Казахский агротехнический исследовательский университет
им. С. Сейфуллина

УДК 004.8:614.39 (043)

На правах рукописи

КАДИРКУЛОВ КУАНЫШ КАЙСАРОВИЧ

Разработка модели искусственного интеллекта по лабораторной диагностике в здравоохранении

8D06101 – Аналитика больших данных

Диссертация на соискание степени
доктора философии (PhD)

Научный консультант
доктор PhD,
ассоциированный профессор
А.А. Исмаилова

Зарубежный научный консультант
доктор PhD,
профессор
Ainuddin Wahid Abdul Wahab
(Малайзия)

Республика Казахстан
Астана, 2023

СОДЕРЖАНИЕ

НОРМАТИВНЫЕ ССЫЛКИ

В настоящей диссертации, использованы ссылки на следующие стандарты:

Инструкция по оформлению диссертации и автореферата, ВАК МОН РК, №377-3ж.

ГОСТ 7.32-2017. Система стандартов по информации, библиотечному и издательскому делу. «Отчет о научно-исследовательской работе. Структура и правила оформления».

СТ РК 34.005-2002. Информационная технология. Основные термины и определения.

СТ РК 34.006-2002. Информационная технология. Базы данных. Основные термины и определения.

СТ РК 34.014-2002. Информационная технология. Комплекс стандартов на автоматизированные системы. Автоматизированные системы. Термины и определения.

СТ РК 34.019-2005. (ISOЛЕС 12207:1995, MOD) Информационная технология. Процессы жизненного цикла программных средств.

Указ Президента Республики Казахстан. О Государственной программе «Информационный Казахстан-2020» и внесении дополнения в Указ Президента Республики Казахстан от 19 марта 2010 года, №957 «Об утверждении Перечня государственных программ»: утв. 8 января 2013 года, №464.

Закон Республики Казахстан. Об информатизации: принят 24 ноября 2015 года, №418.

Постановление Республики Казахстан. Об утверждении Государственной программы «Цифровой Казахстан»: утв. 12 декабря 2017 года, №827.

Закон Республики Казахстан. О науке: принят 18 февраля 2011 года, №407-IV3РК.

ГОСТ 7.1-2003. Библиографическая запись. Библиографическое описание. Общие требования и правила составления.

Кодекс Республики Казахстан. О здоровье народа и системе здравоохранения: принят 7 июля 2020 года, №360-VI ЗРК.

ОПРЕДЕЛЕНИЯ

В настоящей диссертации используются следующие термины с соответствующими определениями:

База данных – систематически организованное или структурированное хранилище индексированной информации, позволяющее легко получать. обновлять, анализировать и выводить данные.

Большие данные – наборы данных, используемые для обозначения огромного объема данных, которые трудно обработать с использованием традиционных методов работы с базами данных и программного обеспечения.

Искусственный интеллект – представляет собой дисциплину в области информатики, направленную на конструирование интеллектуальных вычислительных систем. Под интеллектуальными системами подразумеваются системы, обусловленные способностями, традиционно ассоциируемыми с человеческим интеллектом, включая понимание языка, процесс обучения, способность к логическим рассуждениям, решение проблем и т. д.

Извлечение информации – данные анализируются и просматриваются для извлечения соответствующей информации из источников данных по определенному шаблону. Большая часть извлечения данных происходит из неструктурированных источников данных и различных форматов данных.

Информационный поиск – алгоритм, который используется для извлечения информации, хранящейся в некоторой структуре данных или вычисленной в пространстве поиска проблемной области, либо с дискретными или с непрерывными значениями.

Кластер – представляет собой группу серверов и других ресурсов, которые действуют как единая система и обеспечивают высокую доступность н некоторых случаях, распределение нагрузки и В параллельную обработку.

Машинное обучение – представляет собой подкласс методологий искусственного интеллекта, отличительной характеристикой которых является не прямолинейное решение задачи, а обучение посредством применения решений в рамках множества схожих задач. Конструкция данных методов обуславливается использованием инструментария математической статистики, численных методов, математического анализа, методик оптимизации, теории вероятностей, теории графов, а также разнообразных техник обработки данных в цифровой форме.

Регрессионный анализ – обозначает комплекс статистических процедур, предназначенных для исследования взаимодействия одной или многих независимых переменных с зависимой переменной. В контексте этого анализа, независимые переменные часто именуются как регрессоры или предикторы, в то время как зависимые переменные обозначаются как критериальные или регрессанты. Важно подчеркнуть, что классификация переменных как «зависимые» и «независимые» описывает математическую связь между ними, не указывая на причинно-следственные связи.

Экспертная система – определяется как компьютеризированная система, обладающая способностью частично заменить специалиста-эксперта в контексте разрешения проблемной ситуации. Ключевым элементом экспертной системы выступают базы знаний, которые служат моделями поведения экспертов в специализированной области знаний, используя процедуры логического вывода и принятия решений. Другими словами, базы знаний представляют собой агрегацию фактов и правил для логического вывода в выбранной предметной области.

Лабораторная информационная система – это многопользовательское прикладное программное обеспечение, предназначенная для автоматизации процессов лаборатории от регистрации биоматериалов до выдачи результатов.

Веб-сервисы, или веб-службы – представляют собой программные системы, каждая из которых идентифицируется уникальным веб-адресом (URL-адресом) и оснащена стандартизированными интерфейсами. Они способны осуществлять взаимодействие между собой и с внешними приложениями через передачу сообщений, базирующихся на конкретных протоколах (например, SOAP, XML-RPC и т.д.) и соглашениях (например, REST). Веб-служба функционирует как модульный элемент в контексте использования сервис-ориентированной архитектуры приложения.

Сервис-ориентированная архитектура – это метод разработки программного обеспечения, который использует программные компоненты, называемые сервисами, для создания бизнес-приложений. Каждый сервис предоставляет бизнес-возможности, и сервисы также могут взаимодействовать друг с другом на разных платформах и языках.

Референсные значения – представляют собой термин в медицинской практике, используемый при выполнении и интерпретации результатов лабораторных исследований. Эти значения соответствуют среднему показателю определенного лабораторного критерия, полученному на основе экстенсивных исследований здоровой популяции.

Лабораторные исследования – это физико-химические, биохимические и биологические методы исследования, с помощью которых можно анализировать состав и свойства биологических жидкостей и тканей человека, идентифицировать возбудителей заболеваний.

Лабораторный анализатор – это лабораторное оборудование, предназначенное для выявления антител и антигенов.

Протокол TCP/IP – сетевая модель передачи данных, представленных в цифровом виде. Модель описывает способ передачи данных от источника информации к получателю. Наименование TCP/IP произошло от двух ключевых протоколов данного набора – Transmission Control Protocol (TCP) и Internet Protocol (IP), которые были первоначально разработаны и артикулированы в соответствующем стандарте.

Последовательный порт RS232 – стандарт физического уровня для асинхронного интерфейса. Устройство, поддерживающее этот стандарт, широко известно, как последовательный порт персональных компьютеров.

Используется для подключения к компьютерам широкого спектра оборудования, нетребовательного к скорости обмена данными.

Констелляция – это взаимное расположение и взаимодействие различных факторов.

Аликвотирование – это метод измерения концентрации вещества, количество которого меньше, чем порог чувствительности используемой шкалы

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

БД	– База данных
ВОЗ	– Всемирная организация здарвоохранения
ИТ	– Информационные технологии
ИС	– Информационные системы
ИИ	– Искусственный интеллект
МЗ РК	– Министерство здравоохранения Республики Казахстан
РК	– Республика Казахстан
СУР	– Система управления ресурсами
ТОО	– Товарищество с ограниченной ответственностью
ЛИС	– Лабораторная информационная система
МИС	– Медицинская информационная система
ГОБМ	– Гарантийрованный объем бесплатной медицинской помощи
NN (НС)	– Neural Network (Нейронная сеть)
API	– Application Programming Interface
ML	– Machine Learning
PHP	– Hypertext preprocessor
AI	– Artificial intelligence
КДЛ	– Клинико-диагностическая лаборатория
СУБД	– Система управления базами данных
COVID	– Corona Virus
SQL	– Structured Query Language
БЗ	– База знаний
ПО	– Программное обеспечение
СИ	– Международная система единиц измерений
GB	– Gradient Boosting
RF	– Random Forest
R	– Язык программирования R
JSON	– JavaScript Object Notation, текстовый формат обмена данными, основанный на JavaScript
XML	– Extensible Markup Language, расширяемый язык разметки
CDW	– Clinical Data Warehouse, хранилище клинических данных
MYSQL	– Реляционная система управления базами данных, предоставляемая на условиях свободного программного обеспечения
Nginx	– Веб-сервер и почтовый прокси, который работает под управлением операционных систем семейства Linux/Unix и Microsoft
Apache	– Веб-сервер, который работает под управлением операционных систем семейства Linux/Unix и Microsoft

ВВЕДЕНИЕ

Проблема и актуальность темы исследования. Тема диссертационной работы имеет тесную связь с Государственной программой по цифровизации «Цифровой Казахстан» реализованной в период с 2018 по 2022 гг., которая является стратегической комплексной программой, нацеленной на повышение уровня жизни населения страны за счет использования передовых цифровых технологий [1]. В рамках программы уделяется особое внимание на масштабную реализацию электронного паспорта здоровья населения Республики Казахстан, где автоматизация клинико-диагностических лабораторий занимает немаловажную роль. В основном лаборатории оснащаются автоматическими и полуавтоматическими анализаторами, такими как гематологические, биохимические, иммунологические и другие. Имеет место отметить, что автоматические анализаторы могут проводить более 2000 анализов в час, хотя основные показатели зависят от модели и характеристик оборудования. На практике, не все медицинские организации имеют автоматизированные системы и централизованные хранилища данных, следовательно, каждый результат лабораторных тестов остается на локальных базах данных определенного оборудования. Таким образом, результаты анализов предоставляются на стандартных бланках анализатора (зачастую на английском языке) либо переносятся на бумажный носитель, что может способствовать ошибкам, обусловленным человеческим фактором, несоответствиям показателей и утрате данных. В государственной программе «Цифровой Казахстан» уделяют внимание на электронный паспорт здоровья населения, выдвигается необходимость комплексной автоматизации рутинных процессов медицинских лабораторий с внедрением лабораторной информационной системы (ЛИС).

Пациента-ориентированность является главной основой принципа внедрения цифровизации, что отражена в реализации формирования электронного паспорта здоровья. В мобильном приложении mGov можно посмотреть выписку, паспортные данные, к какому врачу прикреплены, какие медицинские услуги в последнее время получали. Вся эта информация достаточно защищена в соответствии с законодательством об информационной безопасности [2].

Электронный паспорт здоровья содержит данные обо всех результатах обследования и лечения пациента в медицинских учреждениях страны, информацию о перенесенных хронических заболеваниях, список лекарств, вызывающих аллергию у пациента, все рекомендации специалистов которых посещал пациент и другое.

Республиканским центром электронного здравоохранения ведутся работы по интеграции 19 информационных систем МЗ РК и Платформы. Вместе с тем, в рамках элементов мобильного здравоохранения казахстанскими IT-разработчиками внедрены около 20 мобильных приложений, доступные для скачивания в Google Play Market [3].

Президент Республики Казахстан Касым-Жомарт Токаев 4 марта 2020 года провел отчетное заседание по реализации Государственной программы «Цифровой Казахстан», и отметил, что цифровизация направлена не для развития одного сектора, а всей экономики государства и преобразовании общества в целом. Президентом подчеркнуто, что пока поставленная перед государственными органами задача по внедрению технологий не достигнута, что для успешной реализации программы необходима согласованность и скоординированность органов, а также было отмечено важность применения технологий «Big Data», алгоритмов «искусственного интеллекта», что способствует к формированию систем базы данных с наивысшим уровнем достоверности, отбрасывать некорректную или устаревшую информацию, систематизировать данные и выдавать объективную целую картину. Глава государства поручил Правительству разработать комплекс регуляторных и стимулирующих мер по расширению проекта Индустрия 4.0.

Также президент отметил, что в Казахстане имеется зависимость от зарубежных разработок и технологий, в связи с чем, было поручено на законодательном уровне поддержать IT-компании казахстанского производства и обеспечить им приоритетность в конкуренции в государственном секторе [4].

Цифровизация медицинских лабораторий является перспективным направлением, являющимся частью электронного здравоохранения. Медицинские лаборатории аккумулируют огромные объемы информации по результатам лабораторных исследований, однако информация настолько разнообразная что несет в себе элементы «лоскутной цифровизации».

Согласно утверждениям аналитиков McKinsey, пять лет назад, даже динамичное развитие медицинских технологий, сфера здравоохранения с позиции прикладного применения «Big Data» существенно отставала от других отраслей, это связано с отсутствием доверия врачей к предоставляемым выводам сделанных с использованием инструментов информационных технологий [5]. Инновации в data science (перевод с англ. – наука о данных) могут не только помочь больным, но и избавить врачей от части рутинной работы. Плюсы демонстрируют чат боты, которые оптимизируют некоторые процессы в клинической практике [6].

В сфере цифровой лабораторной медицины имеются пробелы, но практика показывает, что даже процессы, которые на первый взгляд не поддаются цифровизации, с бурным развитием информационных технологий многие задачи имеют возможность к автоматизации.

Клинико – лабораторная диагностика является одним из ключевых элементов оказания медицинских услуг населению, тем самым покрывает до 90% всех диагностических мероприятий.

Согласно статистике Всемирной организации здравоохранения (ВОЗ), объем необходимых лабораторных исследований для диагностики пациентов удваивается каждый пенталетний цикл. В номенклатуре ВОЗ представлены сотни различных лабораторных тестов. Современные клинико-диагностические лаборатории (КДЛ), обладая обширными аналитическими возможностями,

активно содействуют удовлетворению потребностей клинической медицины в данных, касающихся состояния здоровья, выявления патологий, диагностики болезней и эффективности терапевтических воздействий на пациентов. Прогресс и переосмысление в области медицины, воспринимаемые как элемент публичного здравоохранения, тесно коррелируют с активной интеграцией информационных технологий и современных технических устройств. Это условие обусловлено значимостью не только финальных результатов в системе здравоохранения, но и стоимостной составляющей для их достижения. Без учета объема финансирования и его источников постоянно существует нереализованный потенциал из-за ограниченности ресурсов. Эффективное управление деятельности медицинских работников, основанная на сборе и анализе объективной информации, становится критически важной для гарантирования высококачественных результатов при минимизации финансовых затрат. Экономический анализ, проведенный на основе данных о расходах на лабораторные исследования, подкрепляет обоснование участия лабораторной службы в экономическом аспекте общего медицинского воздействия учреждения, расчет экономической отдачи и определение ее вклада в тарифах медицинского сервиса. Следовательно, результаты лабораторных исследований обеспечивают основу для многообразных перспектив.

В последние годы уделяется больше внимания к клинической лабораторной диагностике, что, в свою очередь, позволяет глубже понимать патогенез различных заболеваний и разрабатывать новые методики их диагностики и лечения. Развитие лабораторной диагностики, подпираемое успехами в области естествознания, электроники и кибернетики, приносит с собой и определенные трудности.

Расхождение знаний клиницистов и лабораторных специалистов. Лабораторные специалисты, углубляясь в разнообразные аспекты своей области, могут становиться менее «проницаемыми» для клиницистов, которые, в свою очередь, могут не поспевать за растущим объемом и специфичностью лабораторной информации.

Проблемы на разных этапах исследования. Возможны сложности, связанные с неправильным назначением тестов, недостатками пред аналитической фазой исследования (например, неправильный забор биоматериала) и др.

Недопонимание между специалистами. Возможен разрыв в понимании и ценности проведения некоторых новых исследований между клиницистами и лабораторными.

Отсутствие клинического мышления у некоторых лабораторных врачей. Это может влиять на качество исследования и интерпретацию результатов.

Развитие и внедрение информационных систем в деятельность лабораторий безусловно поможет улучшить управление бизнес – процессами, повысить качество исследований и решить ряд проблем, связанных с взаимодействием врачей-лаборантов и врачей клиницистов. Особенно это

касается автоматизации наиболее комплексных исследований и обработки их результатов.

Экспертные системы, направленные на помощь клиницистам в интерпретации лабораторных показателей, являются крайне важными, особенно когда речь идет о сложных и многошаговых диагностических алгоритмах. Тем не менее, использование этих систем должно производиться с учетом того, что они служат лишь инструментом поддержки, а не заменой профессионального суждения и клинического опыта врача.

Разработка автоматизированной системы управления процессом лабораторной диагностики может начаться с отдельных сегментов, с последующим масштабированием и адаптацией успешных практик и решений для всей системы. Этот инкрементальный подход позволит более плавно и безболезненно интегрировать нововведения и обеспечит возможность постепенного обучения и адаптации персонала к новой системе управления и работы.

На мировом рынке существует множество лабораторных информационных систем, которые служат для автоматизации операционных, лечебных процессов в лаборатории своей страны. Каждая лабораторная информационная система адаптирована под законодательные нормы своего государства, и зачастую масштабирование в другие страны представляются крайне сложным процессом так, так адаптация системы может занимать в сроки более одного года. На рынке Республики Казахстан представлены как отечественные разработчики лабораторных информационных систем, так и решения ближнего зарубежье, такие как ЛИС «K-Lab» (ТОО «Meditec», Россия-Казахстан), ЛИС «Ариадна» (ООО «Брегис», Россия), ЛИС «Siroca» (ТОО «SIROCA TECHNOLOGY», Кыргызстан), ЛИС «Info Lab» (ТОО «Inform Medical», Казахстан), ЛИС «CS-Soft» (ОсОО «Мобайл Сервис Групп», Кыргызстан), ЛИС «TerraLab» (ООО «Терралаб СОФТ», Украина), ЛИС «Даму» (КМИС) (ТОО «Центр Информационных Технологий «ДАМУ», Казахстан), ЛИС «SmartLab» (ТОО «SmartLab Kazakhstan», Казахстан). Преимуществом Казахстанских разработчиков является оперативность и гибкость, и разработчики находятся на территории страны, что позволяет быстро осуществлять разработки под потребности медицинских лаборатории и адаптироваться под местные законодательные акты.

В рамках работы было выявлено, отсутствие в Казахстане утвержденных норм и стандартов для лабораторных исследований. Существующие приказы и стандарты фокусируются на работе самой лаборатории, а не на результаты лабораторных исследований. Как правило, большинство лабораторий используют референсные значения, указанные в инструкциях для реактивов, поставляемых с оборудованием, и это приводит к ошибке в их работе. Так как для каждого реактива имеется указание, что независимо от стандартов, указанных в инструкциях к реактиву, лаборатория должна определить свои нормы, в зависимости от региона и популяции. Например, можно отметить выдержки из инструкции к реагентам от производителей:

1. F. Hoffmann-La Roche Ltd. – крупный производитель анализаторов и реактивов всегда рекомендует, что каждая лаборатория должна изучить применимость значений к собственной популяции пациентов и, если необходимо, определить собственные референтные диапазоны [7], (Приложение А).

2. Sysmex Corporation – компания-производитель гематологических анализаторов и реактивов отмечает, чтобы каждая лаборатория определяла собственные ожидаемые референтные интервалы на основе популяции пациентов лаборатории, обслуженных в течение одного дня работы. Ожидаемые референтные интервалы могут меняться в силу различий по полу, возрасту, диете, приему жидкостей, географическому положению и пр. [8], (Приложение Б).

В рамках исследовательской работы было проведено анкетирование для выявления потребностей населения и непосредственных участников цифровизации медицинских лабораторий. В качестве инструмента организации анкетирования выбрана платформа Google, разработаны вопросы и ответы на казахском и русском языках для каждой категории респондентов согласно общей мировой практике и методологии опроса [9].

Анкетирование проводилась по следующей категории респондентов:

– пациенты, получающие услуги медицинских лабораторий;

– врачи медицинских учреждений;

– сотрудники медицинских лаборатории;

Опросные формы были размещены на сайте ТОО «SmartLab Kazakhstan» www.lis.kz с целью предоставления быстрого публичного доступа и проведению опроса. Анализ полученных данных с помощью Google data visualization позволил получить информацию о потребностях конечных пользователей услуг медицинских лабораторий и определить потребность в интерпретации результатов лабораторных исследований.

По всем категориям, и с учетом опросов на казахском и русском языках, были опрошены 913 человек, из них:

– пациенты медицинских лабораторий - 672 человека (146 на государственном, 526 на русском);

– врачи медицинских учреждений – 111 врачей (21 на государственном, 90 на русском);

– сотрудники медицинских лабораторий - 130 (16 на государственном, 114 на русском).

После анализа опроса респондентов, было определено, что 68,12% имеют потребность в получении интерпретации результатов лабораторных исследований (Приложение В).

Результаты опроса свидетельствуют об актуальности и потребности для развития информационных систем путем внедрения экспертных модулей, формирования базы знаний для дальнейшего перехода в применении искусственного интеллекта.

Важность разработки информационных систем и внедрения технологий искусственного интеллекта в современной медицинской практике трудно переоценить. Эти технологии играют ключевую роль в усилении эффективности процесса диагностики и лечения, обеспечивая, в итоге, своевременное и точное лечение пациентов. Обеспечивая необходимую поддержку при принятии медицинских решений, созданные информационные системы и алгоритмы ИИ становятся ценным инструментом для медицинских профессионалов, позволяя им оперативно и точно интерпретировать данные, минимизировать риск ошибок и оптимизировать процессы управления пациентами. Такой подход не только ускоряет процесс постановки диагноза, но и способствует более точному и своевременному назначению лечения, предоставляя, таким образом, значительные социальные и экономические выгоды.

Степень изученности и научной разработанности темы исследования.

В рамках диссертационной работы рассматривались публикации авторов по следующим направлениям:

Исскуственный интеллект - Ketan Paranjape, Michiel Schinkel, Richard D. Hammer, Bo Schouten, R.S. Nannan Panday, Paul W.G. Elbers, Mark H.H. Kramer, Prabath Nanayakkara, [10], Mohaimenul Islam, Tahmina Nasrin Poly, Hsuan-Chia Yang, Yu-Chuan (Jack) Li [11].

Машинное обучение и прогнозирование - Dong Jin Park, Min Woo Park, Homin Lee, Young-Jin Kim, Yeongsic Kim, Young Hoon Park [12], Gregor Gunčar, Matjaž Kukar, Mateja Notar, Miran Brvar, Peter Černelč, Manca Notar Marko Notar [13], Nikita Jain, Srishti Jhunthra, Harshit Garg, Vedika Gupta, Senthilkumar Mohan, Ali Ahmadian, Soheil Salahshour, Massimiliano Ferrara [14], Vincent Looten, Liliane Kong Win Chang, Antoine Neuraz, Marie-Anne Landau-Loriot, Benoit Vedies, Jean-Louis Paul, Laëtitia Mauge, Nadia Rivet, Angela Bonifati, Gilles Chatellier, Anita Burgun, Bastien Rance [15], Jessilyn Dunn, Lukasz Kidzinski, Ryan Runge, Daniel Witt, Jennifer L. Hicks, Sophia Miryam Schossler-Fiorenza Rose, Xiao Li1, Amir Bahmani, Scott L. Delp, Trevor Hastie, Michael P. Snyder [16], Pengtao Xie [17], Ryan J. Crowley, Yuan Jin Tan, John P.A. Ioannidis [18].

Одним из ранних работ является публикация авторов John F. Place, Alain Truchaud, Kyoichi Ozawa, Harry Pardue and Paul Schnipelsky в 1995 году, где авторы описывают что экспертные системы, построенные на стандартных вычислительных методах, в значительной степени зависят от экспертов в данной области и инженеров по знаниям, которые их программируют, а нейронные сети предназначены для эмуляции возможности распознавания образов и параллельной обработки данных человеческого мозга и скорее обучаются, чем программируются. Будущее может заключаться в сочетании способности нейронной сети к распознаванию и способности экспертной системы к рационализации [19].

Как видно из работы Yuan Luo и др., применение методов линейной регрессии, Байесовской линейной регрессии, регрессии на основе случайного леса (RFR) и лассо-регрессии выявило следующие результаты, особенно с

использованием анализа ферритная в подтверждении основного предположения, извлекли клинические лабораторные данные из анализов пациентов и применили различные алгоритмы машинного обучения для предсказания результатов анализа на ферритин, используя результаты других анализов. Был проведено сравнение предсказанных и измеренных результатов и рассмотрены отдельные случаи, чтобы оценить клиническую ценность предсказанного ферритина [20].

Также можно отметить работу исследователей V. Jackins и др., где для определения болезней был использован искусственный интеллект с классификацией Naive Bayes и алгоритм случайного леса для классификации многих наборов данных о заболеваниях, таких как диабет, сердечно-сосудистые заболевания и рак. Анализ производительности данных о заболеваниях для обоих алгоритмов рассчитан и сравнен. Результаты моделирования показали эффективность методов классификации на наборе данных, а также характер и сложность используемого набора данных [21].

Процесс автоматизации бурно развивается в каждой отрасли, и ученые, врачи и специалисты других отраслей должны изменить свои методы работы, чтобы справиться с технологическим прогрессом. Авторы Christopher Naugler, Deirdre L. Church в своей работе отмечают что применение искусственного интеллекта к большим наборам клинических данных, созданных благодаря повышению уровня автоматизации, приведет к разработке новых диагностических и прогностических моделей [22].

Все перечисленные работы авторов сфокусированы на уже имеющие данные, а именно набором данных определенных медицинских учреждений и производят пост прогнозирование, применяя алгоритмы машинного обучения для выявления закономерностей на основе имеющихся диагнозов и результатов лабораторных исследований, то есть производят машинное обучение и выявление закономерностей предварительная зная диагноз. В представленной диссертационной работе, предлагается решение на основе комплексной автоматизации каждого участка получения лабораторных результатов, путем интеграции с лабораторным оборудованием и применения базы знаний согласно медицинских данных, а также с инструкциями для каждого теста от завода производителя реагентов. В результате искусственный интеллект использует структурированные и систематизированные данные.

Анализ аналогичных работ показал, что мировое научное сообщество из года в год вносят вклад в развития искусственного интеллекта в лабораторной диагностике. В основном применялись классические методы машинного обучения и сравнения их результатов друг с другом.

Отличие представленной диссертационной работы является в том, что производится прогнозирование патологии без наличия диагноза. Производится ручное обучение модели согласно медицинской документации по лабораторной диагностике, и в режиме реального времени производится выявление патологии. Представленная модель диссертационной работы реализована в рамках разработки интеллектуальной лабораторной информационной системы,

которая производит автоматизацию медицинских лаборатории от регистрации пациента до выдачи результатов исследований.

Объектом диссертационного исследования является лабораторная диагностика в здравоохранении, медицинское лабораторное оборудование с надлежащей медицинской документацией, искусственный интеллект в лаборатории.

Предметом диссертационного исследования является модели и алгоритмы искусственного интеллекта по выявлению патологии в результатах лабораторных исследований.

Целью данной диссертации является разработка и реализация модели искусственного интеллекта, способной автоматически интерпретировать результаты лабораторных исследований в области здравоохранения.

Задачи диссертационного исследования:

1. Изучение существующих методов и технологий в области лабораторной диагностики в здравоохранении: проанализировать современные методы и подходы к лабораторной диагностике, выявив их преимущества и недостатки.
2. Формирование методики комплексной автоматизации в области интерпретации данных лабораторных исследований на основе больших данных (big data), которая позволит автоматически интерпретировать результаты лабораторных исследований для дальнейшего анализа биохимических, иммунологических и гематологических данных.
3. Формирование единой базы референсных значений лабораторных исследований: создать базу данных, содержащую референсные значения лабораторных параметров, учитывая производителей лабораторного оборудования и региональные особенности Республики Казахстан.
4. Разработка и обучение модели искусственного интеллекта для выявления патологии в результатах лабораторных исследований с последующей интеграцией разработанных модулей интерпретации результатов в лабораторную информационную систему «SmartLab».

Методы исследования. В диссертационной работе на каждом этапе реализации были использованы ряд следующих методов: Метод анкетирования – опрос населения о потребности в интерпретации результатов лабораторных исследований; Метод системного анализа – формирования технических, лабораторных данных, алгоритмов реализации по операционной деятельности медицинской лаборатории; Метод декомпозиции – разработка программного кода, модулей и всего функционала информационной системы; Методы машинного обучения – определение закономерностей при изучении большого массива данных; Методы реляционной алгебры – все операции над данными, включая функционала по выявлению патологии в результатах лабораторных исследований.

Научная новизна диссертационного исследования:

1. Методика комплексной автоматизации в области интерпретации данных лабораторных исследований на основе больших данных (big data),

позволяющая автоматически интерпретировать результаты лабораторных исследований для дальнейшего анализа биохимических, иммунологических и гематологических данных.

2. Методика формирования единой базы референсных значений лабораторных исследований с учетом производителей лабораторного оборудования и региональных особенностей Республики Казахстан.

3. Модель искусственного интеллекта для выявления патологии в результатах лабораторных исследований и интеграция разработанных модулей интерпретации результатов в лабораторную информационную систему «SmartLab».

Основные научные положения, выносимые на защиту и обладающие признаками научной новизны:

1. Разработана методика комплексной автоматизации по интерпретации результатов лабораторных исследований по BIG DATA (биохимических, иммунологических и гематологических тестов).

2. Методика формирования единой базы референсных значений лабораторных исследований по производителям лабораторного оборудования и региональных особенностей Республики Казахстан.

3. Разработана математическая модель по выявлению патологии в результатах лабораторных исследований на основе автоматизации с применением методов машинного обучения.

4. Разработаны модули в лабораторной информационной системе «SmartLab» для автоматического выявления патологии в результатах лабораторных исследований на основе комплекса моделей искусственного интеллекта, интегрируемая с медицинскими лабораторными анализаторами.

Практическая ценность результатов исследования. В результате проведенных исследований сформирована Единая база референсных значений лабораторных исследований по производителям лабораторного оборудования и региональных особенностей Республики Казахстан.

Разработаны алгоритмы, модули для автоматического выявления патологии в результатах лабораторных тестов на основе комплексных моделей ИИ, интегрируемая с медицинскими лабораторными устройствами.

Созданная в результате диссертационного исследования лабораторная информационная система внедрена и используется в следующих организациях (Приложение А):

1. ТОО «LabStar Kazakhstan», лаборатория Tumar (Алматы).

2. ТОО «Медси», лаборатория Medси (Караганда).

Получено авторское свидетельство на 1) Лабораторную информационную систему «SmartLab» (Приложение Г); 2) Лабораторную информационную систему «SmartGene»; 3) Свидетельство о товарном знаке «SmartLab» (Приложение Д).

Апробация результатов диссертационного исследования. Основные результаты диссертационной работы докладывались на научных семинарах кафедры «Информационные системы» КАТУ им. С. Сейфуллина, кафедры

«Информационные системы» ЕНУ им. Л.Н. Гумилева, а также школы информационных технологий и интеллектуальных систем ВКТУ им. Серикбаева и на следующих международных научно-практических конференциях:

– международной научно-теоретической конференции «Сейфуллинские чтения – 16: Молодежная наука новой формации – будущее Казахстана (Астана, 2019) [23];

– международной научно-практической конференции «Интеграция науки, образования и производства – основа реализации Плана нации» (Сагиновские чтения №12) (Караганда, 2020) [24];

– международной научно-теоретической конференции «Сейфуллинские чтения – 17: «Современная аграрная наука: цифровая трансформация» «Посвященной 30-летию Независимости Республики Казахстан (Астана, 2021) [25];

– международной научно-практической конференции «Сейфуллинские чтения – 18: «Молодежь и наука – взгляд в будущее» (Астана, 2022) [26];

– международной научной конференции «Математическая логика и компьютерная наука» (Астана, 2022) [27].

Личный вклад автора. Все приведенные в диссертационной работе научные и практические результаты выполнены лично автором. Им непосредственно разработана модель искусственного интеллекта, которая была заложена в разработанную лабораторную информационную систему (Приложение Е).

Публикации по теме диссертационного исследования. По теме диссертационной работы опубликованы 13 научных работ, в том числе 1 статья в журнале, входящем в базу данных Scopus (процентиль по Cite Score равный 34%), 3 статьи в изданиях, рекомендованных уполномоченным органом МНВО РК, 4 статьи в других изданиях, 5 – в трудах международных конференций. Имеется 2 свидетельства о государственной регистрации программы для ЭВМ, 1 свидетельство о товарном знаке «SmartLab».

Структура и объем диссертационной работы. Представленная диссертационная работа состоит из введения, четырех разделов, заключения, списка использованных источников из 93 наименований, изложенных на 122 страницах компьютерного текста, включает 63 рисунков, 18 таблиц и 6 приложений.

Во введении обосновывается релевантность темы, уровень изученности и научной проработанности. Представлены объект, предмет, цель, задачи и методология диссертационного исследования. Выделяются научная новизна, научные позиции, представленные к защите, и практическая значимость результатов исследования. Приводятся данные об апробации и публикациях результатов исследования, а также указывается персональный вклад автора в научные исследования.

В первом разделе приведен обзор и анализ существующих модели и систем искусственного интеллекта по лабораторной диагностике в здравоохранении, в том числе анализ современного состояния лабораторной диагностики в здравоохранении, анализ и перспектива развития искусственного интеллекта в здравоохранении, приведен обзор работ по разработке модели искусственного интеллекта в лабораторной диагностике, а также сделан обзор лабораторным информационным системам представленные на рынке Республики Казахстан.

В разделе два рассматривается моделирование процессов анализа результатов лабораторных исследований. Описывается применение системного анализа к лабораторной диагностике, автоматизация лабораторных операций и методы оптимального применения лабораторных тестов для диагностики и прогноза патологий. Также рассматриваются создание базы знаний для анализа результатов лабораторных исследований и моделирование интерпретации стандартных значений на основе их отклонений от установленных норм.

В третьем разделе приводится результаты разработки модели искусственного интеллекта по лабораторной диагностике в здравоохранении, применения методов машинного обучения для анализа данных по BIG DATA результатов лабораторных исследований, результаты разработки модели искусственного интеллекта по лабораторной диагностике, логическая модель искусственного интеллекта по интерпретации результатов, а также обоснована математическая модель искусственного интеллекта по интерпретации результатов.

Четвертый раздел посвящен практической реализаций и результатам внедрения информационной системы, приведены технологии реализации ЛИС, архитектура базы данных, модули информационной системы и их взаимодействие, а также результаты внедрения в лабораториях на территории Республики Казахстан.

1 АНАЛИЗ СУЩЕСТВУЮЩИХ МОДЕЛИ И СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПО ЛАБОРАТОРНОЙ ДИАГНОСТИКЕ В ЗДРАВООХРАНЕНИИ

1.1 Анализ современного состояния лабораторной диагностики в здравоохранении

Темпы развития лабораторной медицины привели к стремительному расширению диагностических возможностей, и вместо ограниченного перечня рутинных тестов у клинициста появились десятки, а в некоторых лабораториях сотни тестов, которые можно назначить для оценки состояния здоровья пациента. Крупные сетевые лаборатории предоставляют возможность выполнения тестов уже по нескольким тысячам различных параметров. Прогресс в развитии науки создали центральную роль для лабораторной медицины в диагностике многих, если не большинства заболеваний. По оценкам, 70% решений относительно диагноза, лечения и выписки пациента частично основываются на результатах лабораторных исследований [28]. К сожалению, основной причиной врачебных ошибок является неточная диагностика [29]. А также постоянно растущие объемы, высокие затраты на здравоохранение и необходимость повышения точности требуют постоянной оптимизации лабораторных процессов [22, p. 110]. С переходом здравоохранения и лабораторной медицины в эпоху больших данных и искусственного интеллекта (ИИ) способность предоставлять точные, легкодоступные и контекстуализированные данные имеет решающее значение [30].

Медицина перейдет от реактивного к проактивному подходу в ближайшие годы [31]. В то же время патология и лабораторная медицина будут фундаментально изменены двумя новыми технологическими разработками: автоматизацией и искусственным интеллектом. Первые создадут эффективность, снижение затрат и генерирование все более большие и сложные наборы данных; второй будет использовать эти наборы данных для принятия клинических решений, выявлять скрытые подтипы заболеваний, ассоциации и прогностические маркеры, а также генерировать новые проверяемые гипотезы. Эти разработки обеспечат патологию и лабораторную медицину инструментами для того, чтобы возглавить процесс перехода от реактивной к проактивной медицине.

Клинические лаборатории все больше автоматизируются с внедрением современного лабораторного оборудования с конвейерными линиями, включающих передовую робототехнику наряду с программным обеспечением, обладающим возможностями искусственного интеллекта [32]. Автоматизированное оборудование позволяет роботизированным системам быстро и эффективно выполнять процессы, которые ранее выполнялись врачами лаборантами или техниками. Использование современных автоматизированных систем, выполняющих хранение образцов и анализ анализов, может значительно увеличить объем исследований [33]. Например,

настольные системы обработки жидкостей и дозирования образцов широко распространены во многих областях клинической лаборатории, а роботы для обработки жидкостей и пипетирования доступны не только для крупных лабораторий, но и для лабораторий с низкой и средней пропускной способностью. Большие автоматизированные рабочие станции для высокопроизводительной обработки образцов для биохимии, гематологии или молекулярных исследований (например, 2000/тестов/час) также доминируют в диагностическом ландшафте. Однако развитие и выравнивание компьютеризации, подключение ИТ, искусственный интеллект, робототехника и цифровая визуализация уже позволили внедрить крупные автоматизированные решений для основных лабораторий. Это также приведет к революции в области эффективности, в пропускную способность и организацию клинической лаборатории и превратит ее в важнейший центр «больших данных» системы здравоохранения.

Хотя многие технологические достижения в области автоматизации лабораторий обусловлены потребностями фармацевтических и биотехнологических компаний [34], наблюдается переход в сферу академических исследований и диагностики [35].Эта тенденция обусловлена не усиливающейся нехваткой лабораторных специалистов, но и с растущей доступностью современного лабораторного оборудования, которые, уменьшают количество ошибок и повышают точность [36].

Автоматизация лабораторий обеспечивает высоко консолидированные услуги, позволяющие справиться с прогнозируемым будущим ростом числа анализов и требований к рабочему процессу в режиме 24/7 [37]. Согласно отчетам, из многих источников, которые внедрили высокую степень автоматизации, она способна полностью изменить организацию лаборатории и при этом позволяя одному участку эффективно справляться со значительно возросшим объемом исследований при меньшим количеством обученного персонале [36, p. 1256].

Исторически сложилось так, анализ мочи проводится с помощью химии и микроскопией, в то время как культура мочи (т. е. включает идентификацию патогенов и анализ на чувствительность к антибиотикам) проводилась в отделении микробиологии [38]. При такой раздробленной системе результаты этих анализов не были представлены в едином сводном отчете, и информация о результатах анализа мочи не могли быть легко использованы для определения необходимости проведения анализа мочи на культуру [39]. Автоматизированные приборы для анализа мочи, гематологического анализа крови уже далеко шагнули вперед. Современное лабораторное оборудование уменьшило ручной труд, например, как микроскопический анализ мочи (мочевая станция) и расчет Лейко формулы уже автоматизированы [40].

Автоматизация проведения исследования и внедрение ЛИС (Лабораторная информационная система) потенциально ускорила процесс проведения анализов и выдачи результатов пациентам в легкодоступном формате.

Однако внедрение автоматизированных систем является рутиной задачей, и если не будут разработаны оптимальные стратегии внедрения, то идеальная эффективность, может быть, не достигнута [35, p. 261]. Хотя эти сложные роботизированные системы могут сократить количество рабочей силы, необходимой для обработки возросшей пропускной способности тестов, аналитическая эффективность может быть достигнута только в том случае аналитические показатели могут быть достигнуты только в том случае, если обученный персонал эксплуатирует эти приборы в соответствии со стандартными операционными процедурами и для постоянного мониторинга сбоев в работе системы и общего качества анализа [41]. В период внедрения, когда персонал знакомится с системой полной автоматизации, может также возникнуть период «вовлечения», период, когда увеличивается количество лабораторных ошибок.

В последнее годы в связи с развитием в области информационных технологий, искусственного интеллекта и робототехники, производители лабораторного оборудования начали реализацию крупных автоматизированных систем для лаборатории [42]. Эти системы состоят из нескольких компонентов, работающих параллельно с использованием роботизированной системы слежения и программного обеспечения с искусственным интеллектом для управления образцами без участия лабораторных специалистов [43]. Несколько примеров таких высоко интегрированных комплексных платформ автоматизации лабораторий уже используются в клинических лабораториях. Оборудование компании Cobas (Roche) представляют собой высокоавтоматизированные лабораторные платформы [44], которые, в сочетании с одним или несколькими соединительными модулями, могут производить исследования большого объема с помощью роботизированного конвейера, выполняя сортировку, распаковку [45], проверку качества образцов, аликвотирование и повторное укупоривание пробирок для соединительной диагностики [46]. Также компания Abbott представила высоко автоматизированный оборудование Accelerator для выполнения преаналитических задач, который имеет возможность подключения к мульти инструментальной основной лабораторной платформой [47].

1.2 Анализ и перспектива развития искусственного интеллекта в здравоохранении

Внедрение систем искусственного интеллекта (ИИ) в здравоохранении – это один из важнейших современных направлений мирового здравоохранения. Технологии искусственного интеллекта трансформируют глобальную систему здравоохранения, предоставляя возможности для глубокой перестройки процессов медицинской диагностики и разработки новых фармацевтических продуктов, а также в общем смысле способствуя улучшению качества медицинских услуг. Применение искусственного интеллекта для решения медицинских проблем не является новшеством. Фактически, эта область деятельности возникла более трех десятилетий назад [48].

