

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«СИБИРСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ»
МИНИСТЕРСТВА ЗДРАВООХРАНЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

На правах рукописи

Аллилуев Александр Сергеевич

**ПРОГНОЗИРОВАНИЕ РЕЦИДИВА ТУБЕРКУЛЁЗА ОРГАНОВ ДЫХАНИЯ
С МНОЖЕСТВЕННОЙ ЛЕКАРСТВЕННОЙ УСТОЙЧИВОСТЬЮ**

3.1.26. Фтизиатрия

Диссертация на соискание учёной степени
кандидата медицинских наук

Научный руководитель:
доктор медицинских наук, профессор
Филинюк Ольга Владимировна

Томск – 2021

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.	4
ГЛАВА 1 ОБЗОР ЛИТЕРАТУРЫ.	13
1.1 Рецидив туберкулеза (в том числе с множественной лекарственной устойчивостью возбудителя): эпидемиология, причины возникновения.	13
1.2 Искусственный интеллект и методы машинного обучения: история создания и сфера деятельности.	34
1.3 Методы машинного обучения, применяемые во фтизиатрии.	39
ГЛАВА 2 МАТЕРИАЛ И МЕТОДЫ ИССЛЕДОВАНИЯ.	50
2.1 Организация исследования и схема его проведения.	50
2.2 Материал и методы исследования.	53
ГЛАВА 3 РЕЗУЛЬТАТЫ СОБСТВЕННЫХ ИССЛЕДОВАНИЙ.	58
3.1 Статистический анализ групп наблюдения по выявлению предикторов рецидива туберкулеза с множественной лекарственной устойчивостью.	58
3.2 Регрессионный анализ данных пациентов по выявлению предикторов рецидива туберкулеза с множественной лекарственной устойчивостью.	85
3.3 Алгоритмы машинного обучения в прогнозировании рецидива заболевания у больных туберкулезом с множественной лекарственной устойчивостью.	93
ЗАКЛЮЧЕНИЕ.	110
ВЫВОДЫ.	120
ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ.	122
СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ.	123
СПИСОК ЛИТЕРАТУРЫ.	127
СПИСОК ИЛЛЮСТРАТИВНОГО МАТЕРИАЛА.	150
ПРИЛОЖЕНИЕ А (справочное) Иллюстративный материал этапов работы классификатора на основе метода дерево решений.	154
ПРИЛОЖЕНИЕ Б (справочное) Иллюстративный материал этапов работы классификатора на основе метода случайный лес.	160

ПРИЛОЖЕНИЕ В (справочное) Иллюстративный материал этапов работы классификатора на основе метода XGBoost.	166
ПРИЛОЖЕНИЕ Г (справочное) Иллюстративный материал этапов работы классификатора на основе метода логистическая регрессия.	172
ПРИЛОЖЕНИЕ Д (справочное) Тестируемый набор признаков пациента № 11.	178
ПРИЛОЖЕНИЕ Е (справочное) Тестируемый набор признаков пациента № 56.	180
ПРИЛОЖЕНИЕ Ж (справочное) Программа-калькулятор для оценки риска развития рецидива у больных с множественно лекарственно устойчивым туберкулезом (фрагменты текста программы)	182

ВВЕДЕНИЕ

Актуальность избранной темы

В соответствии с данными ВОЗ туберкулез (ТБ) входит в десятку заболеваний, являющихся основными причинами смертности населения в мире. По оценкам ВОЗ в 2019 г. ТБ заболело 10 млн человек [100]. За последнее десятилетие Россия демонстрирует высокие темпы снижения заболеваемости и смертности от ТБ: число заболевших за период с 2000 по 2019 год снизилось более чем в 2 раза, а умерших от ТБ – более чем в 4 раза, что является результатом приоритетного отношения государства к данной проблеме. При этом наибольшей сложностью в противодействии ТБ в настоящее время является особенность возбудителя заболевания – его множественная лекарственная устойчивость (МЛУ), где микобактерии туберкулеза (МБТ) резистентны, как минимум, к основным противотуберкулезным препаратам (ПТП) – изониазиду и рифампицину. В мире в 2019 году у 465 000 заболевших развился МЛУ туберкулез, еще у 100 000 была выявлена устойчивость к рифампицину. По данным информационного бюллетеня ВОЗ, в Европе каждый пятый случай впервые выявленного ТБ – это МЛУ ТБ, в РФ – каждый третий [6, 198]. Несмотря на всеобщий доступ к ПТП результаты лечения в РФ остаются неоптимальными. В настоящее время показатели эффективного лечения больных с МЛУ ТБ составляют только 57 % случаев в мире, в РФ – 55 % [198]. В то же время эффективное излечение МЛУ ТБ не гарантирует, что через некоторое время вновь не будут выявлены признаки активности туберкулезного процесса (рецидива заболевания). Рецидивом ТБ считается случай у пациента, когда зарегистрирован повторный эпизод заболевания ТБ после эффективного курса химиотерапии (ЭКХТ).

В последние годы результаты исследований показывают, что рецидивы туберкулеза органов дыхания с МЛУ возбудителя (МЛУ ТОД) после успешной химиотерапии (ХТ) нередки и зарегистрированные показатели повторного заболевания при сроке наблюдения до 8 лет варьируют от 2,5 % до 18 % [3, 6, 15,

153, 190]. Ведение пациентов с МЛУ ТБ, а тем более с рецидивом заболевания представляет собой сложную задачу для клинической медицины в связи с тем, что лечение больных с МЛУ возбудителя является длительным (минимум 20 месяцев), осложняется потенциально тяжелыми побочными реакциями и является наиболее дорогостоящим, чем лечение лекарственно-чувствительного ТБ [17, 29]. Учитывая высокий уровень заболеваемости в РФ МЛУ ТБ и низкий уровень излечения, адекватность ведения больных с МЛУ ТБ требует всестороннего исследования, так как случаи рецидива МЛУ ТОД пополняют резервуар резистентной туберкулёзной инфекции, что, в свою очередь, поддерживает в стране высокий уровень инфицированности и первичной МЛУ МБТ. В связи с этим для клинициста крайне важно знать особенности, связанные с рецидивом МЛУ ТОД, и факторы, позволяющие прогнозировать, какие пациенты подвергаются наибольшему риску реактивации заболевания, чтобы провести профилактические мероприятия.

Одновременный прогресс в области инфраструктуры информационных технологий и роста мощностей компьютеров позволяет сегодня анализировать любые наборы биомедицинских данных для решения задач здравоохранения, используя современные методы анализа, такие как методы машинного обучения (МО) [8, 9, 24, 151]. Эти методы могут обнаруживать и идентифицировать внутренние закономерности и взаимосвязи из сложных и многочисленных наборов данных, в то же время они обладают эффективной прогнозирующей способностью и могут применяться для создания стандартных моделей принятия клинических решений [62, 115, 171].

Методы МО стали популярным инструментом для медицинских исследователей во всем мире, в том числе и в проблематике ТБ, особенно в прогнозировании и диагностике, а также в предсказании исходов заболевания [46, 66, 83, 94, 121, 138]. Учитывая вышеизложенное, а также важность персонализированной медицины, исследование, в котором используются методы МО в отношении выделения значимых факторов риска и прогнозирования развития рецидива МЛУ ТОД, является актуальным.

Степень разработанности темы диссертации

В настоящее время в исследованиях определены основные факторы риска развития рецидива ТОД без учета спектра лекарственной устойчивости (ЛУ) МБТ к ПТП [17, 29, 32, 33]. Закономерно возникает вопрос о причинно-следственных связях, ведущих к повторному заболеванию людей, пролеченных по схемам ХТ МЛУ ТБ. Одни авторы отводят основную роль неполноценному первому этапу лечения, частым перерывам в ХТ, по мнению других специалистов, основная причина развития рецидива – наличие исключительно остаточных изменений в легких и наличие сопутствующей патологии. Выявленные особенности не позволяют сделать окончательное заключение об их вкладе в развитие повторного заболевания и не обозначают их абсолютной значимости. Факторы риска рецидива ТОД, особенно с МЛУ возбудителя, нуждаются в дальнейшем исследовании и количественной оценке информативности и прогностичности. Вопросам алгоритмов прогнозирования посвящено значительное число исследований во фтизиатрии в разных аспектах [46, 66, 83, 138]. До настоящего времени как отечественными, так и зарубежными авторами не были предложены математические модели, которые позволили бы создать автоматизированную систему оценки риска развития рецидива у больных МЛУ ТОД для применения во фтизиатрической практике. Вышеизложенное определило цель и задачи настоящего исследования.

Цель исследования

Выявить приоритетные факторы риска развития рецидива у эффективно излеченных больных туберкулезом органов дыхания с множественной лекарственной устойчивостью и разработать с помощью алгоритмов машинного обучения методику прогнозирования.

Задачи исследования

1. Установить частоту возникновения рецидива туберкулеза органов дыхания после успешного лечения больных с множественной лекарственной устойчивостью возбудителя.

2. Охарактеризовать больных, у которых после эффективно завершеного лечения развился рецидив туберкулеза органов дыхания с множественной лекарственной устойчивостью, определить ведущие медицинские и немедицинские факторы риска повторного заболевания.

3. Оценить возможность применения алгоритмов машинного обучения в прогнозировании рецидива у больных туберкулезом органов дыхания с множественной лекарственной устойчивостью.

4. Создать программу ЭВМ для оценки индивидуального риска развития рецидива у больных туберкулезом органов дыхания с множественной лекарственной устойчивостью.

Научная новизна

В результате выполнения данного исследования были получены следующие результаты, характеризующиеся научной новизной:

1. Впервые выявлены основные социальные, клинико-рентгенологические и микробиологические факторы риска, влияющие на развитие рецидива у больных МЛУ ТОД после ЭКХТ в периоде 5-летнего наблюдения.

2. Впервые получены новые знания о частоте, сроках возникновения, характере и причинах развития рецидива ТОД у излеченных больных с МЛУ возбудителя, включающие как медико-социальные характеристики пациентов, так и индивидуальные особенности проводимого лечения, с анализом полноты схем ХТ на начало и на конец лечения, с оценкой влияния амплификации МБТ, сроков абациллирования и характера остаточных изменений в легких.

3. Впервые на основе методов МО предложена методология прогнозирования вероятности развития рецидива МЛУ ТОД: выделение значимых признаков (факторов риска), ранжирование их по вносимому вкладу, построение модели для прогнозирования риска развития рецидива заболевания.

4. Разработанное впервые методами МО программное обеспечение позволяет персонализировано оценить риск развития рецидива у эффективно излеченных больных МЛУ ТОД.

Теоретическая и практическая значимость работы

В результате исследования выявлены и систематизированы социальные, клинико-микробиологические и рентгенологические факторы риска развития рецидива у больных МЛУ ТОД после ЭКХТ. Созданная с помощью методов МО прогностическая модель позволяет предположить возможность развития рецидива у пациентов МЛУ ТБ в начале противотуберкулезной терапии, что позволит совершенствовать организацию лечения с учетом индивидуальных особенностей больных. Способность прогнозировать риск развития рецидива заболевания в ходе лечения пациентов МЛУ ТБ предупреждает их развитие в будущем. Это ведет к рациональному использованию лекарственных средств, сохранению их клинической эффективности и, соответственно, сдерживанию распространения ТОД с множественной и широкой ЛУ возбудителя, что, в свою очередь, лежит в основе новой стратегии ВОЗ по ликвидации ТБ, направленной на своевременное выявление случаев заболевания среди групп высокого риска.

Методология и методы диссертационного исследования

Данное диссертационное исследование представляет собой научное решение в повышении эффективности прогнозирования развития рецидива заболевания у эффективно излеченных пациентов МЛУ ТОД. Объект исследования – больные с рецидивом МЛУ ТОД. Предмет исследования – значимые факторы риска, оказывающие влияние на развитие рецидива после ЭКХТ МЛУ ТБ. Научная идея исследования заключается в выявлении совокупности значимых факторов риска развития рецидива МЛУ ТОД, на основе которых с применением методов традиционной статистики и методов МО получена прогностическая модель. В научной работе использован комплексный подход с применением клинико-рентгенологических и лабораторных методов обследования, экспертной оценки медицинской документации, системного и статистического анализа, математического и компьютерного моделирования, оценки информативности.

Справедливость гипотезы подтверждают результаты анализа социальных, клинических, эпидемиологических, рентгенологических, микробиологических данных ретроспективной выборки 346 пациентов МЛУ ТОД, из которой сформированы две группы. Первая (основная) – лица, у которых наступил рецидив заболевания после ЭКХТ ($n = 35$), вторая (контрольная) – лица без рецидива заболевания ($n = 311$), срок наблюдения минимум 5 лет.

Обработка и статистический анализ полученных результатов проводили с использованием программного продукта IBM SPSS Statistics версия 23.0. Модели прогнозирования рецидива МЛУ ТОД с помощью алгоритмов МО включали в себя дерево решений, случайный лес, XGBoost (градиентный бустинг) и логистическую регрессию, с использованием K-блочной стратифицированной проверки. Использовали инструменты библиотеки scikit-learn Version 0.24.2 в интерактивной облачной среде работы с программным кодом Google Colaboratory.

Положения, выносимые на защиту

1. У каждого десятого излеченного пациента туберкулезом органов дыхания с множественной лекарственной устойчивостью развивается рецидив заболевания, в среднем через два года наблюдения.

2. Факторами риска у больных туберкулезом органов дыхания с множественной лекарственной устойчивостью, определяющими значимое влияние на развитие рецидива заболевания, являются:

- социальные – пациенты пенсионного возраста, имеющие инвалидность по соматическому заболеванию, психические заболевания, табакокурение;

- медицинские – повторное лечение, неполноценность предыдущего курса химиотерапии, остаточные изменения в легких, деструктивные формы туберкулеза с наличием полостей распада в легочной ткани от 2,0 см, бактериовыделение на начало химиотерапии, определяемое микроскопией, наличие у пациентов ВИЧ-инфекции;

- микробиологические – наличие лекарственной устойчивости возбудителя на конец лечения к шести и более лекарственным средствам,

формирование амплификации микобактерии туберкулеза к противотуберкулезным препаратам и замедленные сроки прекращения бактериовыделения на фоне химиотерапии.

3. Логит-регрессионная модель прогнозирования на основе традиционного статистического анализа и прогностическая модель на основе методов машинного обучения позволяют совершенствовать лечебно-диагностический процесс, формируя тактику ведения больных туберкулезом органов дыхания с множественной лекарственной устойчивостью, с учетом имеющихся в каждом конкретном случае социальных, эпидемиологических и клинических особенностей заболевания.

4. Созданная автоматизированная система для оценки индивидуального риска развития рецидива туберкулеза органов дыхания с множественной лекарственной устойчивостью, построенная на основе классификаторов машинного обучения, позволяет обеспечить индивидуальный подход к определенному пациенту в рамках системы поддержки принятия решений в повседневной практике врача-фтизиатра.

Степень достоверности

Достоверность и обоснованность полученных результатов диссертационного исследования обеспечены применением современных методик сбора и обработки данных, корректным подбором объектов исследования, достаточным объемом выборочной совокупности, а также использованием для анализа апробированного математического аппарата с интерпретацией результатов исследования в рамках отечественного и мирового опыта.

Апробация работы

Основные результаты диссертационной работы были представлены в виде научных докладов и обсуждались на: 6-й ежегодной молодежной научной конференции с Международным участием, посвящённой Дню Российской науки (Новосибирск, 2018); 7-й ежегодной научной конференции «Современные

тенденции развития фтизиатрии», посвящённой Дню Российской науки и 100-летию со дня рождения выдающегося Сибирского фтизиохирурга профессора Андрея Илларионовича Боровинского (Новосибирск, 2019); Юбилейной научно-практической конференции «Новые горизонты фтизиатрии», посвященной 75-летию образования ФГБУ «Новосибирский НИИ туберкулеза» Минздрава России (Новосибирск, 2019); 8-й ежегодной научной конференции «Фтизиатрия сегодня и завтра», посвящённой Дню Российской науки (Новосибирск, 2020); Всероссийской научно-практической конференции молодых ученых с международным участием, посвященной Всемирному Дню борьбы с туберкулезом «Современные инновационные технологии в эпидемиологии, диагностике и лечении туберкулеза взрослых и детей» (Москва, 2020); научно-практической конференции с Международным участием «Исследования и инновации в современной фтизиатрии» (Новосибирск, 2021).

Диссертационное исследование апробировано на заседании проблемной комиссии «Внутренние болезни (пульмонология, фтизиатрия, сочетанные формы патологий внутренних органов)» ФГБОУ ВО «Сибирский государственный медицинский университет» Минздрава России (Томск, 2021).

Диссертационная работа выполнена в соответствии с утвержденным направлением научно-исследовательской работы ФГБОУ ВО «Сибирский государственный медицинский университет» Минздрава России по теме: «Фундаментальные аспекты клинической физиологии дыхания. Важнейшие звенья патогенеза острых и хронических заболеваний легких и сочетанных форм патологии». Регистрационный номер АААА-А16-116021010210-5.

Внедрение результатов исследования

Результаты работы внедрены в лечебно-диагностический процесс ОГАУЗ «Томский фтизиопульмонологический медицинский центр», в учебный процесс кафедры фтизиатрии и пульмонологии и кафедры медицинской и биологической кибернетики ФГБОУ ВО «Сибирский государственный медицинский университет» Минздрава России.

Публикации

По теме диссертации опубликовано 7 научных работ, в том числе 1 свидетельство о государственной регистрации программы для электронных вычислительных машин и 4 статьи в научных журналах, включенных в перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, из них 4 статьи в журнале, входящем в международную реферативную базу данных и систем цитирования (Scopus).

Объем и структура работы

Диссертация изложена на 190 страницах машинописного текста и состоит из введения, 3 глав, выводов, практических рекомендаций, списка сокращений и условных обозначений, списка литературы, списка иллюстративного материала и приложений. Список литературы представлен 210 источниками, из которых 158 в зарубежных изданиях. Полученные результаты проиллюстрированы с помощью 38 таблиц и 13 рисунков, в приложениях представлены важные этапы работы алгоритмов МО и часть зарегистрированной программы ЭВМ.

Личный вклад автора

Изучение проблематики и поиск литературных данных, получение исходных данных, анализ медицинской документации, планирование исследования, постановка гипотезы, цели и задач исследования, получение результатов исследования, разработка алгоритма и математических моделей, подтвержденных проверкой, обсуждение результатов исследования на научных конференциях, публикации результатов в рецензируемых научных изданиях и написание диссертации выполнены лично автором.

ГЛАВА 1 ОБЗОР ЛИТЕРАТУРЫ

1.1 Рецидив туберкулеза (в том числе с множественной лекарственной устойчивостью возбудителя): эпидемиология, причины возникновения

За последние два десятилетия достигнут значительный прогресс в глобальном контроле над туберкулезом, а именно в 2019 году успехи в диагностике, лечении и профилактике туберкулеза позволили снизить распространенность данного инфекционного заболевания в мире на 42 %, смертность на 47 %, по сравнению с уровнями 2000 года [99]. Усилия по борьбе с туберкулезом, которые привели к увеличению показателя успешности лечения во всем мире среди новых случаев с 69 % в 2000 году до 84 % в 2018 году, помогли спасти около 63 миллионов жизней на планете [203]. Но время меняется. Прошедший 2019 год был очень тяжелым не только в России, но и на всей планете. Здравоохранение всех стран не было готово к столь быстрому распространению пандемии COVID-19, и, несмотря на то, что медицинский персонал героически вел борьбу с инфекцией, пандемия повсеместно снизила актуальность многих заболеваний и поставила себя во главу угла, унося миллионы жизней людей. Туберкулез, онкология, плановая диспансеризация, терапевтические и хирургические вмешательства и др. – все отодвинулось. Мы были сосредоточены на COVID-19.

Не зря в 2021 г. Всемирный день борьбы с туберкулезом, провозглашенный ВОЗ, проходит под лозунгом «Время на исходе». Этот лозунг призывает, что у человечества истекает время, отпущенное на выполнение принятых мировыми лидерами обязательств по ликвидации эпидемии туберкулеза. Это имеет особое значение в контексте пандемии COVID-19 и увеличении доли больных МЛУ ТБ на 10 %, по сравнению с 2018 годом [100].

В Российской Федерации доля больных с МЛУ ТБ за последние 10 лет, аналогично, увеличивается, причем как среди впервые выявленных (2010 г. – 17,1 %, 2020 г. – 33,4 %), так и среди случаев повторного лечения (в 2010 –

34,7 %, 2020 – 58,8 %).

Распространенность туберкулеза с МЛУ МБТ в Томской области в 2020 году составила 2,4 на 100 тыс. населения, а заболеваемость – 5,7 на 100 тыс. населения. Доля МЛУ ТБ среди обследованных больных туберкулезом органов дыхания с бактериовыделением среди контингентов больных, состоящих на учете в 2020 году, была зарегистрирована на уровне 68,8 %: среди впервые выявленных больных – 34,0 %, среди повторных случаев лечения – 64,0 %.

Страдает и эффективность лечения больных МЛУ ТБ. Она, по данным ВОЗ, достигла в 2019 году уровня 57 %, а в РФ – 51 % [28, 47, 99, 100].

При этом не все, успешно прошедшие курс лечения, больные достигают долгосрочного периода излечения. Ранее рецидив ТБ определялся, как повторное заболевание после того, как пациент считается излеченным с использованием стандартной противотуберкулезной терапии. В нормативных документах Минздрава России это состояние расшифровывается более подробно, рецидив туберкулёза – это возобновление процесса у ранее прошедшего эффективный курс лечения больного, у которого вновь появились признаки активного туберкулёза: по клинико-рентгенологическим данным и/или положительным результатам микроскопии или посева мокроты [39]. В практике клинические, эпидемиологические и/или микробиологические данные не могут дифференцировать ситуацию с пациентами – что же произошло: рецидив или повторное инфицирование МБТ. Так как с точки зрения молекулярной биологии, различие между этими двумя состояниями требует оценки гомологии штаммов МБТ, выделенных во время первого и последующих эпизодов ТБ, с использованием технологий молекулярного генотипирования. В связи с этим есть современное определение рецидива туберкулеза. Это последующий эпизод заболевания, возникающий из-за реактивации или повторного появления исходного инфицирующего штамма МТБ, который определяется оценкой генотипической гомогенности первичных и рецидивирующих штаммов возбудителя туберкулеза. Так как практически отсутствует возможность повсеместного генотипирования МБТ, эпизоды возникновения повторного

туберкулеза регистрируются как случаи «рецидива», без отличия истинных рецидивов от случаев нового заражения [40].

Случаи рецидива являются индикатором, отражающим эффективность реализации программ по лечению туберкулеза, включаются в статистическую отчетность всех стран, и, в данный момент рассматриваются ВОЗ как отрицательный исход химиотерапии туберкулеза [40]. При изучении причин и факторов риска развития повторного туберкулеза используются понятия ранних (отмечаются в сроки от 2 до 5 лет) и поздних рецидивов (5 и более лет после успешного завершения химиотерапии). Некоторые авторы в своих исследованиях сообщают, что большинство рецидивов в рамках реактивации процесса произошли в течение первых 4 лет, в то время как риск повторного заражения был идентифицирован в течение 14 и более лет наблюдения [2, 10, 67]. Другие обнаружили, что частота рецидивов достигает своего пика в ранний период наблюдения после лечения в 70 % случаев у ВИЧ-негативных лиц, среди ВИЧ-позитивных больных – у 80 % [92, 161].

В целом в мире (без учета резистентности возбудителя МБТ) по данным публикаций последнего десятилетия регистрируется частота рецидива туберкулеза в диапазоне 1,3–30 %. По имеющейся информации ВОЗ, в 2015 г. было зарегистрировано 476 107 рецидивов, что составляло 7,7 % от всех зарегистрированных новых и повторных случаев туберкулеза [98]. При этом многие исследователи отмечают, что в странах с высокой заболеваемостью туберкулезом реинфекция является основной причиной рецидивов туберкулеза (до 90 %), напротив, в странах с низкой заболеваемостью ТБ наблюдаются истинные рецидивы заболевания. На это указывают результаты научных работ ученых из регионов с различным уровнем заболеваемости туберкулезом. Так, если опираться на высоко доказательные исследования, с использованием генотипирования изолятов образцов мокроты, определяющих для обоих эпизодов заболевания (впервые выявленные, рецидив) генотипические паттерны, то в США в большинстве случаев (93 %) регистрировались истинные рецидивы [161], а в Китае их доля составляла 52 % [165].

Закономерно, что самый низкий уровень развития рецидива наблюдается в странах Северной Америки, Западной Европы, Австралии, Новой Зеландии и некоторых азиатских странах, таких как Япония и Тайвань, где уровень распространенности туберкулеза низкий. Безусловно, это объясняется социальным благополучием и высоким уровнем жизни населения, лучшим доступом к медицинской помощи и эффективными программами борьбы с туберкулезом [93, 156, 157, 168, 180, 193, 195].

Так, в исследовании В. Afshar и соавт. (2019) были изучены все случаи ТБ жителей Лондона с 2002 по 2015 гг. с использованием молекулярно-генетической (типирование штаммов МБТ MIRU VNTR) и эпидемиологической информации. Из 43 465 случаев туберкулеза только у 618 (1,4 %) был диагностирован рецидив. При этом авторы отмечают отчетливое снижение доли рецидивов ТБ с 2002 г. (2,3 %) по 2015 г. (1,3 %) [180].

В крупных городах Европы, США и Канады – аналогичная ситуация. Так, в Барселоне из 1 823 бактериологически подтвержденных случаев ТБ у 13 (1,3 %) были рецидивы заболевания. Частота рецидивов составила 341 случай на 100 000 человек в год, что в 13 раз выше, чем заболеваемость туберкулезом среди населения в целом. Вероятность рецидива туберкулеза на 1-м, 3-м и 5-м году наблюдения составила 0,1 %, 1,4 % и 1,6 % соответственно [193]. Там же в Испании, чуть позже в Каталонии рассчитали частоту рецидивов ТБ за период 2000–2016 года на человеко-годы через 1, 2, 5 и 10 лет наблюдения. Из 839 пролеченных от ТБ пациентов, 24 развили рецидив (2,9 %), что составило 0,49 на 100 человеко-лет. Вероятность рецидива составила 0,63 % через 1 год наблюдения, 1,35 % через 2 года и 3,69 % через 5 лет наблюдения [195].

В исследовании, которое было проведено на национальном уровне в США [93], Kim L. и соавт. подтвердили низкую частоту рецидивов (1,7 %). При этом авторы обнаружили интересные социальные различия больных. Среди пациентов, родившихся в США, рецидивы были связаны с возрастом от 25 до 44 лет и употреблением психоактивных веществ. Среди эмигрантов, переболевших туберкулезом, рецидивы возникали чаще у ВИЧ-инфицированных

в возрасте 45–64 лет, и, как правило, первоначальная форма туберкулеза органов дыхания была у них с распадом и с положительным мазком мокроты.

В Австралии зарегистрирован самый низкий показатель повторного заболевания. Так, в период с 2002 по 2014 год было пролечено 4 766 больных ТБ и только у 32 (0,7 %) развился рецидив. Специалисты отмечают, что только индивидуально подобранная социальная поддержка приверженности лечению пациентов, направленная на ежедневный прием ПТП и усиленный самоконтроль, приводит к высокой эффективности химиотерапии туберкулеза и безрецидивного течения заболевания [157].

Вышеописанную информацию по поводу рецидивов заболевания в эпоху ликвидации туберкулеза подтверждает систематический обзор, опубликованный A. Rosser и соавт. в 2018 году, где включены 44 статьи, 71 % из которых от ученых из стран с высокоразвитой экономикой. Авторы констатируют, что в исследованиях со средней продолжительностью наблюдения 7,8 лет (с диапазоном от 2 до 33 лет) доля больных туберкулезом, у которых возник эпизод рецидива после завершения лечения, составила 3,4 %. В основном это доказанная эндогенная реактивация – 81 % [168].

Немного другая ситуация в азиатских странах. В исследовании из Вьетнама Bestrashniy J. и соавт. показали, что из 505 пациентов 24 (4,7 %) развили рецидив заболевания через 1–2 года после успешного лечения туберкулеза [156]. Исследование, проведенное в Йемене, показало, что 71 из 751 (9,5 %) больных туберкулезом с положительным мазком мокроты испытали рецидив заболевания после успешного лечения [152]. Аналогичная ситуация в Иране [186].

Интересные, на наш взгляд, исследования по поводу рецидивов ТБ из регионов с высоким уровнем заболеваемости и низкой распространенностью случаев ТБ и ВИЧ-инфекцией. В Китае, в Шанхае провели ретроспективное когортное исследование за 12-летний период наблюдения всех бактериологически подтвержденных случаев ТБ (13 417), которые были успешно вылечены. В целом, у 710 (5,3 %) были диагностированы рецидивы, что составило 7,55 (95 % ДИ 7,01–8,13) эпизодов на 1 000 человеко-лет, что в 18 раз превышает

заболеваемость ТБ среди населения в целом [187]. Позже, ученые из этой же страны, определили, что заболевание рецидивировало у 275 из 4 043 успешно пролеченных больных туберкулезом, что составило 6,8 % (275/4 043). Кроме того, на основании результатов генотипирования для обоих эпизодов заболевания (впервые выявленные, рецидивы), в 69 % случаев доказаны идентичные генотипические паттерны [197]. Одно из самых последних генетически-популяционных исследований случаев туберкулеза с положительной культурой мокроты, с использованием MIRU VNTR, проводилось в Китае в провинции Цзянсу в период с 2013 по 2019 гг. В общей сложности были собраны 1 451 изолятов культуры мокроты, у 30 (2,06 %, 30/1 451) больных ТБ были зарегистрированы повторные эпизоды заболевания. Результаты Китайских исследователей определили уже описанную выше современную закономерность – в 52,17 % случаях наблюдались истинные рецидивы, в 47,83 % – повторное инфицирование [165]. Аналогичное по дизайну исследование проходило в Индии. Из 1 565 впервые выявленных больных с положительным мазком мокроты, 1 207 (77 %) были эффективно излечены, у 158 развивался рецидив туберкулеза. Общая доля рецидивов ТБ в исследовании составила 10,9 %, а частота рецидивов ТБ на 100 человеко-лет составила 12,7, от 5,4 до 30,5 лет. Эндогенная реактивация наблюдалась в 93 % случаев [155].

В странах с высоким уровнем ВИЧ и туберкулеза частота рецидивов значительно выше. По информации литературных источников рецидивирующий туберкулез у лиц с коинфекцией ВИЧ/ТБ возникает в диапазоне от 10 до 30 %, причем у большинства повторное заболевание возникает в течение 2 лет после успешного завершения лечения [74, 108, 140, 148, 194].

Что касается России и стран постсоветского периода, то обобщенной информации когортных исследований, по сравнению с публикациями в мире по данной тематике минимум.

Опираясь на официальные статистические материалы основных показателей противотуберкулезной деятельности в Сибирском и Дальневосточном федеральном округе, подготовленных сотрудниками Новосибирского

научно-исследовательского института туберкулеза, необходимо заключить, что уровень рецидивов туберкулеза в Российской Федерации в 2019 г. снизился по сравнению с 2012 г. (8,3) на 20,5 % и достиг 6,6 на 100 000 населения (ф. № 33) [28]. Несмотря на общую для России позитивную тенденцию, уровень рецидивов туберкулеза в ряде регионов остается высоким, коррелируя с высоким уровнем заболеваемости в них. Так, в Республике Тыва, в Чукотском АО, в Еврейском АО по последнему отчету самый высокий показатель в мире – 41,5 (был 53,3), 29,9 (12,5), 26,9 (40,3) на 100 000 населения, соответственно. В долевого выражении это 30–40 % больных, которые были эффективно излечены (III группа диспансерного учета).

В странах бывшего СССР, по имеющейся информации, уровень рецидивов, аналогично, выше общемирового и достигает: в Молдове – 16,9 % [49], Казахстане – 14,9 % [4], Узбекистане – 9,9 % [158], Украине – 18,7 [192].

Иная и неоднородная ситуация в мире складывается с рецидивами у успешно пролеченных больных МЛУ ТБ. В таблице 1 представлены опубликованные основные по данной тематике исследования за 2010–2020 гг.

Анализируя табличную информацию необходимо отметить, что данные по эпидемиологии рецидивов МЛУ ТБ немногочисленные. Рецидивы у ранее пролеченных пациентов с МЛУ ТБ представлены в ретроспективных когортных исследованиях, в среднем с частотой от 5 до 30 %, в основном в регионах с высоким уровнем МЛУ ТБ (Индия, Китай и Российская Федерация, Южная Африка).

Таблица 1 – Частота рецидивов у больных МЛУ ТБ: анализ публикаций за 2011–2020 гг.

Автор, год публикации (страна)	Период наблюдения	Количество пролеченных	Средний возраст, лет	Сроки наблюдения, мес.	Рецидив (n, %)
Leung E. C. C. et al., 2011 (ГонгКонг) [174]	1997–2006	187	нет данных	24	8 (4,3)
Blondal K. et al., 2012 (Эстония) [79]	2001–2003	129	42,9	98,1	11 (11,5)
Anderson L. F. et al., 2013 (Великобритания) [189]	2004–2007	145	мода 15–44	48	1 (0,7)
Franke M. F. et al., 2013 (Перу) [60]	1999–2002	402	27,5	40,5	26 (6,5)
Aung K. J. et al., 2014 (Бангладеш) [179]	2005–2011	362	мода 31–42	60	4 (0,9)
Piubello A. et al., 2014 (Нигерия) [104]	2008–2010	49	31	48	0
Гайда А. И. и соавт., 2014 (Россия) [6]	2005–2009	782	нет данных	60	41 (5,2)
Елипашев А. А. и соавт., (Россия) 2015 [3]	1992–2004	161	нет данных	более 48	18,01
Kuaban C. et al., 2015 (Камерун) [105]	2008–2011	100	33,7	24	0
Kwak et al., 2015 (Южная Корея) [75]	2006–2010	105	37	38,7	2 (1,9)
Khan F. Ahmad, 2016 (Россия) [61]	2000–2004	399	нет данных	42,4	27 (6,8)
Chen M. Y. et al., 2017 (Тайвань) [153]	2007–2010	295	48	50	10 (2,95)
Sun Y. et al. (Китай), 2017 [112]	2001–2002	234	49	108	69 (29,5)
Saldanha N., et al., (Индия), 2019 [106]	2014–2016	200	41,5	36	18 (8,8)
Zheng X. B. et al., (Китай), 2020 [190]	2014–2015	58	45,4	36	1 (5,8)
Piubello A. et al., 2019 (Нигерия) (короткий курс) [128]	2008–2016	1044	нет данных	12	4 (1,6)
Марьяндышев А. О. и соавт, 2019 (Россия) (короткий курс) [15]	2015–2019	180	39,5	48	6 (4,7)
Schwœbel V. et al., 2020 (девять африканских стран) (короткий курс) [142]	2013–2017	1 006	33,9	24	14 (9,6)

Sun Y. и соавт. (2017) провели ретроспективное исследование в Китае для изучения долгосрочных результатов химиотерапии больных туберкулезом, завершивших лечение, с особым акцентом на рецидивы МЛУ ТБ. В когорте исследуемых пациентов около трети (29,5 %, 69/234) развили рецидив после завершения лечения. Общая частота рецидивов пациентов с МЛУ ТБ составила 65/1 000 пациенто-лет. Среднее время возникновения рецидива среди пациентов с МЛУ ТБ составило 5,7 лет [112]. В России коллектив ученых во главе с Елипашевым А. А. (2015), изучал отдаленные результаты лечения больных ТБ, которым, в том числе, применялась резекционная хирургия. Частота развития рецидивов среди больных с лекарственно-чувствительным и лекарственно-устойчивым туберкулезом была достоверно разная – 18, и 6,71 % соответственно [3].

В трех последних исследованиях, которые представлены в таблице 1, определена доля рецидивов при лечении больных МЛУ ТБ короткими курсами с применением в схемах лечения новых ПТП: бедаквилина, линезалида и клофазимина. Согласно наблюдениям безрецидивного исхода лечения больных МЛУ ТБ, частота развития повторного случая туберкулеза на континентах различная (от 4,7 до 9,6 %), зависит от уровня ко-инфекции с ВИЧ [15, 128, 142].

По данным одного из последних метаанализов индивидуальных данных 12 030 больных МЛУ ТБ, опубликованных в 50-ти исследованиях, датированных 2017 годом, из 25 стран – 7 346 (61 %) имели успешное лечение, 1 017 (8 %) имели неудачу или рецидив и 1 729 (14 %) умерли от туберкулеза. Неудачи и рецидивы были объединены, потому что только в 16 исследованиях были возможности мониторинга пациентов на предмет рецидивов после завершения лечения [77]. Согласно последнему сводному руководству ВОЗ по лечению МЛУ ТБ, где опубликован подобный метаанализ за 2018 год, в который вошли данные по 13 104 случаям МЛУ ТБ, полученные в рамках 53 исследований, проведенных в 40 странах, где многие пациенты находились на режимах лечения, содержащих бедаквилин, уже отдельно представлена информация по рецидивам. Результаты исследования показали, что рецидив произошел у 3,3 % пациентов со сроком

наблюдения более 12 месяцев [40]. Таким образом, анализ литературных источников показал, что в большинстве государств не ведется специальной регистрации рецидивов МЛУ ТБ, в некоторых регионах значительная часть данных случаев не классифицируется ни как новые, ни как ранее пролеченные. Из-за недостатка данных фактически невозможно оценить сравнение рецидивов МЛУ ТБ среди ВИЧ-позитивных и ВИЧ-отрицательных лиц. В связи с этим ситуация по вкладу рецидива резистентного туберкулеза в глобальное бремя туберкулеза недооценена.

Факторы риска развития рецидива туберкулеза хорошо изучены, представлены в главах профильных учебников и в специализированных руководствах, освещены во многих публикациях. Отдельно научных работ по развитию рецидива МЛУ ТБ в доступных источниках не нашлось. В целом предрасполагающие признаки классифицируются по трем факторам риска: связанные с организацией и особенностями лечения случая туберкулеза, сопряженные с индивидуальными характеристиками процесса в легких и окружающей средой.

В таблице 2 на основе обобщенных публикаций, определены критерии каждого из предрасполагающих признаков рецидива туберкулеза.

Таблица 2 – Факторы риска развития рецидива туберкулеза

Организация и особенности лечения туберкулеза	Характеристика туберкулезного процесса и индивидуальная предрасположенность	Факторы внешней среды, способствующие реинфекции
Недоступные методы определения всего спектра чувствительности МБТ к ПТП (дефекты лабораторной службы)	Туберкулезный анамнез: перенесенное ранее заболевание с остаточными изменениями в легких	Высокая популяционная распространенность туберкулеза
Скрытая лекарственная устойчивость МБТ к ПТП	Большая распространенность процесса с наличием деструкции	Загрязнение атмосферы

Продолжение таблицы 2

Организация и особенности лечения туберкулеза	Характеристика туберкулезного процесса и индивидуальная предрасположенность	Факторы внешней среды, способствующие реинфекции
Режим химиотерапии не соответствует спектру ЛУ МБТ (не адекватная схема лечения)	Заражение штаммами, которые имеют повышенную склонность к лекарственной устойчивости (Beijing)	Низкий уровень инфекционной безопасности в больницах, тюрьмах, специализированных клиниках
Несоответствие дозы ПТП к массе тела	Сохранение бактериовыделения через 2 месяца лечения	Неудовлетворительная работа в очагах туберкулезной инфекции
Короткий курс химиотерапии	Пожилой возраст	Перенаселенность или плохие жилищные условия
Плохая обеспеченность ПТП, особенно новыми	Дефицит массы тела	Отсутствие профилактических осмотров
Низкое качество ПТП	Иммунодефициты и ВИЧ-инфекция с низким количеством CD4+	
Особенности фармакодинамики и фармакокинетики ПТП (низкая проникающая способность, извращенный метаболизм лекарственных средств, не стабильные концентрации в очаге инфекции)	Сахарный диабет, хроническая болезнь легких, другие заболевания, влияющие на местную защиту и общую резистентность	
Низкая приверженность лечению	Социальные факторы риска, табакокурение, алкоголизм, употребление психоактивных веществ	

Дефекты в лечении каждого нового случая туберкулеза являются основной предпосылкой развития в будущем рецидива заболевания. Неадекватные схемы лечения с неправильным выбором ПТП, режимы с низкой бактерицидной активностью этиотропных препаратов и недостаточной продолжительностью, плохая приверженность лечению с нерегулярным приемом лекарственных средств, неправильно определенная лекарственная устойчивость возбудителя туберкулеза или нераспознанная резистентность МБТ (отсутствие полного тестирования лекарственной чувствительности) – все это способствует неполному бактериологическому излечению, развитию амплификации и рецидиву заболевания [166, 191, 201].

