на основе ИНС	29
2.2.4	Классификаторы на основе градиентного бустинга и случайного леса	30
2.2.5	Двухэтапный классификатор.....	31
2.2.5.1	Структура двухэтапного классификатора.....	31
2.2.5.2	Настройка внутренних классификаторов двухэтапного классификатора	32
2.2.5.3	Классификатор первого этапа на основе градиентного бустинга.....	33
2.2.5.4	Классификатор структуры на основе градиентного бустинга.....	34
2.2.5.5	Классификатор хронометрии на основе случайного леса	35
2.2.5.6	Формирование диагноза в случае ошибок классификатора первого этапа.....	36
2.2.5.7	Результаты тестирования	36
2.2.6	Результаты исследования.....	37
2.3	Анализ ошибок двухэтапного классификатора	37

2.4	Трёхэтапный классификатор.....	38
2.4.1	Структура трёхэтапного классификатора	38
2.4.2	Настройка внутренних классификаторов трёхэтапного классификатора	39
2.4.3	Уточнение структуры третьего этапа трёхэтапного классификатора	40
3	Разработка модуля поддержки принятия решений	42
3.1	Проектирование модуля поддержки принятия решений.....	42
3.1.1	Описание архитектуры модуля поддержки принятия решений .	42
3.1.2	Описание архитектуры компонента TegClassifier.....	43
3.2	Описание компонентов Train и Predict	44
4	Внедрение модуля в ПО тромбоэластографа	46
4.1	Взаимодействие с компонентом «Train»	46
4.2	Взаимодействие с компонентом «Predict».....	46
5	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение...	48
6	Социальная ответственность	61
	ЗАКЛЮЧЕНИЕ	69
	СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ.....	70
	ПРИЛОЖЕНИЕ А Описание показателей НПТЭГ и их характеристик для набора данных	74
	ПРИЛОЖЕНИЕ Б Рекомендации к диагностике гемостазиологических расстройств	76
	ПРИЛОЖЕНИЕ В Матрица корреляции параметров НПТЭГ	77
	ПРИЛОЖЕНИЕ Г Результаты настройки классификаторов на основе градиентного бустинга и случайного леса	78
	ПРИЛОЖЕНИЕ Д Схемы трёхэтапного классификатора	80

ПРИЛОЖЕНИЕ Е Диаграмма Ганта.....	82
ПРИЛОЖЕНИЕ Ж Акт ввода в эксплуатацию	83
ПРИЛОЖЕНИЕ З Награды по теме работы.....	84

ВВЕДЕНИЕ

Система гемостаза – совокупность биологических и биохимических механизмов, обеспечивающих сохранение жидкого состояния циркулирующей крови, поддержание целостности кровеносных сосудов и купирование кровотечения при их повреждениях [1]. Нарушения в работе любого звена системы гемостаза представляют опасность для здоровья и жизни человека.

Современным способом выявления гемостазиологических расстройств является использование тромбоэластографа. ПО тромбоэластографа представляет результаты исследования в виде графика, отражающего динамику исследуемого процесса, и набора дополнительно регистрируемых параметров. Несмотря на то, что процесс анализа результатов требует глубокого понимания метода исследования, в ПО тромбоэластографа не всегда входят программные модули для помощи в постановке диагноза. Это замедляет обучение специалистов и делает его затруднительным в отдаленных областях, где тромбоэластографы до этого не применялись. Также в процессе постановки диагноза на основе графической и числовой информации может быть полезным дополнительное мнение. Для помощи врачу в быстрой и точной диагностике ПО тромбоэластографа требуется модуль поддержки принятия решений.

Целью данной работы является разработка модуля поддержки принятия решений в процессе диагностики функционального состояния системы гемостаза, а также внедрение модуля в ПО тромбоэластографа.

Для достижения данной цели необходимо выполнить следующие задачи:

- провести анализ предметной области;
- разработать алгоритм классификации диагнозов;
- разработать и выполнить программную реализацию модуля;
- внедрить модуль в ПО тромбоэластографа;

- составить отчетную документацию о выполненной работе.

Практическая значимость разработанного модуля состоит в помощи пользователям тромбоэластографа в своевременном и точном выявлении гемостазиологических расстройств.

1 Анализ предметной области

1.1 Оценка состояния системы гемостаза человека с помощью тромбоэластографа

Тромбоэластограф – медицинский прибор, позволяющий оценить состояние системы гемостаза в кратчайшие сроки без необходимости проведения дорогостоящих лабораторных коагуляционных тестов. Тромбоэластограф включен в перечень приборов, оборудования и медицинского инструментария для клиничко-диагностических лабораторий лечебно-профилактических учреждений [2].

1.1.1 Обзор метода низкочастотной пьезотромбоэластографии

В основе работы тромбоэластографа лежит метод низкочастотной пьезотромбоэластографии (НПТЭГ) цельной крови. Динамика изучаемого процесса определяется изменениями агрегатного состояния исследуемого образца цельной крови. Результат регистрируется в виде интегрированной кривой, каждая точка которой A_i является изменением значения свойств крови в определенный момент времени T_i . [3].

На рисунке 1 представлен пример графика изменения показателей НПТЭГ здорового пациента. По оси ординат – амплитуда исследуемого процесса в относительных единицах, по оси абсцисс – время исследования в минутах.

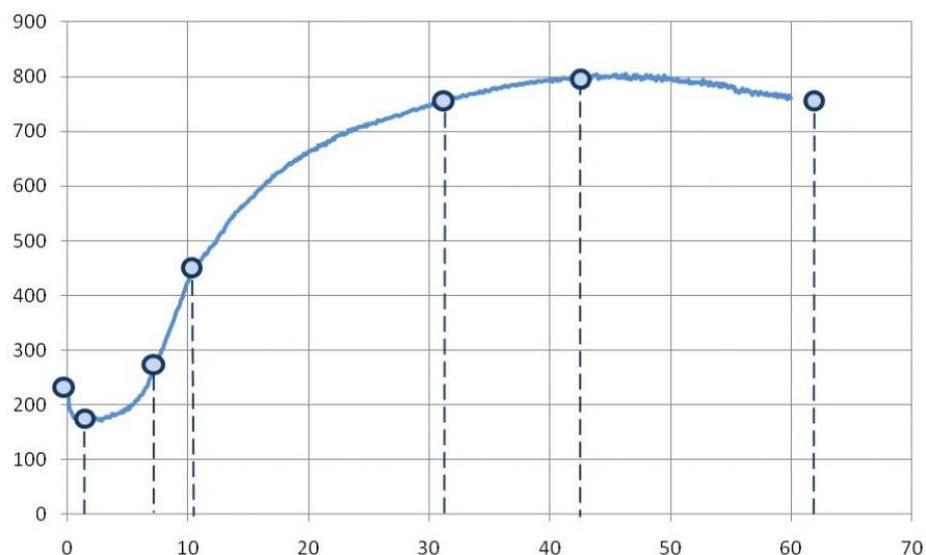


Рисунок 1 – График изменения значения свойств крови НПТЭГ цельной крови здорового человека

В результате исследования образца цельной крови человека записываются 20 числовых параметров НПТЭГ, представленных в таблице 19 приложения А.

1.1.2 Диагностика гемостазиологических расстройств на основе результатов НПТЭГ

По собранным во время проведения исследования образцу цельной крови данным эксперт оценивает ее состояние структуры и хронометрии во время свертывания. Затем, в соответствии с рекомендациями к диагностике, представленными в таблице 18 приложения Б, ставится один из 11 диагнозов: 3 диагноза группы «Гиперкоагуляция», 3 диагноза группы «Гипокоагуляция», 3 условно-нормальных состояния и 2 смешанных диагноза.

Разделение диагнозов «Условная норма», «Норма с тенденцией к гиперкоагуляции» или «Норма с тенденцией к гипокоагуляции» происходит по усмотрению врача.

1.2 Обзор аналогов модуля поддержки принятия решений в процессе оценки состояния системы гемостаза

«ТЕГ 5000» является единственным тромбоэластографом, который способен проводить исследования цельной крови и собирать схожий набор параметров НПТЭГ.

Для помощи врачам в ПО тромбоэластографа «ТЕГ 5000» встроен функционал для сравнения результатов исследования с заранее установленными пользователем нормальными результатами исследования цельной крови человека. На основе сравнения ставится общий диагноз: относится результат анализа к группе диагнозов «Гиперкоагуляция», «Гипокоагуляция» или к условно-нормальным состояниям.

Данный результат рассчитывается на основе сравнения численных значений следующих параметров: «Максимальная плотность сгустка», «Время свертывания крови» и «Лизис через 30 минут после достижения максимальной амплитуды».

1.3 Теоретические основы

1.3.1 Постановка задачи классификации

Проблема классификации заключается в том, чтобы анализируемую совокупность объектов разбить на сравнительно небольшое число однородных групп или классов [4].

Задача постановки диагноза по результатам исследования образца крови человека, проведённого с помощью тромбоэластографа, может быть определена как задача классификации, где в качестве объектов выступают определяемые во время исследования образца крови параметры, а в качестве классов – гемостазиологические диагнозы.

Решение задачи классификации возможно с помощью эмпирических методов. Для этого специалисту требуется построить аналитическую модель, с помощью которой будет происходить классификация.

В случае с анализом параметров НПТЭГ возможно построить модель для оценки общего состояния системы гемостаза – определить, к какой группе относится нарушение: «Гиперкоагуляция», «Гипокоагуляция» или «Норма условная». Часто медицинский специалист останавливается на такой интерпретации результатов исследования образца цельной крови.

Для более подробной диагностики требуется глубокое понимание методов НПТЭГ. Модель, основанная на эмпирических методах, для постановки одиннадцати диагнозов будет слишком сложна, поэтому задача подробной диагностики гемостазиологических расстройств по результатам исследования цельной крови решается с использованием методов машинного обучения.

Так как число классов заранее известно, и присутствует набор данных, в котором каждому объекту заранее сопоставлена метка класса, можно сделать вывод, что рассматриваемая задача – классификация с учителем.

Для решения этой задачи из данных формируется тренировочная выборка, затем на маркированных тренировочных данных обучается классификатор. Результаты его работы могут быть оценены на тестовой выборке, что позволяет сравнить несколько классификаторов и выбрать лучший. Обученный классификатор позволяет делать прогнозы на ранее не встречавшихся данных [5].

1.3.2 Обзор классификаторов

Существует множество алгоритмов машинного обучения, применимых к решению задачи классификации. Например, часто используются ИНС. Они показывают высокую эффективность особенно при решении задач классификации изображений или звука [6].

В случае, если нужно классифицировать менее сложные объекты, качество работы ансамблевых классификаторов (например, на основе случайного леса или градиентного бустинга) сравнимо, а зачастую даже превосходит качество работы классификатора на основе ИНС [7, 8].

Заранее невозможно определить какой классификатор лучше подходит для решения конкретной задачи, поэтому следует исследовать несколько, а затем сравнить их результаты. Далее перечислены классификаторы, используемые в работе.

1.3.2.1 Искусственная нейронная сеть

Искусственной нейронной сетью называют математическую модель, а также её программную или аппаратную реализацию, которая является результатом применения математического аппарата к исследованию работы нервной системы живых существ. ИНС представляет собой систему соединённых и взаимодействующих между собой простых процессоров, называемых формальными нейронами [9].

Под обучением ИНС подразумевается процесс настройки весов межнейронных связей с целью нахождения набора весов, обеспечивающего изменение выходных сигналов в необходимой зависимости от данного вектора входных сигналов. Одним из примеров алгоритмов обучения ИНС с учителем является алгоритм обратного распространения ошибки или «backpropagation». В этом алгоритме происходит распространение ошибки от выходов ИНС к входам, то есть в направлении, обратном распространению сигналов в обычном режиме работы.

1.3.2.2 Случайный лес

Дерево решений – это модель машинного обучения, представляющая собой совокупность правил для принятия решений [10]. Случайный лес – алгоритм машинного обучения, основная идея которого заключается в использовании большого ансамбля деревьев решений. Каждое дерево само по себе даёт невысокое качество классификации, но за счёт их большого количества результат улучшается [7].

1.3.2.3 Градиентный бустинг

Бустинг – объединение ансамбля «слабых» классификаторов с целью получить «сильный» классификатор. Здесь «слабым» называется классификатор, дающий лишь слегка лучший результат, чем случайное угадывание (его предсказания слабо коррелированы с истинным распределением классов). Предсказания же «сильного» классификатора сильно коррелированы с истинным распределением.

Финальный классификатор находится в виде линейной комбинации классификаторов. Поиск оптимальных значений коэффициентов этой линейной комбинации слишком трудоемкая задача, поэтому в градиентном бустинге используется жадный алгоритм постепенного добавления классификаторов [11].

1.3.2.4 Многоэтапные классификаторы

Несколько классификаторов могут быть объединены в цепочки. Тогда результаты работы одного классификатора являются основанием для использования или не использования другого классификатора. Такой классификатор является многоэтапным классификатором.

1.3.3 Настройка классификаторов

Для получения качественного классификатора требуется провести отбор признаков, которые используются в процессе обучения, а также настроить его гиперпараметры.

Гиперпараметры классификатора – параметры, значения которых задаются до начала обучения и не изменяются в процессе обучения. Качество настройки гиперпараметров напрямую влияет на качество классификации.

В случае работы с данными небольшой размерности для отбора признаков часто применяются алгоритм обратного отбора признаков и способ отбора признаков на основе анализа корреляционной матрицы, для «тонкой» настройки гиперпараметров – метод поиска по решетке [5].

1.3.3.1 Алгоритм обратного отбора признаков

В основе алгоритма обратного отбора признаков лежит идея последовательного удаления признаков из полнопризнакового множества, пока новое подпространство признаков не будет содержать нужное их количество [5].

Для определения того, какой признак удалять на каждом шаге, определяется критериальная функция, значение которой алгоритм минимизирует. Вычисленный данной функцией критерий может быть разницей в качестве классификатора после и до удаления отдельно взятого признака, тогда удаляемый на каждом шаге признак определяется как признак, который максимизирует этот критерий.