Первоначальная идея искусственного интеллекта приписывается известному британскому математику Алану Тьюрингу. Описывая в 1950 году «Игру в имитацию» - игру в обман между человеком и машиной: «Если бы человек попытался притвориться машиной, он бы явно показал себя очень плохо. Его бы сразу выдали медлительность и неточность в арифметике. Разве машины не могут выполнять то, что можно назвать мышлением, но что сильно отличается от того, что делает человек?» [49]. Действительно, потенциал искусственного интеллекта в здравоохранении заключается в том, что компьютеры могут выявлять сложные, нелинейные ассоциации, строить лучшие прогностические уравнения и определять субвизуальную информацию на изображениях, используя новые подходы к большим массивам данных: в очень реальном смысле «думая» иначе, чем человек. «Машинное обучение» – это родственный термин, определяемый как «наука о том, как заставить компьютеры учиться и действовать, как люди, и улучшать свое обучение со временем в автономном режиме, предоставляя им данные и информацию в виде наблюдений и взаимодействий в реальном мире» [50]. Понятие «большие данные» обычно используется для описания больших наборов данных, обладающих «4 V»: объем, разнообразие, скорость и ценность (например, медицинские изображения, биометрические данные электронных медицинских карт и т.д.) [51]. Имеются ряд решений для работы с большими данными, включая собственные программы крупных технологических компаний, а также решения с открытым исходным кодом. Решения с открытым исходным кодом обычно используют платформу Hadoop и фреймворк программирования MapReduce [51, p. 22-2]. Ряд решений с открытым исходным кодом доступны для тех, кто заинтересован в собственных разработках и реализации [51, p. 22-6]. Было разработано несколько различных вычислительных подходов к искусственному интеллекту. Четыре наиболее часто используемых – это машины опорных векторов (SVM), деревья решений, K-Nearest Neighbor (K-NN) и искусственные нейронные сети (ANN). Алгоритм K-NN является одной из самых простых моделей классификаторов. Этот алгоритм классификации обычно классифицирует элементы данных как принадлежащие к ближайшему классу, который представлен набором измеренных признаков. Деревья решений создают ветвящуюся структуру классификации, где тесты в узлах и под узлами выполняются по одному признаку (или комбинации признаков) [52]. SVM - популярные и мощные инструменты для классификации. Этот подход находит линейный классификатор, известный как гиперплоскость, которая максимально разделяет два класса (рис. 1), где нелинейно разделяемый набор данных был преобразован в разделяемое высоко размерное пространство, разделенное гиперплоскостью.

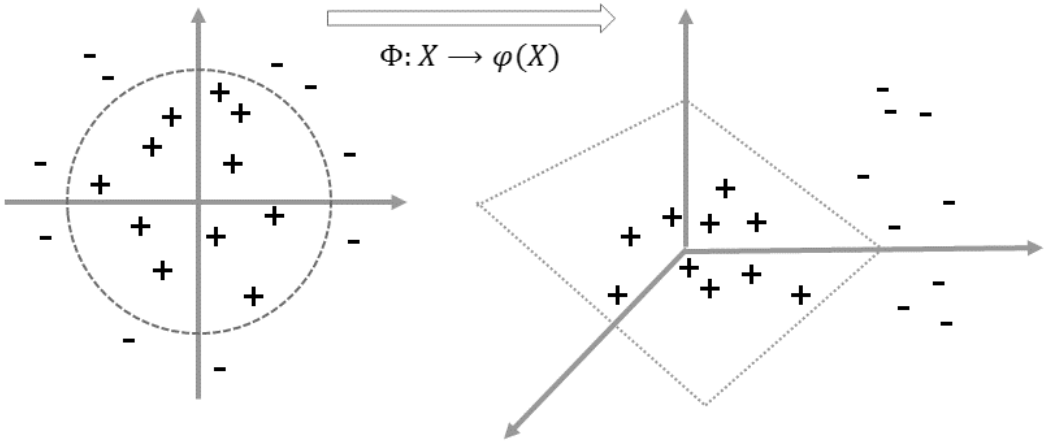


Рисунок 1 – Пример алгоритма машины опорных векторов

ANN использует многоуровневую коллекцию связанных узлов, которые «обучаются» улучшать классификацию, изменяя свои входы и выходы для постоянного улучшения своей работы. Пример нейронной сети показан на рисунке 2.

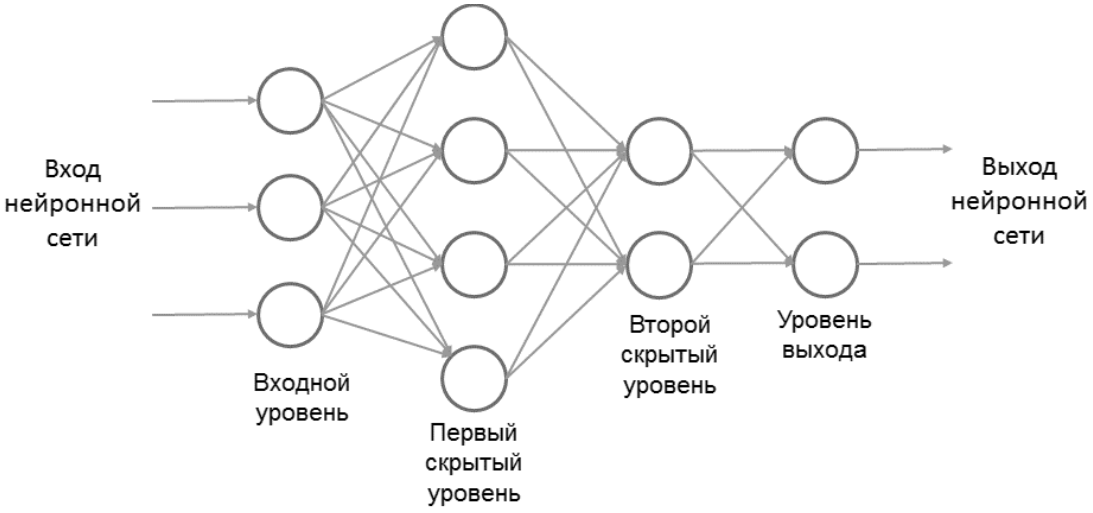


Рисунок 2 - Базовая архитектура искусственной нейронной сети с одним входным уровнем, двумя скрытыми и одним выходным уровнем

Современные нейронные сети могут иметь миллионы единиц и соединений. Такие большие сети в сочетании с усовершенствованными алгоритмами обучения называются глубоким обучением, и в последние годы они стали предметом серьезных исследований. Глубокое обучение (Deep Learning) является одним из наиболее перспективных подходов к использованию искусственного интеллекта в здравоохранении. Другой метод, имеющий особые перспективы в анализе патологических изображений – это конволюционная нейронная сеть (также называемая ConvNet или CNN). Трехмерная архитектура CNN с явным предположением, что входными данными являются изображения, делает ее особенно подходящей для

применения в анализе изображений, например, при анализе микроскопических исследовании.

Некоторые критерии, которые могут быть использованы для классификации инструментов искусственного интеллекта в здравоохранении, показаны таблице 1 [51, p. 22-15].

Таблица 1 – Критерий для классификации инструментов ИИ в здравоохранении

Критерий	Направления	Примеры
По назначению	Для помощи в диагностике	IBM Watson, Ada, Your.MD, Third opinion, DeepMind Health, Face2Gene, Botkin AI
	Для управления организациями здравоохранения	Qventus
	Для поддержания здорового образа жизни/планирования тренировок	Cardiio, Get in Shape, Gymfitty
Путем сбора данных	Сбор данных с помощью датчиков	Cardiio, Gymfitty
	Сбор данных путем опроса	Get In Shape, Ada, Your.MD, Gymfit
По типам пользователей	Для врачей	IBM Watson, Third opinion DeepMind Health, Face2Gene, Botkin AI
	Для пациентов	Ada, Your.MD, Cardiio, Get in Shape, Gymfitty
По видам обрабатываемых данных	Обработка выражений естественного языка	IBM Watson, Ada, Your.MD, Gymfitty
	Обработка изображений	Third opinion, Face2Gene, Botkin AI
	Обработка числовых данных	Get In Shape, Gymfitty

Необходимо отметить некоторые проекты из таблицы 1:

– проект Qventus является ИИ платформой, и был создан в 2012 году в США. Данный проект оптимизирует решения в больницах в режиме реального времени для снижения затрат, повышения качества оказания услуг;

– проект Cardiio – это стартап в области цифрового здравоохранения из США, запущенный в 2012 году. Является платформой для смарта часов мониторинга самочувствия, физической формы и хронических заболеваний;

– существует набор фитнес-приложений для смартфонов, среди которых выделяется Gymfitty, представляющее себя как первого виртуального тренера с искусственным интеллектом. Gymfitty мониторит показатели пользователя в реальном времени и адаптирует тренировочные сессии соответственно. Пользователи получают персонализированные рекомендации, основанные на

различных факторах, включая их упражнения, цели, пульс, обратную связь, анатомию, физическую форму, спортивные умения и историю фитнес-активности (т.е. данные прошлых занятий).

Применение искусственного интеллекта является важным инструментом в развитии точного здравоохранения благодаря способности искусственного интеллекта обрабатывать все более крупные структурированные и неструктурированные данные, а также его способности обнаруживать ценные скрытые ассоциации и классификации заболеваний. Возможно, наиболее важным является то, что подходы к клиническим лабораторным данным могут улучшить краткосрочное и долгосрочное прогнозирование результатов лечения пациентов. В обзоре по этой теме был сделан вывод о том, что растущие объемы данных о состоянии здоровья, а также передовые вычислительные методы означают, что прецизионное общественное здоровье будет становиться все более важной частью здравоохранения в будущем [53].

1.3 Обзор работ по разработке модели искусственного интеллекта в лабораторной диагностике

Большинство лаборатории выдают результаты исследований в виде отдельных числовых или текстовых значений. Однако отдельные результаты исследований, обычно имеют диагностическую ценность. Чтобы использовать результаты тестов в диагностике, врачи должны интегрировать множество отдельно взятых результатов анализов пациента и интерпретировать их в контексте клинических данных и медицинских знаний и опыта. Ручной подход к интерпретации результатов анализов является текущим стандартом в большинстве случаев, однако вычислительные подходы к лабораторной интеграции и анализу данных обладают огромным потенциалом для повышения диагностической ценности. В частности, многие пациенты имеют сотни или тысячи результатов отдельных анализов, зачастую за несколько лет. Как следствие, занятые врачи могут легко пропустить ключевые результаты или важные закономерности и тенденции в наборе лабораторных данных. Более того, важная диагностическая информация иногда может содержаться в закономерностях многочисленных элементов данных, которые могут быть слишком тонкими или сложными для выявления без помощи вычислительных методов [54]. Кроме того, поскольку человеческий мозг сталкивается с большими трудностями при одновременном рассмотрении большого количества точек данных, даже самые опытные врачи могут оказаться неспособными извлечь всю полезную информацию из существующих клинических и лабораторных данных [55]. Поддержка принятия клинических решений представляет собой важный инструмент, позволяющий улучшить интерпретацию результатов анализов и улучшить качество лабораторных данных.

Машинное обучение является основным инструментом анализа данных и выявление закономерностей. Применяя, их исследователи достигают определённых результатов и анонсируют их в научном сообществе.

Обзор научных работ. В своих исследованиях авторы Yuan Luo и др. прогнозировали значения ферритина. Для этого использовались данные результатов исследований, проведенного в Массачусетской больнице общего профиля (MGH), третичном госпитале на 989 коек, в Бостоне, штат Массачусетс, собранные с разрешения институциональным наблюдательным советом больницы. Набор данных включал все амбулаторные результаты ферритина, собранные в течение 3-месячного периода в 2013 году. Каждый результат ферритина был связан с возрастом, полом пациента и результатами «предикторных тестов», выполненных в том же назначений (пациент, дата, время комбинация) [20, p. 778]. Результаты многих лабораторных исследований соответствуют логарифмическому распределению. В регрессионном анализе минимизация среднеквадратичной ошибки эквивалентна оценке максимального правдоподобия, только в предположении, что целевая переменная имеет нормальное распределение. Таким образом, значения ферритина было преобразовано с помощью натурального логарифмического преобразования:

$$y = \ln(1 + x)$$

где y – преобразованное значение ферритина,

x – исходное значение ферритина.

Изучения данных было произведено путем использования методов регрессии, а именно линейная регрессия, Байесовская линейная регрессия, регрессия случайного леса (RFR) и лассо-регрессия (lasso). Для их реализации использовали пакет библиотеку Scikit-learn Python (рисунок 3).

Предполагая, что в некоторых случаях прогнозируемый ферритин может быть более репрезентативным для статуса железа пациента, чем измеренный ферритина, как показано на рисунке 3, было выявлено 26 (1,7%) случаев в тестовом наборе данных, в которых предсказанный ферритин и фактический ферритин сильно расходились, означающее, что фактический и предсказанный ферритин отличались в 10 или более раз. В результате авторы показали, что результаты ферритина предсказуемы с учетом результатов с другими совокупными тестами. Это позволяет предположить, что общие наборы лабораторных результатов могут содержать значительную избыточность информации. В более широком смысле эта работа обеспечивает основу для нового типа поддержки принятия клинических решений.

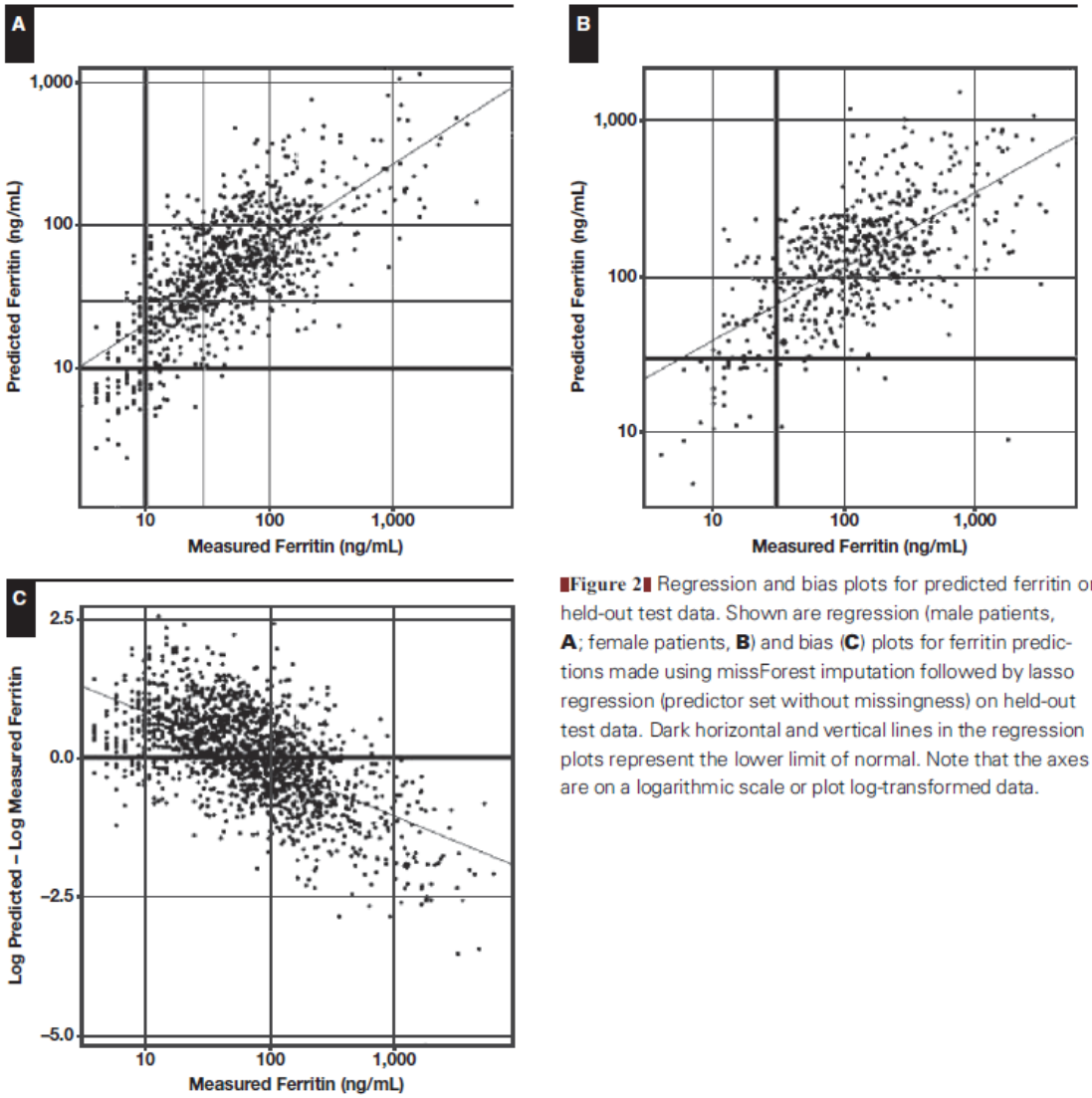


Figure 2 Regression and bias plots for predicted ferritin on held-out test data. Shown are regression (male patients, **A**; female patients, **B**) and bias (**C**) plots for ferritin predictions made using missForest imputation followed by lasso regression (predictor set without missingness) on held-out test data. Dark horizontal and vertical lines in the regression plots represent the lower limit of normal. Note that the axes are on a logarithmic scale or plot log-transformed data.

Рисунок 3 - Результаты проведения регрессионного анализа данных

Примечание – Составлено по источнику [20, p. 777]

В работе следующих авторов V. Jackins и др. рассматривается использование искусственного интеллекта с классификацией Naive Bayes и алгоритмом случайной леса для классификации многих наборов данных о заболеваниях, таких как диабет, сердечно-сосудистые заболевания и рак, где с помощью данных методов проверяли здоровье пациента, на наличие заболеваний [21, p. 5198]. Рассчитан и сравнен анализ производительности данных о заболевании для обоих алгоритмов. Результаты моделирования показали эффективность методов классификации на наборе данных, а также характер и сложность используемого набора данных. Авторы отмечают, что искусственные нейронные сети являются лучшим алгоритмом классификации для прогнозирования медицинской диагностики благодаря своему наилучшему параметру эффективности. А также нейронная сеть состоит из нейронов с тремя слоями, такими как входной слой, скрытый слой и выходной слой для достижения эффективности. Обучающие данные подаются в качестве входного

параметра при поддержке алгоритма обратного распространения. Нейронная сеть с прямой передачей и машиной опорных векторов (SVM) является лучшей методикой для прогнозирования рака [56]. Авторы для решения поставленных задачи использовали гибридный классификатор, объединяющий машину опорных векторов и искусственную нейронную сеть. Типичная нейронная сеть состоит из одного входного слоя, одного или нескольких скрытых слоев и одного выходного слоя (рисунок 4).

Рисунок 4 - Гибридный классификатор

Каждый слой имеет несколько нейронов, и нейроны в одном слое связаны с нейронами в соседнем слое своими собственными весами связи [57]. Нейрон представляет собой узел в сети. Входные признаки подаются на нейроны во входном слое. После применения данного алгоритма авторы выявили, что Байесовской классификации показывает точность 74,46, 82,35 и 63,74% для

данных о диабете, ишемической болезни сердца и раке. Аналогично, классификация с помощью модели Random forest показывает точность 74,03, 83,85 и 92.40%. Результат точности модели Random forest для трех заболеваний является выше, чем точность классификатора Naive Bayes.

Немаловажно отметить работу 1995 года авторов John F. Place и др., где авторы описывают зависимость экспертных системы, построенные на стандартных вычислительных методах от экспертов в данной области и инженеров, которые их программируют и предлагают использование нейронных сетей для эмуляции возможности распознавания образов и параллельной обработки данных [19, p. 373]. Авторами были определены понятия базы знаний, где база знаний разрабатывалась специалистами по знаниям. Важное различие проводится между данными (которые состоят из числовых представлений, уравнений и алгоритмических процедур) и информацией (которая включает в себя факты, концепции, идеи, предыдущий опыт и правила поведения). Специалист по знаниям пополняет системную базу данных из руководств, процедур тестирования и т.д. (факты) и опрашивает экспертов, чтобы узнать, как и когда они принимают решение (правила) и определить, как опыт и интуиция играют роль (эвристика). Способ, с помощью которого специалисты по знаниям интерпретируют знания реального мира в терминах компьютерных структур данных, является важнейшим компонентом системы ИИ и называется «представлением знаний» [58].

Обзорная статья авторов Christopher Naugler и Deirdre L. Church приводят примеры применения искусственного интеллекта в клинической лабораториях и диагностике заболеваний и делают акцент на искусственный интеллект в прогностических моделях [22, p. 98]. Традиционные модели риска обычно основывались на методах регрессии и основаны на предположении, что предикторы действуют одинаково для всех пациентов и одинаково во всем диапазоне их значений. Машинное обучение не страдает от этих предположений и может использоваться для обнаружения новых ассоциаций или моделей стратификации риска [59]. Таким образом, ключевым преимуществом подходов машинного обучения является то, что они не предвзяты заранее существующей гипотезой относительно патофизиологии или прогностических биомаркеров и поэтому дают возможность обнаружить скрытые знания, выявить интересные закономерности или сформулировать новые гипотезы, которые могут быть исследованы в дальнейшем. Популярным направлением применения искусственного интеллекта в прогностических моделях является прогнозирование управления диабетом и его осложнений. Интеграция фенотипических, клинических, экологических, жизненных и лабораторных данных может обеспечить целевые профили риска и рекомендации по лечению.

В своей научной публикации авторы Dong Jin Park и его коллеги представили оптимизированную ансамблевую модель путем объединения модели DNN (глубокой нейронной сети) с двумя моделями ML для прогнозирования заболеваний с использованием результаты лабораторных

тестов (рисунок 5) [12, p. 7567-1]. Для достижения поставленных целей исследователи отобрали 86 тестов из наборов данных на основе подсчета значений, характеристик, связанных с клинической важностью, и отсутствующих значений. Данные были собраны по 5145 случаям, включая 326 686 результатов лабораторных тестов. Было исследовано в общей сложности 39 конкретных заболеваний, основанных на кодах Международной классификации болезней МКБ-10. Эти наборы данных были использованы для построения ML-моделей Light Gradient Boosting Machine (LightGBM) и Extreme Gradient Boosting (XGBoost), а также DNN-модели с использованием TensorFlow. Оптимизированная ансамблевая модель показала F1-score 81% и точность предсказания 92% для пяти наиболее распространенных заболеваний. Глубокое обучение и ML-модели показали различия в предсказательной способности и классификации заболеваний. В рамках анализа данных использована матрица путаницы и проанализирован важность признаков с помощью метода значения SHAP. В результате ML модель достигла высокой эффективности предсказания заболеваний через их классификацию.

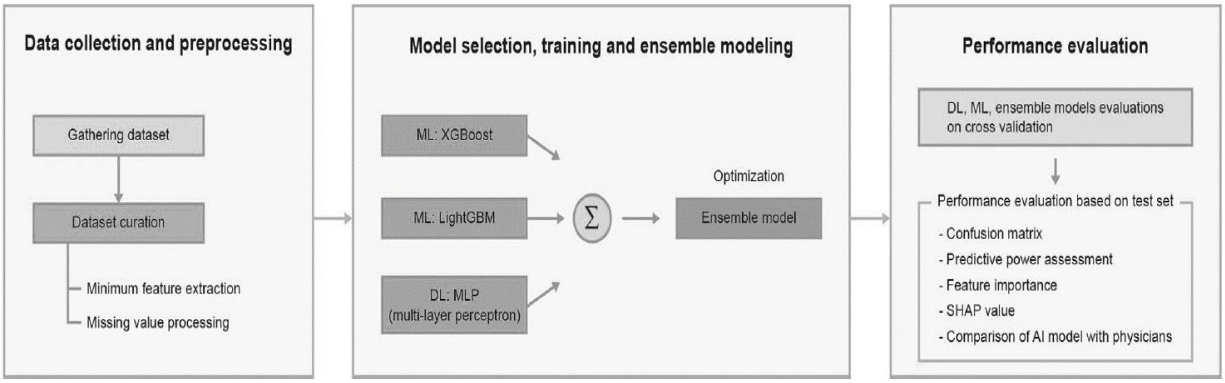


Рисунок 5 – Общая схема модели машинного обучения для диагностического прогнозирования заболеваний на основе лабораторных тестов

Примечание – Составлено по источнику [12, p. 7567-7]

Авторы использовали следующие методы при построении модели:

– сбор и предварительная обработка данных – были проанализированы наборы данных, предоставленные кафедрой медицины, а именно данные пациентов, обратившихся в отделение неотложной помощи и поступивших в Католический университет Кореи St. Vincent's Hospital в Сувоне, в период с 2010 по 2019 год. Все пациенты были не моложе 19 лет. Набор данных состоял из лабораторных анализов, включая результаты анализов крови и мочи, а также окончательный диагноз каждого пациента при выписке. В результате собраны выборочные данные по 5145 случаям, включая 326 686 (73,83%, 326 686/442 470) результатов лабораторных исследований;

– извлечение признаков – извлечение признаков играет важную роль в создании ML-моделей. В результате произвели сравнение DL и ML с

использованием структурированных данных и построили новую ансамблевую модель, объединяющую DL и ML Модель;

– выбор и обучение модели – Выбор DL, так как может аппроксимировать функцию сложной структуры [60]. Технологии Deep Learning (DL), включая сверточные нейронные сети (CNN), стековые автокодировщики (SAE), глубокие сети убеждений (DBN), глубокие нейронные сети (DNN) и рекуррентные нейронные сети (RNN), могут быть применены в различных областях, таких как анализ медицинских изображений, исследование ДНК/РНК и структур белков. В связи с этим исследование в работе проводилось с использованием DNN для структурированных данных. Следующей выбранной моделью было MLP (многослойный перцептрон). MLP состоит как минимум из трех слоев: входной слой, скрытый слой и выходной слой. Кроме входного слоя, каждый слой использует нелинейную функцию активации. Этот метод позволяет анализировать данные, которые не являются линейно разделяемыми. MLP не допускает нулевых значений, поэтому нулевые значения были заменены на медианное значение каждого признака.

В исследовании скрытый слой состоял из двух слоев. Для каждого слоя была использована функция активации Relu (rectified linear unit). К каждому скрытому слою применялась техника отсева, которая является простым методом предотвращения чрезмерной подгонки в нейронных сетях [61].

1. ML: выбор модели ускорения – Ансамблевое моделирование – это метод создания сильных обучаемых путем объединения слабых обучаемых, и в последнее время он широко используется для ML. Boosting и bagging [62]. Bagging создает обобщенную модель путем бутстрапинга (случайной выборки) наборов данных с последующим объединением в различные наборы данных. И boosting, и bagging – это похожие модели обучения, основанные на бутстраппинге. Однако boosting позволяет взвесить и классифицировать данные, которые не были определены на предыдущих этапах. Между этими методами был использован бустинг с помощью алгоритмов XGBoost (рисунок 7) и LightGBM (рисунок 6), которые являются наиболее популярными алгоритмами бустинга.

2. K-кратная перекрестная валидация – в исследовании было разделено 5145 наборов данных в соотношении 8:2 для создания, обучающего и тестового наборов. Было установлено соотношение валидных данных 0,2 для обучающего набора, который оценивался с помощью валидных потерь для оптимизации модели на основе обучающих данных. Количество случаев составило 5145, что является относительно небольшим набором данных. Если размер набора данных невелик, то высокая дисперсия может вызвать проблемы с производительностью при оценке валидности набора данных. Однако если количество проверочных данных увеличивается, количество обучающих данных уменьшается, что приводит к проблеме высокой погрешности. Для решения этих компромиссов, был использован использовали k-кратная перекрестная валидация, чтобы предотвратить потерю данных из обучающего набора.

3. *SHAP (Shapley Adaptive Explanations)* – существует множество способов расчета важности признаков. Среди них значение SHAP имеет хорошую последовательность и точность в вычислении важности признаков. Значения SHAP обеспечивают строгое теоретическое улучшение, устраняя значительные проблемы согласованности (рисунок 6).

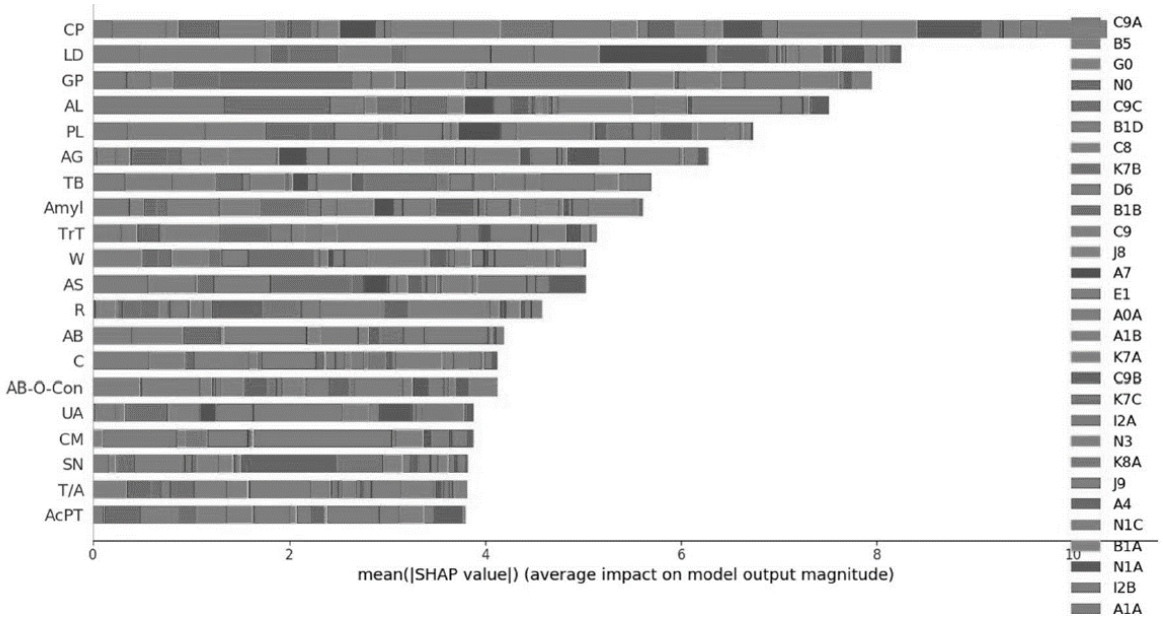


Рисунок 6 – Значение метода SHAP между параметрами и классификации заболеваний с использованием LightGBM

Примечание – Составлено по источнику [12, p. 7567-5]

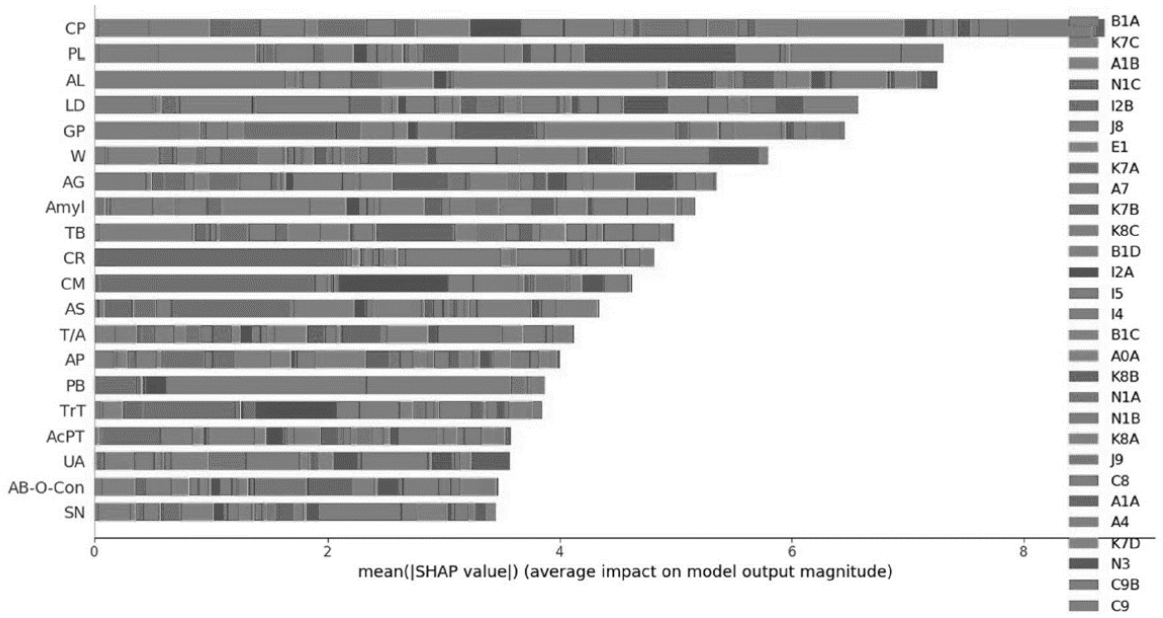


Рисунок 7 – Значение метода SHAP между параметрами и классификации заболеваний с использованием XGBoost

Примечание – Составлено по источнику [12, p. 7567-5]

Искусственный интеллект вызывает большую насторожённость среди врачей, врачей-лаборантов и других медицинских специалистов. В связи с такой ситуацией исследовательская группа во главе с Ketan Paranjape, произвели опрос об использовании ИИ среди участников стратегической консультативной сети компании «Rosche», в которую входят ключевые заинтересованные стороны в области лабораторной медицины [10, p. 823]. В результате из 302 заинтересованных участников ответили на опрос 128 сторон. Большинство участников были практикующими врачами (26%) или руководители лабораторий (22%). ИИ в настоящее время используется в организациях 15,6%, в то время как 66,4% считают, что они могут использовать его в будущем. Большинство не определились с тем, что им нужно для внедрения ИИ в диагностику. Высокие инвестиционные затраты, отсутствие доказанных клинических преимуществ, количество лиц, принимающих решения, и проблемы конфиденциальности были определены как препятствия для внедрения. Исследование показало, что конкретные знания об ИИ в медицинском сообществе недостаточны и что образование в области ИИ крайне необходимо. Одной из стратегий, может быть, внедрение новых инструментов ИИ наряду с существующими [63].

Большие данные всегда имели больше значение в анализе данных. Чем больше данных, тем лучше строятся модели прогнозирования для любых видов алгоритмов. Авторы Vincent LOOTEN и др. [15, с. 104825-1] дали понятие «Хранилище клинических данных (Clinical Data Warehouse)». Исследователи произвели анализ результатов лабораторных исследований за 17 лет. Целью авторов была выявление общих профилей временной эволюции в биологических данных и предложить полу автоматизированный метод для выявления этих закономерностей в хранилище клинических данных (CDW). В качестве хранилища клинических данных выступал Европейская больница Жоржа Помпиду (HEGP), это государственная больница на 700 коек, расположенная в Париже, Франция. HEGP специализируется на онкологии, сердечно-сосудистых заболеваниях и неотложной медицинской помощи. Лаборатория клиники установила лабораторную информационную систему управления лабораторной информацией (LIMS) при открытии больницы в 2000 году [64]. В 2008 году больница разработала CDW, объединяющую все данные, производимые в больнице, включая лабораторные анализы [65]. На момент июля 2017 года Clinical Data Warehouse (CDW) включала в себя свыше 131 миллиона записей лабораторных результатов, относящихся к более чем 445 000 пациентам и связанных с более чем 11 000 различными биологическими показателями. В рамках проведенного исследования данные, интегрированные из разных источников, охватывали период с момента основания больницы в 2000 году и до июля 2017 года. Анализировались анонимные лабораторные данные с указанием дат проведения исследований, причем основное внимание было уделено данным из биохимического и гематологического подразделений лабораторий. Биологические данные были рассмотрены без привязки к типу госпитализации пациентов (например, стационар или амбулаторное лечение). В

качестве критериев отбора использовались биологические показатели, для которых было доступно не менее 10 000 точек данных, представленных в числовом формате. Интервалы времени, в течение которых было зарегистрировано менее 100 значений в течение двух месяцев, были исключены из наборов данных. Были рассмотрены все результаты лабораторных тестов из CDW HEGP, соответствующие критериям отбора, что в итоге образовало 192 «временных ряда», состоящих из пар: временная метка и биологическое значение. Каждый набор данных представлял собой временной ряд ежедневных результатов для конкретного биологического показателя. На рисунке 8 каждая точка символизирует значение для отдельно взятого пациента в конкретный момент времени, в то время как общая диаграмма рассеивания демонстрирует эволюцию распределения отдельного биологического показателя со временем.

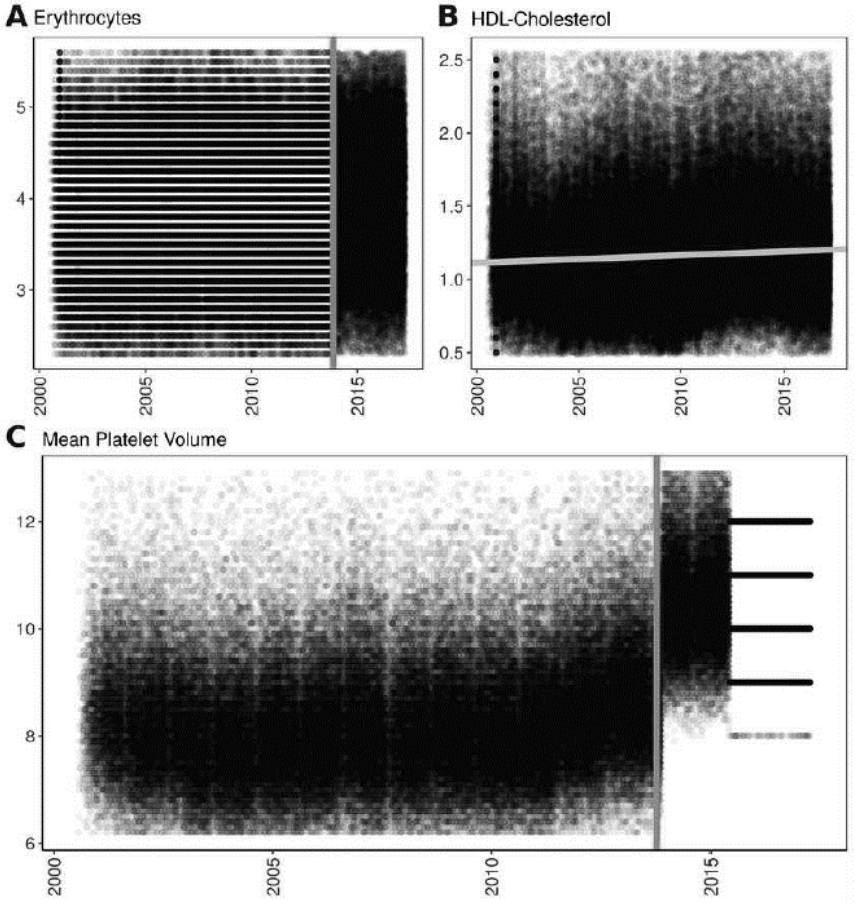


Рисунок 8 – Примеры лабораторных параметров, на которые влияет временная эволюция

Примечание – Составлено по источнику [15, p. 104825-6]

В итоге продольное качество 192 биологических параметров было оценено. Количество значений биологических тестов, покрываемых этими 192 параметрами, достигло отметки в 87 814 136. Глобальное профилирование данных представлено в таблице 2.

Таблица 2 – Оценка продольного качества

Категория	Количество биологических параметров	Наименование параметра (теста)
Дискретизация	32 (16.7%)	Эритроциты
Точки разрыва	30 (15.6%)	Средний уровень тромбоцитов
Тенденции	79 (41.1%)	Холестерин ЛПВП

Следующий автор Pengtao Xie из Carnegie Mellon University опубликовал исследование, где был использован преимущества глубоких генеративных моделей для работы со сложными лабораторными тестами [17]. В частности, автор предлагает сквозную архитектуру, которая включает в себя глубокие генеративные вариативные рекуррентные нейронные сети (VRNN) для обучения надежным и обобщенным характеристикам и дискриминирующая нейронная сеть (НС) для обучения принятию решения о диагнозе, причем эти две модели обучаются совместно. Эксперименты проводились на наборе данных, включающем 46 252 пациента, и 50 наиболее частых тестов используются для предсказания 50 распространенных диагнозов. Результаты показали, что модель, VRNN+NN, значительно ($p < 0:001$) превосходит другие базовые модели. Более того, полученные в результате совместного обучения, более информативны, чем представления, полученные с помощью чистых генеративных моделей.

Пандемия Covid19 стала предметом беспокойства для всего мира. Основной причиной распространения коронавируса являются SARS-CoV и SARS-CoV-2, которые являются входящие в семейство коронавирусов. Таким образом, прогнозирование пациентов, страдающих от таких пандемических заболеваний, поможет сформулировать разницу в неточной и невыполнимой временной продолжительности. В работе авторов Nikita Jain и др. предлагаются различные стратегии ансамблевого обучения, которые оказались полезными при составлении прогнозов [14, p. 103813-1]. Прогнозы делаются с использованием различных моделей машинного обучения. Для прогнозирования и анализе набора данных использовались модели машинного обучения, такие как SVM, Naïve Bayes, K-nearest соседи, AdaBoost, градиентное усиление, XGBoost, случайный лес, ансамбли и нейронные сети. Наиболее точный результат был получен с помощью предложенного алгоритма с показателем для прогнозирования вируса SARS-CoV-2.

В следующей обзорной статье авторов Norah Alballa, Isra Al-Turaiki производится анализ публикации по тематикам связанные с COVID-19, SARS-CoV-2, где были применены алгоритмы машинного обучения [66]. Поиск публикации производился по PubMed, Scopus, IEEE Xplore, and Google Scholar и на английском языке за период с января 2020 года по январь 2021 года. В результате было идентифицировано 52 исследований с 76 моделями машинного обучения.

Таблица 3 – Статистические данные по Covid-19

Возрастная группа	Общее коли чество	В том числе			Из них:							
		коли чество отриц.	*коли чество полож.*	*% соотношений полож-х результатов к общему количеству*	*коли чество Ж*	*Коли чество М.*	*количество отрицательных*		*количество полож.*		*% соотношений полож-х результатов к общему количеству у женщин и мужчин*	
							ж.	*м.*	*ж.*	*м.*	*ж.*	*м.*
От 0 до 5 лет	1 929	1 725	204	10,6	797	1 132	712	1 013	85	119	4,4	6,2
От 6 до 10 лет	772	683	89	11,5	335	437	299	384	36	53	4,7	6,9
От 11 до 15 лет	1 193	1 104	89	7,5	500	693	456	648	44	45	3,7	3,8
От 16 до 20 лет	3 159	2 986	173	5,5	1 928	1 231	1 854	1 132	74	99	2,3	3,1
От 21 до 30 лет	8 968	7 728	1 240	13,8	4 383	4 585	3 831	3 897	552	688	6,2	7,7
От 31 до 40 лет	8 304	6 922	1 382	16,6	3 952	4 352	3 277	3 645	675	707	8,1	8,5
От 41 до 50 лет	6 519	5 301	1 218	18,7	3 553	2 966	2 816	2 485	737	481	11,3	7,4
От 51 до 60 лет	5 735	4 214	1 521	26,5	3 307	2 428	2 363	1 851	944	577	16,5	10,1
От 61 до 70 лет	4 385	3 362	1 023	23,3	2 490	1 895	1 882	1 480	608	415	13,9	9,5
От 71 до 80 лет	1 744	1 312	432	24,8	1 055	689	787	525	268	164	15,4	9,4
Старше 80 лет	731	531	200	27,4	486	245	360	171	126	74	17,2	10,1
Всего:	43 439	35 868	7 571	17,4	22 786	20 653	18 637	17 231	4 149	3 422	9,6	7,9

В период пандемии нами тоже были проведены работы по автоматизации ПЦР лаборатории, и были опубликованы где производился сбор и анализ данных для выявления эпид. ситуации в регионе [67], и практическое применение QR кодирования для защиты от подделок ПЦР результатов [68]. В публикациях были представлены модели централизованного сбора результатов ПЦР на Covid-19 (рисунок 9) и обезличенные статистические данные по заболеваемости (таблица 3).