Необходимо особенно подчеркнуть роль плохой приверженности к противотуберкулезному лечению. В совокупности с асоциальными особенностями больных туберкулезом, способствующими пропуску приема ПТП, они ведут к увеличению продолжительности лечения, изменению схем ПТП, формированию и усилению лекарственной устойчивости МБТ и затягиванию сроков абациллирования. В. Afshar и соавт (2019) подчеркнули в своем недавнем исследовании, что рецидивы чаще встречались среди недавних мигрантов и лиц с социальными факторами риска туберкулеза [180]. На высокую долю малообразованных пациентов, имеющих минимальный прожиточный уровень в группе пациентов, развивающих рецидив туберкулеза, указывают многие исследователи [50, 154, 156, 180, 187]. Низкий социально-экономический статус, как сообщается в научных трудах, является значимым фактором риска рецидива туберкулеза. Так, в исследовании, проведенном на Тайване, было установлено, что люди, зарабатывающие менее 605 долларов США в месяц, подвержены большему риску заболеть туберкулезом в сравнении с зарабатывающими 954 доллара и более в месяц [110]. Другие социально-экономические детерминанты развития рецидива заболевания включают безработицу [55], отсутствие собственного транспорта [160], проживание в мегаполисе [193]. В дополнение к вышесказанному, есть ряд исследований, которые свидетельствуют, что иммиграция также связана с рецидивом туберкулеза [90, 202]. Кроме того,

сообщается, что некоторые этнические группы подвергаются более высокому риску рецидива. К ним относятся коренные и некоренные группы населения [58, 92, 180].

Что касается вопроса о связи контролируемости приема ПТП, приверженности лечению и рецидива, D. Menzies и соавт. (2009) сообщили, что бесконтрольный курс стандартной 6-месячной химиотерапии ведет к прерыванию лечения и увеличивает риск рецидива ТБ [88]. M. Anaam и соавт. (2012) выявили связь между пропусками доз более 80 % в фазе продолжения лечения и риском рецидива (ОШ = 25,7; 95 % ДИ 2,2–297,9) [55]. Ассоциации между контролируемым приемом в сравнении с самостоятельным приемом ПТП пациентами и риском рецидива были оценены двумя исследованиями. H. El-Sahly и соавт. (2004) обнаружили, что контролируемая терапия туберкулеза была связана с более низким уровнем рецидива (ОШ = 0,12; 95 % ДИ 0,06–0,23) [160], тогда как L. Kim и соавт. (2016) не обнаружили данной связи [93].

Вероятность успеха режимов лечения пациентов, особенно длительных курсов с МЛУ/РУ возбудителя зависит, как уже было отмечено, как от пациента, так и от характеристик инфицирующего штамма (есть данные, что рецидив туберкулеза был в значительной степени связан с инфицированием МБТ генотипа Beijing [162, 196]), а также от доступности медицинской помощи и обеспеченности больных в достаточном количестве эффективными ПТП. Известно, что длительные режимы лечения с оптимальным количеством эффективных препаратов повышают вероятность выздоровления и снижают риск смерти у взрослых и детей [172, 173]. Состав длительных режимов определяется выбором препаратов, основанных на тесте лекарственной чувствительности МБТ к ПТП, признанных эффективными для конкретного пациента, а также необходимостью комбинирования достаточного количества лекарственных средств, позволяющих максимально увеличить вероятность безрецидивного лечения без повышения токсичности. Режимы могут иметь стандартизованный (фиксированный) состав или быть индивидуально адаптированными. Продолжительность длительных режимов обычно составляет 18–20 месяцев и

более [40]. Адекватными (в англоязычной литературе – агрессивными) схемами лечения больных МЛУ ТБ считают, если в интенсивной фазе включатся пять эффективных препаратов, используемых в течение не менее 6 месяцев после конверсии мокроты, а в фазе продолжения – не менее четырех. В ретроспективном когортном исследовании взрослых пациентов с МЛУ ТБ F. Ahmad Khan и соавт. (2016) доказали, что у пациентов, которые находились на длительном и агрессивном режиме лечения, частота рецидивов была ниже [61]. Следующее обсервационное исследование тоже указывает, что данная тактика лечения в длительном режиме применения ПТП у больных МЛУ ТБ значительно может снизить частоту рецидивов. F. Qin и соавт. (2016), наблюдая за пациентами с легочным туберкулезом из Калифорнии в 2004–2009 гг., определили зависимость между длительностью курса лечения и развитием рецидива. Они обнаружили, что из 5 680 эффективно пролеченных больных туберкулезом, 483 (8,5 %) имели повторное заболевание, из них у четверти (111 (23 %) было укороченное лечение (менее 270 дней) [150]. При этом немаловажную роль в завершении всего курса химиотерапии играют программы по контролю и повышению приверженности лечению. На это указывают исследователи из России и Тайваня [61, 126, 153].

Другая проблема, связанная с эффективностью лечения и развитием рецидива, касается полной доступности и обеспеченности больных ПТП, включая их качество. Она в основном присутствует в эндемичных странах по ТБ, с низким уровнем финансирования здравоохранения [166].

В причинах неудач лечения туберкулеза и развитии рецидива важно учитывать роль фармакодинамики и фармакокинетики лекарственных средств. Общеизвестный факт, что концентрация и фармакокинетика ПТП у пациентов находится в индивидуальной вариабельности, связана с генетическими особенностями функционирования ферментов, метаболизирующих лекарственные средства и белках-переносчиках. Дозировка противотуберкулезных препаратов зависит от массы тела пациента, она обозначена во всех национальных руководствах по лечению ТБ. При превышении

ее из-за токсичности развиваются побочные реакции, а при использовании субоптимальных концентраций, развивается лекарственная устойчивость возбудителя и, как следствие, – рецидивы [78, 130, 135].

Индивидуальные характеристики пациентов, особенности туберкулезного процесса в легких являются независимыми от применяемого лечения факторами риска рецидива. Прежде всего, это возрастно-половые различия и антропометрические данные пациентов. Есть данные, что у пациентов с массой тела 50–69 кг низкие риски развития повторного заболевания, чем у пациентов с массой тела менее 50 кг (ОШ = 0,53; 95 % ДИ 0,33–0,85) [93]. В дополнение в другом исследовании сообщалось, что потеря веса более 10 % от нормы на момент постановки диагноза была связана с высоким риском рецидива ТБ (ОШ = 7,2; 95 % ДИ 2,4–21,9) [152]. А. Khan и соавт. (2006) в течение 2-х лет наблюдали за 857 пациентами, у 61 из 857 (7,1 %) возник рецидив. Риск рецидива был высоким при индексе массы тела пациентов менее 18,5 кг/м² на момент постановки диагноза. Кроме того, на риск рецидива влияла неспособность набрать более 5 % массы тела в течение первых 2 месяцев лечения, что независимо было показано в трех исследованиях [120].

Имеются некоторые противоречия в вопросе о связи между возрастом и повторным развитием заболевания. J. P. Crofts и соавт. (2010) обнаружили, что у детей в возрасте до 14 лет риск рецидива был ниже, чем у пациентов в возрасте 15–44 лет (ОШ = 0,37; ДИ 95 % 0,22–0,62) [58]. Golub J. E. и соавт. (2008) сообщали, что у пациентов 40–49 лет риск повторного заболевания был ниже, чем в возрасте 30 лет (ОШ = 0,47; 95 % ДИ 0,25–0,89) [159], и, наоборот, согласно L. Kim и соавт. (2016), пациенты в возрасте более 65 лет подвергались большему риску повторного заболевания, чем в возрасте 45 лет (ОШ = 1,88; 95 % ДИ 1,21–2,92) [93]. Гусейнов Г. К., Гиреев Т. Г. (2016) определили возраст пациентов с поздними рецидивами. В 70 % случаев он был более 50 лет [10]. Q. Chen и соавт. (2019) в своем исследовании с помощью MIRU-VNTR обнаружили у 23-х (15,2 %) из 151 пациента с РУ ТБ рецидив, тогда как в группе с чувствительностью к рифампицину только в 5,1 % случаях (у 24 из 466).

Сравнивая факторы риска, авторы подтвердили зависимость рецидива у больных РУ ТБ от возраста, а также от уровня образования, ВИЧ-статуса, наличия полости в легких, пола и употребления алкоголя [154]. Принадлежность к мужскому полу во многих исследованиях является сильным независимым предиктором рецидива туберкулеза. Так, С. L. Hung и соавт. (2015) обнаружили, что, у мужчин вероятность рецидива в два раза выше (ОШ = 2,23; 95 % ДИ 1,40–3,53), чем у женщин [110]. J. P. Millet и соавт. (2013), аналогично, отметили гендерную предрасположенность к повторному заболеванию у мужчин (ОШ = 4,3; 95 % ДИ 1,3–14,6) [193]. Они же предположили, что наблюдаемая связь между мужским полом и рецидивом может быть связана с влиянием других факторов риска, присущих мужчинам, например, с таким фактором, как курение табака [193].

Доказан факт в систематическом обзоре, что табачный дым неизменно связан с повышенным риском заражения туберкулезом. По сравнению с некурящими, у курящих пациентов увеличивается риск развития активного туберкулеза и смертности [122, 176]. Кроме того, табакокурение увеличивает и частоту рецидивов [68, 177]. С. Leung и соавт. (2015) провели проспективное исследование в Гонконге с 2001 по 2003 года и наблюдали за пациентами для определения результатов лечения. Впоследствии пациентов отслеживали через регистр уведомлений о туберкулезе на предмет рецидива до конца 2012 года. Авторы обнаружили у больных на исходном уровне ассоциативную связь между курением и обширным деструктивным поражением в легких с выраженным бактериовыделением. Данные свидетельствуют о том, что курение отрицательно влияет на тяжесть и распространенность заболевания, что затягивает процесс прекращения бактериовыделения, влияет на исход лечения и рецидивы [176]. Исследование D. Ahmad и соавт. (2016), проведенное в Пакистане, подтвердило учащение рецидивов туберкулеза легких у курящих мужчин [68]. При этом в метаанализе, где были рассмотрены 17 исследований, посвященных курению и рецидивам ТБ, M. Underner и соавт. (2016) было определено, что активное курение увеличивает риск рецидива не только легочного, но и внелегочного

туберкулеза. Кроме того, они обнаружили, что курильщики с меньшей вероятностью будут придерживаться противотуберкулезного лечения [177]. Равно как пациенты, злоупотребляющие алкоголем и психоактивными веществами. На это указывают авторы из различных стран [68, 110, 191], а недавно проведенный метааналитический обзор (2020) влияния употребления алкоголя на результаты лечения туберкулеза доказал, что существует высокий риск неблагоприятных исходов лечения (включая рецидивы), как при лекарственной чувствительности МБТ, так и при МЛУ [184]. В РФ результаты работы С. Л. Плиевой (2013) подтверждают, что в группе рецидивов сосредоточены высокие доли больных, злоупотребляющих алкоголем, не имеющих постоянной работы, инвалидов со стойкой утратой трудоспособности в связи с тяжелыми сопутствующими заболеваниями [31, 32, 33].

Доминирующим коррелятом в рисках развития повторного случая рецидива является распространенный и деструктивный туберкулезный процесс в легких, ведущий к формированию больших остаточных изменений. Наличие двухстороннего поражения легких и полостей распада до начала лечения, оцененное с помощью рентгенографии грудной клетки, было тесно связано с рецидивом в исследовании M. S. Anaam и соавт. (2012) (ОШ = 6,1; 95 % ДИ 2,2–16,9) [55]. Подобную ассоциацию между площадью поражения туберкулезным процессом легочной ткани и развитием в будущем рецидива обнаружили многие исследователи, причем с различным ВИЧ-статусом [50, 110, 185].

Несмотря на то, что систематический обзор и метаанализ, проведенный в 2010 году D. J. Horne и соавт., с целью оценки точности ассоциации между наличием у пациентов до начала лечения положительных результатов мазков или посева мокроты и неэффективностью терапии или наступлением рецидива, выявил низкую чувствительность и умеренную специфичность, все-таки на результаты современных когортных исследований следует обратить внимание [178]. M. Moosazadeh и соавт. (2015) ретроспективно определили рецидив в когорте из 1 271 пациента в Иране в 8,3 % случаях. Частота рецидивов была

значительно выше у курильщиков с обильным бактериовыделением по микроскопии мазка мокроты до начала лечения, и у тех пациентов, у которых мазок на кислотоустойчивые микобактерии (КУМ) продолжал оставаться положительным в конце 2-го месяца лечения [186]. G. Muzanyı и соавт. (2019) в проспективном когортном исследовании установили, что из 630 больных туберкулезом с подтвержденным посевом в постлечебном наблюдении в течение восемнадцати месяцев 57 (9 %) заболели рецидивирующим ТБ. Из них при однофакторном анализе данных было установлено, что до лечения у 4,7 % мазок мокроты был с низкой степенью бактериовыделения (1+), у 8,8 % – с высокой (3+) (ОШ 0,31; 95 % ДИ 0,10–0,93) [147]. Аналогично, по данным P. G. T. Cudahy и соавт. (2020) из 500 впервые выявленных больных 333 (79 %) успешно завершили лечение туберкулеза, у 35 пациентов (11 %) был бактериологически подтвержденный рецидив заболевания. На основе модели пропорциональных рисков Кокса авторы установили, что мазок мокроты КУМ 3+ был достоверно связан с повторным заболеванием (на основе генотипирования МБТ) с ОШ = 3,33; 95 % ДИ 1,44–7,7 [80]. Таким образом, есть сведения, доказывающие зависимость между массивностью бактериовыделения до начала лечения и повторным заболеванием.

Разумеется, что больные с рецидивами туберкулеза легких представляют большую эпидемическую опасность ведь доля фиброзно-кавернозного туберкулеза, казеозной пневмонии в их структуре составляет, соответственно, 3,8 % и 7,1 % [2]. На высокую частоту распространенных тяжелых клинических форм с деструкцией легочной ткани, таких как казеозная пневмония, фиброзно-кавернозный, цирротический туберкулез при рецидивах указывают многие, особенно российские исследователи [18, 19, 26, 31, 32, 33]. Так, например, Л. Г. Тарасова, в своем исследовании установила, что ограниченные процессы у данных пациентов встречаются лишь в 9,1 % эпизодах, а деструкция – в 84,8 %. Достоверно чаще выявляется при рецидивах фиброзно-кавернозный туберкулез [43].

В нескольких исследованиях изучалось влияние лекарственной

устойчивости до лечения на частоту рецидивов туберкулеза [112]. Было установлено, что резистентность к пиразинамиду может быть связана с рецидивом туберкулеза. Как сообщается, среди успешно пролеченных пациентов с МЛУ ТБ, профили резистентности пре-ШЛУ ТБ и ШЛУ ТБ (ОШ = 7,3; 95 % ДИ 1,2–44,0) сопряжены с рецидивом заболевания [153]. Интересные данные были получены Плиевой С. Л. [32]. Автор установила, что ранние рецидивы отличались от поздней большей частоты множественной и широкой лекарственной устойчивости МБТ. Из числа бактериовыделителей (520) с ранними рецидивами туберкулеза органов дыхания 113 (65,7 %) пациентов имели лекарственную устойчивость МБТ, в том числе МЛУ в 60,2 % случаях, полирезистентность в 20,3 %, ШЛУ – 12,4 % и монорезистентность только в 7,1 %. Среди бактериовыделителей с поздними рецидивами 148 (42,5 %) пациентов имели лекарственную устойчивость МБТ, в том числе множественную лекарственную устойчивость (52,7 %), полирезистентность (30,5 %), монорезистентность (12,1 %) и широкую лекарственную устойчивость (4,7 %). Позже Алексеенко С. Н., Дробот Н. Н. (2019), аналогично, подтвердили ситуацию с больными из группы с рецидивом туберкулеза в РФ: у 62,3 % бактериовыделителей с ранними рецидивами туберкулеза легких установлена ЛУ МБТ, в структуре лекарственной устойчивости МЛУ и ШЛУ составили 58,1 % и 12,7 % соответственно. В группе пациентов с поздними рецидивами доля пациентов с ЛУ МБТ среди бактериовыделителей составила 43,7 %, в том числе МЛУ МБТ – 26,7 %, ШЛУ МБТ – 7,3 % [2].

Распространенность посттуберкулезных изменений в легких, наличие остаточных полостей после завершения курса лечения являются предпосылками в активизации эндогенной туберкулезной инфекции в будущем. Именно на этом основании лиц с неактивными посттуберкулезными остаточными изменениями в лёгких относят к группам с повышенным риском рецидива. Как сообщают российские и постсоветские специалисты, пациенты с рецидивом туберкулеза легких в 80–90 % случаев имеют большие остаточные изменения после излеченного в прошлом туберкулеза [4, 5, 10, 11, 16, 22, 30, 38, 43, 44, 45].

Обобщая результаты данных публикаций, определяем клинический сценарий развития события: у большинства больных (90 %) рецидивы заболевания выявлены по обращаемости при появлении клинических симптомов туберкулеза; повторное заболевание возникает в течение первого года после окончания курса лечения в 10–15 % случаев, через 2–5 лет – у, примерно, 30 %, через 6–10 лет – у 20 %, наиболее часто возобновление туберкулеза отмечается у 35 % по истечении 10 лет и более. Рецидивы возникали в первый раз у 80–90 % пациентов, второй-третий эпизод заболевания – у 3–8 %, четвертый и более – в 2 % случаях.

О влиянии сопутствующей патологии на развитие рецидива туберкулеза в обзоре уже упоминалось. На первом месте стоит ВИЧ-инфекция 1 типа, которая во многих исследованиях ассоциировалась с рецидивом [74, 76, 88, 106, 108, 132, 185, 191, 194, 196]. При этом авторы утверждают, что связь между ВИЧ и реинфекцией присутствовала даже в условиях с низкой заболеваемостью туберкулезом. А. С. Pettit и соавт. (2011) обнаружили, что ВИЧ-инфекция была связана с рецидивом вследствие реинфекции, а не вследствие эндогенной реактивации [76]. Однако другое исследование не смогло подтвердить данную особенность [91]. Многие исследователи, включая российских, ассоциировали повторное заболевание со степенью иммуносупрессии. Количество CD4-лимфоцитов в периферической крови ассоциируется с риском рецидива. J. E. Golub и соавт. (2008) выявили, что уровень CD4+ лимфоцитов в диапазоне от 200 до 349 клеток/мм³, определенный во время постановки диагноза ТБ, был связан с более низким риском рецидива, чем данный показатель ниже 200 клеток/мм³ (ОШ = 0,35; 95 % ДИ 0,20–0,60) [159]. О том, что число рецидивов достоверно возрастает у лиц, живущих с ВИЧ на фоне выраженной иммуносупрессии при уровне CD4+ < 200 клеток/мкл, а диагностика рецидивов туберкулеза у этого контингента имеет свои особенности и отличается от таковой у пациентов с ВИЧ-негативным статусом, указывают и российские ученые [41, 42]. В частности, А. М. Сенин (2018) установил, что среди пациентов с ВИЧ-инфекцией с уровнем CD4+ ≥ 350 клеток/мкл на момент констатации клинического излечения безрецидивная длительность достоверно не отличалась

от таковой у ВИЧ-негативных пациентов [42]. Им же, а также другими авторами было изучено влияние антиретровирусной терапии на рецидив туберкулеза. J. E. Golub и соавт. (2008) показал, что у пациентов, получающих АРТ, вероятность рецидива была ниже (ОШ = 0,5; 95 % ДИ 0,28–0,89) [41, 159]. Российский ученый заключил, что уровень $CD4+ \geq 350$ клеток/мкл, а также эффективная АРТ являются благоприятными прогностическими факторами длительности ремиссии при туберкулезе и могут быть использованы как дополнительные критерии при определении показаний к констатации клинического излечения туберкулеза у пациентов с наличием ВИЧ-инфекции. В двух ретроспективных обсервационных когортных исследованиях изучалась связь между продолжительностью противотуберкулезной терапии и риском развития рецидива у ВИЧ-инфицированных пациентов. Химиотерапия более 9,5 месяцев на фоне предшествующей АРТ показала связь с более низким риском рецидива [163, 164]. Кроме исходного количества $CD4+$ лимфоцитов у пациентов с ко-инфекцией с рецидивом туберкулеза ассоциируется также вирусная нагрузка, положительный туберкулиновый кожный тест, неоднократные эпизоды заболевания, обширный туберкулезный процесс в легких с поражением более трех сегментов и анемия. Интересно, что возраст и антиретровирусный статус не были определены как факторы риска повторного заболевания [133].

Сахарный диабет (СД) независимо связан с рецидивом туберкулеза [55, 84, 152]. P. Lee и соавт. (2014) исследовали влияние уровня сахара в крови (измеренный по уровню гликированного гемоглобина) на риск рецидива. Они обнаружили, что пациенты без контроля гликированного гемоглобина имели более высокий риск рецидива (ОШ = 1,98; 95 % ДИ 1,13–3,45), чем те, у кого было хотя бы одно измерение. Авторы предположили, что их результаты были обусловлены лучшим гликемическим контролем и высокой приверженностью к лечению больных СД [84].

Хронические заболевания легких, включая хроническую обструктивную болезнь легких (ХОБЛ), бронхиальную астму и интерстициальные заболевания легких, признаны факторами риска развития туберкулеза. Как показано у A. Pettit

и соавт. (2011) (ОШ = 5,28; 95 % ДИ 1,16–24,04) [90] и С. Hung и соавт. (2015) (ОШ = 1,59; 95 % ДИ 1,08–2,36) [110] хронические неспецифические заболевания легких независимо связаны с рецидивом туберкулеза. Напротив, последний систематический отчет предполагает, что наличие ТБ в анамнезе может привести к хроническим заболеваниям легких, особенно ХОБЛ и бронхоэктазам [149]. Н. Fan и соавт. (2021) сообщили о сильной и устойчивой положительной связи между туберкулезом в анамнезе и наличием хронических респираторных заболеваний, включая ХОБЛ и бронхоэктазы. Это говорит о том, что развитие хронического заболевания легких после туберкулеза увеличивает риск рецидива туберкулезной инфекции. Кроме того, как уже было отмечено выше, курение табака является фактором риска развития ХОБЛ и, таким образом, может быть связующим звеном в развитии туберкулеза и, в дальнейшем его рецидива [149]. Ученые под руководством А. Pettit обнаружили значительную взаимосвязь повторного туберкулеза с курением. Пациенты, которые курили более 20 сигарет в день, были более склонны к рецидивам заболевания, чем те, кто курил менее 20 сигарет в день (ОШ = 9,4; 95 % ДИ 1,1–83,9) [76]. Помимо табачного дыма, агрессивное воздействие окружающей среды также связано с повышенным риском развития туберкулеза. Доказано, что атмосферное загрязнение и оксид углерода вызывают эндогенную реактивацию МБТ и увеличивают заболеваемость туберкулезом [67, 86]. Роль этих факторов при рецидиве ТБ требует дальнейшего изучения.

1.2 Искусственный интеллект и методы машинного обучения: история создания и сфера деятельности

Искусственный интеллект (ИИ) и методы машинного обучения (Machine Learning) (МО) совершили революцию во всем мире. Ни в одной другой отрасли не было такого масштаба разработок, как в области ИИ и МО [8, 109]. Термин «Искусственный интеллект» был создан Джоном Маккарти и само словосочетание впервые было использовано на семинаре, организованном

Маккарти в Дартмутском колледже летом 1956 года. [73, 103]. Термин «машинное обучение» было впервые введен в 1959 году выдающимся ученым Артуром Самуэлем. Понятия ИИ и МО часто используются взаимозаменяемо, хотя это не совсем точно: ИИ относится больше к группе наук, объединяющих концепцию всеобъемлющей «мыслящей машины», способной выводить логическое заключение и технологии интеллектуального, автоматического принятия решений, в то время как согласно представлениям самого А. Самуэля, МО – это возможность компьютерным системам учиться автоматически и совершенствовать свой опыт без программирования [102, 131]. Иными словами искусственный интеллект – это область компьютерных наук, которая стремится имитировать мыслительные процессы человека, способность к обучению и хранению знаний, а ее методы применяются для выявления предикторов развития, например, заболеваний, прогнозирования их клинического течения, стратификации риска, повышения эффективности лечения и оптимизации расходов здравоохранения.

Первоначально концепция ИИ восходит к 1950-м годам и связана с «обучающей машиной» А. Тьюринга и разработкой первой нейронной сети, сфокусированной в основном на военных экспериментах [167, 199]. К концу 1960-х годов уже было ясно, как методы ИИ и МО могут быть использованы для решения конкретных проблемных областей, в том числе в естественных науках. В Стэнфордском университете проект DENDRAL генетика Джошуа Ледерберга и ученого в области теории вычислительных систем Эдуарда Фейгенбаума в сотрудничестве с химиком-органиком Карлом Джерасси и ученым в области информатики Брюсом Бьюкененом содержал алгоритмы, которые выводили вероятные химические структуры органических соединений из данных масс-спектрометра, используя правила для кодирования, основанные на знаниях опытных химиков [95].

В 1970-х годах взрывной интерес к биомедицинским применениям ИИ привел к созданию вычислительного ресурса SUMEX-AIM, который был подключен к только что созданному сетевому ресурсу ARPANET (сеть Агентства

перспективных исследовательских проектов). Это способствовало сотрудничеству и формированию нового научного сообщества, работы которого стали более доступными, и позволило координировать свою деятельность ведущим университетам для предоставления ресурсов исследователям в области ИИ в медицине, первоначально в США, а затем и на международном уровне [175, 183]. Так были созданы первые экспертные системы при поддержке методов ИИ и МО: INTERNIST-1, CASNET, CADIAG-2 и MYCIN [63, 119, 169]. Система MYCIN, к примеру, использовалась для видовой идентификации бактерий и подбора необходимого количества антибактериальных препаратов при тяжелых инфекционных процессах [64].

В 1980-х годах наличие таких вычислительных ресурсов, как ARPANET и NSFNET (сеть Национального научного фонда), усилило данное взаимодействие и позволило исследователям решать медицинские задачи с помощью методов ИИ и МО, в том числе и по созданию новых экспертных систем [183]. Ученые всего мира осознали, что вычислительные мощности компьютеров стали полезны в медицине, как в фундаментальных, так и в клинических направлениях. Так в 1985 году в Италии впервые была организована конференция по ИИ в медицине (AIME), посвященная проблемам, которые связывали информатику, медицину и биологию [175, 183]. В последующие годы методы ИИ и МО развивались на основе инженерии знаний, создания правил, до все более ориентированных данных биоинформатики и семантических технологий, причем в основном претерпевали изменения методы, объединяющие математическую статистику и теорию вероятностей [169]. С ростом ресурсоемкости по хранению информации, интеграции интеллектуального анализа данных и анализа больших данных в общую парадигму произошло еще большее расширение сферы и потенциала ИИ и МО [119].

Методы МО и ИИ стали еще более актуальными в медицине с момента появления доступа к большим наборам биомедицинских и клинических данных, включая использование электронных медицинских карт, а также наличие современных вычислительных машин, обладающих высокими мощностями [63].

В последнее время различные коллаборации во всем мире используют возможности ИИ и МО в медицине и здравоохранении для достижения более точных диагностических алгоритмов, индивидуального подхода в лечении и для решения важных задач прогнозирования в практическом здравоохранении [64, 65, 69, 85].

Как уже было отмечено выше, МО рассматривается как подраздел ИИ [169]. Основная цель – это изучение данных путем использования эффективных вычислительных алгоритмов, которые позволяют определять в них скрытые закономерности, зависимости и взаимосвязи, которые было бы трудно сделать с помощью традиционных статистических методов, и построить из массивных наборов таких данных прогнозные модели, включающие в себя эвристические правила, основанные на человеческой логике. Во многом МО базируется на известных математических дисциплинах и статистических законах [72]. Однако эти методы все же отличаются от методов традиционной статистики, они находят нелинейные зависимости: алгоритмы МО, обрабатывая огромное количество данных, извлекают информацию путем интеллектуального анализа, математически фокусируясь на прогнозирующей эффективности моделей с процедурой оценивания обобщающей способности (кросс-валидация) и итеративного улучшения алгоритма [141, 143].

Методы МО являются основным инструментом ИИ и представляют собой обучаемые автоматизированные системы, которые, используя различные виды моделирования и опираясь на доступные данные, принимают необходимые решения для реализации предсказательной функции. Алгоритмы МО используются в смежных областях: когнитивных вычислениях, компьютерном зрении и робототехнике. В отличие от традиционных статистических методов, которые обеспечивают исследования взаимосвязей между ограниченным числом переменных, методы МО предоставляют возможность для обработки и анализа больших сложных разнородных данных. Кроме того, алгоритмы МО основаны на меньшем количестве допущений и имеют более высокую прогностическую точность. В дальнейшем создание интеллектуальных машин на основе

искусственного интеллекта происходило по образу и подобию работы человеческого мозга, искусственные нейронные сети которых способны обрабатывать огромное количество данных, чтобы «натренировать» или обучить систему четко и точно классифицировать данные.

На сегодняшний день ИИ-системы задействованы во многих сферах жизнедеятельности человека, включая медицину [8]. В частности преуспели в маркетинге по работе с клиентопотоком [1, 20, 37, 48, 127], логистике [23], управлении персоналом [51], юриспруденции [14], в банковской деятельности [27] и других областях. В медицине разработки в этом направлении активно ведутся и за рубежом, и в России, а новые проекты и «умные» медицинские продукты, сервисы и процессы появляются практически каждый месяц. Преуспели в этом направлении такие компании как IBM, Google, Apple, Microsoft, General Electric и многие другие. Это означает, что активное использование алгоритмов машинного обучения вовсе не туманные перспективы, а реалии сегодняшнего дня. По данным аналитиков Research and Markets объем мирового рынка технологий искусственного интеллекта, используемого в здравоохранении, по итогам 2020 года достиг \$ 4,2 млрд. По их словам, продажи медицинских ИИ-решений интенсивно растут. Так, до 2025 года они будут увеличиваться на 45,3 % ежегодно и составят \$ 27,2 млрд.

В здравоохранении созданные интеллектуальные продукты помогают распознать многие заболевания. На основании опросников жалоб, анамнестических сведений, данных объективного осмотра и результатов обследования системы собирают, анализируют параметры состояния здоровья человека и ищут в базе данных информацию о возникшей проблеме. Так, например, есть сведения по применению ИИ-алгоритмов в распознавании злокачественных новообразований [7], кардиологических заболеваний [35, 36], COVID-19 [9], в акушерстве и гинекологии [12], в терапевтической [25] и хирургической практике [13]. Кроме обработки данных пациентов, алгоритмы искусственного интеллекта ведут поиск оптимального лечения пациентов (с учетом функционирования внутренних систем организма и генетического

анализа), а также создания качественно новых лекарственных средств на основе поиска оптимальной химической структуры [8]. В терапии, пульмонологии, онкологии и фтизиатрии созданные ИИ-системы определяют перспективы повышения точности диагностики заболеваний по данным рентгенологического исследования. В России Институтом развития интернета на основе обучения тысяч медицинских снимков и данных компьютерной томографии органов дыхания создана платформа первой медицинской нейросети под названием «Третье мнение», которая после загрузки изображений в систему и анализа данных выделяет ряд специфических признаков, определяющих приоритет в дифференциально-диагностическом ряде: туберкулез, рак, типичная и атипичная пневмония. Безусловно, цифровые подсказки позволят врачу поставить правильный диагноз, сократив время на дифференциальную диагностику.

1.3 Методы машинного обучения, применяемые во фтизиатрии

Методы машинного обучения оперируют с данными, представленными в виде точек в многомерном пространстве. Такие точки называются векторами признаков (англ. feature vectors), а пространство – пространством векторов признаков (англ. feature vector space). Среди методов машинного обучения выделяют обучение с учителем (англ. Supervised learning), обучение без учителя (англ. Unsupervised learning) и обучение с частичным привлечением учителя (англ. Semi-Supervised learning) (Рисунок 1).

Обучение с учителем использует размеченные экспертом данные для построения модели анализируемых данных. Обучение без учителя использует априорную информацию о близости векторов признаков в пространстве, либо о количестве классов данных. Обучение с частичным привлечением учителя опирается на комбинацию размеченных и неразмеченных данных. Хотя задачи регрессионного анализа и классификации схожи, особенно в области искусственных нейронных сетей, все же существуют различия. Так, на выходе регрессионной модели имеется непрерывный вектор данных, тогда как в

классификации на выходе – метки классов [81].

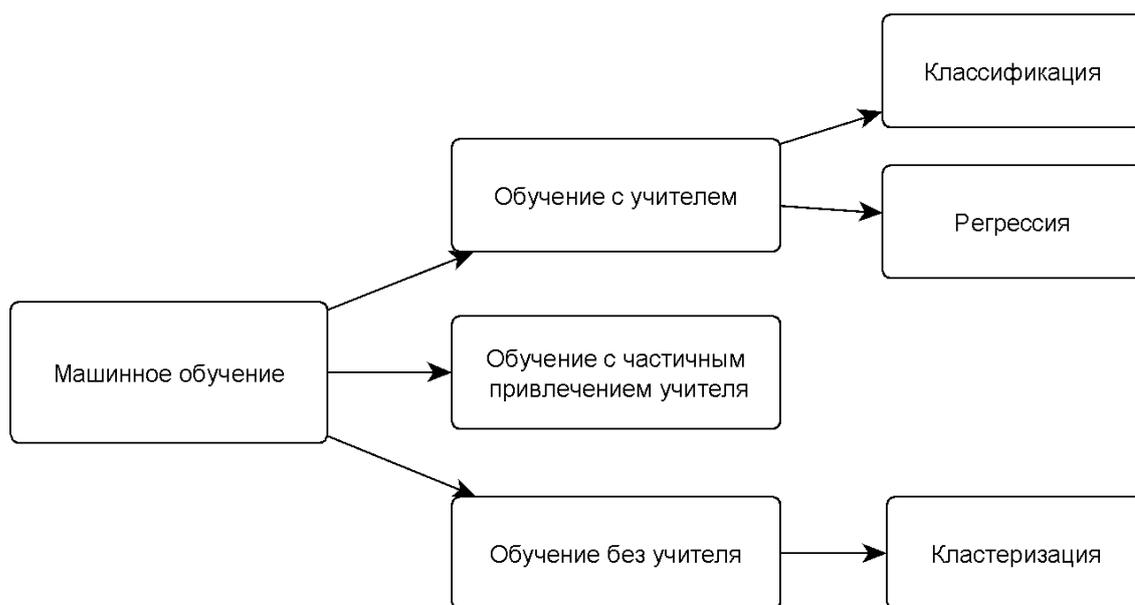


Рисунок 1 – Классификация методов машинного обучения

Методы выделения информативных признаков с учителем опираются на имеющиеся размеченные данные и могут быть грубо разделены на методы-фильтры, методы-обертки и методы-вложения [101]. Методы-фильтры опираются на классические статистические подходы, например, корреляцию Спирмена для попарного исследования предикторов и зависимой переменной. Достоинство данных методов – в вычислительной эффективности и робастности к переобучению. Недостаток методов состоит в игнорировании связей между предикторами. Для учета связей между предикторами используются методы-обертки. Для реализации данных методов требуется сформулировать некоторый критерий оптимальности выбора подмножества признаков. Таким критерием может служить, например, точность прогнозной модели, построенной на заданном подмножестве признаков по отношению к точности в референсном наборе. Недостаток данных методов – в росте риска переобучения и вычислительной сложности. Методы-вложения используют выделение информативных признаков как часть алгоритма классификации данных.

Одной из трудностей применения МО в анализе медицинских данных

является то, что полученные модели данных зачастую представляют собой «черный ящик» и оценить «разумность» полученных результатов сложно. Тем не менее, необходимо остановиться кратко на каждом из основных методов.

Логистическая регрессия (Logistic Regression, LR) является статистической линейной моделью классификации и позволяет предсказывать апостериорные вероятности классов с помощью логистической кривой. Данный метод используется для построения моделей классификации [71]. С помощью LR можно проверить, связаны ли две переменные линейно, и вычислить силу линейной зависимости [129]. К достоинствам логистической регрессии можно отнести относительную быстроту обучения, простоту интерпретации и малое количество настраиваемых параметров. В связи с этим в медицинских исследованиях LR используется для прогнозирования риска развития определенного заболевания на основе введенных признаков пациента [71]. Данный метод хорошо себя зарекомендовал как в когортных исследованиях по прогнозированию развития туберкулеза [117, 129], так и в исследованиях случай-контроль. Так, A. Rosser и соавт. (2018) из Великобритании у 82 больных с повторным случаем заболевания и у 4 546 больных с безрецидивным течением (контроль) с помощью LR выявили предикторы рецидива туберкулеза. По результатам анализа курение (ОШ = 3,8; $p \leq 0,04$), неблагоприятные побочные реакции на химиотерапию 3-4 степени тяжести (ОШ = 5,6; $p \leq 0,02$), этническая индийская принадлежность (ОШ = 8,5; $p \leq 0,01$) и прием иммунодепрессантов (ОШ = 6,8; $p \leq 0,01$) были независимыми факторами риска рецидива туберкулеза [54].

Линейный дискриминантный анализ (Linear Discriminant Analysis, LDA), также известный как линейный дискриминант Фишера [170], применим, если выборка удовлетворяет следующим гипотезам: классы распределены по нормальному закону, и матрицы ковариаций классов равны. Линейный дискриминант Фишера является упрощением квадратичного дискриминанта. В случае двух классов в двумерном пространстве разделяющей поверхностью, построенной с помощью этого метода, будет прямая. В случае большего числа классов разделяющая поверхность будет кусочно-линейной. Этот метод работает

с меньшим подмножеством данных и сравнивает его с размером исходной выборки данных [52]. LDA может справиться с проблемой дисбаланса между классами набора данных и максимизирует долю дисперсии между классами для дисперсии внутри класса в любом наборе данных, тем самым обеспечивая максимальную разделяемость признаков. В медицине дискриминантный анализ, например, может оценить течение заболевания по первичным признакам состояния больного и дать прогноз на полное, частичное выздоровление или на неблагоприятный исход терапии [53]. Во фтизиатрии есть исследование, посвященное поиску иммунных маркеров плазмы крови для прогнозирования терапевтического ответа при лечении МЛУ ТБ. Из всех возможных комбинаций аналитов LDA определил комбинацию подмножества аналитов, которые лучше всего предсказывают прекращение бактериовыделения. В частности это – IP-10 / CXCL10, VEGF-A, SAA и CRP [53].

Метод К ближайших соседей (K-Nearest Neighbors, KNN) опирается на одно важное предположение, называемое гипотезой компактности: если мера сходства объектов введена достаточно удачно, то схожие объекты гораздо чаще лежат в одном классе, чем в разных. В этом случае граница между классами имеет достаточно простую форму, а классы образуют компактно локализованные области в пространстве объектов. В KNN объект относится к тому классу, которому принадлежит большинство из его соседей, классификатор будет извлекать ближайших соседей и использовать их для прогнозной модели. Это метод непараметрической классификации, используется как для классификации, так и регрессии. Алгоритм может классифицировать пациентов по разным показателям, основываясь на данных пациентов предыдущих периодов. Есть данные использования KNN в распознавании заболеваний по рентгенологическим снимкам. Чувствительность метода по сравнению с другими методиками в выявлении туберкулеза по рентгенологическим снимкам невысокая, достигает 70 % [134].

Наивный байесовский классификатор (Naive Bayes, NB) – это простой вероятностный классификатор, основанный на теореме Байеса [181].

Классификатор предполагает, что каждый признак независимо влияет на классификацию, и значение каждого признака не зависит от значения любого другого признака. Классификатор NB быстр, прост в создании и полезен для классификации больших наборов данных. Классификатору требуется только небольшое количество данных для целей обучения, и он хорошо зарекомендовал себя во многих задачах классификации, включая оценку прогноза течения туберкулеза на фоне химиотерапии [117].

К алгоритмам логической классификации, основанным на построении иерархических композиций простых правил, относится метод дерева решений (Decision Trees, DT) [52].