1.3.3.2 Анализ матрицы корреляции

Если две величины связаны между собой, то между ними есть корреляция. Для выяснения вопроса о наличии связи между двумя величинами определяется, существует ли соответствие между большими и малыми значениями одной величины и соответствующими значениями другой величины или такой связи не обнаруживается [12].

Значение каждого элемента матрицы корреляции определяется величиной меры связи между двумя величинами. Слишком большое значение элемента матрицы говорит о наличии сильной зависимости между двумя величинами, слишком маленькое значение говорит об обратной зависимости.

1.3.3.3 Поиск по решётке

Традиционный метод оптимизации гиперпараметров, который реализует полный перебор заданному вручную подмножеству пространства гиперпараметров обучающего алгоритма [13].

1.3.4 Оценка качества работы классификатора

Существует множество способов оценки качества модели. Далее перечислены метрики и методы графического представления результата, которые часто используются при решении задачи классификации.

1.3.4.1 Матрица несоответствий

Матрица несоответствий – матрица, в которой для каждого класса наблюдений приводится количество наблюдений, отнесенных моделью к этому и другим классам. Её общий вид представлен на рисунке 2.

		<i>P</i>	<i>N</i>
True label	<i>P</i>	True Positives (TP)	False Negatives (FN)
	<i>N</i>	False Positives (FP)	True Negatives (TN)
		Predicted label	

Рисунок 2 – Общий вид матрицы несоответствий

Диагональное преобладание в матрице несоответствий означает, что модель устойчива и классы разделимы.

1.3.4.2 Точность тестовой выборки

Точность тестовой выборки определяется путём обучения классификатора на одной части данных (обучаемой выборке) и вычисления значения точности на другой (тестовой выборке) по формуле 1 [14].

$$\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN}, \quad (1)$$

1.3.4.3 Точность перекрёстной проверки

Точность перекрёстной проверки определяется следующим образом:

1. Данные разделяются на несколько равных частей;

2. Все части данных, кроме одной, становятся обучающей выборкой, неиспользованная часть – тестовой выборкой;
3. Вычисляется точность тестовой выборки по формуле 1;
4. Пункты 2-3 повторяются для всех возможных вариантов разбиения данных на обучающую и тестовую выборку;
5. Точностью перекрестной проверки является среднее арифметическое точностей всех возможных тестовых выборок.

1.3.5 Оценка важности признаков в процессе классификации

Для оценки важности признаков (степени влияния признаков) использовалась библиотека «shap». Для это рассчитываются значения Шэпли – оптимальность распределения выигрыша между игроками в задачах теории кооперативных игр – с и без признака, и на основании полученных результатов делается вывод о степени влияния признака [15].

1.4 Программное обеспечение тромбозластографа

Алгоритм классификации реализуется в модуле поддержки принятия решений, который предназначен для работы в ПО тромбозластографа. Программное обеспечение тромбозластографа работает на ПК, который поставляется вместе с прибором. ПК использует ОС Windows OS и не обладает большими вычислительными мощностями.

ПО тромбозластографа написано на языке программирования Delphi. На рисунке 3 представлен пример пользовательского интерфейса ПО тромбозластографа.

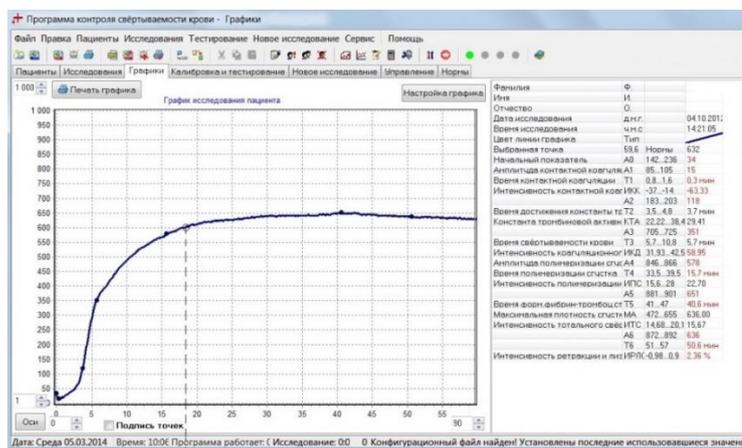


Рисунок 3 – Пользовательский интерфейс ПО тромбоэластографа

Данное ПО служит для хранения и представления результатов исследования крови, проведенных с помощью тромбоэластографа.

1.5 Программные инструменты

Для разработки, исследования классификаторов и реализации модуля поддержки принятия решений использовался стандартный набор инструментов для анализа данных и машинного обучения:

- ЯП Python – высокоуровневый язык программирования, ориентированный на повышение производительности разработчика и читаемости кода. Python отличается большим набором различных инструментов, лаконичным синтаксисом;
- «numpy» – библиотека языка Python для быстрой работы с многомерными массивами;
- «pandas» – библиотека языка Python, содержащая инструменты для работы с данными;
- «scikit-learn» – библиотека языка Python, содержащая большое количество алгоритмов машинного обучения, в том числе реализации вышеупомянутых классификаторов;
- «Keras» в связке с «TensorFlow» – библиотека языка Python для работы с ИНС.

Для внедрения модуля в ПО тромбоэластографа использовались средства ЯП Delphi.

2 Разработка классификатора

2.1 Анализ данных исследований крови с помощью тромбоэластографа

В медицинских учреждениях РФ был собран набор из 1243 реальных исследований образцов цельной крови, проведённых с помощью тромбоэластографа. Каждая запись состоит из 20 показателей НПТЭГ и соответствующего им диагноза. Описания показателей НПТЭГ и их характеристик представлены в таблице 19 приложения А.

Были замечены следующие особенности данных:

- Ошибки в данных.

Для некоторых исследований отметки времени располагаются не в хронологическом порядке. Такая запись результатов возможна, если отключить тромбоэластограф до конца исследования и запустить заново, не очищая память прибора. Ошибочно записанные исследования были удалены.

- Наличие отсутствующих значений.

Чаще всего отсутствуют записи амплитуды НПТЭГ, времени последней точки графика исследования («Аб» и «Тб»), интенсивности лизиса и ретракции сгустка («IRLS»). Это объясняется тем, что получение значений этих параметров требует более продолжительного исследования образца крови, а также тем, что врач способен определить диагноз, не опираясь на значения этих параметров.

Признаки «Аб», «Тб» и «IRLS» для которых отсутствует большая часть значений, были исключены.

- Значения параметров находятся в разных диапазонах.

Средние, минимальные и максимальные значения разных параметров сильно различаются. Это может отрицательно влиять на работу алгоритмов классификации, поэтому перед обучением и тестированием значения признаков масштабируются.

- Несбалансированность данных.

Количество исследований для разных диагнозов значительно различается. Для устранения данной несбалансированности при обучении классификаторов указываются веса классов. Веса классов рассчитываются как отношение количества исследований класса и общего количества исследований.

- Наличие сильно коррелирующих признаков.

Матрица корреляции признаков представлена на рисунке 14 приложения В.

Можно заметить, что существуют параметры, коэффициент корреляции которых больше 0.9, что говорит о высокой зависимости значения одного параметра от значения другого. Это подтверждается назначениями параметров. Например, параметр «Т4» («Время через 10 минут после свертывания крови») зависит от параметра «Т3» («Время свертывания крови»). Наличие сильно коррелирующих параметров может отрицательно влиять на качество модели, поэтому для обучения каждого классификатора отбирается подходящий для него набор признаков.

После подготовки данных набор содержит 1186 записи и 17 параметров.

2.2 Исследование классификаторов

На основе анализа предметной области были выбраны следующие классификаторы для исследования:

- ИНС;
- случайный лес;
- градиентный бустинг;
- двухэтапный, повторяющий процесс диагностики и используемый медицинскими экспертами.

Для каждого классификатора были отобраны показатели НПТЭГ (признаки), настроены гиперпараметры и получены результаты, которые после сравнивались.

2.2.1 Алгоритм настройки классификаторов

Настройка гиперпараметров и отбор признаков для классификатора происходили следующим образом:

1. Настройка гиперпараметров классификатора методом решетки;
2. Отбор признаков с помощью алгоритма обратного отбора признаков или анализа матрицы корреляции;
3. Настройка гиперпараметров классификатора, работающего с отобраным подмножеством признаков.

При необходимости процесс повторялся.

2.2.2 Выбор метрик

Прежде чем приступить к работе с классификаторами требовалось определить метрики, которые следует максимизировать во время настройки и отбора признаков и использовать для сравнения качества работы классификаторов.

Были выбраны следующие метрики качества:

- Точность перекрестной проверки.
- Точность перекрестной проверки с допустимыми ошибками – разработанная для данной задачи точность перекрестной проверки, при которой не учитываются такие ошибки, где тенденция к определенному типу диагноза, например, «Норма в гиперкоагуляции», предсказывается как один из диагнозов данного типа, например, «Гиперкоагуляция».
- Точность перекрестной проверки с допустимыми ошибками и штрафами – разработанная для данной задачи точность, является точностью перекрестной проверки с допустимыми ошибками с добавленными штрафами, которые назначаются, если заболевший человек определяется как здоровый.
- Время, которое требуется на классификацию одного исследования образца крови.

Для визуализации результатов работы классификатора использовалась матрица несоответствий. Для вычисления метрик перекрестной проверки данные случайным образом разделялись на пять частей. Для вычисления матрицы несоответствий классификатор обучался на 70% данных и тестировался на оставшихся 30%.

2.2.3 Классификатор на основе ИНС

Отбор признаков для нейронной сети происходил на основе анализа матрицы корреляции, представленной в приложении В. В результате были исключены четыре признака.

Параметры ИНС представлены в таблице 1.

Таблица 1 – Параметры ИНС

Параметр	Значение
Тип сети	Прямого распространения
Количество полносвязанных слоёв	5
Количество слоёв исключения	2
Способ обучения	С учителем
Количество нейронов на входном слое	13
Количество нейронов на скрытых слоях	35
Количество нейронов на выходном слое	11
Функция активации	categorical_crossentropy

Точность перекрестной проверки классификатора на основе ИНС составила 0.9089, точность перекрестной проверки с допустимыми ошибками – 0.9266, точность перекрестной проверки с допустимыми ошибками и штрафами – 0.9203. Время работы классификатора на конечной машине – приблизительно 15 секунд. Такое большое значение времени работы объясняется характеристиками ПК, и особенностями запуска TensorFlow на ОС Windows.

2.2.4 Классификаторы на основе градиентного бустинга и случайного леса

Признаки для ансамблевых классификаторов на основе градиентного бустинга и случайного леса были отобраны с помощью алгоритма обратного отбора признаков.

С помощью поиска по решетке настраивались следующие гиперпараметры деревьев решений, ансамбль которых используется в классификаторах: «Минимальное число объектов, при котором выполняется расщепление», «Максимальная глубина деревьев». Также для градиентного бустинга настраивался параметр «Скорость обучения».

Параметры классификаторов на основе градиентного бустинга и случайного леса для работы с 11 диагнозами представлены в приложении Г в таблицах 21 и 22 соответственно. Результаты тестирования приведены в таблице 2.

Таблица 2 – Сравнение результатов работы классификаторов

	Время работы на конечной машине, сек	Точность перекрестной проверки	Точность перекрестной проверки с допустимыми ошибками	Точность перекрестной проверки с допустимыми ошибками и штрафами
Классификатор на основе градиентного бустинга	2	0.9072	0.9174	0.9102
Классификатор на основе случайного леса	2	0.9182	0.9309	0.9246

2.2.5 Двухэтапный классификатор

2.2.5.1 Структура двухэтапного классификатора

Врач диагностирует гемостазиологическое расстройство на основе анализа двух характеристик свёртывания крови: структуры и хронометрии крови. Данный подход лёг в основу двухэтапного классификатора.

Структура двухэтапного классификатора представлена на рисунке 4.



Рисунок 4 – Структура двухэтапного классификатора

Так как метод диагностики, основанный на отдельной оценке исследования крови по структуре и хронометрии не способен разделить диагнозы «Норма с тенденцией к гиперкоагуляции», «Норма условная» и «Норма с тенденцией к гипокоагуляции», классификация данных диагнозов и их отделение от остальных диагнозов происходит на первом этапе.

На втором этапе происходит оценка исследования образца крови по структуре и хронометрии. После второго этапа формируется общий диагноз на основе рекомендаций, описанных в таблице 17 приложения А: Если патология одинаковая для хронометрии и структуры, например, «Гиперкоагуляция» и для хронометрии и для структуры, она является результирующим диагнозом. Если патология присутствует только для хронометрии или для структуры, характер патологии уточняется, например, «Гиперкоагуляция хронометрическая». Иначе диагноз формируется путём

объединения результатов, например, «Структурная гиперкоагуляция – Хронометрическая гипокоагуляция».

2.2.5.2 Настройка внутренних классификаторов двухэтапного классификатора

В качестве внутренних классификаторов для двухэтапного классификатора использовались алгоритмы на основе случайного леса и градиентного бустинга. В данном случае от использования ИНС было решено отказаться в связи с их неприемлемым временем работы на конечной машине, замеченным ранее.

Настройка внутренних классификаторов происходила по алгоритму, описанному в пункте 2.2.1. В качестве результирующей метрики была выбрана точность перекрестной проверки.

Результаты настройки классификаторов представлены в таблицах 21 и 22 приложения Г. Результаты тестирования представлены в таблице 3. Выделенные ячейки показывают выбранные классификаторы.

Таблица 3 – Выбор классификаторов для двухэтапного классификатора

Классификатор	Случайный лес, точность перекрёстной проверки	Градиентный бустинг, точность перекрёстной проверки
Первого этапа	0.9772	0.9791
Второго этапа – Состояние структуры	0.9326	0.9393
Второго этапа – Состояние хронометрия	0.9886	0.9857

Далее, с целью проанализировать работу двухэтапного классификатора, подробно рассмотрены классификаторы, показавшие наивысшую точность.

2.2.5.3 Классификатор первого этапа на основе градиентного бустинга

На рисунке 4 представлен график зависимости точности перекрестной проверки от количества признаков.