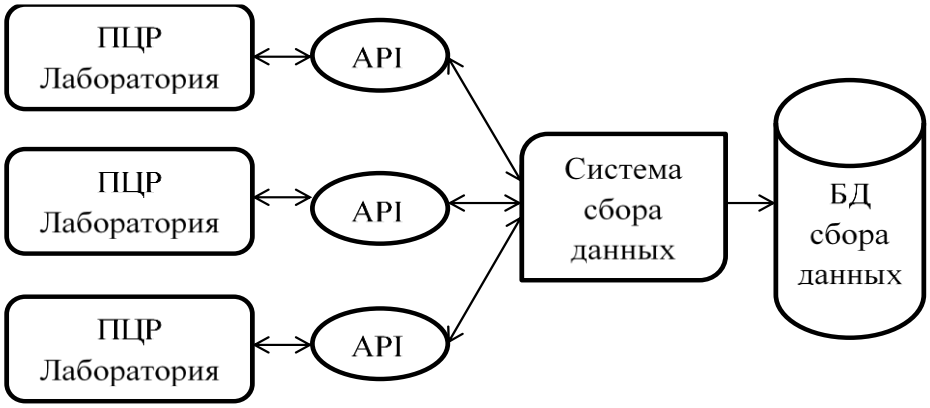


Рисунок 9 – Модель централизованного сбора данных по результатам анализов на Covid-19

В таблице 3 отражена информация об общем количестве результатов по возрастным категориям с разделением чисел отрицательных и положительных определений по половому признаку. Также отражено процентное соотношение положительных результатов к общему количеству, прошедших исследование в разрезе мужского и женского пола.

В данном разделе были рассмотрены работы по искусственному интеллекту, машинному обучению, а также обзорных работ о преимуществе искусственного интеллекта в лабораторной диагностике.

1.4 Лабораторные информационные системы (ЛИС) представленные на рынке Республики Казахстан

Рынок IT-решений в секторе здравоохранении Республики Казахстан является достаточно новым, он активен и темпы появления новых IT-компаний и стартапов высоки. Пандемия Covid-19 дал огромный толчок для развития и расширения IT-рынка Казахстана.

На рынке Республики Казахстан представлены ряд лабораторных информационных систем, но только часть из них являются отечественными продуктами. На рисунке 10 представлены перечень компаний и их продуктов предназначенных для автоматизации медицинских лабораторий различного профиля, такие как клинико-диагностические, бактериологические, генетические и др.

Программный продукт компании
ЛИС «K-Lab», Россия-Казахстан, www.k-lab.kz
ЛИС «Ариадна», Россия, www.bregis.ru
ЛИС «Siroca», Кыргызстан, www.siroca.com
ЛИС «Info Lab», Казахстан, www.medico.kz
«ЛИС CS-Soft», Кыргызстан, www.333.kg
ЛИС «TerraLab», Украина, www.limsterralab.com
ЛИС «Даму» (КМИС), Казахстан, www.damumed.kz
ЛИС «АЛИСА», Россия, www.galen.ru
ЛИС «SmartLab», Казахстан, www.lis.kz

Рисунок 10 – ЛИС на рынке Республики Казахстан

На сегодняшний день государство Республики Казахстан уделяет особое внимание на развития отечественных разработок в области информационных систем, в связи с этим, главным преимуществом является то, что ТОО «SmartLab Kazakhstan» является компанией с 100% казахстанским содержанием. В таблице 4 представлен сравнительный анализ функционала представленных лабораторных информационных систем. Критериями оценки служат технические и функциональные возможности каждой из систем. В качестве оценки были определены следующие критерии:

– кроссплатформенная архитектура серверной части системы; применение SOA технологии;

– универсальный обучаемый интегратор с лабораторными оборудованиями любого производителя;

– база знаний по клинической диагностике бактериологии, микробиологии, ПЦР диагностике, генетике;

– пополняемая база знаний референс значений производителей лабораторного оборудования;

– автоматическая идентификация отклонений результатов лабораторных исследований;

– история результатов пациента по качественным и количественным исследованиям;

– интерпретация причин отклонения результатов лабораторных исследований;

– учетная система генерации штрих кодов для регистрации проб; соответствие функционала ISO-15189; электронные журналы внутри лабораторного контроля показателей и санитарных мероприятий;

– интеграция с платежными системами фискального учета;

– интеграция с мессенджерами whatsapp, telegram;

– использование технологии оцифровки изображений с микроскопов;

– отечественный производитель программного обеспечения.

По каждому критерию оценки производилась маркировка «+» и «-», и в результате производилась сумма плюсов и минусов.

Таблица 4 – Сравнительная таблица функционала лабораторных информационных систем

Критерии оценки	Перечень лабораторных информационных систем								
	ЛИС «K-Lab» Россия-Казахстан	*ЛИС «Ариадна»* Россия	*ЛИС «Siroca»* Кыргызстан	*ЛИС «Info Lab»* Казахстан	*«ЛИС CS-Soft»* Кыргызстан	*ЛИС «TerraLab»* Украина	*ЛИС «Даму» (КМИС)* Казахстан	*ЛИС «АЛИСА»* Россия	*ЛИС «SmartLab»* Казахстан
Кроссплатформенная архитектура серверной части системы	-	-	-	-	-	-	-	-	+
Применение SOA технологии	+	-	-	-	-	-	+	-	+
Универсальный обучаемый интегратор с лабораторными оборудованиями любого производителя	-	-	-	-	-	-	-	-	+
База знаний по клинической диагностике, бактериологии, микробиологии, ПЦР диагностике, генетике	+	-	-	-	-	-	-	-	+
Пополняемая база знаний референс значений производителей лабораторного оборудования	+	-	+	-	-	-	+	-	+
Автоматическая идентификация отклонений результатов лабораторных исследований	+	-	-	+	-	+	+	-	+
История результатов пациента по качественным и количественным исследованиям	-	-	-	-	-	-	-	-	+
Интерпретация причин отклонения результатов лабораторных исследований	-	-	-	-	-	-	-	-	+
Учетная система генерации штрих кодов для регистрации проб	-	-	-	-	-	-	-	-	+
Соответствие функционала ISO-15189	+	-	-	-	-	-	-	-	+
Электронные журналы внутри лабораторного контроля показателей и санитарных мероприятий	+	-	-	-	-	-	-	-	+
Интеграция с платежными системами фискального учета	+	-	-	-	-	-	+	-	+
Интеграция с мессенджерами whatsapp, telegram	+	-	-	-	-	-	+	-	+
Использование технологии оцифровки изображений с микроскопов	-	-	-	-	-	-	-	-	+
Отечественный производитель программного обеспечения	-	-	-	+	-	-	+	-	+

В результате подсчета критериев оценки, можно увидеть преимущество ЛИС SmartLab перед другими поставщиками лабораторных информационных систем представленных на рисунке 10. Согласно критериям оценки в сравнительной таблице 4 можно увидеть, что функционал и технологии решения ЛИС «SmartLab» превосходит другие системы.

Выводы по разделу

В рамках диссертационной работы был проведен обзор опыта ученых мира, где исследователи, используя данные разного характера применяют машинное обучение для дальнейшего его использования искусственным интеллектом. Также осуществляли опрос готовности медицинских лаборатории к использованию искусственного интеллекта для принятия решений. Основные модели, описанные в разделе, основываются на классических методах машинного обучения, а именно Дерево принятия решений, наивная байесовская классификация, Метод наименьших квадратов, логистическая регрессия, метод опорных векторов (SVM), Метод ансамблей, Метод главных компонент (PCA), Сингулярное разложение, Анализ независимых компонент (ICA). Исследователи, применяя машинное обучение обучали модели, анализировали выявляли закономерности, и также применяли глубокое обучения использую несколько методов машинного обучения. Наборы данных для моделирования процессов, анализа брались из свободных источников, из архива больниц.

Была проведены работы по анализу функциональных возможностей лабораторных информационных систем на рынке Казахстана и составлена сравнительная таблица.

Представлены алгоритмы и решения, которые были апробированы в период пандемии Covid-19.

2 МОДЕЛИРОВАНИЕ ПРОЦЕССОВ ИНТЕРПРЕТАЦИИ РЕЗУЛЬТАТОВ ЛАБОРАТОРНЫХ ИССЛЕДОВАНИЙ

2.1 Реализация системно-ориентированного подхода в процессах лабораторной диагностической деятельности

Лабораторная диагностика, когда-то представлявшая собой узконаправленный комплект тестов, теперь преобразовалась в сложную научную дисциплину. Специалисты в этой сфере сегодня сталкиваются с все более многообразным спектром профессиональных задач. К их обязанностям теперь относятся: управленческие вопросы, реализация научно-исследовательских программ, консультирование врачей-клиницистов, анализ новых методов в оборудовании, реагентах и процедурах, а также вопросы образования и подготовки специалистов и студентов по медицине.

Вместе с трансформацией профессиональных обязанностей медицинских лаборантов наблюдается непрерывное увеличение объема знаний и технологических инноваций в этой сфере.

Методы системного анализа [69] можно использовать для оптимизации процесса клинической лабораторной диагностики, предоставляя высоко организованный, концептуальный взгляд на лабораторную диагностику в общем, осматривая её с «верху вниз». Такой метод подчёркивает глобальную структуру или инфраструктуру анализируемого диагностического процесса.

Предложенная методика основывается на принципе, что разнообразные элементы диагностического процесса можно классифицировать на две основные группы: функциональные процессы и операционные методики.

Функциональные процессы в медицинских лабораториях служат стабильным основанием или «каркасом», предоставляя устойчивую инфраструктуру для обеспечения последовательности и стандартизации. Эти процессы, как правило, меняются медленно и служат для поддержания установленных стандартов и норм в области управления и регулирования. С другой стороны, оперативные подходы обладают гибкостью и адаптивностью, обеспечивая способность системы подстраиваться под новые технологии, изменения в клинической практике, потребности пациентов и другие переменные факторы. Несмотря на то, что функциональные процессы предоставляют некий неизменный «скелет», управление и оптимизация оперативных подходов остаются критически важными для поддержания актуальности и качества услуг, предоставляемых клиническими лабораториями.

Функциональные процессы в контексте лабораторной диагностики могут быть организованы в иерархические уровни. На самом начальном уровне важно определить минимальное количество функций, которое будет в совокупности охватывать все аспекты профессиональных обязанностей лаборантов в процессе лабораторной диагностики. Группировка должна быть свободной от избыточности, однако, также должна обеспечивать полноту, чтобы любая функция, не представленная явно, могла быть интерпретирована как

подкатегория одной из перечисленных функций. Предполагается, что пять основных функций: обслуживание, исследование, разработка, управление и образование, могут удовлетворять этим критериям, охватывая все ключевые аспекты деятельности в этой сфере.

Исполнение функции сервиса напрямую связано с реализацией четырех других функций. Следовательно, чтобы обеспечить эффективное обслуживание в сфере лабораторной диагностики, необходимо тщательно управлять (управление) всеми процессами, активно проводить исследования (исследование) для постоянного развития методов и техник, продолжать разрабатывать (разработка) новые технологии и методики, а также обеспечивать постоянное образование и обучение (образование) специалистов. Таким образом, сервис в лабораторной диагностике оказывается качественным и продуктивным, когда остальные четыре функции активно и эффективно выполняются. Эти четыре функции взаимосвязаны и взаимозависимы, и успешное выполнение каждой из них способствует улучшению качества сервиса, предоставляемого пациентам и медицинским специалистам (рисунок 11).

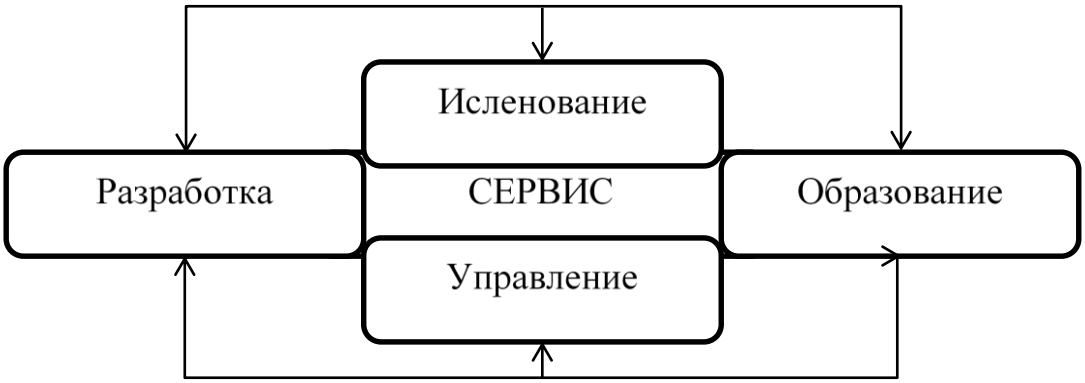


Рисунок 11 – Модель сервиса лабораторной диагностики

Не всегда возможно провести четкую линию разделения между вышеописанными функциями, но роль каждой из них в организации процесса лабораторной диагностики является достаточно ясной и определенной. Содержание второго уровня, цель которого – описать функции, выделенные на первом уровне, можно иллюстрировать, рассмотрев функцию сервиса в качестве примера. Аналогично любой другой основной функции, функция сервиса может быть разбита на подуровни с иерархической перспективой. Это может включать в себя различные аспекты, такие как прием и обработка образцов, выполнение тестов, интерпретация и сообщение результатов, а также консультирование клинических коллег. Каждый из этих подуровней может быть дальше разбит на более детализированные этапы или аспекты, подчеркивающие многомерность и сложность процессов, лежащих в основе функции сервиса в контексте клинической лабораторной диагностики. Это деление может продолжаться на более низких уровнях иерархии, позволяя

анализировать и оптимизировать каждый аспект деятельности для обеспечения наиболее эффективного и высококачественного сервиса (рисунок 12).

Рисунок 12 – Структура функции сервиса

Постановка диагноза или прогнозирование состояния здоровья пациента представляет собой итеративный процесс, в котором участвуют как врачи-клиницисты, так и врачи-лаборанты.

Процесс начинается с оценки состояния здоровья пациента, что включает в себя сбор информации, постановку предварительного диагноза. Последующим производится сбор дополнительных данных для подтверждения диагноза и планирование исследований. После проведения лабораторных исследований информация сопоставляется с первичным диагнозом, и при необходимости корректируется или уточняется диагноз. Этот циклический процесс может повторяться несколько раз для окончательной постановки диагноза и мониторинга состояния пациента на протяжении лечения.

На третьем уровне функциональных процессов рассматривается количественное определение, и одной из ключевых функций здесь является функция исследования. Эта функция играет важную роль в лабораторном диагностическом процессе и включает два основных аспекта:

– анализ биологических материалов пациента;

– осмысление результатов тестов, их анализ и извлечение значимой информации.

Четвертый уровень функциональных процессов может быть наглядно проиллюстрирован на примере функции обработки проб. Эта функция включает в себя ряд важных операций, таких как транспортировка, физическое и химическое воздействие, смешивание и разделение биологических материалов. Для начала материалы пациентов должны быть перевезены в специальное помещение, где будут проводиться последующие этапы обработки.

Оперативные подходы – это разнообразные методы и способы реализации функциональных процессов. Давайте рассмотрим конкретные примеры системного подхода к разработке оперативных методов, применительно к функции взятия и обработки проб. Для этого представим несколько оперативных методов, связанных с этой функцией. Перед тем как взять пробу, важно определить, в какой части системы организма она будет взята.

При идентификации системы забора проб возможны два ключевых варианта выбора области забора. Первый вариант – это выбор области, которая представляет основное свойство всей системы, то есть какого-то параметра или состояния, характерного для всего организма или системы. Второй вариант – это выбор области, которая представляет локальное свойство специфической части системы, например, определенного органа или ткани. При этом важно учесть несколько ключевых моментов. Во-первых, необходимо определить место взятия проб – это может быть внутри самой системы (внутренний забор проб) или удаленно от нее (внешний забор проб). Во-вторых, нужно решить, как часто будут браться пробы – это может быть непрерывный процесс, когда пробы берутся постоянно, или через определенные временные интервалы. И, в-третьих, необходимо определить тип процесса взятия проб - будет ли он инвазивным (требующим внедрения в тело) или не инвазивным, и деструктивным (разрушительным) или не деструктивным. Эти аспекты могут существенно влиять на характер и качество получаемых проб и, соответственно, на их информативность для последующего анализа.

Системно-ориентированный подход к лабораторной диагностике предоставляет унифицированную и структурированную схему, в которой различные компоненты и этапы процесса лабораторной диагностики рассматриваются как разнообразные технологии и функции.

2.2 Имплементация процессов автоматизации в лабораторной практике и разработка методики рационализации использования диагностических тестов для уточнения и прогноза патологических состояний

Современная клинико-диагностическая лаборатория (КДЛ) представляет собой сложную и высокотехнологичную систему, в которой выполняются

сотни различных технологических процессов. В данном контексте автоматизация лабораторных операций становится неотъемлемой частью создания современных систем управления КДЛ. Основной целью такой автоматизации является обеспечение высокого качества результатов исследований, сокращение затрат и обеспечение безопасности пациентов. Внедрение автоматизированных инструментов является единственным практически осуществимым способом достижения этих целей. Для эффективного функционирования современной лаборатории необходима поддержка информационных систем. Большинство клинико-диагностических лабораторий уже используют какие-либо формы автоматизированных информационных систем. Тем не менее, не все из них внедрили единую лабораторную информационную систему, которая объединяла бы все аспекты работы различных подразделений лаборатории в единую систему. Иными словами, автоматизация и использование единой лабораторной информационной системы стали необходимыми для современных клинико-диагностических лабораторий, чтобы обеспечить высокий уровень качества, эффективность и безопасность в проведении лабораторных исследований.

Автоматизированные лабораторные системы действительно существенно улучшили точность и качество проводимых анализов. Однако они также порождали огромные объемы данных, что представляло новые вызовы для лабораторий в обработке и интерпретации этой информации. В некоторых аспектах автоматизация лабораторных тестов усугубила сложности, связанные с управлением данными в лабораториях. В ответ на эти вызовы началось активное развитие лабораторных информационных систем (ЛИС), и компьютеры стали использоваться для сбора данных от лабораторных анализаторов, их обработки, хранения и выдачи результатов исследований. Клиническая лаборатория оказалась основным местом для внедрения компьютерных технологий, поскольку компьютеры могут обрабатывать большие объемы данных быстро и точно. Это позволило лабораториям справляться с растущими потоками информации, повышая эффективность и минимизирую возможные ошибки в анализах и интерпретации результатов.

Лабораторные информационные системы представляют собой комплексное решение, включающее в себя программное обеспечение, аппаратное оборудование, базы данных и знаний, разработанное для автоматизации и оптимизации процессов, которые происходят в лабораториях. Они созданы с целью удовлетворения потребностей специалистов и сотрудников лаборатории в доступе к систематизированной информации, касающейся различных аспектов их работы. Эта информация предоставляется в целях облегчения принятия решений, которые способствуют улучшению функционирования лаборатории и повышению качества результатов лабораторных анализов.

Современные лабораторные системы представляют собой компонентную систему (Модульную), где каждый блок выполняет свои функциональные задачи (рисунок 13).

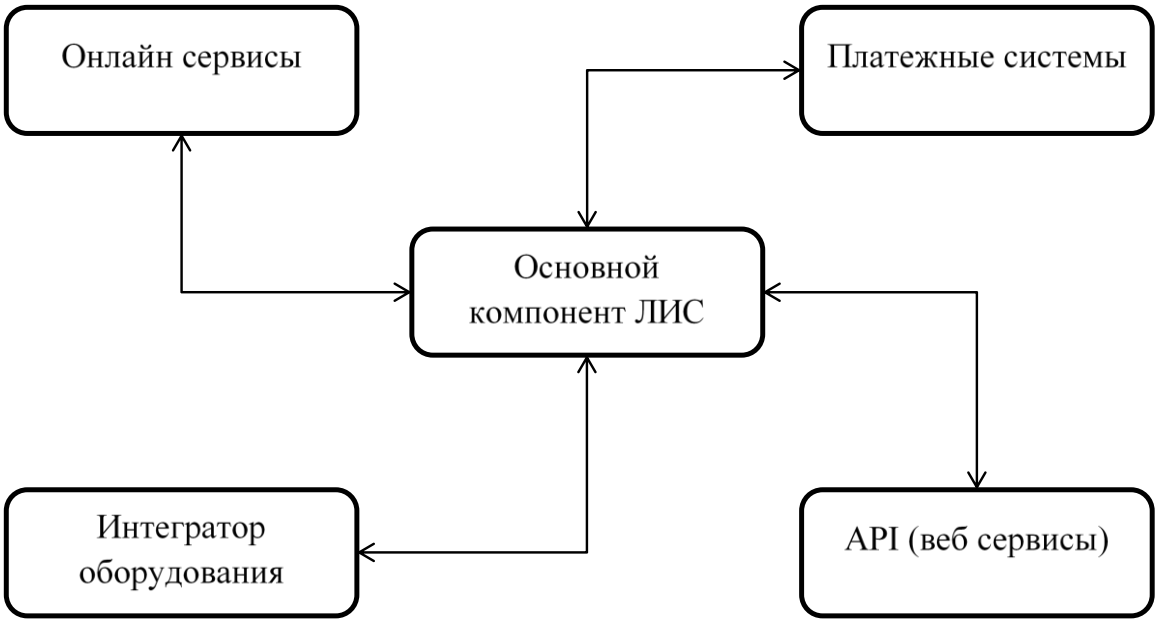


Рисунок 13 – Структура современных лабораторных информационных систем

Рассмотрим каждый компонент структуры, приведенный на рисунок 12, где:

– основной компонент ЛИС – является ядром, которая содержит все основные функции по автоматизации деятельности лаборатории и базы знания. Регистрация пациентов, назначения исследований, справочники, базы знаний по исследованиям и тестам, нормативные величины, рабочие листы врачей-лаборантов, выдача результатов и другие необходимые инструменты для работы лаборатории;

– онлайн сервисы – данный блок представляет собой онлайн личный кабинет для пациентов и для получателей услуг лабораторий. То есть через данные сервисы пациенты могут получить результаты лабораторных исследований онлайн, с использованием логина и пароля. А также потребители услуг могут осуществлять регистрацию проб, получения результатов. Данный функционал разгружает сотрудников лабораторий от лишних рутинных процедур;

– интегратор оборудований – данный блок служит для подключения лабораторного оборудования к системе. Это позволит в режиме реального времени взаимодействовать с анализаторами, получать сделанные результаты, давать задание, тем самым исключая ручную работу специалистов лаборатории;

– платежные системы – данный блок предоставляет возможность интегрироваться с платежными и учетными системами как онлайн кассы, процессинговые центры обработки платежей, банковские терминалы;

– API (веб сервисы) – служат для взаимодействия со всеми внешними системами, такими медицинские информационные системы, государственные

порталы. Например, в период пандемии все результаты ПЦР тестирования направлялись портал Национального центра экспертизы. А также КМИС ДАМУ, где с помощью веб-сервисов производится электронный оборот лабораторных исследований, в результате которого все лабораторные анализы отражаются в электронной карте пациента.

Внедрение более широкого спектра лабораторных исследований и использование констелляционного подхода значительно увеличило объем лабораторной данных. Что привело к улучшению диагностики и более точному определению скрытых форм патологии. Однако с накоплением большого объема данных возникают вопросы о том, насколько эффективно эта информация используется в процессе диагностики и насколько целесообразно расширять список лабораторных тестов для каждого пациента. Иногда избыточность информации может привести к затруднениям для врачей и отрицательно повлиять на точность постановки диагноза.

Для решения такого рода проблем, предлагается подход дифференциально-диагностических программ, которые имеют специальные логические алгоритмы и модели. Последовательное выполнение этих алгоритмов позволяет наиболее эффективно различать близкие формы патологии и идентифицировать патологическое состояние. Эти программы разрабатываются с учетом поэтапного обследования пациентов, и в их создании активно участвуют медицинские специалисты различного профиля, работая сообща для достижения основных медицинских целей (рисунок 14).

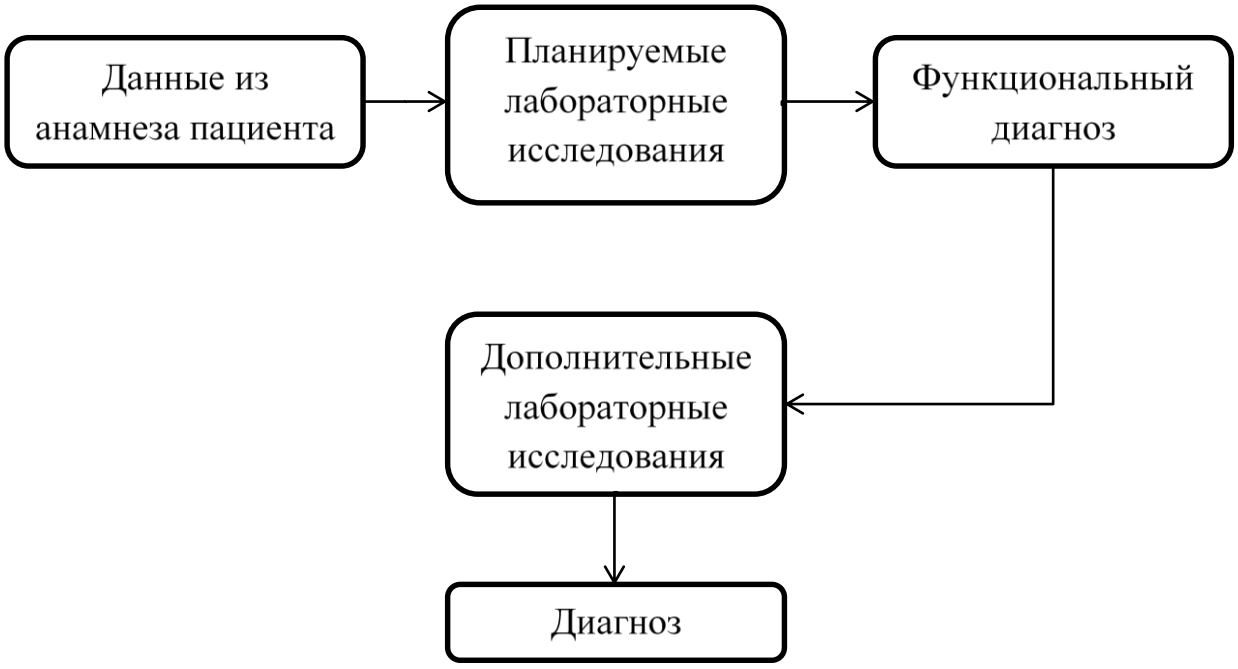


Рисунок 14 – Схема обследования пациента по этапам

Для реализации цели исследования предусматривается создание алгоритмов и логических структур, способных оптимизировать и улучшить

процесс лабораторной диагностики. Указанные алгоритмы и структуры будут соответствовать заданиям, определенным врачом-клиницистом при выборе лабораторных тестов в процессе осмотра пациента, таким как:

– диагностирование определенной нозологии с использованием лабораторных данных;

– идентификация прогностически негативных комбинированных показателей на основании лабораторных данных в процессе отслеживания динамики нозологии или реакции на терапевтическое воздействие;

– оценка степени тяжести нозологии и составление прогноза на основе лабораторных показателей.

Такие алгоритмы и модели помогут врачам сделать более обоснованный выбор лабораторных тестов и улучшить качество диагностики и лечения.

Пример алгоритма лабораторной диагностики щитовидной железы. Аномалии щитовидной железы наблюдаются у большого количества людей и отвечают за 30% всех эндокринных нарушений. В сочетании с клинической оценкой биохимические диагностические маркеры помогают исследовать этиологию дисфункции щитовидной железы и контролировать ее состояние после терапевтических вмешательств. Алгоритм представляет собой каскадное тестирования, предложенной компанией Roche Diagnostics и состоит из следующих тестов (таблица 5).

Таблица 5 – Таблица тестов для каскадного тестирования щитовидной железы

Наименование	Описание
ТТГ	Тиреотропный гормон – основной регулятор функции щитовидной железы
fT4/T4	Тироксин свободный (fT4) и общий (T4)
fT3/T3	Трийодтиронин свободный (fT3) и общий (T3)
Anti-TSHR	Антитела к рецепторам ТТГ
Anti-TPO	Антитела к ТПО – это белковые соединения, чье действие направлено против ферментов, отвечающих за формирование активной формы йода, необходимой для синтеза тиреоидных гормонов
Anti-Tg	Антитела к тиреоглобулину специфические иммуноглобулины, направленные против предшественника гормонов щитовидной железы

Алгоритм каскадного тестирование имеет следующий вид (рисунок 15).

Текущая корректность и скорость установления диагноза, а также мониторинг эффективности терапии, определяются не только компетентностью и знаниями врача, но и в значительной мере от качества работы лаборатории. Развитие в секторе лабораторной диагностики, как дисциплины, основывающейся на достижениях различных областей науки и информационных технологий, влечет за собой то, что специалисты в области лабораторной диагностики, обладая знаниями в этих многоаспектных областях, удаляются от практики клинического лечения. Тем временем, клиницисты не всегда способны успевать ассимилировать объем информации касательно

специфичности и чувствительности постоянно увеличивающегося перечня лабораторных исследований.

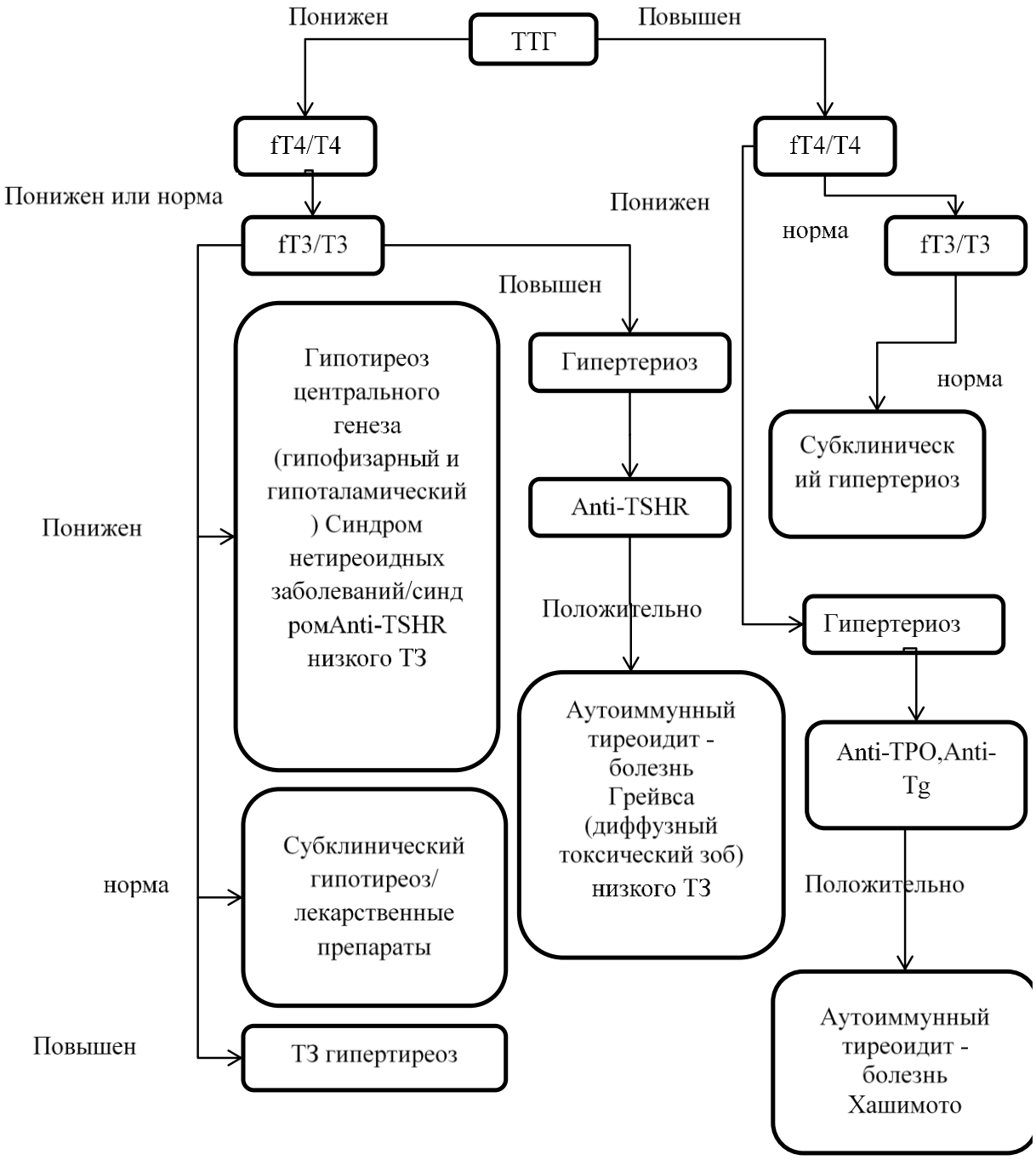


Рисунок 15 – Алгоритм каскадного тестирование щитовидной железы

Исходя из вышеуказанного и учитывая концепцию взаимодействия между клиницистами и лабораторными врачами в рамках диагностирования и прогнозирования заболеваний, а также мониторинга состояния пациента в ходе терапии, и применяя методологию, базирующуюся на логическом моделировании лабораторной диагностики, появляется возможность представить унифицированные основания метода для рационального использования лабораторных анализов с целью предоставления

интеллектуальной поддержки медицинскому персоналу (включая как клиницистов, так и специалистов лабораторий).

Оценка диагностической значимости лабораторных тестов. Поскольку сегодня акцент смещается не столько на фундаментальное признание клинической эффективности лабораторных исследований, сколько на информационную стоимость результатов таковых, актуален выделение критериев для измерения данного параметра. К ключевым из них можно отнести: диагностическую содержательность тестов, надежность лабораторных данных и оперативность предоставления информации клинической практике. Один из наиболее применяемых подходов в настоящий момент — вычисление диагностической чувствительности и специфичности лабораторного анализа [70], предложенное R.S. Galen и S. R. Gambing (1975) и основанное на применении теоремы Байеса.

Диагностический тест включает в себя определение диагностического индикатора через специфический лабораторный метод, причем его аналитические характеристики остаются стабильными при условии контроля качества выполнения теста.

Результат применения диагностического теста выражается в получении информации. Независимо от характера получаемой информации (качественной или количественной), результаты всех тестов можно разделить на положительные и отрицательные. В роли такой условной границы является одна из границ нормы, хотя могут быть выбраны и другие. Эти альтернативные границы могут быть применены для решения различных клинических задач, например, для определения прогностических факторов, факторов риска или для контроля терапии.

При оценке аналитических параметров теста он применяется к двум исследуемым группам: пациентам с заболеванием и контрольной группе. Соотношение между результатами теста и верным диагнозом иллюстрируется в следующей диаграмме (рисунок 16).

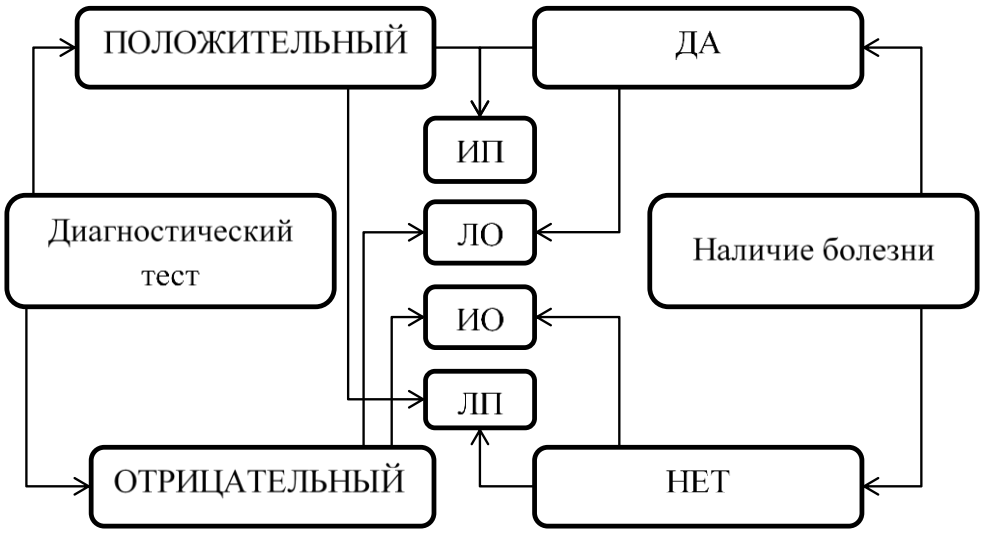


Рисунок 16 – Диаграмма тестирования и постановка корректного диагноза

Где, истинно положительный (ИП) – когда тест правильно идентифицирует наличие заболевания или состояния, истинно отрицательный (ИО) - когда тест правильно определяет отсутствие заболевания или состояния, ложноположительный (ЛП) – когда тест неверно указывает на наличие заболевания или состояния, которого на самом деле нет, ложноотрицательный (ЛО) - когда тест неверно указывает на отсутствие заболевания или состояния, которое на самом деле присутствует.

Результаты теста могут быть проинтерпретированы в четырех вариантах - два из них являются верными, а два – неверными. Самые значимые характеристики любого теста включают в себя аналитическую точность, чувствительность и специфичность (таблица 6).

Таблица 6 – Таблица имплементации результатов тестов

Наименование	Формула
Точность	Точность = (ИП+ИО)/(ИП+ИО+ЛП+ЛО)
Чувствительность (Ч)	Ч= ИП/(ИП +ЛО)
Специфичность (С)	С = ИО/(ИО+ЛП)
Фактор риска	Фактор риска = Ч / (1-С)
Прогноз положительного результата (ППР)	ППР = ИП/(ИО+ЛП)
Прогноз отрицательного результата (ПОР)	ПОР = ИО/(ИО+ЛП)

Для сравнения различных диагностических тестов между собой используется понятие аналитической точности, которая представляет собой отношение числа верных результатов к общему числу результатов обследования. Этот показатель позволяет сопоставить разные методы определения одного показателя при исследовании одной группы пациентов. Чувствительность – это отношение числа верных положительных результатов к общему числу больных, проходивших обследование. Если существует риск пропустить заболевание при неясной диагностической картине или если нужно исключить несколько распространенных причин с помощью высокочувствительных тестов, то предпочтительным выбором будет чувствительный тест. В этом случае, когда тест дает отрицательный результат, он наиболее информативен для врача. Специфичность указывает на отношение числа верных отрицательных результатов к общему числу контрольной группы.

Чувствительность теста зависит от выбранной границы нормы. Если граница нормы установлена на более низком уровне, то тест будет более чувствительным, и он сможет выявлять больше случаев заболевания, включая тех, у кого болезнь находится в начальной стадии или имеет более легкое течение. Однако при этом могут возрастать ложноположительные результаты у здоровых людей, что может привести к ненужным обследованиям и беспокойству пациентов. Понятие нормы и здоровья, как вы отметили, действительно условно и может сильно варьироваться в зависимости от конкретных параметров и контекста исследования. Поэтому при разработке и оценке лабораторных тестов важно учитывать клинические и эпидемиологические данные, чтобы определить оптимальные границы нормы и

баланс между чувствительностью и специфичностью, чтобы тест был наиболее информативным и полезным в конкретной медицинской практике.

«Нормальные границы» теста могут быть рассмотрены как биологическое и медицинское понятие, и они имеют различные аспекты:

1. Биологическая норма: Она определяется биологической вариацией изучаемого параметра в популяции здоровых людей. Биологическая норма может варьироваться в зависимости от возраста, пола, этнической группы и других факторов. Она основывается на научных данных и исследованиях, которые определяют, какой диапазон значений считается типичным для данной популяции.

2. Медицинская норма: это понятие более специфичное и связано с клиническими задачами и целями диагностики. Медицинская норма учитывает не только биологическую вариацию, но и то, какие значения параметра могут быть связаны с определенными медицинскими состояниями или рисками. Медицинская норма может изменяться с течением времени и развитием медицинской науки, поскольку новые исследования могут обнаруживать связи между значениями параметра и конкретными заболеваниями.

3. Соотношение с риском неблагоприятного результата: медицинский стандарт также может быть установлен на основе эпидемиологических данных о состоянии здоровья аналогичной клинической группы. Оценка вероятности неблагоприятного исхода ассоциируется с тем, какие значения параметра могут быть связаны с повышенным риском появления конкретных болезней или компликаций.

Использование комбинаций тестов может значительно повысить эффективность диагностики. Существует два основных способа применения комбинаций тестов: параллельный и последовательный.

Параллельный способ (Parallel Testing) - при использовании этого метода несколько тестов проводятся одновременно или параллельно у одного и того же пациента. Результат считается положительным, если хотя бы один из тестов является положительным. Этот метод обычно применяется, когда требуется высокая чувствительность, и важно не упустить ни одного случая болезни. Он может использоваться для скрининговых тестов или в случаях, когда пропущенный диагноз более нежелателен, чем ложноположительный результат.

Последовательный способ (Sequential Testing) – при этом методе различные тесты применяются последовательно. Если первый тест положительный, то может быть проведен второй тест для подтверждения диагноза. Этот метод обычно используется, когда более важно минимизировать ложноположительные результаты. Он может уменьшить число ложноположительных результатов, но может также увеличить вероятность ложноотрицательных результатов, так как положительный результат первого теста может привести к пропуску второго теста.

Выбор между параллельным и последовательным способами зависит от конкретных клинических условий и целей тестирования. Параллельный метод

может быть предпочтителен в случаях, когда важно выявить как можно больше случаев заболевания, даже при наличии ложноположительных результатов. Последовательный метод может быть полезен, когда цель - минимизировать ложноположительные результаты, но это может привести к более высокой вероятности ложноотрицательных результатов.