Структура дерева состоит из так называемых ветвей и листьев, в ветвях записаны признаки, от которых зависит целевая функция, а в листьях – значения целевой функции. Деревья решений пригодны как для решения задач классификации, так и решения задач регрессии. В случае задачи классификации в листьях дерева записывается класс, которому принадлежит объект классификации, в случае задач регрессии – действительное число. Схема построения дерева выглядит следующим образом:

- 1) выбирается очередной признак $f_i(x)$;
- 2) значения выбранного признака разбиваются на несколько подгрупп, тем самым разделяя признаковое пространство на несколько подпространств;
- 3) процедура рекурсивно повторяется, пока не будет достигнут один из критериев остановки – максимальная глубина дерева, количество листьев, либо максимальный размер каждого листа. В зависимости от того, по какому принципу выбирается очередной признак для разбиения, и как он разбивается, существует несколько вариантов данного алгоритма: по критерию прироста информации, по критерию нормализованного прироста информации, и, так называемый, CART-алгоритм (classification and regression tree) [52, 117]. На практике, во избежание эффекта переобучения после построения дерева решений, некоторые его ветви усекают для поддержания лучшей обобщающей способности, данная процедура называется *pruning*.

Метод дерева решений используются для решения задач, как классификации, так и регрессии. Для каждого нового элемента в тестовом наборе дерево решений необходимо пройти от корня до одного из его листьев, таким образом, каждый узел в дереве должен быть проверен, и, в зависимости от значения, он должен быть назначен одному из подчиненных элементов. В обзоре уже была упомянута работа S. R. N. Kalhori и соавт. (2013) [117] по оценке и сравнению различных методов машинного обучения для прогнозирования результатов курса лечения туберкулеза. В данном исследовании были применены на обучающихся (4 515) и тестовых (1 935) выборках следующие алгоритмы МО: дерево решений (DT), искусственная нейронная сеть (ANN), логистическая регрессия (LR), радиальная базисная функция (RBF), наивный байесовский классификатор (NB), метод опорных векторов (SVM). Было идентифицировано семнадцать значимых коррелированных признаков ($p = 0,004$; 95 % ДИ = 0,001–0,007). DT оказался лучшим алгоритмом с точностью предсказания 74,21 % по сравнению с ANN, BN, LR, RBF и SVM с 62,06 %, 57,88 %, 57,31 %, 53,74 % и 51,36 % соответственно.

Метод опорных векторов (Support Vector Machine, SVM) был разработан для решения задачи бинарной классификации [118]. Идея метода опорных векторов заключается в поиске оптимальной разделяющей гиперплоскости для двух классов векторов признаков. Критерий оптимальности здесь состоит в максимизации расстояния от всех точек каждого из класса до гиперплоскости. SVM – это набор контролируемых методов обучения, которые анализируют данные и распознают закономерности. Он обычно используется для классификации и регрессионного анализа [118]. SVM основан на критерии минимизации структурного риска, и его цель – найти оптимальную разделяющую гиперплоскость. Этот подход улучшает способность обучающейся машины к обобщению и эффективен при решении таких проблем, как нелинейное разделение данных большой размерности и задачи классификации, для которых не хватает предварительных знаний [118]. Данный метод зарекомендовал себя в прогнозировании исходов лечения туберкулеза и диагностики [94, 111, 117].

Метод случайных лесов (Random Forest, RF) представляет собой

разновидность ансамблевых методов обучения с широкой областью применения: классификация, регрессия, кластеризация, поиск аномалий, выделение информативных признаков. RF представляет собой классификатор, состоящий из множества древовидных классификаторов, где из множества случайных независимо распределенных векторов признаков каждый древовидный классификатор «голосует» за самый популярный класс для входного вектора. Если решается задача регрессии, голоса каждого древовидного классификатора усредняются, в случае задачи классификации реализуется схема голосования по большинству [70]. В настоящее время RF является одним из наиболее часто используемых алгоритмов машинного обучения среди методов интеллектуального анализа данных, так как это метод, который можно использовать как для прогнозирования, так и для классификации, и его относительно легко обучить. Такое предпочтение объясняется его высокой производительностью обучения и низкими требованиями в отношении для подготовки ввода и настройки гиперпараметров [59]. По сути, это метод, который объединяет несколько деревьев решений, относящихся к входным данным базы данных. Таким образом, классификатор состоит из N деревьев, где N – количество выращиваемых деревьев, которое может быть любым значением, определяемым пользователем. Чтобы классифицировать новый набор данных, каждый случай наборов данных передается каждому из N деревьев. Лес выбирает класс с максимальным N голосов [59]. Он широко использовался для выявления туберкулеза и в трех выявленных исследованиях прогноза туберкулеза [21, 94, 111, 117].

Градиентный бустинг (Gradient Boosting, GB) – это техника машинного обучения для задач классификации и регрессии, которая строит модель предсказания в форме ансамбля слабых предсказывающих моделей для контролируемых задач, иными словами, алгоритм объединяет несколько слабых учеников для создания сильного ансамбля. В GB процедура обучения последовательно соответствует новым моделям, чтобы обеспечить более точную оценку переменной отклика. Основная идея, лежащая в основе этого алгоритма, состоит в том, чтобы построить новые базовые обучающиеся, которые будут

максимально коррелированы с отрицательным градиентом функции потерь, связанной со всем ансамблем [70, 136].

Многослойный перцептрон (Multilayer Perceptron, MLP) – это модель машинного обучения, используемая как для классификации, так и для регрессии [209], и была исследована для использования при прогнозировании туберкулеза [117]. По сути, это модель перцептрона с одним или несколькими скрытыми слоями, каждый из которых имеет определенное количество нейронов, связанных весами. Данные независимых переменных вставляются в нейроны входного слоя и обрабатываются в скрытом слое. В конечном итоге результат MLP представлен в выходном слое.

В отличие от одного классификатора, методы ансамбля пытаются построить набор моделей и объединить их. Ансамблевое обучение также называется обучением на основе комитетов или нескольких систем классификаторов обучения [210]. Комбинация моделей обучения традиционно может производиться тремя способами: в среднем, путем голосования или модели обучения. Среднее значение обычно применяется при обработке числовых выходных данных, среднее значение получается в качестве выходных данных классификаторами. По голосованию производится подсчет выходных данных классификаторов на основе частоты появления класса, а класс с наибольшим количеством голосов используется в качестве входных данных для новой модели обучения. Модель обучения использует результат, полученный в результате комбинации других моделей, и передает его другой модели обучения, которая будет учиться на этих моделях, чтобы предоставить свое собственное прогнозирование [210].

Итак, основные методики машинного обучения определены, они широко используются во фтизиатрии, и в целом направлены на: эпидемиологию, диагностику туберкулеза, оценку риска заболеваемости или смертности пациентов, прогнозировании шансов развития заболевания после первичного инфицирования (по геномному анализу РНК) и в выявлении МЛУ у инфицирующего штамма. В таблице 3 представлен краткий обзор исследований, посвященный использованию методов МО во фтизиатрии.

Таблица 3 – Обзор литературных источников по применению алгоритмов машинного обучения во фтизиатрической практике

Автор, год	Цель исследования	Методы МО	Количество человек (n)	Оценка прогнозирования
Sergeev R. S. и соавт., 2014 [96]	Выявление мутаций МБТ, связанных с лекарственной устойчивостью	LMM; LR	144	Точность более 90 % по выбранным 9-ти препаратам
Melendez J. и соавт., 2015 [56]	Выявление туберкулеза	KNN; SVM	917; 869; 850	KNN: ROC – 0,84; 0,78; 0,82 SVM: ROC – 0,88; 0,79; 0,85
Alcantara M. F. И соавт., 2017 [114]	Выявление туберкулеза	CNN	4701	Точность – 62,07 %
Lopes U. K. и соавт., 2017 [125]	Выявление туберкулеза	CNN	138; 662	Точность – 82,6 %; 92,6 % ROC – 84,7 %; 92,6 %
Yahiaoui A. и соавт., 2017 [207]	Выявление туберкулеза	SVM	150	Точность – 96,68 %
Évora L. H. R. и соавт., 2017 [89]	Выявление туберкулеза с множественной лекарственной устойчивостью	NN	280	Чувствительность – 95,1 %; Специфичность – 85 %
Thompson E. G. и соавт., 2017 [107]	Прогнозирование неэффективности лечения туберкулеза	LR	153	Чувствительность – 83–100 %; Специфичность – 26–100 %
Sambareu A. и соавт., 2017 [200]	Дифференциальная диагностика между туберкулезом и ВИЧ	KNN	54	Точность – 79–93 %
Mamiya H. и соавт., 2015 [188]	Прогнозирование недавней передачи туберкулеза	LR	1 552	Чувствительность – 53 %; Специфичность – 67 %

Продолжение таблицы 3

Автор, год	Цель исследования	Методы МО	Количество человек (n)	Оценка прогнозирования
JoãoFilho B. D. O. и соавт., 2016 [57]	Выявление туберкулеза легких с отрицательным мазком мокроты	MLP	136	Чувствительность – 100 %; Специфичность – 80 %; Точность – 88 %; ROC – 91,8 %
<p>Примечания:</p> <p>МО – машинное обучение;</p> <p>LMM – линейная смешанная модель;</p> <p>LR – логистическая регрессия;</p> <p>KNN – метод К ближайших соседей;</p> <p>SVM – метод опорных векторов;</p> <p>CNN – глубокие сверточные нейронные сети;</p> <p>NN – нейронная сеть;</p> <p>MLP – многослойный перцептрон;</p> <p>ROC – анализ, оценивающий эффективности моделей с помощью двух показателей – чувствительности (Se) и специфичности (Sp).</p>				

РЕЗЮМЕ

По данным литературных источников необходимо заключить, что даже в странах с низким уровнем заболеваемости, рецидив у больных туберкулезом после эффективной химиотерапии является естественным явлением. Ученые согласуются во мнении, что у успешно излеченных от туберкулеза пациентов риск повторного заболевания в несколько раз выше, чем у непереболевших людей. В связи с этим рецидивирующий туберкулез представляет собой серьезную угрозу для мировых и государственных программ по ликвидации туберкулеза, особенно с учетом роста доли пациентов с ВИЧ-инфекцией/туберкулез, более высокой концентрацией в данной когорте больных резистентных штаммов МБТ, особенно с МЛУ. Одновременный прогресс в области инфраструктуры информационных технологий и роста мощностей компьютеров позволяет сегодня анализировать любые наборы биомедицинских данных для решения задач здравоохранения, используя современные методы анализа, такие как методы машинного обучения. Эти методы могут обнаруживать и идентифицировать внутренние закономерности и взаимосвязи из сложных и многочисленных наборов данных, в то же время они обладают эффективной прогнозирующей способностью и могут применяться для создания стандартных моделей принятия клинических решений. Анализ литературных источников показал, что в основном все исследования во фтизиатрии, основанные на искусственном интеллекте, носят поисковый характер, не описывают практическое применение. Несмотря на это данная область науки только зарождается и она перспективна, в частности, в глобальном здравоохранении, направленном на искоренение туберкулеза.

ГЛАВА 2 МАТЕРИАЛ И МЕТОДЫ ИССЛЕДОВАНИЯ

2.1 Организация исследования и схема его проведения

Диссертационная работа выполнена на кафедре фтизиатрии и пульмонологии ФГБОУ ВО «Сибирский государственный медицинский университет» Минздрава России с 2017 по 2020 гг., клинической базой которой является ОГАУЗ «Томский фтизиопульмонологический медицинский центр». Проведенное исследование одобрено локальным этическим комитетом ФГБОУ ВО «Сибирский государственный медицинский университет» Минздрава России (протокол № 5554 от 16.10.2017).

Дизайн исследования построен по типу «исследование случай-контроль», набор данных произведен ретроспективно.

Исследование содержит данные о 346 пациентах эффективно закончивших лечение с различными клиническими формами туберкулеза органов дыхания, которые получали лечение по IV режиму химиотерапии с января 2009 г. по декабрь 2011 г. в соответствии с действующими рекомендациями по лечению больных МЛУ ТБ.

Изначально была проанализирована когорта всех лиц, получающих ПТП 2 ряда в период с 01.01.2009 года по 31.12.2011 года в количестве 786 пациентов. Из них были выделены случаи в количестве 606 человек с верифицированным диагнозом МЛУ ТБ и получающие лечение по IV режиму химиотерапии. Формирование выборки пациентов в исследование проведено согласно критериев включения и исключения.

Критериями включения явились:

- возраст старше 18 лет;
- туберкулез органов дыхания;
- оба пола;
- диагноз, подтвержденный клинически, рентгенологически и культуральным исследованием (посев на плотные питательные среды) с

устойчивостью МБТ минимум к изониазиду и рифампицину;

- эффективно завершённый курс химиотерапии МЛУ ТБ;

Критерии исключения:

- возраст младше 18 лет;
- внелегочные формы туберкулеза;
- неэффективный курс химиотерапии;
- прерывание курса химиотерапии;
- выбывшие из административной территории до завершения диспансерного наблюдения;
- смерть от туберкулеза и от других причин.

Для реализации поставленной цели исследования и решения задач были сформированы две группы пациентов с туберкулезом органов дыхания. Первая (основная группа) – больные, у которых наступил рецидив заболевания после эффективного курса химиотерапии туберкулеза (ЭКХТ) с МЛУ возбудителя ($n = 35$), вторая (контрольная группа) – пациенты с МЛУ ТБ, успешно завершившие курс химиотерапии, в постлечебном наблюдении не развивали рецидив заболевания ($n = 311$). Группы сформированы ретроспективно. На рисунке 2 представлена методика включения пациентов в исследование, деления их на группы, а также их численный состав. Обследование, назначение курсов химиотерапии, а также диспансерное наблюдение пациентов обеих групп проходили согласно действующим нормативным документам на период лечения больных (в соответствии с приказом Минздрава России от 21.03.2003 № 109 «О совершенствовании противотуберкулезных мероприятий в Российской Федерации»).



Рисунок 2 – Схема исследования ретроспективного когортного анализа данных больных МЛУ ТБ после эффективного курса химиотерапии

Выбор ПТП проводился на основе результатов теста лекарственной чувствительности (ТЛЧ) МБТ к лекарственным средствам (посев на плотные среды). Назначение режима химиотерапии МЛУ ТБ, организация и продолжительность лечения, выбор ПТП, перевод с интенсивной фазы лечения на фазу продолжения, назначение иных мероприятий производилось на врачебной комиссии ОГАУЗ «Томского фтизиопульмонологического медицинского центра». Параллельно в схему лечения больных МЛУ ТБ включались витамины группы В (В₁, 6), проводилось симптоматическое лечение и корригирующая терапия сопутствующих заболеваний. Ежедневный прием ПТП проходил под непосредственным контролем медицинского персонала.

2.2 Материал и методы исследования

Для анализа данных была создана исследовательская база с обезличенными персональными данными пациентов (данные на пациентов были закодированы), куда вносились анамнестические сведения, результаты клинических и параклинических исследований, схемы химиотерапии, изменения ПТП в режиме лечения, информация о спектре ЛУ МБТ, данные о характере остаточных изменений, данные о поражении и распространенности процесса и др.

Источниками информации явилась следующая медицинская документация: форма № 089/у-туб «Извещение о больном с впервые в жизни установленным диагнозом активного туберкулеза, с рецидивом туберкулеза» в модификации ОГАУЗ «Томского фтизиопульмонологического медицинского центра», форма ФСН № 8 «Сведения о заболеваниях активным туберкулезом», форма ФСН № 33 «Сведения о больных туберкулезом», форма № 01-ТБ/у «Медицинская карта лечения больного», форма № 03-ТБ/у «Журнал регистрации больных туберкулезом», форма № 2-ТБ «Сведения о больных, зарегистрированных для лечения», форма № 7-ТБ «Сведения о впервые выявленных больных и рецидивах заболеваний туберкулеза», форма № 8-ТБ «Сведения о результатах курсов химиотерапии больных туберкулезом легких», форма № 10-ТБ «Сведения о результатах интенсивной фазы лечения (по микроскопии мокроты)», форма № 030-4/у «Контрольная карта диспансерного наблюдения контингентов противотуберкулезных учреждений», форма № 081/у «Медицинская карта больного туберкулезом», форма № 003/у «Медицинская карта стационарного больного», форма №106/у-98 «Медицинское свидетельство о смерти».

Учитывая большое количество агрегированных данных о бактериологических исследованиях и сведениях о курсах химиотерапии, использовались базы учета контингентов организационно – методического отдела ОГАУЗ «ТФМЦ» СУБД «Dbase» за соответствующий период исследовательских когорт. Для исключения ошибок, возникающих в процессе сбора данных (записи в исследовательскую форму) и в процессе их внесения, был осуществлен

независимый двойной ввод в базу данных, программное сравнение двух вводов с исправлением выявленных несовпадений и ошибок.

Общая часть исследовательской базы включала в себя возраст, индекс массы тела ($\text{кг}/\text{м}^2$), место жительства, социальный статус, пребывание ранее в местах заключения, табакокурение, злоупотребление алкоголем, тип пациента (впервые выявленный, повторное лечение), наличие туберкулезного контакта, наличие сопутствующих заболеваний (в том числе ВИЧ-инфекция, гепатиты, сахарный диабет и др.). Также, в анализ включены клинические формы заболевания, распространенность процесса на основании данных рентгенологических методов исследования, присутствие фазы распада, наличие остаточных изменений. Расширенная специфическая часть по бактериологическим исследованиям СУБД «Dbase», которая включала в себя спектр ЛУ на начало лечения, в процессе и на конец лечения, данные об амплификации и всех изменениях в курсах химиотерапии, а также назначенные схемы лечения на начало и конец химиотерапии, количество курсов, приверженность пациента лечению.

Таким образом, полученная на основе данных медицинской документации и СУБД «Dbase» составленная обезличенная база информации о пациентах имела количественные, качественные и порядковые признаки.

Для расчета выборки необходимого объема было принято во внимание, что частота рецидивов после эффективной химиотерапии у больных МЛУ ТБ в мире колеблется от 0 % до 8,5 %, а средняя продолжительность наблюдения варьировала от 19,2 до 98,1 месяца. В РФ, в том числе и на территории Томской области, данная частота составляет до 6,8 % при среднем сроке наблюдения 42,4 месяца [61]. В Томской области после ЭКХТ из числа зарегистрированных больных на лечение с 01.01.2009 по 31.12.2011 в ОГАУЗ «ТФМЦ» при среднем сроке наблюдения в 60 месяцев (за всеми пациентами, включенными в исследование продолжалось наблюдение до 2017 года) данный показатель составил 10,1 %. Таким образом, для обеспечения репрезентативности данных и достоверности выводов, полученных в ходе нашего исследования для Томской

области, размером выборки выступила генеральная совокупность лиц, развивших повторное заболевание после ЭКХТ, т. е. исследованию подверглась вся популяция больных с рецидивами МЛУ ТБ, зарегистрированными за указанный период постлечебного наблюдения.

Рецидив фиксировался, как появление признаков активного туберкулеза у лиц, ранее перенесших туберкулез и излеченных от него, наблюдающихся в III группе или снятых с учета в связи с выздоровлением. В связи с этим в исследование включены больные, успешно завершившие курс лечения, у которых вновь появились признаки активного туберкулеза: получены положительные результаты микроскопии или посева мокроты и/или были обнаружены клиничко-рентгенологические признаки туберкулеза. Когорты больных по группам пациентов формировались согласно приказу Минздрава России от 13.02.2004 № 50 «О введении в действие учетной и отчетной документации мониторинга туберкулеза». Исходы лечения больных нормативно дифференцировались на «эффективный курс химиотерапии», «неэффективный курс химиотерапии», «умер», «прерывание курса химиотерапии», «выбыл», и «диагноз туберкулеза снят».

Данные были реализованы в табличном виде с использованием пакета Microsoft Office Excel 2016. Обработка и статистический анализ (процедуры описательной статистики, множественный регрессионный анализ) полученных результатов проводили с использованием программного обеспечения IBM SPSS Statistics версия 23.

Проверка на нормальность закона распределения проводили с помощью критериев: W-критерий Шапиро – Уилка (группа рецидива), критерий Колмогорова – Смирнова (контрольная группа). Поскольку количественные данные часто не описывались нормальным законом распределения, для описательной статистики, наряду со средним значением M и стандартным отклонением σ , использовали медиану Me и квартили Q_1 и Q_3 . Достоверность различий в двух независимых группах количественных данных, соответствующих нормальному закону распределения, проверяли при помощи t-критерия

Стьюдента. При сравнении порядковых признаков, а также количественных с законом распределения, отличным от нормального, использовали непараметрический критерий Манна – Уитни (U). Качественные признаки представлены в виде абсолютных и относительных частот n (%). Для сравнения качественных признаков использовали критерий χ^2 Пирсона (частота признака от 10 и выше). В случае, когда частоты ограничивали применение данного критерия, применяли поправку Йетса (χ^2*) и точный тест Фишера (ТТФ), в случае многопольных таблиц с малыми частотами встречаемого признака – критерий Фишера по Монте-Карло (F). Различия считали статистически значимыми при $p < 0,05$. В таблицах сопряженности 2×2 при обнаружении значимых различий определяли ОШ и 95 % доверительный интервал (ДИ 95 %) по методу Уилсона. Для исследования сравнения в динамике лекарственной устойчивости внутри группы рассчитывали критерий χ^2 Мак-Немара ($\chi^2_{\text{мн}}$).

Для получения уравнения и оценки вклада факторов риска в развитие рецидива МЛУ ТБ использовался многофакторный анализ в модуле бинарной логистической регрессии. Для построения модели применялся пошаговый отбор включения переменных в уравнение прямой с определением минимального набора предикторов по оценке значения коэффициента детерминации (R^2), показывающего долю влияния всех предикторов модели на дисперсию зависимой переменной. Адекватность полученной модели оценивали с помощью критериев Хосмера – Лемешева, Вальда. Различия считали статистически значимыми при $p < 0,05$. Проверку значимости модели осуществляли при помощи критерия χ^2 . При значении $p < 0,05$ принималась гипотеза о согласованности модели. Интерпретацию параметров логистической регрессии производили на основе величины $\exp(b)$: если коэффициент b является положительным, то $\exp(b)$ больше 1 и шансы наступления прогнозируемого события возрастают, отрицательное значение коэффициента b указывает на снижение шансов ($\exp(b)$). Чувствительность и специфичность предикторов оценивали при помощи ROC-анализа. Количественная интерпретация результатов проводилась по ROC-кривым с оценкой показателя AUC (Area under ROC curve – площадь под

ROC-кривой).

Методы машинного обучения были реализованы на высокоуровневом языке Python версии 3.8.6.

В качестве классификаторов использовали: дерево решений, случайный лес, XGBoost (градиентный бустинг) и логистическая регрессия. При построении моделей прогнозирования применяли инструменты библиотеки scikit-learn Version 0.24.2 в интерактивной облачной среде работы с программным кодом Google Colaboratory.

Для учета несбалансированности размеров выборки использовался подход с разбиением исходной обучающей выборки на 5 случайных подвыборок с сохранением пропорций классов (функция StratifiedKFold пакета sklearn). При обучении использовалась опция задания балансирующих коэффициентов для данных из каждого класса (параметр `class_weight="balanced"` для классификатора). Полученные оценки точности, специфичности, чувствительности и площади под ROC кривой усреднялись по всем 5 выборкам. Информативность признаков оценивали как среднее по всем 5 выборкам.

ГЛАВА 3 РЕЗУЛЬТАТЫ СОБСТВЕННЫХ ИССЛЕДОВАНИЙ

3.1 Статистический анализ групп наблюдения по выявлению предикторов рецидива туберкулеза с множественной лекарственной устойчивостью

Для установления частоты возникновения рецидива туберкулеза органов дыхания после успешного лечения больных с МЛУ возбудителя, а также для расчета выборки необходимого объема было принято во внимание, что частота рецидивов в мире варьирует от 0 % до 8,5 % после эффективной химиотерапии у больных МЛУ ТБ при средней продолжительности наблюдения от 19,2 до 98,1 месяцев. Из немногочисленных исследований установлено, что в РФ, в том числе и на территории Томской области данная частота составляет около 6,8 % при среднем сроке наблюдения в 42,4 месяца [61]. Для обеспечения репрезентативности данных и достоверности выводов, полученных в ходе нашего исследования, было принято решение учесть все случаи развития рецидивов МЛУ ТБ после ЭКХТ на территории Томской области из числа зарегистрированных больных на лечение IV режиму химиотерапии с 01.01.2009 по 31.12.2011 в ОГАУЗ «ТФМЦ» при сроке наблюдения в 60 месяцев. Таким образом, размером выборки выступила генеральная совокупность лиц, развивших повторное заболевание за указанный период постлечебного наблюдения (Таблица 4).

Таблица 4 – Анализ когорты зарегистрированных больных МЛУ ТБ после ЭКХТ

Год регистрации когорты	Пациенты, зарегистрированные с ЭКХТ (n, абс.)	Рецидивы (из ст. 2) (n, абс.)	Годы наблюдения (n, абс.):					Частота возникновения рецидива (n, %)
			1	2	3	4	5	
01.01–31.12.2009	97	11	3	4	1	0	3	11,3
01.01–31.12.2010	140	16	1	7	3	2	3	11,4
01.01–31.12.2011	109	8	3	2	3	0	0	7,3
01.01.2009–31.12.2011	346	35	7	13	7	2	6	10,1
Me (Q ₁ ; Q ₃):			1,9 (1,22; 2,71)					

Из анализа таблицы 4 установлено, что показатель частоты возникновения рецидивов после успешного лечения больных МЛУ ТБ органов дыхания за период наблюдения составил 10,1 %. Рецидивы МЛУ ТБ органов дыхания в 77,1 % (27 пациентов) возникают в течение первых 3-х лет наблюдения. Средний срок наступления рецидива заболевания после эффективного курса химиотерапии больных МЛУ ТБ составляет 1,9 (1,22; 2,71) года.

Характеризуя пациентов, было установлено, что в обеих группах больных преобладали мужчины – 26 (74,3 %) и 224 (72,0 %) соответственно ($p = 0,933$, χ^2 *). Выраженного дефицита массы тела не наблюдалось, его средние значения колебались в пределах 20,9 ($p = 0,813$, U) (Таблица 5). По среднему возрасту пациентов различий выявлено не было ($p = 0,255$, U): 42 (32; 50) и 38 (30; 50) лет соответственно. В равных частях это были жители г. Томска и Томской области ($p = 0,666$, χ^2), пациентов без определенного места жительства было 6 (17,1 %) и 26 (8,4 %) соответственно ($p = 0,164$, χ^2 *), без определенного места жительства было 32 человека.

При анализе социального статуса больных МЛУ ТБ установлено, что безработных в группе с рецидивом было 27 человек (77,1 %), в группе контроля – 235 (75,6 %), доля работающих составила в исследовании 22,9 и 24,4 % соответственно ($p = 0,999$, χ^2 *), доля лиц с инвалидностью по соматическому заболеванию преобладала в группе с рецидивом – 31,4 % и в группе сравнения – 8,7 % ($p < 0,001$, χ^2). Шансы возникновения рецидива туберкулеза у пациентов с инвалидностью по соматическому заболеванию выше (ОШ = 4,82; 95 % ДИ 2,13–10,90). Каждый пятый пациент, развивший рецидив, был пенсионного возраста (65 и более лет) – 7 (20,0 %) и в группе сравнения 25 (8,0 %) ($p = 0,045$, χ^2 *). У пациентов пенсионного возраста шансы возникновения рецидива были более высокими (ОШ = 2,86; 95 % ДИ 1,14–7,20).

Различий между больными по возрасту и индексу массы тела (ИМТ) не обнаружено, соотношение частот по месту проживания (город/село), семейного положения (одинок/семейный), наличие трудоустройства, статуса без определенного места жительства, пребывание в местах заключения в группах

было одинаково. Важными параметрами для прогноза возникновения рецидива в будущем являются: пенсионный возраст и наличие у пациента инвалидности по соматическому заболеванию.

Таблица 5 – Характеристика больных МЛУ ТБ

Параметры		Группы наблюдения (n = 346)				p
		основная группа (n = 35)		контрольная группа (n = 311)		
Возраст, лет, Me (Q ₁ ; Q ₃)	18–40	42 (32; 50)	16 (45,7)	38 (30; 50)	179 (57,6)	0,255*
	41–55		15 (42,9)		98 (31,5)	
	56 и более		4 (11,4)		34 (10,9)	
Индекс массы тела, кг/м ²	16 и менее	20,9 (19,8; 22,1)	0	20,9 (19,2; 22,8)	9 (2,9)	0,813*
	16,1–18,5		3 (8,6)		40 (12,9)	
	18,6–24,99		29 (82,8)		236 (75,9)	
	25,0 и более		3 (8,6)		26 (8,3)	
Пол (n, %)	мужчины	26 (74,3)		224 (72,0)		0,933****
	женщины	9 (25,7)		87 (28,0)		
Место жительства (n, %)	город	17 (48,6)		163 (52,4)		0,666**
	область	18 (51,4)		148 (47,6)		
Семейное положение (n, %)	семейный	19 (54,3)		163 (52,4)		0,833**
	одинокый	16 (45,7)		148 (47,6)		
Социальный статус (n, %)	безработный	27 (77,1)		235 (75,6)		0,999****
	работает	8 (22,9)		76 (24,4)		
	пенсионер	7 (20,0)		25 (8,0)		0,045****
	БОМЖ	6 (17,1)		26 (8,4)		0,164****
	инвалидность	11 (31,4)		27 (8,7)		< 0,001**

Продолжение таблицы 5

Параметры		Группы наблюдения (n = 346)		p
		основная группа (n = 35)	контрольная группа (n = 311)	
Пребывание в МЛС		14 (40,0)	80 (25,7)	0,072**
Туберкулезны й контакт (n, %)	да	9 (25,7)	52 (16,7)	0,276***
	нет	26 (74,3)	259 (83,3)	
Примечания: БОМЖ – без определенного места жительства; МЛС – места лишения свободы; Me – медиана, Q ₁ ; Q ₃ – квартили первого и третьего порядка; * – U-критерий Манна – Уитни; ** – χ^2 Пирсона; *** – χ^2 с учетом поправки Йетса;				

Таблица 6 – Сопутствующая патология у больных МЛУ ТБ

Сопутствующее заболевание	Группы наблюдения (n = 346)		p
	основная группа (n = 35)	контрольная группа (n = 311)	
	n (%)	n (%)	
ВИЧ	7 (20,0)	4 (1,3)	< 0,001***
ХНЗЛ	7 (20,0)	42 (13,5)	0,430**
Алкоголизм	24 (68,6)	172 (55,3)	0,133*
Опиоидная зависимость	7 (20,0)	35 (11,3)	0,219**
Сахарный диабет	3 (8,6)	12 (3,9)	0,185***
Психические заболевания	8 (22,9)	15 (4,8)	< 0,001**
Гепатит (В, С, В + С)	12 (34,3)	76 (24,4)	0,205*
Гастрит, ЯБЖ и ДПК, колит	2 (5,7)	24 (7,7)	1,000***
Почечная недостаточность	1 (2,9)	7 (2,3)	0,578***
Сердечная недостаточность	1 (2,9)	11 (3,5)	1,000***

Продолжение таблицы 6

Сопутствующее заболевание	Группы наблюдения (n = 346)		p
	основная группа (n = 35)	контрольная группа (n = 311)	
	n (%)	n (%)	
Табакокурение	31 (88,6)	221 (71,1)	0,027***
Дыхательная недостаточность	1 (2,9)	1 (0,32)	0,192***
Оперативные вмешательства	1 (2,9)	10 (3,2)	1,000***

Примечания: ВИЧ – вирус иммунодефицита человека; ХНЗЛ – хронические неспецифические заболевания легких; ЯБЖ и ДПК – язвенная болезнь желудка и двенадцатиперстной кишки; * – χ^2 Пирсона; ** – χ^2 с учетом поправки Йетса; *** – ТТФ.

Были обнаружены значимые различия между группами по наличию сопутствующей патологии, а именно: в группе рецидива чаще встречались ВИЧ (7 (20,0 %) пациентов против группы контроля – 4 (1,3 %), ($p < 0,001$, ТТФ), шансы возникновения рецидива у пациента с ВИЧ значительно выше (ОШ = 19,19; 95 % ДИ 5,29–69,56); психические заболевания (8 (22,9) и 15 (4,8 %) соответственно, ($p < 0,001$, χ^2 *), шансы заболеть туберкулезом повторно при наличии психического заболевания выше (ОШ = 5,85; 95 % ДИ 2,27–15,03); и табакокурение (31 (88,6 %) – пациент с рецидивом ТБ, 221 (71,1 %) – контрольная группа, ($p = 0,027$, ТТФ), шансы рецидива туберкулеза выше, если пациент является табакокурительщиком (ОШ = 3,16; 95 % ДИ 1,08–9,20) (Таблица 6).

Результаты общего и биохимического анализов крови представлены в таблицах 7 и 8. Различий между группами в показателях общего и биохимического анализов крови до проведения химиотерапии МЛУ ТБ не наблюдалось.

Таблица 7 – Показатели общего анализа крови до лечения больных МЛУ ТБ

Показатели и единицы измерения	Группы наблюдения (n = 346)		p*
	основная группа (n = 35)	контрольная группа (n = 311)	
	Me (Q ₁ ; Q ₃)		
Гемоглобин, г/л	132,0 (126,0; 140,0)	136,0 (122,0; 144,0)	0,706
Эритроциты, × 10 ¹² /л	4,0 (3,8; 4,4)	4,2 (3,8; 4,6)	0,175
Лейкоциты, × 10 ⁹ /л	7,8 (6,3; 8,7)	8,0 (6,0; 9,8)	0,613
Нейтрофилы п/я, %	8,0 (4,0; 10,0)	5,0 (2,0; 9,0)	0,123
Нейтрофилы с/я, %	53,0 (44,0; 60,0)	53,0 (47,0; 60,0)	0,574
Эозинофилы, %	3,0 (0,0; 5,0)	3,0 (1,0; 6,0)	0,479
Лимфоциты, %	23,0 (16,0; 30,0)	25,0 (17,0; 34,0)	0,301
Моноциты, %	10,0 (9,0; 16,0)	11,0 (7,0; 16,0)	0,419
СОЭ, мм/ч	20,0 (5,0; 28,0)	15,0 (6,0; 30,0)	0,817
Примечание: СОЭ – скорость оседания эритроцитов; * – U-критерий Манна – Уитни.			

В целом изменения в ОАК коснулись количества лейкоцитов, моноцитов и СОЭ, показатели были незначительно выше общепринятой нормы или на верхней границе нормы.

Таблица 8 – Показатели биохимического анализа крови до лечения больных МЛУ ТБ

Показатели и единицы измерения	Группы наблюдения (n=346)		p
	Me (Q ₁ ; Q ₃), M ± σ		
	основная группа (n = 35)	контрольная группа (n = 311)	
Общий белок, г/л	70,3 (65,3; 73,0)	70,0 (63,3; 74,1)	0,943*
Глюкоза, ммоль/л	5,0 (4,6; 5,5)	5,0 (4,5; 5,6)	0,825
Мочевина, ммоль/л	4,4 ± 1,6	4,4 ± 1,2	0,989**
Билирубин, мкмоль/л	7,2 (6,7; 8,6)	7,3 (6,5; 9,0)	0,886*
АСТ, Ме/л	0,4 (0,2; 0,6)	0,4 (0,2; 0,8)	0,976*

Продолжение таблицы 8

Показатели и единицы измерения	Группы наблюдения (n=346) Me (Q ₁ ; Q ₃), M ± σ		p
	основная группа (n = 35)	контрольная группа (n = 311)	
АЛТ, Ме/л	0,4 (0,2; 0,7)	0,3 (0,1; 0,8)	0,879*
Тимоловая проба	2,2 (2,0; 4,8)	2,8 (1,5; 3,4)	0,525*
Креатинин, мкмоль/л	81,4 ± 17,5	84,0 ± 16,9	0,401**
Мочевая кислота, мкмоль/л	163,0 (101,0; 337)	323,0 (101,0; 405,0)	0,313*
ТТГ, мкМЕ/мл	1,6 (1,1; 2,3)	1,5 (0,9; 2,1)	0,670*
Калий, ммоль/л	4,5 (4,1; 4,8)	4,5 (4,0; 4,9)	0,636*
Примечание: * – точный критерий Фишера; ** – t-критерий Стьюдента; М – среднее значение, σ – стандартное отклонение.			

Все пациенты были разделены на три основные группы согласно их типу с учетом регистрации данных больных на лечение по IV режиму химиотерапии. При анализе сравниваемых групп было выявлено, что число пациентов с впервые выявленным диагнозом туберкулеза в первой группе было 4 человека (11,4 %), в контрольной группе – 190 (61,1 %) ($p < 0,001$, ТТФ) (Таблица 9). Шанс заболеть повторно рецидивом у впервые выявленного пациента значительно ниже (ОШ = 0,08; 95 % ДИ 0,03–0,24).

У 31 пациента (88,6 %) из группы с рецидивом заболевания и у 121 человека (38,9 %) среди безрецидивных больных наблюдалось несколько курсов лечения туберкулеза ($p < 0,001$, ТТФ). Шанс возникновения рецидива туберкулеза выше у больных с несколькими курсами лечения (ОШ = 12,17; 95 % ДИ 4,19–35,34). Среди них повторное лечение было назначено пациентам из группы с рецидивом туберкулеза – 25 (71,4 %) и 96 (30,9 %) в группе контроля соответственно ($p < 0,001$, χ^2), шанс заболеть повторно у пациента с рецидивом туберкулеза выше (ОШ = 5,60; 95 % ДИ 2,59–12,11); хроническое течение

заболевания наблюдалось у 6 (17,1 %) и 25 (8,0 %) пациентов соответственно ($p = 0,140, \chi^2*$).

Таблица 9 – Характеристика больных в группах наблюдения по анамнезу заболевания

Тип пациента	Группы наблюдения (n = 346)		p	
	основная группа (n = 35)	контрольная группа (n = 311)		
	n (%)	n (%)		
Впервые выявленный	4 (11,4)	190 (61,1)	< 0,001*	< 0,001**
Рецидив	25 (71,4)	96 (30,9)		< 0,001***
Хроническое течение	6 (17,1)	25 (8,0)		0,140****
Примечания: * – критерий Фишера по Монте-Карло; ** – ТТФ; *** – χ^2 Пирсона; **** – χ^2 с учетом поправки Йетса.				

В связи с этим в группе безрецидивных больных случаев повторного лечения значительно меньше по сравнению с группой с рецидивом заболевания. Важную роль в возникновении рецидива имеет наличие нескольких курсов лечения.

При рассмотрении наблюдаемых типов пациентов различий не найдено ($p = 0,200, F$): после неудачи – 18 (51,4 %) и 172 (55,3 %) соответственно; лечение после прерывания курса химиотерапии встречалось только во второй группе у 7 (2,3 %) пациентов, заболевание в фазе обострения наблюдалось в группе рецидива у 1 больного (2,9 %), прибывших из другой территории было 1 (2,9 %) и 9 (2,9 %) пациентов, поступивших из мест заключения – 2 (5,7 %) и 9 (2,9 %) соответственно.

В таблице 10 представлено общее число курсов лечения туберкулеза у наблюдаемых пациентов.

Табличные данные демонстрируют, что рецидивы МЛУ ТБ возникают у пациентов с большим количеством предыдущих курсов лечения (3 курса лечения ($p < 0,001$, χ^2 , ОШ = 3,40; 95 % ДИ 1,60–7,23) 4 курса ($p < 0,001$, ТТФ, ОШ = 47,06; 95 % ДИ 12,29–180,16), 5 курсов ($p < 0,001$, ТТФ), в группе безрецидивного течения МЛУ ТБ у большинства повторное лечение не превышало 2 курсов.

Таблица 10 – Общее число курсов лечения больных МЛУ ТБ в группах наблюдения

Общее число курсов химиотерапии туберкулеза	Группы наблюдения (n = 346)		p
	основная группа (n = 35)	контрольная группа (n = 311)	
	n (%)	n (%)	
0	0	4 (1,2)	1,000**
1	0	100 (32,2)	< 0,001**
2	2 (5,7)	158 (50,8)	< 0,001**
3	13 (37,14)	46 (14,8)	< 0,001***
4	11 (31,4)	3 (1,0)	< 0,001**
5	8 (22,9)	0	< 0,001**
6	1 (2,9)	0	0,100**

Примечания:
 * – U-критерий Манна – Уитни;
 ** – ТТФ;
 *** – χ^2 Пирсона.

В таблице 11 представлены данные больных МЛУ ТБ по количеству курсов химиотерапии по IV режиму, содержащих ПТП резервного ряда.