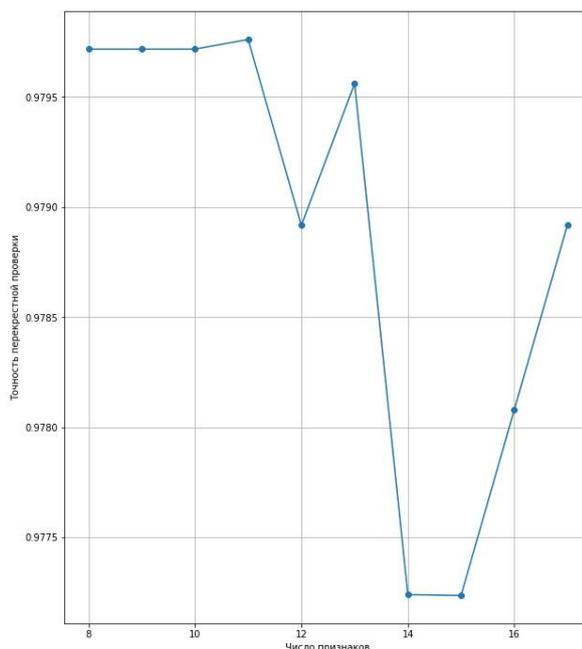


Рисунок 4 – Зависимость точности перекрестной проверки от числа признаков для классификатора первого этапа

Классификатор на основе градиентного бустинга для первого этапа работает с набором из 11 признаков. После настройки гиперпараметров его точность достигла 97.91%. Анализ важности признаков с помощью библиотеки «shap» показал, что на процесс классификации наибольшее влияния оказывают следующие признаки (в порядке убывания степени влияния):

- время свертывания крови;
- максимальная плотность сгустка;
- интенсивность контактной коагуляции.

2.2.5.4 Классификатор структуры на основе градиентного бустинга

На рисунке 5 представлен график зависимости точности перекрестной проверки от количества признаков.

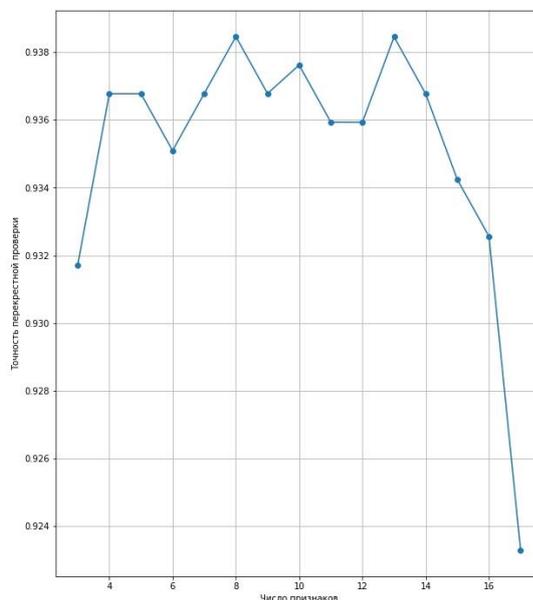


Рисунок 5 – Зависимость точности перекрестной проверки от числа признаков для классификатора структуры

Классификатор на основе градиентного бустинга для структуры показал наилучшие результаты при работе с набором из 8 признаков. После настройки гиперпараметров его точность достигла 93.93%.

Основное влияние во время квалификации оказывают следующие признаки (в порядке убывания степени важности):

- максимальная плотность сгустка;
- время свертывания крови;
- интенсивность контактной фазы коагуляции;
- значение амплитуды НПТЭГ во время свертывания крови;
- интенсивность тотального свертывания;
- константа тромбиновой активности;
- интенсивность полимеризации сгустка.

2.2.5.5 Классификатор хронометрии на основе случайного леса

На рисунке 6 представлен график зависимости точности перекрестной проверки от количества признаков.

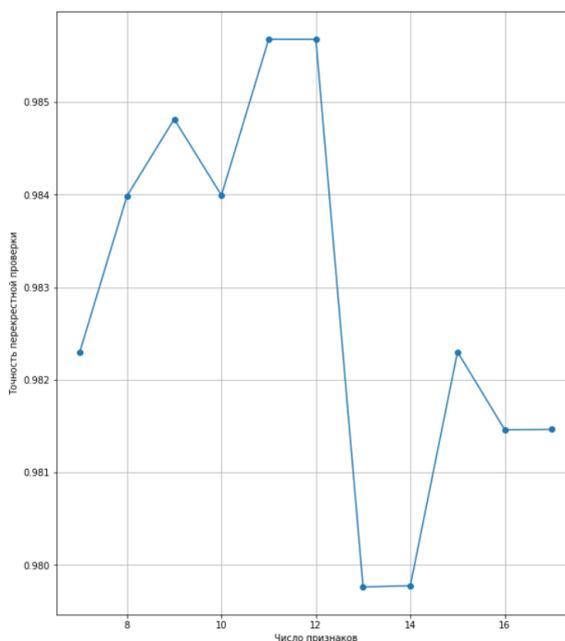


Рисунок 6 – Зависимость точности перекрестной проверки от числа признаков для классификатора хронометрии

Классификатор для хронометрии показал наилучшие результаты при работе с набором из 12 признаками. После настройки гиперпараметров его точность достигла 98.86%. Анализ важности признаков с помощью библиотеки «shap» показал, что на процесс классификации наибольшее влияния оказывают следующие признаки (в порядке убывания степени влияния):

- Время свертывания крови;
- Время через 10 минут после свертывания крови;
- Время достижения амплитуды «A2» (Увеличение амплитуды НПТЭГ на 100 о.е.);
- Интенсивность контактной коагуляции;
- Константа тромбиновой активности;
- Максимальная плотность сгустка.

2.5.5.6 Формирование диагноза в случае ошибок классификатора первого этапа

На рисунке 7 представлена матрица несоответствий для классификатора первого этапа.

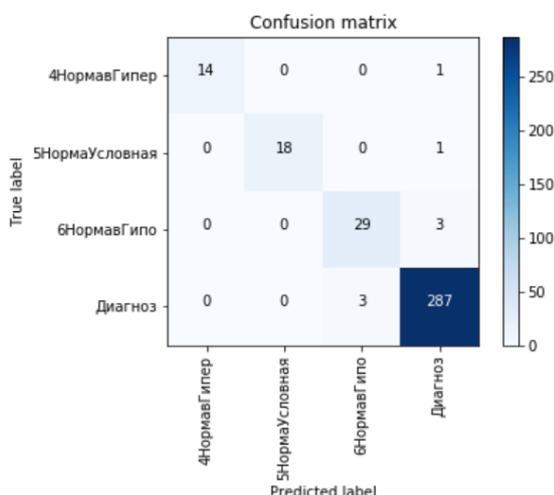


Рисунок 7 – Матрица несоответствий для классификатора на основе случайного леса для первого этапа

Можно заметить, что точность работы классификатора первого этапа не является идеальной, поэтому на этапе формирования диагноза может оказаться исследование с результатами «Норма» и для структуры, и для хронометрии. Так как классификатор первого этапа чаще всего по ошибке отправляет на второй этап исследования с диагнозом «Норма в гипокоагуляции», что видно на матрице несоответствий для классификатора первого этапа (рисунок 4), данный диагноз является результирующим в случае двух показателей «Норма» на этапе формирования диагноза.

2.5.5.7 Результаты тестирования

Время постановки диагноза на конечной машине с помощью двухэтапного классификатора составило 4 секунды. Точность перекрестной проверки, точность перекрёстной проверки с допустимыми ошибками и точность перекрестной проверки с допустимыми ошибками и штрафами составили 0.9309, 0.9384 и 0.933 соответственно.

2.2.6 Результаты исследования

Сравнение результатов работы классификаторов представлено в таблице 4.

Таблица 4 – Сравнение результатов работы классификаторов

	Время работы на конечной машине, сек	Точность перекрестной проверки	Точность перекрестной проверки с допустимыми ошибками	Точность перекрестной проверки с допустимыми ошибками и штрафами
Классификатор на основе ИНС	15	0.9089	0.9266	0.9203
Классификатор на основе градиентного бустинга	2	0.9072	0.9174	0.9102
Классификатор на основе случайного леса	2	0.9182	0.9309	0.9246
Двухэтапный классификатор	4	0.9309	0.9384	0.933

Классификаторы на основе градиентного бустинга и случайного леса показали лучшее время работы на конечной машине, но наивысшие значения метрик перекрестной проверки показал двухэтапный классификатор, описание которого представлено в следующем разделе. Данный классификатор был выбран для дальнейшего исследования.

2.3 Анализ ошибок двухэтапного классификатора

На рисунке 8 представлена матрица несоответствий для двухэтапного классификатора.

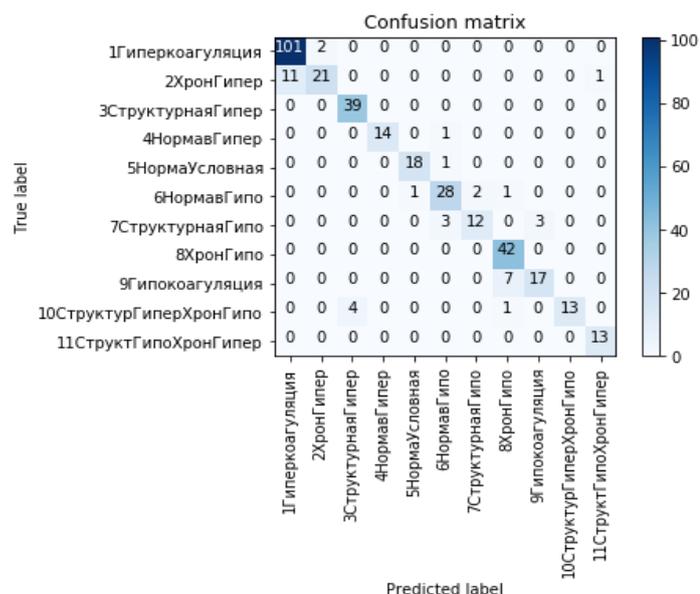


Рисунок 8 – Матрица несоответствий для двухэтапного классификатора

На основе анализа матрицы несоответствия были выделены пары диагнозов, которые чаще всего определяются неверно.

- **1 и 2:** «Гиперкоагуляция» (Метка диагноза – 1) и «Хронометрическая гиперкоагуляция» (Метка диагноза – 2);
- **6 и 7:** «Норма в гипокоагуляции» (Метка диагноза – 6) и «Структурная гипокоагуляции» (Метка диагноза – 7);
- **8 и 9:** «Гипокоагуляция хронометрическая» (Метка диагноза – 8) «Гипокоагуляция» (Метка диагноза – 9).

Возникло предположение, что при добавлении классификаторов уточнения в качестве третьего этапа, метрики, оценивающие работу многоэтапного классификатора, улучшатся.

2.4 Трёхэтапный классификатор

2.4.1 Структура трёхэтапного классификатора

Структура трёхэтапного классификатора представлена на рисунке 9.

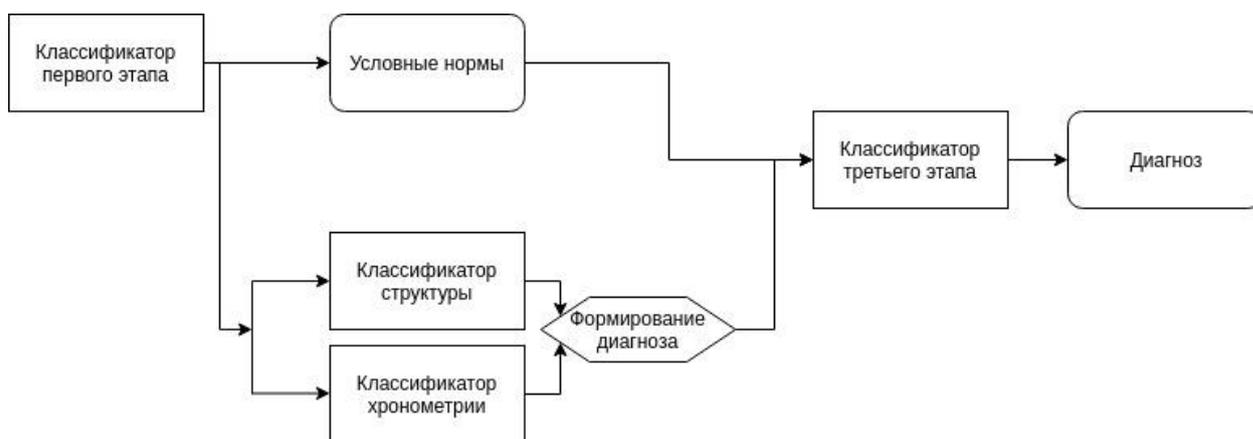


Рисунок 9 – Структура трёхэтапного классификатора

Назначение первых двух этапов и процесс формирования диагноза не изменились по сравнению с двухэтапным классификатором. Третий этап служит для уточнения спорных диагнозов. Структура третьего этапа определялась во время исследования.

2.4.2 Настройка внутренних классификаторов трёхэтапного классификатора

Настройка внутренних классификаторов для трёхэтапного классификатора происходила аналогично настройке внутренних классификаторов для двухэтапного классификатора.

Результаты настройки представлены в таблицах 21 и 22 приложения Г. Результаты тестирования представлены в таблице 3. Выделенные ячейки показывают выбранные классификаторы для пар классов.

Таблица 5 – Выбор классификаторов для трёхэтапного классификатора

Классификатор третьего этапа	Случайный лес, точность перекрёстной проверки	Градиентный бустинг, точность перекрёстной проверки
Для пары диагнозов 1 и 2	0.9588	0.9546
Для пары диагнозов 6 и 7	0.8900	0.8962
Для пары диагнозов 8 и 9	0.9344	0.9295

2.4.3 Уточнение структуры третьего этапа трёхэтапного классификатора

Результаты работы трёхэтапного классификатора с различным набором классификаторов третьего этапа представлены в таблице 4.