Для оперативной оценки состояния пациента осуществляется параллельное назначение тестов. Целью параллельно назначаемых тестов является увеличение претестовых шансов или вероятности того, что пациент имеет определенное заболевание. Это может быть полезно для идентификации критических случаев и принятия быстрых медицинских решений. Однако такие тесты могут быть менее специфичными и давать больше ложноположительных результатов, что может потребовать последующего более детального и специфического обследования для окончательного диагноза. Параллельное назначение тестов может быть полезным инструментом для начальной оценки состояния пациента и повышения вероятности выявления критических состояний, но требует дополнительного анализа и дальнейшего диагностического процесса для установления окончательного диагноза.

Параллельное тестирование обеспечивает высокую чувствительность и прогнозируемую отрицательную ценность (ПОР), что означает способность выявлять больных с заболеванием, особенно при его высокой распространенности.

В результате применения параллельного тестирования в специализированных лабораториях может возникать тенденция к диагностированию заболеваний, которые не были выявлены врачами поликлиник. Это может быть полезным в случаях, когда необходимо выявить скрытые или редкие заболевания. Однако при слишком широком и необоснованном использовании параллельных тестов может возникать гипердиагностика, что означает диагностирование заболеваний у пациентов, которые фактически не нуждаются в лечении.

При последовательном присвоении тестов целесообразно классифицировать их в две категории: скрининговые и подтверждающие, исходя из их лабораторных спецификаций. Скрининговые тесты должны характеризоваться высокой чувствительностью, предполагающей высокую способность обнаруживать заболевание, и, следовательно, высокой отрицательной прогностической ценностью. Напротив, подтверждающие тесты проявляют высокую специфичность, что подразумевает их способность эффективно различать наличие и отсутствие заболевания, и, таким образом, имеют высокую положительную прогностическую ценность.

Последовательное применение диагностических тестов особо ценно в ситуациях, когда ни одно исследование не отличается высокой специфичностью. Таким образом, тесты, проводимые последовательно, должны исследовать одинаковый биологический феномен и их результаты должны быть консистентными друг с другом. Эти последовательно выполняемые тесты

могут быть интегрированы в диагностические алгоритмы, включающие тесты с разнообразными клиническими и лабораторными показателями.

2.3 Формирование базы знаний по интерпретации результатов лабораторных исследований

База знаний (БЗ; англ. knowledge base) - база данных, содержащая правила ввода и вывода информации о человеческом опыте и знаниях в некоторой предметной области (ISO/IEC/IEEE 24765–2010 [71], ISO/IEC 2382-1:1993 [72] ISO/IEC/IEEE 24765-2010, Systems and software engineering).

Современные БЗ тесно взаимодействуют с системами поиска и извлечения информации. Для эффективной работы таких систем требуется определенная модель классификации понятий и структура представления знаний. Одним из способов иерархического представления знаний в БЗ является онтология. Онтология представляет собой иерархию понятий и их взаимосвязей в определенной области знаний. Часто термины «онтология» и «база знаний» используются взаимозаменяемо, так как онтология определяет структуру знаний и связи между ними. Полноценные базы знаний [73], в отличие от обычных баз данных, содержат не только фактическую информацию, но и правила вывода, которые позволяют автоматически делать умозаключения на основе имеющихся или вновь вводимых фактов. Это обеспечивает семантическую обработку информации и позволяет программам принимать осмысленные решения. Базы знаний играют важную роль в интеллектуальных системах, и одним из наиболее известных классов таких систем являются экспертные системы. Они используют БЗ для поиска решений в определенной предметной области на основе знаний, заключенных в базе, и на основе ввода пользователя, описывающего ситуацию или проблему. Экспертные системы способствуют автоматизации процесса принятия решений и решению задач, связанных с определенными областями экспертизы [74].

Простые базы знаний могут быть использованы для создания интеллектуальных информационных систем внутри организации, где они служат для хранения данных, таких как документация, руководства и технические статьи. Основная цель создания таких баз - помочь менее опытным сотрудникам быстро найти существующие решения для различных проблем и задач. Процесс поддержания и актуализации баз знаний внутри корпоративных информационных систем предприятий является довольно трудоемким. Он включает в себя выполнение множества поисковых операций как внутри корпоративных сетей, так и во внешних источниках, включая интернет [75]. Это необходимо для обновления информации, добавления новых данных и обеспечения актуальности знаний в базе. Важно отметить, что такие базы знаний играют ключевую роль в обеспечении доступа к информации и опыту организации, что способствует более эффективному решению задач и улучшению качества работы сотрудников.

Основной задачей клинической лабораторной диагностики является выявление или подтверждение наличия патологии, которую невозможно

однозначно определить с помощью органолептических методов исследования. Для достижения этой цели могут применяться различные вспомогательные методы. Эти методы условно можно разделить на следующие группы [76]:

– методы, расширяющие возможности визуального восприятия человека, такие как оптическая микроскопия;

– методы, основанные на использовании характерных биохимических особенностей организма, такие как серологические методы диагностики;

– методы, опирающиеся на характерные особенности патологического агента, такие как биологические и культуральные методы исследования.

В лабораторной диагностике базой знания является база данных постоянно пополняющиеся данными об исследованиях, тестах, единицах измерений, нормативных величинах к тестам, биоматериалами, и множество других параметров, которые участвуют в проведение исследований и интерпретации результатов.

Лабораторные исследования можно разделить на следующие группы по видам результатов:

1. Качественные – где результаты лабораторных исследований представляются в виде чисел и интерпретируются путем отклонения от нормативных величин.

2. Количественные – где результаты представляется в виде текста, а именно прямым отклонением от нормы. Например, «положительно, отрицательно, выявлено, не выявлено»

3. Количественно-качественные – где результаты исследовании представляются в виде чисел и текстовом выражений.

В зависимости от видов исследования производится наполнения базы знаний, и устанавливаются все необходимые параметры и связи. Логическая модель базы знаний представлен на следующем рисунке 17.

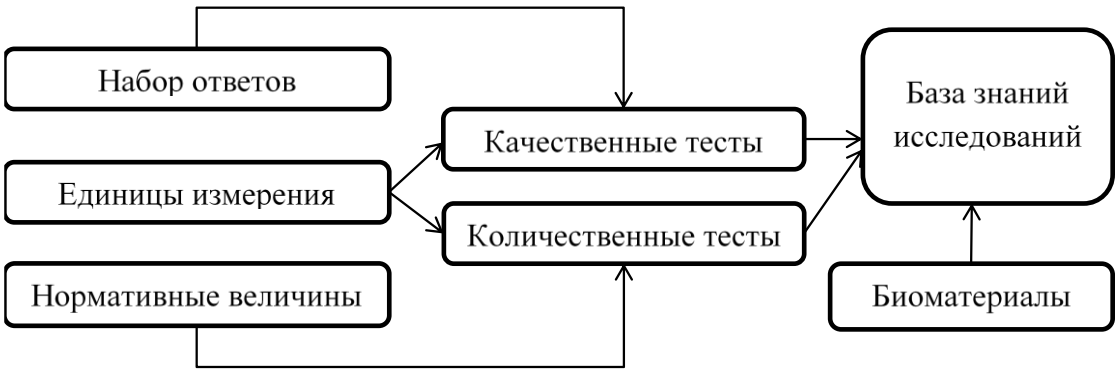


Рисунок 17 – Схема формирования базы знаний для тестов

Согласно рисунку 16, формирование база знаний исследований состоит из шести основных блоков, где каждый блок также представляет собой определенный набор данных:

– набор ответов – представляет справочник преднастроенных ответов для выбора при формирование заключения результата лабораторного теста (таблица 7).

Таблица 7 - Набор ответов для результата качественных тестов

Наименование	Код	Альтернативное название
O(I) первая	O(I) первая	Group O
A(II) вторая	A(II) вторая	Group. A
B(III) третья	B(III) третья	Group B
AB(IV) четвертая	AB(IV) четвертая	Group AB
Отрицательный	Отрицательный	Negative
Положительный	Положительный	Positive
Отрицательно	Отрицательно	Терiс
Положительно	Положительно	Оң

– единицы измерения – представляет собой справочник единиц измерений по международном системе единиц СИ (таблица 8) [77].

Таблица 8 – Пример единиц измерения в системе единиц СИ

Наименование	Код	Мнемоника	Альтернативное название
Мкг	Масс003	Мкг	ug
мкг/дл	11	мкг/дл	ug/dL
мкг/л	Масс/Об006	мкг/л	ug/L
мкг/мкл	Масс/Об008	мкг/мкл	ug/uL
мкг/мл	Масс/Об007	мкг/мл	ug/ml
мкЕд/мл	8	мкЕд/мл	uU/ml

– биоматериалы – представляет собой справочник биологических материалов с помощью которых производится лабораторные исследования. Для каждого исследования прикрепляется надлежащие биоматериалы. Для некоторых видов исследовании (микробиологические) могут устанавливаться множество биоматериалов, такие как:

- кровь ЭДТА;
- сыворотка;
- RW-сыв.;
- кровь;
- соскоб;
- плазма (цитрат Na);
- моча.

– нормативные величины – представляет собой набор данных с величинами определящее отклонения теста установленных границ. Для определения величин используются следующие параметры:

– группа пациентов: Возраст [от, до] (день, неделя, месяц, год, лет);

– пол (мужской, женский);

– Интервал границ, где устанавливаются нормативные величины для каждого теста согласно данным производителя реагента (таблица 9).

Таблица 9 – Представление интервалов для количественных тестов

Критический низкий	Нижняя норма	Верхняя норма	Критический высокий
1	2	3	4

– качественные тесты – набор данных о тестах, которые были сформированы согласно составным элементам, где результатами будут текстовые ответы.

– количественные тесты – набор данных о тестах, которые были сформированы согласно составным элементам, где результатами будут числовые данные.

Каждый тест и исследования самостоятельно являются деревом решений, так как в зависимости от методов проведения анализов они имеют различия по химическим составляющим. Классификация дерева теста принимает следующий вид (рисунок 18).

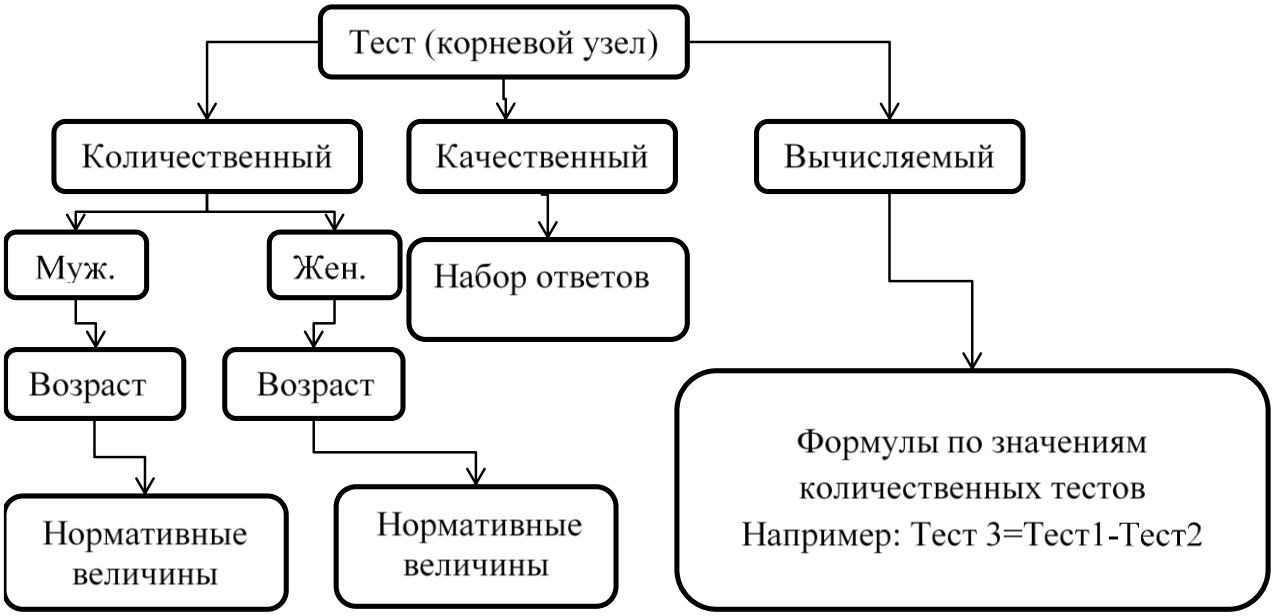


Рисунок 18 – Классификация дерева формирования теста

Где, тест является корневым узлом, виды тестов являются узлами, и распределяются по свойствам и параметрам.

Сформированная база знаний представляет собой набор данных, имеющих множество связей между тестами и исследованиями (рисунок 19).

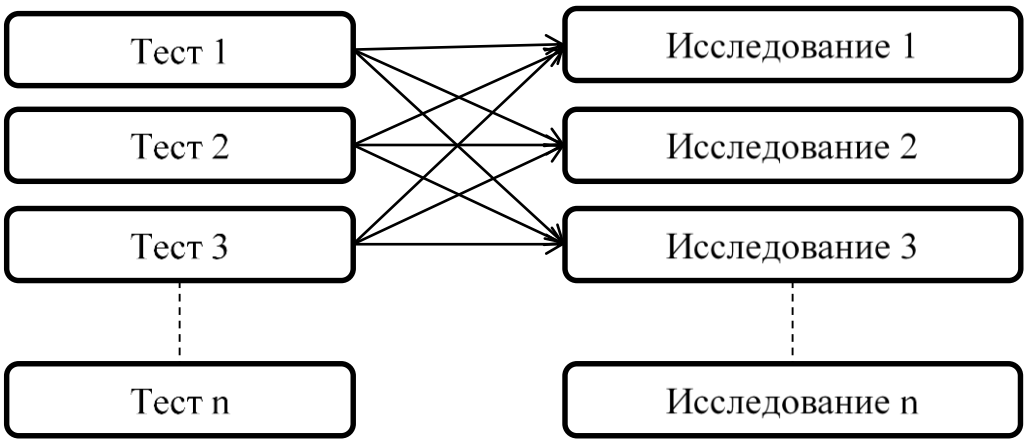


Рисунок 19 – Связи тестов и исследований

База знаний формируется учителем, то есть все взаимосвязи определяются согласно медицинской документации определенного исследования.

Выявление новых закономерностей производится методами машинного обучения, такими как Байесовский классификатор, Дерево решений, TF-IDF анализ. Данные в базах данных являются систематизированными и структурированными, а также пополнение данных о результатах пациентов также являются чистыми, что позволяет максимально точно производить анализ данных.

2.4 Моделирование интерпретации референсных значений по отклонению от нормативных величин

Нормативные величины или референс значения являются ядром определение отклонений результатов лабораторных исследований. Референс значения определяются врачами клиницистами, врачами лаборантами согласно медицинской документации к каждому виду исследования. В рамках диссертационной работы была сформирована база знаний боле 4500 референс значений для более 2000 тестов и 1400 исследований. Каждый тест в базе знаний имеет свой уникальный идентификатор, который позволяет определить тест и осуществлять по нему поиск зависимых данных в базе знаний. С помощью уникального идентификатора производится интерпретация для всех видов тестов.

Каждый вид теста имеет свои данные по интерпретации, также имеются различия по видам исследований. Интерпретация качественных видов исследований производится путем сравнения полученного результата с набором предварительного множества данных, среди которых по умолчанию должен быть отрицательный показатель, иными словами показатель отрицающая патологию.

Интерпретация качественных исследований производится по следующей логической модели (рисунок 20).

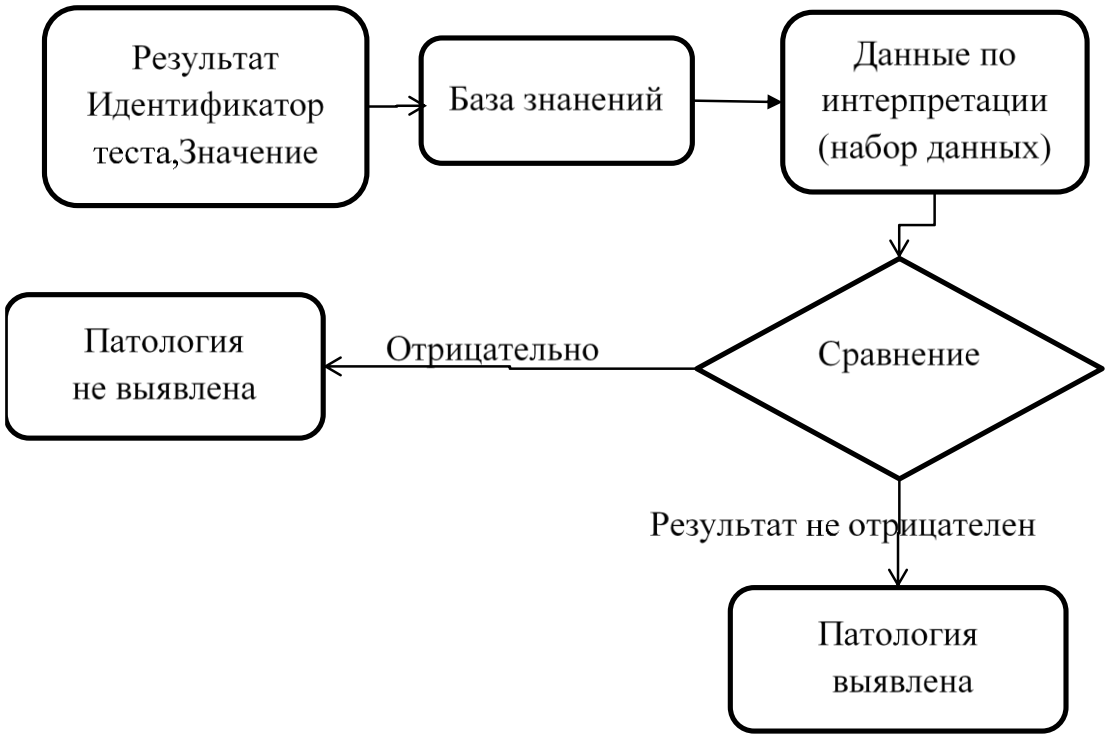


Рисунок 20 – Алгоритм интерпретации качественных тестов

Согласно логической модели, идентификатор теста и его результат проходит проверку в базе знаний, после получения нормативных данных, производится сравнение и предоставляется информация об отклонениях.

Интерпретация количественных и вычислительных тестов производится путем анализа нормативных величин, такие как верхний и нижний порог значений. Информация о референсных значениях извлекаются путем обращения к базе знаний через уникальный идентификатор теста и его количественного результата, а также возраст и пол пациента влияет на получение корректных нормативных величин. Процесс определения отклонений для количественных тестов описаны и отражен в следующей логической модели (рисунок 21).

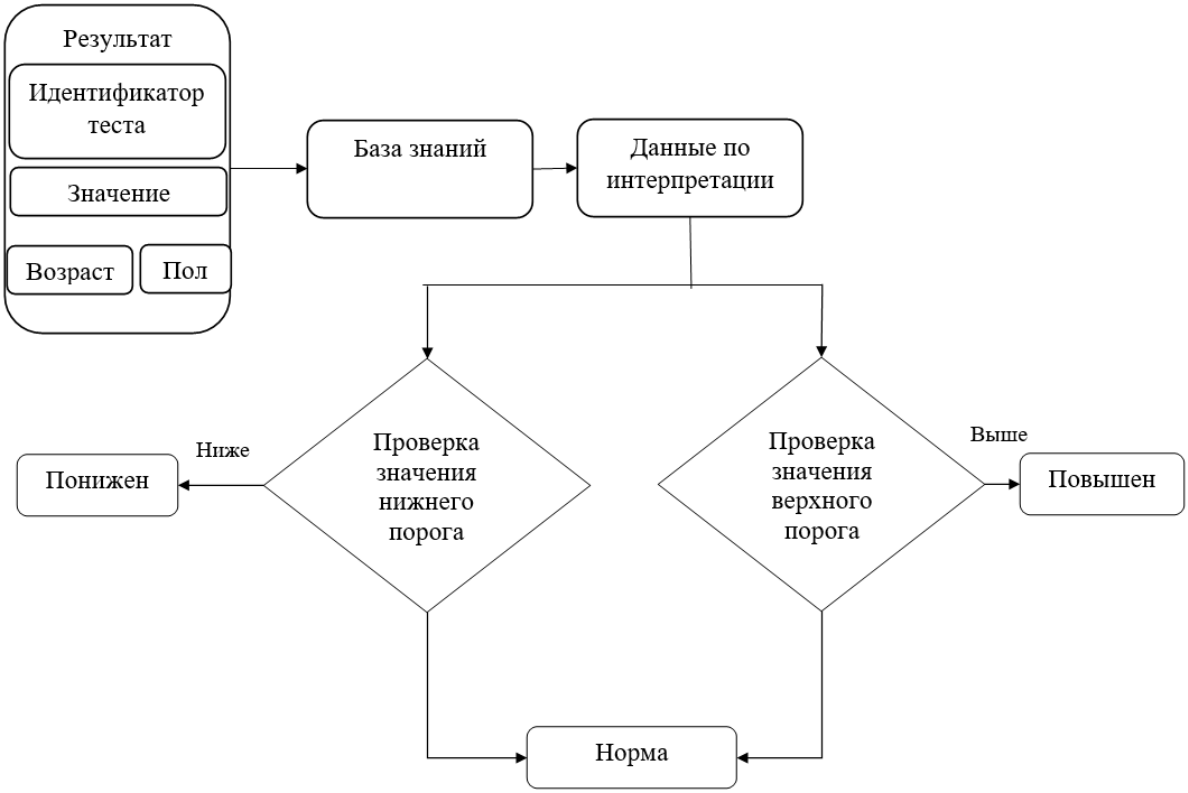


Рисунок 21 – Модель интерпретации количественных тестов

Согласно представленной модели, значение результата лабораторного исследования проверяется в интервале данных, то есть между значениями нижнего и верхнего порога. При отклонении полученного результата, а именно если значение больше верхнего или ниже нижнего порога, то это будет отклонением от нормальных величин.

Процесс интерпретации отклонении от нормальных величин результатов лабораторных исследовании производится в режиме реального времени, то есть с использованием технологии подключения лабораторного оборудования к информационной системе. Что позволяет производить автоматическую интерпретацию отклонений без участия специалистов. Процесс автоматизации описан ниже в 4 разделе.

Определение отклонений производится с применением языка структурированных запросов SQL, с помощью которого происходит мгновенный запрос в базу по входящим параметрам и определяются отклонения от референс значений. В математическом представлений определяют несколько основных понятий для набора данных:

1. A (конечное) множество значений столбцов S_i, индексируемое некоторым (конечным) множеством I, где Каждый набор столбцов представляет некоторый тип данных (или область данных): конечное сравнимое по равенству множество возможных значений данных.

2. A (конечное) множество отношений R: из A идентификаторов столбцов и значений для некоторого подмножества столбцов.

Работа с данными осуществляется в терминах конечного предикатного исчисления первого порядка. В этих терминах отношения обычно определяются как функция истинности:

$$\prod_i S_i \xrightarrow{f} \{T, F\} \quad (1)$$

где $f\big((s_1 \ldots . s_n)\big) = T \Leftrightarrow (s_1 \ldots . s_n) \in R$

Фундаментально, все конечные данные можно рассматривать как сложные утверждения о данных, которые могут быть сформулированы как предикатные утверждения о функции истинности.

2.5 Разработка моделей для анализа и интерпретации данных лабораторных исследований

Интерпретация результатов лабораторных исследований является конечной точкой по определению причины отклонения, и в дальнейшем наступает следствие для формирования дополнительных контрольных лабораторных анализов, проведение других диагностических исследований для уточнения патологии пациента.

Модель определения отклонении было описано в разделе 2.4., с помощью которого выявляются патологичные результаты. Для получения информации об отклонения строится дерево решений по причинам отклонения согласно клиническим показателям каждого теста. Вероятность наступления причины определяется применением Формула Байеса формула (2). По формуле Байеса вероятность, как и в самых простых случаях, вычисляется как отношение «одного ко всем»:

$$P(B_j|A) = \frac{P\big(A\big|B_j\big)\times P(B_j)}{\sum_{j=1}^{n} P(A|B_j)\times P(B_j)} \quad (2)$$

где, знаменатель полная вероятность события A, а числители для каждого отдельного случая равны первому, второму, и так далее до n - го, слагаемому суммы, находящейся в знаменателе.

Для каждого теста формируется база знаний по причинам повышения и понижения нормативных величин, с привязкой к дополнительным диагностическим тестам, где также формируется свои взаимосвязи, в результате которого получается связь многие ко многим между причинами и тестами (рисунок 22).

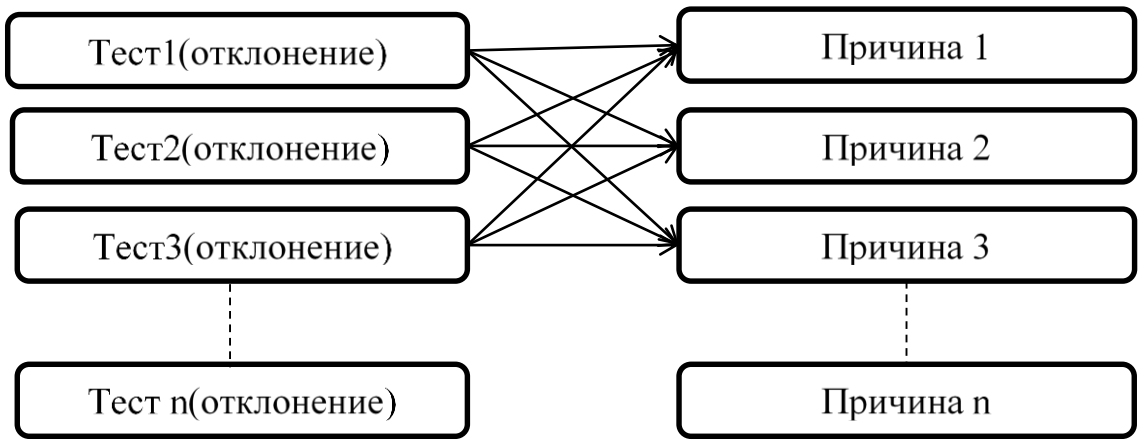


Рисунок 22 – Взаимосвязь отклонении тестов

На рисунке описана общая логическая модель причины, так как взаимосвязи для причины понижения и повышения строятся по отдельности. Причины отклонения тесно связаны клиническими данными пациента, и интерпретация может наиболее точной при наличии полного объема проведенных анализов. Процесс формирования базы знаний причин отклонений и взаимосвязанных тестов описан в разделе 4.

С помощью множественных взаимосвязей между тестами и причинами отклонении будет производиться определение предварительной интерпретации результатов лабораторных исследований путем пересечения причин (рисунок 23).

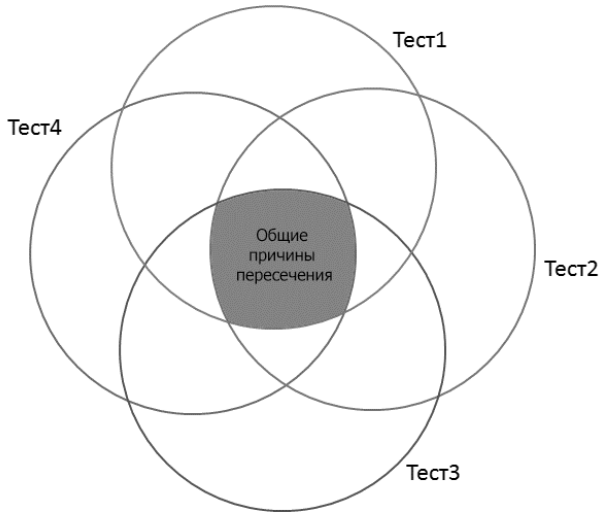


Рисунок 23 – Пересечения областей по отклонению

Метод пересечения является элементом реляционной алгебры, где отношение с тем же заголовком, что и у отношений A и B, и телом, состоящим из кортежей, принадлежащих одновременно обоим отношениям A и B, и в частном случае описывается $A \bigcap B = \{x: x \in A \text{ и } x \in B\}$, а пересечение более чем

двух множеств (обобщенное пересечение) может быть записано как формула (4):

$$\bigcap_{i=1}^{n} A_i \quad (4)$$

После определения пересечения данных будет формироваться рейтинг пересеченных причин, от большого количество до меньшего (таблица 10).

Таблица 10 – Вероятность выявления причин отклонения

Причина	Вероятность, %
Причина 1	100
Причина 2	90
Причина 3	80
Причина 4	70
	60
Причина n	N

В результате модель предоставляет диагностическую карту интерпретации с вероятными патологиями.

Выводы по разделу

В данной главе мы определи все необходимые данные по формированию базы знаний, модели интерпретации отклонений, а также модель по интерпретации результатов лабораторных исследований. Были описаны методы, представлены логические модели и иллюстрации, которые будут неотъемлемой частью последующих разделов диссертационной работы. Были рассмотрены разносторонние аспекты использования алгоритмов для достижения поставленной цели. Модели, описанные в данном разделе, иллюстрируют пошаговый подход, для интерпретации результатов лабораторных исследований, так как применение алгоритмов по структурированным данным, и наличие предварительно обученной базы знаний эффективно ускорит процесс автоматизации и выявление патологии при диагностировании лабораторных проб. Были определены понятие как диагностическая карта пациента, содержащий результаты лабораторных исследований, проведенных у пациента. В ней содержится полный обзор состояния здоровья пациента, включая информацию о его жизненных показателях, истории результатов лабораторных исследований, и текущем состоянии здоровья. Диагностическая карта помогает медицинским работникам принимать обоснованные решения о лечении и ведении пациента.

3 РАЗРАБОТКА МОДЕЛИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПО ЛАБОРАТОРНОЙ ДИАГНОСТИКЕ В ЗДРАВООХРАНЕНИИ

3.1 Использование методик машинного обучения в анализе больших данных, полученных в результате лабораторных исследований

Машинное обучение (англ. machine learning, ML) представляет собой категорию методов искусственного интеллекта, которые отличаются от традиционных методов тем, что они не решают задачи напрямую, а, скорее, обучаются на основе опыта и данных, используя знания и решения, полученные из множества схожих задач. Для построения и применения методов машинного обучения используются разнообразные математические и вычислительные инструменты, включая методы математической статистики, численные методы, математический анализ, методы оптимизации, теорию вероятностей, теорию графов и различные методики обработки цифровых данных. В машинном обучении различают два основных типа обучения:

– обучение по прецедентам (индуктивное обучение): Этот тип обучения основан на анализе и использовании эмпирических данных. Модель машинного обучения изучает закономерности в предоставленных данных и пытается обобщить эту информацию, чтобы делать прогнозы или принимать решения на основе новых, ранее не виденных данных. Примерами индуктивного обучения являются алгоритмы классификации, регрессии и кластеризации;

– дедуктивное обучение: Этот подход предполагает формализацию знаний экспертов и перенос их в компьютерную систему в виде базы знаний. В этом случае модель машинного обучения использует заранее определенные правила и знания для решения задачи. Дедуктивное обучение часто используется в экспертных системах, где экспертные знания формализованы и могут быть использованы для принятия решений.

Оба типа обучения имеют свои применения и ориентированы на разные сценарии. Индуктивное обучение широко используется для анализа данных и создания прогностических моделей, в то время как дедуктивное обучение полезно там, где существует явная формализованная база знаний.

Дедуктивное обучение традиционно ассоциируется с областью экспертных систем, в которых знания экспертов формализуются и используются для решения конкретных задач. В контексте машинного обучения термины «машинное обучение» и «обучение по прецедентам» часто используются как синонимы, особенно когда речь идет об индуктивном обучении.

Многие индуктивные методы обучения были разработаны как альтернатива традиционным статистическим методам. Множество таких методов имеют тесную связь с задачами извлечения информации и интеллектуальным анализом данных [78].

Классические задачи, которые решаются с использованием машинного обучения, включают в себя следующие:

– классификация – выполняется с применением обучения с учителем на этапе обучения. Задачей является разделение данных на различные классы или категории;

– кластеризация – выполняется с использованием обучения без учителя. Задачей является группировка данных на основе их сходства без заранее известных категорий;

– регрессия – выполняется с помощью обучения с учителем на этапе тестирования и представляет собой частный случай задачи прогнозирования. Задачей является предсказание числовых значений;

– понижение размерности данных и их визуализация – выполняется с помощью обучения без учителя и позволяет уменьшить размерность данных и визуализировать их, чтобы облегчить анализ;

– восстановление плотности распределения вероятности – позволяет оценить вероятностное распределение данных на основе имеющихся наблюдений;

– одно классовая классификация и выявление новизны - задачей является определение аномалий или новых, несоответствующих известным классам данных;

– построение ранговых зависимостей – позволяет определить и оценить связи и зависимости между переменными в данных.

– обнаружение аномалий – задачей является выявление аномальных или необычных событий, или объектов в данных.

Большие данные (англ. big data, [ˈbɪg ˈdeɪtə]) – обозначение структурированных и неструктурированных данных огромных объёмов и значительного многообразия, эффективно обрабатываемых горизонтально масштабируемыми программными инструментами, появившимися в конце 2000-х годов и альтернативных традиционным системам управления базами данных и решениям класса Business Intelligence [79].

В качестве определяющих характеристик для больших данных традиционно выделяют «три V»: объём (англ. volume, величина физического объёма), скорость (velocity скорости прироста, так и необходимости высокоскоростной обработки и получения результатов), многообразие (variety, возможности одновременной обработки различных типов, структурированных и полуструктурированных данных) [80];

В диссертационной работе используются обезличенные данные более 7 млн. 600 тысяч результатов лабораторных исследований по количественным тестам, по подразделениям гематология, биохимия и иммунология. Данные содержат 439 уникальных исследований, выбранных за период с января 2020 года по октябрь 2022 года. Обработка и анализ данных будет производиться по патологичным результатам (примерно 1 890 000 результатов), имеющих отклонение от нормативных величин, модель которого был представлен во второй главе диссертационной работы.

Целью анализа данных является определение закономерности патологичных результатов и сравнение их действующими медицинскими

протоколами диагностики заболеваний. В качестве медицинских протоколов диагностики будут использованы определение анемии и дефицита щитовидной железы, алгоритм определения которых был описан в разделе 2.2.

Из методов машинного обучения использован Наивный байесовский алгоритм классификации, деревья решений, логистическая регрессия, нейронные сети.

В диссертационной работе будут использоваться методы обучения с учителем, так как в лабораторной диагностике все процедуры диагностики регламентированы правилами и приказами министерства здравоохранения и международных организации в сфере здравоохранения.

Наивный байесовский алгоритм классификации. Для построения модели обучения необходимо подготовить предварительные данные. Обработка первичных производится на языке SQL, анализ будет данных с использованием платформы R-studio (язык R).

В качестве предварительных данных произведем выборку тестов связанные для диагностирования анемии. Анемия представляет собой медицинское состояние, характеризующееся снижением концентрации гемоглобина в крови и, в большинстве случаев, уменьшением количества эритроцитов и гематокрита в единице объема крови.

Классификация анемии выглядит следующим образом:

– железодифицитная анемия;
– талассемия;
– сидеробластная анемия;
– анемия хронических заболеваний;
– почечная недостаточность;
– миелодиспластический синдром;
– острая геморрагическая анемия;
– гемолитическая анемия;
– В-12 (фолиево) дифицитная анемия.

В диссертационной работе для диагностики анемии были выбраны следующие тесты (таблица 11).

Таблица 11 – Перечень исследуемых тестов по диагностике анемии

Код	Наименование теста
1	2
10715	Гемоглобин
10716	Эритроциты
10718	Гематокрит
10719	Средний объем эритроцитов
10720	Среднее содержание Hb в одном эритроците
10721	Средняя концентрация Hb в эритроцитах
10722	Расчетная ширина распределения эритроцитов по объёму (коэф, вариации)
10724	Тромбоциты
10728	Лейкоциты

Продолжение таблицы 11

1	2
10765	Ретикулоциты
10907	Железо
10908	Латентная (ненасыщенная) железосвязываюшая способность сыворотки
10909	Трансферрин
10910	Ферритин
10938	Фолиевая кислота (витамин B9)
27171	Витамин B12

Анализ данных патологичных результатам, определения веса тестов по понижению и повышению от нормальных величин в наборе данных. Для удобства обработки данных вместо наименование тестов выбраны их кодировка в базе (таблица 12). Набор данных состоит из 496 979 записей, и отобран по следующим параметрам:

– наименование теста;
– статус отклонения (повышение или понижение от нормы);
– пол;
– возраст.

После импорта данных в среду обработки R-studio, получаем информацию о загруженных данных и атрибутов для дальнейшего анализа.
Следующим шагом производится построение матрицы и осуществляется расчет количество результатов тестов с повышенными и пониженными показателями, где 0 - понижение от нормы, 1 - повышение от нормы (рисунок 24).

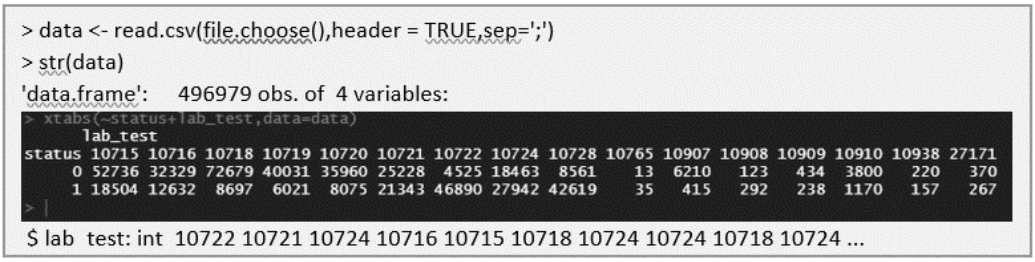


Рисунок 24 – Перечень анализируемых тестов по отклонению от нормативных величин

На рисунке 24 можно увидить соотношение матрицы статуса к тестам, и на пересечении количество результатов соответственно.

Таблица 12 – Перечень анализируемых тестов с кодами

Код	Наименование	Кол-результатов с пониженными значениями	Кол-результатов с повышенными значениями
1	2	3	4
10715	Гемоглобин	52 736	18504

Продолжение таблицы 12

1	2	3	4
10716	Эритроциты	32 329	12632
10718	Гематокрит	72 679	8697
10719	Средний объем эритроцитов	40 031	6021
10720	Среднее содержание Hb в одном эритроците	35 960	8075
10721	Средняя концентрация Hb в эритроцитах	25228	21343
10722	Расчетная ширина распределения эритроцитов по объёму (коэф, вариации)	4525	46890
10724	Тромбоциты	18463	27942
10728	Лейкоциты	8561	42619
10765	Ретикулоциты	13	35
10907	Железо	6210	415
10908	Латентная (ненасыщенная) железосвязывающая способность сыворотки	123	292
10909	Трансферрин	434	238
10910	Ферритин	3800	1170
10938	Фолиевая кислота (витамин B9)	220	157
27171	Витамин B12	370	267

Следующим шагом производится классификация данных по байесовскому алгоритму. В R-studio программный код модели будет выглядеть следующим образом:

```
set.seed(1234)
ind <- sample(2,nrow(data),replace = TRUE,prob=c(0.8,0.2))
train <- data[ind==1,]
test <- data[ind==2,]
model<-naive_bayes(status ~ .,laplace = 1,data=train)
model
```

Было построено две модели на осное тестов и статусов отклонении. После выполнения программного кода априорная вероятность по статусам отклонения имеет следующий вид (рисунок 25).

```
================================================================================ Naive Bayes =======

 Call:
naive_bayes.formula(formula = status ~ ., data = train, laplace = 1)

--------------------------------------------------------------------------------------------------------

Laplace smoothing: 1

--------------------------------------------------------------------------------------------------------

 A priori probabilities:

        0         1
0.6064137 0.3935863

--------------------------------------------------------------------------------------------------------
```

Рисунок 25 – Априорная вероятность

Где анализ показывает, что индикация результат с пониженными показателями больше чем с повышенными. По другим параметрам исследования можно также увидеть вероятности Бернулли и ранги по каждому тесту (рисунок 26).

```
 ::: lab_test (Categorical)
------------------------------------------------------------

lab_test            0           1
   10715 1.742166e-01 9.444906e-02
   10716 1.069056e-01 6.465027e-02
   10718 2.409513e-01 4.445225e-02
   10719 1.329842e-01 3.075056e-02
   10720 1.189955e-01 4.133504e-02
   10721 8.381635e-02 1.091600e-01
   10722 1.487599e-02 2.394890e-01
   10724 6.141115e-02 1.436346e-01
   10728 2.873206e-02 2.187416e-01
   10765 5.804456e-05 1.916321e-04
   10907 2.058094e-02 2.159055e-03
   10908 4.228961e-04 1.539444e-03
   10909 1.426238e-03 1.232833e-03
   10910 1.262055e-02 6.055573e-03
   10938 7.504333e-04 8.240179e-04
   27171 1.252104e-03 1.335037e-03

------------------------------------------------------------
 ::: age (Gaussian)
------------------------------------------------------------

age           0        1
  mean 29.59402 33.41782
  sd   21.98807 22.95964

------------------------------------------------------------
 ::: sex (Bernoulli)
------------------------------------------------------------

sex         0         1
  F 0.6647939 0.6229046
  M 0.3352061 0.3770954
```

Рисунок 26 – Результаты индикации

Если осуществить прогнозирование по выбранной модели, то получаем следующие вероятности из всего набора данных (рисунок 27).

	0	1	lab_test	status	age	sex
1	0.1020949	0.8979051	10722	1	22	F
2	0.5857144	0.4142856	10721	0	21	F
3	0.4404791	0.5595209	10724	1	21	F
4	0.6753547	0.3246453	10716	0	46	M
6	0.8721083	0.1278917	10718	0	46	M
7	0.4311920	0.5688080	10724	1	27	F

Рисунок 27 – Результаты прогнозирования

После сопоставления кодов теста с их наименованиями, то рисунок 26, принимает следующий табличный вид (таблица 13).

Таблица 13 – Результаты построения модели наивного Байеса

Вероятность пониженных результатов	Вероятность повышенных результатов	Наименование теста	Возраст	Пол
0.58484515	0.41515485	Средняя концентрация Hb в эритроцитах	21	Ж.
0.44660929	0.55339071	Тромбоциты	21	Ж.
0.67583130	0.32416870	Эритроциты	46	М.
0.70088415	0.29911585	Гемоглобин	46	М.
0.87285092	0.12714908	Гематокрит	46	М.
0.90716483	0.09283517	Гематокрит	27	Ж.
0.43729106	0.56270894	Тромбоциты	27	Ж.
0.88485051	0.11514949	Средний объем эритроцитов	26	Ж.
0.10119538	0.89880462	Расчетная ширина распределения эрит роцитов по объёму (коэф, вариации)	26	Ж.
0.88678888	0.11321112	Средний объем эритроцитов	23	Ж.
0.74665262	0.25334738	Эритроциты	28	Ж.
0.34438166	0.65561834	Тромбоциты	65	Ж.
0.34438166	0.65561834	Тромбоциты	65	Ж.
0.57889421	0.42110579	Средняя концентрация Hb в эритроцитах	25	Ж.
0.75046560	0.24953440	Эритроциты	25	Ж.
0.90828082	0.09171918	Гематокрит	25	Ж.
0.77169894	0.22830106	Гемоглобин	25	Ж.
0.77169894	0.22830106	Гемоглобин	25	Ж.
0.90828082	0.09171918	Гематокрит	25	Ж.
0.19166315	0.80833685	Лейкоциты	23	Ж.