Таблица 11 – Количество курсов химиотерапии по IV режиму больных МЛУ ТБ в группах наблюдения

Номер курса лечения ПТП 2 ряда	Группы наблюдения (n = 346)		p*
	основная группа (n = 35)	контрольная группа (n = 311)	
	n (%)	n (%)	
Не было предыдущих курсов лечения	29 (82,9)	270 (86,8)	0,690
Два курса лечения	5 (14,3)	36 (11,6)	
Три курса лечения	0	5 (1,6)	
Четыре курса лечения	1 (2,9)	0	
Примечание: * – U-критерий Манна – Уитни.			

В сравниваемых группах не обнаружено разницы в частоте использования ранее ПТП 2-го ряда. Только чуть более 10 % в обеих группах принимали ПТП резерва ($p = 0,690$, U).

Клинические формы туберкулеза легких у больных в группах наблюдались с одинаковой частотой (Таблица 12).

Таблица 12 – Клинические формы туберкулеза у больных МЛУ ТБ сравниваемых групп

Клиническая форма туберкулеза, рентгенологическая характеристика процесса	Группы наблюдения (n = 346)		p
	основная группа (n = 35)	контрольная группа (n = 311)	
	n (%)	n (%)	
Инфильтративная	15 (42,9)	173 (55,6)	0,526*
Диссеминированная	11 (31,4)	70 (22,5)	
Очаговая	3 (8,6)	25 (8,0)	
Туберкулёма	3 (8,6)	19 (6,1)	

Продолжение таблицы 12

Клиническая форма туберкулеза, рентгенологическая характеристика процесса		Группы наблюдения (n = 346)		p
		основная группа (n = 35)	контрольная группа (n = 311)	
		n (%)	n (%)	
Фиброзно-кавернозная и кавернозная		2 (5,7)	13 (4,2)	
Казеозная пневмония		1 (2,9)	3 (1,0)	
Туберкулез верхних дыхательных путей, трахеи, bronхов, плеврит, ТВГЛУ		0 (0,0)	7 (2,3)	
Поражение обоих легких		26 (74,3)	192 (61,7)	0,203**
Наличие фазы распада	да	33 (94,3)	214 (68,8)	< 0,001***
	нет	2 (5,7)	97 (31,2)	
Примечания: ТВГЛУ – туберкулез внутригрудных лимфатических узлов; * – F-критерий Фишера по Монте-Карло; ** – χ^2 с учетом поправки Йетса; *** – ТТФ.				

В целом у наблюдаемых больных превалировала инфильтративная форма туберкулеза легких. Диссеминированный туберкулез наблюдался в 31,4 и 22,5 % случаях, доли больных с фиброзно-кавернозным и кавернозным туберкулезом составили 5,7 % (1 группа) и 4,2 % (2 группа). Очаговый туберкулез и туберкулемы диагностировались на уровне 6–8 % в обеих группах, у 7 человек в группе контроля был туберкулез внутригрудных лимфатических узлов, трахеи, бронхов, осложненный экссудативным плевритом, у единиц наблюдалась казеозная пневмония (в общей сумме у 4 пациентов). У большинства пациентов МЛУ ТБ патологический процесс в легочной ткани протекал с поражением обоих легких ($p = 0,203$, χ^2). Деструкция легочной ткани, выявляемая рентгенологически, чаще встречалась в группе, развившей рецидив (94,3 % и 68,8 %, $p < 0,001$, ТТФ, ОШ = 7,48; 95 % ДИ 1,76–31,80). При этом размеры

деструктивного компонента в легочной ткани были одинаковыми ($p = 0,422$, U) (Таблица 13). Иными словами на возникновение рецидива влияет лишь наличие в легких каверн независимо от их размеров.

Таблица 13 – Размеры полостей распада в легких у больных МЛУ ТБ в начале лечения

Диаметр полости распада в легких, см	Группы наблюдения		p*
	основная группа n = 33	контрольная группа n = 214	
	n (%)	n (%)	
До 4	22 (66,7)	160 (74,8)	0,422
4–5	9 (27,3)	48 (22,4)	
5–6	2 (6,0)	6 (2,8)	
Примечание: * – U-критерий Манна – Уитни.			

Как уже было отмечено выше, у 31 пациента (88,6 %) из группы с рецидивом заболевания и у 121 (38,9 %) среди безрецидивных больных наблюдалось несколько курсов лечения туберкулеза. В связи с этим предполагалось, что у большинства из них будут остаточные изменения в легких. Данное предположение подтвердилось анализом рентгенологических исследований: все пациенты с рецидивом МЛУ ТБ имели те или иные рентгенологические критерии ранее перенесенного туберкулеза. В группе контроля их доля была немного ниже – 85,9 % ($p = 0,013$, ТТФ).

Известно, что условно все остаточные изменения в легких делятся на малые и большие. К малым остаточным изменениям относятся немногочисленные уплотненные очаги и кальцинаты, ограниченный фиброз в пределах одного сегмента, запаянные синусы, междолевые шварты, плевродиафрагмальные и плевромедиастинальные сращения, плевроапикальные и плеврокостальные наслоения шириной до 1 см (с кальцинацией плевры или без нее). К большим остаточным изменениям относятся множественные кальцинированные лимфатические узлы, множественные очаги и фокусы, распространенный фиброз

(более 1 сегмента), цирротические изменения любой протяженности, санированные полости и каверны, фиброторакс, плевропневмосклероз с бронхоэктазами, массивные плевральные наслоения шириной более 1 см (с кальцинацией плевры или без нее). Анализ рентгенограмм легких больных МЛУ ТБ показал, что большие остаточные изменения после эффективного лечения противотуберкулезными препаратами преобладают в группе рецидива ($p = 0,001$, χ^2 , ОШ = 3,38; 95 % ДИ 1,64–6,98) (Таблица 14).

Таблица 14 – Наличие остаточных изменений в легких после эффективного курса химиотерапии больных МЛУ ТБ сравниваемых групп

Признак	Группы наблюдения (n = 346)		P
	основная группа (n = 35)	контрольная группа (n = 311)	
Наличие остаточных изменений в легких	35 (100,0)	267 (85,9)	0,013*
Малые остаточные изменения	14 (40,0)	185 (69,3)	0,001**
Большие остаточные изменения	21 (60,0)	82 (30,7)	
Примечания: * – ТТФ; ** – χ^2 Пирсона.			

Влияние роли остаточных изменений после предыдущих курсов химиотерапии на развитие рецидива подтвердилось в нашем исследовании: у всех пациентов из первой группы были зафиксированы остаточные изменения, в группе сравнения они наблюдались в 85,9 % случаев ($p = 0,013$) (Таблица 15).

Таблица 15 – Характеристика малых остаточных изменений в легких после эффективного курса химиотерапии больных МЛУ ТБ сравниваемых групп

Остаточные изменения в легких	Группы наблюдения (n = 199)		p*
	основная группа n = 14	контрольная группа, n = 185	
Апикальный пневмофиброз с одной или с двух сторон	3 (21,4)	19 (10,3)	0,327
Немногочисленные уплотненные очаги с одной или с двух сторон	11 (78,6)	153 (82,7)	
Плевральные спайки, шварты с одной или с двух сторон	0	13 (7,0)	
Примечание: * – F-критерий Фишера по Монте-Карло.			

Однако при детальном анализе посттуберкулезных изменений необходимо заключить, что конкретно те или иные малые ($p = 0,327$, F) или большие ($p = 0,080$, F) остаточные изменения в легких не влияют на развитие у пациента рецидива туберкулеза (Таблица 16).

Таблица 16 – Характеристика больших остаточных изменений в легких после эффективного курса химиотерапии больных МЛУ ТБ сравниваемых групп

Остаточные изменения в легких	Группы наблюдения (n = 103)		p*
	основная группа n = 21	контрольная группа n = 82	
Множественные уплотненные очаги, фокусы	0	5 (6,0)	0,080
Санитарованные полости, каверны до 2 см	1 (4,8)	18 (22,0)	
Санитарованные полости, каверны более 2 см	20 (95,2)	59 (72,0)	
Примечание: * – F-критерий Фишера по Монте-Карло.			

У пациентов в группе с рецидивом МЛУ ТБ результаты положительной микроскопии на начало лечения встречались чаще, чем у пациентов контрольной группы. Положительный микроскопический анализ на КУМ был установлен у

27 (77,1 %) и 167 (53,7 %) пациентов соответственно ($p = 0,014$, χ^2* , ОШ = 2,91; 95 % ДИ 1,28–6,61) (Таблица 17). При этом разброс частот по степени бактериовыделения на основе положительного посева на начало лечения в группах примерно был одинаков, различий не обнаружено ($p = 0,180$, U) (Таблица 18).

Таблица 17 – Результаты исследования мокроты методом микроскопии больных МЛУ ТБ на начало лечения

Результаты микроскопии	Группы наблюдения (n = 346)		p
	основная группа (n = 35)	контрольная группа (n = 311)	
КУМ +	27 (77,1)	167 (53,7)	0,014*
Скудное	16 (59,3)	74 (23,8)	0,320**
Умеренное	2 (7,4)	29 (9,3)	
Обильное	9 (33,3)	64 (20,6)	
Примечание: * – χ^2 с учетом поправки Йетса; ** – U-критерий Манна – Уитни.			

Таблица 18 – Результаты исследования мокроты больных МЛУ ТБ на начало лечения методом посева

Результаты посева	Группы наблюдения (n = 346)		p*
	основная группа (n = 35)	контрольная группа (n = 311)	
Скудное	18 (51,4)	114 (36,7)	0,180
Умеренное	5 (14,3)	64 (20,6)	
Обильное	12 (34,3)	133 (42,8)	
Примечание: * – U-критерий Манна – Уитни			

Из анамнеза заболевания было выяснено, что у больных двух групп наблюдения давность заболевания и сроки назначения химиотерапии по IV режиму после лабораторной диагностики МЛУ МБТ не различались ($p = 0,126$, U) (Таблица 19).

Таблица 19 – Стаж заболевания, сроки назначения химиотерапии после обнаружения МЛУ МБТ у больных сравниваемых групп

Параметр	Группы наблюдения (n = 346)		p*
	основная группа (n = 35)	контрольная группа (n = 311)	
	Me (Q ₁ ; Q ₂)		
Давность заболевания, годы	2,03 (0,33; 7,47)	0,60 (0,23; 5,90)	0,126
Давность обнаружения МЛУ МБТ, дни	0,06 (0,02; 0,42)	0,06 (0,02; 0,25)	0,790
Примечание: * – U-критерий Манна – Уитни.			

Определяя влияние лекарственной устойчивости МБТ к ПТП на возникновение рецидива было установлено, что первоначально спектр лекарственной устойчивости МБТ у больных в группах наблюдения до начала лечения был одинаков, без установленных различий по каждому противотуберкулезному препарату (Таблица 20). При этом до начала лечения у всех больных выявлялась достаточно высокая частота лекарственной устойчивости МБТ к следующим препаратам: к этионамиду/протионамиду в 40,0 % и 37,8 % случаях, к канамицину/амикацину в 25,7 % и 13,6 %, к капреомицину в 25,7 % и 14,0 % соответственно. Лекарственная устойчивость МБТ к фторхинолонам в группах наблюдения не достигала 20 %, к ПАСК около 5 %, к циклосерину не превышала 1 %. Анализируя в группах наблюдения результаты бактериологического исследования мокроты на фоне лечения больных МЛУ ТБ, были выявлены следующие особенности. В последних положительных бактериограммах мокроты спектр лекарственной устойчивости к ПТП увеличивался по шести ПТП ($p < 0,001$) (Таблица 20).

Таблица 20 – Лекарственная устойчивость МБТ к ПТП у больных МЛУ ТБ сравнимых групп

Лекарственная устойчивость МБТ к ПТП на момент выявления МЛУ		Группы наблюдения (n = 346)		p
		основная группа	контрольная группа	
		n (%)	n (%)	
E	До лечения n ₁ = 35, n ₂ = 308	9 (25,7)	51 (16,6)	0,264**
	В последнем посеве n ₁ = 35, n ₂ = 311	24(68,6)	64 (20,6)	< 0,001*
	p****	< 0,001	0,002	—
Km + Am	До лечения n ₁ = 35, n ₂ = 308	9 (25,7)	42 (13,6)	0,076**
	В последнем посеве n ₁ = 35, n ₂ = 311	19 (54,3)	52 (17,0)	< 0,001*
	p****	0,002	0,022	—
Cm	До лечения n ₁ = 35, n ₂ = 301	9 (25,7)	42 (14,0)	0,113**
	В последнем посеве n ₁ = 35, n ₂ = 306	18 (51,4)	52 (17,0)	< 0,001*
	p****	0,004	0,008	—
Lfx	До лечения n ₁ = 35, n ₂ = 301	6 (17,1)	35 (11,6)	0,502**
	В последнем посеве n ₁ = 35, n ₂ = 306	19 (54,3)	45 (14,7)	< 0,001*
	p****	< 0,001	0,008	—
Eto + Pto	До лечения n ₁ = 35, n ₂ = 299	14 (40,0)	113 (37,8)	0,800*
	В последнем посеве n ₁ = 35, n ₂ = 304	26 (74,3)	125 (41,1)	< 0,001**
	p****	< 0,001	0,035	
PAS	До лечения n ₁ = 35, n ₂ = 300	2 (5,7)	15 (5,0)	0,694***
	В последнем посеве n ₁ = 35, n ₂ = 305	15 (42,9)	24 (7,9)	<0,001*
	p****	< 0,001	0,008	—
Cs	До лечения n ₁ = 35, n ₂ = 301	0 (0,0)	2 (0,7)	1,000***
	В последнем посеве n ₁ = 35, n ₂ = 306	2 (5,7)	5 (1,6)	0,154***
	p****	0,500	0,625	—

Примечание:

ПТП – противотуберкулезный препарат, Km – канамицин, Cm – капреомицин, Lfx – левофлоксацин, E – этамбутол, Eto – этионамид, Pto – протионамид, Cs – циклосерин, PAS – ПАСК, Fq – фторхинолоны

* – χ^2 Пирсона;

** – χ^2 с учетом поправки Йетса;

*** – ТТФ;

**** – χ^2 Мак-Немара.

При этом в процессе лечения в первой группе отмечалась более выраженная отрицательная динамика. Амплификация лекарственной устойчивости возбудителя туберкулеза наблюдалась к: канамицину/амикацину (25,7 % и 54,3 % соответственно ($p < 0,001$, χ^2 , ОШ = 5,91; 95 % ДИ 2,85–12,26); капреомицину (25,7 % и 51,4 %, ($p < 0,001$, χ^2 , ОШ = 5,27; 95 % ДИ 2,55–10,91), этамбутолу (25,7 % и 68,6 %, ($p < 0,001$, χ^2 , ОШ = 0,12; 95 % ДИ 0,06–0,26), этионамиду/протионамиду (40,0 % и 74,3 %, ($p < 0,001$, χ^{2*} , ОШ = 4,30; 95 % ДИ 1,95–9,48), левофлоксацину (17,1 % до 54,3 %, ($p < 0,001$, χ^2 , ОШ = 7,02; 95 % ДИ 3,36–14,66). Необходимо отметить самые высокие темпы амплификации ПАСК. У данных пациентов частота лекарственной устойчивости МБТ к ПАСК увеличивалась в процессе химиотерапии в восемь раз (5,7 % и 42,9 % ($p < 0,001$, χ^2 , ОШ = 8,97; 95 % ДИ 4,08–19,73). В группе сравнения амплификация лекарственной устойчивости МБТ по всем вышеуказанным препаратам на фоне лечения, аналогично, присутствовала у ряда пациентов, но в динамике лечения доля таких пациентов существенно не нарастала, увеличение частот лекарственной устойчивости МБТ к ПТП было в пределах 3–4 %.

Продолжая анализ амплификации МБТ у больных МЛУ ТБ на фоне лечения, необходимо заключить, что в целом у 21 (60,0 %) пациента с рецидивом туберкулеза в процессе лечения наблюдалась амплификация к противотуберкулезным препаратам, что не только усложнило лечение текущего курса химиотерапии, но и повлияло на возникновение рецидива в будущем ($p < 0,001$, χ^2 , ОШ = 12,22; 95 % ДИ 5,69–26,24) (Таблица 21). Подобная тенденция в группе контроля наблюдалась у 34 (10,9 %) больных.

Таблица 21 – Амплификация МБТ к ПТП во время лечения у больных МЛУ ТБ сравниваемых групп

Признак	Группы наблюдения (n = 346)		p*
	основная группа (n = 35)	контрольная группа (n = 311)	
Амплификация во время лечения	21 (60,0)	34 (10,9)	< 0,001
Примечание: * – χ^2 Пирсона.			

Амплификация в группе с рецидивом МЛУ ТБ по сравнению с контрольной группой достоверно чаще наблюдалась к 2 ($p < 0,001$, χ^2 *, ОШ = 6,94; 95 % ДИ 2,31–20,88), 3 ($p = 0,005$, ТТФ, ОШ = 9,90; 95 % ДИ 2,36–41,56), 4 ($p < 0,001$, ТТФ, ОШ = 25,75; 95 % ДИ 4,79–138,46), 5 ($p = 0,028$, ТТФ, ОШ = 18,79; 95 % ДИ 1,66–212,81) противотуберкулезным препаратам. В группе контроля амплификация встречалась реже. Так, например, одновременное нарастание лекарственной устойчивости к 4-м ПТП в группе пациентов, развивающих рецидив заболевания, встречалось в 23,8 % случаях, против 5,9 % в контрольной группе (Таблица 22).

Таблица 22 – Число препаратов, к которым произошла амплификация во время лечения у больных МЛУ ТБ сравниваемых групп

Число ПТП, к которым произошло нарастание лекарственной устойчивости МБТ в процессе лечения	Группы наблюдения (n = 346)		p
	основная группа (n = 21)	контрольная группа (n = 34)	
	n (%)	n (%)	
Амплификация к 1 ПТП	4 (19,1)	17 (50,0)	0,250**
Амплификация к 2 ПТП	6 (28,6)	9 (26,5)	< 0,001*
Амплификация к 3 ПТП	4 (19,1)	4 (11,8)	0,005**
Амплификация к 4 ПТП	5 (23,8)	2 (5,9)	< 0,001**
Амплификация к 5 ПТП	2 (9,5)	1 (2,9)	0,028**
Амплификация к 7 ПТП	0	1 (0,3)	1,000**
Примечания: * – χ^2 с учетом поправки Йетса; ** – ТТФ.			

В таблицах 23 и 24 представлены частота использования и основные схемы ПТП на начало лечения.

Таблица 23 – Частота первоначального использования противотуберкулезных препаратов (ПТП) в режимах химиотерапии больных МЛУ ТБ (n = 346)

ПТП	Z	E	Km	Cm	Lfx/ Mfx	Pto/ Eto	Cs	PAS	Trd
Частота использования, n (%)	332 (95,6)	238 (68,8)	205 (59,2)	120 (34,7)	336 (97,1)	289 (83,5)	289 (83,5)	158 (45,7)	11 (3,2)
Примечания: Z – пиразинамид, Km – канамицин, Cm – капреомицин, Lfx – левофлоксацин, E – этамбутол, Mfx – моксифлоксацин, Eto – этионамид, Pto – протионамид, Cs – циклосерин, PAS – ПАСК, Trd- теризидон.									

Таблица 24 – Основные первоначальные схемы лечения больных МЛУ ТБ сравниваемых групп

Схема ПТП	Группы наблюдения (n = 346)				p*
	основная группа (n = 35)		контрольная группа (n = 311)		
	n	%	n	%	
Cm/Km Cs E Lfx/Mfx Eto/Pto Z	11	31,4	118	37,9	0,600
Cm/Km Cs Lfx/Mfx Pas Eto/Pto Z	7	20,0	71	22,8	
Cm/Km Cs E Lfx/Mfx Pas Z	2	5,7	26	8,4	
Cm/Km E Lfx/Mfx Eto/Pto Z	1	2,9	18	5,8	
Cm/Km E Lfx/Mfx Pas Eto/Pto Z	2	5,7	10	3,2	
Cm/Km Cs E Lfx/Mfx Pas Eto/Pto	2	5,7	6	1,9	
Cm/Km Cs Lfx/Mfx Pas Z	1	2,9	5	1,6	
Cm/Km Cs Lfx/Mfx Eto/Pto Z	0	0,0	5	1,6	
Cm/Km Cs E Eto/Pto Z	0	0,0	3	1,0	
Cs E Lfx/Mfx Pas Z	0	0,0	3	1,0	
Другие	9	25,7	46	14,8	
Примечания: Z – пиразинамид, Km – канамицин, Cm – капреомицин, Lfx – левофлоксацин, E – этамбутол, Mfx – моксифлоксацин, Eto – этионамид, Pto – протионамид, Cs – циклосерин, Pas – ПАСК.					
* – F-критерий Фишера по Монте-Карло.					

Наличие различного сочетания ЛУ МБТ к ПТП у больных с МЛУ ТБ привело к разнообразию схем химиотерапии. Анализируя схемы назначенного лечения, было определено, что фторхинолоны (Fq) ($p = 0,228$, ТТФ) и инъекционные ($p = 0,706$, ТТФ) ПТП применялись у большинства больных двух групп наблюдения (Таблица 25).

Таблица 25 – Наличие фторхинолонов и инъекционных препаратов в первоначальной схеме лечения больных МЛУ ТБ сравниваемых групп

Параметры	Группы наблюдения (n = 346)		p*
	основная группа (n = 35)	контрольная группа (n = 311)	
В курсе лечения присутствовал инъекционный ПТП (n, %)	34 (97,1)	292 (93,9)	0,706
В курсе лечения присутствовал фторхинолон ПТП (n, %)	33 (94,3)	304 (97,7)	0,228
Примечание: * точный критерий Фишера.			

При этом число ПТП в схемах химиотерапии значимо не различалось ($p = 0,940$, U) у пациентов в группах наблюдения (Таблица 26).

Таблица 26 – Количество ПТП в схемах лечения больных МЛУ ТБ сравниваемых групп

Число препаратов в схемах химиотерапии	Группы наблюдения (n = 346)		p*
	основная группа (n = 35)	контрольная группа (n = 311)	
3	0 (0,0)	2 (0,6)	0,940
4	2(5,7)	11 (3,5)	
5	6 (17,1)	56 (18,0)	
6	26 (74,3)	240 (77,2)	
7	1 (2,9)	2 (0,6)	
Примечание: * – U-критерий Манна – Уитни.			

Если в схемах лечения объединить фторхинолоны (Fq) и инъекционные ПТП, то преобладающим в химиотерапии больных МЛЮ ТБ был следующий набор ПТП: Cm/Km Cs E Lfx/Mfx Eto/Pto Z (в 31,4 и 37,9 % в первой и второй группах соответственно) (Таблица 24). При лечении пациентов обеих групп во всех случаях применялись инъекционные препараты – 94,2 % случаев (у 326 больных из 346): группа рецидива – 97,1 %, группа контрольная – 93,9 %. У 337 (97,4 %) пациентов применялись фторхинолоны (Lfx или Mfx) (Таблица 23). Кроме того, у данных больных в большинстве случаев, у 332 (96,0 %), в лечении применялся Z, у 158 пациентов (45,7 %) применялся PAS, у 289 пациентов (83,5 %) – Pto, у 238 пациентов (68,8 %) – E.

В связи с тем, что в начале лечения у пациентов готовы лишь результаты микроскопии мокроты, схема ПТП могла быть назначена некорректно, как по числу, так и по спектру ЛУ МБТ. Соответственно после получения результатов посева фтизиатры корректировали химиотерапию пациенту.

Статистический анализ не выявил зависимости между наступлением рецидива и первоначальной некорректной схемой лечения больных МЛЮ ТБ ($p = 0,655$, ТТФ). По результатам обработки данных оказалось, что в большинстве случаев схема была полной и соответствовала спектру ЛУ МБТ у больных обеих групп – у 33 (94,3 %) и 298 (95,8 %), соответственно (Таблица 27).

Таблица 27 – Схема химиотерапии на начало лечения у больных МЛЮ ТБ сравниваемых групп

Параметр	Группы наблюдения (n = 346)		p*
	основная группа (n = 35)	контрольная группа (n = 311)	
	n (%)	n (%)	
Схема лечения соответствует спектру ЛУ МБТ	33 (94,3)	298 (95,8)	0,655

В таблице 20 были представлены данные по первоначальному спектру ЛУ МБТ к ПТП при впервые установленной МЛЮ возбудителя у наблюдаемых пациентов. Ниже, в таблице 28, приводятся данные результатов тестов на лекарственную чувствительность МБТ в последнем положительном посева у

пациентов сравниваемых групп в процессе лечения.

Таблица 28 – Количество препаратов, к которым сохранялась лекарственная чувствительность МБТ в последнем положительном посеве мокроты больных МЛУ ТБ сравниваемых групп

Количество ПТП	Группы наблюдения (n = 346)		p
	основная группа (n = 35)	контрольная группа (n = 311)	
0	2 (5,7)	2 (0,64)	0,052**
1	8 (22,9)	4 (1,3)	< 0,001**
2	7 (20,0)	7 (2,3)	< 0,001***
3	3 (8,6)	20 (6,4)	0,716**
4	2 (5,7)	19 (6,1)	1,000**
5	2 (5,7)	42 (13,5)	0,283**
6	4 (11,4)	87 (28,0)	0,041**
7	7 (20,0)	123 (39,6)	0,038***
8	0	7 (2,3 %)	1,000**

Примечания:
 * – U-критерий Манна – Уитни;
 ** – точный критерий Фишера;
 *** – χ^2 с учетом поправки Йетса.

Табличные сведения демонстрируют, что пациенты с сохраненной лекарственной чувствительностью МБТ к 6-ти ($p = 0,041$, ТТФ, ОШ = 0,33; 95 % ДИ 0,11–0,97) и 7-ми ПТП ($p = 0,038$, χ^2 *, ОШ = 0,38; 95 % ДИ 0,16–0,90) излечивались от туберкулеза и относились к группе безрецидивных больных, в то время, как в группе пациентов, развивающих рецидив в будущем, у каждого второго резерв ПТП к концу лечения ограничивался лишь одним ($p < 0,001$, ТТФ, ОШ = 22,74; 95 % ДИ 6,43–80,42) или двумя ПТП ($p < 0,001$, χ^2 *, ОШ = 10,86; 95 % ДИ 3,55–33,17) (Таблица 28).

Заканчивали они лечение на неполной схеме химиотерапии (Таблица 29). Только лишь у 19 (54,3 %) больных в первой группе наблюдения схема химиотерапии соответствовала действующим в то время клиническим

рекомендациям по лечению МЛУ ТБ, в группе контроля, наоборот, у 293 (94,2 %) пациентов она была корректной ($p < 0,001$, χ^2 , ОШ = 0,07; 95 % ДИ 0,03–0,17).

Таблица 29 – Схема химиотерапии в конце лечения у больных МЛУ ТБ сравниваемых групп

Параметр	Группы наблюдения (n = 346)		p*
	основная группа (n = 35)	контрольная группа (n = 311)	
Схема лечения соответствует спектру ЛУ МБТ	19 (54,3)	293 (94,2)	< 0,001
Примечание: * – χ^2 Пирсона.			

Сравнительный анализ первоначальных и последующих схем химиотерапии больных МЛУ ТБ четко определил установленную выше высокую скорость формирования амплификации МБТ к ПТП в группе пациентов, развивающих рецидив заболевания. Так, частота первоначального назначения больным МЛУ ТБ двух групп наблюдения ПТП, включающих инъекционные, фторхинолоны и ПТП первого ряда, значимо не различалась – в 65,7 % и 78,1 % соответственно. В то время как в динамике лечения подобные схемы химиотерапии встречались в два раза реже у пациентов первой группы наблюдения (34,3 %), в группе сравнения частота их применения не изменилась (79,2 %). Аналогичные трансформации происходили по другим схемам химиотерапии. В связи с этим в группе пациентов, развивающих рецидив МЛУ ТБ, назначались менее эффективные ПТП, чем в контрольной группе, а иногда их было недостаточно (см. Таблицы 28, 29 и 30). Так, при анализе применения схем ПТП в режимах лечения больных на конец лечения, в группе безрецидивного течения заболевания почти в половине случаев использовалась эффективная схема, включающая Cm/Km/Am Cs E Eto/Pto Lfx/Mfx Z (в 17,1 % случаях в основной и в 41,2 % контрольной, $p < 0,010$, χ^{2*} , ОШ = 0,30; 95 % ДИ 0,12–0,73). У 18 (51,4 %) пациентов группы рецидива и 33-х (10,6 %) группы контроля были назначены индивидуальные схемы химиотерапии

с числом препаратов менее 5, что было сопряжено с высокими шансами развития рецидива ($p < 0,010$, χ^2 , ОШ = 8,92; 95 % ДИ 4,19–18,97) (см. Таблицу 30).

Таблица 30 –Схемы химиотерапии в конце курса лечения больных МЛУ ТБ сравнимых групп

Схема ПТП	Группы наблюдения (n = 346)		p
	основная группа, n = 35	контрольная группа, n = 311	
	n (%)	n (%)	
См/Км/Ам Cs E Eto/Pto Lfx/Mfx Z	6 (17,1)	128 (41,2)	0,010***
См/Км/Ам Cs E Eto/Pto Lfx/Mfx Pas Z	3 (8,6)	48 (15,4)	0,448*****
См/Км/Ам Cs E Lfx/Mfx Pas Z	1 (2,9)	26 (8,4)	0,500*****
См/Км/Ам Cs Eto/Pto Lfx/Mfx Pas Z	0	26 (8,4)	0,091*****
См/Км/Ам Cs E Eto/Pto Z	0	10 (3,2)	0,607*****
См/Км/Ам Cs Eto/Pto Lfx/Mfx Z	2 (5,7)	8 (2,6)	0,268*****
См/Км/Ам Cs Eto/Pto Pas Z	2 (5,7)	5 (1,6)	0,151*****
См/Км/Ам Cs E Eto/Pto Pas Z	0	6 (1,9)	1,000*****
См/Км/Ам Cs Lfx/Mfx Pas Z	1 (2,9)	5 (1,6)	0,475*****
Cs E Eto/Pto Lfx/Mfx Pas Z	0	6 (1,9)	1,000*****
Cs Eto/Pto Lfx/Mfx Pas Z	2 (5,7)	4 (1,3)	0,115*****
См/Км/Ам Cs E Eto/Pto Lfx/Mfx Trd Z	0	3 (1,0)	1,000*****
Cs E Eto/Pto Lfx/Mfx Z	0	3 (1,0)	1,000*****
Другие	18 (51,4)	33 (10,6)	< 0,001**

< 0,001*

Примечания:
 * – F-критерий Фишера по Монте-Карло;
 ** – χ^2 Пирсона;
 *** – χ^2 с учетом поправки Йетса;
 **** – ТТФ.

Одним из главных критериев эффективности химиотерапии больных туберкулезом является прекращение бактериовыделения. Анализируя сроки абациллирования у пациентов МЛУ ТБ (Таблица 31), которые эффективно завершили лечение и в дальнейшем развили рецидив туберкулеза, было установлено, что для них характерен больший период бактериовыделения и, соответственно, средний срок лечения. Так, через два месяца лечения в группе сравнения (без рецидива) прекратили выделять МБТ большая часть больных – 239 (76,8 %), в первой группе абациллированных пациентов было в два раза меньше – 14 (40,0 %) ($p < 0,001$, χ^2* , ОШ = 4,98; 95 % ДИ 2,41–10,29). Соответственно и средний срок лечения пациентов в основной группе наблюдения составил по медиане 21,1 мес. лечения, в группе сравнения – 20,0 мес. ($p < 0,001$, U).

Таблица 31 – Время прекращения бактериовыделения по микроскопии мазка мокроты и посева на плотные питательные среды (мес.) и средний срок лечения больных МЛУ ТБ

Параметры		Группы наблюдения (n = 346)		p
		основная группа, n = 35	контрольная группа, n = 311	
		Me (Q ₁ ; Q ₂)		
Время прекращения бактериовыделения, мес., Me (Q ₁ ; Q ₂)***	мазок	4 (0; 18)	1 (0; 9)	0,090*
	посев	8,4 (6,2; 13,9)	5,9 (3,9; 9,4)	< 0,001*
Прекращение бактериовыделения (посев) Me (Q ₁ ; Q ₂)	до 2 мес.	14 (40,0)	239 (76,9)	< 0,001**
	более 2-х мес.	21 (60,0)	72 (23,2)	
Срок лечения, мес.		21,1 (20,3; 22,9)	20,0 (19,0; 20,9)	< 0,001*
Примечания: * – U-критерий Манна – Уитни; ** – χ^2 с учетом поправки Йетса.				

С учетом высоких темпов амплификации МБТ у больных основной группы и, несмотря на очевидность, в исследовании не подтвердилось влияние приверженности лечению на наступление в будущем рецидива МЛУ ТБ. Анализируя данные таблиц 32 и 33, необходимо заключить, что в основной группе частота больных с нарушением приема ПТП (37,1 %) превышает значения группы контроля (23,2 %), но значимых различий с контрольной группой выявлено не было, аналогично, количество пациентов, которые самостоятельно прервали лечение, значимо не различалось в группах, хотя значения частоты перерывов в лечении существенно различались – 20,0 % и 8,7 соответственно ($p = 0,067$, χ^2 *). В целом процент принятых доз у пациентов основной и группы сравнения практически был идентичен (Таблица 33).

Таблица 32 – Частота прерывания лечения/нарушение режима у больных МЛУ ТБ сравниваемых групп

Тип прерывания лечения	Группы наблюдения (n = 346)		p
	основная группа, n = 35	контрольная группа, n = 311	
	n (%)	n (%)	
Нарушение режима	13 (37,1)	72 (23,2)	0,068*
Самостоятельный отрыв от лечения	7 (20,0)	27 (8,7)	0,067**
Примечания: * – χ^2 Пирсона; ** – χ^2 с учетом поправки Йетса.			

Таблица 33 – Доля принятых доз ПТП от должного у больных МЛУ ТБ сравниваемых групп

Группы наблюдения (n = 346)		p*
основная группа, n = 35	контрольная группа, n = 311	
Me (Q ₁ ; Q ₂)		
93,00 (89,0; 96,0)	92,0 (86,0; 96,0)	0,526
Примечание: * – U-критерий Манна – Уитни.		

3.2 Регрессионный анализ данных пациентов по выявлению предикторов рецидива туберкулеза с множественной лекарственной устойчивостью

В регрессионном анализе рассматривались параметры, предикторная роль которых в возникновении рецидива у излеченных пациентов с МЛУ ТБ была доказана с помощью однофакторного анализа при уровне значимости $p < 0,05$. С целью исключения мультиколлинеарности при построении бинарной логистической регрессии из переменных, характеризующих один и тот же параметр, были отобраны факторы с более высоким значением ОШ и уровнем значимости стремящемуся к 0. Такими величинами в исследовании стали: пенсионный возраст пациента, наличие инвалидности, болезни, вызванной вирусом иммунодефицита человека, психического заболевания, полости распада, остаточных изменений на начало курса и по завершению лечения, количество курсов противотуберкулезной терапии, возникновение амплификации к принимаемым препаратам, лекарственная чувствительность МБТ к препаратам в последнем посеве, а также тип туберкулезного процесса, табакокурение, срок лечения пациента и время абациллирования. Для построения модели применялся пошаговый отбор включения переменных в уравнение прямой. Зависимой переменной являлся исход: возникновение рецидива или безрецидивное излечение. Порог отсечения был принят 0,5: если при расчете вероятность возникновения события более 0,5, считалось, что событие – «рецидив туберкулеза» наступит. Адекватность полученной модели оценивали с помощью критерия Хосмера – Лемешова. Проверка значимости отличия коэффициентов от нуля (решение о включении предиктора в модель) проводилась при помощи статистики Вальда, использующей распределение χ^2 -квадрат, которая представляет собой квадрат отношения соответствующего коэффициента к его стандартной ошибке. Качество приближения регрессионной модели оценивалось при помощи функции правдоподобия, мерой правдоподобия является отрицательное удвоенное значение логарифма этой функции.

Для оценки качества модели использовались: качество подгонки модели

(χ^2), мера Нэйджелкерка, значения коэффициента конкордации, чувствительности и специфичности, ROC-анализ.

Результаты многофакторного анализа и параметры регрессионной модели представлены в таблице 34.

Таблица 34 – Параметры многофакторного анализа признаков больных МЛУ ТБ

Показатели		b	Среднеквадратическая ошибка	ОШ	exp (b)	95 % ДИ (exp b)	Критерий Вальда
Константа	B ₀	-46,940	11,687	—	—	—	16,131 p < 0,001
Пенсионер	x ₁	2,257	0,906	ОШ = 2,86 [1,14–7,20]	9,550	1,617– 56,402	6,202 p = 0,013
ВИЧ	x ₂	5,370	1,779	ОШ = 19,19 [5,29–69,56]	214775	6,567– 7 024,022	9,106 p = 0,003
Количество предыдущих курсов лечения	x ₃	3,172	0,682	U = 8,61 P = 0,00	23,857	6,273– 90,726	21,664 p < 0,001
Наличие амплификации	x ₄	2,057	0,924	P = 0,000 ОШ = 12,22 [5,69–26,24]	7,824	1,279– 47,847	4,958 p = 0,026
Деструкция на начало лечения	x ₅	2,882	1,560	ОШ = 7,48 [1,76–31,80]	17,854	0,839– 380,047	3,412 p = 0,065
Срок лечения	x ₆	1,528	0,465	U = -5,38 p < 0,001	4,607	7,853– 11,459	10,801 p = 0,001
Примечание: b – коэффициент регрессии; exp (b) – экспонированное значение шансов; U-критерий Манна – Уитни							

Оценивались весовые коэффициенты 6 основных факторов, имеющих большой вклад в формирование рецидива у пациентов с МЛУ ТБ, параметры включения предикторов в уравнение, качество предсказания модели.

Уравнение регрессии:

$$y = 2,257 \times x_1 + 5,370 \times x_2 + 3,172 \times x_3 + 2,057 \times x_4 + 2,882 \times x_5 + 1,528 \times x_6 - 46,940$$

Параметры регрессии:

- качество подгонки модели $\chi^2 = 181,134$ $p < 0,001$;
- R^2 (Нэйджелкерка) = 0,848;
- критерий Хосмера – Лемешева $\chi^2 = 0,575$ $p = 1,000$;
- снижение коэффициента логарифмического правдоподобия до 45,576.

Вероятность возникновения события рассчитывается по следующей формуле:

$$p = \frac{1}{1 + e^{-y}}$$

Далее полученное значение вероятности сравнивается с порогом отсечения равным 0,5.

Интерпретация полученных в ходе регрессионного анализа параметров.

На основании расчета коэффициентов регрессии (b), используемых в уравнении для предикторов, а также экспонированного значения шансов (exp (b)), который показывает во сколько раз изменяется шанс возникновения изучаемого события, если значения одного из предикторов изменяется на единицу при фиксированных значениях прочих предикторов, факторы риска определены следующим образом (Таблица 34). Они представлены ниже в порядке уменьшения влияния их на рецидив заболевания:

- наличие у пациента ВИЧ,
- число предыдущих курсов химиотерапии,
- деструкция в легочной ткани на начало лечения,

- пенсионный возраст,
- нарастание спектра лекарственной устойчивости МБТ к ПТП в период лечения (амплификация),
- срок лечения.

Несмотря на то, что критерий Вальда показывает, что предиктор «деструкция в легких на начало лечения» стоит исключить из модели, было принято решение обратное, обоснованием является значительное снижение показателей, характеризующих качество модели. Одной из таких характеристик является мера определенности Нэйджелкерка (другое название «коэффициент детерминации»), обозначение – R^2 . Она характеризует часть дисперсии, объясняемой с помощью построенной модели логистической регрессии. В нашем случае коэффициент детерминации составил 84,8 %. Еще одним параметром для оценки качества модели является процент конкордации, отражающий долю верно классифицированных пациентов.

Также в данном метаматематическом анализе оценивались такие параметры, как: чувствительность (Se) и специфичность (Sp) классифицируемой модели. Чувствительность представляет собой долю истинно положительных случаев (процент верно классифицируемых пациентов с рецидивом). Специфичность – доля истинно отрицательных случаев, т. е. сколько пациентов с отсутствием рецидива заболевания верно соотнесено моделью к группе контроля. Однако, не стоит забывать о том, что оба показателя должны иметь достаточно высокие значения, в обратном случае есть риск получить ложноположительные (модель распознает безрецидивного как больного с рецидивом) – «гипердиагностика» и ложноотрицательные результаты классификации (рецидивный больной воспринимается моделью как безрецидивный) – «гиподиагностика».