Таблица 6 – Результаты работы трёхэтапного классификатора с различным набором классификаторов третьего этапа

Набор классификаторов третьего этапа	Точность перекрестной проверки	Точность перекрестной проверки с допустимыми ошибками	Точность перекрестной проверки с допустимыми ошибками и штрафами
Классификаторы для пары диагнозов 1-2	0.9283	0.9359	0.9304
Классификаторы для пары диагнозов 6-7	0.9266	0.9469	0.9456
Классификаторы для пары диагнозов 8-9	0.93	0.9376	0.9321
Классификаторы для пар диагнозов 1-2, 5-6	0.9241	0.9443	0.9431
Классификаторы для пар диагнозов 1-2, 8-9	0.9275	0.9351	0.9296
Классификаторы для пар диагнозов 5-6, 8-9	0.9228	0.926	0.9428
Классификаторы для пар диагнозов 1 и 2, 5 и 6, 8 и 9	0.9233	0.9435	0.9422
Двухэтапный классификатор	0.9309	0.9384	0.933

В результате наилучшее значение точности перекрёстной проверки показал двухэтапный классификатор, но значения метрик с допустимыми ошибками и штрафами значительно выше у трёхэтапного классификатора с классификаторами уточнения для диагнозов 6-7. Это объясняется тем, что пара диагнозов 6-7 располагаются на границе условно-нормальных состояний и заболеваний: «Норма в гипокоагуляции» – это разновидность нормы, «Структурная гипокоагуляции» – гемостазиологическое расстройство. Улучшение делимости данных классов или возможность в случаях, когда точная диагностика затруднительна, направить ошибку в сторону ложноположительного диагноза, значительно повлияла на точность перекрёстной проверки со штрафами.

Трёхэтапный классификатор с классификатором уточнения для диагнозов 6 и 7 в качестве третьего этапа был выбран для реализации в модуле поддержки принятия решений, так как его общая точность перекрёстной проверки отличается от максимального значения данной точности незначительно, а значения метрик перекрёстной проверки с допустимыми ошибками и штрафами являются лучшими. Данный классификатор менее подвержен таким видам ошибок, как классификация заболевшего человека как здорового. Именно такой вид ошибок является наиболее опасным в медицинской диагностике.

Уточнённая структура трёхэтапного классификатора представлена в приложении Д.

3 Разработка модуля поддержки принятия решений

3.1 Проектирование модуля поддержки принятия решений

3.1.1 Описание архитектуры модуля поддержки принятия решений

Диаграмма компонентов модуля поддержки принятия решений в процессе оценки состояния системы гемостаза представлена на рисунке 9.

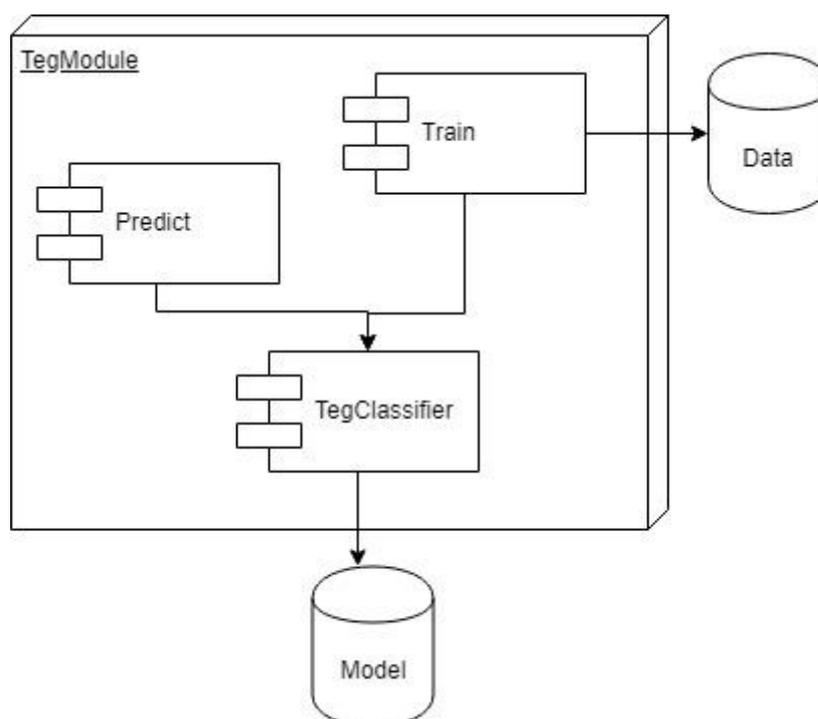


Рисунок 10 – Диаграмма компонентов

Описание компонентов:

- «Data» – файл данных для обучения классификатора.
- «Model» – файл модели, сохраненный обученный классификатор, который используется для постановки диагноза.
- «TegClassifier» – компонент, в котором реализован классификатор на основе выбранного алгоритма трёхэтапной классификации и функционал для обучения, сохранения и использования файла модели «Model».
- «Train» – компонент, с которым возможно взаимодействие извне модуля, использует функционал компонента «TegClassifier» для обучения и сохранения классификатора с помощью набора данных «Data».

– «Predict» – компонент, с которым возможно взаимодействие извне модуля, использует функционал компонента «TegClassifier» для постановки диагноза.

3.1.2 Описание архитектуры компонента TegClassifier

Архитектура компонента TegClassifier представлена при помощи UML-диаграммы классов на рисунке 10.

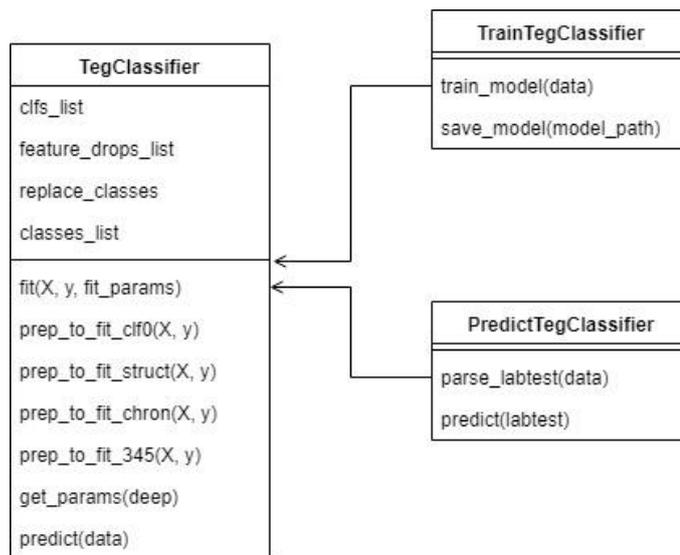


Рисунок 11 – UML диаграмма классов компонента TegClassifier

Описание классов.

– Класс «TegClassifier».

«TegClassifier» представляет собой трёхэтапный классификатор. В качестве классификаторов первого этапа, хронометрии, структуры, третьего этапа могут быть использованы любые классификаторы из библиотеки «scikit-learn». Они передаются «TegClassifier» в соответствующем порядке в виде списка «clfs_list».

Исключаемые признаки для каждого классификатора указываются в списках «feature_drops_list». Какими классами для классификаторов хронометрии и структуры заменяются оригинальные классы указывается в словаре «replace_classes». С какими классами какие классификаторы третьего этапа работают указывается в списках «classes_list».

Метод «fit» служит для обучения трёхэтапного классификатора. Метод «predict» служит для получения результата как одиночного исследования крови, так и набора исследований. Метод «get_params» возвращает подробный список параметров трёхэтапного классификатора.

Метод prepare_to_fit_clf0, prepare_to_fit_struct, prepare_to_fit_chron, prepare_to_fit_345 являются вспомогательными функциями для подготовки обучающей выборки во время обучения определённого классификатора.

– Класс «TrainTegClassifier».

Класс «TrainTegClassifier» предназначен для обучения модели трёхэтапного классификатора. Методы «train_model» и «save_model» служат для тренировки и сохранения модели соответственно.

– Класс «PredictTegClassifier».

Класс «PredictTegClassifier» предназначен для использования сохраненной модели трёхэтапного классификатора. Метод «parse_labtest» служит для преобразования параметров исследования из вида строки в список значений. Метод «predict» служит для диагностики исследования крови.

3.2 Описание компонентов Train и Predict

Компоненты «Train» и «Predict» служат для взаимодействия с модулем. Характеристики их реализации описаны в таблице 7.

Таблица 7 – Характеристики реализации компонентов для взаимодействия с модулем

Характеристика	Train	Predict
Тип приложения	Консольное	Консольное
Язык реализации	Python	Python
Совместимость с ОС	Windows, Linux	Windows, Linux

Продолжение таблицы 5

Входные параметры	—	Строка – результаты исследования крови, проведённого с помощью тромбоэластографа
Выходные параметры	—	Метка диагноза

4 Внедрение модуля в ПО тромбоэластографа

Сложность внедрения модуля поддержки принятия решений для классификации гемостазиологических расстройств заключалась в необходимости наладить взаимодействия между модулем, реализованным на ЯП Python и ПО тромбоэластографа, реализованным на ЯП Delphi.

4.1 Взаимодействие с компонентом «Train»

Так как обучение классификатора происходит одновременно и ПО тромбоэластографа настраивается специалистом для каждого прибора отдельно, возможность запуска обучения в ПО тромбоэластографа не внедрялась.

Во время установки ПО тромбоэластографа специалист, используя компонент «Train» модуля поддержки принятия решений и набор данных, обучает классификатор, который сохраняется в модуле в виде файла модели.

4.2 Взаимодействие с компонентом «Predict»

Возможным способом внедрения функционала для постановки диагноза рассматривалось использование библиотеки «Python4Delphi», которая предоставляет доступ к функциям и классам, реализованным на ЯП Python через код ЯП Delphi, но возникли проблемы при загрузке библиотеки «numpy», используемой модулем для быстрой работы с многомерными массивами и другими библиотеками, исходный код которых написан на ЯП C++.

В результате было решено реализовать взаимодействие между модулем и ПО тромбоэластографа через запуск компонента «Predict» в терминале ОС Windows в скрытом режиме.

Для взаимодействия между ПО тромбоэластографа и модулем поддержки принятия решений был реализован метод на ЯП Delphi. Данный метод получает на вход показатели НПТЭГ исследования цельной крови,

полученные с помощью тромбоэластографа, и выполняет следующие действия:

1. Преобразует показатели НПТЭГ цельной крови в строку;
2. Запускает терминал ОС Windows;
3. Вызывает компонент «Predict» модуля поддержки принятия решений;
4. Ожидает результат;
5. Преобразует результат к виду метки диагноза;
6. Возвращает результат.

Схема взаимодействия ПО тромбоэластографа и модуля поддержки принятия решений представлена на рисунке 12.

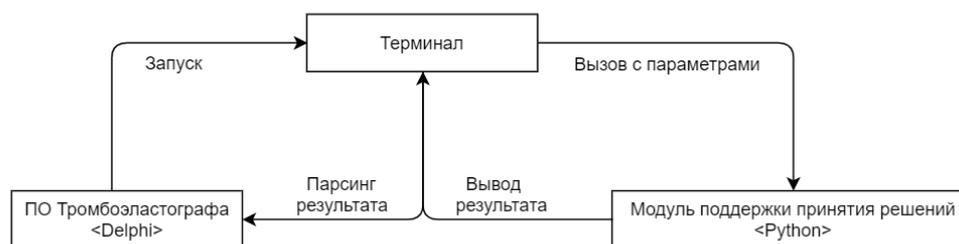


Рисунок 12 – Схема взаимодействия ПО тромбоэластографа и модуля поддержки принятия решений

5 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

5.1 Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения

5.1.1 Потенциальные потребители результатов исследования

Разработанный модуль поддержки принятия решений в процессе оценки состояния системы гемостаза после внедрения является частью ПО тромбоэластографа. Модуль способен с высокой точностью определять одиннадцать различных диагнозов. Метод тромбоэластографии на сегодняшний день является наиболее информативным способом диагностики нарушений системы гемостаза, но его использование требует высокой квалификации медицинских сотрудников. Внедрение модуля в ПО тромбоэластографа позволило ускорить обучение специалистов и улучшить качество диагностики гемостазиологических расстройств.

Потенциальным потребителем разработанного модуля являются пользователи тромбоэластографа. Так как метод тромбоэластографии применяется в хирургии, акушерстве, гинекологии, гематологии, кардиологии, неврологии, сердечнососудистой хирургии и в других областях медицины, потенциальными потребителями разработанного модуля являются мед учреждения перечисленных специальностей.

5.1.2 Анализ конкурентных технических решений

Как правило, в ПО тромбоэластографа не входит функционал для поддержки принятия решений в процессе постановки диагноза. Единственным конкурентом разработки может являться встроенный в ПО «TEG Analytical Software» тромбоэластографа «TEG 5000» функционал для помощи в диагностике гемостазиологических расстройств. В данном ПО существует функционал для сравнений анализа крови человека с заранее

установленным нормальным анализом крови человека или анализами крови людей из базы данных.

Оценочная карта для сравнения «TEG Analytical Software» и разработанного модуля представлена в таблице 8.

Таблица 8 – Оценочная карта для сравнения конкурентных технических решений

№ п/п	Конкурен ты	Факторы конкурентоспособности						Итого вая оценк а
		Точност ь диагнос тики	Детальн ость диагнос тики	Отсутств ие участия пользо вателя в процессе диагнос тики	Возможно сть точной диагнос тики для пациен тов с особеннос тями	Постано вка диагноз а на основе размече нной эксперта ми базы анализо в	Скорост ь диагнос тики	
1	TEG Analytical Software	10	3	5	8	7	10	7
2	Разработа нный модуль	9	10	10	6	8	8	8.76
	b_j	5	5	5	3	3	4	25
	w_j	0.2	0.2	0.2	0.12	0.12	0.16	-

Итоговая оценка конкурентного решения составляет 7, итоговая оценка разработанного модуля составляет 8.76, что значительно выше. Исходя из этого, можно сделать вывод о высокой конкурентоспособности разработанного модуля.

Многоугольник конкурентоспособности представлен на рисунке 13.