Повторно выполнив тот же программный код на основе теста, были полученые следующие результаты (рисунок 28).

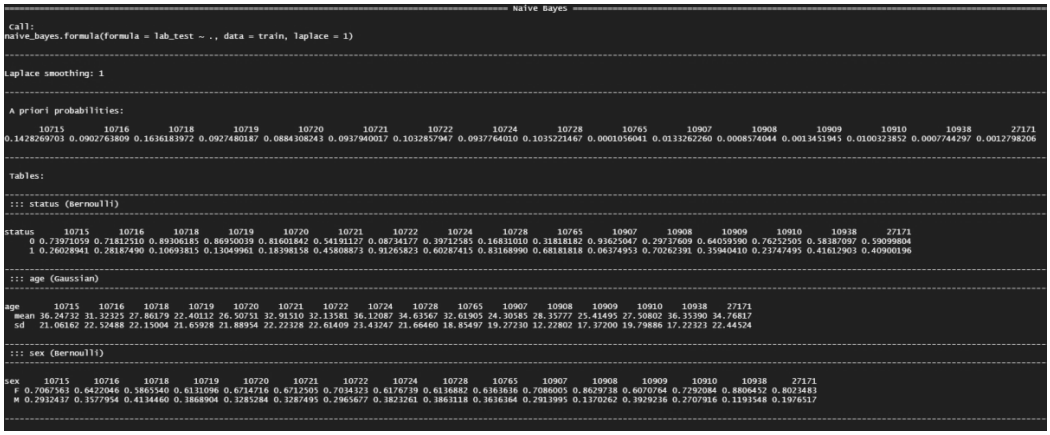


Рисунок 28 – Результаты повторного анализа

Если сравнить расчеты с предыдущими расчетами, то получаем что возрастные вероятности также остаются неизменными (рисунок 29).

```
> data$lab_test <- as.factor(data$lab_test)
> train %>% filter (status=="0") %>%
+             summarise(mean(age),sd(age))
  mean(age)  sd(age)
1  29.59402 21.98807
>
> train %>% filter (status=="1") %>%
+    summarise(mean(age),sd(age))
  mean(age)  sd(age)
1  33.41782 22.95964
>
```

Рисунок 29 – Суммарные расчеты по тестам по возрасту

В результате повторное вычисление с выбором параметра в виде теста, получаются следующие вероятности из всего набора данных (рисунок 30).

```
> head(cbind(p,train))
      10715      10716      10718      10719      10720      10721      10722      10724      10728       10765      10907       10908      10909      10910      10938       27171 lab_test status age sex
1 0.09692937 0.06504536 0.04369857 0.03345878 0.04772413 0.11235554 0.25902745 0.12138942 0.20095631 2.025482e-04 0.003029146 0.0036276450 0.001618628 0.008245257 0.0011372371 0.0015546082    10722      1  22   F
2 0.16959729 0.10345417 0.22928792 0.14162271 0.13332800 0.08266686 0.01545339 0.04955807 0.02517053 5.831073e-05 0.028102883 0.0009334389 0.001812637 0.016595126 0.0009660275 0.0013926440    10721      0  21   F
3 0.09592870 0.06527386 0.04413339 0.03416703 0.04832042 0.11232796 0.25956427 0.12093385 0.19993028 2.008522e-04 0.003075890 0.0035451755 0.001634725 0.008307670 0.0011067145 0.0015492247    10724      1  21   F
4 0.17921612 0.11294944 0.26453049 0.10791136 0.09880423 0.08574507 0.01329591 0.07539962 0.03672055 6.832216e-05 0.013577859 0.0001368539 0.001309953 0.009173933 0.0003632789 0.0007970109    10716      0  46   M
6 0.17921612 0.11294944 0.26453049 0.10791136 0.09880423 0.08574507 0.01329591 0.07539962 0.03672055 6.832216e-05 0.013577859 0.0001368539 0.001309953 0.009173933 0.0003632789 0.0007970109    10718      0  46   M
7 0.10171618 0.06395109 0.04156956 0.03007627 0.04479866 0.11246652 0.25656521 0.12404164 0.20578965 2.087128e-04 0.002777245 0.0037938349 0.001510636 0.007875830 0.0012765958 0.0015823729    10724      1  27   F
>
```

Рисунок 30 – Результаты вычисления по тестам

На рисунке 30 отображены данные, по которым можно сказать, что результаты не изменились, то есть классическая теория влияния гемоглобина подтвердилась, и является одним из важных компонентов исследования для дальнейшей диагностики анемии. Но было определено, что при классификации данных был классифирован тест «Гематокрит», который при понижениях указывает на «на дефицит железа, витамина B12 и фолиевой кислоты, заболевания почек или костного мозга, таких как лейкемия, лимфома, множественная миелома».

В результате анализа данных с помощью алгоритма Байеса по патологичным данным были классифицированы тесты, которые также рекомендуются в диагностических сценариях медицинских протоколов.

Алгоритм деревья решений. Дерево решений – это графическое отображение, который представляет варианты и их результаты в форме дерева. Узлы на графике представляют события или варианты, а края графика представляют правила или условия принятия решения. В данном анализе будут использованы тот же набор данных надо которыми производили байесовском классификацию. В качестве платформы будет также использован R-studio.

Для анализа данных и построения деревья решений было отобрано 10 тыс. записей с нормальными и патологичными значениями. Было определено 3 параметра, такие как:

– код теста;
– статусы (0 понижение, 1 повышение, 2 норма);
– значение результатов.

Сформированный набор данных был импортирован в среду анализа:

```
> data <- read.csv(file.choose(),header = TRUE,sep=';')
> str(data)
'data.frame':	10000 obs. of  3 variables:
 $ lab_test: int  10728 10716 10715 10718 10719 10720 10721 10724 10722 10907 ...
 $ status  : int  2 2 2 2 2 2 2 2 2 2 ...
 $ value   : num  4.17 4.88 142 41.5 85 ...
```

Далее набор данных был предварительно обработан, для построения модели:

```
data_tree <- read.csv(file.choose(),header = TRUE,sep=';')
str(data_tree)
data_tree$lab_test <- as.factor(data_tree$lab_test)
## data partition
set.seed(1234)
ind <- sample(2,
              nrow(data_tree),
              replace=TRUE,
              prob=c(0.8,0.2)
              )
train <-data_tree[ind==1,]
test <-data_tree[ind==2,]
##decision tree
library(party)
tree          <-ctree(value~.,train,controls=ctree_control(mincriterion=0.9999, minsplit = 200))
tree
```

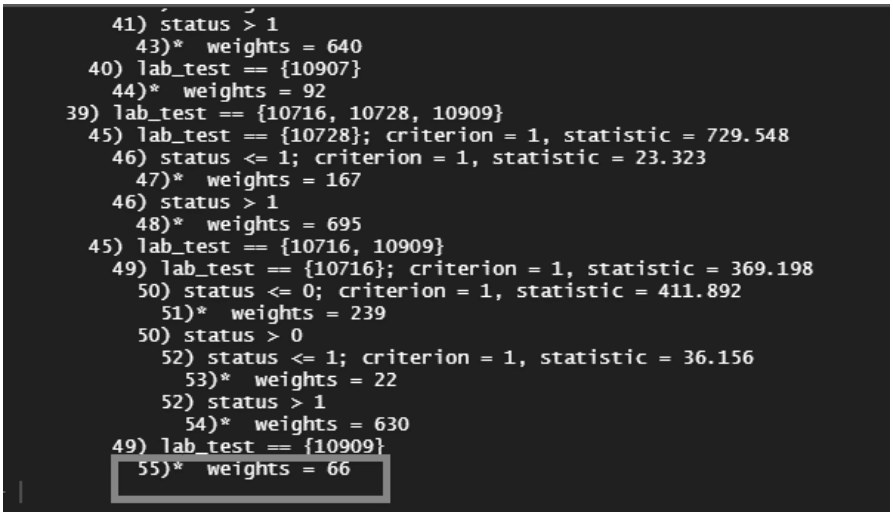


Рисунок 31 – Дерево решений

И как видно на рисунке 31, было построено дерево, и оно состояло из 55 узлов, из расчетами веса и критерии. Изначально намеренно было получено меньше данных, чтобы было эффективно визуализировать результаты построения дерево.

В результате визуализации мы получили распределения наших тестов по веткам деревья (рисунок 32).

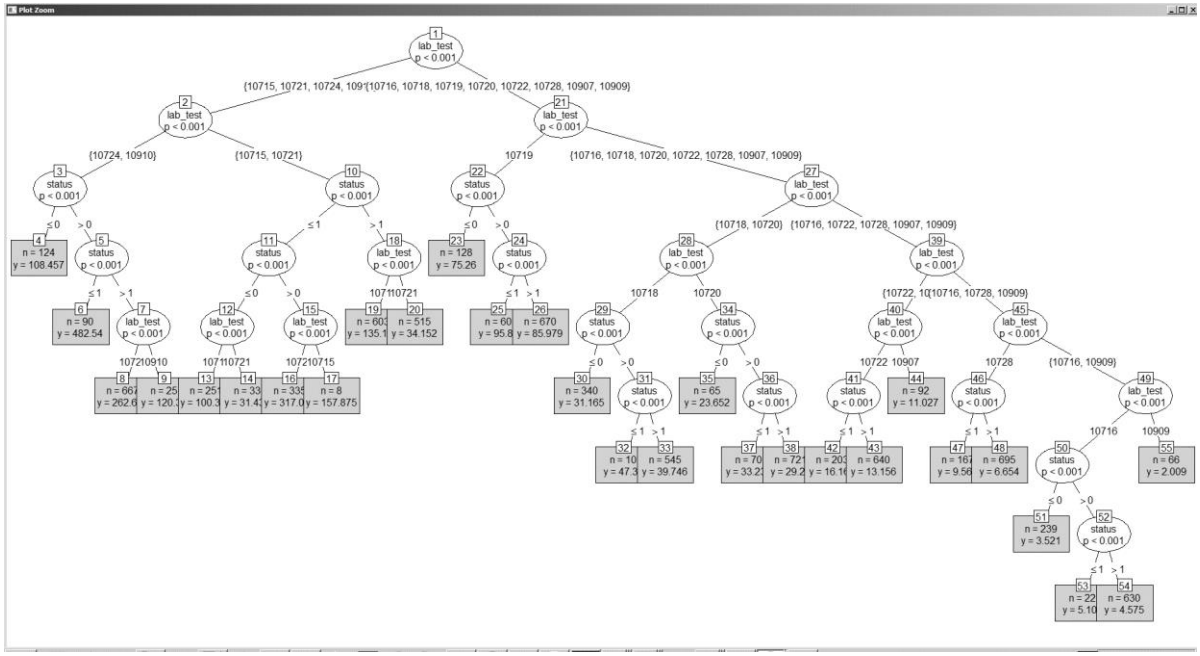


Рисунок 32 – Визуализация дерево решений

Далее визуализируем оптимизированное дерево с помощью библиотеки «rpart» (рисунок 33).

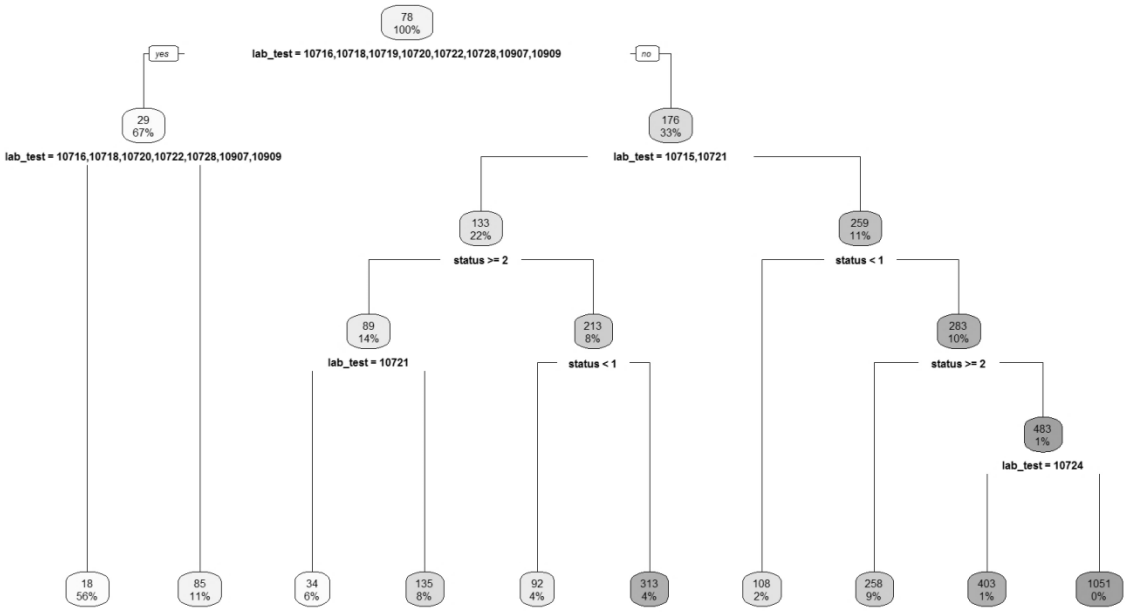


Рисунок 33 – Визуализация оптимизированного дерева решений

Здесь мы получаем логическое распределения тестов по статусу и его значений. В результате построения, оптимизированного дерево, видим, веса распределения тестов по нормальным и патологичным значениям. В том числе производится распределение по повышенным и пониженным тестам. Дерево решений помогает нам визуально увидеть группу тестов и их число необходимых для дальнейшего анализа.

Метод TF-IDF. Данный статистический метод является так же популярным при анализе документов, который можно применить и для анализа данных с помощью применения языка SQL.

TF (term frequency – частота слова) – отношение числа вхождений некоторого слова к общему числу слов документа. Таким образом, оценивается важность слова в пределах одного документа.

$$tf(t,d) = \frac{n_t}{\sum_k n_k}$$

где n_t является числом вхождения слово t в документ, а в знаменателе общее число слов в документе.

IDF (inverse document frequency – обратная частота документа) – инверсия частоты, с которой некоторое слово встречается в документах коллекции. Основоположником данной концепции является Карен Спарк Джонс [81]. Учёт IDF уменьшает вес широкоупотребительных слов. Для каждого уникального слова в пределах конкретной коллекции документов существует только одно значение IDF.

$$idf(t,D) = log\frac{|D|}{|\{d_i \in D|\ t \in d_i\}|}$$

где |D| – число документов в коллекции;
$|\{d_i \in D \mid t \in d_i\}|$ – число документов из коллекции D, в которых встречается t (когда $n_t \neq 0$).

Выбор основания логарифма в формуле не имеет значения, поскольку изменение основания приводит к изменению веса каждого слова на постоянный множитель, что не влияет на соотношение весов.

Таким образом, мера TF-IDF является произведением двух сомножителей:

$$tf - idf(t,d,D) = tf(t,d) \times idf(t,D)$$

Большой вес в TF-IDF получат слова с высокой частотой в пределах конкретного документа и с низкой частотой употреблений в других документах.

При использовании данного метода к данным размещенных в базе данных увеличивает скорость обработки информации, так как анализ будет

производиться по структурированным данным. В качестве «слово» будет рассматриваться лабораторный тест, а в качестве «коллекции документов» будет рассматривать единая база данных с результатами лабораторных исследований, который так же может быть разделен в коллекцию по определенным метрикам выборки. Анализ данных будет производиться по единой коллекции, то есть по набору данных в 7 млн. 600 тысяч результатов лабораторных исследований.

Произведем анализ TF всех данных. Анализ производился с использованием следующих скриптов:

```
--Переменные для расчета TF
declare @count_all integer
--Получаем количество записей в коллекции
select @count_all=count(*) from data_table where status<>'norm'
--Используем временную таблицу для облегчения обращения к данным
IF OBJECT_ID(N'tempdb..##temp') IS NOT null drop table ##temp
--Вычисляем количество тестов в коллекции
select test_mname, count(*) as count_
into ##temp
from data_table
where status<>'norm'
group by test_mname
order by count_ desc
--Производим расчет TF с ранжировкой от высокого к низкому
select test_mname,count_,1.0*count_/@count_all as tf from ##temp
order by count_ desc
```

Results | Messages

	test_mname	count_	tf
1	Средний объем эритроцитов	282642	0.036829183976
2	Гемоглобин	268914	0.035040380339
3	Цветной показатель	268835	0.035030086379
4	Гематокрит	268700	0.035012495434
5	Лейкоциты	268456	0.034980701430
6	Тромбоциты	268312	0.034961937755
7	Эритроциты	267712	0.034883755778
8	Среднее содержание Hb в одном эритроците	239262	0.031176627028
9	Средняя концентрация Hb в эритроцитах	239238	0.031173499749
10	Средний объем тромбоцитов	238752	0.031110172348
11	Тромбокрит	238511	0.031078769254
12	Расчетная ширина распределения эритроцито...	191706	0.024979923519
13	Расчетная ширина распределения эритроцито...	191682	0.024976796240
14	Базофилы (относит. количество)	173285	0.022579606517
15	Базофилы (абс.количество)	173154	0.022562536786
16	Распределение тромбоцитов по объёму (стан...	173094	0.022554718588
17	Лимфоциты (абс.количество)	173056	0.022549767063
18	Эозинофилы (относит. количество)	173035	0.022547030694
19	Лимфоциты (относит. количество)	173018	0.022544815538
20	Эозинофилы (абс.количество)	172957	0.022536867037
21	Моноциты (относит. количество)	172908	0.022530482175
22	Нейтрофилы (абс. количество)	172844	0.022522142764
23	Нейтрофилы (относит. количество)	172461	0.022472236602
24	Моноциты (абс.количество)	132102	0.017213325909

Query executed successfully. | KKK\SSDB (12.0 SP3) | KKK\KKuanysh (53) | phd_analyse | 00:02:43 | 406 rows

Рисунок 34 – Результаты вычисления TF

В результате вычисление, было определено 406 уникальных тестов от наибольшего к наименьшему (рисунок 34).

На рисунке 34 можно обратить внимание, что в выбранной коллекции данных из 7 млн. 600 тысяч записей 11 тестов имеют частоту более 0,03. Так как у нас одна коллекция то IDF будет равен единице. В результате за основу будет использоваться только TF.

Произведем анализ тестов в коллекции данных на патологичные результаты, то есть, где есть отклонения от нормативных величин. В анализе данных используется только патологичные результаты 1 млн. 898 тыс. записей. В результате было определено 348 уникальных тестов, и TF у 2-х тестов имели частоту 0,05 (Гематокрит, Гемоглобин) (рисунок 35), которые так же входили в перечень анализа по всем данным.

Results | Messages

	test_mname	count_	tf
1	Гематокрит	107338	0.056540910750
2	Гемоглобин	106076	0.055876144969
3	Тромбокрит	80332	0.042315344448
4	Лимфоциты (относит. количество)	78771	0.041493078692
5	Цветной показатель	77813	0.040988446665
6	Лейкоциты	74460	0.039222234571
7	Тромбоциты	66313	0.034930755320
8	Эритроциты	64235	0.033836156832
9	Средний объем эритроцитов	61391	0.032338063424
10	Расчетная ширина распределения эритроцитов по об...	56195	0.029601040447
11	Лимфоциты (абс.количество)	51519	0.027137930471
12	Расчетная ширина распределения эритроцитов по об...	51415	0.027083147871
13	Средняя концентрация Hb в эритроцитах	46571	0.024531542925
14	Относительное распределение тромбоцитов по объёму	44719	0.023555991241
15	Среднее содержание Hb в одном эритроците	44035	0.023195690294
16	Нейтрофилы (относит. количество)	42259	0.022260172048
17	Глюкоза (сахар крови)	34366	0.018102488762
18	Скорость оседания эритроцитов (СОЭ)	34325	0.018080891776
19	Моноциты (абс,количество)	31910	0.016808776593
20	Базофилы (абс.количество)	31796	0.016748726436
21	Базофилы (относит. количество)	30863	0.016257263303
22	Нейтрофилы (абс. количество)	29768	0.015680465736
23	Эозинофилы (абс.количество)	29674	0.015630950694
24	Холестерин общий	24495	0.012902882565

Query execute... | KKK\SSDB (12.0 SP3) | KKK\KKuanysh (58) | phd_analyse | 00:00:02 | 348 rows

Рисунок 35 – Результаты вычисления TF по патологичным результатам

Результаты анализов показали, что согласно методу, TF-IDF целесообразно обучать систему по данным исследованиям указанных на рисунке 35, так как они имеют вес частоты больше, чем другие анализы.

С помощью TF/IDF метода был определен перечень более патологичных тестов, в дальнейшем над которыми произведем логистический регрессионный анализ.

Логистический регрессионный анализ. Логистическая регрессия — это статистический метод анализа, который используется для предсказания вероятности наступления определенного события на основе набора входных переменных. Этот метод обычно используется в машинном обучении для классификации двух категорий.

Формула логистической регрессии выглядит следующим образом:

$$P(z) = \frac{1}{1 + e^{-z}}$$

где P – вероятность наступления события;

z – линейная комбинация входных переменных и соответствующих весов модели:

$$z = w_0 + w_1x_1 + w_2x_2 + ... + w_nx_n$$

где w_0, w_1, w_2, ..., w_n – веса модели;

x_1, x_2, ..., x_n – входные переменные.

Таким образом, логистическая регрессия применяет логистическую функцию к линейной комбинации входных переменных и весов модели, чтобы получить вероятность наступления события.

Обучение модели логистической регрессии заключается в настройке весов модели (w_0, w_1, w_2, ..., w_n) для минимизации функции потерь (loss function), которая измеряет разницу между предсказанной вероятностью и реальной меткой класса. Обычно используется функция кросс-энтропии:

$$L = \frac{-1}{N}\sum_{i=1}^{N} y_i * log(P_i) + (1 - y_i) * log(1 - P_i)$$

где N – количество примеров в обучающей выборке;

y_i – истинная метка класса (0 или 1) для i-го примера;

P_i – предсказанная вероятность для i-го примера.

Для настройки параметров модели применяется алгоритм градиентного спуска (gradient descent), который осуществляет обновление весов модели в направлении, обеспечивающем наиболее быстрое уменьшение значения функции потерь. На рисунке 34 представлен рейтинг различных лабораторных тестов на основе их оценки TF. Тесты перечислены в порядке убывания их оценки TF, причем тест с самой высокой оценкой стоит первым, а тест с самой низкой оценкой - последним. Каждая строка таблицы представляет один лабораторный тест и включает следующую информацию:

1. Название теста: Название лабораторного теста.

2. Количество: Количество раз, когда проводился тест.

3. TF: показатель частоты терминов, который является мерой важности или актуальности каждого теста.

Цель ранжирования - представить четкое и ясное понимание того, какие лабораторные тесты считаются наиболее значимыми и актуальными, исходя из их оценки по TF. Количественная колонка в таблице предоставляет дополнительную информацию, указывая частоту, с которой проводилось каждое испытание. Для проведения логистической регрессии были выбраны 49 тестов для первой итерации. Была построена модель для изучения взаимосвязи между возрастом, отклонением (выше или ниже нормы) (рисунок 36).

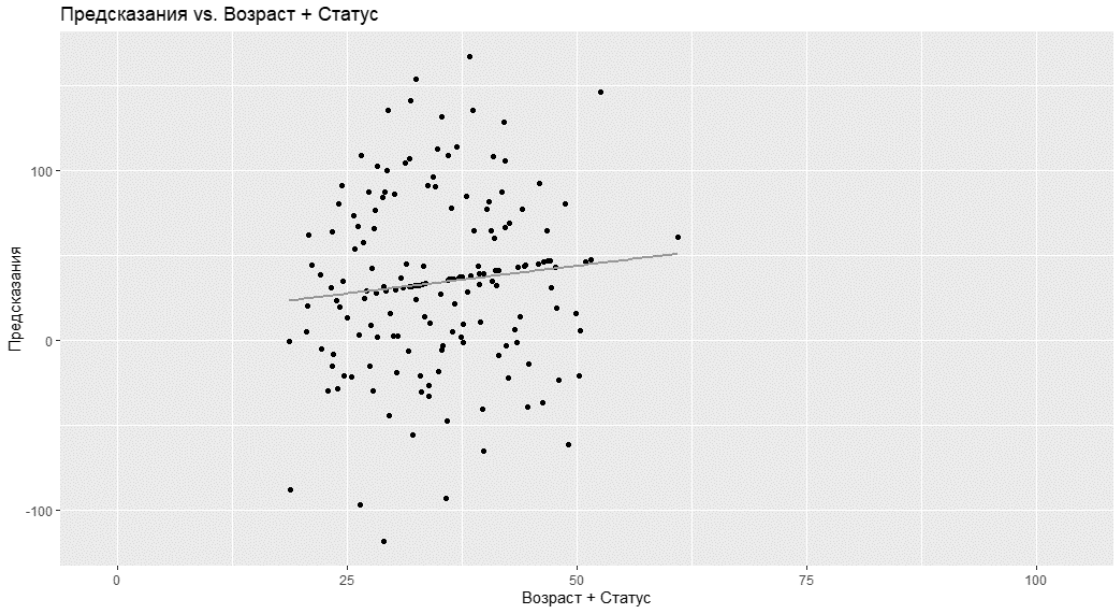


Рисунок 36 – Анализ без обработки данных ранжировки

Результаты, представленные на рисунке 35, имели плохую связь между патологическими тестами. Для улучшения модели из первоначальных 49 тестов были удалены тесты с отрицательными результатами, и затем связь была установлена с помощью следующих тестов: TSH (гормон щитовидной железы), эозинофилы (абсолютное количество), нейтрофилы (относительное количество), базофилы (абсолютное количество), креатинин, лимфоциты (относительное количество), AST (аспартатаминотрансфераза), глюкоза (сахар крови) и эозинофилы (абсолютное количество) (рисунок 37).

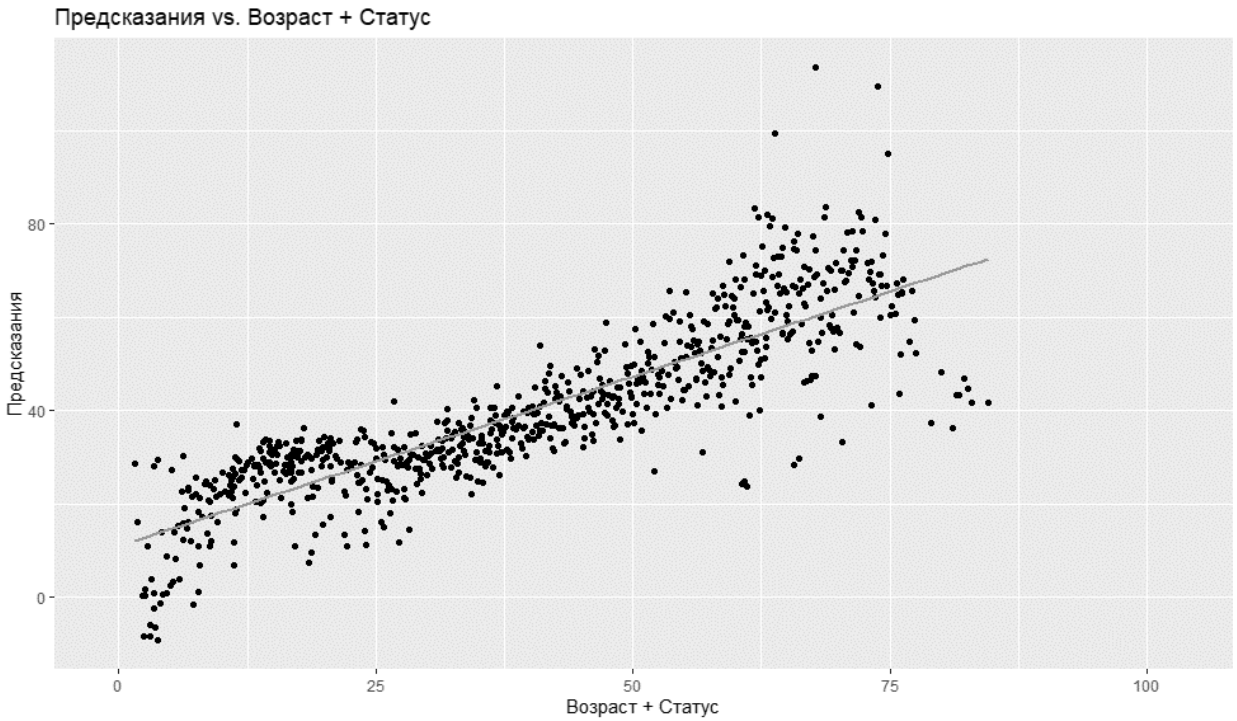


Рисунок 37 – Анализ данных после очистки выбросов

В результате, как показано на рисунке 37, модель продемонстрировала хорошую взаимосвязь между возрастом пациента и патологическими параметрами, и тестами.

Следующим шагом было определение того, какой патологический тест связан с другими патологическими тестами. Были проведены многочисленные комбинации тестов, и было установлено, что тест TSH (тиреотропный гормон) в сочетании с возрастом связан со следующими тестами: Гематокрит, Эритроциты, Эозинофилы (абсолютное количество), Моноциты (абсолютное количество), Средняя концентрация Hb в эритроцитах, Нейтрофилы (относительное количество), Базофилы (абсолютное количество), Лимфоциты (относительное количество), Лимфоциты (относительное число 1), АСТ (аспартатаминотрансфераза), Глюкоза (сахар крови), Среднее содержание Hb на эритроцит, Моноциты (абсолютное количество), Моноциты (относительное количество), ХГЧ (хорионический гонадотропин), Эозинофилы (относительное количество), Общий билирубин, Расчетное стандартное отклонение ширины распределения объема эритроцитов, Средний объем эритроцитов, Аланин-аминотрансфераза (АЛТ), Эозинофилы (абсолютное количество), Нейтрофилы (абсолютное количество), Лейкоциты, Общий холестерин, Гликозилированный гемоглобин, Моноциты (относительное количество), Тромбоциты, Базофилы (относительное количество 1), Мочевина, Общий белок и Скорость оседания эритроцитов (скорость оседания на анализаторе), как показано на рисунке 38.

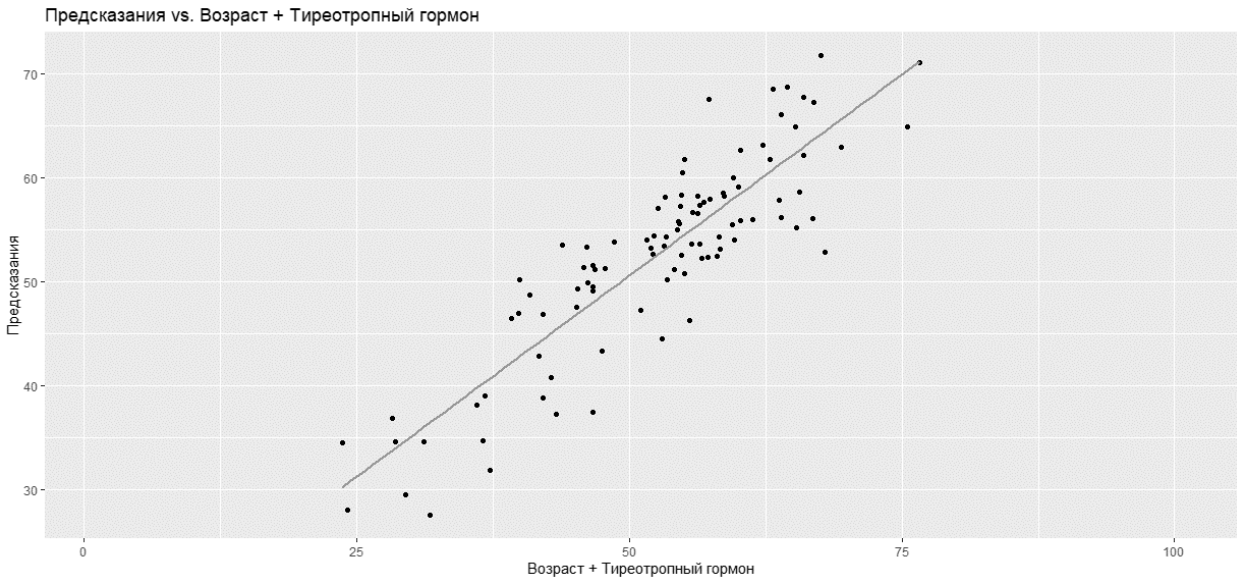


Рисунок 38 – Ассоциация между TSH (тиреотропный гормон) + возраст с другими тестами

В данном исследовании мы провели логистический регрессионный анализ для прогнозирования зависимой переменной. Целью анализа было понять взаимосвязь между независимыми переменными и зависимой переменной и сделать прогнозы относительно зависимой переменной на основе независимых переменных.

В следующем рисунке 39 представлена сводная статистика прогнозируемых значений зависимой переменной.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
31.75	46.38	53.62	52.47	57.27	69.45

Рисунок 39 – Сводная статистика прогнозируемых значений

Эти сводные статистические данные дают нам представление о распределении прогнозируемых значений. Минимальное прогнозируемое значение составляет 31,75, первый квартиль - 46,38, медиана - 53,62, среднее - 52,47, третий квартиль - 57,27, а максимальное прогнозируемое значение - 69,45.

Чтобы лучше понять распределение предсказанных значений, мы построили гистограмму и график плотности. Гистограмма показывает, что прогнозируемые значения распределены вокруг среднего значения с небольшим положительным перекосом. График плотности подтверждает это распределение: наибольшая плотность предсказанных значений находится вблизи среднего значения. Анализ предсказанных значений в логистической регрессии показывает, что предсказанные значения распределены вокруг среднего значения с небольшим положительным перекосом. Также существует умеренная положительная связь между прогнозируемыми значениями и одной из независимых переменных. Эти результаты могут быть использованы для

прогнозирования зависимой переменной и для дальнейшего анализа и принятия решений.

Обучение Нейронной сети. Следующим этапом производилась обучение нейронной сети. В качестве данных были отобран набор данных с параметрами необходимых для определения патологии щитовидной железы. Это тесты тиреотропный гормон (ТТГ) в комбинации с трийодтиронина (Т3) и тетрайодтиронина (Т4). ТТГ особенно важен для поддержания гормонального баланса, поскольку уровень TSH в крови тесно связан с уровнем Т3 и Т4. Если уровни Т3 и Т4 низкие, гипофиз увеличивает выработку ТТГ, что стимулирует щитовидную железу к усилению синтеза и высвобождению тиреоидных гормонов. Если уровни Т3 и Т4 слишком высокие, гипофиз снижает выработку ТТГ, что приводит к уменьшению активности щитовидной железы. Этот процесс регулирования называется обратной связью. Если уровень ТТГ понижен, это может указывать на несколько состояний. Самые распространенные причины пониженного ТТГ включают:

– гипертиреоз: Низкий уровень ТТГ может быть свидетельством гипертиреоза, состояния, при котором щитовидная железа производит избыточное количество тиреоидных гормонов Т3 и Т4. Гипертиреоз может быть вызван различными причинами, такими как Болезнь Грейвса, токсический аутоиммунный тиреоидит или токсическую аденому;

– вторичный гипотиреоз: Редко пониженный ТТГ может указывать на вторичный гипотиреоз, когда недостаток тиреоидных гормонов вызван нарушением в гипофизе или гипоталамусе, которые регулируют выработку ТТГ. Это может быть вызвано опухолями, радиотерапией, инфекциями или травмами в области головного мозга;

– недавнее лечение гипертиреоза: Пациенты, которые недавно прошли лечение гипертиреоза (радиоактивным йодом, антитиреоидными препаратами или хирургическим удалением части щитовидной железы), могут временно иметь пониженные уровни ТТГ, пока организм не адаптируется к изменениям;

– прием лекарств: Некоторые лекарства, такие как левотироксин (синтетический аналог Т4), могут вызывать снижение уровня ТТГ при переизбытке дозировки или индивидуальной чувствительности.

Набор выбранных данных состоял из следующих параметров:

– возраст;
– пол;
– значение ТТГ;
– значение Т3;
– значение Т4;
– признак патологии.

Модель нейронной сети был построен по 3 скрытым слоям, и в результате обучения модели были получены следующие результаты (рисунок 40).

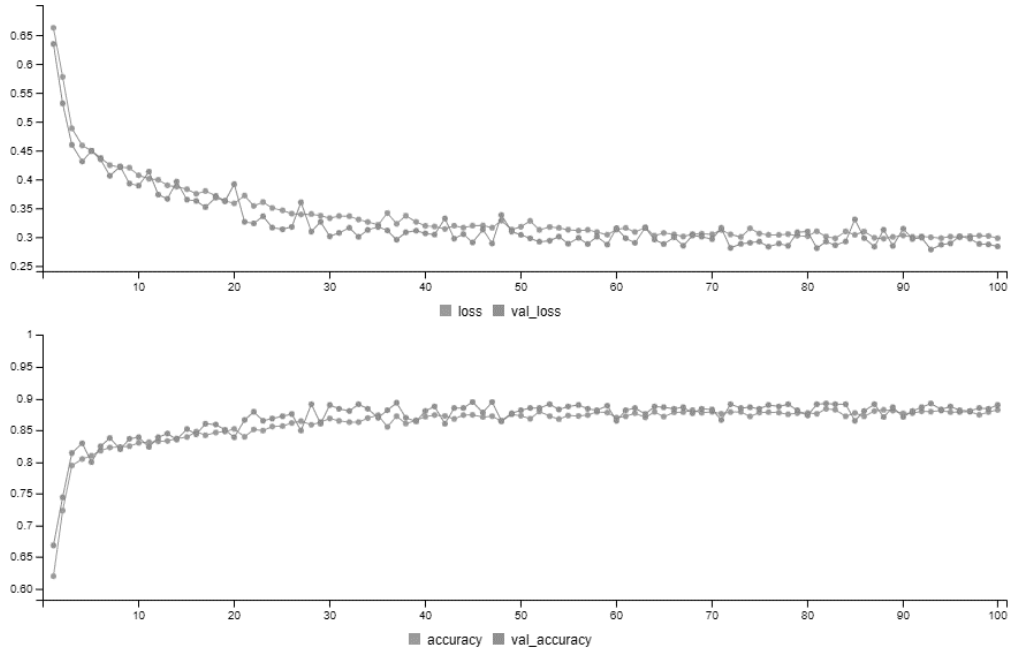


Рисунок 40 – График обучения нейронной сети

История обучения модели демонстрирует следующие показатели:

– потери при обучении: 0,2732;
– точность обучения: 0,8949;
– потери при проверке: 0,338;
– точность валидации: 0,8612.

Эти значения показывают, что производительность модели на тренировочных данных лучше, чем производительность на валидных данных. Точность обучения выше, чем точность проверки, а потери при обучении ниже, чем потери при проверке. Это говорит о том, что модель может слишком хорошо подходить к тренировочным данным, то есть быть слишком сложной и плохо обобщаться на новые, невидимые данные.

Модель показала, что необходимо использовать другие методы машинного обучения или использовать ансамблевый подход путем усиления нейронной сети с использованием методов градиентного бустинга (GB, gradient boosting), случайного леса (RF, random forest) и метод опорных векторов (SVM, support vector machine). Ансамблевый метод в машинном обучении относится к стратегии, когда несколько моделей обучаются решению одной задачи, после чего их выходные данные объединяются с намерением повысить эффективность. Основной посыл состоит в том, что вывод, полученный от множества моделей, будет более точным по сравнению с выводом единственной модели.

Для построения ансамблевого метода необходимо сравнить предлагаемые методы между собой, чтобы определить наилучший из них. Сравнение начинается с построения трех моделей на тех же данных, после строится матрица ошибок и производится сравнение производительности каждой из модели (рисунок 41).

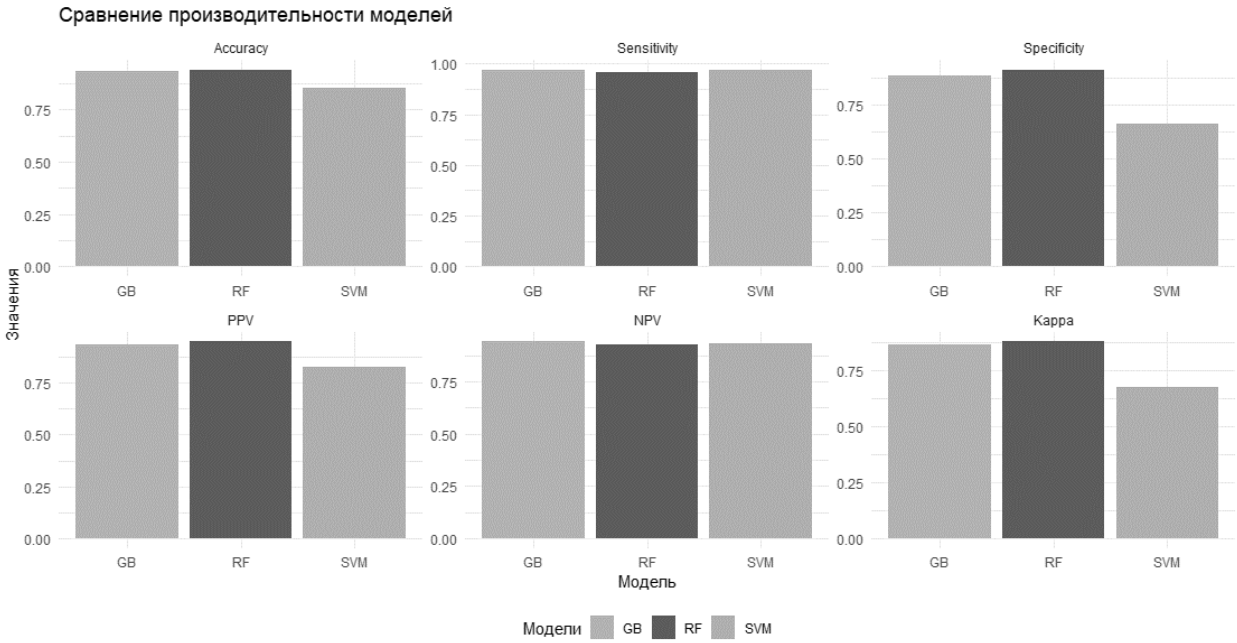


Рисунок 41 – Сравнение производительности моделей

Из рисунка 41 можно выделить следующую таблицу результатов сравнительного анализа (таблица 14).