В таблице 35 представлены результаты классификации модели, где коэффициент конкордации составил 97,7 %.

Таблица 35 – Параметры классификации регрессионного анализа

Наблюдаемые значения	Предсказанные значения		
	рецидив	нет рецидива	доля правильных ответов
Рецидив	31	4	88,6 (Se)
Нет рецидива	4	307	98,7 (Sp)
Доля корректно предсказанных признаков			97,7

Чувствительность метода 88,6 %, специфичность – 98,7 %. В связи с этим предложенная модель позволяет распознать 31 пациента с рецидивом заболевания.

Самым наглядным методом (Рисунок 3), позволяющим оценить качество модели, является построение ROC-кривой на основании ранее рассчитанных показателей: специфичность и чувствительность. Данный график показывает зависимость количества верно классифицированных истинно положительных случаев от количества классифицированных ложноотрицательных. После построения графика вычисляется площадь под кривой (AUC – Area Under Curve). Чем ближе значение этого параметра к единице (идеальный тест), тем лучше качество модели. Для трактовки показателя AUC существуют классификации по данному параметру, используемый нами пример классификации модели представлен в таблице 36.

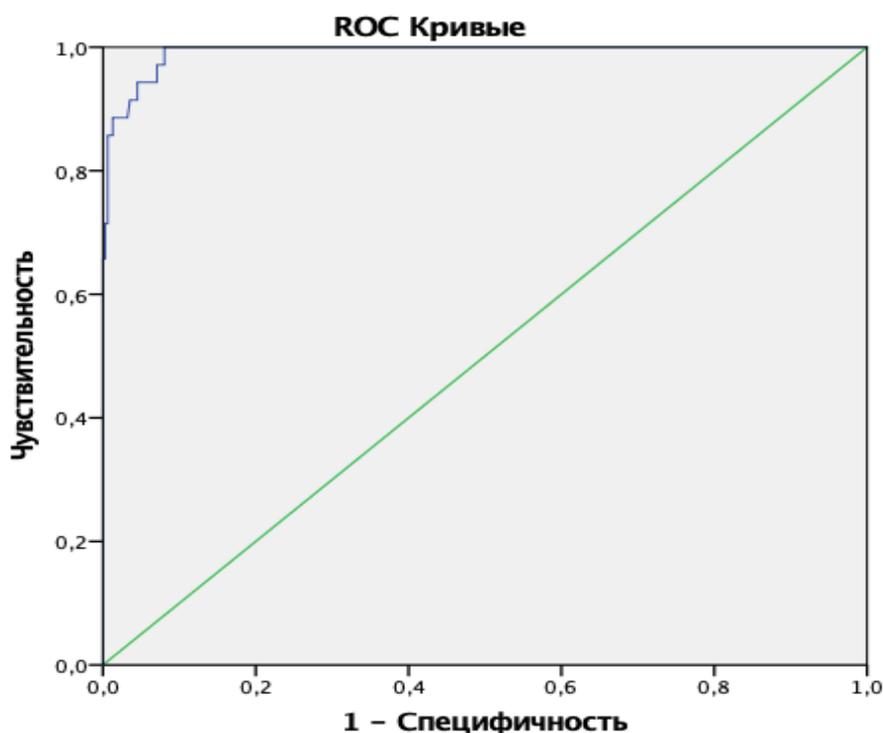


Рисунок 3 – ROC-кривая, построенная на предсказанных значениях вероятности развития рецидива МЛУ ТБ на основе регрессионного анализа

Таблица 36 – Оценка качества построенной модели на основе ROC-анализа

Качество модели	Значение AUC
Отличное	0,9–1,0
Очень хорошее	0,8–0,9
Хорошее	0,7–0,8
Среднее	0,6–0,7
Неудовлетворительное	0,5–0,6

Полученная регрессионная модель является адекватной с высоким уровнем значимости ($\chi^2 = 181,134$ $p < 0,001$), результат теста Хосмера – Лемешова – $\chi^2 = 0,575$; значимость соответствия – $p = 1,000$. Процент объясненной дисперсии – 84,8.

Таким образом, результаты ROC-анализа демонстрируют площадь под кривой AUC равной 0,992, что говорит о высоком (отличном) качестве модели, следовательно, полученное уравнение регрессии обоснованно рекомендует ее к

использованию в практике для оценки вероятности развития рецидива у излеченных пациентов с МЛУ ТБ.

Созданная методом логистической регрессии, модель классификации была протестирована на двух случайным образом выбранных пациентах, принадлежащих к группам исследования.

Пациент № 11

Мужчина, 45 лет, неработающий, имеет инвалидность в связи с наличием психического заболевания. Житель Томской области, одинокий. Контакта с больным туберкулезом зафиксировано не было. Мужчина употреблял алкоголь, постоянный курильщик. ИМТ – 21,2. В анамнезе болел туберкулезом, три курса химиотерапии (стаж заболевания – 12,2 года, давность обнаружения МЛУ МБТ – 5,8 года). Сопутствующих заболеваний нет. Пациент был зарегистрирован с рецидивом туберкулеза с диагнозом: инфильтративный туберкулез легких S₂, S₃, S₆ правого легкого в фазе распада. При рентгенологическом обследовании пациента наблюдалось поражение двух долей правого легкого. На начало результаты микроскопии отрицательные. В первом положительном посеве мокроты имел лекарственную устойчивость (помимо изониазида и рифампицина) к стрептомицину, этамбутолу, этионамиду. Получал лечение в стационаре в течение 24,2 мес. по IV режиму химиотерапии. За время химиотерапии амплификации к ПТП не наблюдалось. Последний положительный посев был зафиксирован через 16,7 мес. лечения. На начало лечения схема химиотерапии включала этионамид, в связи с этим была сформирована некорректно (минимальное число (3) эффективно работающих в режиме лечения ПТП), в дальнейшем схема ПТП была скорректирована и включала инъекционный препарат и фторхинолон. Несмотря на нарушение режима в период нахождения в лечебном учреждении, пациент был эффективно излечен с минимальными остаточными изменениями в легких в виде уплотненных очагов, приверженность лечению составила 99 %.

Через 4,9 года у данного пациента было зафиксировано развитие позднего рецидива МЛУ ТБ.

Применение модели логистической регрессии.

Входные данные в уравнение прямой: пенсионный возраст = 0; наличие ВИЧ = 0; число курсов = 3; наличие амплификации = 0; деструкция на начало лечения = 1; средний срок лечения = 24,2. При расчете вероятность возникновения рецидива туберкулеза у данного пациента составляет 92 % ($p = 0,918$).

Пациент № 56

Мужчина, 35 лет, работающий. Житель Томской области, проживает вместе с семьей. Контакт с больным туберкулезом не зафиксирован. Пациент без вредных привычек. Ранее туберкулезом не болел, в анамнезе два курса химиотерапии (давность обнаружения заболевания – 0,6 года, давность обнаружения МЛУ – 0,04 года), сопутствующей патологией неотягощен, ИМТ – 27,8. Пациент зарегистрирован с диагнозом – инфильтративный туберкулез S₁, S₂ в фазе инфильтрации. Результаты рентгенологического исследования: поражение одной доли правого легкого без распада. На начало результаты микроскопии отрицательные. В первом положительном посеве мокроты имел лекарственную устойчивость (помимо изониазида и рифампицина) к стрептомицину. Пациент лечился амбулаторно (ФАП) в течение 20,6 мес. За время лечения амплификация произошла к 1 ПТП – этионамиду. Последний положительный посев наблюдался спустя 8,8 мес. лечения, после него достигнуто стойкое абациллирование. На начало лечения схема была подобрана некорректно (число ПТП, принимаемых пациентом, к которым не было зафиксировано ЛУ было минимальным (2), в дальнейшем схема ПТП была скорректирована и включала инъекционный препарат и фторхинолон. Нарушение режима в период лечения не зафиксировано, пациент был эффективно излечен без остаточных изменений в легких, приверженность лечению составила 99 %. За время диспансерного наблюдения рецидив туберкулеза у данного пациента не зафиксирован.

Применение модели логистической регрессии

Входные данные в уравнение прямой: пенсионный возраст = 0; наличие

ВИЧ = 0; число курсов = 2; наличие амплификации = 1; деструкция на начало лечения = 0; средний срок лечения = 20,6. Вероятность возникновения рецидива ТБ у данного пациента составляет 0,1 % ($p = 0,001$).

Таким образом, метод логистической регрессии определил прогностическую функцию: оба пациента были классифицированы по группам наблюдения верно.

По результатам тестирования модели можно сделать вывод, что регрессионный анализ обладает достаточно хорошей прогностической способностью и может быть использован для предсказания возникновения рецидива у больных туберкулезом органов дыхания с МЛУ возбудителя.

Для достижения большей чувствительности в исследовании далее были рассмотрены другие методики предсказания возникновения события, которые в отличие от данной логистической регрессии анализируют полный объем входных данных пациента.

3.3 Алгоритмы машинного обучения в прогнозировании рецидива заболевания у больных туберкулезом с множественной лекарственной устойчивостью

Выделение факторов риска рецидива МЛУ ТБ с целью создания прогностических моделей на основе алгоритмов машинного обучения традиционно начинался с подготовки данных. Выбирались метрики, подбирались алгоритмы и определялись их гиперпараметры, с помощью которых оптимизировалась их работа. Первоначальный этап работы состоял из очистки данных, при котором из выборки удалялись выбросы – значения, сильно выбивающиеся из общего тренда или распределения. Далее проводилась нормализация и стандартизация для численных данных, кодирование для категориальных данных. Данный этап был необходим, так как большинство алгоритмов машинного обучения способно работать только с численными данными. Изначально, было выбрано 106 признаков (исключены дублирующие и

несущественные, например, гражданство), в итоге в анализ включены 97 признаков.

Когда данные были обработаны, их делили на две группы: целевой признак – рецидив/нет рецидива, и признаки, которые будут обрабатываться алгоритмами машинного обучения для прогнозирования целевого признака.

Для построения классифицирующих моделей были выбраны следующие методы: дерево решений, случайный лес, XGBoost (градиентный бустинг) и логистическая регрессия. При построении моделей использовались инструменты библиотеки `scikit-learn` Version 0.24.2 в интерактивной облачной среде работы с программным кодом `Google Colaboratory`.

К достоинствам деревьев решений можно отнести простоту в применении и интерпретации, что является крайне важным для клиницистов, а также возможность обработки любых (непрерывных и категориальных) признаков и быстрой оценки их значимости. Важность признака при использовании этого подхода определяется на основе изменения доли неоднородности в результате расщепления признакового пространства при использовании правила, ассоциированного с этим признаком [82, 97].

Модель случайного леса представляет собой ансамбль оценщиков, которыми являются деревья решений, каждый из которых настроен на наборе из случайным образом выбранных с возвратом из первоначальной выборки тренировочных образцов [146]. Предполагается, что использование серии деревьев с разными параметрами делает модель более устойчивой к так называемым шумам, которые могут повлиять на правила, получаемые в случае построения классификатора только одним деревом. Важность признаков для модели случайного леса выражается как усредненное уменьшение неоднородности, вычисленное из всех оценщиков в лесе.

Алгоритм XGBoost представляет собой также ансамбль методов деревьев, но использующий принцип бустинга слабых учеников при помощи архитектуры градиентного спуска. Многие современные практические приложения и исследования в здравоохранении используют этот подход для решения

диагностических задач [113, 144, 204, 205]. Для оценки важности признаков чаще используются подходы, основанные на уменьшении неоднородности, как и в предыдущем алгоритме, а также путем вычисления значений Шепли [116, 145, 206,].

Логистическая регрессия является популярным и неоднократно испытанным инструментом бинарной классификации для исследований в здравоохранении. Из-за того, что алгоритм использует только числовые данные, их переводят в категориальные признаки путем прямого кодирования или заданием правил в случае порядковых категорий. При приведении признаков к единой шкале оценивается значимость каждого из признаков, как значение веса, ассоциированное с ним в уравнении регрессии [123].

Для объяснения важности признака на прогноз рецидива имеет решение также знак, стоящий перед коэффициентом w_i в уравнении регрессии:

$$f(x_1, x_2, \dots, x_n) = \frac{1}{1 + \exp - (\sum_{i=1}^n w_i x_i)},$$

где $f(x_1, x_2, \dots, x_n)$ – вероятность прогноза рецидива для пациента с предикторами (x_1, x_2, \dots, x_n) ;

x_i – значение i -го предиктора исследуемого пациента. При этом отрицательные коэффициенты w_i связаны с признаками x_i повышающими вероятность рецидивов, а положительные – уменьшающие.

Для расчета важности для моделей, использующих деревья, применяется анализ на основе целевой функции прироста информации при каждом расщеплении:

$$IG(D_p, f) = I(D_p) - \sum_{j=1}^m \frac{N_j}{N_p} I(D_j),$$

где f – признак, использующийся для расщепления признакового пространства;

D_p и D_j – набор данных родительского и j -того дочернего узла соответственно;

$I(D_p)$ – мера неоднородности в узле p , N_j и N_p – число образцов в родительском и j -том дочернем узле соответственно;

$m = 2$ – число регионов, полученных при расщеплении.

Чувствительность и специфичность моделей классификаторов оценивалась при помощи ROC-анализа. Количественная интерпретация результатов проводилась по ROC-кривым с оценкой показателя AUC (Area under ROC curve – площадь под ROC-кривой).

В первом эксперименте вся выборка была поделена в отношении 70 % и 30 % с сохранением баланса классов для обучающего и тестового набора, соответственно. Данные из тестовой выборки не участвовали в процессе обучения. Это необходимо было для валидации полученных моделей и исключения случая переобучения, т. е. ситуации, когда модель хорошо описывает данные из обучающей выборки, но плохо применима для тестовых данных.

Гиперпараметры, каждого из используемых алгоритмов, были отобраны с помощью сеточного поиска на наборе значений, приведенных в таблице 37.

Таблица 37 – Гиперпараметры классификаторов, выбранные с помощью сеточного поиска

Классификатор	Гиперпараметр: значения использованные при анализе	Выбранное значение
Дерево решений	Критерий расщепления: энтропия, коэффициент Джини	Энтропия
	Максимальная глубина дерева: 3–10	6
	Пороговое значение критерия расщепления для разделения данных в узле: 0.3, 0.2, 0.1	0.2
Случайный лес	Количество оценщиков: 5–19	13
	Критерий расщепления: энтропия, коэффициент Джини	Энтропия
	Минимальное количество примеров в узле для расщепления: 2–6	2
	Максимальная глубина одного дерева в модели: 3–10, не ограничена	Не ограничена

Продолжение таблицы 27

Классификатор	Гиперпараметр: значения использованные при анализе	Выбранное значение
XGBoost	Алгоритм бустинга: gbtree – настройка на основе слабых учеников, gblinear – настройка с использованием линейной регрессии с L1 и L2 сжатием, dart – метод обучения с использованием прореживания	gbtree – настройка на основе слабых учеников
	Уменьшение размера шага η : 0.1, 0.3, 0.5, 0.7	0.1
	Количество деревьев градиентного бустинга: 5–19	11
	Коэффициент L2-регуляризации λ : 0.01, 0.1, 0.0, 1.0, 10.0	1.0
	Коэффициент L1-регуляризации α : 0.01, 0.1, 0.0, 1.0, 10.0	0.0
	Максимальная глубина дерева: 3–10	3
Логистическая регрессия	Инверсия коэффициента регуляризации C: 0.001, 0.01, 0.1, 1.0, 100.0, 1000.0	1.0
	Функция штрафа: L1-регуляризация, L2-регуляризация, ElasticNet	L2-регуляризация
	Алгоритм оптимизации: алгоритм Ньютона – сопряженных градиентов, алгоритм BFGS с ограниченной памятью, LIBLINEAR оценка линейной комбинации признаков	алгоритм BFGS с ограниченной памятью

Для расчета степени неоднородности методом сеточного поиска выбрана мера неоднородности – энтропия, используемая для оценки важности признаков:

$$I_H = - \sum_{i=1}^c p(i|t) \log_2 p(i|t),$$

где $p(i|t)$ – доля образцов, которая принадлежит классу i для узла t ;

$c = 2$ – количество классов (‘Рецидив’, ‘Нет рецидива’).

По итогам первого эксперимента были получены следующие значения метрик качества моделей для каждого из алгоритмов МО: чувствительность 0,73, 0,55, 0,64 и 0,57, специфичность 0,97, 1,0, 1,0, 0,98 для моделей дерева решений,

случайного леса, XGBoost и логистической регрессии соответственно.

Необходимо отметить, что все деревья МО строятся независимо друг от друга по следующей процедуре: генерируется случайная подвыборка с повторением размером n из обучающей выборки; строится решающее дерево, классифицирующее параметры данной подвыборки, причём в ходе создания очередного узла дерева будет выбираться признак, на основе которого производится разбиение, не из всех M признаков, а лишь из m случайно выбранных; дерево строится до полного исчерпания подвыборки и не подвергается процедуре прунинга (англ. pruning – отсечение ветвей).

Таким образом, на рисунке 4 представлено дерево решения, выбранное первым алгоритмом, который использовался в исследовании для установки факторов рисков развития рецидива, обладающих конечным вариантом развития течения заболевания у больных МЛУ ТБ.

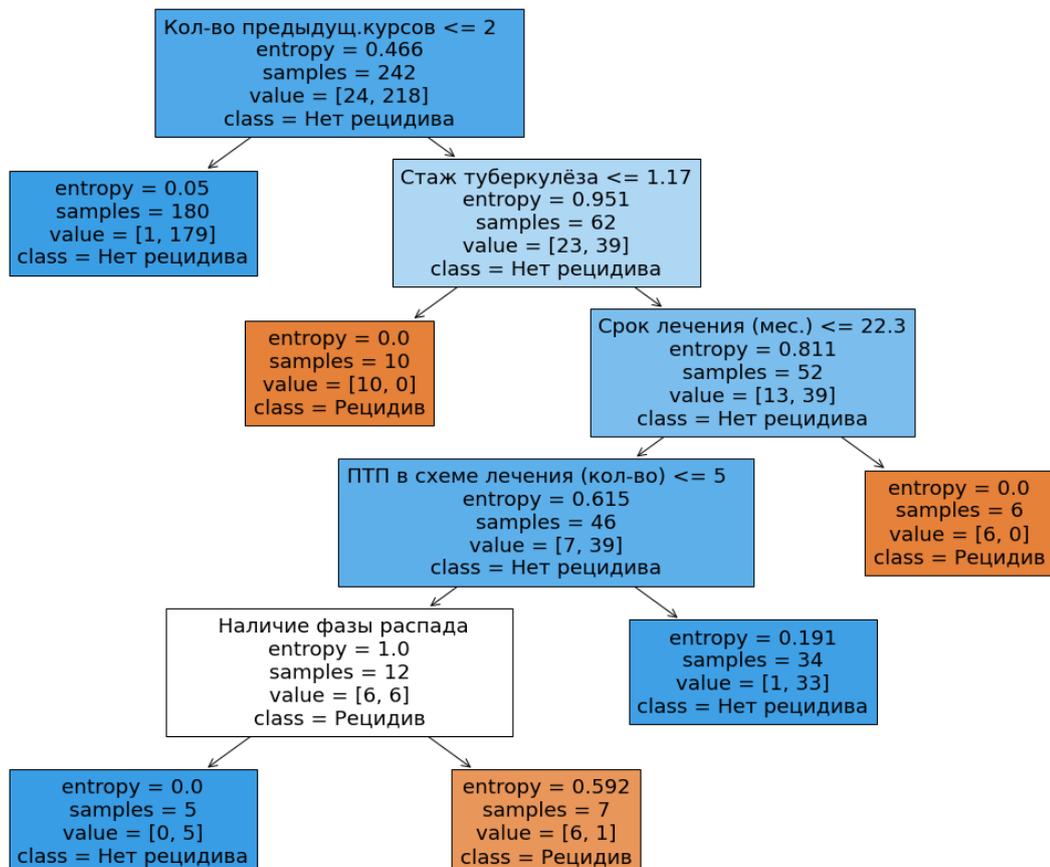


Рисунок 4 – Дерево решений прогнозирования рецидива МЛУ ТБ

Дерево иллюстрирует ситуацию, когда имеют место наличие у пациентов несколько последовательных признака, сопряженных с рецидивом МЛУ ТБ, причем последующие решения включения признака в прогнозную модель основываются на результатах предыдущих. Дерево решений, представленное на рисунке 4 в виде графа, вершина – частные признаки, а ветви дерева – различные характеристики пациента, которые могут иметь место в определенном частном состоянии. Каждой ветви присваиваются различные числовые характеристики признака, с указанием степени неоднородности распределения признака в ветви дерева (entropy), количестве примеров, принадлежащих предсказываемым классам (количестве пациентов с рецидивом или без рецидива в ветви дерева (value) и непосредственно прогнозируемое состояние (class). Причем числовое значение entropy указывает на оценку важности признаков.

Представленная модель прогноза рецидива у больных МЛУ ТБ интерпретируется следующим образом. Лица, имеющие в анамнезе туберкулез чуть более года (случаи впервые выявленного МЛУ ТБ), а также случаи повторного лечения, но не более 2-х эффективных курсов химиотерапии туберкулеза, не развивают рецидив МЛУ ТБ как минимум в течение 5 лет наблюдения. Однако если у пациентов наблюдались неоднократные курсы химиотерапии (два и более), диагностируется деструктивный туберкулез, общий срок лечения не достигает 22,3 мес. и в схеме химиотерапии присутствуют менее 5 ПТП, рецидив заболевания более вероятен.

Необходимо отметить, что, как правило, в случае алгоритмов, использующих деревья, для построения классифицирующей модели, включаются не все признаки, имеющиеся в тестируемом наборе, а только те, которые получаются в правилах расщепления признакового пространства для получения однородных областей. Таким образом, существует значительная часть признаков, не попадающих в правила, формирующие деревья, и их информативность по критерию прироста информации в модели, равна нулю. В представленном исследовании определены 20 неинформативных признаков. Их анализ показал, что в основном они связаны с лекарственной устойчивостью МБТ к тому или

иному ПТП при впервые установленной МЛУ возбудителя (канамицин, амикацин, капреомицин, этамбутол, пипразинамид, ПАСК), а значит применяемой первоначальной схемой химиотерапии по IV-му режиму. Кроме того, классификатор не выявил глубинной связи между рецидивом МЛУ ТБ и коморбидностью у пациентов с язвенной болезнью желудка и двенадцатиперстной кишки, сахарным диабетом и сердечно-сосудистыми заболеваниями, а также с зависимостями в употреблении психоактивных веществ и табакокурении.

Путем исключения данных признаков из исходного набора получена информативная база, которая использовалась повторно в исследовании при настройке модели логистической регрессии при тех же параметрах, определенных сеточным поиском. В результате значение чувствительности в данном алгоритме увеличилось до 0,73, а значение специфичности осталось прежним.

На втором этапе для демонстрации надёжности выявленных прогностических моделей была использована стратифицированная K-блочная перекрёстная проверка для каждого метода МО, включенного в исследование. Всего исследуемый набор данных был поделен на 5 блоков, использующихся при построении набора классификаторов. Для моделей использующих деревья решений, снова был использован полный набор признаков, а для логистической регрессии – набор информативных критериев, выбранный ранее в первом эксперименте.

Полученные оценки информативности признаков, специфичности, чувствительности и площади под ROC кривой усреднялись по всем 5 выборкам (Приложения А, Б, В и Г).

На рисунках 5, 6, 7 и 8 показаны гистограммы распределения значимости признаков, усредненной по 5 блокам перекрестной проверки, для каждого из выбранных в исследовании алгоритмов МО.

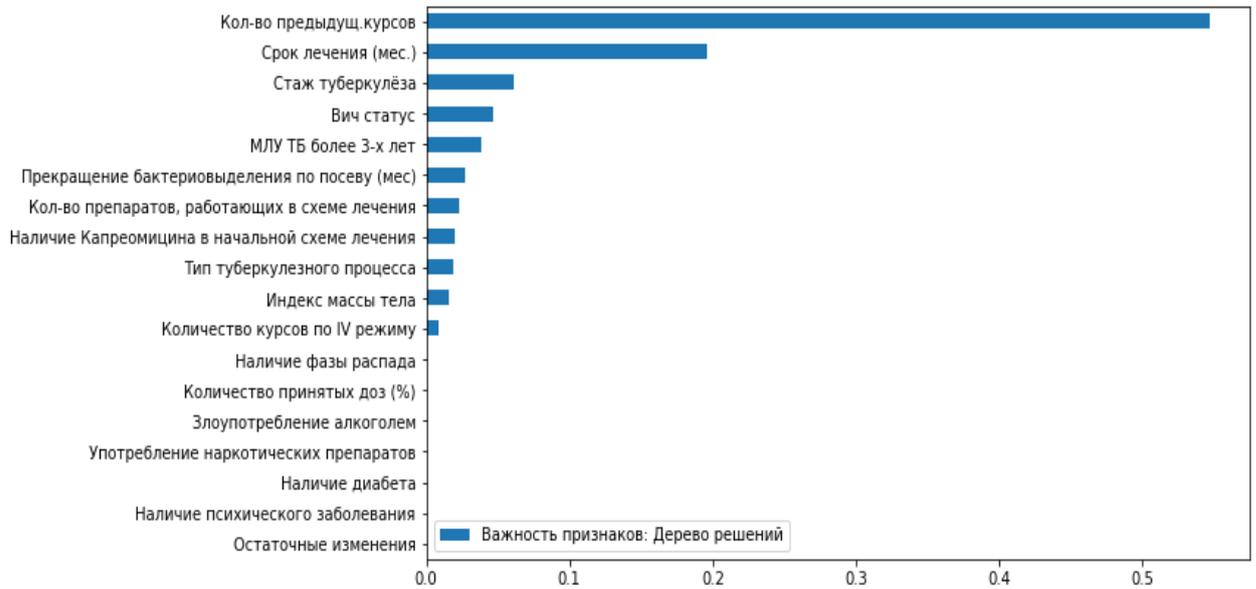


Рисунок 5 – Усредненная значимость признаков алгоритма классификации пациентов, построенного методом дерево решений

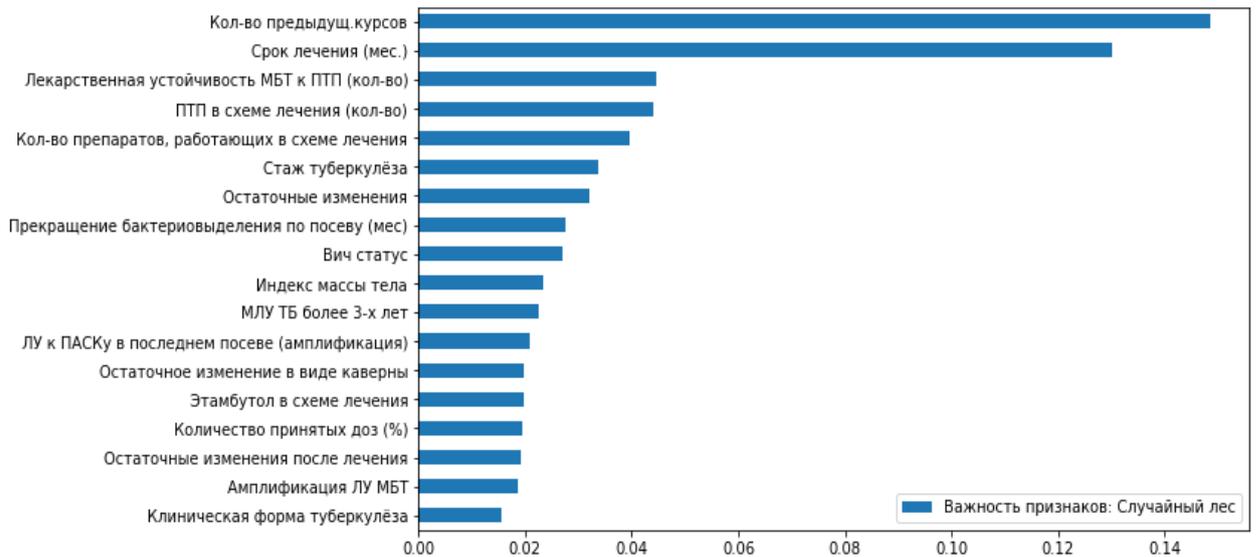


Рисунок 6 – Усредненная значимость признаков при использовании прогностической модели классификации рецидива больных МЛУ ТБ, полученной методом случайный лес

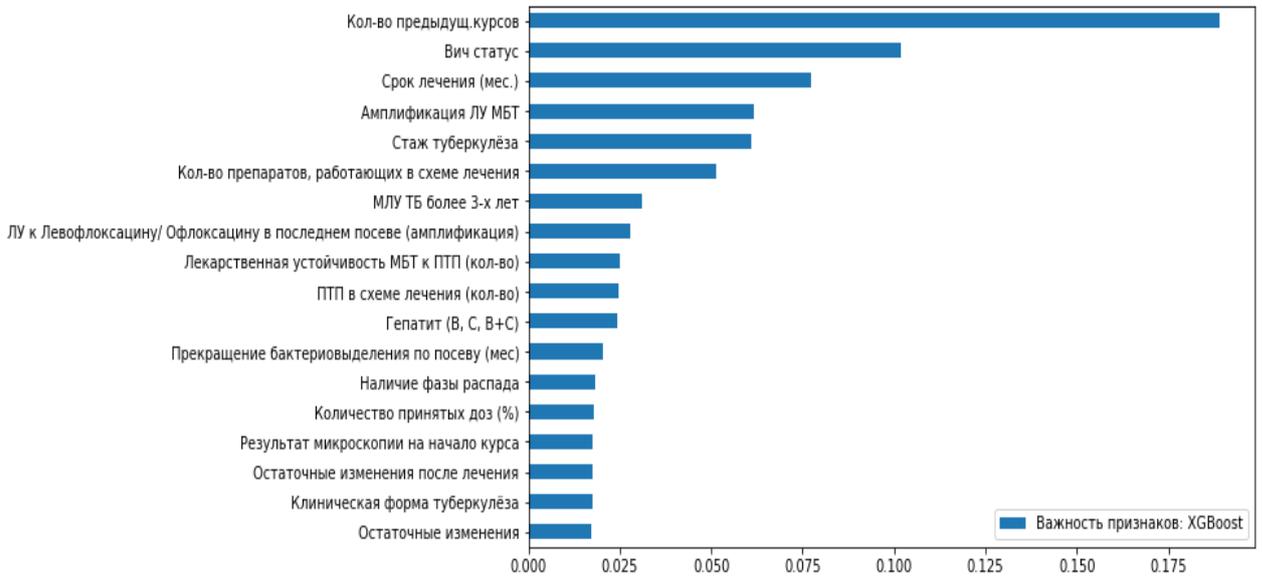


Рисунок 7 – Усредненная значимость признаков при использовании прогностической модели классификации рецидива больных МЛУ ТБ, построенной методом XGBoost

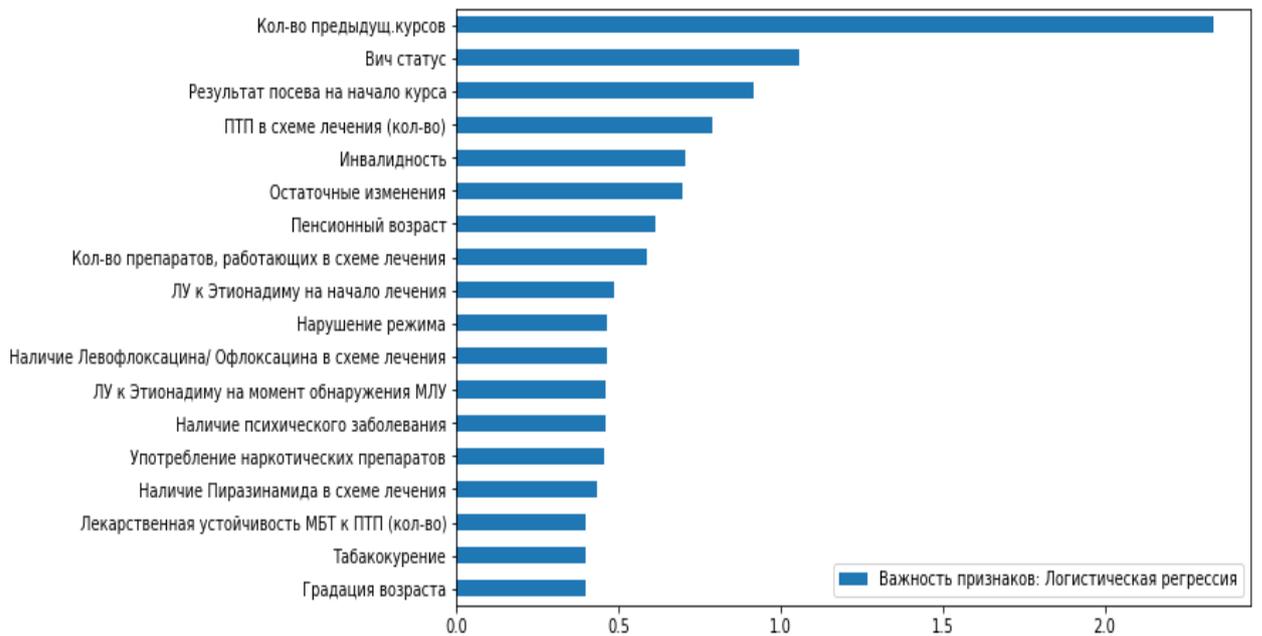


Рисунок 8 – Усредненная значимость признаков при использовании прогностической модели классификации рецидива больных МЛУ ТБ, полученного методом логистической регрессии

Всего на рисунках 5, 6, 7 и 8 представлено десять важных признаков, определенных классификаторами, позволяющих выделить их в ведущие и второстепенные факторы риска развития рецидива у успешно пролеченных больных МЛУ ТБ.

Лидирующим признаком, как уже было отмечено выше, который был определен всеми алгоритмами МО, включенными в исследование, был фактор, указывающий на наличие у пациентов повторных курсов химиотерапии, сопряженных с достаточно продолжительным стажем заболевания, причем с МЛУ возбудителя более трех лет. На это указывают три модели классификаторов признаков, за исключением логистической регрессии. В лидирующей пятерке признаков, определенных этими же тремя методами МО, включены продолжительность лечения МЛУ ТБ и наличие у пациентов ВИЧ инфекции. Из коморбидностей, сопряженных с рецидивом МЛУ ТБ, включены больные с наличием вирусных гепатитов В, С, В + С. Данный фактор определен только алгоритмом XGBoost.

К менее важным из десяти признаков, но введенных в модель прогноза большинством методов, относятся признаки, влияющие на эффективность лечения – количество работающих ПТП в схеме лечения, сроки абациллирования, амплификация МБТ к левофлоксацину на фоне лечения. Однократно в модели случайный лес определен вклад наличия в легочной ткани остаточных изменений после ранее перенесенного туберкулеза и деревом решений – индекс массы тела. В улучшенной модели логистической регрессии «высвечиваются» и социальные факторы риска в виде наличия у пациентов инвалидности по туберкулезу и пенсионный возраст.

Самым наглядным методом, позволяющим оценить качество построенных моделей прогноза, является построение ROC-кривой на основании рассчитанных показателей специфичности и чувствительности. На рисунке 9 приведены ROC-кривые для каждой из разработанных моделей (см. Приложения А, Б, В и Г).

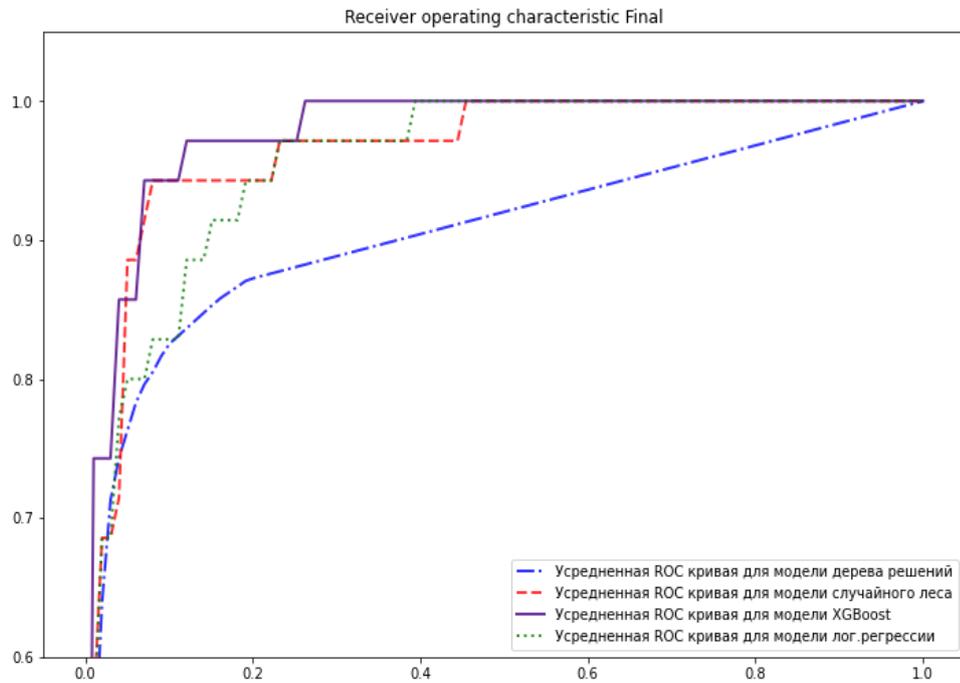


Рисунок 9 – ROC кривые для моделей классификаторов, определяющих признаки, сопряженные с рецидивом МЛУ ТБ органов дыхания

Данный комплексный график показывает зависимости количества верно классифицированных истинно положительных признаков от количества классифицированных ложноотрицательных в используемых методиках. После построения графика вычисляется площадь под кривой (AUC). Чем ближе значение этого параметра к единице (идеальный тест), тем идеальнее качество модели.

Числовые трактовки показателя AUC, представленные в таблице 38, свидетельствуют о высоком качестве прогнозирующих моделей, следовательно, методы МО могут быть использованы в практике для оценки вероятности развития рецидива у излеченных пациентов с МЛУ ТБ.

Таблица 38 – Метрики качества работы классификаторов, построенных с помощью К-блочной стратифицированной проверки в прогнозировании рецидива МЛУ ТБ

Алгоритм	Чувствительность, отклонение	Специфичность, отклонение	AUC, отклонение
Дерево решений	0,74 ± 0,167	0,97 ± 0,03	0,9 ± 0,1
Случайный лес	0,91 ± 0,17	0,98 ± 0,02	0,96 ± 0,03
XGBoost	0,91 ± 0,14	0,98 ± 0,02	0,98 ± 0,02
Логистическая регрессия	0,91 ± 0,16	0,98 ± 0,02	0,95 ± 0,04

Наиболее оптимальным алгоритмом МО из представленных в таблице 38, является метод случайный лес. На его основе построена модель классификации и проведено ее тестирование на описанных ранее клинических случаях двух пациентов, данные которых расположились в исследовательской базе в строках 11 и 56.

Был произведен импорт данных пациентов (набор признаков пациентов в количестве 97) с последующей интеграцией в формате MS Excel в интерактивную облачную среду с программным кодом Google Colaboratory (Приложения Д и Е).

Для оценки факторов риска далее происходил выбор пациента в облачной среде путем введения следующих команд и настройки параметров:

Ввод данных и команд представлен на примере пациента № 11:

```
[In]
dataRow = testDfX.loc[11]
testPatient = pd.DataFrame(columns = testDfX.columns)
testPatient = testPatient.append(dataRow, ignore_index=True,
```

где «In» – ячейка в которую вводится код команды пользователем;
«Out» – ячейка для отображения результата работы программы.