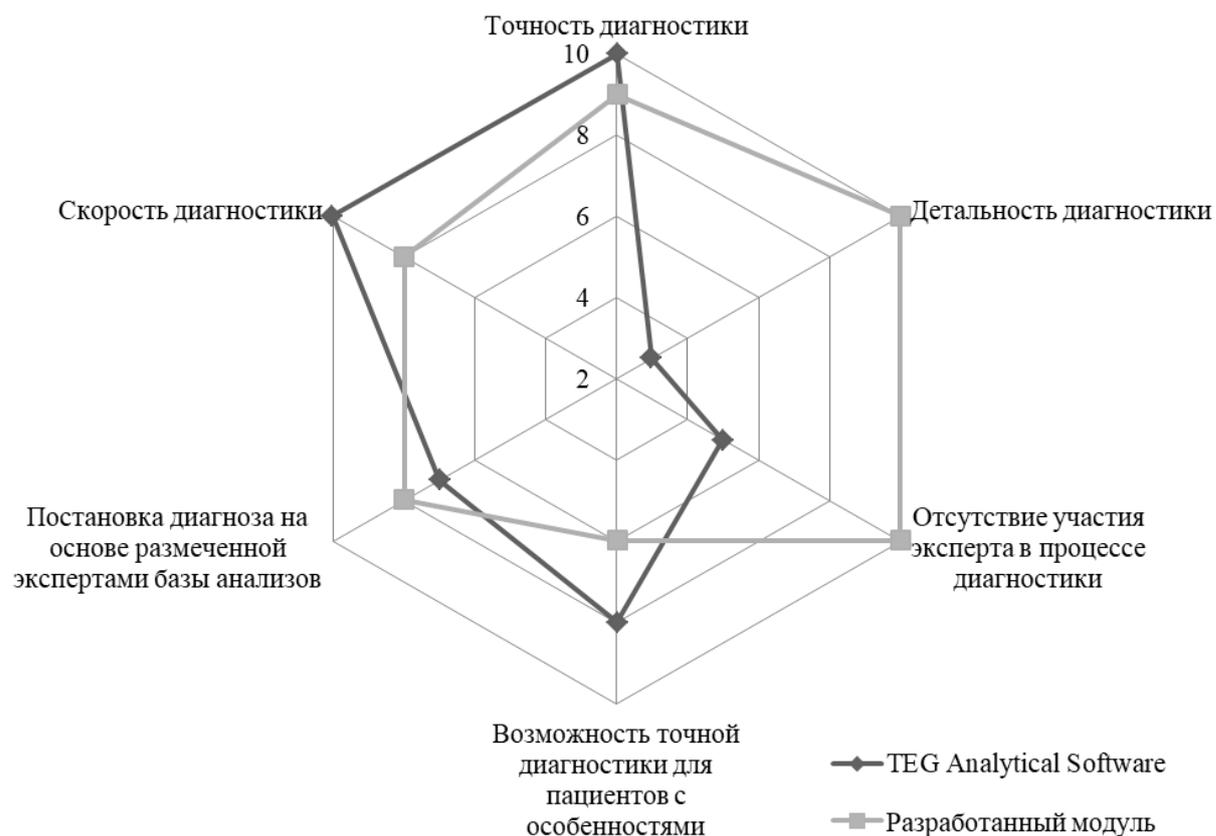


Рисунок 13 – Многоугольник конкурентоспособности

Можно заметить, что «TEG Analytical Software» отлично справляется с узкими задачами: дать общую оценку анализу крови на основе заданного в настройках нормального диагноза и предоставить эксперту возможность сравнить исследуемый анализ крови с базой анализов. Разработанный модуль уступает «TEG Analytical Software» по времени работы, незначительно уступает по точности диагностики, но способен определить детальный диагноз на основе анализа базы данных и построения модели машинного обучения, также не требует вмешательства пользователя. Разработанный модуль является более подходящим решением для обучения специалистов и помощи в постановке комплексных диагнозов.

Итоговая оценка разработки и анализ многоугольника конкурентоспособности позволяют говорить о целесообразности разработки и ее дальнейшего улучшения.

5.1.3 SWOT-анализ

SWOT-анализ – метод планирования, который заключается в выявлении внутренних и внешних факторов среды объекта планирования, которые разделяются на следующие категории:

- Strength (сильные стороны);
- Weaknesses (слабые стороны);
- Opportunities (возможности);
- Threats (угрозы).

Для исследуемой разработки SWOT-анализ приведен в таблице 9.

Таблица 9 – SWOT-анализ

Внутренние факторы		
	Сильные стороны: 1. Высокая точность постановки диагноза. 2. Отсутствие участия пользователя в процессе диагностики. 3. Простой способ внедрения в любую версию ПО тромбоэластографа.	Слабые стороны: 1. Скорость постановки диагноза. 2. Неточности в данных, на которых обучается модуль. 3. Отсутствие в данных, на которых обучается модуль, информации об уникальных пациентах (беременные, курящие и пр.).

Внешние факторы	<p>Возможности:</p> <p>1. Добавление возможности точной диагностики для уникальных пациентов.</p> <p>2. Добавление функционала для поддержки принятия решений в процессе выбора стратегии коррекции гемостазиологических расстройств.</p>	<p>Предполагаемое приоритетное направление развития – добавление возможности точной диагностики для уникальных пациентов.</p> <p>Используя уже разработанные методы, можно обучить модуль диагностировать пациентов с уникальными показателями нормы с такой же точностью.</p>	<p>Для добавления функционала требуется расширение базы данных анализов крови человека. Для расширения базы данных требуется обсуждение с медицинскими экспертами.</p>
	<p>Угрозы:</p> <p>1. Появление более качественных аналогов</p>	<p>Достоинства разработки и тромбоэластографа, в который она внедрена, могут помочь в конкурентной борьбе с аналогами, которые, возможно, будут разработаны в будущем.</p>	<p>Для того чтобы нивелировать угрозу появления конкурентов, необходимо регулярно расширять базу данных анализов крови человека и на ее основе расширять функционал или улучшать качество работы.</p>

В ходе SWOT-анализа были определены приоритетные направления развития разработки, а также способы для нивелирования слабых сторон разработки и противодействия возможным угрозам.

5.2 Планирование научно-исследовательских работ

5.2.1 Структура работ в рамках научного исследования

Планирование комплекса предполагаемых работ имеет следующий вид:

- определение работ в разработки;
- определение участников и исполнителей каждой работы;
- определение продолжительности работ;
- построение планируемого графика разработки.

В таблице 10 приведен порядок работ, выполняемых в ходе разработки, и исполнитель каждой работы.

Таблица 10 – Перечень работ и исполнителей при разработке модуля

№ работы	Наименование работы	Исполнители работы
1	Выбор научного руководителя бакалаврской работы	Чурсина Елена Андреевна
2	Составление и утверждение темы бакалаврской работы	Кудинов Антон Викторович, Чурсина Елена Андреевна
3	Составление календарного плана-графика выполнения бакалаврской работы	Кудинов Антон Викторович, Чурсина Елена Андреевна
4	Подбор и изучение литературы по теме бакалаврской работы	Чурсина Елена Андреевна
5	Анализ предметной области	Чурсина Елена Андреевна
6	Разработка алгоритма классификации гемостазиологических расстройств	Чурсина Елена Андреевна
7	Проектирование модуля поддержки принятия решений на основе разработанного алгоритма	Чурсина Елена Андреевна
8	Реализация модуля	Чурсина Елена Андреевна
9	Внедрение модуля в ПО тромбоэластографа	Чурсина Елена Андреевна
10	Согласование выполненной работы с научным руководителем	Кудинов Антон Викторович
11	Выполнение других частей работы (финансовый менеджмент, социальная ответственность)	Чурсина Елена Андреевна

12	Подведение итогов, оформление работы	Кудинов Антон Викторович, Чурсина Елена Андреевна
----	--------------------------------------	---------------------------------------------------

5.2.2 Разработка графика проведения научного исследования

Для определения ожидаемого значения трудоемкости для конкретной работы используется следующая формула:

$$t_{ож_i} = \frac{3 \cdot t_{min_i} + 2 \cdot t_{max_i}}{5}, \quad (2)$$

где t_{min_i} – минимальная трудоемкость i -ой работы, человеко-дни;

t_{max_i} – максимальная трудоемкость i -ой работы, человеко-дни.

Продолжительность каждой работы определяется по следующей формуле:

$$T_{pi} = \frac{t_{ож_i}}{c_i}, \quad (3)$$

где c_i – количество человек, выполняющих i -ую работу.

Для перевода длительности каждой работы из рабочих дней в календарные дни используется формула:

$$T_{ki} = T_{pi} \cdot k_{кал}, \quad (4)$$

где $k_{кал}$ – коэффициент календарности.

Для шестидневной рабочей недели коэффициент календарности определяется как:

$$k_{кал} = \frac{T_{кал}}{T_{кал} - T_{вых} - T_{пр}} = \frac{365}{365 - 66} = 1.22, \quad (5)$$

где $T_{кал}$ – количество календарных дней в году,

$T_{вых}$ – количество выходных дней в году,

$T_{пр}$ – количество праздничных дней в году.

Таблица 11 – Временные показатели реализации разработки

Наименование работы	Исполнители работы	Трудоемкость работ, человеко-дни			Длительность работ, дни	
		t_{min}	t_{max}	$t_{ож}$	T_p	T_k
Выбор научного руководителя бакалаврской работы	Чурсина Елена Андреевна	1	5	2.6	3	3
Составление и утверждение темы бакалаврской работы	Кудинов Антон Викторович	2	3	2.4	2	3
	Чурсина Елена Андреевна	1	2	1.4	1	2
Составление календарного плана-графика выполнения бакалаврской работы	Кудинов Антон Викторович	1	2	1.4	1	2
	Чурсина Елена Андреевна	2	5	3.2	3	4
Подбор и изучение литературы по теме бакалаврской работы	Чурсина Елена Андреевна	7	10	8.2	8	10
Анализ предметной области	Чурсина Елена Андреевна	10	14	11.6	12	14
Разработка алгоритма классификации гемостазиологических расстройств	Чурсина Елена Андреевна	10	12	10.8	11	13
Проектирование модуля поддержки принятия решений на основе разработанного алгоритма	Чурсина Елена Андреевна	7	9	7.8	8	10
Реализация модуля	Чурсина Елена Андреевна	4	8	5.6	6	7

Продолжение таблицы 11

Внедрение модуля в ПО тромбозластографа	Чурсина Елена Андреевна	4	8	5.6	6	7
Согласование выполненной работы с научным руководителем	Кудинов Антон Викторович	2	4	2.8	3	3
Выполнение других частей работы (финансовый менеджмент, социальная ответственность)	Чурсина Елена Андреевна	13	16	14.2	14	17
Подведение итогов, оформление работы	Чурсина Елена Андреевна	7	10	8.2	8	10
	Кудинов Антон Викторович	2	3	2.4	2	3

На основании данных из таблицы 11 была построена диаграмма Ганта выполненных работ, представленная на рисунке 15 в приложении В. Расшифровка цветовых обозначений на диаграмме приведена в таблице 19 в приложении Е.

5.2.3 Бюджет научно-технического исследования

5.2.3.1 Расчет материальных затрат научно-технического исследования

В материальные затраты научно-технического исследования вошли затраты на канцелярские принадлежности в сумме 2000 рублей. Таким образом, общая сумма материальных затрат составляет 2000 рублей.

5.2.3.2 Расчет затрат на специальное оборудование для научных (экспериментальных) целей

Поскольку необходимые лицензии на программное обеспечение, которое было использовано при реализации разработки, уже были приобретены компанией, затраты на оборудование включают в себя только затраты на амортизацию оборудования студента.

Норма амортизации для использованного оборудования составляет

$$A_n = \frac{1}{n} * 100\% = \frac{1}{3} \times 100\% = 33.33\%, \quad (6)$$

где n – срок полезного использования оборудования.

Тогда, с учетом того, что продолжительность работы составляет четыре месяца, формула для расчета итоговой суммы амортизации основных средств выглядит следующим образом:

$$A = \frac{C \cdot A_n \cdot 4}{12 \cdot 100\%}, \quad (7)$$

где C – стоимость оборудования.

Расчет амортизации приведен в таблице 12.

Таблица 12 – Расчет затрат на амортизацию

Наименование	Стоимость, руб	Затраты, руб.
Амортизация ПК	76000	8444.44
Итого:	76000	8444.44

Во время проведения научного исследования использовался ПК стоимостью 65000 рублей и свободное ПО стоимостью 0 рублей. Расчёт затрат на амортизацию представлен в таблице 12.

5.2.3.3 Основная заработная плата исполнителей темы

Для расчета основной заработной платы студента берется оклад, равный окладу ассистента без степени, т.е. 21760 руб. Зарплата преподавателя – кандидата технических наук – составляет 33664 руб.

Основная заработная плата рассчитывается по формуле:

$$Z_{\text{осн}} = Z_{\text{дн}} \cdot T_p \cdot (1 + K_{\text{пр}} + K_{\text{д}}) \cdot K_p, \quad (8)$$

где $Z_{\text{дн}}$ – среднедневная зарплата, руб.;

$K_{\text{пр}}$ – премиальный коэффициент (0,3);

$K_{\text{д}}$ – коэффициент доплат и надбавок (0,2-0,5);

K_p – районный коэффициент (1.3);

T_p – продолжительность работ, выполняемых работником, раб. дни.

Среднемесячная заработная плата рассчитывается по формуле:

$$Z_{\text{дн}} = \frac{Z_m \cdot M}{F_d}, \quad (9)$$

где Z_m – месячный оклад работника, руб.;

M – количество месяцев работы без отпуска в течение года: (10.4);

F_d – действительный годовой фонд рабочего времени персонала: (243).

Среднемесячная зарплата студента составляет:

$$Z_{\text{дн}} = \frac{Z_m \cdot M}{F_d} = \frac{21760 \cdot 10.4}{243} \approx 931.29 \text{ рублей.} \quad (10)$$

Среднемесячная зарплата руководителя составляет:

$$Z_{\text{дн}} = \frac{Z_m \cdot M}{F_d} = \frac{33664 \cdot 10.4}{243} \approx 1440.76 \text{ рублей.} \quad (11)$$

Расчет основной заработной платы исполнителей представлен в таблице 13.