Таблица 14 – Сравнительный анализ моделей машинного обучения

Модель	Точность	Чувствительность	Специфичность	PPV	NPV	Каппа
SVM	0,8556	0,9709	0,6619	0,8283	0,9312	0,6722
GB	0,9369	0,9673	0,8859	0,9344	0,9416	0,8635
RF	0,9422	0,9588	0,9145	0,9496	0,9296	0,8761

Таблица 14 содержит метрики производительности для трех различных моделей: метод опорных векторов (SVM), градиентный бустинг (GB) и случайный лес (RF). Метрики включают точность (Accuracy), чувствительность (Sensitivity), специфичность (Specificity), положительную прогностическую ценность (PPV), отрицательную прогностическую ценность (NPV) и коэффициент Каппа (Kappa). Из таблицы видно, что модель случайного леса имеет наивысшую точность (0,9422) и коэффициент Каппа (0,8761), что указывает на ее превосходную производительность в данном сравнении. Модель градиентного бустинга также показывает хорошую производительность с точностью 0,9369 и коэффициентом Каппа 0,8635. Модель метода опорных векторов имеет наименьшую производительность с точки зрения точности (0,8556) и коэффициента Каппа (0,6722), но обладает наивысшей чувствительностью (0,9709), что может быть полезно в определенных приложениях, где высокая вероятность истинно положительных результатов имеет большое значение.

Сравнительная таблица показала, что для представленных данных более подходящим являются методы машинного обучения случайный лес и

градиентный бустинг. Следовательно, используя данные методы можно построить ансамбль [GB+NN] и [RF+NN]. При построении ансамблевой модели производим обучение модели по GB и RF. Далее строим нейронную сеть и анализируем производительность.

Ансамблевый метод градиентного бустинга и нейронной сети (GB+NN). В данном подходе обучение данных происходит в начале методом градиентного бустинга:

```
gbm_model <- train (
  status ~.,
  data = train_set,
  method = «gbm»,
  trControl = trainControl (method = «cv», number = 5),
  tuneLength = 5
)
//Подготовка данных для нейронной сети
train_set$gbm_preds <- predict (gbm_model, train_set)
test_set$gbm_preds <- predict (gbm_model, test_set)
```

Следующим этапом производится построение модели нейронной сети:

```
model <- keras_model_sequential() %>%
  layer_dense(units = 32, activation = «relu», input_shape = ncol(train_set) - 1) %>%
  layer_dense(units = 16, activation = «relu») %>%
  layer_dense(units = 1, activation = «sigmoid»)

model %>% compile(
  optimizer = optimizer_adam(learning_rate = 0.001),
  loss = «binary_crossentropy»,
  metrics = c(«accuracy»)
)
```

После построение модели обучаем нейронную сеть:

```
train_set$status <- as.numeric(train_set$status) - 1
test_set$status <- as.numeric(test_set$status) - 1

//Обучаем нейронную сеть
history <- model %>% fit(
  as.matrix(train_set[, -ncol(train_set)]), train_set$status,
  epochs = 100,
  batch_size = 32,
  validation_split = 0.1
)
```

По завершению обучения строим прогнозы проблем и классы:

```
predicted_probs <- predict (model, as.matrix(test_set[, -ncol(test_set)]))
predicted_classes <- ifelse (predicted_probs > 0.5, 1, 0)
```

Строим матрицу ошибок и выводим результаты построения нейронной сети:

true_labels <- factor (test_set$status, levels = c(0, 1))
nn_preds <- factor (predicted_classes, levels = c(0, 1))

cm <- confusionMatrix (nn_preds, true_labels)
print(cm)

В результате ансамбль моделей градиент бустинг (GB) и нейронная сеть (NN) демонстрируют значительное улучшение производительности по сравнению с отдельными моделями, чему следуют следующие результаты:

1. Точность: (TP+TN)/(TP+TN+FP+FN) = (296+800)/ (296+800+25+195) = 0,8328, что означает, что модель правильно классифицировала 83,28% случаев.

2. 95% ДИ: (0,8115, 0,8526) представляет собой доверительный интервал для точности; мы можем быть на 95% уверены, что истинная точность лежит между 81,15 и 85,26%.

3. Коэффициент отсутствия информации (NIR): 0,6269 - базовая точность, если мы всегда предсказываем класс большинства (в данном случае класс 0).

4. P-Value [Acc> NIR]: <2.2e-16, вероятность соблюдения точности модели, если бы модель не имела информации. Очень маленькое p-значение указывает на то, что модель значительно лучше, чем случайное угадывание.

5. Kappa: 0,6157, мера согласия между предсказанными и фактическими значениями с учетом случайности. Значение, близкое к 0, указывает на плохое согласие, а значение, близкое к 1, указывает на идеальное согласие.

Ансамблевый метод случайного леса и нейронной сети (RF+NN). Произведем аналогичное построение нейронной сети на базе обученных данных с помощью модели случайный лес. В результате мы получили показатели хуже, чем модель GB+NN:

1. Точность: (TP+TN)/(TP+TN+FP+FN) = (128+815)/ (128+815+10+363) = 0,7166, что означает, что модель правильно классифицировала 71,66% случаев.

2. 95% CI: (0,6914, 0,7408) представляет собой доверительный интервал для точности; мы можем быть на 95% уверены, что истинная точность лежит между 69,14 и 74,08%.

3. Коэффициент отсутствия информации (NIR): 0,6269 - базовая точность, если мы всегда предсказываем класс большинства (в данном случае класс 0).

4. P-Value [Acc> NIR]: 4.289e-12, вероятность соблюдения точности модели, если бы модель не имела информации. Очень маленькое p-значение указывает на то, что модель значительно лучше, чем случайное угадывание.

5. Kappa: 0,2909, мера согласия между предсказанными и фактическими значениями с учетом случайности. Значение, близкое к 0, указывает на плохое согласие, а значение, близкое к 1, указывает на идеальное согласие.

6. P-значение теста Макнемара: <2,2e-16, мера разницы между FP и FN. Небольшое значение p-value указывает на значительную разницу между ними.

Наше исследование показало, что методы машинного обучения для набора данных по выявлению отклонению при диагностике щитовидной

железы могут работать как отдельная модель, так и с применением методики ансамбля. В результате согласно рисунку 40 можно сказать, что высокие показатели были у отдельных моделей как случайный лес и градиентный бустинг, и методика ансамбль между градиентного бустинга и нейронной сети.

3.2 Разработка модели искусственного интеллекта по лабораторной диагностике

3.2.1 Логическая модель искусственного интеллекта по интерпретации результатов

Логическая модель искусственного интеллекта по интерпретации результатов лабораторных исследований представляет описание алгоритмов и структурных схем работы искусственного интеллекта.

Основной логической моделью является описание участков информационной системы, производящее обучение системы для конечной интерпретации результатов лабораторных исследований (рисунок 42).

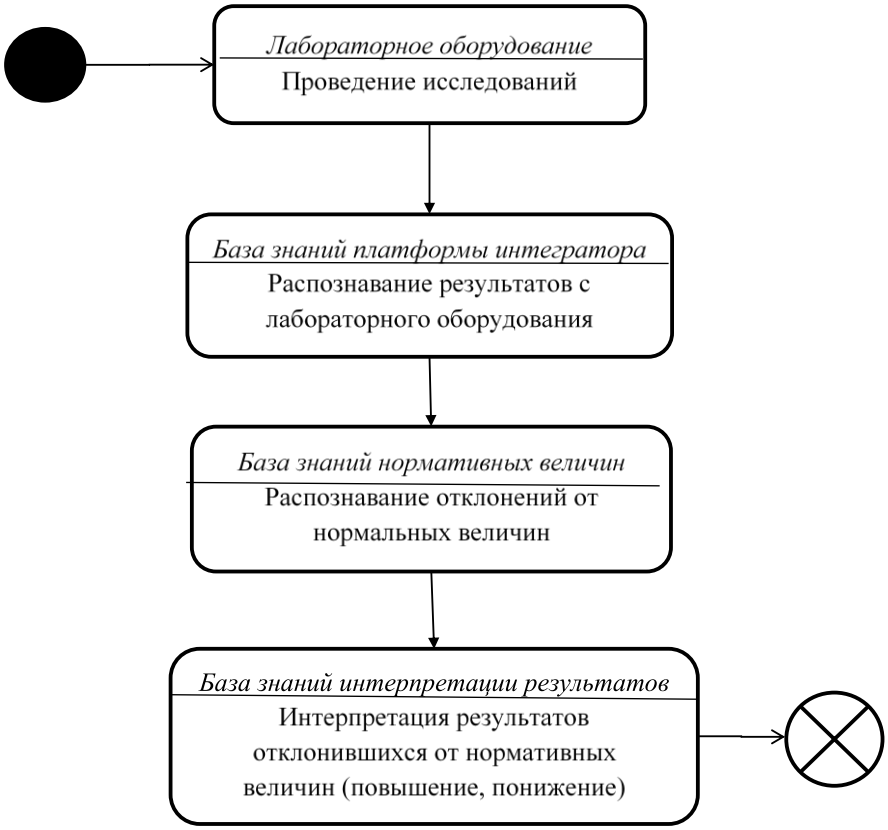


Рисунок 42 – Логическая модель интерпретации результатов

На рисунке 42 отражены 4 блока, которые проходят обучение с учителем и непосредственно взаимодействуют от получения результатов с лабораторного оборудования до интерпретации. Все взаимодействие происходят без участия человека, и система производит интерпретацию в режиме реального времени.

Взаимодействие с большинством лабораторным оборудованием происходит путем использования международных протоколов обмена медицинскими данными, такие HL7 (на англ. health level 7 – «Седьмой уровень») и ASTM (англ. American Society for Testing and Materials – «Американское общество по испытанию материалов»). Но ряд современных лабораторных оборудований имеют свои локальные стандарты, и не следуют международным стандартам. В этих случаях производится обучение системы распознавать сторонние формы и стандарты, для взаимодействия с оборудованием. Технический взаимодействие происходит по следующим протоколам:

1. Протокол TCP/IP (Transmission Control Protocol/Internet Protocol) используется для установления связи между лабораторным анализатором и хостом соединения посредством IP-адресов. При этом важно учитывать роли участников в соединении:

TCP/IP сервер: Лабораторное оборудование действует в качестве сервера, и хост, действующий как клиент TCP/IP, устанавливает соединение с оборудованием через определенный порт, например, 5100 или 5050.

TCP/IP клиент: Лабораторное оборудование действует как клиент TCP/IP, а хост, выступающий в роли сервера, принимает подключения по определенному порту. При этом необходимо учитывать, что использование стандартных зарезервированных портов (например, 80 или 443) может вызвать конфликты, поэтому следует избегать их использования во избежание проблем при взаимодействии оборудования с хостом

2. Соединение по протоколу RS232 (Recommended Standard 232) осуществляется через классический COM-порт.

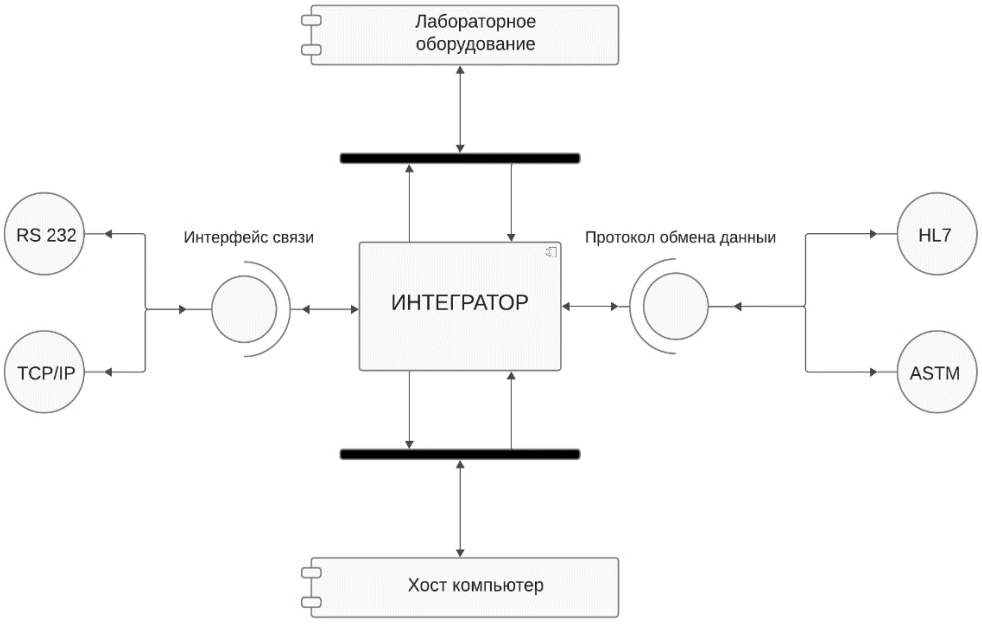


Рисунок 43 – Логическая модель взаимодействия с лабораторным оборудованием

В соответствии с рисунком 43, в результате модель взаимодействия с лабораторным анализатором имеет следующую логическую модель.

Согласно рисунку 43, получаем универсальную схему взаимодействия с анализаторами, которая будет обучаться для распознавания новых оборудовании путем добавления их в базу знаний модели.

Распознавание полученных результатов и системы производится путем обучения тестов системы с уникальными кодами тестов лабораторного оборудования. Это производится путем синхронизации один к одному кодов тестов системы с тестами анализатора (рисунок 44).

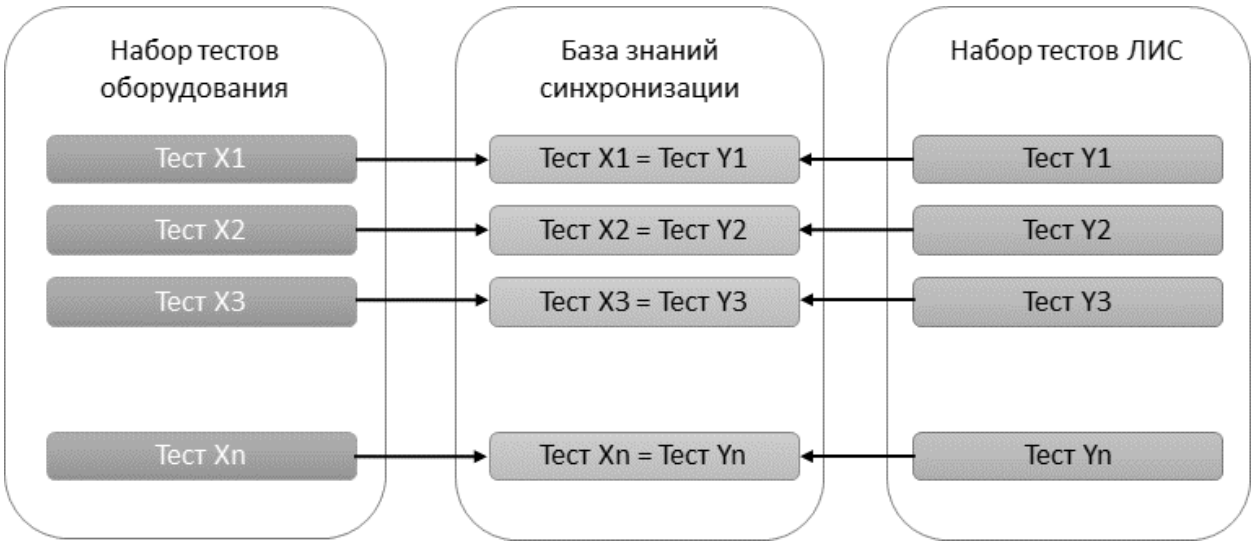


Рисунок 44 – Модель формирования базы знаний для взаимодействия с лабораторным оборудованием

Связь между тестами производится один к одному, чтобы система в дальнейшем быстро могла осуществлять распознавания при получении данных с анализатора.

Описание моделей распознавания отклонений и интерпретации результатов лабораторных исследований были описаны в подразделах 2.3 и 2.5.

Описываемая модель искусственного интеллекта по интерпретации результатов лабораторных исследований является комплексом, где каждые ключевые блоки обучаются с учителем и представляют собой цельный рабочий механизм (систему), которая функционирует и выполняет задачи без участия человека (рисунок 45).

Рисунок 45 – Алгоритм интерпретации результатов

В результате мы получаем комплексную модель, которая состоит из-под структур и является частью общего механизма интерпретации результатов, так как конечной целью работы является предварительное определение патологии.

3.2.2 Математическая модель искусственного интеллекта по интерпретации результатов

При разработке интеллектуальных информационных систем в медицинской деятельности, важно придерживаться общей концепции, которая может быть представлена в виде пирамиды с четырьмя уровнями:

Первый уровень: На этом уровне осуществляется создание информационных систем, которые включают в себя сбор, предварительную обработку и хранение первичных данных. Этот уровень может быть построен с использованием модульного подхода.

Второй уровень: Этот уровень представляет собой интегрированную базу данных, в которой общие данные хранятся в иерархически структурированной форме.

Третий уровень: На этом уровне осуществляется анализ данных с применением различных методов моделирования и прогнозирования.

Четвертый уровень: Самым важным уровнем являются интеллектуальные системы, которые принимают решения на основе анализа данных с предыдущих уровней.

Ключевым требованием при формировании уровней является гарантирование их хорошей связи, что критично для эффективной работы всей системы. Основой для структурирования и создания почти любой автоматизированной системы становится выбор схемы декомпозиции общесистемной задачи. Эта схема должна обеспечивать координацию локальных модулей, их согласование по целям, критериям, ограничениям и методам управления.

Функциональная декомпозиция становится основным видом разложения в этом контексте. Это связано с тем, что при исследовании и разработке информационных систем приоритетной задачей является анализ функционального предназначения системы. Исходя из общей функции системы, предназначенной для удовлетворения специфических потребностей путем выполнения определенных видов работ, необходимо проанализировать ее функциональную структуру. Функциональный (или системный) анализ позволяет разработать функциональную модель объекта, описывающую функции, исполняемые всеми элементами системы и включенными в ее структуру, с намерением достижения общей цели ее работы.

В процессе функциональной декомпозиции и усечения содержимого функциональных блоков появляется множество узкоспециализированных задач, способствующих решению как конкретных, так и глобальных проблем системы, при условии их синхронизированного и координационного выполнения. Одним из исходных этапов в этой области является группировка локальных задач в блоки, включающие в себя задачи, связанные и зависимые по своему характеру. Эти блоки создаются, принимая во внимание функциональные и целеполагающие аспекты, что значит, задачи, имеющие общую цель и способствующие выполнению функций для достижения этой цели, объединяются внутри блока. При группировке задач также учитывается их фокусировка на определенный объект или проблемную область. К тому же, необходимо учесть временные характеристики, чтобы гарантировать синхронизацию и периодическое выполнение задач.

Применение метода декомпозиции на практике действительно позволяет создать «дерево целей», выстраивая иерархию задач, которые система должна решить для достижения своей конечной цели. Это дерево, затем, трансформируется в «дерево функций», представляющее собой структуру, отражающую, как каждая функция служит достижению поставленных целей.

Предположим, когда цель X установлена для системы S и идентифицирован набор функций $\{F\}$, который должен быть выполнен для достижения этой цели, одним из следующих логичных шагов будет декомпозиция каждой функции в $\{F\}$, чтобы обеспечить более детальное и однозначное понимание того, как эти функции должны быть выполнены. Процесс декомпозиции показан на рисунке 46.

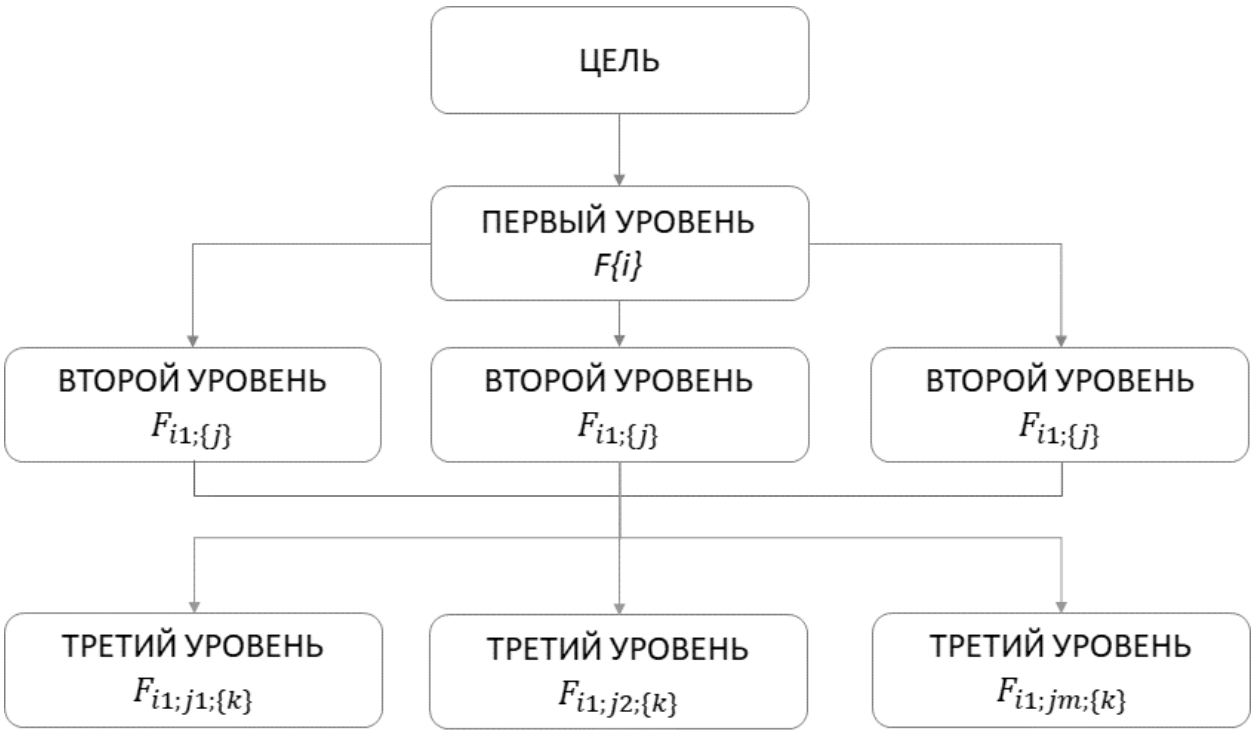


Рисунок 46 – Диаграмма декомпозиции функции

Учитывая, что внутри системы F могут существовать разные объекты, такие как i_1, i_2, ..., i_n, функции первого уровня относятся к функциям, связанным с этими конкретными объектами. Над каждым объектом из набора {i} могут выполняться различные операции, например, j_1, j_2, ..., j_n. С учетом этого функции второго уровня относятся к функциям, связанным с выполнением определенных операций, характерных для каждого из этих объектов. Например, функции второго уровня могут относиться к объекту i_1 при выполнении операции j_1, которая включает в себя ряд так называемых переходов, таких как k_1, k_2, ..., k_n, и так далее.

Следовательно, процесс разложения может продолжаться, пока это оправдано с точки зрения наиболее эффективного достижения заданной цели. Иными словами, этот процесс должен завершиться на самом низком уровне системы, на котором функции физического мониторинга состояния каждого из объектов системы являются конечными. Путем декомпозиции функций системы в общем виде можно представить, как:

$$Q^F = \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{l} \dots q_{i,j,k} \dots \qquad (6)$$

где Q^F – представляет собой общее количество элементарных функций, с помощью которых достигается главная цель X.

n – представляет количество i-х объектов в системе S, где i=1, n;

m – представляет количество j-х операций, которые выполняются над i-м объектом, где j=1, m;

l – представляет количество k-х переходов, на которые можно разделить j-ую операцию, где k=1, l;

$q_{i,j,k}$;… – обозначает количество функций управления, связанных с i-м объектом при прохождении им k-го перехода j-ой операции.

Если Q^F это построенная интеллектуальная система, то одним из функции является функция интерпретации результатов, которая зависит от Q^F. А функция интерпретации в теории реляционной алгебры описан в разделе 2.5, с помощью которого будет происходить выявление причин отклонений из базы знаний.

Предсказание интерпретации производится с использованием машинного обучения, а именно методов случайный лес, градиентный бустинг и нейронные сети с использованием ансамблевой методики.

Методы ансамбля – это методы машинного обучения, которые объединяют прогнозы нескольких моделей для улучшения общей производительности. В данном случае ансамбль состоит из модели градиентного бустинга и модели нейронной сети. Общая идея заключается в том, чтобы использовать сильные стороны обеих моделей для достижения лучшей производительности, чем у каждой из них в отдельности.

Градиентная бустинг - это техника, которая последовательно строит ансамбль деревьев решений. Каждое последующее дерево строится для исправления ошибок своего предшественника. Окончательный прогноз представляет собой взвешенную сумму прогнозов каждого дерева. Математически модель можно представить следующим образом:

$$F(x) = \sum_{x=0}^{\infty} \alpha k * hk(x)$$

где F(x) – окончательный прогноз;

αk – вес k-го дерева;

hk(x) – прогноз k-го дерева, и сумма по всем деревьям в ансамбле.

Нейронные сети – это класс моделей машинного обучения, которые состоят из взаимосвязанных слоев искусственных нейронов. Эти нейроны организованы во входной, скрытый и выходной слои. Модель изучает сложные закономерности в данных с помощью процесса, называемого обратным распространением, который регулирует веса связей между нейронами. Выход NN является функцией входных данных и весов связей:

$$y(x) = f(W * x + b)$$

где y(x) – выходное предсказание,

W – матрица весов,

x – входные данные,

b – член смещения,

f – функция активации.

Чтобы создать ансамбль моделей GB и NN, производится объединение прогнозов с помощью средневзвешенного значения:

$$A(x) = w1 * F(x) + w2 * y(x)$$

где A(x) – прогноз ансамбля;

w1 и w2 – веса, присвоенные прогнозам моделей GB и NN, соответственно;

F(x) и y(x) – прогнозы моделей GB и NN.

Выводы по разделу

В данном разделе был описаны логические и математические модели, алгоритмы и технологические описания по взаимодействию с лабораторным оборудованием. Рассматривались технические аспекты коммуникации с анализаторами, а также протоколы их взаимодействие. Были описаны методы применения базы знаний по лабораторным исследованиям, которые применялись по автоматической интерпретации отклонений от нормальных величин.

Рассматривались модели функциональной декомпозиции, который является основной моделью построения информационной системы, и всего функционала лабораторной информационной системы. Описаны методы реляционной алгебры, которые используются при выявлении причин и следствии при интерпретации результатов лабораторных исследований.

Были продемонстрированы алгоритмы машинного обучения как дерево решений, наивный Байес и статистический метод TF-IDF, градиентный бустинг, случайный лес, нейронные сети. Применяя алгоритмы, производились анализ данных и выявления патологических исследований, на которые необходимо произвести акцент, и учесть факторы при дальнейшей интерпретации. Были построены визуальные представления данных, таблицы с данными, про иллюстрирована программные коды, а также результаты обработок в виде экранных снимков.

4 ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ И РЕЗУЛЬТАТЫ ВНЕДРЕНИЯ

4.1 Архитектура информационной системы

Разработка интеллектуальной лабораторной информационной системы производилась с использованием web технологии и классического приложения. Была создана гибридная саморазвивающаяся трехзвенная платформа, с помощью которого создавались все формы, скрипты и структура базы данных. Все созданные объекты платформы размещаются на сервере приложений. Платформа состоит из двух основных частей:

– серверная часть, ядро системы (Backend), которая функционирует под управлением опера система Linux Ubuntu 22. В качестве сервера приложений выступают веб сервисы платформы (API), функционирующая под управлением Nginx (Apache). Языком разработки и функционирования платформы используются PHP 8.4. Системой управления базами данных является MYSQL (MariaDB). Форматом данных между бэкендом и клиентом используются XML и JSON;

– клиентская часть системы разрабатывалась на лицензионном Embarcadero Rad Studio 10.4, с использованием лицензионных компонентов.

Данная архитектура позволяет функционировать системе в любых решениях, как виртуальный хостинг, выделенные сервера, физические севера, системы виртуализации и т.д., а также пропорционально распределят нагрузку между клиентской и серверной части, что позволяет работать на обычных виртуальных хостингах.

Клиентская часть системы реализована под операционные системы Windows XP и выше для 64 и 32 разрядных систем.

Архитектура системы показано на следующем рисунке 47.

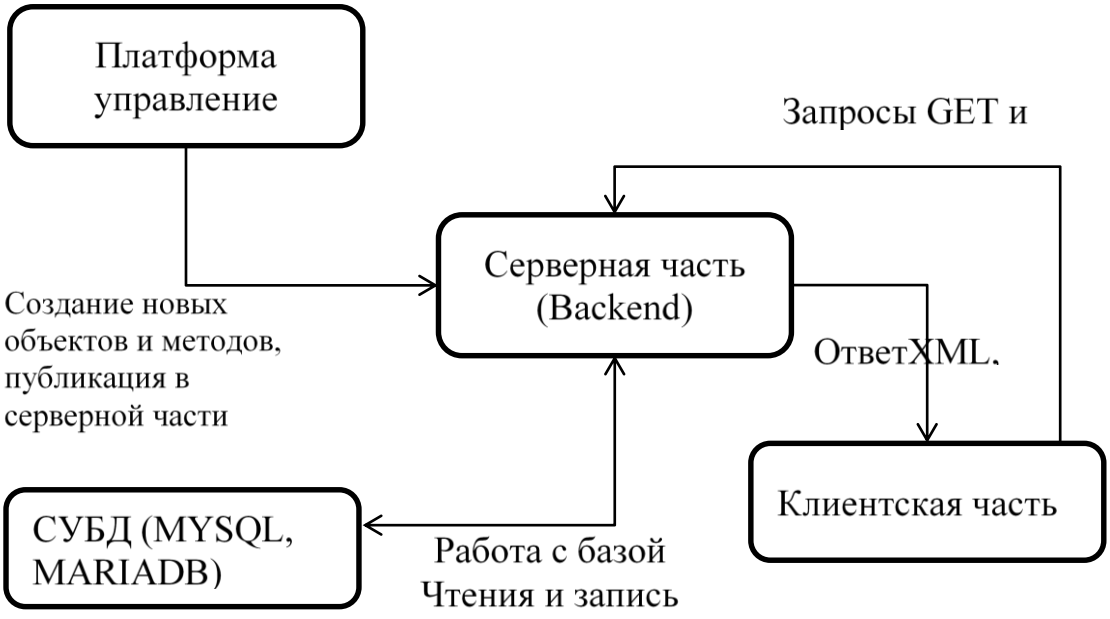


Рисунок 47 – Архитектура интеллектуальной системы

Данная архитектура является безопасной, так как отсутствует прямой доступ к данным и элементам платформы. Для обращения к сервисам требуется авторизации BASIC, далее вторичная авторизация в систему.

Архитектура базы данных полностью реляционная, и соответствует классическим 5 нормальным формам организации базы данных. Физический архитектура решения состоит из двух баз данных:

– системная – содержаться все настройки, объекты, пользователи и дрг. системные таблицы для функционирования системы. База данных содержит 25 таблиц;

– контентная – содержит все таблицы с данными. База данных содержит 184 таблицы;

– контентная база разделяется на логические группы таблиц, такие как:

группа основных транзакционных таблиц, которые содержат все данные по операционным процессам системы, а именно регистрация пациентов, заказы, выполнения исследований, результаты, прием платежей, интеграция с внешними системами и др. (рисунок 48, таблица 15).

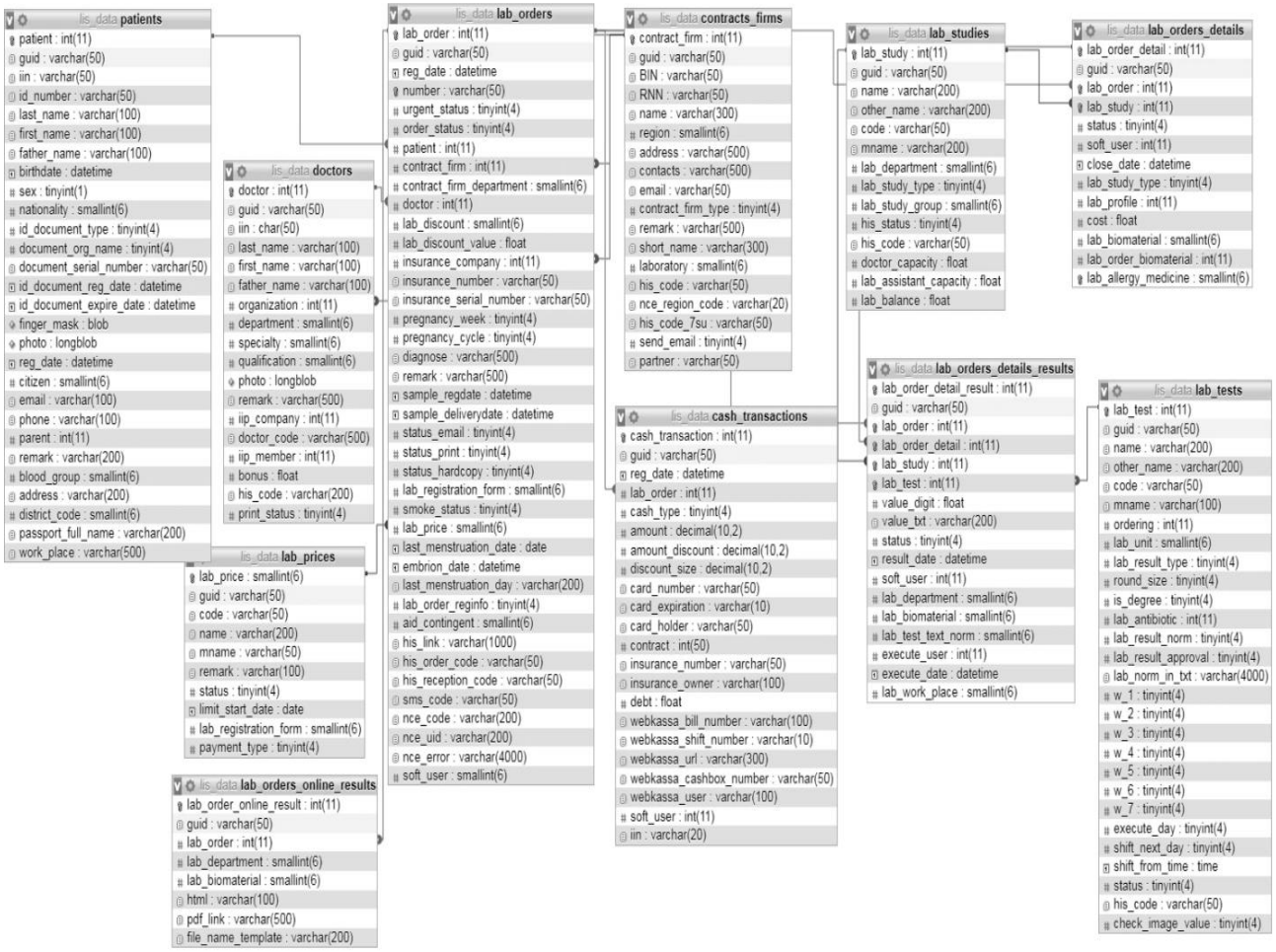


Рисунок 48 – Структура и связи основных транзакционных таблиц

Таблица 15 – Описание основных транзакционных таблиц

Наименование таблицы	Описание
Lab_orders	Регистрация всех заказов на оказание лабораторных исследований
Lab_orders_details	Содержит перечень исследований каждого из заказа из таблицы регистрации
Lab_orders_details_results	Содержит результаты выполнения исследований
Contract_firms	Справочник организации направляющие заказы на выполнения лабораторных исследований
Patients	Справочник пациентов, для которых выполняются лабораторные исследования
Doctors	Справочник врачей направителей на выполнения исследований

Группа тестов и исследований – таблицы содержащие данные о тестах и об исследованиях, а также их взаимосвязей. Данные группы таблицы содержат базу знаний по интерпретации результатов лабораторных исследований, распознавание кодов лабораторного оборудования, а также перечень оборудования подключенных той или иной конфигурации системы (рисунок 49).

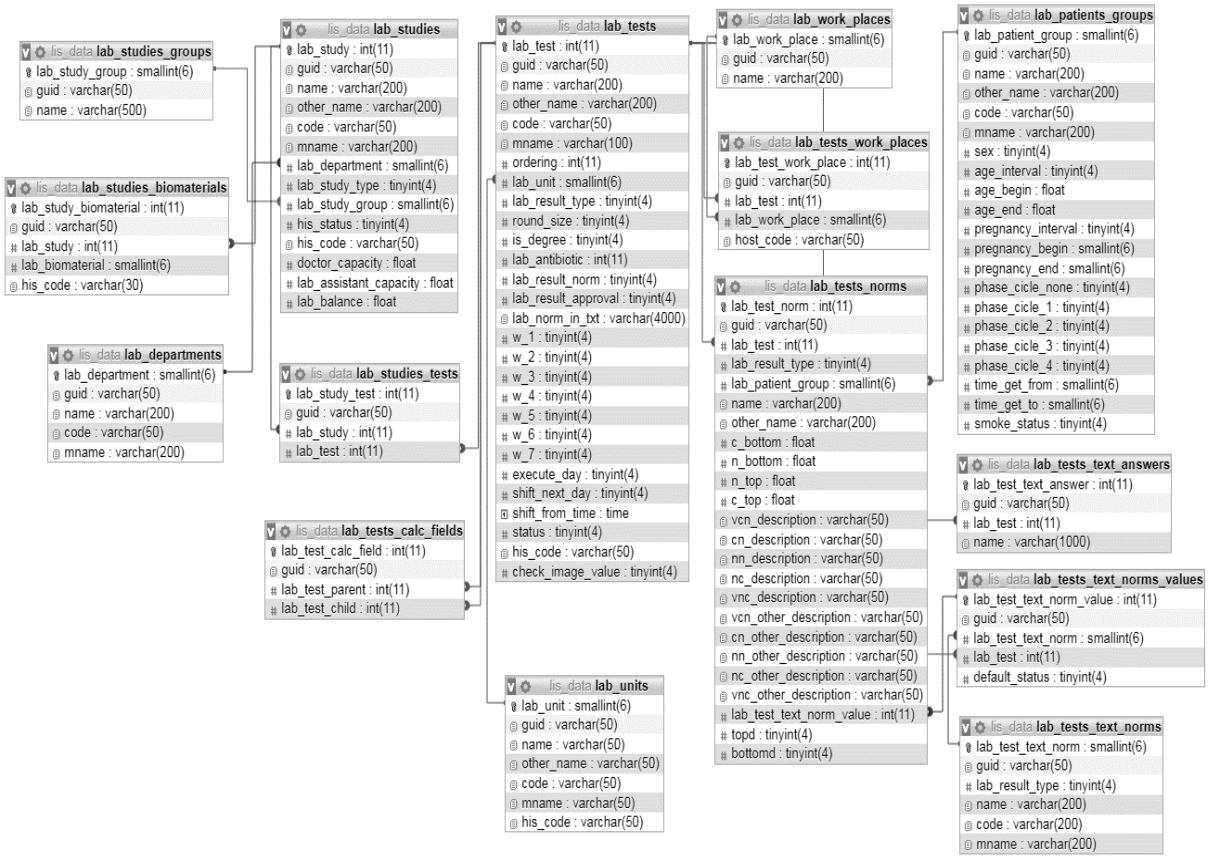


Рисунок 49 – Группа таблиц тестов и исследований со взаимосвязями

Таблица 16 – Группа таблиц базы знаний тестов

Наименование таблицы	Описание
Lab_tests	Все тесты необходимые для проведения исследований
Lab_studies	Исследования необходимые для проведения лабораторных анализов
Lab_studies_tests	Связку тестов с исследованиями
Lab_tests_norms	Нормативные величины для количественных тестов
Lab_tests_text_norms	Текстовые представления норм для качественных тестов
Lab_tests_text_norms_values	Значения текстовых представлении норм для качественных тестов
Lab_tests_text_answers	Ответы для качественных тестов
Lab_biomaterials	Перечень биоматериалов
Lab_studies_biomaterials	Связка между исследованиями и биоматериалами

Таблицы базы данных описанные в таблице 16 являются источниками базы знанийй по интепретации отклонений от нормативных величин для количественных и качественных тестов.

Группа справочников – это вспомогательные таблицы, которые дополняются основные таблицы, и участвуют в систематизации данных и позволяет исключить избыточность данных. К таким таблицам можно отнести справочник пациентов, врачей, контрагентов, прайс-листов, единиц измерения, регионов и т.д.

Следующая группа таблиц являются базой знаний по интепретации результатов лабораторных исследований (рисунок 50).

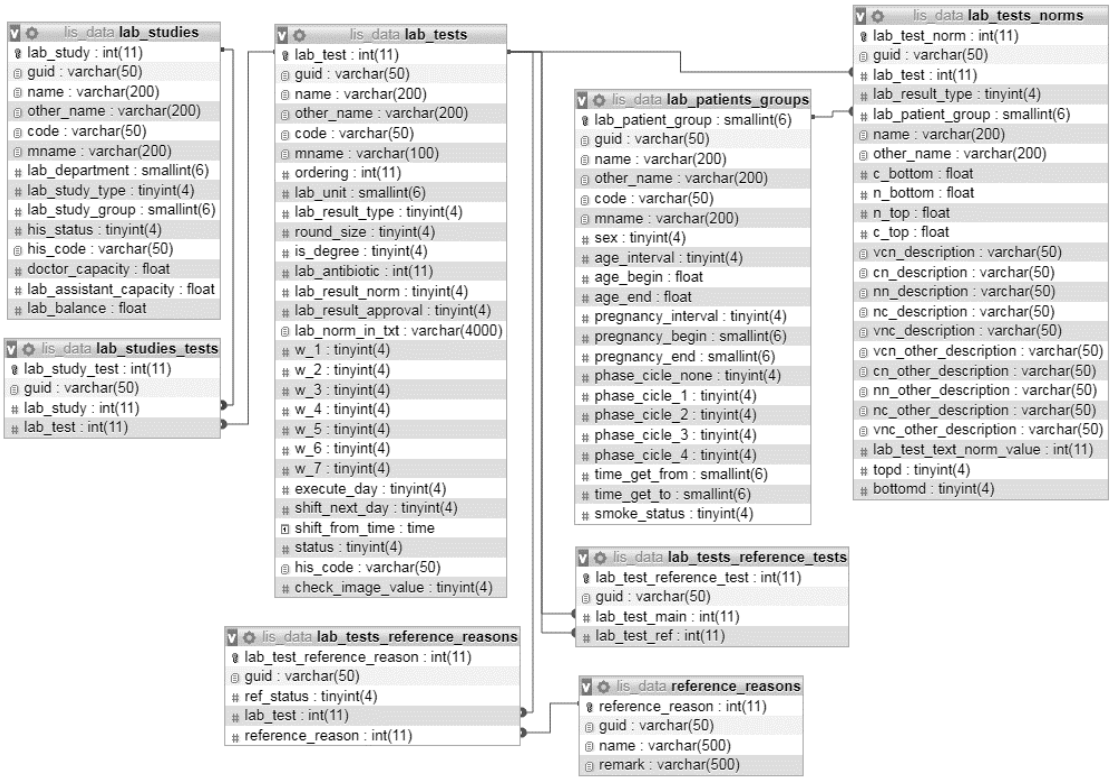


Рисунок 50 – Группа таблиц по интепретации результатов лабораторных исследований

Как показано на рисунке 49, таблицы lab_tests_reference_reasons, reference_reasons, lab_tests_reference_tests в связке с группой таблиц описанные на рисунке 50, осуществляют интепретацию результатов по причинам повышения и понижения. Интерпретация происходит с использованием методом пересечения результатов тестов и их связей с другими тестами.