Затем была произведена установка модуля, анализирующего признаки пациента, улучшающий модель классификации с отображением диаграммы вклада признаков (красные фрагменты диаграммы голосуют «за» прогноз развития рецидива, синие – «против») (Рисунок 10):

```
[In]
explainerForest = dx.Explainer(gridForest.best_estimator_, dfX, y,
    label = 'Влияние признаков на оценку рисков')
bd_tb = explainerForest.predict_parts(testPatient, type = 'shap_wrapper')
bd_tb.plot()
```

[Out]

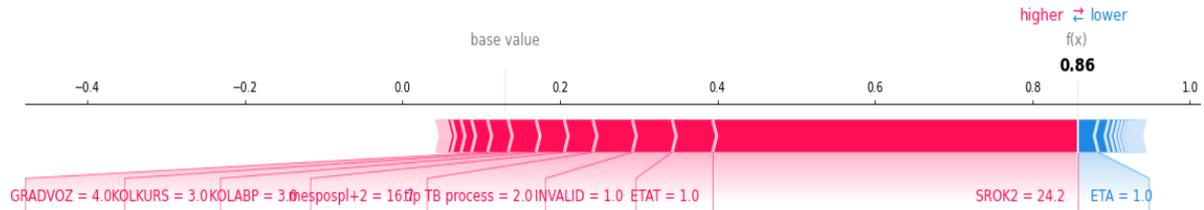


Рисунок 10 – Диаграмма предикторов и вероятность возникновения рецидива больного № 11

На основании обработки признаков получен результат модели, представленный на рисунке 10, который интерпретируется следующим образом: вероятность возникновения рецидива МЛУ ТБ у данного пациента составляет 86 %, базовое значение функции – $f(x) = 0,86$.

Для более детального рассмотрения важности каждого фактора риска в прогностической модели была выполнена команда системы для отображения вклада каждого признака с помощью диаграммы конкурирующих предикторов:

```
[In]
bd_tb = explainerForest.predict_parts(testPatient, type = 'break_down_interactions')
bd_tb.plot()
```

Далее произведен вывод диаграммы на экран компьютера с указанием веса каждого предиктора, где зеленым цветом указаны факторы, вес которых обладает положительным вкладом в наступлении рецидива МЛУ ТБ (Рисунок 11).

[Out]

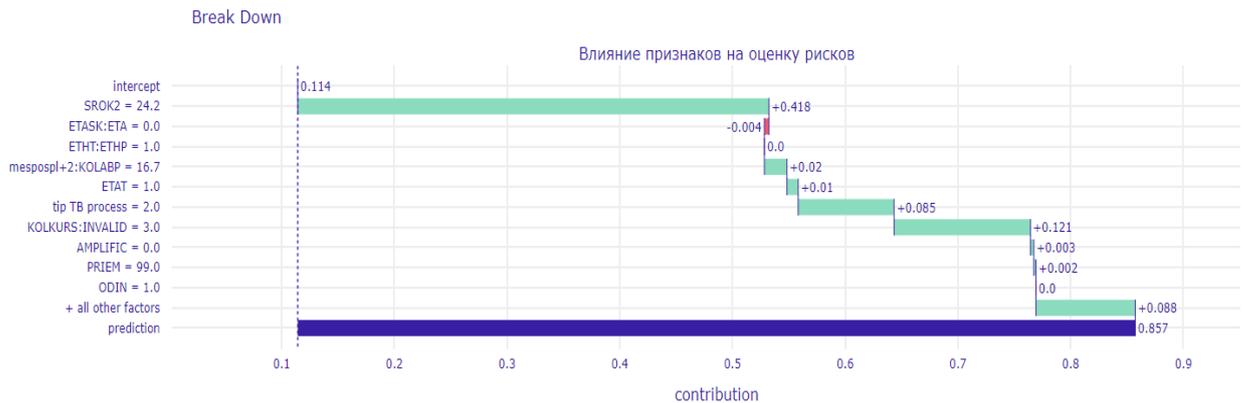


Рисунок 11 – Диаграмма конкурирующих предикторов возникновения рецидива больного № 11

Анализируя информацию на рисунке 11, необходимо заключить, что у пациента № 11 вероятность развития рецидива МЛУ ТБ составляет 86 % (параметр «prediction» обозначен синим цветом). На основании коэффициентов предикции на развитие повторного заболевания у данного пациента влияют следующие факторы: средний срок лечения (0,418), тип туберкулезного процесса (0,085), количество курсов химиотерапии и наличие инвалидности (совместный вклад) (0,121), прочие факторы (0,088 суммарно).

Таким образом, построенная методом случайный лес прогностическая модель классифицирует пациента № 11 верно.

По аналогии были проведены этапы работы классификатора по пациенту № 56.

```
[In]
dataRow = testDfX.loc[56]
testPatient = pd.DataFrame(columns = testDfX.columns)
testPatient = testPatient.append(dataRow, ignore_index=True)
```

```
[In]
explainerForest = dx.Explainer(gridForest.best_estimator_, dfX, y,
    label = 'Влияние признаков на оценку рисков')
bd_tb = explainerForest.predict_parts(testPatient, type = 'shap_wrapper')
bd_tb.plot()
```

Получен прогноз развития рецидива у пациента № 56, который на основе базового значения функции составляет 8 % ($f(x) = 0,08$) (Рисунок 12).

[Out]



Рисунок 12 – Диаграмма предикторов и вероятность возникновения рецидива больного № 56

Вклад каждого предиктора в прогностическую модель представлен ниже.

[In]

```
bd_tb = explainerForest.predict_parts(testPatient, type = 'break_down_interactions')
bd_tb.plot()
```

[Out]

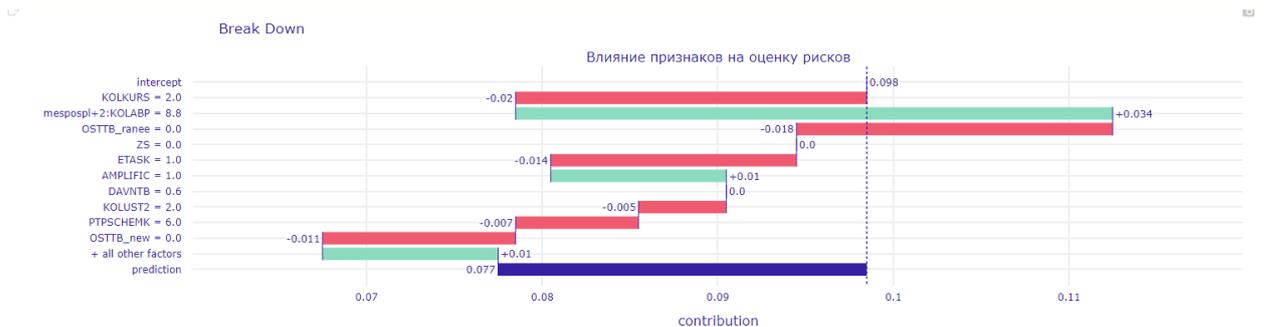


Рисунок 13 – Диаграмма конкурирующих предикторов возникновения рецидива больного № 56

Анализируя информацию, представленную на рисунке 13, необходимо заключить, что у пациента № 56 вероятность развития рецидива МЛУ ТБ составляет 8 %.

На диаграмме рисунка 13 указаны веса каждого предиктора, где зеленым цветом показаны факторы, вес которых обладает положительным вкладом в наступлении рецидива МЛУ ТБ. Необходимо отметить, что в данном случае наблюдается большой разброс по значениям влияния признаков на развитие рецидива.

Так, основными предикторами в пользу безрецидивного течения заболевания являются признаки с отрицательными значениями: количество курсов предыдущего лечения ($-0,02$), число ПТП в схеме лечения ($-0,007$), отсутствие остаточных изменений в легких ($-0,011$). При этом регистрировались признаки с положительными значениями: срок абациллирования и число препаратов, к которым произошла амплификация ($0,034$), факт амплификации во время лечения ($0,01$), прочие признаки ($0,01$ суммарно).

Таким образом, оба пациента были классифицированы верно.

В заключение необходимо отметить, что, несмотря на хорошее качество модели регрессионного анализа традиционной статистики и на основании параметров, характеризующих качество моделей (Se , Sp , AUC , таблицы 35 и 38), а также по причине использования полного объема данных для предсказания рецидива заболевания, модель случайный лес, равно, как и остальные методы МО, применяемые в исследовании, обладает наилучшей прогностической способностью.

На основании результатов диссертационного исследования создана программа и получено свидетельство о государственной регистрации программы для ЭВМ № 2020619181 «Программа-калькулятор для оценки риска развития рецидива у больных с множественно лекарственно устойчивым туберкулезом» (Приложение Ж).

ЗАКЛЮЧЕНИЕ

Ежегодно более 10 миллионов человек заболевают туберкулезом, который является ведущей причиной смерти среди инфекционных болезней. Установлено, что пациенты, которые ранее болели туберкулезом, имеют значительно повышенный риск развития рецидивов заболевания. Частота рецидивов в течение первого года после завершения лечения туберкулезом составляет 2–10 %, что в несколько раз превышает уровень заболеваемости среди тех же групп населения, не болевших туберкулезом. Таким образом, рецидивирующее течение туберкулеза вносит существенный вклад в заболеваемость, особенно в странах, где он эндемичный. Еще одним сложным фактором во фтизиатрии является туберкулез с МЛУ возбудителя. По данным глобальных отчетов ВОЗ за 2019 г. Российская Федерация занимает третье место по общенациональному бремени МЛУ ТБ после Индии и Китая. Учитывая высокое бремя и низкий уровень излечения от МЛУ ТБ (около 50 %), адекватность ведения больных туберкулезом органов дыхания с МЛУ возбудителя требует всестороннего изучения. Влияние МЛУ МБТ на рецидивы заболевания, равно как и развитие рецидивов с МЛУ ТБ после эффективной химиотерапии в РФ исследовались мало.

Частота рецидивов туберкулеза органов дыхания является одним из показателей, характеризующих эффективность противотуберкулезных мероприятий. При рецидивах туберкулеза высок уровень тяжелых и распространенных клинических форм, гораздо реже наступает полное излечение больных от туберкулеза без остаточных изменений, чаще – инвалидность из-за развивающихся морфологических и функциональных расстройств. Лечение пациентов с рецидивами оказывается более длительным и дорогостоящим, и зачастую менее эффективным. Среди них более высока смертность от туберкулеза и его последствий. Выявление рецидивов туберкулеза зачастую бывает несвоевременным, что требует разработки новых научно-обоснованных подходов в решении данной проблемы.

В условиях наметившегося улучшения эпидемиологической обстановки по

туберкулезу ведущую роль в профилактике ранних и поздних рецидивов должно играть их прогнозирование путем оценки влияния различных факторов риска и их сочетаний на лиц, перенесших туберкулез, и разработки персонифицированных мер профилактики в отношении тех, кто имеет высокую вероятность развития рецидивов. Значение прогноза в клинической практике определяется основными задачами современной медицины – предупреждение заболеваний и лечение больных. Важнейшее значение для получения достоверного прогноза имеют научные знания об особенностях течения и исхода того или иного заболевания, которые позволяют предвидеть, как может или должно измениться состояние пациента в будущем. Знания этиологии, патогенеза конкретного заболевания позволяют сделать более достоверный и убедительный прогноз.

Современные исследования показывают наличие нескольких направлений создания прогностического моделирования в области фтизиатрии. Во-первых, математические расчёты, прогнозирующие развитие эпидемической обстановки, которые основываются на статистическом анализе различных эпидемиологических показателей туберкулеза. Во-вторых, это оценка факторов риска развития заболевания (прямые – МБТ и косвенные предрасполагающие), а также прогноза эффективности применяемых методов терапии. Третьим направлением является создание математических моделей, оценивающих прогноз выживаемости пациентов на основе анализа клинических данных и результатов лабораторно-инструментальных исследований.

В немногочисленных отечественных публикациях чаще всего описывался один из методов математического моделирования, применяемый во фтизиатрии – множественной логистической регрессии. Он использовался в исследованиях для прогноза степени риска развития рецидива после завершения лечения больных туберкулезом, инвалидности, побочных эффектов противотуберкулезной химиотерапии, в оценке факторов риска, влияющих на формирование основных эпидемиологических показателей по туберкулезу. Ранговым дисперсионным анализом определялся набор предикторов неблагоприятного исхода впервые выявленного туберкулеза легких. Совершенно недавно появились работы по

прогнозированию рецидивов туберкулеза с применением последовательного анализа Вальда и байесовского метода [34], но первое научное исследование описывало лишь использование традиционного математического аппарата, второе исследование было посвящено иерархической структуре математической модели исхода коинфекции среди лиц ВИЧ/туберкулёз [46].

Прогностические возможности значительно расширяются с применением алгоритмов машинного обучения, отличающихся высокой точностью и позволяющих анализировать большие объемы данных пациентов. Главное преимущество этих методов заключается в получении обоснованного диагноза или прогноза заболевания на основе оценки значимости признаков. Данные методы используются с высокой эффективностью в прогнозировании ряда заболеваний или исходов, таких как рак легких, заболеваний сердечно-сосудистой системы, инсультов и других, включая туберкулез [62, 66, 72, 81, 83, 94, 101, 115, 121, 137, 138, 139, 171]. Стоит отметить, что использование нескольких методик математического анализа при изучении многомерных данных на проявление закономерностей в них, как способа прогноза рецидивов, до настоящего времени не велись и не изучены.

Для прогноза рецидивов туберкулеза органов дыхания методы МО практически не использовались. Применение данных методов может оказаться перспективным в прогнозировании рецидивов туберкулеза органов дыхания и не только, и в доступной литературе не имеет аналогов. Это определяет актуальность диссертационного исследования.

Выявить приоритетные факторы риска рецидива у эффективно излеченных больных МЛУ ТБ органов дыхания и разработать с помощью алгоритмов машинного обучения методику прогнозирования – основная практическая цель диссертационного исследования.

В задачи исследования входили изучение частоты возникновения рецидива после успешного лечения больных МЛУ ТБ, проживающих на территории Томской области в период с 2009–2011 гг., в генеральной совокупности со сроком наблюдения минимум 5 лет, определить ведущие медицинские и немедицинские

факторы риска развития рецидива, характеристики и возможности применения методов машинного обучения с целью прогнозирования повторного МЛУ ТБ. Заключительной задачей явилась разработка программы ЭВМ для оценки индивидуального риска развития рецидива у больных МЛУ ТБ.

При определении показателя рецидива органов дыхания в генеральной совокупности пациентов МЛУ ТБ использовались отчеты за 2009–2011 гг. по Томской области (формы № 8 «Сведения о заболеваниях активным туберкулезом» и № 33 «Сведения о больных туберкулезом (годовая)» по Томской области за указанный период времени). Для оценки влияния значимых факторов риска развития рецидива были изучены сведения о 346 взрослых больных МЛУ ТБ с эффективно завершенным курсом химиотерапии по IV режиму. Из них 35 были с ранними и поздними рецидивами туберкулеза органов дыхания и 311 излеченных лиц, в том числе снятых с диспансерного наблюдения в связи с выздоровлением, не имевших рецидивов в течение пяти лет наблюдения. В ретроспективный анализ вошли данные медицинской документации лабораторной диагностики, лечения и диспансерного наблюдения пациентов, а также сведения базы учета контингентов организационно-методического отдела Томского фтизиопульмонологического медицинского центра. Все больные, включенные в исследование, были обследованы и пролечены в соответствии с действующими на тот момент Федеральными клиническими рекомендациями. Исследовательская база формировалась в Microsoft Office Excel 2016 с обезличенными данными пациентов со смешанными признаками – количественными, качественными и порядковыми. На первом этапе обработка и статистический анализ (процедуры описательной статистики, множественный регрессионный анализ) полученных результатов проводили с использованием программного обеспечения IBM SPSS Statistics версия 23.

Анализировали следующие переменные: место проживания, возраст, пол, уровень образования, статус занятости, социальные факторы риска (включая курение, алкоголизм, пребывание в местах заключения и т. д.), наличие туберкулезных контактов, индекс массы тела, локализация процесса в легких,

клиническая форма туберкулеза органов дыхания, наличие и размер полости, поражение плевры, результаты посева, результаты микроскопии, информация о спектре лекарственной устойчивости МБТ, схемы химиотерапии, изменения в лечении туберкулеза, а также данные о сопутствующей патологии, о характере остаточных изменений в легких после предыдущих курсов лечения.

Проведенный анализ определил, что частота возникновения рецидива после успешного лечения больных МЛУ ТБ за 5 последующих лет достигает 10,1 %; рецидивы в большинстве случаев (77,1 %) возникают в течение первых 3-х лет наблюдения; средний срок наступления рецидива после эффективного курса химиотерапии составил 1,9 года. Сравнивая данный показатель с имеющимися публикациями в РФ необходимо констатировать факт, что данная клиническая проблема фтизиатрии в нашей стране недооценена. Во-первых, отсутствует нормативная отчетность по данным пациента, позволяющая анализировать и принимать клинические решения по профилактике у них рецидива. Во-вторых, высокие показатели рецидивов МЛУ ТБ в течение 2-х лет наблюдения определяют создание новых программ по активному выявлению случаев заболевания среди эффективно пролеченных пациентов, у которых существуют факторы риска.

Статистический анализ данных пациентов двух групп наблюдения выявил некоторые особенности. Сравнительно характеризуя пациентов двух групп наблюдения МЛУ ТБ, было установлено, что в обеих группах преобладали мужчины (74,3 % и 72,0 % соответственно), средний возраст в группе рецидивов составил 42 (32; 50) и 38 (30; 50) лет во второй, без выраженного дефицита массы тела (индекс массы тела – 20,9). В равных частях это были жители г. Томска и Томской области, пациентов без определенного места жительства было 6 (17,1 %) и 26 (8,4 %) соответственно ($p = 0,164$). Доля больных в первой группе наблюдения, пребывающих ранее в местах лишения свободы была существенной (40,0 %), но без значимых различий в группе сравнения (25,7 %, $p = 0,072$), равно как и наличие туберкулезного контакта (25,7 % и 16,7 % соответственно, $p = 0,276$).

Социальный статус больных МЛУ ТБ не повлиял на наступление в будущем рецидива. Безработных в группе развивших рецидив заболевания было 27 человек (77,1 %), в группе без рецидива – 235 (75,6 %). Обращает на себя внимание, что каждый третий пациент успешно излечившийся от МЛУ ТБ и развивший рецидив заболевания имел инвалидность по соматическим заболеваниям – 31,4 и в группе сравнения 8,7 % ($\chi^2 = 16,65$, ОШ = 4,82; 95 % ДИ 2,13–10,90, $p < 0,001$), каждый пятый был пенсионного возраста – 20,0 и в группе сравнения 8,0 % ($\chi^2 = 4,03$, ОШ = 2,86; 95 % ДИ 1,14–7,20, $p = 0,045$).

Значимыми факторами формирования рецидива явились: присутствие у больных ВИЧ инфекции (ОШ = 19,19; 95 % ДИ 5,29–69,56); психических заболеваний ($\chi^2 = 13,71$, ОШ = 5,85; 95 % ДИ 2,27–15,03, $p < 0,001$) и табакокурение (ОШ = 3,16; 95 % ДИ 1,08–9,20). Примечательно, что наличие коморбидных заболеваний, таких как хроническое неспецифическое заболевание легких (ХОБЛ, бронхиальная астма, эмфизема), сахарного диабета, вирусных гепатитов В, С, В + С, заболеваний желудочно-кишечного тракта, а также зависимостей (алкогольной, опиоидной) не повлияло на возникновение в будущем рецидива после успешной химиотерапии больных МЛУ ТБ.

Закономерно в группе рецидива преобладали пациенты с повторными курсами лечения туберкулеза – 31 (88,6 %), в нее вошли пациенты как ранее леченные, так и с хроническим течением заболевания, тогда как в группе сравнения, наоборот, в 61,1 % были впервые выявленные случаи.

Распределение больных по клиническим формам туберкулеза легких было одинаковым, в целом у половины больных в группах наблюдения отмечалась инфильтративная форма туберкулеза легких (42,9 и 55,6 % соответственно). Диссеминированный туберкулез наблюдался в 31,4 и 22,5 % случаях, доли больных с фиброзно-кавернозным туберкулезом составили 5,7 % (1 группа) и 4,2 % (2 группа). Очаговый туберкулез и туберкулемы в обеих группах диагностировались на уровне 6,1–8,6 %.

Деструкция легочной ткани, выявляемая рентгенологически, чаще встречалась в группе рецидива (в 94,3 и 68,8 %, ОШ = 7,48; 95 % ДИ 1,76–31,80).

Также, как и выявление микроскопически КУМ в мокроте чаще встречалось в группе больных, развивших рецидив заболевания – у 27 (77,1 %) и 167 (53,7 %), ($\chi^2 = 6,10$, ОШ = 2,91; 95 % ДИ 1,28–6,61, $p = 0,014$).

Влияние роли остаточных изменений после предыдущих курсов химиотерапии на развитие рецидива подтвердилось и в нашем исследовании: у всех пациентов из первой группы были зафиксированы остаточные изменения, в группе сравнения они наблюдались в 85,9 % случаев ($p = 0,013$).

В исследовании выявлена тенденция влияния приверженности лечению больных на формирование рецидива, частота самостоятельных отрывов от химиотерапии в первой группе больных была почти в три раза выше, но различий не обнаружено (20,0 % и 8,7 % соответственно, $p = 0,067$).

Косвенно на данную связь указывает высокая частота возникающей амплификации возбудителя туберкулеза, которая определяется в последнем положительном анализе посева мокроты у больных, развивающих рецидив. При этом до начала лечения у всех больных выявлялась достаточно высокая частота ЛУ МБТ к следующим препаратам: к этионамиду/протионамиду в 40,0 % и 37,8 % случаях, к канамицину/амикацину в 25,7 % и 13,6 %, к капреомицину в 25,7 % и 14,0 % соответственно. ЛУ МБТ к фторхинолонам в группах наблюдения не достигала 20 %, к ПАСК около 5 %, к циклосерину не превышала 1 %. В последних положительных бактериограммах мокроты спектр ЛУ к ПТП увеличивался по шести ПТП ($p < 0,01$). Амплификация возбудителя ТБ наблюдалась к: канамицину/амикацину (25,7 % и 54,3 % соответственно, (ОШ = 5,91; 95 % ДИ 2,85–12,26); капреомицину (25,7 % и 51,4 %, (ОШ = 5,27; 95 % ДИ 2,55–10,91), этамбутолу (25,7 % и 68,6 %, (ОШ = 0,12; 95 % ДИ 0,06–0,26), этионамиду/протионамиду (40,0 % и 74,3 %, (ОШ = 4,30; 95 % ДИ 1,95–9,48), левофлоксацину (17,1 % до 54,3 %, (ОШ = 7,02; 95 % ДИ 3,36–14,66). Частота ЛУ МБТ к ПАСК увеличивалась в процессе химиотерапии в восемь раз (5,7 % и 42,9 % (ОШ = 8,97; 95 % ДИ 4,08–19,73). В группе сравнения амплификация ЛУ МБТ по всем вышеуказанным препаратам на фоне лечения, присутствовала, но в динамике существенно не нарастала,

увеличение частот ЛУ МБТ к ПТП было в пределах 3–4 %.

Вероятнее всего этим и объясняются разные сроки прекращения бактериовыделения, определяемые как при микроскопии мазка мокроты, так и посеву. Так, через два месяца лечения в группе сравнения (без рецидива) прекратили выделять МБТ большая часть больных 239 (76,8 %), в первой группе абациллированных пациентов было в два раза меньше (14 (40,0 %), $\chi^2 = 21,73$, ОШ = 4,98; 95 % ДИ 2,41–10,29, $p < 0,001$).

Согласно задачам исследования первоначально в целях прогнозирования риска развития рецидивов МЛУ ТБ был применен регрессионный анализ значимых признаков, которые были определены в однофакторном анализе базы данных пациентов, с расчетом показателей Хосмера – Лемешова, Вальда. Для оценки качества модели использовались: качество подгонки модели (χ^2), мера Нэйджелкерка, значения коэффициента конкордации, чувствительности и специфичности, ROC-анализ.

В результате выделены 6 факторов риска, имеющих наибольший вклад в формирование рецидива у пациентов с МЛУ ТБ. На основании расчета коэффициентов регрессии (b), используемых в уравнении для предикторов, а также экспонированного значения шансов (Exp (b), который показывает во сколько раз изменяется шанс возникновения изучаемого события, если значения одного из предикторов изменятся на единицу при фиксированных значениях прочих предикторов, факторы риска определены следующим образом: наличие у пациента ВИЧ-инфекции (b = 5,370, Exp (b) = 214775, $p = 0,003$); несколько курсов химиотерапии (b = 3,172, Exp (b) = 23,857, $p < 0,001$), деструкции в легочной ткани на начало лечения (b = 2,882, Exp (b) = 17,854, $p = 0,065$), пенсионный возраст (b = 2,257, Exp (b) = 9,550, $p = 0,013$), нарастание спектра лекарственной устойчивости МБТ к ПТП в период лечения (амплификация) (b = 2,057, Exp (b) = 7,824, $p = 0,026$), срок лечения (b = 1,528, Exp (b) = 4,607, $p = 0,001$).

Как уже было описано выше, в последние годы активно развиваются особые методы анализа данных, так называемый Data mining (добыча данных,

интеллектуальный анализ данных), использующий весь потенциал современных методов классификации, моделирования и прогнозирования для выявления в изучаемом наборе данных новых сведений, которые не могут быть получены путем простого статистического анализа. Другими словами, методы Data mining и алгоритмы машинного обучения являются средством извлечения из набора данных ранее неизвестных скрытых взаимосвязей между переменными, характеризующими объект, предоставляя исследователю новые, в практическом смысле полезные и доступные для интерпретации знания, необходимые для принятия решений в различных сферах человеческой деятельности, в том числе и в медицинской диагностике.

Фундаментальной основой в прогностической аналитике является математическое моделирование, в которое входит создание моделей с учетом набора правил, а также формул и уравнений, полученных из исходных данных, позволяющих в дальнейшем генерировать предсказания. Одним из наиболее распространенных методов построения прогнозной модели является метод классификации, который используется в системах поддержки принятия решения и клинического прогнозирования в оценке вероятности или возникновения события. Во многих областях медицины используются различные методики моделирования и алгоритмы машинного обучения.

Для построения классифицирующих моделей в данном исследовании были выбраны следующие методы МО: дерево решений, случайный лес, XGBoost (градиентный бустинг) и логистическая регрессия. Данные методы показали высокую информативность в медицине.

Таким образом, в проведенном исследовании с помощью математических методов, с применением алгоритмов МО определена прогностическая значимость наиболее информативных возрастно-половых, эпидемиологических, социальных, медико-биологических факторов риска и их сочетаний в развитии рецидивов МЛУ ТБ. Результаты исследования доказали, что инструменты искусственного интеллекта могут применяться к получению набора значимых признаков и прогностических моделей, характеризующих эффективность химиотерапии

больных МЛУ ТБ в контексте безрецидивного течения заболевания в течение 5 лет наблюдения.

С помощью математического моделирования, с применением классификаторов дерево решений, случайный лес, XGBoost (градиентный бустинг), логистической регрессии, настоящее исследование выделило предикторы развития рецидива у успешно пролеченных больных МЛУ ТБ, которые включают в себя пять основных признаков глубинно связанных с факторами риска, определенных при описательной статистике: наличие у пациентов повторных курсов химиотерапии, сопряженных со стажем заболевания, присутствием в легочной ткани полости распада более 2,0 см в диаметре, химиотерапия менее 5 ПТП со сроком лечения менее 22 месяцев. Алгоритмы позволили с высокой точностью и чувствительностью определить риск развития рецидива МЛУ ТБ и дополнительно продемонстрировали сложную взаимосвязь развития рецидива МЛУ ТБ между особенностями возбудителя туберкулеза, способного мутировать в неоднократных курсах химиотерапии, отягощая лекарственную устойчивость, клинической характеристикой пациента (возраст, наличие сопутствующей патологии) и возможностями применяемого лечения.

На основе полученных данных создана методика прогнозирования индивидуального риска рецидивов у лиц, перенесших МЛУ ТБ и успешно вылеченных (см. Приложение Ж).

Таким образом, данное исследование показало эффективность моделей алгоритмов искусственного интеллекта, как полезных инструментов для принятия решений в клинической практике фтизиатров, направленных на снижение заболеваемости туберкулезом. Выявленные факторы риска помогут установить группу больных, нуждающихся в профилактических мероприятиях по предотвращению у них в будущем рецидива МЛУ ТБ. Результаты исследования позволяют заключить, что разработан новый инструмент прогнозирования риска рецидива МЛУ ТБ, основанный на искусственном интеллекте с использованием методов машинного обучения в виде дерево решений, случайный лес, XGBoost (градиентный бустинг) и логистическая регрессия.

ВЫВОДЫ

1. Частота возникновения рецидивов после успешного лечения больных туберкулезом органов дыхания с множественной лекарственной устойчивостью за 5 последующих лет наблюдения равна 10,1 %; в большинстве случаев (77,1 %) туберкулез повторно возникает в течение первых трех лет наблюдения; средний срок наступления рецидива после эффективного курса химиотерапии больных туберкулезом органов дыхания с множественной лекарственной устойчивостью составляет 1,9 года.

2. Значимые медико-социальные факторы риска развития рецидивов у больных туберкулезом органов дыхания с множественной лекарственной устойчивостью: инвалидность по соматическим заболеваниям (ОШ = 4,82; 95 % ДИ 2,13–10,90); пенсионный возраст (ОШ = 2,86; 95 % ДИ 1,14–7,20); повторное лечение (ОШ = 5,60; 95 % ДИ 2,59–12,11); ВИЧ-инфекция (ОШ = 19,19; 95 % ДИ 5,29–69,56); наличие психических заболеваний (ОШ = 5,85; 95 % ДИ 2,27–15,03); табакокурение (ОШ = 3,16; 95 % ДИ 1,08–9,20).

3. Рецидив туберкулеза органов дыхания с множественной лекарственной устойчивостью потенциально возможен у лиц с деструктивными формами заболевания (ОШ = 7,48; 95 % ДИ 1,76–31,80), повторно леченных с остаточными изменениями в легких (ОШ = 3,38; 95 % ДИ 1,64–6,98), имеющих бактериовыделение, определяемое при микроскопии (ОШ = 2,91; 95 % ДИ 1,28–6,61), с лекарственной устойчивостью возбудителя к шести и семи противотуберкулезным препаратам, с формированием в процессе лечения амплификации к следующим препаратам: канамицин/амикацин (ОШ = 5,91; 95 % ДИ 2,85–12,26), капреомицин (ОШ = 5,27; 95 % ДИ 2,55–10,91), левофлоксацин (ОШ = 7,02; 95 % ДИ 3,36–14,66), этионамид/протионамид (ОШ = 4,30; 95 % ДИ 1,95–9,48), ПАСК (ОШ = 8,97; 95 % ДИ 4,08–19,73).

4. Прекращение бактериовыделения до 2 мес. (посев) (ОШ = 4,98; 95 % ДИ 2,41–10,29) и полноценность химиотерапии в фазе продолжения (ОШ = 0,07; 95 % ДИ 0,03–0,17) являются критериями отсутствия рецидива заболевания у

больных туберкулезом органов дыхания с множественной лекарственной устойчивостью.

5. Анализ характеристик моделей прогнозирования с помощью четырех алгоритмов машинного обучения, включающих дерево решений, случайный лес, XGBoost (градиентный бустинг) и логистическую регрессию, с использованием К-блочной стратифицированной проверки выявил высокую чувствительность ($0,74 \pm 0,167$, $0,91 \pm 0,17$, $0,91 \pm 0,14$, $0,91 \pm 0,16$ соответственно) и специфичность ($0,97 \pm 0,03$, $0,98 \pm 0,02$, $0,98 \pm 0,02$, $0,98 \pm 0,02$ соответственно).

6. С помощью алгоритмов дерево решений, случайный лес, XGBoost и логистическая регрессия установлены 5 основных предикторов развития рецидивов у успешно пролеченных больных туберкулезом органов дыхания с множественной лекарственной устойчивостью: неоднократные курсы лечения, давность заболевания, деструктивный туберкулез, общий срок лечения не достигает 22 мес. и в схеме химиотерапии присутствуют менее пяти противотуберкулезных препаратов.

7. Создана программа ЭВМ для оценки индивидуального риска развития рецидива у больных туберкулезом органов дыхания с множественной лекарственной устойчивостью.

ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ

1. Разработанную программу ЭВМ по прогнозированию индивидуального риска развития рецидива туберкулеза органов дыхания с множественной лекарственной устойчивостью с целью проведения профилактических мероприятий целесообразно применять в клинической практике врачей фтизиатров и специалистов первичной медико-санитарной помощи для выделения группы пациентов с наиболее высоким риском повторного заболевания и как стандартную модель в поддержке принятия клинических решений.

2. Быстро возникающая амплификация к инъекционным препаратам, а также к левофлоксацину, этионамиду/протионамиду и ПАСКу на фоне химиотерапии больных туберкулезом органов дыхания с множественной лекарственной устойчивостью должна быть учтена фтизиатрами как фактор риска рецидива заболевания, где крайне важно соблюдать сроки лечения и количество применяемых противотуберкулезных препаратов.

3. Ориентиром для безрецидивного течения заболевания у больных туберкулезом органов дыхания с множественной лекарственной устойчивостью служит прекращение бактериовыделения в течение 2 мес. включительно, определяемое методом посева.

4. На территории Российской Федерации необходимо введение мониторинга рецидивов больных туберкулезом органов дыхания с множественной лекарственной устойчивостью и создание индикатора по эффективности МЛУ-программы с внедрением данного показателя в форму федерального статистического наблюдения № 33 «Сведения о больных туберкулезом».

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

АО	– Автономный округ
АЛТ	– аланинаминотрасфераза
АРТ	– антиретровирусная терапия
АСТ	– аспартатаминотрансфераза
БОМЖ	– без определенного места жительства
ВИЧ	– вирус иммунодефицита человека
ВОЗ	– Всемирная организация здравоохранения
ДИ	– доверительный интервал
ДПК	– двенадцатиперстная кишка
ИИ	– искусственный интеллект
ИМТ	– индекс массы тела
КУМ	– кислотоустойчивые микобактерии
КХТ	– курс химиотерапии
ЛУ	– лекарственная устойчивость
МБТ	– микобактерия туберкулеза
МЛС	– места лишения свободы
МЛУ	– множественная лекарственная устойчивость
МО	– машинное обучение
ОАК	– общий анализ крови
ОГАУЗ	– Областное государственное автономное учреждение
«ТФМЦ»	здравоохранения «Томский фтизиопульмонологический медицинский центр»
ОШ	– отношение шансов
ПАСК	– пара-аминосалицилат натрия
ПТП	– противотуберкулезные препараты
РНК	– рибонуклеиновая кислота
РУ	– устойчивость к рифампицину
РФ	– Российская Федерация

СД	– сахарный диабет
СССР	– Союз советских социалистических республик
СОЭ	– Скорость оседания эритроцитов
США	– Соединённые Штаты Америки
СУБД	– система управления базами данных
ТБ	– туберкулез
ТВГЛУ	– туберкулез внутригрудных лимфатических узлов
ТЛЧ	– тест лекарственной чувствительности
ТТГ	– тиреотропный гормон
ТТФ	– точный тест Фишера
ФСН	– Федеральное статистическое наблюдение
ХНЗЛ	– хронические неспецифические заболевания легких
ХОБЛ	– хроническая обструктивная болезнь легких
ШЛУ	– широкая лекарственная устойчивость
ЭВМ	– электронно-вычислительная машина
ЭКХТ	– эффективный курс химиотерапии туберкулеза
ЯБЖ	– язвенная болезнь желудка
ANN	– Artificial neural network – искусственная нейронная сеть
AUC	– Area under ROC curve – площадь под ROC-кривой
В	– коэффициент предикторов в регрессии
CART- алгоритм	– classification and regression tree – критерий нормализованного прироста информации
CD4	– cluster of differentiation 4 – кластер дифференцировки 4
Сm	– капреомицин
CNN	– convolutional neural network – сверточная нейронная сеть
Cs	– циклосерин
DT	– Decision Trees – дереворешений
Е	– этамбутол
Еto	– этионамид
Exp (b)	– экспонированное отношение шансов

F	– критерий Фишера по Монте-Карло
Fq	– фторхинолоны
GB	– Gradient Boosting – градиентный бустинг
Km	– канамицин
KNN	– k-nearest neighbors algorithm – метод К ближайших соседей
Lfx	– левофлоксацин
LR	– Logistic regression – логистическая регрессия
LDA	– Linear Discriminant Analysis – линейный дискриминантный анализ
LMM	– large matrix multiplication – алгоритм перемножения матриц
Me	– медиана
Mfx	– моксифлоксацин
MIRU	– mycobacterial interspersed repetitive units – микобактериальные рассеянные повторяющиеся единицы
MLP	– multilayer perceptron – многослойный перцептрон
N	– абсолютная частота
NB	– naïve Bayes – наивный байесовский классификатор
PAS	– ПАСК
Pto	– протионамид
Q ₁ , Q ₃	– первый и третий квартили
R ²	– коэффициент детерминации
RBF	– radial basis function – радиальная базисная функция
ROC	– Receiver Operator Characteristic – рабочая характеристика моделей
Se	– Sensitivity – чувствительность
Sp	– Specificity – специфичность
SVM	– Support Vector Machines – метод опорных векторов
Trd	– теризидон
VNTR	– variable number of tandem repeats – полиморфизм числа tandemных повторов
χ^2	– критерий Хи-квадрат Пирсона
χ^{2*}	– критерий Хи-квадрат с поправкой Йетса

- χ^2_{MH} – критерий «Хи-квадрат» Мак-Немара
- XGBoost – eXtreme Gradient Boosting – градиентный бустинг
- Z – пиразинамид

СПИСОК ЛИТЕРАТУРЫ

1. 4 практических совета по применению машинного обучения в маркетинге. – Текст : электронный. – URL: <http://datareview.info/article/4-prakticheskixsoveta-po-primeneniyu-mashinnogo-obucheniya-vmarketinge/>.
2. Алексеенко, С. Н. Ранние рецидивы туберкулеза легких – эпидемиологические и экономические проблемы / С. Н. Алексеенко, Н. Н. Дробот. – Текст : электронный // Современные проблемы науки и образования. – 2019. – № 2. – С. 156. – URL: <http://www.science-education.ru/ru/article/view?id=28739> (дата обращения: 20.04.2021).
3. Анализ осложнений и рецидивов хирургического лечения больных с ограниченным легочным туберкулезом при разной морфологической активности специфического воспаления / А. А. Елипашев, В. О. Никольский, А. С. Шпрыков, А. В. Елькин // Туберкулез и болезни легких. – 2015. – № 10. – С. 20–23.
4. Анализ причин и сроков развития рецидивов туберкулеза в республике Казахстан / Ж. К. Исмаилов, Э. А. Берикова, М. А. Жапаркулова, Б. В. Ким // Наука, новые технологии и инновации Кыргызстана. – 2017. – № 8. – С. 92–97.
5. Андрианова, А. Ю. Результаты отдаленных наблюдений за лицами с остаточными туберкулезными изменениями в легких, состоящими на учете в поликлиниках общей медицинской сети / А. Ю. Андрианова, В. Ю. Мишин // Туберкулез и болезни легких. – 2011. – № 4. – С. 34–35.
6. Гайда, А. И. Отдаленные результаты лечения больных с множественной лекарственной устойчивостью микобактерий туберкулеза, прервавших курс химиотерапии / А. И. Гайда, Е. И. Никишова, А. О. Марьяндышев // Туберкулез и болезни легких. – 2014. – № 12. – С. 47–52.
7. Ганцев, Ш. Х. Искусственный интеллект в онкологии: взгляд в будущее / Ш. Х. Ганцев, М. В. Франц // Практическая онкология. – 2019. – Т. 20 (1). – С. 21–28. – DOI: 10.31917/2001021.
8. Гусев, А. В. Перспективы нейронных сетей и глубокого машинного

обучения в создании решений для здравоохранения / А. В. Гусев // Врач и информационные технологии. – 2017. – № 3. – С. 92–105.

9. Гусев, А. В. Технологии прогнозной аналитики в борьбе с пандемией COVID-19 / А. В. Гусев, Р. Э. Новицкий // Врач и информационные технологии. – 2020. – № 4. – С. 24–33

10. Гусейнов, Г. К. Поздние рецидивы у лиц с остаточными туберкулезными изменениями в легких / Г. К. Гусейнов, Т. Г. Гиреев // Туберкулез и болезни легких. – 2016. – № 94 (2). – С. 26–28.

11. Давлетбаева, Н. В. Рецидивы туберкулеза в Республике Башкортостан / Н. В. Давлетбаева, И. Н. Аталипова // Туберкулез и болезни легких. – 2015. – № 7. – С. 34–35.

12. Ившин, А. А. Искусственный интеллект: предиктивная аналитика перинатального риска / А. А. Ившин, А. В. Гусев, Р. Э. Новицкий // Вопросы гинекологии, акушерства и перинатологии. – 2020. – № 19 (6). – С. 133–144. DOI: 10.20953/1726-1678-2020-6-133-144.