Таблица 13 – Расчет основной заработной платы

Исполнители	$Z_{\text{дн}}$, руб.	$K_{\text{пр}}$	$K_{\text{д}}$	K_p	T_p , дни	$Z_{\text{осн}}$, руб.
Чурсина Елена Андреевна	931.29	0.3	0.2	1.3	80	145281.2
Кудинов Антон Викторович	1440.76	0.3	0.2	1.3	8	22475.86
Итого:						167757.096

5.2.3.4 Дополнительные расходы

Дополнительная заработная плата работника составляет 12% от его основной заработной платы и рассчитывается по формуле:

$$З_{\text{доп}} = 0.12 \cdot З_{\text{осн}}. \quad (12)$$

Страховые отчисления составляют 30% от суммарной заработной платы работника и рассчитываются по формуле:

$$O_{\text{страх}} = 0.3 \cdot (З_{\text{осн}} + З_{\text{доп}}). \quad (13)$$

Расчет дополнительной заработной платы и страховых отчислений с заработной платы исполнителей представлен в таблице 14.

Таблица 14 – Расчет дополнительной з/п и страховых взносов

Исполнители работ	З _{осн} , руб.	З _{доп} , руб.	З _{общ} , руб.	O _{страх} , руб.
Чурсина Елена Андреевна	145281.2	17433.75	162715	48814.5
Кудинов Антон Викторович	22475.86	2697.103	25172.96	7551.888
Итого:	167757.1	20130.85	187887.9	56366.38

5.2.3.5 Формирование бюджета затрат научно-исследовательского проекта

В таблице 15 представлен расчет накладных расходов на проведение исследования.

Таблица 15 – Расчет накладных расходов

Наименование	Сумма, руб.
Материальные затраты	2000
Затраты на специальное оборудование	8444.44
Затраты на основную заработную плату	167757.1
Затраты на дополнительную заработную плату	20130.85
Страховые взносы	56366.38
Итого (P _{сум}):	254698.77

Накладные расходы составляют 16% от общей суммы расходов и рассчитываются по формуле:

$$P_{\text{накл}} = 0.16 \cdot P_{\text{сум}}. \quad (14)$$

Статьи расходов и их удельный вес в бюджете затрат исследования представлены в таблице 16.

Таблица 16 – Расчет бюджета затрат НТИ

Наименование	Сумма, руб.	Удельный вес, %
Материальные затраты	2000	0.38
Затраты на специальное оборудование	8444.44	2.76
Затраты на основную заработную плату	167757.1	57.05
Затраты на дополнительную заработную плату	20130.85	6.85
Страховые взносы	56366.38	19.17
Накладные расходы	40751.80	13.79
Общий бюджет	295450.57	100.00

5.4 Определение потенциального эффекта исследования

В результате проведения исследования по разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение» были определены:

- Потенциальные потребители результатов разработки – пользователи тромбоэластографа;
- Итоговая оценка конкурентности решения – 8.76;
- Приоритетное направление развития разработки – добавление возможности высокой точности диагностики уникальных пациентов;
- Возможные угрозы для разработки и способы нивелирования их воздействия на нее;
- Работы, необходимые для реализации разработки;
- Общая длительность исследования, равная 88 рабочих день;
- Общий бюджет НТИ, составляющий 295450.57 рублей, основную часть которого обеспечивает Томский политехнический университет.

Рассматриваемая разработка позволит улучшить качество технологию оценки системы гемостаза человека с помощью тромбоэластографа и ускорить обучение специалистов.

6 Социальная ответственность

Результатом выпускной квалификационной работы является модуль поддержки принятия решений в процессе оценки состояния системы гемостаза. Разработанный в результате выполнения выпускной квалификационной работы модуль после внедрения является частью ПО тромбоэластографа. Модуль способен с высокой точностью определять одиннадцать различных гемостазиологических диагнозов на основе анализа крови человека, проведенного с помощью тромбоэластографа. Внедрение модуля в ПО тромбоэластографа позволяет ускорить обучение специалистов и улучшить качество диагностики гемостазиологических расстройств.

Тромбоэластографы активно находят применение в современной медицине. Потребителями результатов данной выпускной квалификационной работы являются медицинские специалисты в нескольких субъектах Российской Федерации. Разработка модуля и эксплуатация ПО, в состав которого входит модуль, осуществляется на ПЭВМ.

6.1 Правовые и организационные вопросы обеспечения безопасности

Требования для выбранного рода деятельности задаются правовыми и организационными вопросами. Так как эксплуатация модуля происходит на ПЭВМ, требуется соблюдение правовых норм к ПЭВМ и организации рабочего места. Согласно СанПиН 2.2.2/2.4.1340-03 [16], при организации рабочего места должны быть соблюдены следующие требования:

- Требования к ПЭВМ;
- Требования к помещениям для работы с ПЭВМ;
- Требования к микроклимату, содержанию аэроионов и вредных химических веществ в воздухе на рабочих местах, оборудованных ПЭВМ;
- Требования к уровням шума и вибрации на рабочих местах, оборудованных ПЭВМ;

- Требования к освещению на рабочих местах, оборудованных ПЭВМ;
- Требования к уровням электромагнитных полей на рабочих местах, оборудованных ПЭВМ;
- Общие требования к организации рабочих мест пользователей ПЭВМ;
- Требования к организации медицинского обслуживания пользователей ПЭВМ;
- Требования к проведению государственного санитарно-эпидемиологического надзора и производственного контроля.

Регулирование требований к продолжительности рабочего времени и оплаты труда осуществляется законодательством РФ, а именно трудовым кодексом РФ [17].

Нормальная продолжительность рабочего времени не может превышать 40 часов в неделю. Продолжительность рабочего времени и размер оплаты труда для конкретного работника устанавливается трудовым договором.

Разработка модуля и эксплуатация ПО, в состав которого входит модуль, осуществляется на ПЭВМ с соблюдением всех вышеперечисленных требований.

6.2 Производственная безопасность

Во время разработки модуля эксплуатации ПО, в состав которого входит модуль, могут возникать вредные и опасные факторы. Выявленные факторы представлены в таблице 17.

Таблица 17 – Возможные опасные и вредные факторы

Факторы (ГОСТ 12.0.003-2015)	Этапы работы		Нормативные документы
	Разр.	Эксп.	
1. Отклонение показателей микроклимата	+	+	СанПиН 2.2.4.548-96 [18]
2. Превышение уровня шума на рабочем месте	+	+	СанПиН 2.2.4/2.1.8.562-96 [19], ГОСТ 12.1.029-80 [20]
3. Недостаточная освещенность рабочей зоны	+	+	ГОСТ Р 55710-2013 [21]
4. Нарушение предельно допустимых значений напряжений прикосновения и токов	+	+	ГОСТ 12.1.038-82 ССБТ [22]

6.2.1 Анализ выявленных вредных и опасных факторов

6.2.1.1 Отклонение параметров микроклимата

Комфортные условия для работы создаются оптимальным сочетанием температуры, относительной влажности и скорости движения воздуха. На рабочих местах пользователей персональных компьютеров должны обеспечиваться оптимальные параметры микроклимата в соответствии с СанПин 2.2.2/2.4.1340-03 [16].

Согласно этому документу должны быть соблюдены следующие требования:

- Температура воздуха должна быть в холодный период года не более 22-24 °С, в теплый период года – 20-25 °С;
- Относительная влажность должна составлять 40-60%;

- Скорость движения воздуха должна быть 0.1 м/с.

Для поддержания оптимальных значений микроклимата используется система отопления и кондиционирования воздуха. В кабинетах, где выполнялась разработка и работают пользователи ПО тромбоэластографа, параметры микроклимата соответствуют требованиям СанПиН.

6.2.1.2 Превышение уровня шума на рабочем месте

Шум является одним из распространенных в производстве вредных факторов. Его создают работающее оборудование, преобразователи напряжения, работающие осветительные приборы дневного света и другие источники шума.

Шум может стать причиной снижения работоспособности и повышенной утомляемости. Значительные превышения уровня шума на рабочем месте вызывают необратимые изменения в органах слуха человека, также оказывают неблагоприятное влияние на весь организм человека через нервную систему. В результате ослабляется внимание, ухудшается память, снижается реакция, что вызывает увеличение числа ошибок при работе.

Требования к допустимому уровню шума описаны в СанПиН 2.2.2/2.4.1340-03 [23]. Согласно данному документу допустимый уровень шума составляет не более 50 дБА.

Допустимый уровень шума на рабочих местах пользователей персональных компьютеров не превышает установленного значения. Шумящее оборудование (АЦПУ, принтеры и т.п.), уровни шума которого превышают нормированные, находится вне помещения с ПЭВМ.

6.2.1.3 Недостаточная освещенность рабочей зоны

Естественное и искусственное освещение рабочего места оказывает влияние на физическое и психологическое состояние пользователя, что неблагоприятно сказывается на его работе. Не надлежащего качество освещения может привести к ухудшению зрения.

Согласно СанПиН 2.2.2/2.4.1340–03 [16] должны быть соблюдены следующие требования к освещенности рабочей зоны:

- На поверхности рабочего стола пользователя ПЭВМ освещенность должна быть 300 – 500 лк;
- При освещении блики должны отсутствовать на поверхности экрана;
- Поверхность экрана должна быть до 300 лк;

Далее перечислены общие требования и рекомендации к организации освещения на рабочем месте.

- Система общего равномерного должно регулировать искусственное освещение в помещениях для эксплуатации ПЭВМ;
- Рабочие места следует размещать таким образом, чтобы естественный свет падал преимущественно слева, а дисплеи монитора были ориентированы боковой стороной к световым проемам.

Вышеперечисленные меры полностью соблюдаются, что позволяет сохранить зрение и избежать пагубного воздействия на глаза во время разработки и эксплуатации результатов выпускной квалификационной работы.

6.2.1.4 Нарушение предельно допустимых значений напряжений прикосновения и токов

Напряжение прикосновения – напряжение, появляющееся на теле человека при одновременном прикосновении к двум точкам проводников или проводящих частей, в том числе при повреждении изоляции. Напряжение может возникнуть на корпусах и каркасах оборудования, в частности ПЭВМ, в случае полного или частичного повреждения электрической изоляции оборудования или питающих оборудование кабелей.

Воздействие напряжения отрицательно влияет на здоровье человека. Стандарт ГОСТ 12.1.038-82 ССБТ [16] устанавливает предельно допустимые

значения напряжений прикосновений и токов, протекающих через тело человека.

Напряжения прикосновения и токи, протекающие через тело человека при нормальном (неаварийном) режиме работы, не должны превышать значений, указанных в таблице 18.

Таблица 18 – Возможные опасные и вредные факторы

Род тока	U, В	I, mA
	Не более	
Переменный, 50 Гц	2.0	0.3
Переменный, 400 Гц	3.0	0.4
Постоянный	8.0	1.0

Для контроля предельно допустимых значений напряжений прикосновения и токов измеряют напряжения и токи в местах, где может произойти замыкание электрической цепи через тело человека.

В кабинетах, где происходила разработка и происходит эксплуатация модуля, осуществляется регулярный контроль предельно допустимых значений напряжений прикосновения и токов на корпусах ПЭВМ и питающих их кабельных линиях. Все значения удовлетворяют требованиям.

6.2.2 Обоснование мероприятий по снижению воздействия опасных и вредных факторов на работающего

При организации рабочего места и режима труда следует учитывать перечисленные требования безопасности и промышленной санитарии. Невыполнение этих требований может привести к повышению уровня воздействия опасных и вредных факторов на работающего.

6.3 Экологическая безопасность

Непосредственно разработка модуля и эксплуатация ПО, в состав которого входит модуль, не оказывают воздействие на окружающую среду. Негативное влияние на атмосферу, гидросферу и литосферу могут оказывать

используемые для разработки и эксплуатации вещества. Требования к обращению с отходами описаны в ГОСТ Р 53692-2009 [23].

ПЭВМ, батарейки, оргтехнику в случае выхода из строя, либо по причине морального износа необходимо утилизировать через специализированные организации, имеющие лицензию на данную деятельность.

Отходы от канцелярских товаров, продуктов питания и личной гигиены и другие твердые отходы в случае их неправильной утилизации могут оказывать негативное воздействие на почечный покров. Защита почечного покрова от данного воздействия реализуется за счет сбора, сортировки и утилизации отходов и их организованного захоронения.

Все отходы, образующиеся в ходе разработки и эксплуатации результатов выпускной квалификационной работы, утилизируются без оказания негативного влияния на экологию.

6.4 Безопасность в чрезвычайных ситуациях

К чрезвычайным ситуациям, которые могут возникнуть во время разработки или эксплуатации можно отнести возможность возникновения и распространения пожара, так как работа происходит на ПЭВМ. В современных ПЭВМ очень высокая плотность размещения элементов электронных схем, также в непосредственной близости друг от друга располагаются соединительные провода и кабели. При протекании по ним электрического тока выделяется значительное количество теплоты, при этом возможно оплавление изоляции и возникновение возгорания. Возникновение других видов чрезвычайных ситуаций – маловероятно.

Необходимо соблюдение техники пожарной безопасности для предотвращения возникновения и распространения пожара. Общие требования к пожарной безопасности описаны в ГОСТ 12.1.004-91 [24].

Далее перечислены правила поведения в случае возникновения данной чрезвычайной ситуации.

– В случае возникновения пожара люди должны покинуть помещение в течение минимального времени согласно плану эвакуации.

– В помещениях с компьютерной техникой, недопустимо применение воды и пены ввиду опасности повреждения или полного выхода из строя дорогостоящего электронного оборудования.

– Для тушения пожаров необходимо применять углекислотные и порошковые огнетушители, которые обладают высокой скоростью тушения, большим временем действия, возможностью тушения электроустановок, высокой эффективностью борьбы с огнем. Воду разрешено применять только во вспомогательных помещениях.

– Помещение должно быть оборудовано пожарными извещателями, которые позволяют оповестить дежурный персонал о пожаре.

В кабинетах, где происходила разработка и происходит эксплуатация модуля, были соблюдены все правила техники безопасности. Сотрудникам были объяснены правила поведения в случае возникновения и распространения пожара.