Все запросы в базу данных осуществляется путем использования SQL запросов, которые вызываются только методами веб сервисов.

4.2 Модули информационной системы и их взаимодействие

Интеллектуальная лабораторная информационная система содержит следующие основные разделы и их модули:

Раздел «Работа» – данный раздел содержит основные транзакционные модули генерирующие данные в системе, и состоит из следующих модулей:

– каталог пациентов – справочник пациентов в системе;

– организации – справочник организации с которыми происходит взаимодействие;

– журнал заявок – журнал регистрации всех заявок на проведение лабораторных исследований;

– прайс-листы – содержит прейскурант цен по взаимодействию со сторонними организациями на проведение лабораторных исследований по государственному объему бесплатной медицинской помощи (ГОБМП), медицинское страхование (ОСМС), и платные услуги;

– журнал штрихкодов – является учетным модулем генерации штрих кодов для биоматериалов. Генерация штрих кодов производится по международному стандарту Code39;

– пользователи – учет пользователей системы, а также присвоение ролей и доступа к определенным модулем, ограничения по транзакционным действиям в системе.

– группы и роли – группы и роли для работы пользователей.

– внешние заявки – учет внешних заявок, а именно лабораторные исследования, направляющиеся в сторонние организации для выполнения лабораторных исследований;

– экспорт в НЦЭ (Национальный центр экспертизы) – модуль взаимодействия с Национальным центром экспертизы по направлению результатов анализов COVID 19, для дальнейшей активации системе «ASHYK» и «Путешествую без Covid19»;

– сортировка – модуль сортировки биоматериалов в лаборатории, который позволяет вести строгий учет биоматериалов, доставленных в лабраторию.

Раздел «Действия» – данный раздел содержит функционалы по работе с данными, такие «добавление», «редактирование», «удаление», «сортировка», «печать», «экспорт» и др.

Раздел «Отчеты» – данный раздел содержит преднастроенные отчетные модули для ведения статистики по операционно-хозяйственной деятельности лаборатории;

Раздел «Справочники» – данный раздел содержит базу знаний системы;

Раздел «Документация» – содержит модули по автоматизации ручных методик проведения анализов в лаборатории.

Модуль регистрации является точкой входа данных в интеллектуальную лабораторную информационную систему. В систему информация о проведении лабораторных исследований производится следующими способами (рисунок 51).

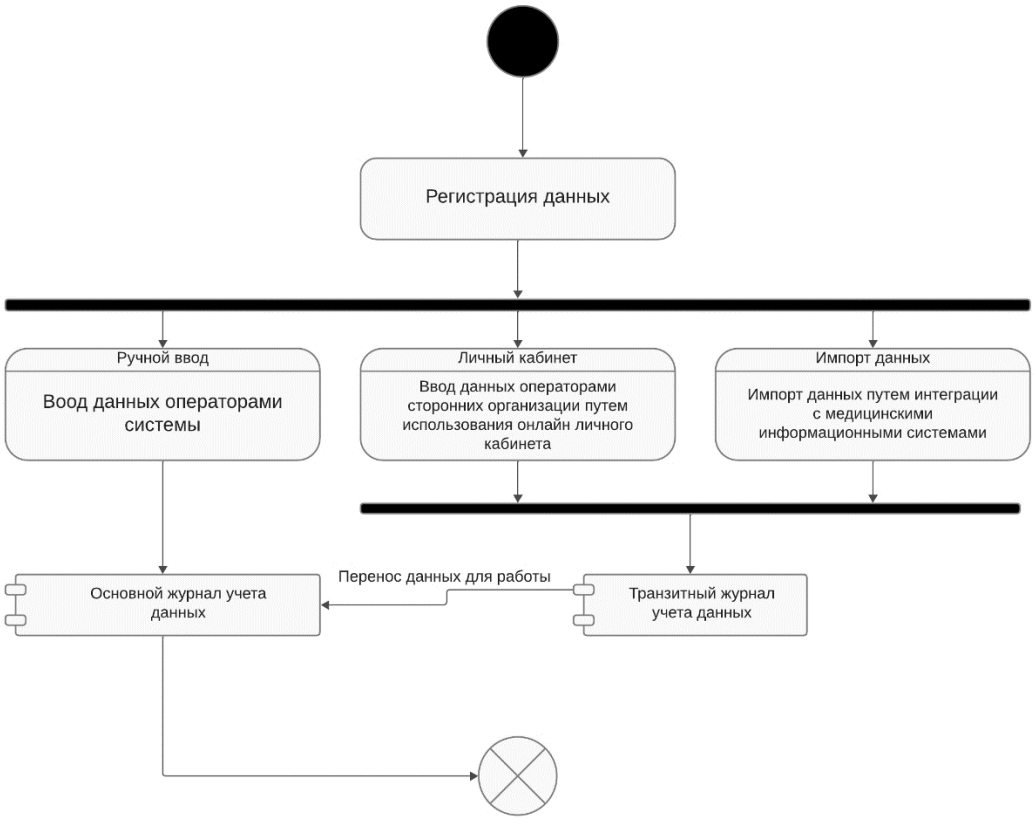


Рисунок 51 – Логика регистрации данных в системе

Ручной ввод данных – производится ручная регистрация пациента и анализов регистраторами системы, где заполняются все необходимые поля и выбираются исследования для проведения анализов.

При вводе ИИН система автоматический производит поиск пациента по Регистрационной базе прикрепленного населения, что минимизирует ошибки человеческого фактора. После ввода регистрационных данных выбираются исследования.

И после ввода всех необходимых данных, производится прием оплаты при необходимости. Информационная система имеет прямую интеграции с оператором фискальных данных WebKassa [82].

Интеграция с медицинскими информационными системами – в данном виде регистрации данных, пользователь производит импорт существующей

заявки из транзитного журнала (буфера) от медицинской информационной системы. Транзитный журнал является точкой контроля доставки биоматериалов в лабораторию, и после наличия материала, производится импорт в основной журнал, для гарантийного выполнения анализов. В настоящий момент времени система имеет интеграцию с медицинскими информационными системами «КМИС ДАМУ» [83], «МИС Жетысу» [84], «МИС Авиценна» [85], МИС Надежда [86].

Транзитный журнал позволяет отслеживать наличие доставленных биоматериалов, а также планируемые для проведения анализов. Интеграция с медицинскими информационными системами производится путем вызова веб-сервисов, разработанного для каждой МИС отдельно, что систематизирует и структурирует поток данных в лабораторию (рисунок 52).

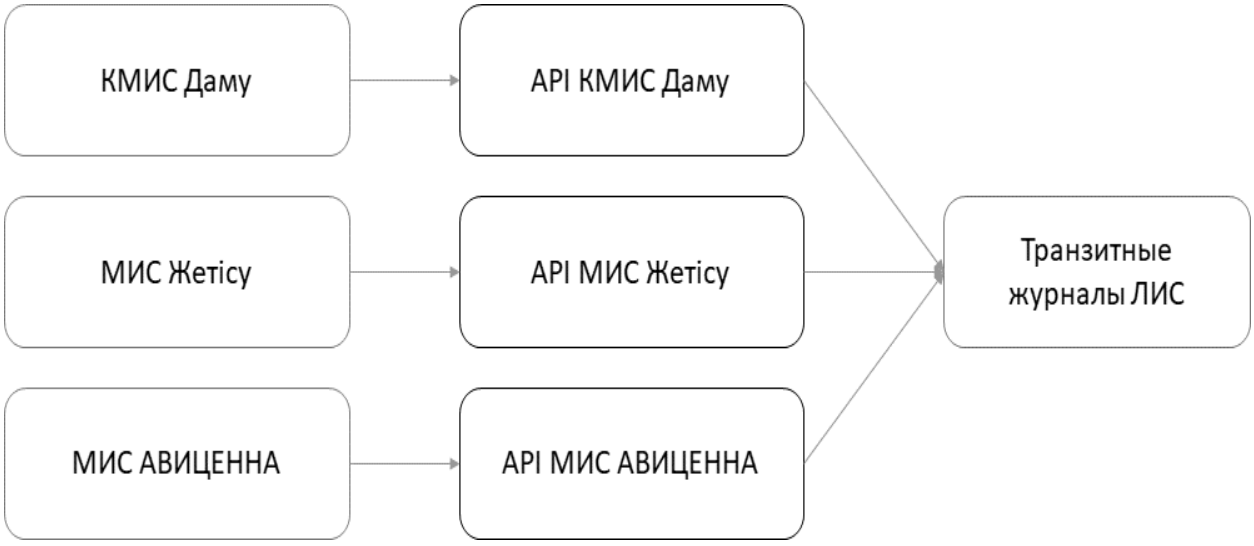


Рисунок 52 – Логическая модель интеграции

Личный кабинет контрагента – данный способ является прямой регистрации заявки в системе партнерами лаборатории, или же удаленными пунктами забора. Регистрация производится через специализированный web-сайт лаборатории, путем использования веб-сервисов системы. Личный кабинет оптимизирует работу персонала в самой лаборатории, и разгружает их от рутинных операционных процедур регистрации заявок.

Раздел справочников

Данный раздел является базой знаний интеллектуальной лабораторной информационной системы. Раздел состоит из следующих основных справочников:

– биоматериалы – данный справочник содержит перечень биоматериалов, используемых при проведениях лабораторных анализов. В данном справочник также указывается первичный контейнер, в котором производится транспортировка биоматериала;

– подразделения лаборатории – в данном справочнике содержится информация о внутренних подразделениях лаборатории;

– исследования – данный справочник содержит все исследования в системе, которые используются при формировании заказов на лабораторные анализы (рисунок 53).

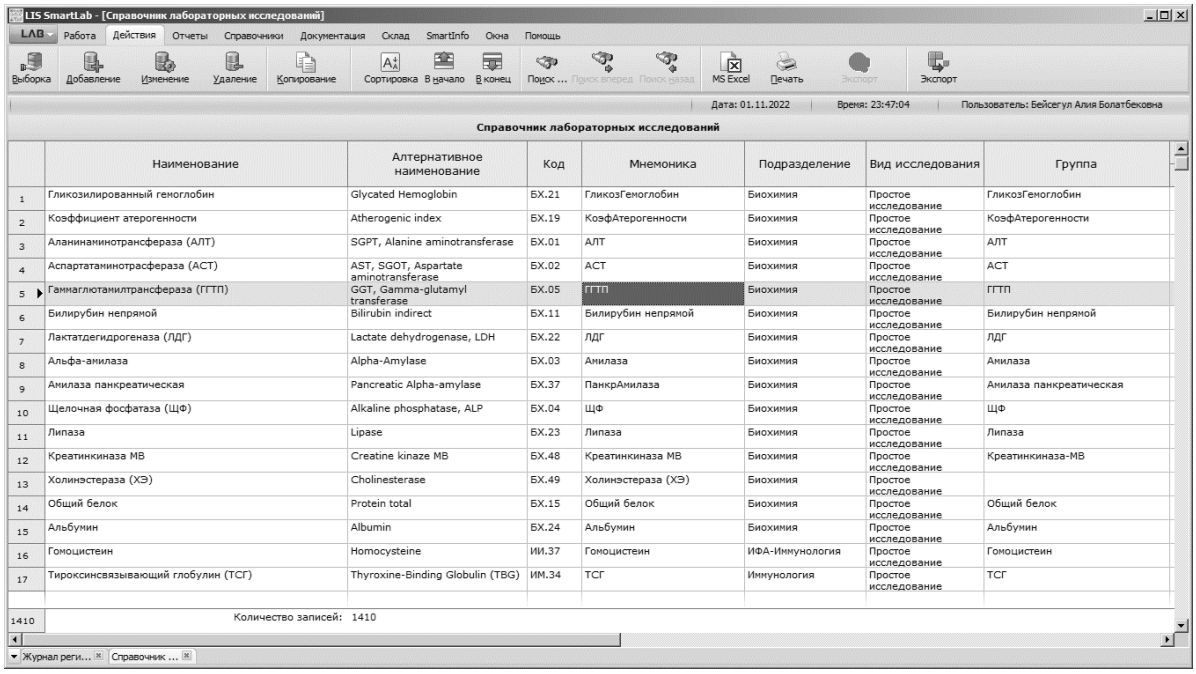


Рисунок 53 – Справочник исследований

Справочник является одним из ключевых справочников в базе знаний системы и имеет свои взаимосвязи с другими справочниками системы (рисунок 54).

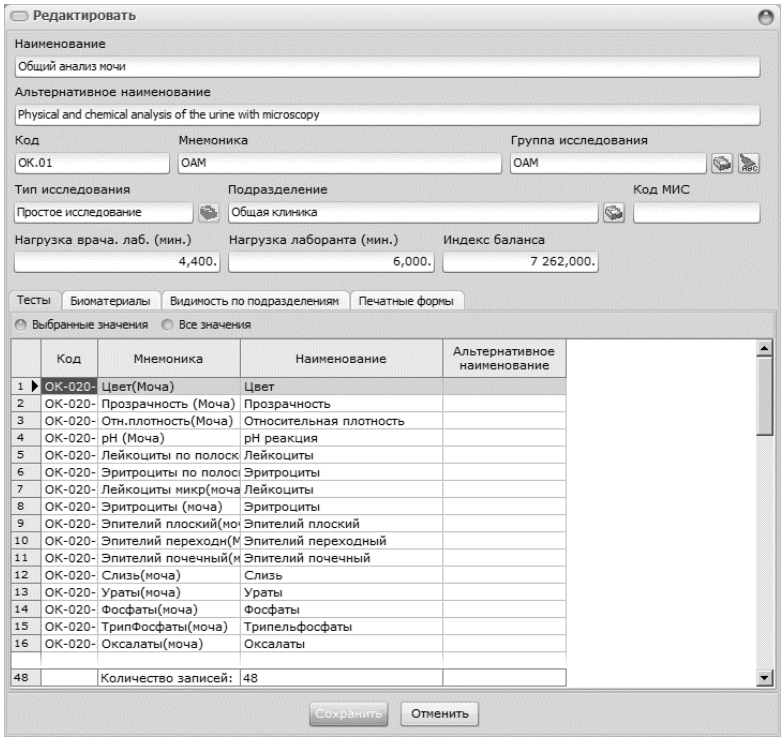


Рисунок 54 – Параметры исследования

Тесты – данный содержит перечень тестов для выполнения лабораторных анализов и является также одним из ключевых справочников.

Справочник определяет вид исследования (качественный, количественный), а также все нормативные величины имеют прямую привязку к тестам. Тесты имеют многочисленные взаимосвязи и параметры учета, и является корневой базой знаний в ЛИС (рисунок 55).

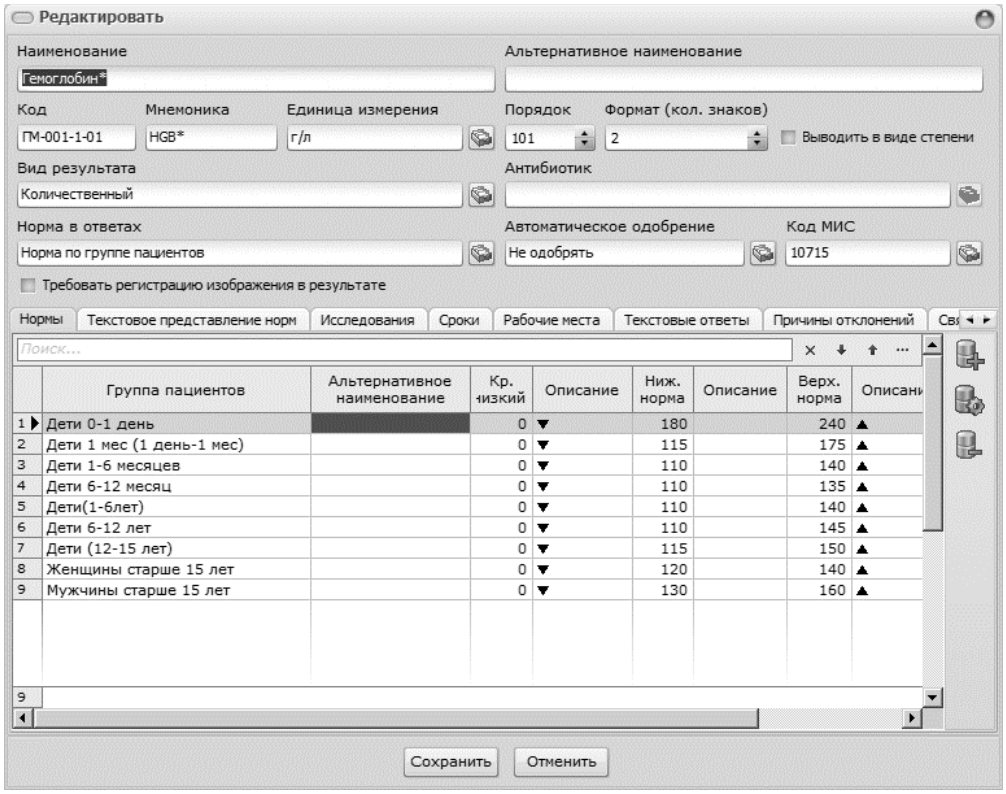


Рисунок 55 – Учетные параметры теста

Как видно на рисунке 54, что каждый тест в системе имеет множество параметров, которые определяют вид теста, роль теста, привязку к МИС, округления количественных результатов. В зависимости от вида теста определяется нормы (рисунок 56).

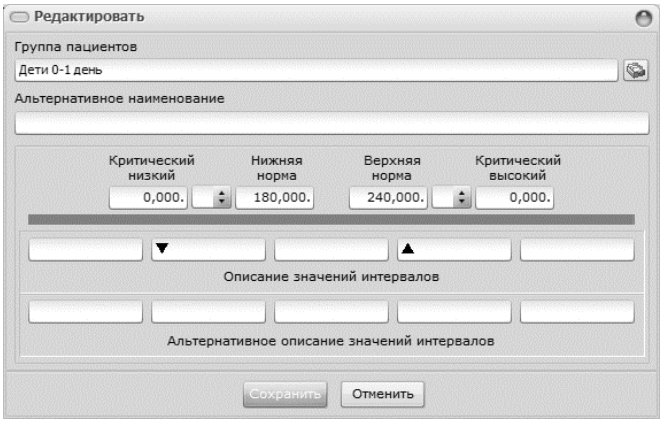


Рисунок 56 – Форма представление количественных норм

На рисунке 56 видны все параметры по нормативным величинам, которые были описаны в разделах выше:

– единицы измерения – справочник единиц измерений, используется при формировании справочника тестов. В системе учет единиц измерений производится согласно международной система единиц СИ [87] (фр. Système international d'unités, SI) – система единиц физических величин, современный вариант метрической системы. СИ является наиболее широко используемой системой единиц в мире – как в повседневной жизни, так и в науке и технике;

– группы пациентов – данный справочник используется в определении базы знаний по нормативным величинам по возрасту и полу и другим параметрам учета;

– антибиотики – данный справочник содержит перечень антибиотиков по группам и используется в микробиологии для определения антибиотика чувствительности микроорганизма.

Справочник также классифицирует антибиотики по международным кодам АТС. Анатомо-терапевтическо-химическая классификация – международная система классификации лекарственных средств. Используются сокращения: латиницей АТС (от Anatomical Therapeutic Chemical) [88].

Микроорганизмы – данный справочник содержит данные по микроорганизмам и используется в определении бактерии. Информация по микроорганизмам в основном представляются на латинице. Система также имеет базу знаний по интеграции с автоматизированными микробиологическими лабораторными анализаторами (Vitek, BactAlert) [89].

Регистрационные формы – данный справочник является конструктором отображения исследований при ручной регистрации в системе, с привязкой к прайс-листу.

Печатные формы – данный справочник также является базой знаний для бланков результатов, где в зависимости от вида исследования производится настройка печатной формы.

Печатная форма имеет физический шаблон, а также перечень исследований, которые будут использоваться при формировании заключения результатов лабораторного исследования.

Рабочие места – данный справочник содержит рабочие места для лабораторного оборудования. С помощью данного справочника производится подготовка базы знаний по взаимодействию с анализатором.

Шаблоны печатных форм – данный справочник содержит все шаблоны печатных форм. Шаблоны в дальнейшем используются в печатных формах и других разделах системы. Данный модуль разработан в виде встроенного дизайнера шаблонов на базе FastReport версии 5.4. [90]. Добавление новых шаблонов в систему производится путем выбора пред настроенного шаблона, или же путем заимствования с других шаблонов.

Браки – данный справочник содержит причины браков биоматериалов. Установка брака материалов происходит по многим причинам, например, «сгусток», «не правильный забор» и дрг. Каждая запись справочника имеет

шаблон ответа и привязку к справочнику биоматериалов, что позволяет системе следовать преднастроенным правилам.

Интеграции с лабораторным оборудованием. Интеграция с лабораторным производится по следующей схеме (рисунок 57).

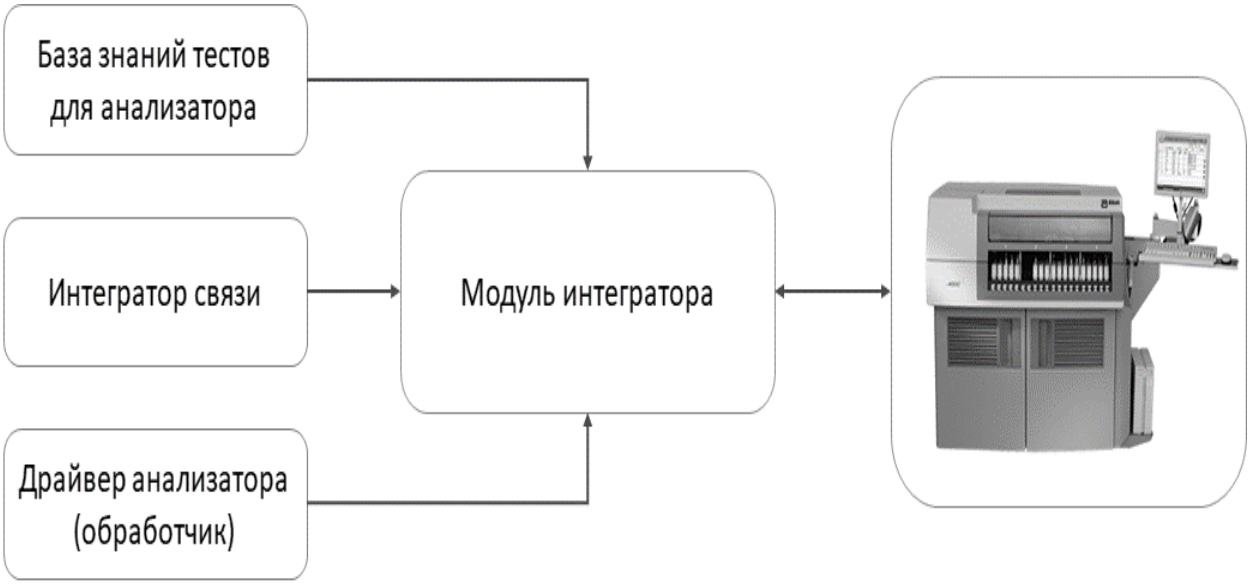


Рисунок 57 – Схема интеграции с лабораторным оборудованием

Согласно схеме, на рисунке 56, модуль интеграции состоит из взаимодействия трех блоков системы, такие как база знаний, интегратор связи и драйвер анализатора (обработчик).

Справочник рабочих мест, который был кратко описан в разделе 4.2, является базой знаний и содержит перечень анализаторов, шаблоны и тесты по взаимодействию с лабораторным анализатором. Каждая запись из справочника имеет свою базу знаний по шаблонам данных и тестов. У каждого теста указывается ХОСТ код, который является ключем к взаимодействию с тестами лабораторного оборудования.

Шаблоны содержат формат и порядок данных, наименование полей для обмена с анализатором. Настройка параметров производится путем выбора параметров из базы знаний для каждого анализатора. Формирование тестов производится путем выбора из справочника тестов. Для каждого рабочего места определяются свои тесты и соответственно свои хост коды (рисунок 58).

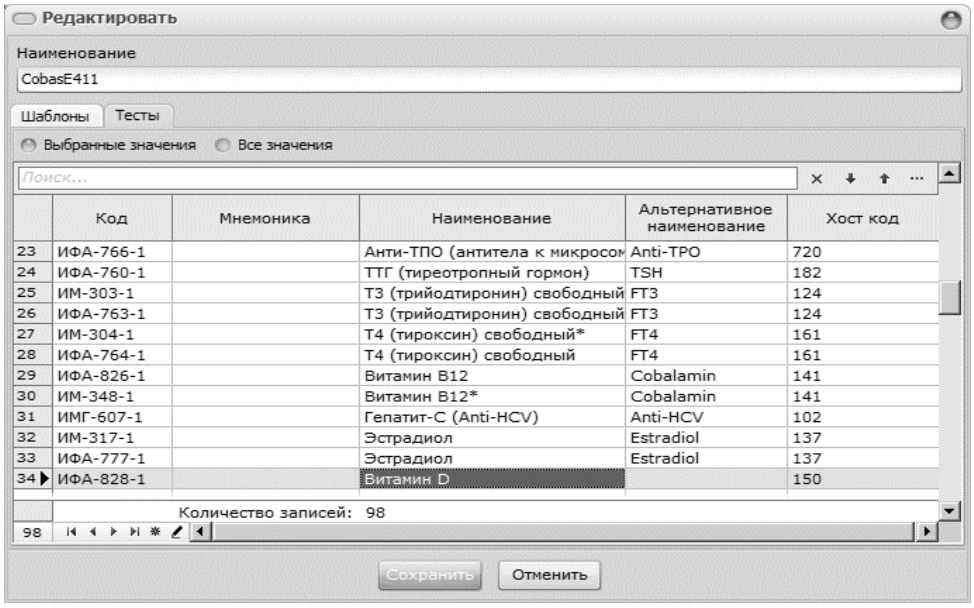


Рисунок 58 – Формирование тестов с хост кодами

Для качественных тестов, производится обучение каждого из возможного значения, возвращаемого анализатором (рисунок 59).

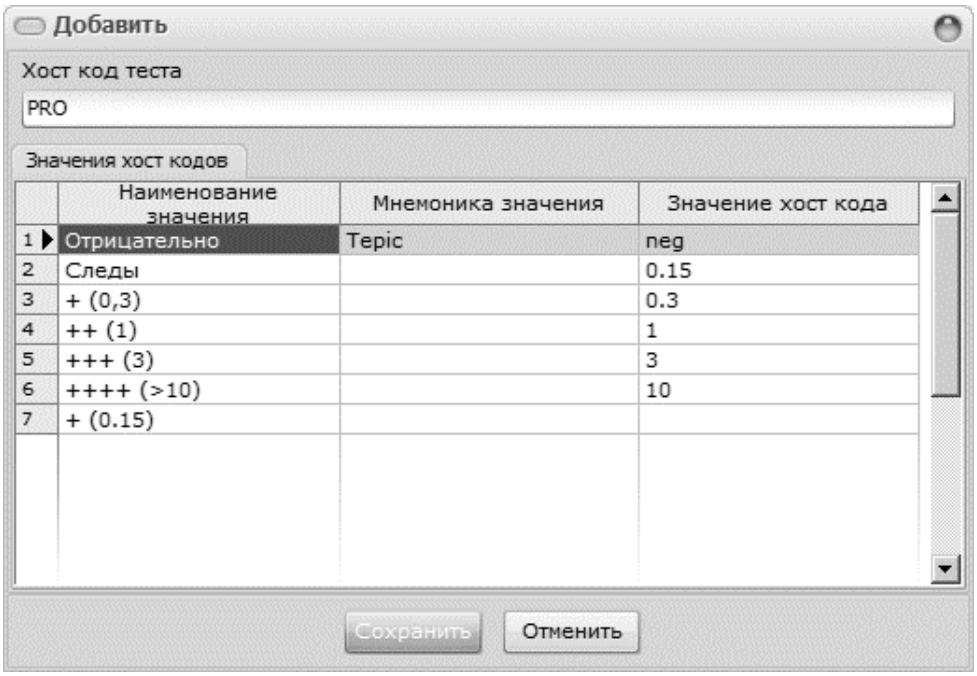


Рисунок 59 – Настройка значений качественных тестов

Следующим из блоков модуля интеграции является компонент интеграции связи. Данный компонент представляет собой универсальный модуль, с помощью которого определяется протокол взаимодействия с лабораторным анализатором (рисунок 60). Протокол взаимодействия происходит путем использования TCP/IP протокола или RS232 (com порт).

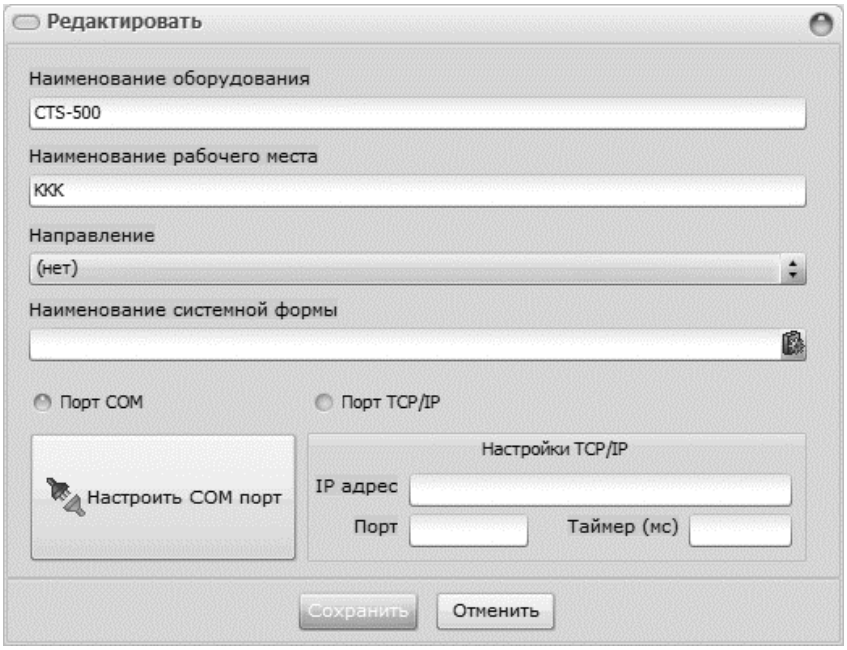


Рисунок 60 – Модуль интегратора связи

В модуле интегратора связи производится выбор драйвера, а именно системной формы который непосредственно взаимодействовать лабораторным оборудованием с предопределенными параметрами связи и стандартов.

Драйвера определяются для каждого лабораторного анализатора по отдельности. В системе содержится более 80 системных драйверов для различного вида оборудования.

После настройки рабочего места для работы с анализатором система готова полноценно взаимодействовать с лабораторным оборудованием. Одно физическое рабочее место может интегрироваться с несколькими лабораторными оборудованиями, а также позволяет визуально наблюдать обмен данными с анализатором.

В результате взаимодействия трех блоков производится полноценная интеграция с лабораторным оборудованием, а также дальнейшее интерпретация значений по отклонению от нормативных величин.

4.3 Интерпретации результатов лабораторных исследований

В системе интерпретация отклонений производится по качественным и количественным тестам. Интерпретация производится в режиме реального времени согласно правилам базы знаний. Качественные тесты интерпретируются путем сравнения полученного результата с результатом с нормальным значением (рисунок 61).

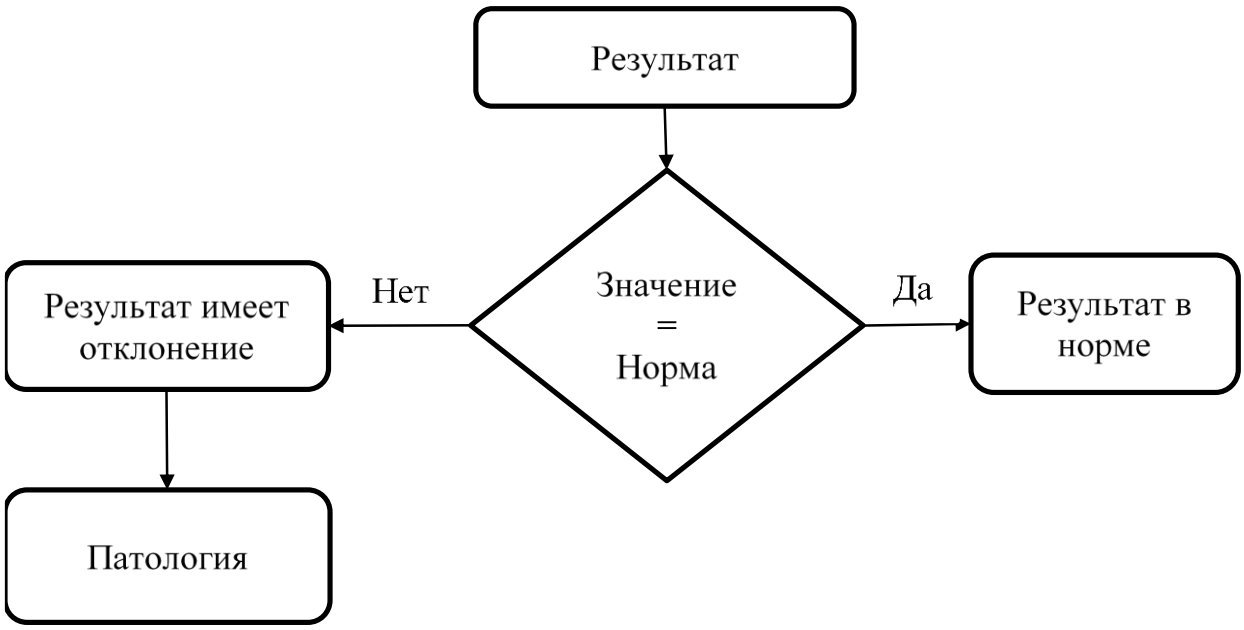


Рисунок 61 – Интерпретация качественных тестов

При интерпретации качественных тестов ряд тестов автоматический показывают патологию, так как исследование производится на наличие патологии или нет. Например, исследование на Sars Covid-19, где положительный результат автоматический описывает наличие вируса в организме.

В системе качественные тесты имеющие отклонения автоматический идентифицируются индикаторами (рисунок 62).

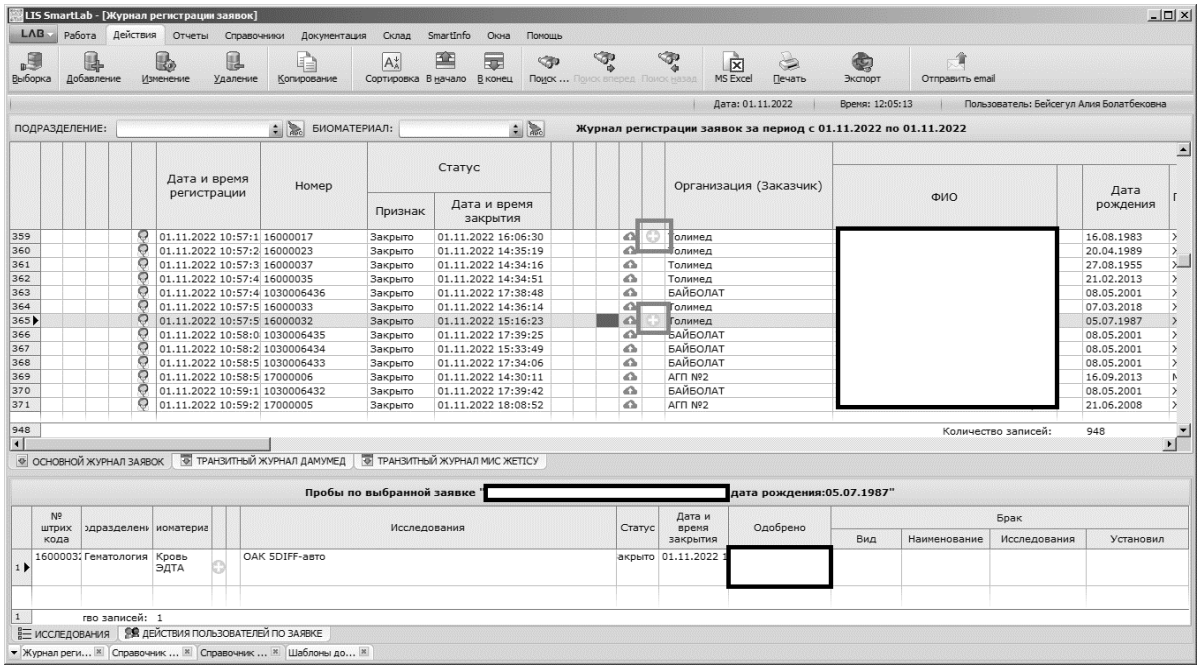


Рисунок 62 – Идентификация качественных тестов

Интерпретация количественных результатов производится путем сравнения результатов с граничными (нормативными) значениями тестов. Выявление отклонений производится автоматический по преднастроенными данными в базе знаний (рисунок 63).

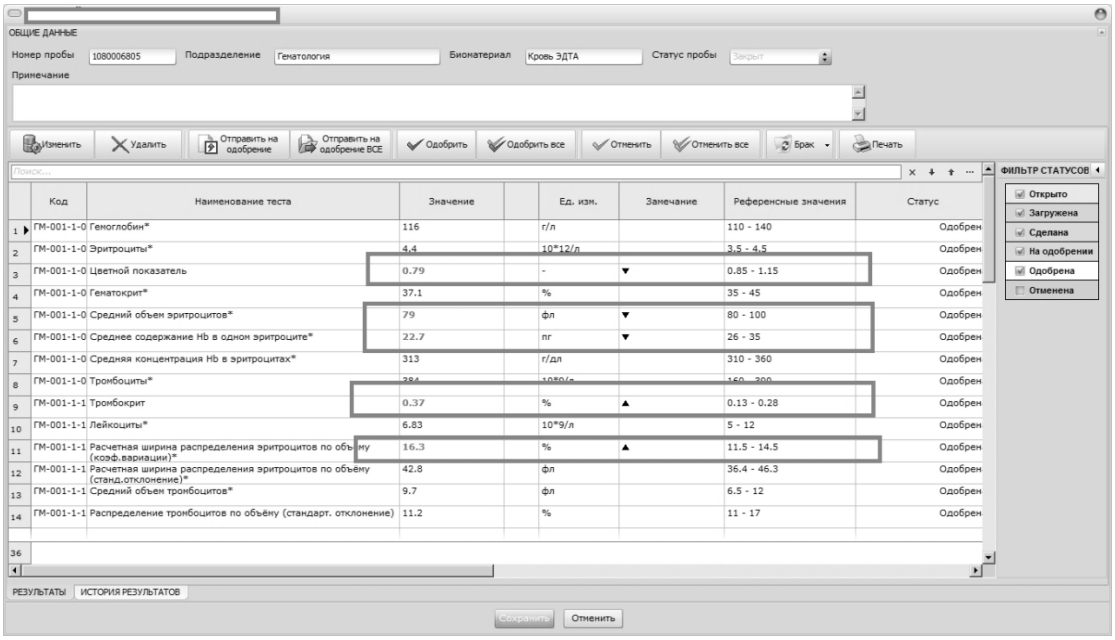


Рисунок 63 – Выявления отклонения от нормативных величин

На рисунке 63 можно увидеть позиции с повышенными и пониженными значениям от нормативных величин. Они выделяются красным цветом, тем самым сигнализируя наличие патологии.

Дальнейшем при формировании бланка результата, система автоматический получает шаблон, выявляет отклонения и генерирует заключение.

Интерпретация результатов производится для количественных методов исследования. Данный модуль является конечным этапом при формировании заключения для пациентов.

Для формирования патологии производится обучение системы данными, такими как связанные тесты и причины повышения и понижения.

Полноценная интерпретация производится с использованием искусственного интеллекта OpenAI [91] с применением chatGPT бота. Интеграция с OpenAI производилась путем использования публичных API методов [92], которые позволяют использовать базу знаний искусственного интеллекта. Обращение к ИИ производилось, используя вопросы касательно получившихся отклонений тестов от нормальных величин, отправляя запрос в формате JSON:

["model"=>"gpt-3.5-turbo",
"messages"=> [
[
"role"=>"user",

```
            "content"=> $question
        ]
            ]
];
```

В результате отправки запроса, ИИ возвращает необходимую информацию о причинах отклонения, методы лечения и другую информацию. Можно осуществлять множественные запросы, что позволит получить больше информации.