13. Искусственный интеллект в онкохирургической практике / П. В. Мельников, В. Н. Доведов, Д. Ю. Каннер, И. Л. Черниковский // Тазовая хирургия и онкология. – 2020. – № 10 (3-4). – С. 60–64.

14. Искусственный интеллект победил юристов в соревновании по оценке договоров. – Текст : электронный. – URL: <https://rb.ru/news/juristi-ne-nuzni/>.

15. Использование укороченных схем лечения туберкулеза с множественной лекарственной устойчивостью в Архангельской, Мурманской, Белгородской областях / А. О. Марьяндышев, А. И. Кулижская, Е. С. Химова [и др.] // Туберкулёз и болезни лёгких. – 2019. – Т. 97, № 7. – С. 5–10.

16. Клинико-эпидемиологический анализ и причины рецидивов туберкулеза / А. Б. Цепке, Е. А. Хегай, Б. К. Ермагамбетова, С. В. Пономарева // Клиническая медицина Казахстана. – 2012. – Т. 24, № 1 (24). – С. 58–61.

17. Клиническая характеристика рецидивов туберкулезного процесса в Кемеровской области / Т. В. Пьянзова, Н. В. Лузина, И. Ф. Копылова [и др.] // Туберкулез и болезни легких – 2013. – № 9. – С. 25–28.

18. Клинические и микробиологические особенности рецидивов туберкулеза органов дыхания / О. В. Рукосуева, И. А. Васильева, В. А. Пузанов [и др.] // Проблемы туберкулеза. – 2008. – № 10. – С. 28–31.

19. Коломиец, В. М. Рецидивы туберкулеза органов дыхания / В. М. Коломиец. – Санкт-Петербург. – 2014 – 263 с.

20. Кораблев, А. Ю. Машинное обучение в бизнесе / А. Ю. Кораблев, Р. Б. Булатов // Азимут научных исследований: экономика и управление. – 2018. – Т. 7.– № 2. – С. 68–72.

21. Костин, К. А. Классификация патологий диссеминированного туберкулёза лёгких с помощью методов машинного обучения / К. А. Костин, Т. С. Ламонова // Сборник избранных статей научной сессии ТУСУР. – 2018. – Т. 1. – №. 3. – С. 129–132.

22. Максимова, О. М. Особенности развития и течения рецидивов туберкулеза легких : специальность 14.00.26 «Фтизиатрия» : автореферат диссертации на соискание кандидата медицинских наук / Максимова Ольга Михайловна ; Якутский научно-исследовательский институт туберкулеза. – Якутск, 2004. – 24 с. : ил. – Библиогр.: с. 20–22. – Место защиты: Центральный научно-исследовательский институт туберкулеза. – Текст : непосредственный.

23. Машинное обучение в транспортных перевозках и логистике. – Текст : электронный. – URL: <https://blog.heyaml.com/машинное-обучение-втранспортных-перевозках-и-логистике-d134d3ee2e69>.

24. Мелдо, А. А. Искусственный интеллект в медицине: современное состояние и основные направления развития интеллектуальной диагностики / А. А. Мелдо, Л. В. Уткин, Т. Н. Трофимова // Лучевая диагностика и терапия. – 2020. – № 11 (1). – С. 9–17. DOI: 10.22328/2079-5343-2020-11-1-9-17.

25. Методы машинного обучения как инструмент диагностических и прогностических исследований при ишемической болезни сердца. Российский кардиологический журнал / Б. И. Гельцер, М. М. Циванюк, К. И. Шахгельдян, В. Ю. Рублев. – 2020. – № 25 (12). – С. 3999. DOI: 10.15829/1560-4071-2020-3999.

26. Мониторинг рецидивов туберкулеза: причины и факторы,

способствующие развитию рецидивов / Е. И. Скачкова, О. Б. Нечаева, Э. В. Арефьева, Н. В. Ясинецкая // Проблемы туберкулеза. – 2008. – № 8. – С. 16–19.

27. Нейросети для транзакций: как на деле работают «большие данные» в российских банках? – Текст : электронный. – URL: <http://www.forbes.ru/tehnologii/341517-neyroseti-dlya-tranzakciy-kak-na-delerabotayut-bolshie-dannye-v-rossiyskih-bankah>.

28. Основные показатели противотуберкулезной деятельности в Сибирском и Дальневосточном федеральном округе (статистические материалы). – Новосибирск : Издательско-полиграфический центр НГМУ, 2020. – 112 с.

29. Особенности течения и диагностики рецидивов туберкулёза лёгких / О. В. Рукосуева, И. А. Васильева, В. А. Пузанов [и др.] // Туберкулез и болезни лёгких. – 2011. – № 7. – С. 138–139.

30. Парпиева, Н. Н. Частота и характер излеченного туберкулеза легких / Н. Н. Парпиева, С. В. Костромцов, К. С. Мухамедов // Туберкулез и болезни легких. – 2011. – № 5. – С. 102–103.

31. Плиева, С. Л. О различиях клинической структуры ранних и поздних рецидивов туберкулеза органов дыхания и впервые выявленных больных туберкулезом / С. Л. Плиева // Пермский медицинский журнал. – 2011. – Т. 28, № 6. – С. 31–35.

32. Плиева, С. Л. Основные факторы риска развития ранних рецидивов туберкулеза органов дыхания / С. Л. Плиева // Пермский медицинский журнал. – 2013. – Т. 30, № 2. – С. 43–49.

33. Плиева, С. Л. Особенности ранних и поздних рецидивов туберкулеза органов дыхания / С. Л. Плиева, П. П. Сельцовский // Туберкулез и болезни легких. – 2011. – № 6. – С. 23–27.

34. Плиева, С. Л. Прогнозирование рецидивов туберкулеза органов дыхания в современных условиях : специальность 14.01.16 «Фтизиатрия» : автореферат диссертации на соискание ученой степени кандидата медицинских наук / Плиева Светлана Леонидовна ; Российская медицинская академия непрерывного медицинского образования. – Москва, 2017. – 27 с. : ил. –

Библиогр. с. 25–26. – Место защиты: Центральный научно-исследовательский институт туберкулеза. – Текст : непосредственный.

35. Правильность оценки сердечно-сосудистого риска в повседневной клинической практике / Д. В. Гаврилов, А. В. Гусев, А. В. Никулина [и др.] // Профилактическая медицина. – 2021. – № 24 (4). – С. 69–75. DOI: 10.17116/profmed20212404169.

36. Предсказание сердечно-сосудистых событий при помощи комплексной оценки факторов риска с использованием методов машинного обучения / Д. В. Гаврилов, Л. М. Серова, И. Н. Корсаков [и др.] // Врач. – 2020. – № 5. – С. 41–45.

37. Применение машинного интеллекта в российском бизнесе. – Текст : электронный. – URL: [https:// spark.ru/startup/digital-contact/blog/38210/botirovvsyudukogda-lyudi-stanut-ne-nuzhni](https://spark.ru/startup/digital-contact/blog/38210/botirovvsyudukogda-lyudi-stanut-ne-nuzhni).

38. Причины формирования и структура клинических форм туберкулеза ранних и поздних рецидивов заболевания / Т. Ж. Сагалбаева, А. В. Мордык, А. Л. Кортусова [и др.] // Туберкулез и болезни легких. – 2015. – № 5. – С. 163–164.

39. Российская Федерация. Приказы. О введении в действие учётной и отчётной документации мониторинга туберкулёза : Приказ Минздрава Российской Федерации от 13.02.2004 г. № 50 : [утвержден Министром здравоохранения Российской Федерации]. – Москва, 2004. – 35 с. – Доступ из справочно-правовой системы КонсультантПлюс.

40. Сводное руководство ВОЗ по туберкулезу. Модуль 4: лечение (лечение лекарственно-устойчивого туберкулеза) : [сайт ВОЗ]. – Копенгаген: Европейское региональное бюро ВОЗ, 2021. – URL: <https://apps.who.int/iris/bitstream/handle/10665/339991/9789289054966-rus.pdf>. – Текст : электронный.

41. Сенин А. М. Сравнение особенностей выявления рецидивов туберкулеза у пациентов с ВИЧ-инфекцией и без наличия таковой. / А. М. Сенин, Н. В. Эйсмонт // Фтизиатрия и пульмонология. – 2017. – № 2 (15). – С. 116–133.

42. Сенин, А. М. Особенности рецидивов туберкулеза у пациентов с ВИЧ-инфекцией / А. М. Сенин, И. Д. Медвинский // Туберкулёз и болезни лёгких. – 2018. – Т. 96, № 10. – С. 41–47. DOI: 10.21292/2075-1230-2018-96-10-41-47.

43. Тарасова, Л. Г. Причины рецидивов туберкулеза легких / Л. Г. Тарасова, Е. Н. Стрельцова // Астраханский медицинский журнал. – 2014. – № 1. – С. 125–130.

44. Токторгазиева, К. А. Клинико-эпидемиологическая характеристика рецидивов туберкулеза легких по данным отделения легочного туберкулеза № 2 национального центра фтизиатрии / К. А. Токторгазиева, А. Б. Азыкова, Н. К. Курманова // Наука, новые технологии и инновации Кыргызстана. – 2017. – № 8. – С. 89–91.

45. Факторы и риски развития рецидива туберкулеза легких / К. А. Шишкина, Е. М. Богородская, Е. М. Алексеева [и др.] // Туберкулёз и болезни легких. – 2011. – № 5. – С. 238–239.

46. Факторы риска госпитальной летальности больных с сочетанием туберкулеза и ВИЧ-инфекции в учреждениях уголовно-исполнительной системы / М. Е. Вострокнутов, Е. В. Дюжева, А. В. Кузнецова, О. В. Сенько // Туберкулёз и болезни легких. – 2019. – № 97 (7). – С. 34–41.

47. Факторы, влияющие на эффективность лечения больных туберкулезом с множественной и широкой лекарственной устойчивостью / И. С. Лапшина, Б. У. Салихов, Т. В. Мякишева [и др.] // Туберкулёз и болезни лёгких. – 2020. – Т. 98, № 10. – С. 28-32.

48. X5 автоматизировала целевой маркетинг в «Перекрестке» с помощью технологии машинного обучения. – Текст : электронный. – URL: https://www.vedomosti.ru/press_releases/2017/11/01/x5-avtomatizirovalatselevoi-marketing-v-perekrestke-s-pomoschyu-tehnologiiimashinnogo-obucheniya.

49. Характеристика больных, заболевших туберкулезом, в г. Кишинёве в 2015 г. / Э. Лесник, А. Нигуляну, А. Малик, К. Яворский // Туберкулёз и болезни лёгких. – 2018. – Т. 96, № 7. – С. 46–49. DOI: 10.21292/2075-1230-2018-96-7-46-49.

50. Характеристика рецидивов туберкулеза у пациентов с ВИЧ-

инфекцией / Е. Ю. Зоркальцева, Ю. О. Егорова, О. А. Воробьева, Е. В. Батунова // *Забайкальский медицинский вестник*. – 2018. – № 4. – С. 14–18.

51. Цифровая трансформация производства. – Текст : электронный. – URL: <http://www.dupont.ru/products-and-services/consulting-services-processtechnologies/articles/digital-transformation-productionoperations.html>.

52. Шитиков, В. К. Дискриминантные функции для классификации многомерных объектов / В. К. Шитиков, Г. С. Розенберг, Т. Д. Зинченко. – Текст : электронный. – URL: <http://www.ievbras.ru/ecostat/Kiril/Library/Book1/Content383/Content383.htm>.

53. A combination of baseline plasma immune markers can predict therapeutic response in multidrug resistant tuberculosis / S. Ferriani, C. Manca, S. Lubbe [et al.] // *PLoS ONE*. – 2017. – Vol. 12 (5). – P. e0176660. DOI: 10.1371/journal.pone.0176660.

54. A nested case-control study of predictors for tuberculosis recurrence in a large UK Centre / A. Rosser, M. Richardson, M. J. Wiselka [et al.] // *BMC Infect Dis*. – 2018. – Vol. 18 (1). – P. 94. DOI: 10.1186/s12879-017-2933-4.

55. A nested case-control study on relapse predictors among tuberculosis patients treated in Yemen's NTCP / M. S. Anaam, M. I. Ibrahim, A. W. Al Serouri [et al.] // *Public Health Action*. – 2012. – Vol. 2. – P. 168–173.

56. A novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest X-rays. *IEEE Trans* / J. Melendez, B. van Ginneken, P. Maduskar [et al.] // *Med. Imaging*. – 2015. – Vol. 34. – P. 179–192.

57. A screening system for smear-negative pulmonary tuberculosis using artificial neural networks / B. D. O. JoãoFilho, J. M. de Seixas, R. Galliez [et al.] // *Int. J. Infect. Dis*. – 2016. – Vol. 49. – P. 33–39.

58. Abubakar Risk factors for recurrent tuberculosis in England and Wales, 1998–2005 / J. P. Crofts, N. J. Andrews, R. D. Barker, I. V. Delpech // *Thorax*. – 2010. – Vol. 65. – P. 310–314.

59. Adaptive random forests for evolving data stream classification / H. M. Gomes, A. Bifet, J. Read [et al.] // *Mach. Learn*. – 2017. – Vol. 106. – P. 1469–1495.

60. Aggressive regimens for multidrug-resistant tuberculosis reduce recurrence / M. F. Franke, S. C. Appleton, C. D. Mitnick [et al.] // *Clin Infect Dis.* – 2013. – Vol. 56 (6). – P. 770–776. DOI: 10.1093/cid/cis1008.
61. Aggressive Regimens Reduce Risk of Recurrence After Successful Treatment of MDR-TB / F. A. Khan, I. Y. Gelmanova, M. F. Franke [et al.] // *Clin Infect Dis.* – 2016. – Vol. 63 (2). – P. 214–220. DOI: 10.1093/cid/ciw276.
62. Application of machine learning techniques to tuberculosis drug resistance analysis / S. Kouchaki, Y. Yang, T. M. Walker [et al.] // *Bioinformatics.* – 2019. – Vol. 35 (13). – P. 2276–2282. DOI: 10.1093/bioinformatics/bty949.
63. Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review / Z. Ahmad, S. Rahim, M. Zubair [et al.] // *Diagn Pathol.* – 2021. – Vol. 16. – P. 24. DOI: 10.1186/s13000-021-01085-4.
64. Artificial Intelligence Applications in Dermatology: Where Do We Stand? / A. Gomolin, E. Netchiporouk, R. Gniadecki, I. V. Litvinov // *Front Med (Lausanne).* – 2020. – Vol. 7. – P. 100. DOI: 10.3389/fmed.2020.00100.
65. Artificial Intelligence: A New Paradigm in Obstetrics and Gynecology Research and Clinical Practice / P. Iftikhar, M. V. Kuijpers, A. Khayyat [et al.] // *Cureus.* – 2020. – Vol. 12 (2). – P. e7124. DOI: 10.7759/cureus.7124.
66. Artificial neural networks for prediction of tuberculosis disease / M. T. Khan, A. C. Kaushik, L. Ji [et al.] // *Front Microbiol.* – 2019. – Vol. 10. – P. 395.
67. Association Between Ambient Air Pollution and Elevated Risk of Tuberculosis Development / Y. J. Lin, H. C. Lin, Y. F. Yang [et al.] // *Infect Drug Resist.* – 2019. – Vol. 12. – P. 3835–3847. DOI: 10.2147/IDR.S227823.
68. Association of smoking with recurrence of pulmonary Kochs; after completion of antituberculous treatment / A. Khan, F. Aslam, S. Abbas, Q. U. Elahi // *J Ayub Med Coll Abbottabad.* – 2016. – Vol. 28. – P. 781–787. PMID: 28586615.
69. Basile, A. O. Artificial Intelligence for Drug Toxicity and Safety /

A. O. Basile, A. Yahi, N. P. Tatonetti // *Trends Pharmacol Sci.* – 2019. – Vol. 40 (9). – P. 624–635. DOI: 10.1016/j.tips.2019.07.005.

70. Benchmarking Machine Learning Models to Assist in the Prognosis of Tuberculosis / L. F. da Silva Barros, G. Oliveira Alves, S. L. Morais Florêncio [et al.] // *Informatics.* – 2021. – Vol. 8. – P. 27.

71. Bonte, C. Privacy-preserving logistic regression training / C. Bonte, F. Vercauteren // *BMC Med. Genom.* – 2018. – Vol. 11. – P. 86.

72. Breiman, L. Statistical modeling: the two cultures (with comments and a rejoinder by the author) / L. Breiman // *Statist Sci* 2001. – Vol. 16. – P. 199–231.

73. Buchanan, B. G. A (very) brief history of artificial intelligence / B. G. Buchanan // *AI Magazine.* – 2005. – Vol. 26 (4). – P. 53.

74. Chaisson, R. E. Recurrent tuberculosis: relapse, reinfection, and HIV / R. E. Chaisson, G. J. Churchyard // *J Infect Dis.* – 2010. – Vol. 201 (5). – P. 653–655. DOI: 10.1086/650531.

75. Changes in treatment outcomes of multidrug-resistant tuberculosis / N. Kwak, H. R. Kim, C. G. Yoo [et al.] // *Int J Tuberc Lung Dis.* – 2015. – Vol. 19. – P. 525–530. DOI: 10.5588/ijtld.14.0739.

76. Chronic lung disease and HIV infection are risk factors for recurrent tuberculosis in a low-incidence setting / A. C. Pettit, L. A. Kaltenbach, F. Maruri [et al.] // *Int J Tuberc Lung Dis.* – 2011. – Vol. 15. – P. 906–911.

77. Collaborative Group for the Meta-Analysis of Individual Patient Data in MDR-TB treatment–2017. Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis / N. Ahmad, S. D. Ahuja, O. W. Akkerman [et al.] // *Lancet.* – 2018. – Vol. 392 (10150). – P. 821–834. DOI: 10.1016/S0140-6736(18)31644-1.

78. Correlation of plasma anti-tuberculosis drug levels with subsequent development of hepatotoxicity / A. Satyaraddi, T. Velpandian, S. K. Sharma [et al.] // *The International Journal of Tuberculosis and Lung Disease.* – 2014. – Vol. 18. – P. 188–195.

79. Countrywide management of pulmonary tuberculosis reverses increasing

incidence / K. Blöndal, P. Viiklepp, P. Blöndal, A. Altraja // *Int J Tuberc Lung Dis.* – 2011. – Vol. 15 (7). – P. 892–898. DOI: 10.5588/ijtld.10.0601.

80. Cudahy, P. G. T. Risk factors for recurrent tuberculosis after successful treatment in a high burden setting: a cohort study / P. G. T. Cudahy, D. Wilson, T. Cohen // *BMC Infect Dis.* – 2020. – Vol. 20. – P. 789. (2020). DOI: 10.1186/s12879-020-05515-4.

81. Das, K. A survey on machine learning: Concept, algorithms and applications / K. Das, R. N. Behera // *Int. J. Innov. Res. Comput. Commun. Eng.* – 2017. – Vol. 5. – P. 1301–1309.

82. Decision Tree-initialised Neuro-fuzzy Approach for Clinical Decision Support / T. Chen, C. Shang, P. Su [et al.] // *Artificial Intelligence in Medicine.* – 2020. – Vol. 111. – P. 101986. DOI: 10.1016/j.artmed.2020.101986.

83. Development of a risk score for prediction of poor treatment outcomes among patients with multidrug-resistant tuberculosis / K. A. Alene, K. Viney, D. J. Gray [et al.] // *PLOS ONE.* – 2020. – Vol. 15 (1). – P. e0227100. DOI: 10.1371/journal.pone.0227100.

84. Diabetes and risk of tuberculosis relapse: nationwide nested case-control study / P. H. Lee, H. C. Lin, A. S. Huang [et al.] // *PLOS ONE.* – 2014. – Vol. 9. – P. 92623.

85. Diagnostic Accuracy of Different Machine Learning Algorithms for Breast Cancer Risk Calculation: a Meta-Analysis / R. D. Nindrea, T. Aryandono, L. Lazuardi, I. Dwiprahasto // *Asian Pac J Cancer Prev.* – 2018. – Vol. 19 (7). – P. 1747–1752. DOI: 10.22034/APJCP.2018.19.7.1747.

86. Dimala, C. A. The association between ambient air pollution and pulmonary tuberculosis: a systematic review protocol / C. A. Dimala, B. M. Kadia, A. Hansell // *Environ Evid.* – 2020. – Vol. 9. – P.29. DOI: 10.1186/s13750-020-00213-9.

87. eDoctor: machine learning and the future of medicine / G. S. Handelman, H. K. Kok, R. V. Chandra [et al.] // *J Intern Med.* – 2018. – Vol. 284 (6). – P. 603–619. DOI: 10.1111/joim.12822.

88. Effect of duration and intermittency of rifampin on tuberculosis treatment

outcomes: a systematic review and meta-analysis / D. Menzies, A. Benedetti, A. Paydar [et al.] // *PLOS Med.* – 2009. – Vol. 6. – P. e1000146.

89. Évora, L. H. R. A. Neural network models for supporting drug and multidrug resistant tuberculosis screening diagnosis / L. H. R. A. Évora, J. M. Seixas, A. L. Kritski // *Neurocomputing.* – 2017. – Vol. 265. – P. 116–126.

90. Exogenous reinfection as a cause of late recurrent tuberculosis in the United States / J. D. Interrante, M. B. Haddad, L. Kim, N. R. Gandhi // *Ann Am Thorac Soc.* – 2015. – Vol. 12. – P. 1619–1626.

91. Exogenous reinfection of tuberculosis in a low-burden area / C. Schioli, M. Carugati, F. Zanini [et al.] // *Infection.* – 2015. – Vol. 43. – P. 647–653.

92. Factors associated with diagnostic delay in recurrent TB / Z. Xie, T. Wang, H. Chen [et al.] // *BMC Public Health.* – 2020. – Vol. 20 (1). – P. 1207. DOI: 10.1186/s12889-020-09005-9.

93. Factors associated with recurrent tuberculosis more than 12 months after treatment completion / L. Kim, P. K. Moonan, C. M. Heilig [et al.] // *Int J Tuberc Lung Dis.* – 2016. – Vol. 20. – P. 49–56.

94. Feature selection and prediction of treatment failure in tuberculosis / C. M. Sauer, D. Sasson, K. E. Paik [et al.] // *PLoS One.* – 2018. – Vol. 13 (11). – P. e0207491. DOI: 10.1371/journal.pone.0207491.

95. Feigenbaum, E. A. DENDRAL and Meta-DENDRAL: roots of knowledge systems and expert system applications / E. A. Feigenbaum, B. G. Buchanan // *Artificial Intelligence.* – 1993. – Vol. 59 (1-2). – P. 233–240. DOI: 10.1016/0004-3702(93)90191-d.

96. Genome-wide Analysis of MDR and XDR Tuberculosis from Belarus: Machine-learning Approach / R. S. Sergeev, I. Kavaliou, U. Sataneuski [et al.] // *IEEE/ACM Trans. Comput. Biol. Bioinform.* – 2017. – Vol. 16 (4). – P. 1398–1408. DOI: 10.1109/TCBB.2017.2720669.

97. Ghiasi, M. M. Decision tree-based methodology to select a proper approach for wart treatment / M. M. Ghiasi, S. Zendejboudi // *Computers in Biology and Medicine.* – 2019. – Vol. 108. – P. 400–409. DOI:

10.1016/j.compbimed.2019.04.001.

98. Global tuberculosis database : [сайт WHO]. – Geneva, Switzerland : WHO, 2017. – URL: <http://www.who.int/tb/data/en/>. Accessed November 2017. – Текст : электронный.

99. Global tuberculosis report 2019 : [сайт WHO] / World Health Organization : Printed in France, 2019. – 279 p. – URL: http://www.who.int/tb/publications/global_report/en/ (дата обращения: 20.07.2020). – Текст : электронный.

100. Global tuberculosis report 2020 : [сайт WHO] / World Health Organization. Printed in France, 2020. – 232 p. – URL: <https://apps.who.int/iris/bitstream/handle/10665/336069/9789240013131-eng.pdf> (дата обращения 20.04.2021). – Текст : электронный.

101. Guyon, I. An Introduction to Variable and Feature Selection / I. Guyon, A. Elisseeff // Journal of Machine Learning Research. – 2003. – Vol. 3. – P. 1157–1182.

102. Hayes, M. T. Machine Learning / M. T. Hayes . – New York, NY, USA : McGraw Hill, 1997. – 432 p.

103. Hayes, P. J. On John McCarthy's 80th birthday, in honor of his contributions / P. J. Hayes, L. Morgenstern // AI Magazine. – 2007. – Vol. 28, № 4. – P. 93.

104. High cure rate with standardised short-course multidrug-resistant tuberculosis treatment in Niger: no relapses / A. Piubello, S. H. Harouna, M. B. Souleymane [et al.] // Int J Tuberc Lung Dis. – 2014. – Vol. 18 (10). – P. 1188–1194. DOI: 10.5588/ijtld.13.0075.

105. High effectiveness of a 12-month regimen for MDR-TB patients in Cameroon / C. Kuaban, J. Noeske, H. L. Rieder [et al.] // Int J Tuberc Lung Dis. – 2015. – Vol. 19 (5). – P. 517–524. DOI: 10.5588/ijtld.14.0535.

106. High prevalence of multi drug resistant tuberculosis in people living with HIV in Western India / N. Saldanha, K. Runwal, C. Ghanekar [et al.] // BMC Infect Dis. – 2019. – Vol. 19. – P. 391. DOI: 10.1186/s12879-019-4042-z.

107. Host blood RNA signatures predict the outcome of tuberculosis treatment / E. G. Thompson, Y. Du, S. T. Malherbe [et al.] // Tuberculosis. – 2017. – Vol. 107. –

P. 48–58.

108. How soon after infection with HIV does the risk of tuberculosis start to increase? a retrospective cohort study in South African gold miners / P. Sonnenberg, J. R. Glynn, K. Fielding [et al.] // *Journal of Infectious Diseases*. – 2005. – Vol. 191. – P. 150–158. DOI: 10.1086/426827.

109. Hsu, W. Section Editors for the IMIA Yearbook Section on Sensors, Signals, and Imaging Informatics. Advancing Artificial Intelligence in Sensors, Signals, and Imaging Informatics / W. Hsu, C. Baumgartner, T. Deserno // *Yearb Med Inform.* – 2019. – Vol. 28 (1). – P. 115–117. DOI: 10.1055/s-0039-1677943.

110. Hung, C. L. Associated factors for tuberculosis recurrence in Taiwan: a nationwide nested casecontrol study from 1998 to 2010 / C. L. Hung, J. Y. Chien, C. Y. Ou // *PLOS ONE*. – 2015. – Vol. 10. – P. e0124822.

111. Hussain, O. A. Predicting treatment outcome of drug-susceptible tuberculosis patients using machine-learning models / O. A. Hussain, K. N. Junejo // *Inform. Health Soc. Care*. – 2019. – Vol. 44. – P. 135–151.

112. Impact of Multidrug Resistance on Tuberculosis Recurrence and Long-Term Outcome in China / Y. Sun, D. Harley, H. Vally, A. Sleigh // *PLoS One*. – 2017. – Vol. 12 (1). – P. e0168865. DOI: 10.1371/journal.pone.0168865.

113. Improving protein-protein interactions prediction accuracy using XGBoost feature selection and stacked ensemble classifier / C. Chen, Q. Zhang, B. Yu [et al.] // *Comput Biol Med*. – 2020. – Vol. 123. – P. 103899. DOI: 10.1016/j.combiomed.2020.103899.

114. Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor communities in Perú / M. F. Alcantara, Y. Cao, C. Liu [et al.] // *Smart Health*. – 2017. – Vol. 1. – P. 66–76.

115. Introduction to machine and deep learning for medical physicists / S. Cui, H. H. Tseng, J. Pakela [et al.] // *Med Phys*. – 2020. – Vol. 47. – P. e127–1247.

116. Irimatea, K. M. Partitioned GMM logistic regression models for longitudinal data / K. M. Irimatea, J. Broatch, J. R. Wilson // *Stat Med*. – 2019. – Vol. 38 (12). – P. 2171–2183. DOI: 10.1002/sim.8099.

117. Kalhori, S. R. N. Evaluation and comparison of different machine learning methods to predict outcome of tuberculosis treatment course / S. R. N. Kalhori, X. J. Zeng // *J. Intell. Learn. Syst. Appl.* – 2013. – Vol. 5. – P. 10.
118. K-SVM: An Effective SVM Algorithm Based on K-means Clustering / Y. Yao, Y. Liu, Y. Yu [et al.] // *JCP.* – 2013. – Vol. 8. – P. 2632–2639.
119. Kulikowski, C. A. Beginnings of artificial intelligence in medicine (AIM): computational artifices assisting scientific inquiry and clinical art—with reflections on present aim challenges / C. A. Kulikowski // *Yearbook Med Inform.* – 2019. – Vol. 28. – P. 249.
120. Lack of weight gain and relapse risk in a large tuberculosis treatment trial / A. Khan, T. R. Sterling, R. Reves [et al.] // *Am J Respir Crit Care Med.* – 2006. – Vol. 174. – P. 344–348.
121. Lakhani, P. Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks / P. Lakhani, B. Sundaram // *Radiology.* – 2017. – Vol. 284. – P. 574–582.
122. Lin, H. H. Tobacco smoke, indoor air pollution and tuberculosis: A systematic review and meta-analysis / H. H. Lin, M. Ezzati, M. Murray // *PLoS Medicine.* – 2007. – Vol. 4 (1). – P. e20.
123. Logistic regression analyses of factors affecting fertility of intrauterine adhesions patients / X. Zhao, Y. Liu, A. Zhang [et al.] // *Ann Transl Med.* – 2020. – Vol. 8 (4). – P. 49. DOI: 10.21037/atm.2019.11.115.
124. Long term outcome of multidrug-resistant TB patients in Fukujuji Hospital in Japan / T. Yoshiyama, K Morimoto, M Okumura [et al.] // *Trans R Soc Trop Med Hyg.* – 2014. – Vol. 108 (9). – P. 589–590. DOI: 10.1093/trstmh/tru080.
125. Lopes, U. K. Pre-trained convolutional neural networks as feature extractors for tuberculosis detection / U. K. Lopes, J. F. Valiati // *Comput. Biol. Med.* – 2017. – Vol. 89. – P. 135–143.
126. Low rates of recurrence after successful treatment of multidrug-resistant tuberculosis in Tomsk, Russia / I. Y. Gelmanova, F. Ahmad Khan, M. C. Becerra [et al.] // *Int J Tuberc Lung Dis.* – 2015. – Vol. 19. – P. 399–405.

127. Machine Learning в коммерческих банках. – Текст : электронный. – URL: <https://www.codeinstinct.pro/2015/07/big-data-in-banking.html>.
128. Management of multidrug-resistant tuberculosis with shorter treatment regimen in Niger: Nationwide programmatic achievements / A. Piubello, M. B. Souleymane, S. Hassane-Harouna [et al.] // *Respir Med.* – 2020. – Vol. 161. – P. 105844. DOI: 10.1016/j.rmed.2019.105844.
129. Menard, S. *Applied Logistic Regression Analysis* / S. Menard. – SAGE : Thousand Oaks, CA, USA, 2002. – Vol. 106. – 128 p.
130. Meta-analysis of clinical studies supports the pharmacokinetic variability hypothesis for acquired drug resistance and failure of antituberculosis therapy / J. G. Pasipanodya, S. Srivastava, T. Gumbo [et al.] // *Clinical Infectious Diseases.* – 2012. – Vol. 55. – P. 169–177.
131. Mintz, Y. Introduction to artificial intelligence in medicine / Y. Mintz, R. Brodie // *Minim Invasive Ther Allied Technol.* – 2019. – Vol. 28 (2). – P. 73–81. DOI: 10.1080/13645706.2019.1575882.
132. Mirsaedi, M. Patients at high risk of tuberculosis recurrence / M. Mirsaedi, R. T. Sadikot // *Int J Mycobacteriol.* – 2018. – Vol. 7. – P. 1–6. DOI: 10.4103/ijmy.ijmy_164_17.
133. Moodley, Y. A systematic review of published literature describing factors associated with tuberculosis recurrence in people living with HIV in Africa / Y. Moodley, K. Govender // *Afr Health Sci.* – 2015. – Vol. 15. – P. 1239–1246.
134. Muhathir, F. Analysis K-Nearest Neighbors (KNN) in Identifying Tuberculosis Disease (Tb) By Utilizing Hog Feature Extraction / M. Muhathir, T. T. S. Sibarani, A. K. Al-Khowarizmi // *Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal.* – 2020. – Vol. 1 (1). – P. 33–38.
135. Multidrug-resistant tuberculosis not due to noncompliance but to between-patient pharmacokinetic variability / S. Srivastava, J. G. Pasipanodya, C. Meek [et al.] // *The Journal of Infectious Diseases.* – 2011. – Vol. 204. – P. 1951–1959.
136. Natekin, A. Gradient boosting machines, a tutorial / A. Natekin // *Front. Neurobot.* – 2013. – Vol. 7. – P. 21.

137. Nguyen, D. T. Development and validation of a prognostic score to predict tuberculosis mortality / D. T. Nguyen, E. A. Graviss // *J. Infect.* – 2018. – Vol. 77. – P. 283–290.

138. Nguyen, D. T. Development and validation of a risk score to predict mortality during TB treatment in patients with TB-diabetes comorbidity / D. T. Nguyen, E. A. Graviss // *BMC Infect. Dis.* – 2019. – Vol. 19. – P. 1–8.

139. Nguyen, D. T. Prognostic score to predict mortality during TB treatment in TB/HIV co-infected patients / D. T. Nguyen, H. E. Jenkins, E. A. Graviss // *PLoS ONE.* – 2018. – Vol. 13. – P. e0196022.

140. Nunn P Effects of human immunodeficiency virus infection on recurrence of tuberculosis after rifampin-based treatment: an analytical review / E. L. Korenromp, F. Scano, B. G. Williams, C. Dye // *Clin Infect Dis.* – 2003. – Vol. 37. – P. 101–112.

141. Obermeyer, Z. Predicting the Future – Big Data, Machine Learning, and Clinical Medicine / Z. Obermeyer, E. J. Emanuel // *N Engl J Med.* – 2016. – Vol. 375 (13). – P. 1216–1219. DOI: 10.1056/NEJMp1606181.

142. Outcomes of a nine-month regimen for rifampicin-resistant tuberculosis up to 24 months after treatment completion in nine African countries / V. Schwöebel, A. Trébucq, Z. Kashongwe [et al.] // *EClinicalMedicine.* – 2020. – Vol. 20. – P. 100268. DOI: 10.1016/j.eclinm.2020.100268.

143. Peering Into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods / G. S. Handelman, H. K. Kok, R. V. Chandra [et al.] // [published correction appears in *AJR Am J Roentgenol.* – 2019. – Vol. 212 (2). – P. 479]. *AJR Am J Roentgenol.* – 2019. – Vol. 212 (1). – P. 38–43. DOI: 10.2214/AJR.18.20224.

144. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost / N. Hou, M. Li, L. He [et al.] // *J Transl Med.* – 2020. – Vol. 18 (1). – P. 462. DOI: 10.1186/s12967-020-02620-5.

145. Prediction of Acute Kidney Injury after Liver Transplantation: Machine Learning Approaches vs. Logistic Regression Model / H. C. Lee, S. B. Yoon, S. M. Yang [et al.] // *J Clin Med.* – 2018. – Vol. 7 (11). – P. 428.

DOI: 10.3390/jcm7110428.

146. Prediction of Progression in Idiopathic Pulmonary Fibrosis using CT Scans at Baseline: A Quantum Particle Swarm Optimization – Random Forest Approach Artificial Intelligence in Medicine / Y. Shi, W. K. Wong, J. G. Goldin [et al.] // *Artif Intell Med.* – 2019. – Vol. 100. – P. 101709. DOI: 10.1016/j.artmed.2019.101709.

147. Predictors of recurrent TB in sputum smear and culture positive adults: a prospective cohort study / G. Muzanyi, Y. Mulumba, P. Mubiri [et al.] // *Afr Health Sci.* – 2019. – Vol. 19 (2). – P. 2091–2099. DOI: 10.4314/ahs.v19i2.33.

148. Prevalent infectious tuberculosis in Harare, Zimbabwe: burden, risk factors and implications for control / E. L. Corbett, T. Bandason, Y. B. Cheung [et al.] // *Int J Tuberc Lung Dis.* – 2009. – Vol. 13. – P. 1231–1237.

149. Pulmonary tuberculosis as a risk factor for chronic obstructive pulmonary disease: a systematic review and meta-analysis / H. Fan, F. Wu, J. Liu [et al.] // *Ann Transl Med.* – 2021. – Vol. 9 (5). – P. 390. DOI: 10.21037/atm-20-4576.

150. Qin, F. Factors associated with extended treatment among tuberculosis patients at risk of relapse in California / F. Qin, P. M. Barry, L. Pascopella // *Int J Tuberc Lung Dis.* – 2016. – Vol. 20. – P. 363–369.

151. Rajkomar, A. Machine Learning in Medicine / A. Rajkomar, J. Dean, I. Kohane // *N Engl J Med.* – 2019. – Vol. 380 (14). – P. 1347–1358. DOI: 10.1056/NEJMra1814259.

152. Rate and risk factors of recurrent tuberculosis in Yemen: a 5-year prospective study / M. S. Anaam, A. A. Alrasheedy, S. Alsaahli [et al.] // *Infect Dis (Lond).* – 2020. – Vol. 52 (3). – P. 161–169. DOI: 10.1080/23744235.2019.1690162.

153. Recurrence after Successful Treatment of Multidrug-Resistant Tuberculosis in Taiwan / M. Y. Chen, Y. C. Lo, W. C. Chen [et al.] // *PLoS One.* – 2017. – Vol. 12 (1). – P. e0170980. DOI: 10.1371/journal.pone.0170980.

154. Recurrence Is a Noticeable Cause of Rifampicin-Resistant Mycobacterium tuberculosis in the Elderly Population in Jiangxi, China / Q. Chen, L. Peng, G. Xiong [et al.] // *Front Public Health.* – 2019. – Vol. 7. – P. 182. DOI: 10.3389/fpubh.2019.00182.

155. Recurrence of tuberculosis among newly diagnosed sputum positive pulmonary tuberculosis patients treated under the Revised National Tuberculosis Control Programme, India: A multi-centric prospective study / B. Velayutham, V. K. Chadha, N. Singla [et al.] // *PLoS One*. – 2018. – Vol. 13 (7). – P. e0200150. DOI: 10.1371/journal.pone.0200150.

156. Recurrence of tuberculosis among patients following treatment completion in eight provinces of Vietnam: A nested case-control study / J. R. B. M. Bestrashniy, V. N. Nguyen, T. L. Nguyen [et al.] // *Int J Infect*. –2018. – Vol. 74. – P. 31–37. DOI: 10.1016/j.ijid.2018.06.013.

157. Recurrence of tuberculosis in a low-incidence setting without directly observed treatment: Victoria, Australia, 2002–2014 / K. D. Dale, M. Globan, E. L. Tay [et al.] // *Int J Tuberc Lung Dis*. – 2017. – Vol. 21. – P. 550–555.

158. Recurrent tuberculosis and associated factors: A five - year country wide study in Uzbekistan / J. Dadoev, D. Asadov, A. D. Harries [et al.] // *PLoS ONE*. – 2017. – Vol. 12(5). – P. e0176473. DOI: 10.1371/journal.pone.

159. Recurrent tuberculosis in HIV-infected patients in Rio de Janeiro, Brazil / J. E. Golub, B. Durovni, B. S. King [et al.] // *AIDS*. – 2008. – Vol. 22. – P. 2527–2533.

160. Recurrent tuberculosis in Houston, Texas: a population-based study / H. M. El-Sahly, J. A. Wright, H. Soini [et al.] // *Int J Tuberc Lung Dis*. – 2004. – Vol. 8. – P. 333–340.

161. Recurrent tuberculosis in the United States and Canada: relapse or reinfection? / R. M. Jasmer, L. Bozeman, K. Schwartzman [et al.] // *American Journal of Respiratory and Critical Care Medicine*. – 2004. – Vol. 170 (12). – P. 1360–1366. PMID: 15477492.

162. Relapse associated with active disease caused by Beijing strain of *Mycobacterium tuberculosis* / W. J. Burman, E. E. Bliven, L. Cowan [et al.] // *Emerg Infect Dis*. – 2009. – Vol. 15. – P. 1061–1067.