6.5 Заключение по разделу

В рамках раздела «Социальная ответственность» были определены требования, необходимые для соблюдения, а также методы, позволяющие минимизировать влияние разработки и эксплуатации модуля на экологию, а также создать комфортные и безопасные условия для разработки модуля и эксплуатации ПО, в состав которого входит модуль.

ЗАКЛЮЧЕНИЕ

В данной выпускной квалификационной работе были выполнены следующие задачи:

- изучен принцип оценки состояния системы гемостаза с помощью тромбоэластографа, изучены алгоритмы машинного обучения для решения задачи классификации;
- разработан алгоритм классификации гемостазиологических диагнозов;
- реализован модуль поддержки принятия решений;
- модуль внедрён в ПО тромбоэластографа;
- составлена отчетная документация о выполненной работе.

Таким образом, цель данной работы, разработка модуля поддержки принятия решений в процессе оценки состояния системы гемостаза, достигнута.

Модуль был внедрен в ПО клинико-диагностического комплекса АРП-01М «МЕДНОРД». На его основе была разработана система поддержки принятия решений «Вектор», используемая в клинической практике лечебных учреждений 56 субъектов РФ. Получен акт о внедрении, представленный в приложении Ж.

Результаты по теме ВКР были представлены на XVI Международной научно-практической конференции студентов, аспирантов и молодых ученых «МСИТ-2018». Доклад отмечен дипломом, приведенным в приложении З.

В ходе выполнения данной работы был получен дополнительный опыт анализа данных, разработки алгоритмов машинного обучения, навыки работы со сторонним ПО и опыт работы в компании ООО «Рубиус-Групп» в роли программиста.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Шишонок А.И., Щербакова И.Г., Гребенникова И.В. СОВРЕМЕННЫЕ АСПЕКТЫ ГЕМОСТАЗА // Международный студенческий научный вестник. – 2015. – № 2-1.;
2. Приказ Минздрава РФ ОТ 25.12.97 N 380 "О СОСТОЯНИИ И МЕРАХ ПО СОВЕРШЕНСТВОВАНИЮ ЛАБОРАТОРНОГО ОБЕСПЕЧЕНИЯ ДИАГНОСТИКИ И ЛЕЧЕНИЯ ПАЦИЕНТОВ В УЧРЕЖДЕНИЯХ ЗДРАВООХРАНЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ" // zakonbase [Электронный ресурс]. <https://zakonbase.ru/content/base/70698>
3. Тютрин И. И. Низкочастотная пьезотромбоэластография цельной крови: алгоритмы диагностики и коррекции гемостазиологических расстройств / И. И. Тютрин, В. В. Удут ; Нац. исслед. Том. гос. ун-т, Том. нац. исслед. мед. центр РАН НПО "Меднорд" (г. Томск). – Томск: Издательский Дом Томского государственного университета, 2016.
4. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. – М.: Финансы и статистика, 1989.
5. Рашка С. Python и машинное обучение / пер. с англ. А. В. Логунова. М.: ДМК Пресс, 2017. – 418 с.
6. Breiman L. Bagging predictors // Machine Learning, 1996, vol. 24, no. 2, pp. 123–140.
7. Открытый курс машинного обучения. Тема 5. Композиции: бэггинг, случайный лес. [Электронный ресурс]. URL: <https://habr.com/ru/company/ods/blog/324402/#3-ocenka-vazhnosti-priznakov>
8. Открытый курс машинного обучения. Тема 10. Градиентный бустинг. [Электронный ресурс]. URL: <https://habr.com/ru/company/ods/blog/327250/#4-itog-pro-teoriyu-gbm>

9. Применение нейронных сетей для задач классификации // basegroup. [Электронный ресурс]. URL: <https://basegroup.ru/community/articles/classification>
10. Метод деревьев решений для задачи классификации // edu.kpfu. [Электронный ресурс]. URL: https://edu.kpfu.ru/pluginfile.php/91556/mod_resource/content/3/Decision%20trees_1.pdf
11. Leo Breiman. Random Forests // Machine Learning, October 2001, Volume 45, Issue 1, pp 5-32.
12. Окунь Я. Факторный анализ. Перевод с польского Г. З. Давидовича. М., 1974 г. – 100 с.
13. К. В. Воронцов. Лекции по алгоритмическим композициям // 7 октября 2012 г.
14. Метрики в задачах машинного обучения. [Электронный ресурс]. URL: <https://habr.com/ru/company/ods/blog/328372/>
15. SHAP [Электронный ресурс]. URL: <https://github.com/slundberg/shap>.
16. СанПиН 2.2.2/2.4.1340-03. Гигиенические требования к персональным электронно-вычислительным машинам и организации работы: Санитарно-эпидемиологические правила и нормы. – М.: Федеральный центр госсанэпиднадзора Минздрава России, 2003. – 54 с.
17. ГОСТ 12.4.011-89. Система стандартов безопасности труда. Средства защиты работающих. Общие требования и классификация. – М: ИПК издательство стандартов, 2001.
18. СанПиН 2.2.4.548-96. Гигиенические требования к микроклимату производственных помещений: Санитарные правила и нормы: – М.: Информационно-издательский центр Минздрава России, 2001. – 20 с
19. СанПиН 2.2.4/2.1.8.562-96. Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой

застройки. – М.: Информационно-издательский центр Минздрава России, 1997.

20. ГОСТ 12.1.029-80. Средства и методы защиты от шума. М.: Система стандартов безопасности труда: Сб. ГОСТов. – М.: ИПК Издательство стандартов, 2001.

21. ГОСТ Р 55710-2013. Освещение рабочих мест внутри зданий. – М.: Официальное издание. М.: Стандартиформ, 2016.

22. ГОСТ 12.1.038-82 ССБТ. Электробезопасность. Предельно допустимые значения напряжений прикосновения и токов. – М.: ИПК Издательство стандартов, 2001.

23. ГОСТ Р 53692-2009. Ресурсосбережение. Обращение с отходами. Этапы технологического цикла отходов. – М.: ИПК Издательство стандартов, 2001 год.

24. ГОСТ 12.1.004-91. ССБТ Пожарная безопасность. Общие требования. – М.: ИПК Издательство стандартов, 2001.

25. Davy Cielen, Arno D.V. Meysman, Mohamed Ali Introducing Data Sciences – Manning Publications Co., 2016.

26. Конспекты курса МФТИ «Введение в машинное обучение», специализация – «Машинное обучение и анализ данных». [Электронный ресурс]. URL: <http://www.machinelearning.ru>

27. Удут В.В., Тютрин И.И. Технология низкочастотной пьезотромбоэластографии в оценке гемостатического потенциала // Вестник новых медицинских технологий, электронный журнал – 2016.

28. Спицын В.Г. Применение искусственных нейронных сетей для обработки информации: Методические указания к лабораторным работам / Спицын В.Г., Цой Ю.Р. – Томск: Издательство ТПУ, 2007 – 32 с.

29. Mark R. Segal. Machine Learning Benchmarks and Random Forest Regression // Division of Biostatistics, University of California, San Francisco, CA 94143-0560, April 14, 2003.

30. Caruana R., Niculescu-Mizil A. An Empirical Comparison of Supervised Learning Algorithms // Department of Computer Science, Cornell University, Ithaca, NY 14853 USA.

31. L. Mason , J. Baxter , P. Bartlett , M. Fren. Boosting Algorithms as Gradient Descent // 2000

**ПРИЛОЖЕНИЕ А Описание показателей НПТЭГ и их характеристик
для набора данных**

Таблица 19 – Описание показателей НПТЭГ и их характеристик для набора данных о реальных исследованиях образцов цельной крови

Имя	Назначение	Количество отсутствующих значений	Среднее значение	Мин. значение	Макс. значение
A0	Начальная амплитуда НПТЭГ	0	149.8568	17	301
T1	Время достижения минимальной амплитуды НПТЭГ	1	1.7787	0	13.5
A1	Минимальная амплитуда НПТЭГ	1	120.2874	-3	288
T2	Время достижения амплитуды «A2»	0	5.2504	0.1	30.3
A2	Увеличение амплитуды НПТЭГ на 100 о.е.	0	220.6138	60	636
T3	Время свертывания крови	1	9.7028	1	59.4
A3	Значение амплитуды НПТЭГ во время свертывания крови	1	437.3188	96	982
T4	Время через 10 минут после свертывания крови	10	19.7225	11	69.4
A4	Амплитуда через 10 минут после свертывания крови	10	609.4307	25	1000
T5	Время достижения максимальной амплитуды	1	39.3893	1.8	90
A5	Максимальная амплитуда НПТЭГ	1	706.5990	102	1046
T6	Время через 10 минут после достижения максимального значения амплитуды НПТЭГ	669	48.6155	0	85
A6	Значение амплитуды НПТЭГ через 10 минут после достижения максимального значения	669	663.7230	0	1040
IRLS	Интенсивность лизиса и ретракции сгустка	669	2.3998	-1.17	127.96
MA	Максимальная плотность сгустка	1	586.3382	19	980

Продолжение таблицы 19

ITS	Интенсивность тотального свертывания	1	16.5636	3.5	201.9 1
ИКК	Интенсивность контактной фазы коагуляции	1	-20.7492	-145	80
IKD	Интенсивность контактной коагуляции	1	48.4449	1.79	936
КТА	Константа тромбиновой активности	0	48.8539	0	1000
IPS	Интенсивность полимеризации сгустка	10	17.1401	-13.1	36.5

ПРИЛОЖЕНИЕ Б Рекомендации к диагностике гемостазиологических расстройств

Таблица 20 – Рекомендации к диагностике гемостазиологических расстройств

Состояние структуры свертывания крови	Состояние хронометрии свертывания крови	Диагноз	Метка диагноза
Гиперкоагуляция	Гиперкоагуляция	Гиперкоагуляция	1
Норма или тенденции	Гиперкоагуляция	Гиперкоагуляция хронометрическая	2
Гиперкоагуляция	Норма или тенденции	Гиперкоагуляция структурная	3
Норма или тенденции	Норма или тенденции	Норма с тенденцией к гиперкоагуляции	4
Норма или тенденции	Норма или тенденции	Условная норма	5
Норма или тенденции	Норма или тенденции	Норма с тенденцией к гипокоагуляции	6
Гипокоагуляция	Норма или тенденции	Гипокоагуляция структурная	7
Норма или тенденции	Гипокоагуляция	Гипокоагуляция хронометрическая	8
Гипокоагуляция	Гипокоагуляция	Гипокоагуляция	9
Гиперкоагуляция	Гипокоагуляция	Структурная гиперкоагуляция - Хронометрическая гипокоагуляция	10
Гипокоагуляция	Гиперкоагуляция	Структурная гипокоагуляция - Хронометрическая гиперкоагуляция	11

ПРИЛОЖЕНИЕ В Матрица корреляции параметров НПТЭГ

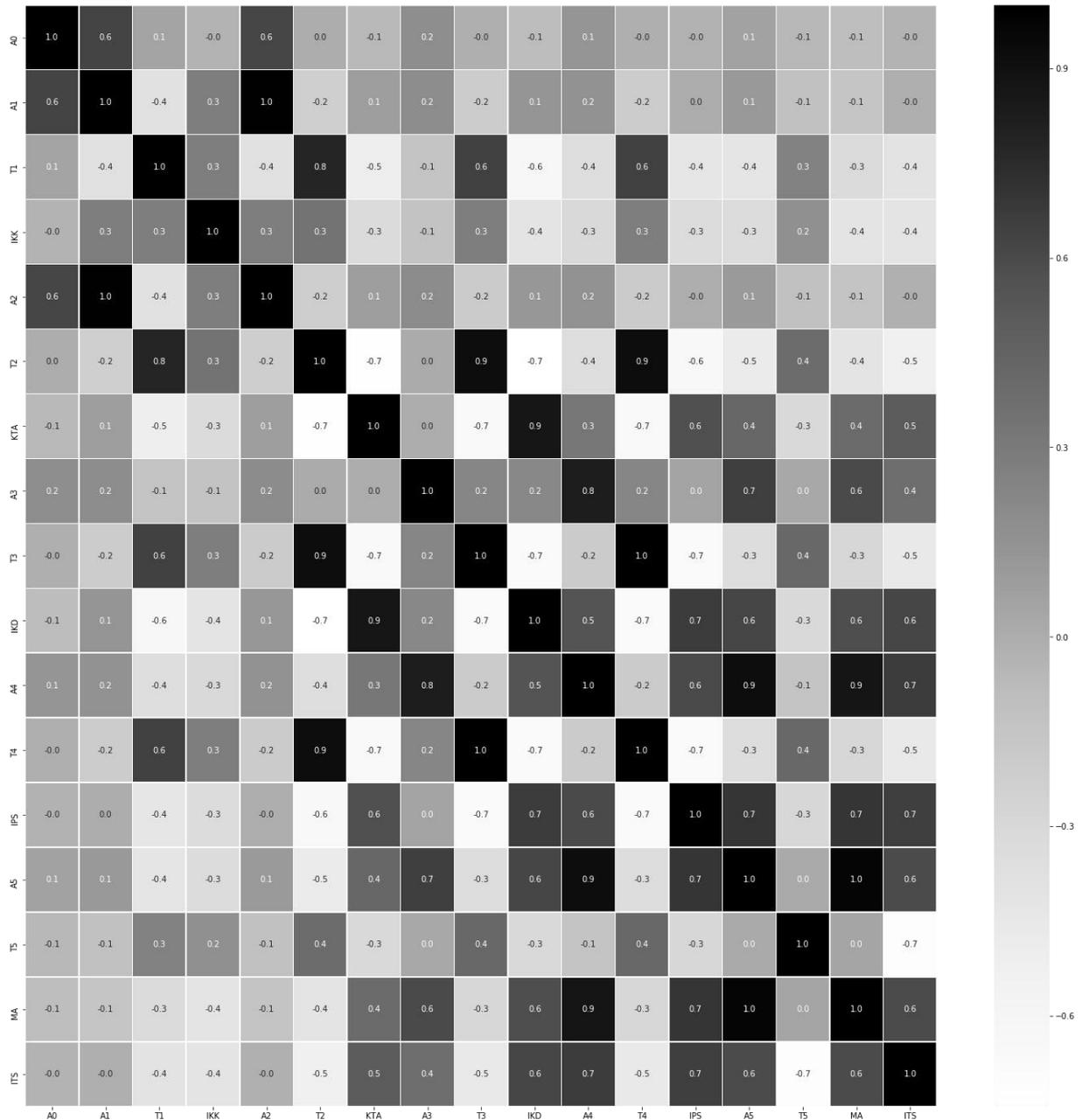


Рисунок 14 – Матрица корреляции параметров НПТЭГ

ПРИЛОЖЕНИЕ Г Результаты настройки классификаторов на основе градиентного бустинга и случайного леса