Для работы с ИИ был разработан интеграционный скрипт в виде класса на языке PHP с применением CURL библиотеки. Разработанный класс можно применять и для других целей, так как он является универсальным по обращение к ИИ.

```
<?php
$path=«/var/www/vhosts/kkk.kz/httpdocs/ai.smartlab.kz/»;
require_once $path . 'lib/smart_token.php';//Класс по работе с токенами
require_once $path . 'lib/db.php'; //Класс по работе с базами
require_once $path . 'lib/smart_ai.php'; //Класс по работе с OpenAI
header('Content-Type: text/html; charset=utf-8');

//Получаем параметры запроса
$params = trim(file_get_contents(«php://input»));

$ai=new smart_ai();

$question=$params[«lab_test_name_combination»];

//Переделываем комбинацию для организации запроса
$question=str_replace('0','ниже нормы',$question);
$question=str_replace('1','выше нормы',$question);
if ($params[«gender»]==«1») $gender=«мужской»;
if ($params[«gender»]==«2») $gender=«женский»;

$request=[«model»=>«gpt-3.5-turbo»,
«messages»=>[[
«role»=>«user»,
«content»=> $question]
                                        ]
                    ];

 //Формируем запрос
 $ai_result=$ai->GetExplanation($request);

if ($ai_result==«error») //Ошибка сервера. Такое бывает
```

```
{
 //Возвращаем сообщение что сбой интерпретации
 $result=[«status»=>«0»,
 «message»=>«Ошибка формирования интерпретации. По пробуйте еще раз.»,
 «explanation»=>««
 ];
        }
else
{
$ai_result=json_decode($ai_result,true);

//Записываем знание в базе
$dataset->WriteKnowledge([
«lab_test_name_combination»=>$params[«lab_test_name_combination»],
«lab_test_combination»=>$params[«lab_test_combination»],

«explanation»=>str_replace('\n','',$ai_result[«choices»][0][«message»][«content»]
),
«user»=>$token[«user_id»],
«age»=>$params[«age»],
«gender»=>$params[«gender»]
                                                    ]
                                                   );
 //Возвращаем клиенту интерпретацию
 $result=[«status»=>«1»,
  «message»=>«Интерпретация сформирована!»,

«explanation»=>str_replace('\n','',$ai_result[«choices»][0][«message»][«content»]
),
«question»=>$question
 ];
}

}
//Освобождаем объекты
unset($dataset);
unset($ai);
unset($token);
?>
```

Логическая схема взаимодействия выглядит следующим образом (рисунок 64).

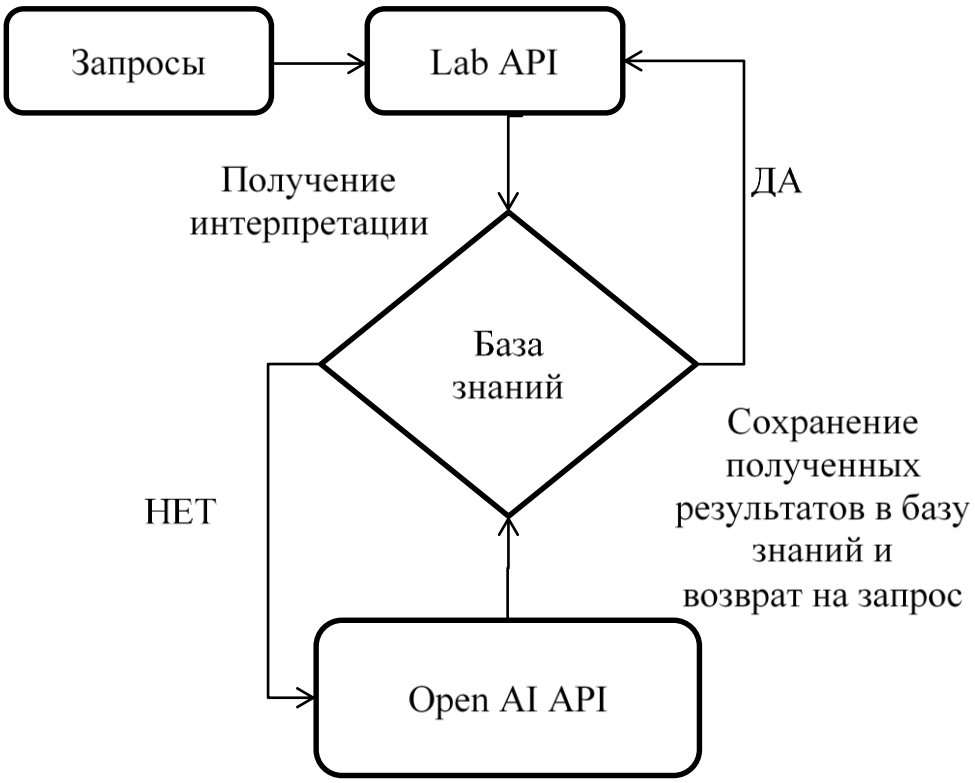


Рисунок 64 – Логическая схема алгоритма взаимодействия

На рисунке 64 представлен процесс обучения собственного ИИ путем обучения с помощью Open AI, то есть при отсутствии необходимых знаний в локальной базе данных, система обращается к Open AI и записывает в базе данных полученные знания.

OpenAI – это технологическая платформа, предоставляющая передовые решения в области искусственного интеллекта для различных отраслей, включая здравоохранение. Для медицинских лабораторий OpenAI может помочь оптимизировать рабочие процессы, автоматизировать повторяющиеся задачи и повысить точность результатов анализов. Интегрируя алгоритмы машинного обучения и возможности обработки естественного языка, OpenAI может помочь медицинским лабораториям предоставлять более эффективные и результативные услуги своим пациентам. Некоторые потенциальные области применения OpenAI в медицинских лабораториях включают:

– автоматизированный анализ и интерпретация результатов лабораторных тестов;

– поддержка принятия клинических решений, которая может помочь медицинским работникам принимать более обоснованные решения на основе последних медицинских исследований и доказательств.

Результаты внедрения. В период с 2020 по 2022 года, в рамках реализации практической части диссертационной работы, интеллектуальная лабораторная информационная система была внедрена и апробирована в следующих организациях (таблица 17).

Таблица 17 – Перечень организации, где внедрена лабораторная информационная система

Организация	Годы
ТОО «LabStar Kazakhstan», лаборатория «Tumar», г. Алматы	2022
ТОО МЕДСИ, лаборатория «Medcи», г. Караганда	2022

Выводы по разделу

В данной разделе было представлено архитектура, описание, лабораторной информационной системы «SmartLab». Представлены визуальные и инфологические иллюстрации, схемы и технологии взаимодействия с лабораторным оборудованием. Были продемнострированы программные коды по взаимодействию с сервисами Open AI. Разработаны веб-сервисы по самостоятельному обучению и пополнению базы знаний по интерпретации результатов лабораторных исследований. А также были описаны алгоритмы применения искусственного интеллекта, применение OpenAI, наименование и адрес веб-сервисов, с указанием методов взаимодействия.

Представлены медицинские лаборатории, которые используют у себя данное решение, и практическим опытом производят обучение системы новыми знаниями по интерпретации результатов лабораторных исследований.

ЗАКЛЮЧЕНИЕ

Интерпретация результатов лабораторных исследований с помощью комплексной автоматизации производится путем использования искусственного интеллекта (ИИ) и других цифровых технологий для автоматического анализа и интерпретации результатов лабораторных исследований. Этот подход направлен на рационализацию процесса интерпретации результатов лабораторных исследований и предоставление более точных, последовательных и своевременных результатов медицинским работникам.

Целью диссертации являлась разработка модели искусственного интеллекта по интерпретации результатов лабораторных исследований в здравоохранении, для его практического применения в медицинских лаборатории с помощью комплексной автоматизации, и участие в Государственной программе «Цифровой Казахстан».

В соответствии с поставленной целью выполнены следующие задачи:

1. Изучены аналоги подобных систем и алгоритмов.
2. Сформировано BIG DATA результатов лабораторных исследований.
3. Сформирована единая базы референсных значений лабораторных исследований.
4. Разработана модель по интерпретации результатов лабораторных исследований по BIG DATA.
5. Разработана модель искусственного интеллекта по выявлению патологии в результатах лабораторных исследований.

В рамках выполнения диссертационной работы также были разработаны следующие дополнительные возможности:

– внедрена система QR кодирования, что позволило населению верифицировать полученные результаты лабораторных исследований и внести вклад в улучшение оказания услуг лабораторной медицины. Данный функционал на практике позволил выявить несколько случаев подделки результатов лабораторных исследований на Covid-19;

– осуществлена интеграция с государственным единым порталом по мониторингу эпидемиологической обстановки по Covid-19, что способствовало реализации задач поставленной Государственной программы «Цифровой Казахстан»;

– внедрение модернизации оказало возможность участия в государственном проекте Ashyq [93], где с помощью мобильного приложения осуществляется допуск в объекты участников программы;

Разработка логических моделей, алгоритмов, ориентированных на интеллектуальную поддержку процессов диагностики на уровне клинико-диагностической лаборатории, способствует оптимизации и совершенствованию диагностических и прогностических методов, связанных с использованием лабораторных данных. Это позволяет создать среду для совместной работы специалистов из лабораторий и клинических отделений,

улучшая общую эффективность и качество диагностических процедур, неотъемлемо связанных с анализом лабораторной информации.

По рамках диссертационной работы были получены следующие результаты:

– изучены аналоги в мировой практике медицинской лабораторной диагностики;

– сформирована BIG DATA результатов лабораторных исследований по биохимическим, иммунологическим и гематологическим результатам лабораторных исследований;

– сформирована единая база референсных значений лабораторных исследований по производителям лабораторного оборудования и региональных особенностей Республики Казахстан;

– разработана методика по комплексной автоматизации по интерпретации результатов лабораторных исследований по BIG DATA;

– разработана модель искусственного интеллекта по выявлению патологии в результатах лабораторных исследований на основе комплексной автоматизации.

– разработана платформа, модулей интеллектуальной лабораторной информационой системы;

– разработана система информационной поддержки рабочих процессов в лабораторной диагностике и верификации клинических патологии;

– разработана интеграционная платформа по взаимодействую с лабораторным оборудованием для автоматического взаимодействия интеллектуальной системы без участия человека;

– разработана интеллектуальная информационная система, реализующее предложенные модели и алгоритмы в режиме реального времени;

– результаты диссертационной работы внедрены и используются государственными и частными медицинскими лабораториями на территории Республики Казахстан.

Комплексная автоматизация интерпретации результатов лабораторных исследований потенциально может повысить эффективность и точность интерпретации результатов лабораторных исследований, что приведет к улучшению состояния пациентов и принятию более правильных клинических решений. Однако важно отметить, что модели ИИ не совершенны и все еще могут допускать ошибки, и что медицинские работники должны всегда просматривать результаты автоматизированной интерпретации перед принятием диагноза или решения о лечении.

Методики, алгоритмы, разработанные в рамках диссертационной работы, имеют большой потенциал применения в других отраслях автоматизации. Так как алгоритмы интеграции с лабораторным оборудованием могут быть использованы во всех сферах автоматизации, где требуется обмен данными с анализаторами. Алгоритмы работы с OpenAI, так же имеет трансформации в другие сферы как образование, здравоохранение, агротехнологии и др.

Следующим этапом развития темы диссертации является применение ИИ в распознавании образов, то есть определение злокачественных микроорганизмов, вирусов на снимках с микроскопов. Что позволит врачам-лаборантам анализировать более точнее, быстрее и эффективней выполнять микроскопические исследования.

Примером трансфера технологии является продукт интеллектуальной лабораторной системы для генетических лабораторий «SmartGene». Решение по автоматизации генетических лабораторий результатом которого является внедрение в трех генетических лабораториях Республики Казахстан и получения Свидетельства о внесении сведений в государственные права на объекты, охраняемые авторским правом №14755 от 29 января 2021 года выданного РГП «Национальный институт интеллектуальной собственности» Министерства Юстиций Республики Казахстан (Приложение Г).

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 Постановление Правительства Республики Казахстан. Об утверждении государственной программы «Цифровой Казахстан»: утв. 12 декабря 2017 года, №827 // https://adilet.zan.kz/rus/docs/P1700000827. 08.09.2020.

2 Электронный паспорт здоровья можно посмотреть на платформах e-Gov и mGov // https://www.primeminister.kz/ru/news/press. 15.07.2021.

3 Госпрограмма развития здравоохранения «Денсаулық»: ключевые показатели за 2018 год // https://primeminister.kz/ru. 01.11.2019.

4 Глава государства провел совещание по реализации государственной программы «Цифровой Казахстан» // https://www.akorda.kz/ru. 13.03.2020.

5 Аптекман А., Калабин В., Клинцов В. и др. Цифровая Россия: новая реальность. – М., 2017. – 132 с.

6 McKinsey. Digital Russia Report // https://www.mckinsey. 20.05.2022.

7 E-labdoc // http://e-labdoc.roche.com/. 20.05.2022.

8 Sysmex // http://www.sysmex.co.jp/en/. 01.11.2019.

9 Добреньков В.И., Кравченко А.И. Методы социологического исследования: учеб. – М.: ИНФРА-М, 2024. – 768 с.

10 Paranjape K., Schinkel M., Hammer R.D. et al. The Value of Artificial Intelligence in Laboratory Medicine // Am J Clin Pathol. – 2021. – Vol. 155, Issue 6. – P. 823-831.

11 Islam M.M., Poly T.N., Yang H.-C. et al. Deep into Laboratory: An Artificial Intelligence Approach to Recommend Laboratory Tests // Diagnostics. – 2021. – Vol. 11, Issue 6. – P. 990-1990-13.

12 Park D.J., Park M.W., Lee H. et al. Development of machine learning model for diagnostic disease prediction based on laboratory tests // Scientific Reports. – 2021. – Vol. 11, Issue 1. – P. 7567-1-7567-11.

13 Gunčar G., Kukar M., Notar M. et al. An application of machine learning to haematological diagnosis // Sci Rep. – 2018. – Vol. 8, Issue 1. – P. 411-1-411-12.

14 Jain N., Jhunthra S., Garg H. et al. Prediction modelling of COVID using machine learning methods from B-cell dataset // Results in Physics. – 2021. – Vol. 21. – P. 103813-1-103813-19.

15 Looten V., Chang L.K.W., Neuraz A. et al. What can Millions of laboratory test results tell us about the temporal aspect of data quality? study of data spanning 17 years in a clinical data warehouse // Computer Methods and Programs in Biomedicine. – 2019. – Vol. 181. – P. 104825-1-104825-6.

16 Dunn J., Kidzinski L., Runge R. et al. Wearable sensors enable personalized predictions of clinical laboratory measurements // Nature Medicine. – 2021. – Vol. 27, Issue 6. – P. 1105-1112.

17 Zhang S. et al. Medical Diagnosis From Laboratory Tests by Combining Generative and Discriminative Learning // https://deepai.org/publication. 01.11.2019.

18 Crowley R.J., Tan Y.J., Ioannidis J.P.A. Empirical assessment of bias in machine learning diagnostic test accuracy studies // Journal of the American Medical Informatics Association. – 2020. – Vol. 27, Issue 7. – P. 1092-1101.

19 Place J.F., Truchaud A., Ozawa K. et al. Use of artificial intelligence in analytical systems for the clinical laboratory // Journal Clin Biochem. – 1995. – Vol. 28, Issue 4. – P. 373-389.

20 Luo Y., Szolovits P. et al. Using Machine Learning to Predict Laboratory Test Results // American Journal of Clinical Pathology. – 2016. – Vol. 145, Issue 6. – P. 778-788.

21 Jackins V., Vimal S., Kaliappan M. et al. AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes // The Journal of Supercomputing. – 2020. – Vol. 77, Issue 5. – P. 5198-5219.

22 Naugler C., Church D.L. Automation and artificial intelligence in the clinical laboratory // Critical Reviews in Clinical Laboratory Sciences. – 2019. – Vol. 56, Issue 2. – P. 98-110.

23 Кадиркулов К.К., Исмаилова А.А. Модель интерпретации результатов лабораторных исследований // Матер. междунар. науч.-теорет. конф. «Сейфуллинские чтения – 16: Молодежная наука новой формации – будущее Казахстана». – Астана, 2020. – С. 153-155.

24 Кадиркулов К.К., Мустафин С.Н., Исмаилова А.А. Мониторинг и анализ динамики цен на лекарственных средств (ЛС) с помощью аналитических сервисов // Матер. междунар. науч.-практ. онлайн-конф. «Интеграция науки, образования и производства - основа реализации Плана нации». – Карагнада, 2020. – С. 1038-1040.

25 Кадиркулов К., Исмаилова А. Практическое применение QR кодов в результатах лабораторных исследований // Современная аграрная наука: цифровая трансформация (С. Сейфуллинские чтения №17). – Астана, 2021. – С. 221-224

26 Кадиркулов К. Процесс автоматизации интерпретации результатов исследований путем интеграции с лабораторным оборудованием BACT/ALERT 3D // Молодежь и наука – взгляд в будущее (Сейфуллинские чтения – 18). – Астана, 2022. – С. 52-55.

27 Кадиркулов К., Исмаилова А. Автоматизация исследований медицинских лабораторий // Математическая логика и компьютерная наука: матер. междунар. науч. конф. – Астана, 2022. – С. 272-275.

28 Badrick T. Evidence-based laboratory medicine // Clin Biochem Rev. – 2013. – Vol. 34. – P. 43-46.

29 Makary M.A., Daniel M. Medical error – the third leading cause of death in the US // BMJ. – 2016. – Vol. 353. – P. i2139.

30 Tizhoosh H.R., Pantanowitz L. Artificial intelligence and digital pathology: challenges and opportunities // J Pathol Inform. – 2018. – Vol. 9. – P. 38-1-38-6.

31 Hood L., Friend S.H. Predictive, personalized, preventive, participatory (P4) cancer medicine // Nat Rev Clin Oncol. – 2011. – Vol. 8. – P. 184-187.

32 Park J.Y., Kricka L.J. One hundred years of clinical laboratory automation: 1967–2067 // Clin Biochem. – 2017. – Vol. 50. – P. 639-644.

33 Chapman T. Lab automation and robotics: automation on the move // Nature. – 2003. – Vol. 421. – P. 661-666.

34 Rutherford M.L., Stinger T. Recent trends in laboratory automation in the pharmaceutical industry // Curr Opin Drug Dis. Dev. – 2001. – Vol. 4. – P. 343-346.

35 Genzen J.R., Burnham C.D., Felder R.A. et al. Challenges and opportunities in implementing total laboratory automation // Clin Chem. – 2018. – Vol. 64. – P. 259-264.

36 Lou A.H., Elnenaei M.O., Sadek I. et al. Evaluation of the impact of a total automation system in a large core laboratory on turnaround time // Clin Biochem. – 2016. – Vol. 49. – P. 1254-1258.

37 Ialongo C., Porzio O., Giambini I. et al. Total automation for the core laboratory: improving the turnaround time helps to reduce the volume of ordered STAT tests // J Lab Autom. – 2016. – Vol. 21. – P. 451-458.

38 Armbruster D.A., Overcash D.R., Reyes J. Clinical Chemistry Laboratory Automation in the 21st Century – Amat Victoria curam (Victory loves careful preparation) // Clin Biochem Rev. – 2014. – Vol. 35. – P. 143-153.

39 Dauwalder O., Landrieve L., Laurent F. et al. Does bacteriology laboratory automation reduce time to results and increase quality management? // Clin Microbiol Infect. – 2016. – Vol. 22. – P. 236-243.

40 Archetti C., Montanelli A., Finazzi D. et al. Clinical laboratory automation: a case study // J Public Health Res. – 2017. – Vol. 6. – P. 31-36.

41 Zaninotto M., Plebani M. The 'hospital central laboratory': automation, integration and clinical usefulness // Clin Chem Lab Med. – 2010. – Vol. 48. – P. 911-917.

42 Novak S.M., Marlowe E.M. Automation in the clinical microbiology laboratory // Clin Lab Med. – 2013. – Vol. 33. – P. 567-588.

43 Lam C.W. et al. Implementing a laboratory automation system: experience of a large clinical laboratory // J Lab Autom. – 2012. – Vol. 17. – P. 16-23.

44 Roche Diagnostics // https://usdiagnostics.roche.com. 01.04.2022.

45 Siemens Healthcare Diagnostics // https://www.healthcare. 15.07.2022.

46 Beckman Coulter // https://www.beckmancoulter.com. 10.03.2022.

47 Abbott Core Laboratory // https://www.corelaboratory.abbott. 20.04.2022.

48 Peek N., Combi C., Marin R. et al. Thirty years of artificial intelligence in medicine (AIME) conferences: a review of research themes // Artif Intell Med. – 2015. – Vol. 65. – P. 61-73.

49 Turing A.M. Computing machinery and intelligence // Mind. – 1950. – Vol. 59. – P. 433-460.

50 TechEmergence // https://www.techemergence.com. 01.10.2022.

51 Mohammed E.A., Far B.H., Naugler C. Applications of the MapReduce programming framework to clinical big data analysis: current landscape and future trends // BioData Min. 2014. – Vol. 7. – P. 22-1-22-23.

52 The Platform for Big Data and the Leading Solution for Apache Hadoop in the Enterprise Cloudera // https://codecondo.com/top-10-big-data. 01.112022.

53 Iliashenko O., Bikkulova Z., Dubgorn A. Opportunities and challenges of artificial intelligence in healthcare // E3S Web of Conferences. – 2019. – Vol. 110. – P. 02028-1-02028-8.

54 Dolley S. Big data's role in precision public health // Front Public Health. – 2018. – Vol. 6. – P. 68-1-68-12.

55 Baron J.M., Dighe A.S., Arnaout R. et al. The 2013 symposium on pathology data integration and clinical decision support and the current state of the field // J Pathol Inform. – 2014. – Vol. 5. – P. 2-1-2-12.

56 Wu H., Yang S., Huang Z. et al. Type 2 diabetes mellitus prediction model based on data mining // Inform Med Unlocked. – 2018. – Vol. 10. – P. 100-107.

57 Sarwar A., Sharma V. Intelligent Naïve Bayes approach to diagnose diabetes Type-2 // International Journal of Computer Applications. – 2012. – Vol. 3. – P. 14-16.

58 Undre P., Kaur H., Patil P. Improvement in prediction rate and accuracy of diabetic diagnosis system using fuzzy logic hybrid combination // Procced. Internat. conf. on Pervasive Computing (ICPC). – Pune, 2015. – P. 1-4

59 Place J.F., Truchaud A., Ozawa K. et al. Use of artificial intelligence in analytical systems for the clinical laboratory // Clinical Biochemistry. – 1995. – Vol. 28, Issue 4. – P. 373-389.

60 Goldstein B.A., Navar A.M., Carter R.E. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges // Eur Heart J. – 2017. – Vol. 38. – P. 1805-1814.

61 Castro W, Oblitas J, Santa-Cruz R, Avila-George H. «Multilayer perceptron architecture optimization using parallel computing techniques» // PLoS ONE. – 2017. – Vol. 12. – P. e0189369-1-e0189369-17.

62 Abdelhafiz D., Yang C., Ammar R. et al. Deep convolutional neural networks for mammography: Advances, challenges and applications // BMC Bioinform. – 2019. – Vol. 20. – P. 75-94.

63 Cannas M., Arpino B. A comparison of machine learning algorithms and covariate balance measures for propensity score matching and weighting // Biom. J. – 2019. – Vol. 61. – P. 1049-1072.

64 Degoulet P. The HEGP component-based clinical information system // Int. J. Med. Inform. – 2003. – Vol. 69. – P. 115-126.

65 Zapletal E., Rodon N., Grabar N. et al. Methodology of integration of a clinical data warehouse with a clinical information system: the HEGP case // Stud. Health Technol. Inform. – 2010. – Vol. 160. – P. 193-197.

66 Alballa N., Al-Turaiki I. Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review // Informatics in Medicine Unlocked. – 2021. – Vol. 24. – P. 100564-1-100564-18.

67 Кадиркулов К.К, Исмаилова А.А. Централизованный сбор и анализ результатов лабораторных исследований на Covid-19 // Scientific Journal of Astana IT University. – 2020. – Т. 4, №4. – С. 58-66.

68 Кадиркулов К.К, Исмаилова А.А. QR verification of laboratory studies results // Известия НАН РК. – 2021. – №2. – С. 96-101.

69 Системный анализ // https://ru.wikipedia.org/wiki. 01.07.2022.

70 Чувствительность и специфичность // https://ru.wikipedia.org/wiki. 01.07.2022.

71 ISO/IEC/IEEE 24765. Systems and software engineering – Vocabulary. – Impl. 2010-12-15. – Geneva, 2010. – 418 p.

72 ISO/IEC 2382-1:1993. Information technology. Vocabulary. Part 1: Fundamental terms // https://www.iso.org/standard/7229.html. 01.07.2022.

73 Новые информационные технологии в образовании: применение технологий '1С' для развития компетенций цифровой экономики: сб. науч.-практ. конф. / под ред. Д.В. Чистова. – М., 2018. – 484 с.

74 Дмитрий Н.Д., Барабанов В.Ф. Использование экспертной системы с продукционной базой знаний и нейронной сети для распознавания сигналов // Актуальные проблемы прикладной математики, информатики и механики: сб. тр. междунар. науч. конф. – Воронеж, 2020. – С. 438-442.

75 Одинцов Б.Е. Роботизация процесса актуализации баз знаний интеллектуальных информационных систем предприятия // Новые информационные технологии в образовании: сб. науч. тр. 20-й междунар. науч.-практ. конф. – М., 2020. – Ч. 1. – С. 42-45.

76 Лабораторная диагностика // https://ru.wikipedia.org/wiki. 01.06.2022.

77 Международная система единиц // https://ru.wikipedia.org/wiki. 01.06.2022.

78 Машинное обучение // https://ru.wikipedia.org/wiki. 01.06.2022.

79 Manyika J., Chui M., Brown B. et al. McKinsey Global Institute. Big Data: The Next Frontier for Innovation, Competition, and Productivity. – Chicago: McKinsey, 2011. – 143p.

80 Канаракус К. Машина Больших Данных // http://www.osp.ru/. 01.06.2022.

81 Jones K.S. A statistical interpretation of term specificity and its application in retrieval // Journal of Documentation. – 2004. – Vol. 60, Issue 5. – P. 493-502.

82 Онлайн служба фискализации чеков // webkassa.kz. 01.07.2022.

83 Медицинская информационная система Даму // https://damu.kz. 01.07.2022.

84 Медицинская информационная система Жетісу // ttps://e-mis.kz/. 01.07.2022.

85 Медицинская информационная система Авиценна // https://avicenna.online/. 01.07.2022.

86 Медицинская информационная система Надежда // https://xn--80aalaeg9b.kz/. 01.07.2022.

87 Международное бюро мер и весов // https://ru.wikipedia.org. 01.07.2022.

88 Анатомо-терапевтическо-химическая классификация // https://ru.wikipedia.org/wiki. 01. 07.2022.

89 Кадиркулов К.К, Исмаилова А.А. Процесс автоматизации интерпретации результатов исследований путем интеграции с лабораторным оборудованием Vitek2 // Тр. Университета. – 2022. – №2(87). – С. 318-324.

90 Система генерации отчетов // https://www.fast-report.com. 01.08.2022.

91 Об OpenAI // https://en.wikipedia.org/wiki/OpenAI. 01.11.2022.

92 API Open AI // https://platform.openai.com. 01.11.2022.

93 Единая платформа Ashyq по контролю вакцинации на Covid-19 // https://ashyq.kz/. 01.11.2022.

ПРИЛОЖЕНИЕ А

Рекомендации из инструкции к реактиву ТТГ для анализаторов Cobas производства Roche

07028091501V1.0

Elecsys TSH

cobas®

Drug	Concentration tested mg/L
Amiodarone	200
Prednisolone	100
Hydrocortisone	200
Fluocortolone	100
Octreotide	0.300
Levothyroxine	0.250
Liothyronine	0.015

The presence of autoantibodies may induce high molecular weight complexes (macro-TSH) which may cause unexpected high values of TSH.[7]

In rare cases, interference due to extremely high titers of antibodies to analyte-specific antibodies, streptavidin or ruthenium can occur. These effects are minimized by suitable test design.

For diagnostic purposes, the results should always be assessed in conjunction with the patient's medical history, clinical examination and other findings.

Limits and ranges

Measuring range

0.005-100 µIU/mL (defined by the Limit of Detection and the maximum of the master curve). Values below the Limit of Detection are reported as < 0.005 µIU/mL. Values above the measuring range are reported as > 100 µIU/mL (or up to 1000 µIU/mL for 10-fold diluted samples).

Lower limits of measurement

Limit of Blank, Limit of Detection and Limit of Quantitation

Limit of Blank = 0.0025 µIU/mL

Limit of Detection = 0.005 µIU/mL

Limit of Quantitation = 0.005 µIU/mL

The Limit of Blank, Limit of Detection and Limit of Quantitation were determined in accordance with the CLSI (Clinical and Laboratory Standards Institute) EP17-A2 requirements.

The Limit of Blank is the 95th percentile value from n ≥ 60 measurements of analyte-free samples over several independent series. The Limit of Blank corresponds to the concentration below which analyte-free samples are found with a probability of 95 %.

The Limit of Detection is determined based on the Limit of Blank and the standard deviation of low concentration samples. The Limit of Detection corresponds to the lowest analyte concentration which can be detected (value above the Limit of Blank with a probability of 95 %).

The Limit of Quantitation is the lowest analyte concentration that can be reproducibly measured with an intermediate precision CV of ≤ 20 %.

Dilution

Samples with TSH concentrations above the measuring range can be diluted with Diluent MultiAssay. The recommended dilution is 1:10 (either automatically by the analyzer or manually). The concentration of the diluted sample must be ≥ 10 µIU/mL.

After manual dilution, multiply the result by the dilution factor.

After dilution by the analyzer, the software automatically takes the dilution into account when calculating the sample concentration.

Expected values

0.270-4.20 µIU/mL

These values correspond to the 2.5th and 97.5th percentiles of results obtained from a total of 516 healthy test subjects examined.

Each laboratory should investigate the transferability of the expected values to its own patient population and if necessary determine its own reference ranges.

Specific performance data

Representative performance data on the analyzer is given below. Results obtained in individual laboratories may differ.

Precision

Precision was determined using Elecsys reagents, samples and controls in a protocol (EP05-A3) of the CLSI (Clinical and Laboratory Standards Institute): 2 runs per day in duplicate each for 21 days (n = 84). The following results were obtained:

	cobas e 801 analyzer				
		Repeatability		**Intermediate precision**	
Sample	Mean µIU/mL	SD µIU/mL	CV %	SD µIU/mL	CV %
Human serum 1	0.009	0.001	7.1	0.001	11.3
Human serum 2	0.209	0.003	1.6	0.005	2.5
Human serum 3	1.88	0.026	1.4	0.043	2.3
Human serum 4	51.8	0.653	1.3	1.05	2.0
Human serum 5	90.0	1.24	1.4	1.75	1.9
PC[b] Universal 1	1.41	0.020	1.4	0.030	2.1
PC Universal 2	8.18	0.132	1.6	0.207	2.5
PC TS	0.184	0.003	1.8	0.004	2.3

b) PC = PreciControl

Method comparison

A comparison of the Elecsys TSH assay on the **cobas e** 801 analyzer (y) with the same assay on the Elecsys 2010 analyzer (x) using clinical samples gave the following correlations (µIU/mL):

Number of samples measured: 130

Passing/Bablok[8]	Linear regression
y = 0.936x - 0.003	y = 0.958x - 0.052
r = 0.989	r = 0.999

The sample concentrations were between 0.009 and 92.6 µIU/mL.

Analytical specificity

The following cross-reactivities were found, tested with TSH concentrations of 0.3 µIU/mL and 8 µIU/mL.

Cross-reactant	Concentration tested mU/mL	Cross-reactivity %
LH	10000	< 0.038
FSH	10000	< 0.008
hGH	1000	0.000
hCG	50000	0.000

References

1 Kronenberg HM, Melmed S, Polonsky KS, et al. Williams Textbook of Endocrinology. Saunders Elsevier, Philadelphia, 12th edition, 2011, chapter 10, p. 301-318.

2 Wu AHB. Tietz Clinical Guide To Laboratory Tests. Saunders Elsevier, Philadelphia, 4th edition, 2006, section II, p. 1040-1043.

3 Surks MI, Chopra IJ, Mariash CN, et al. American Thyroid Association Guidelines for the Use of Laboratory Tests in Thyroid Disorders. JAMA 1990;263:1529-1532.

4 Keffer JH. Preanalytical Considerations in Testing Thyroid Function. Clin Chem 1996;42(1):125-135.

5 Ladenson PW. Optimal laboratory testing for diagnosis and monitoring of thyroid nodules, goiter and thyroid cancer. Clin Chem 1996;42:(1):183-187.

6 Nicoloff JT, Spencer CA. The use and misuse of the sensitive thyrotropin assays. J Clin Endocr Metab 1990;71:553-558.

7 Sakai H, Fukuda G, Suzuki N, et al. Falsely Elevated Thyroid-Stimulating Hormone (TSH) Level Due to Macro-TSH. Endocr J 2009;56(3):435-440.

8 Bablok W, Passing H, Bender R, et al. A general regression procedure for method transformation. Application of linear regression procedures for method comparison studies in clinical chemistry, Part III. J Clin Chem Clin Biochem 1988 Nov;26(11):783-790.

ПРИЛОЖЕНИЕ Б

Рекомендованные компанией SYSMEX референтные интервалы для общего анализа крови (ОАК)

1.5 Reference Intervals

Reference intervals (Normal Population Reference Ranges) were developed for the XT-200*i*/XT-1800*i* using normal individuals. The range for each parameter is calculated for 95% confidence intervals.

Parameter	Range for Females n = 133	Range for Males n = 182
WBC	3.98 - 10.04	4.23 - 9.07
Neut%	34.0 - 71.1	34.0 - 67.9
Lymph%	19.3 - 51.7	21.8 - 53.1
Mono%	4.7 - 12.5	5.3 - 12.2
Eo%	0.7 - 5.8	0.8 - 7.0
Baso%	0.1 - 1.2	0.2 - 1.2
Neut#	1.56 - 6.13	1.78 - 5.38
Lymph#	1.18 - 3.74	1.32 - 3.57
Mono#	0.24 - 0.36	0.30 - 0.82
Eo#	0.04 - 0.36	0.04 - 0.54
Baso#	0.01 - 0.08	0.01 - 0.08
RBC	3.93 - 5.22	4.63 - 6.08
HGB	11.2 - 15.7	13.7 - 17.5
HCT	34.1 - 44.9	40.1 - 51.0
MCV	79.4 - 94.8	79.0 - 92.2
MCH	25.6 - 32.2	25.7 - 32.2
MCHC	32.2 - 35.5	32.3 - 36.5
RDW-CV	11.7 - 14.4	11.6 - 14.4
RDW-SD	36.4 - 46.3	35.1 - 43.9
RET% (XT-2000i only)	0.5 - 1.7	0.51 - 1.81
RET# (XT-2000i only)	1.64 - 7.76	2.6 - 9.5
IRF (XT-2000i only)	3.0 - 15.9	2.3 - 13.4
PLT	182 - 369	163 - 337
MPV	9.4 - 12.3	9.4 - 12.4

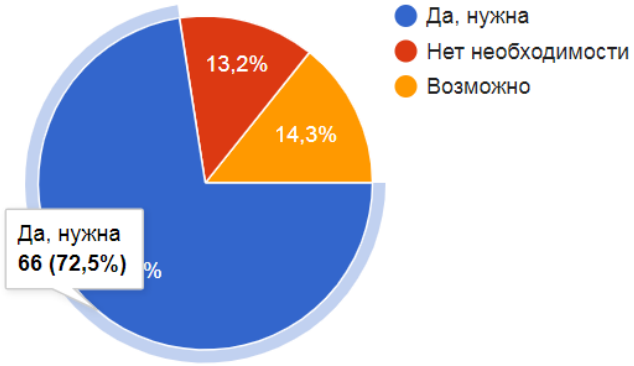
* The age range for females was 21 - 41 years with a mean age of 24.5.

Note:

Sysmex recommends that each laboratory establish its own expected reference intervals based upon the laboratory's patient population encountered during daily operation. Expected reference intervals may vary due to the differences in sex, age, diet, fluid intake, geographic location, etc. The NCCLS Document C28-A "How to Define and Determine Reference Intervals in the Clinical Laboratory; Approved Guideline" contains guidelines for determining reference values and intervals for quantitative clinical laboratory tests.

ПРИЛОЖЕНИЕ В

а



б

б

а – Пациенты медицинских лабораторий; ə – Врачи медицинских учреждений; б – Сотрудники медицинских лабораторий

Рисунок В.1 – Потребность в интерпретации лабораторных результатов

ПРИЛОЖЕНИЕ Г

Авторские свидетельства

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ

РЕСПУБЛИКА КАЗАХСТАН

СВИДЕТЕЛЬСТВО

О ВНЕСЕНИИ СВЕДЕНИЙ В ГОСУДАРСТВЕННЫЙ РЕЕСТР ПРАВ НА ОБЪЕКТЫ, ОХРАНЯЕМЫЕ АВТОРСКИМ ПРАВОМ

№ 32850 от «21» февраля 2023 года

Фамилия, имя, отчество, (если оно указано в документе, удостоверяющем личность) автора (ов):
КАДИРКУЛОВ КУАНЫШ КАЙСАРОВИЧ, БЕЙСЕҒҰЛ ӘЛИЯ БОЛАТБЕККЫЗЫ

Вид объекта авторского права: **программа для ЭВМ**

Название объекта: **Лабораторная информационная система «SmartLab»**

Дата создания объекта: **14.01.2023**

Қоғат түпнұсқалығын https://www.kazpatent.kz/ru сайтының "Авторлық құқық" бөлімінде тексеруге болады https://copyright.kazpatent.kz

Подлинность документа возможно проверить на сайте kazpatent.kz в разделе «Авторское право» https://copyright.kazpatent.kz

Подписано ЭЦП

Е. Оспанов

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ

РЕСПУБЛИКА КАЗАХСТАН

СВИДЕТЕЛЬСТВО

О ВНЕСЕНИИ СВЕДЕНИЙ В ГОСУДАРСТВЕННЫЙ РЕЕСТР ПРАВ НА ОБЪЕКТЫ, ОХРАНЯЕМЫЕ АВТОРСКИМ ПРАВОМ

№ 14755 от «29» января 2021 года

Фамилия, имя, отчество, (если оно указано в документе, удостоверяющем личность) автора (ов):
БЕЙСЕГУЛ ӘЛИЯ БОЛАТБЕККЫЗЫ, КАДИРКУЛОВ КУАНЫШ КАЙСАРОВИЧ

Вид объекта авторского права: **программа для ЭВМ**

Название объекта: **Лабораторная информационная система «SmartGene»**

Дата создания объекта: **10.12.2019**

Қазақстан Республикасының "Авторлық құқық" бөлімінде тексеруге болады https://copyright.kazpatent.kz

Подлинность документа возможно проверить на сайте kazpatent.kz в разделе «Авторское право» https://copyright.kazpatent.kz

Подписано ЭЦП

Осланов Е.К.

ПРИЛОЖЕНИЕ Д

Свидетельство о регистрации товарного знака

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ РЕСПУБЛИКА КАЗАХСТАН

REPUBLIC OF KAZAKHSTAN

ТАУАР ТАҢБАСЫНА КУӘЛІК
СВИДЕТЕЛЬСТВО НА ТОВАРНЫЙ ЗНАК
CERTIFICATE OF TRADEMARK REGISTRATION

№ 83957

(151) 29.05.2023
(450) 02.06.2023
(210) 111553
(220) 10.10.2022
(730) «SmartLab Kazakhstan» жауапкершілігі шектеулі серіктестігі (KZ)
Товарищество с ограниченной ответственностью «SmartLab Kazakhstan» (KZ)
«SmartLab Kazakhstan» Limited Liability Partnership (KZ)

ПРИЛОЖЕНИЕ Е

Акт внедрения

«УТВЕРЖДАЮ»
Директор
ТОО «Labstar Kazakhstan»
Сыдыкова К.Н. _____
«18» мая 2023

АКТ
внедрения результатов научно-исследовательской деятельности PhD докторанта по по направлению подготовки кадров «8D061 – Информационно-коммуникационные технологии» Образовательной программы «Аналитика больших данных» Казахского агротехнического университета им. С. Сейфуллина КАДИРКУЛОВА КУАНЫША КАЙСАРОВИЧА

Мы нижеподписавшиеся, заведующий лабораторией Батырбаева Динара Жармухановна, специалист лаборатории Алибаева Жазира Сергазиевна, специалист лаборатории Кенесбаев Махсат Кадырович составили настоящий акт о том, что результаты диссертационной работы Лабораторная информационная система SmartLab (далее – ЛИС), разработанная в рамках докторской диссертации Кадиркулова К.К., прошла производственную проверку и внедрена в процесс лабораторной диагностики.

Модули ЛИС позволяет автоматизировать весь процесс лабораторной диагностики от регистрации биоматериала, до выдачи результатов пациентам в электронном виде, со степенью защиты от подделок путем использования технологии QR кодирования. Наличие искусственного интеллекта по интерпретации результатов позволяет производить раннюю диагностику здоровья пациента, и рекомендует своевременно обратиться к врачам при возникновении патологичных результатов.

Члены комиссии:
Батырбаева Д.Ж. _____ заведующий лабораторией, к.м.н.

Алибаева Ж.С. _____ специалист лаборатории

Кенесбаев М.К. _____ специалист лаборатории

«УТВЕРЖДАЮ»
Директор ТОО «MEDCИ»
Мартынов П.С. _____
«19» 06 2023 г.

АКТ
о внедрении результатов
диссертационного исследования
КАДИРКУЛОВА КУАНЫША КАЙСАРОВИЧА

Комиссия в составе:
Председатель: директор - Мартынов Павел Серафимович
Члены комиссии: заведующей лаборатории - Явстер Светлана Сергеевна, врач-лаборант - Коппе Ирина Васильевна, инженер - Соохизо Руслан Геннадьевич

Составили настоящий акт о том, что результаты диссертационной работы «Разработка модели искусственного интеллекта по лабораторной диагностике в здравоохранении», представленной на соискание степени доктора философии (PhD) по направлению подготовки кадров «8D061 – Информационно-коммуникационные технологии», использованы в деятельности ТОО «MEDCИ» в виде:

1. Автоматизация обработки результатов лабораторных исследований путем полной интеграция с лабораторным оборудованием;
2. Предварительное выявление патологичных результатов;
3. Использование искусственного интеллекта для оценки полученных результатов с определением предварительного диагноза;
4. Интеграция с внешними системами как КМИС «Даму», МИС «Надежда» позволяет оперативно доставлять результаты пациентам, что вносит вклад в формировании электронного паспорта здоровья населения Республики Казахстан.

Использование указанных результатов позволяет: повысить качество оказания лабораторных услуг, предварительное выявление патологии у пациентов для контрольной проверки биоматериала.

Председатель комиссии
Мартынов Павел Серафимович _____ Директор

Члены комиссии:
Явстер С.С._____ Заведующий лабораторией
Коппе И.В._____ Врач-лаборант
Соохизо Р.Г._____ Инженер