Таблица 21 – Результаты настройки классификаторов на основе градиентного бустинга

Параметр	Для 11 диагнозов	Для первого этапа многоэтапного классификатора	Для структуры (второй этап многоэтапного классификатора)	Для хронометрии (второй этап многоэтапного классификатора)	Для пар диагнозов 1 и 2	Для пар диагнозов 5 и 6	Для пар диагнозов 8 и 9
Количество используемых признаков	9	11	8	13	10	4	9
Скорость обучения	0.2	0.3	0.1	0.1	0.2	0.1	0.2
Минимальное число объектов, при котором выполняется расщепление	52	152	102	102	2	102	102
Максимальная глубина деревьев	8	6	6	6	4	2	4

Таблица 22 – Результаты настройки классификаторов на основе случайного леса

Параметр	Для 11 диагнозов	Для первого этапа многоэтапного классификатора	Для структуры (второй этап многоэтапного классификатора)	Для хронометрии (второй этап многоэтапного классификатора)	Для пар диагнозов 1 и 2	Для пар диагнозов 5 и 6	Для пар диагнозов 8 и 9
Количество используемых признаков	10	16	9	8	14	12	9
Минимальное число объектов, при котором выполняется расщепление	0.2	2	4	6	18	12	2
Максимальная глубина деревьев	17	11	14	8	8	5	5

ПРИЛОЖЕНИЕ Д Схемы трёхэтапного классификатора

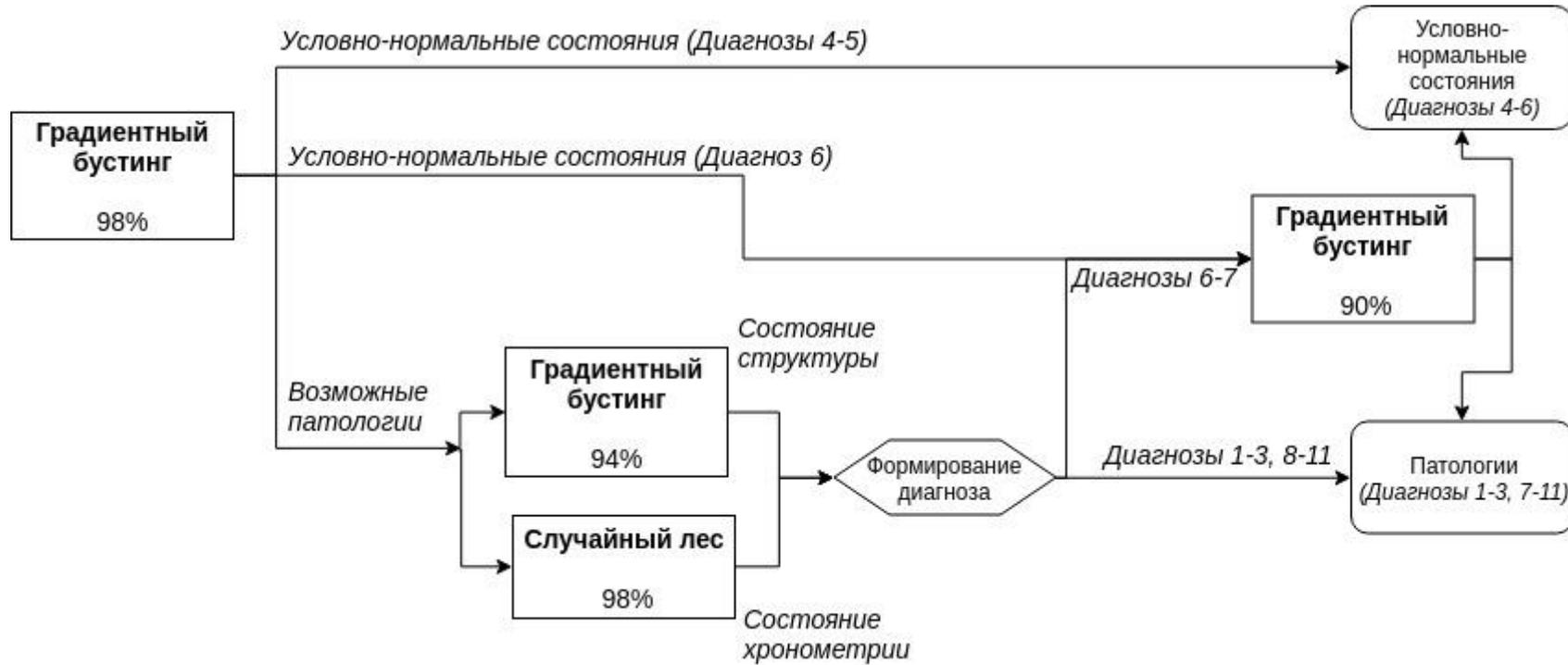


Рисунок 15 – Схема трёхэтапного классификатора

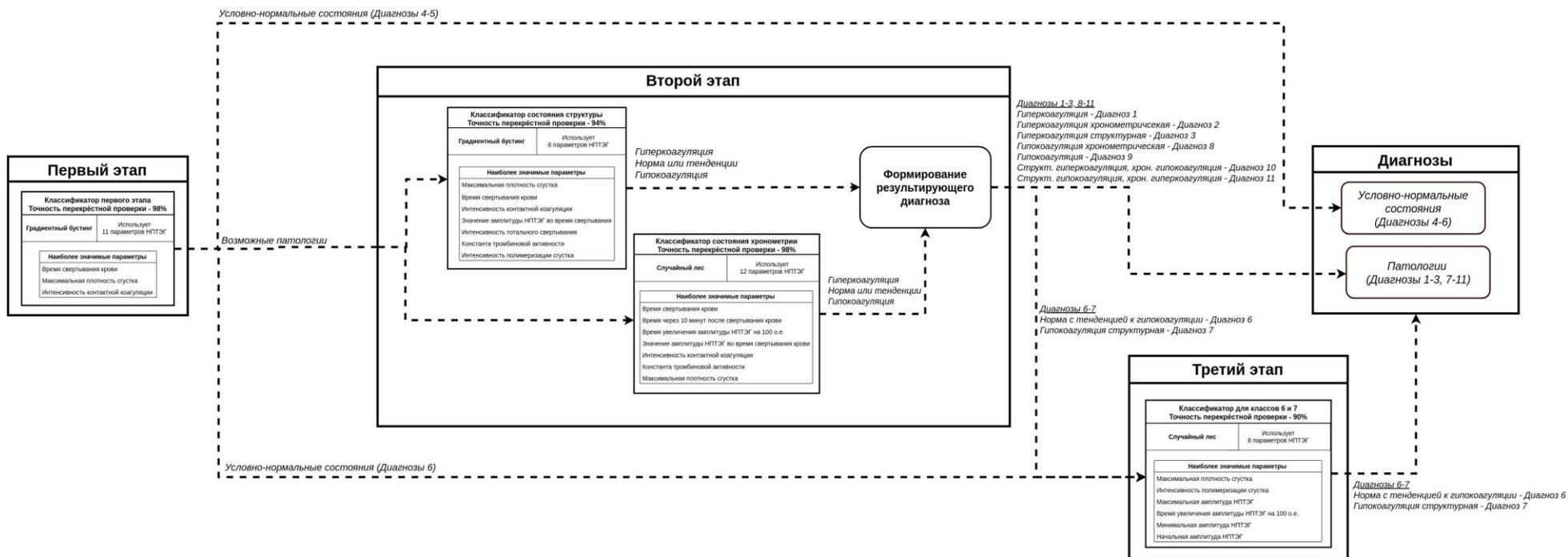


Рисунок 16 – Подробная схема трёхэтапного классификатора

ПРИЛОЖЕНИЕ Е Диаграмма Ганта

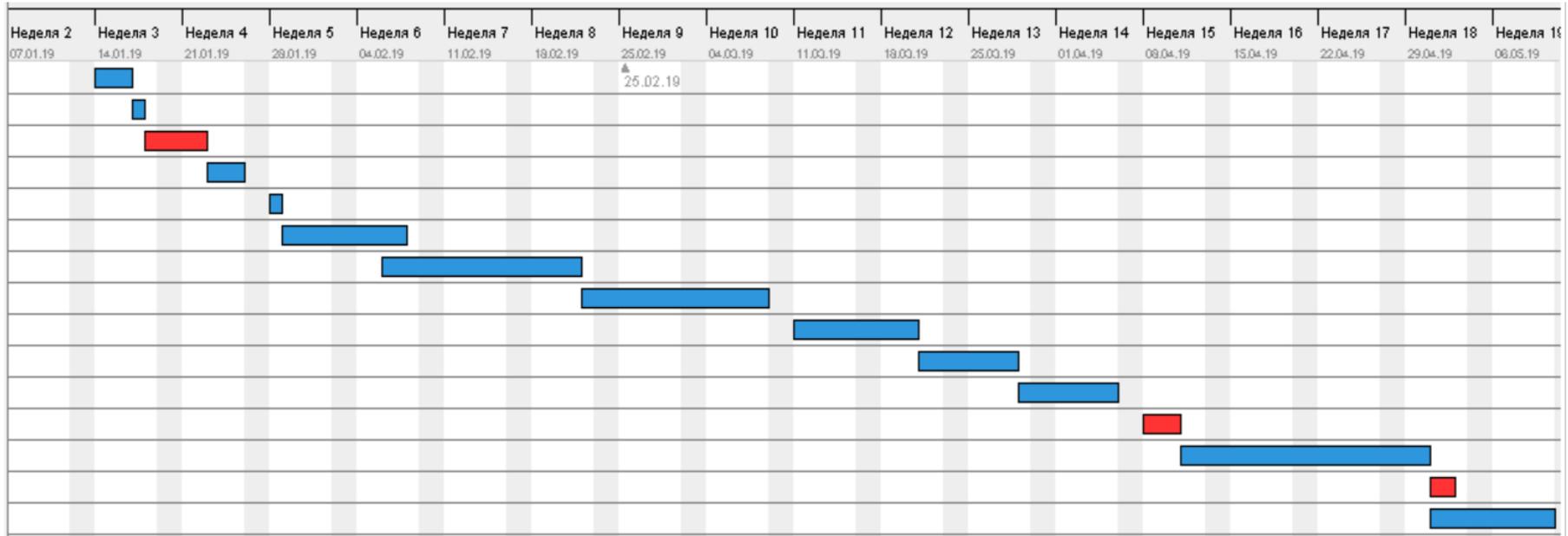


Рисунок 17 – Диаграмма Ганта

Таблица 19 – Расшифровка цветowych обозначений

Цветовое обозначение	Расшифровка
	Кудинов Антон Викторович
	Чурсина Елена Андреевна

ПРИЛОЖЕНИЕ Ж Акт ввода в эксплуатацию

Общество с ограниченной ответственностью
«Меднорд-Техника»
634029, Россия, г. Томск, ул. Белинского, д.38
Тел./Факс.: 8 (3822) 52-75-22
E-mail: info@mednord-t.ru
ОГРН: 1107017013135
ИНН/КПП: 7017266072/701701001

АКТ ВНЕДРЕНИЯ

27 мая 2019 г.

Мы, нижеподписавшиеся, профессор, д-р мед. наук, научный руководитель ООО «Меднорд-Техника», профессор кафедры Анестезиологии и реаниматологии СибГМУ Тютрин Иван Илларионович, технический директор ООО «Рубиус Групп» канд. техн. наук Кудинов Антон Викторович и студент ФГАОУ ВО «Национальный исследовательский Томский политехнический университет» Чурсина Елена Андреевна составили акт о внедрении результатов выпускной квалификационной работы на тему «Разработка модуля поддержки принятия решений в процессе диагностики функционального состояния системы гемостаза» в аппаратно-программный комплекс для клинико-диагностических исследований реологических свойств крови АРП-01М «МЕДНОРД» (Регистрационное удостоверение № ФСР 2010/09767).

Объект внедрения: Модуль поддержки принятия решений в процессе диагностики функционального состояния системы гемостаза.

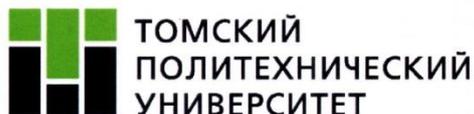
Характеристика полученного результата:

- впервые разработана программа, позволяющая создавать базы данных как условно здоровых лиц различных популяций, так и больных с различной патологией (Свидетельство о государственной регистрации № 2019620555 «База данных пользователей, характеризующих состояние гемостатического потенциала, условно здоровых добровольцев сибирской популяции»);
- на основании данного модуля разработана система поддержки принятия решений «Вектор», внедренная в клиническую практику лечебных учреждений 56 субъектов Российской Федерации (Свидетельство о государственной регистрации № 2019615166 «Система поддержки принятия решений «Вектор»»).

Научный руководитель ООО «Меднорд-Техника»
Технический директор ООО «Рубиус Групп»
Студент ФГАОУ ВО «НИ ТПУ»



ПРИЛОЖЕНИЕ 3 Награды по теме работы



ДИПЛОМ

II степени

НАГРАЖДАЕТСЯ

ЧУРСИНА ЕЛЕНА АНДРЕЕВНА

студентка Томского политехнического университета

за доклад, представленный на

XVI Международной научно-практической конференции
студентов, аспирантов и молодых ученых
«Молодёжь и современные информационные технологии»

«РАЗРАБОТКА СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЯ ДЛЯ
ОЦЕНКИ СИСТЕМЫ РЕГУЛЯЦИИ АГРЕГАТНОГО СОСТОЯНИЯ
КРОВИ»

3-7 декабря 2018 г.

г. Томск

Проректор по научной работе и
инновациям

И.Б. Степанов