163. Relapse in persons treated for drug-susceptible tuberculosis in a population with high coinfection with human immunodeficiency virus in New York City / C. R. Driver, S. S. Munsiff, J. Li [et al.] // *Clin Infect Dis*. – 2001. – Vol. 33. –

P. 1762–1769.

164. Relapse of tuberculosis after treatment in human immunodeficiency virus-infected patients / F. Pulido, J. M. Pena, R. Rubio [et al.] // *Arch Intern Med.* – 1997. – Vol. 157. – P. 227–232.

165. Relapse or Re-Infection, the Situation of Recurrent Tuberculosis in Eastern China / Y. Shao, H. Song, G. Li [et al.] // *Front Cell Infect Microbiol.* – 2021. – Vol. 11. – P. 638990. DOI: 10.3389/fcimb.2021.638990.

166. Risk factors for recurrence in tuberculosis / P. D. Picon, S. L. Basanesi, M. L. A. Caramori [et al.] // *Jornal Brasileiro de Pneumologia.* – 2007. – Vol. 33 (5). – P. 572–578.

167. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain / F. Rosenblatt // *Psychological Review.* – 1958. – Vol. 65 (6). – P. 386–408. DOI: 10.1037/h0042519.

168. Rosser, A. Recurrent tuberculosis in the pre-elimination era. *The International Journal of Tuberculosis and Lung Disease* / A. Rosser, F. M. Marx, M. Pareek // *the Official Journal of the International Union Against Tuberculosis and Lung Disease.* – 2018. – Vol. 22 (2). – P. 139–150. DOI: 10.5588/ijtld.17.0590.

169. Sarker, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT* / I. H. Sarker // *SCI.* – 2021. – Vol. 2. – P. 160. DOI: 10.1007/s42979-021-00592-x.

170. Scholkopf, B. Fisher discriminant analysis with kernels / B. Scholkopf, K. R. Mullert // *Neural networks for signal processing IX.* – 1999. – Vol. 1 (1). – P. 41–48.

171. Schwalbe, N. Artificial intelligence and the future of global health / N. Schwalbe, B. Wahl // *Lancet.* – 2020. – Vol. 395 (10236). – P. 1579–1586. DOI: 0.1016/S0140-6736(20)30226-9.

172. Seddon, J. A. Drug-resistant tuberculosis and advances in the treatment of childhood tuberculosis / J. A. Seddon, H. S. Schaaf // *Pneumonia.* – 2016. – Vol. 8. – P. 20. DOI: 10.1186/s41479-016-0019-5.

173. Seung, K. J. Multidrug-Resistant Tuberculosis and Extensively Drug-Resistant Tuberculosis / K. J. Seung, S. Keshavjee, M. L. Rich // *Cold Spring Harb*

Perspect Med. – 2015. – Vol. 5 (9). – P. a017863. DOI: 10.1101/cshperspect.a017863.

174. Shorter treatment duration for selected patients with multidrug-resistant tuberculosis / E. C. C. Leung, W. W. Yew, C. C. Leung [et al.] // *European Respiratory Journal*. – 2011. – Vol. 38 (1). – P. 227–230. DOI: 10.1183/09031936.00186310.

175. Shortliffe, E. H. Artificial Intelligence in Medicine: Weighing the Accomplishments, Hype, and Promise / E. H. Shortliffe // *Yearb Med Inform.* – 2019. – Vol. 28 (1). – P. 257–262. DOI: 10.1055/s-0039-1677891.

176. Smoking adversely affects treatment response, outcome and relapse in tuberculosis / C. C. Leung, W. W. Yew, C. K. Chan [et al.] // *The European Respiratory Journal*. – 2015. – Vol. 45 (3). – P. 738–745.

177. Smoking and adherence to anti-tuberculosis treatment / M. Underner, J. Perriot, G. Peiffer [et al.] // *Rev Mal Respir.* – 2016. – Vol. 33. – P. 128–144.

178. Sputum monitoring during tuberculosis treatment for predicting outcome: Systematic review and meta-analysis / D. J. Horne, S. E. Royce, L. Gooze [et al.] // *The Lancet Infectious Diseases*. – 2010. – Vol. 10 (6). – P. 387–394.

179. Successful '9-month Bangladesh regimen' for multidrug-resistant tuberculosis among over 500 consecutive patients / K. J. Aung, A. Van Deun, E. Declercq [et al.] // *Int J Tuberc Lung Dis.* – 2014. – Vol. 18(10). – P. 1180–1187. DOI: 10.5588/ijtld.14.0100. PMID: 25216831.

180. Surveillance of tuberculosis (TB) cases attributable to relapse or reinfection in London, 2002–2015 / B. Afshar, J. Carless, A. Roche [et al.] // *PLoS ONE*/ – 2019. – Vol. 14 (2). – P. e0211972. DOI: 10.1371/journal.pone.0211972.

181. Talita, A. Naïve Bayes Classifier and Particle Swarm Optimization Feature Selection Method for Classifying Intrusion Detection System Dataset / A. Talita, O. Nataza, Z. Rustam // *J. Phys. Conf. Ser. IOP Publ.* – 2021. – Vol. 1752. – P. 012021.

182. Temporal Factors and Missed Doses of Tuberculosis Treatment. A Causal Associations Approach to Analyses of Digital Adherence Data / H. R. Stagg, J. J. Lewis, X. Liu [et al.] // *Ann Am Thorac Soc.* – 2020. – Vol. 17 (4). – P. 438–449. DOI: 10.1513/AnnalsATS.201905-394OC.

183. The coming of age of artificial intelligence in medicine / V. L. Patel,

E. H. Shortliffe, M. Stefanelli [et al.] // *Artif Intell Med.* – 2009. – Vol. 46 (1). – P. 5–17. DOI: 10.1016/j.artmed.2008.07.017.

184. The impact of alcohol use on tuberculosis treatment outcomes: a systematic review and meta-analysis / E. Ragan, M. Kleinman, B. Sweigart [et al.] // *Int J Tuberc Lung Dis.* – 2020. – Vol. 24 (1). – P. 73–82. DOI: 10.5588/ijtld.19.0080.

185. The impact of HIV infection on recurrence of tuberculosis in South African gold miners / K. F. Mallory, G. J. Churchyard, I. Kleinschmidt [et al.] // *International Journal of Tuberculosis and Lung Disease.* – 2000. – Vol. 4. – P. 455–462.

186. The incidence of recurrence of tuberculosis and its related factors in smear-positive pulmonary tuberculosis patients in Iran: A retrospective cohort study / M. Moosazadeh, A. Bahrampour, M. Nasehi, N. Khanjani // *Lung India.* – 2015. – Vol. 32. – P. 557–560.

187. The study recurrent tuberculosis and exogenous reinfection, Shanghai, China / G. Shen, Z. Xue, X. Shen [et al.] // *Emerg Infect Dis.* – 2006. – Vol. 12. – P. 1776–1778.

188. Towards probabilistic decision support in public health practice: Predicting recent transmission of tuberculosis from patient attributes / H. Mamiya, K. Schwartzman, A. Verma [et al.] // *J. Biomed. Inform.* – 2015. – Vol. 53. – P. 237–242.

189. Treatment outcome of multi-drug resistant tuberculosis in the United Kingdom: retrospective-prospective cohort study from 2004 to 2007 / L. F. Anderson, S. Tamne, J. P. Watson [et al.] // *Euro Surveill.* – 2013. – Vol. 18 (40). – P. 20601. DOI: 10.2807/1560-7917.es2013.18.40.20601.

190. Treatment quality and outcome for multidrug-resistant tuberculosis patients in four regions of China: a cohort study / X. B. Zheng, V. K. Diwan, Q. Zhao [et al.] // *Infect Dis Poverty.* – 2020. – Vol. 9 (1). – P. 97. DOI: 10.1186/s40249-020-00719-x.

191. Tuberculosis and HIV co-infection-focus on the Asia-Pacific region / Q. M. Trinh, H. L. Nguyen, V. N. Nguyen [et al.] // *Int J Infect Dis.* – 2015. – Vol. 32. – P. 170–178. DOI: 10.1016/j.ijid.2014.11.023.

192. Tuberculosis in Ukraine: Analytical and Statistical Reference Book. The

State Institution «Public Health Center of the Ministry of Health of Ukraine» The State Institution «Center for Medical Statistics of the Ministry of Health of Ukraine». – Kyiv, 2017. – 215 p.

193. Tuberculosis recurrence after completion treatment in a European city: reinfection or relapse? / J. P. Millet, E. Shaw, A. Orcau [et al.] // PLOS ONE. – 2013. – Vol. 8. – P. e64898. DOI: 10.1371/journal.pone.0064898.

194. Tuberculosis recurrence in a high incidence setting for HIV and tuberculosis in Brazil / G. Unis, A. W. Ribeiro, L. S. Esteves [et al.] // BMC Infect Dis. – 2014. – Vol. 14. – P. 548. PMID: 25338623.

195. Tuberculosis recurrences and predictive factors in a vulnerable population in Catalonia / S. Brugueras, V. I. Molina, X. Casas [et al.] // PLoS One. – 2020. – Vol. 15 (1). – P. :e0227291. DOI: 10.1371/journal.pone.0227291.

196. Tuberculosis relapse in Vietnam is significantly associated with Mycobacterium tuberculosis Beijing genotype infections / M. N. Huyen, T. N. Buu, E. Tiemersma [et al.] // J Infect Dis. – 2013. – Vol. 207. – P. 1516–1524.

197. Tuberculosis relapse is more common than reinfection in Beijing, China / Y. Liu, X. X. Zhang, J. J. Yu [et al.] // Infectious Diseases. – 2020. – Vol. 52 (12). – P. 858–865. DOI: 10.1080/23744235.2020.1794027.

198. Tuberculosis surveillance and monitoring in Europe 2021 : [сайт WHO] – URL: <https://www.euro.who.int/en/publications/abstracts/tuberculosis-surveillance-and-monitoring-in-europe-2021-2019-data>. – Текст : электронный.

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

200. Unbiased Identification of Blood-based Biomarkers for Pulmonary Tuberculosis by Modeling and Mining Molecular Interaction Networks / A. Sambarey, A. Devaprasad, A. Mohan [et al.] // EBioMedicine. – 2017. – Vol. 15. – P. 112–126.

201. Vieira, A. A. Tuberculosis recurrence in a priority city in the state of São Paulo, Brazil / A. A. Vieira, D. T. Leite, S. Adreoni // J Bras Pneumol. – 2017. – Vol. 43. – P. 106–112.

202. When tuberculosis comes back: who develops recurrent tuberculosis in

Californian? / L Pascopella, K. Deriemer, J. P. Watt, J. M. Flood // PLOS ONE. – 2011. – Vol. 6. – P. e26541.

203. World Tuberculosis Day 2021. THE CLOCK IS TICKING : [сайт WHO]. – URL: <https://www.who.int/campaigns/world-tb-day/world-tb-day-2021>. – Текст : электронный.

204. XGBoost Improves Classification of MGMT Promoter Methylation Status in IDH1 Wildtype Glioblastoma / N. Q. K. Le, D. T. Do, F. Y. Chiu [et al.] // J Pers Med. – 2020. – Vol. 10 (3). – P. 128. DOI: 10.3390/jpm10030128.

205. XGBoost, a machine learning method, predicts neurological recovery in patients with cervical spinal cord injury / T. Inoue, D. Ichikawa, T. Ueno [et al.] // Neurotrauma Reports. – 2020. – Vol. 1 (1). – P. 8–16. DOI: 10.1089/neur.2020.0009.

206. XG-SF: An XGBoost Classifier Based on Shapelet Features for Time Series Classification / C. Ji, X. Zou, Y. Hu [et al.] // Procedia Computer Science. – 2019. – Vol. 147. – P. 24–28. DOI: 10.1016/j.procs.2019.01.179.

207. Yahiaoui, A. A new method of automatic recognition for tuberculosis disease diagnosis using support vector machines / A. Yahiaoui, O. Er, N. Yumusak // Biomed. Res. – 2017. – Vol. 28. – P. 4208–4212.

208. Yim, J. J. MDR-TB recurrence after successful treatment: additional studies using molecular genotyping are needed / J. J. Yim, W. J. Koh // Int J Tuberc Lung Dis. – 2015. – Vol. 19 (4). – P. 371. DOI: 10.5588/ijtld.15.0117.

209. Zanyaty, E. Support vector machines (SVMs) versus multilayer perception (MLP) in data classification / E. Zanyaty // Egypt. Inform. J. – 2012. – Vol. 13. – P. 177–183.

210. Zhou, Z. H. Ensemble Methods: Foundations and Algorithms / Z. H. Zhou. – CRC Press : Boca Raton, FL, USA, 2012.

СПИСОК ИЛЛЮСТРАТИВНОГО МАТЕРИАЛА

1. Рисунок 1 – Классификация методов машинного обучения. С. 40
2. Рисунок 2 – Схема исследования ретроспективного когортного анализа данных больных МЛУ ТБ после эффективного курса химиотерапии. С. 52
3. Рисунок 3 – ROC-кривая, построенная на предсказанных значениях вероятности развития рецидива МЛУ ТБ на основе регрессионного анализа. С. 90
4. Рисунок 4 – Дерево решений прогнозирования рецидива МЛУ ТБ. . С. 98
5. Рисунок 5 – Усредненная значимость признаков алгоритма классификации пациентов, построенного методом дерево решений. С. 101
6. Рисунок 6 – Усредненная значимость признаков при использовании прогностической модели классификации рецидива ТБ, полученной методом случайный лес. С. 101
7. Рисунок 7 – Усредненная значимость признаков для прогностической модели классификации рецидива ТБ, построенной методом XGBoost. С. 102
8. Рисунок 8 – Усредненная значимость признаков для алгоритма классификации рецидива ТБ, полученного методом логистической регрессии. С. 102
9. Рисунок 9 – ROC кривые для моделей классификаторов, определяющих признаки, сопряженные с рецидивом МЛУ ТБ органов дыхания. С. 104
10. Рисунок 10 – Диаграмма предикторов и вероятность возникновения рецидива больного № 11. С. 106
11. Рисунок 11 – Диаграмма конкурирующих предикторов возникновения рецидива больного № 11. С. 107
12. Рисунок 12 – Диаграмма предикторов и вероятность возникновения рецидива больного № 56. С. 108

13. Рисунок 13 – Диаграмма конкурирующих предикторов возникновения рецидива больного № 56. С. 108
14. Таблица 1 – Частота рецидивов у больных МЛУ ТБ: анализ публикаций за 2011–2020 гг. С. 20
15. Таблица 2 – Факторы риска развития рецидива туберкулеза. С. 22
16. Таблица 3 – Обзор литературных источников по применению алгоритмов машинного обучения во фтизиатрии. С. 47
17. Таблица 4 – Анализ когорты зарегистрированных больных МЛУ ТБ после ЭКХТ. С. 58
18. Таблица 5 – Характеристика больных МЛУ ТБ сравниваемых групп. С. 60
19. Таблица 6 – Сопутствующая патология у больных МЛУ ТБ сравниваемых групп. С. 61
20. Таблица 7 – Показатели общего анализа крови до лечения больных МЛУ ТБ. С. 63
21. Таблица 8 – Показатели биохимического анализа крови до лечения больных МЛУ ТБ. С. 63
22. Таблица 9 – Характеристика больных в группах наблюдения по анамнезу заболевания. С. 65
23. Таблица 10 – Общее число курсов лечения больных МЛУ ТБ в группах наблюдения. С. 66
24. Таблица 11 – Количество курсов химиотерапии по IV режиму больных МЛУ ТБ в группах наблюдения. С. 67
25. Таблица 12 – Клинические формы туберкулеза у больных МЛУ ТБ сравниваемых групп. С. 67
26. Таблица 13 – Размеры полостей распада у больных МЛУ ТБ в начале лечения. С. 69
27. Таблица 14 – Наличие остаточных изменений в легких после эффективного курса химиотерапии больных МЛУ ТБ сравниваемых групп. С. 70

28. Таблица 15 – Характеристика малых остаточных изменений в легких после эффективного курса химиотерапии больных МЛУ ТБ сравнимых групп. С. 71
29. Таблица 16 – Характеристика больших остаточных изменений в легких после эффективного курса химиотерапии больных МЛУ ТБ сравнимых групп. С. 71
30. Таблица 17 – Результаты исследования мокроты методом микроскопии больных МЛУ ТБ на начало лечения. С. 72
31. Таблица 18 – Результаты исследования мокроты больных МЛУ ТБ на начало лечения методом посева. С. 72
32. Таблица 19 – Стаж заболевания, сроки назначения химиотерапии после обнаружения МЛУ МБТ у больных сравнимых групп. . . . С. 73
33. Таблица 20 – Лекарственная устойчивость МБТ к ПТП у больных МЛУ ТБ сравнимых групп. С. 74
34. Таблица 21 – Амплификация МБТ к ПТП во время лечения у больных МЛУ ТБ сравнимых групп. С. 75
35. Таблица 22 – Число препаратов, к которым произошла амплификация во время лечения у больных МЛУ ТБ сравнимых групп. С. 76
36. Таблица 23 – Частота первоначального использования противотуберкулезных препаратов (ПТП) в режимах химиотерапии больных МЛУ ТБ (n = 346) С. 77
37. Таблица 24 – Основные первоначальные схемы лечения больных МЛУ ТБ сравнимых групп. С. 77
38. Таблица 25 – Наличие фторхинолонов и инъекционных препаратов в первоначальной схеме лечения больных МЛУ ТБ сравнимых групп. С. 78
39. Таблица 26 – Количество ПТП в схемах лечения больных МЛУ ТБ сравнимых групп. С. 78
40. Таблица 27 – Схема химиотерапии на начало лечения у больных

	МЛУ ТБ сравниваемых групп.	С. 79
41.	Таблица 28 – Количество препаратов, к которым сохранялась лекарственная чувствительность МБТ в последнем положительном посеве мокроты больных МЛУ ТБ сравниваемых групп.	С. 80
42.	Таблица 29 – Схема химиотерапии в конце лечения у больных МЛУ ТБ сравниваемых групп.	С. 81
43.	Таблица 30 – Схемы химиотерапии в конце курса лечения больных МЛУ ТБ сравниваемых групп.	С. 82
44.	Таблица 31 – Время прекращения бактериовыделения по микроскопии мазка мокроты и посева на плотные питательные среды (мес.) и средний срок лечения больных МЛУ ТБ.	С. 83
45.	Таблица 32 – Частота прерывания лечения/нарушение режима у больных МЛУ ТБ сравниваемых групп.	С. 84
46.	Таблица 33 – Доля принятых доз ПТП от должного у больных МЛУ ТБ сравниваемых групп.	С. 84
47.	Таблица 34 – Параметры многофакторного анализа признаков больных МЛУ ТБ.	С. 86
48.	Таблица 35 – Параметры классификации регрессионного анализа.	С. 89
49.	Таблица 36 – Оценка качества построенной модели на основе ROC-анализа.	С. 90
50.	Таблица 37 – Гиперпараметры классификаторов, выбранные с помощью сеточного поиска.	С. 96
51.	Таблица 38 – Метрики качества работы классификаторов, построенных с помощью K-блочной стратифицированной проверки в прогнозировании рецидива МЛУ ТБ.	С. 105

ПРИЛОЖЕНИЕ А

(справочное)

Иллюстративный материал этапов работы классификатора на основе
метода дерево решений

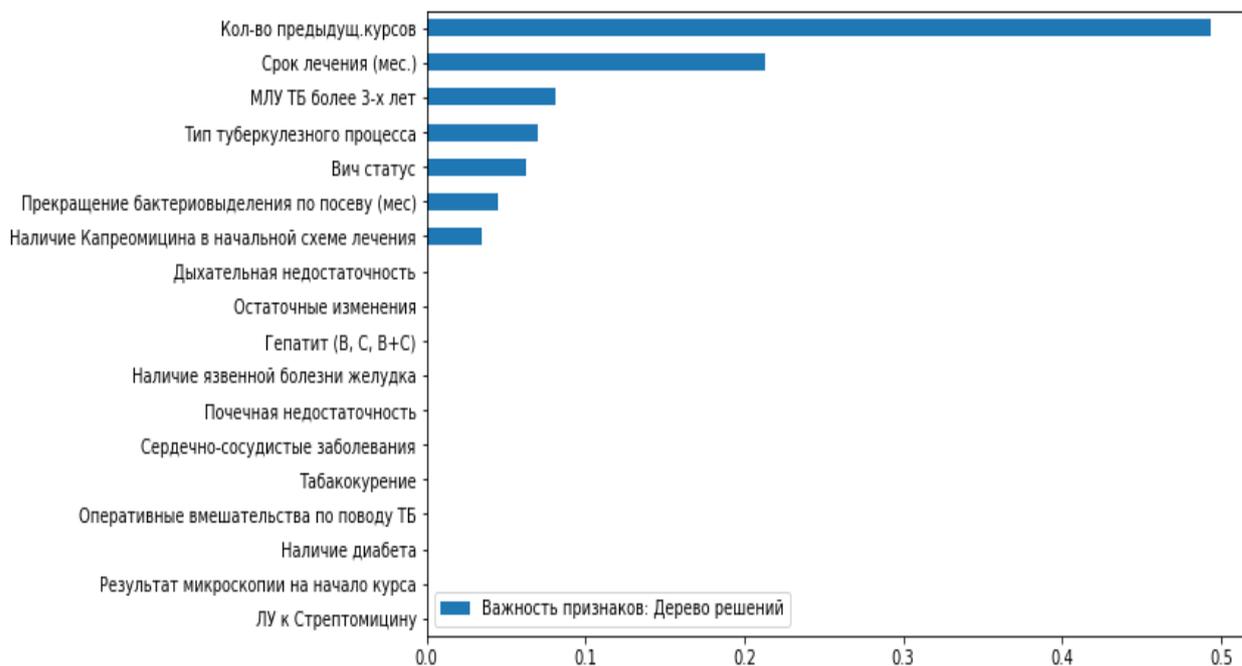


Рисунок 1 – Важность признаков при использовании метода дерево решений
(блок 1 перекрестной проверки)

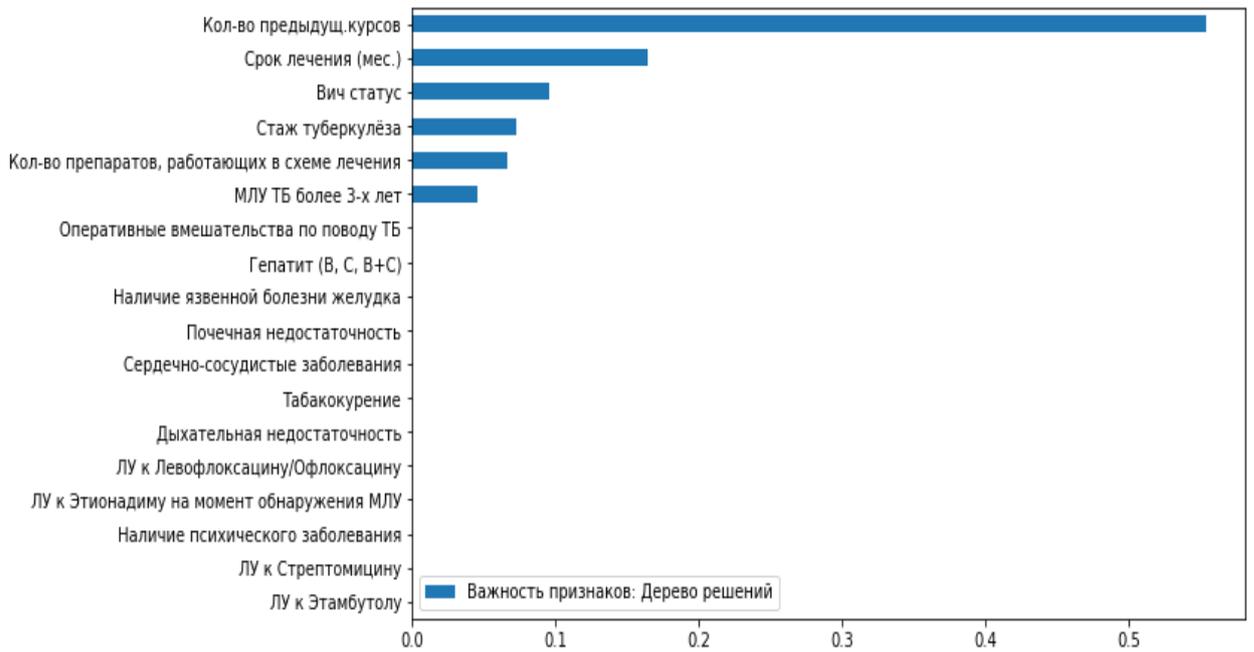


Рисунок 2 – Важность признаков при использовании метода дерева решений (блок 2 перекрестной проверки)

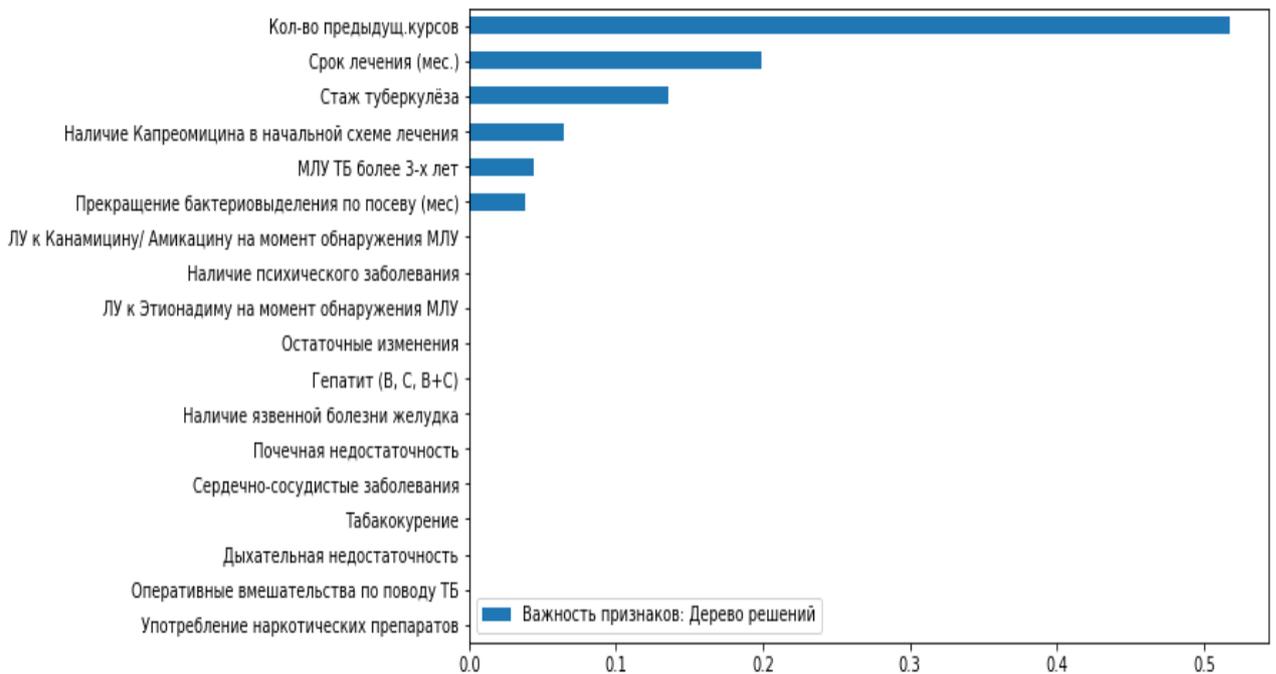


Рисунок 3 – Важность признаков при использовании метода дерева решений (блок 3 перекрестной проверки)

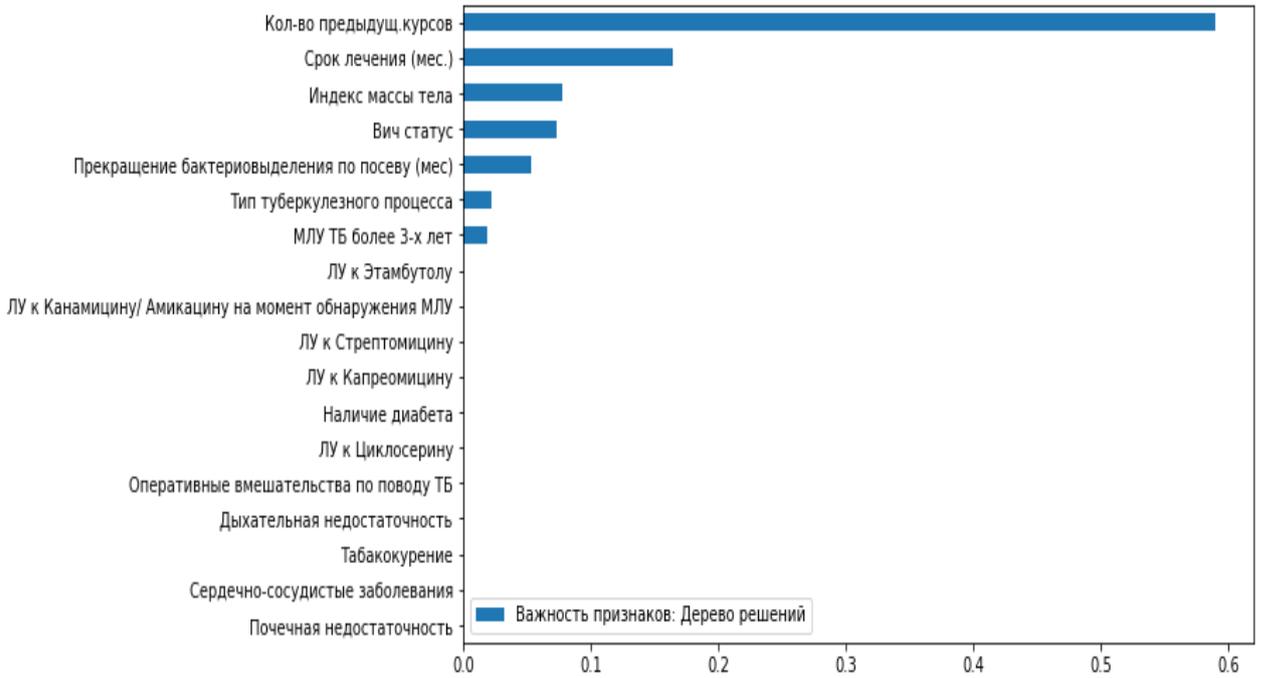


Рисунок 4 – Важность признаков при использовании метода дерево решений (блок 4 перекрестной проверки)

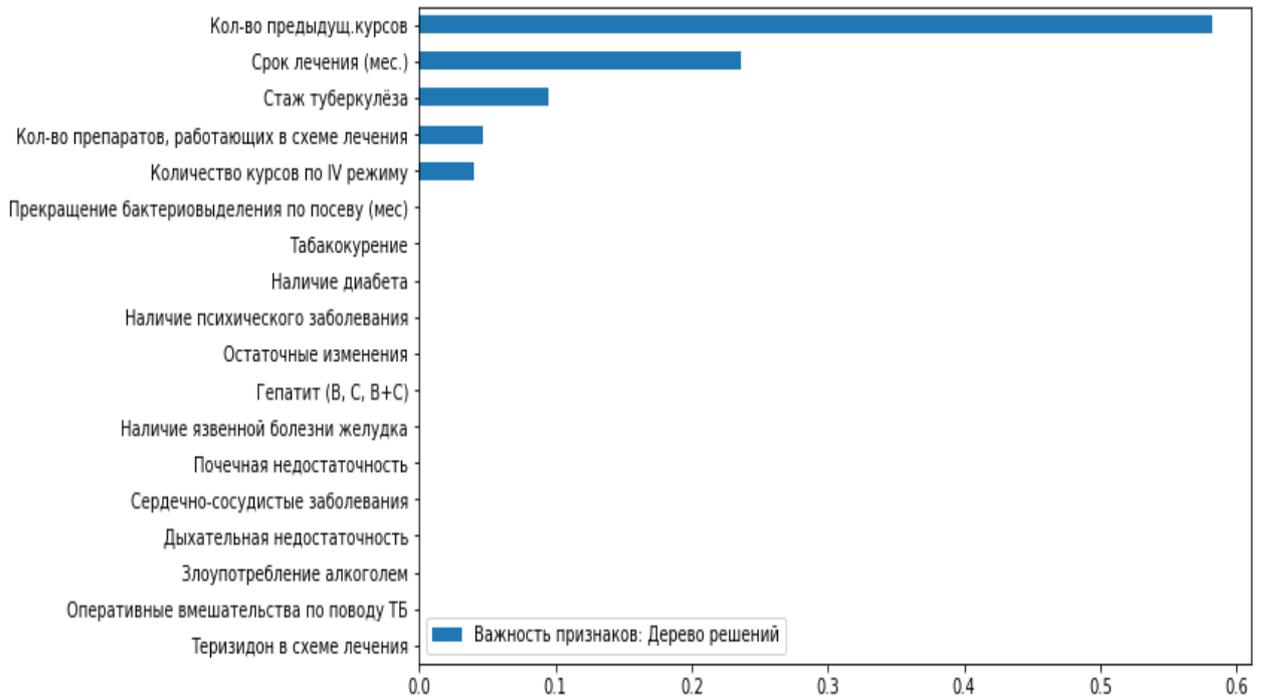


Рисунок 5 – Важность признаков при использовании метода дерево решений (блок 5 перекрестной проверки)

Важность признаков, усредненная по 5 блокам перекрестной проверки для алгоритма дерева решений

Важность признаков: Дерево решений	0.547409
Кол-во предыдущ. курсов	0.195558
Срок лечения (мес.)	0.060629
Стаж туберкулёза	0.046401
Вич статус	0.038145
МЛУ ТБ более 3-х лет	0.027215
Прекращение бактериовыделения по посеву (мес.)	0.022814
Кол-во препаратов, работающих в схеме лечения	0.019856
Наличие Капреомицина в начальной схеме лечения	0.018360
Тип туберкулезного процесса	0.015522
Индекс массы тела	0.008093
Количество курсов по IV режиму	0.000000
Контакт с пациентом с МЛУ ТБ	0.000000
Работающий пациент	0.000000
Начилие Моксифлоксацина в схеме лечения	0.000000
Нарушение режима	0.000000
Наличие Этамбутола в схеме лечения	0.000000
Наличие Циклосерина в схеме лечения	0.000000
Наличие Капреомицина в схеме лечения	0.000000
ЛУ к ПАСКу в последнем посеве (амплификация)	0.000000
ЛУ к Левофлоксацину/ Офлоксацину в последнем по...	0.000000
ЛУ к Этионамиду в последнем посеве (амплификация)	0.000000
ЛУ к Циклосерину в последнем посеве (амплификация)	0.000000
ЛУ к Капреомицину в последнем посеве (амплифика...	0.000000
ЛУ к Канамицину/Амикацину	0.000000
ЛУ к Этамбутолу в последнем посеве (амплификация)	0.000000
ЛУ к Стрептомицину в последнем посеве (амплифик...	0.000000
Пол пациента	0.000000
Амплификация ЛУ МБТ	0.000000
Наличие Левофлоксацина/Офлоксацина в схеме леч...	0.000000
Наличие ПАСКа в схеме лечения	0.000000
Наличие Протионамида в схеме лечения	0.000000
Остаточные изменения после лечения	0.000000
ПТП в схеме лечения (кол-во)	0.000000
Пиразинамид в схеме лечения	0.000000
Теризидон в схеме лечения	0.000000
Этамбутол в схеме лечения	0.000000
Результат посева на начало курса	0.000000
Лекарственная устойчивость МБТ к ПТП (кол-во)	0.000000
Остаточное изменение в виде каверны	0.000000
Наличие Резонизада в схеме лечения	0.000000
Диаметр полости в легких (см.)	0.000000
Количество принятых доз (%)	0.000000
Наличие фазы распада	0.000000
Клиническая форма туберкулёза	0.000000
Код начального места лечения	0.000000
Наличие Пиразинамида в схеме лечения	0.000000
Перерывы в лечении в анамнезе	0.000000
ЛУ к ПАСКу на начало лечения	0.000000
ЛУ к Левофлоксацину/ Офлоксацину на начало лечения	0.000000
Употребление наркотических препаратов	0.000000
Почечная недостаточность	0.000000
Наличие язвенной болезни желудка	0.000000
Гепатит (В, С, В+С)	0.000000
Остаточные изменения	0.000000
Наличие психического заболевания	0.000000
Наличие диабета	0.000000
Злоупотребление алкоголем	0.000000
Табакокурение	0.000000
Хронические неспецифические заболевания легких	0.000000
Семейное положение	0.000000
Пациент поступил из мест лишения свободы	0.000000
Инвалидность	0.000000
Пенсионный возраст	0.000000
Бездомный	0.000000
Сердечно-сосудистые заболевания	0.000000

Дыхательная недостаточность	0.000000
ЛУ к Этионадиму на начало лечения	0.000000
ЛУ к Левофлоксацину/Офлоксацину	0.000000
ЛУ к Циклосерину на начало лечения	0.000000
ЛУ к Капреомицину на начало лечения	0.000000
ЛУ к Амикацину/ Канамицину на начало лечения	0.000000
Результат микроскопии на начало курса	0.000000
ЛУ к Стрептомицину на начало лечения	0.000000
ЛУ к ПАСК	0.000000
ЛУ к Этионадиму на момент обнаружения МЛУ	0.000000
Оперативные вмешательства по поводу ТБ	0.000000
ЛУ к Циклосерину	0.000000
ЛУ к Капреомицину	0.000000
ЛУ к Канамицину/ Амикацину на момент обнаружения...	0.000000
ЛУ к Этамбутолу	0.000000
ЛУ к Стрептомицину	0.000000
Градация возраста	0.000000
ЛУ к Этамбутола на начало лечения	0.000000

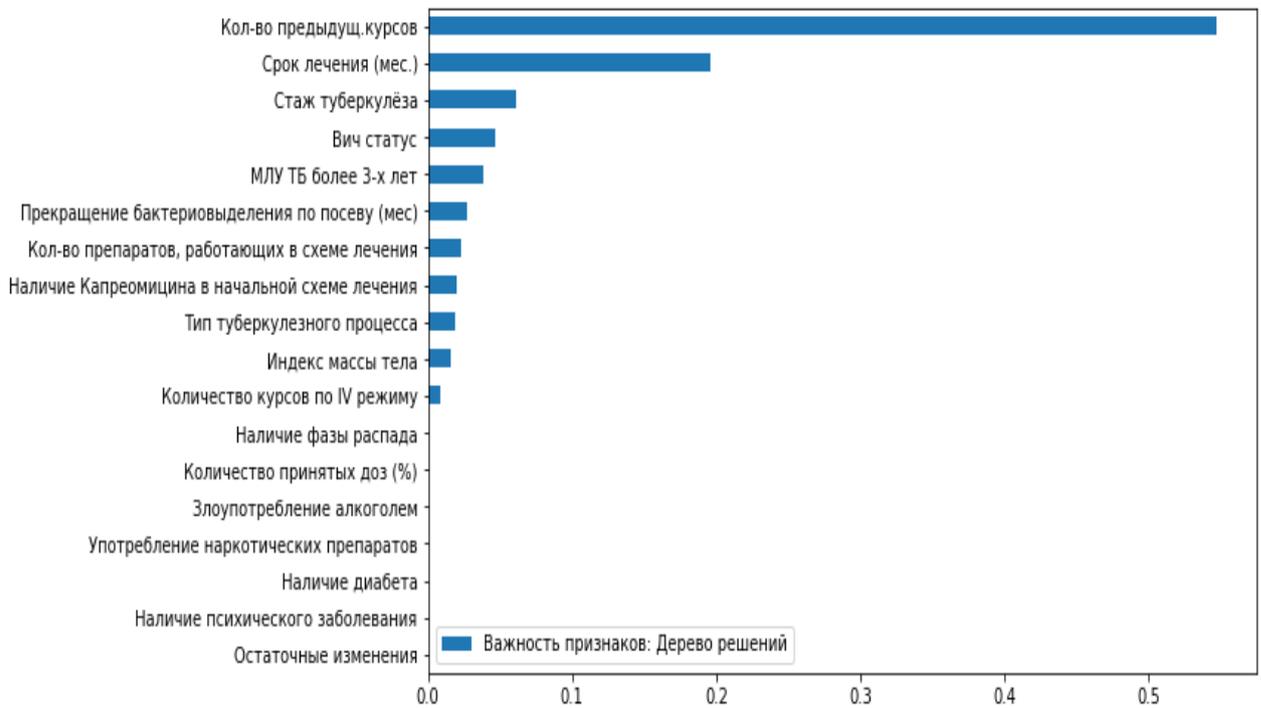


Рисунок 6 – Усредненная важность признаков по 5 блокам проверки при использовании метода дерева решений

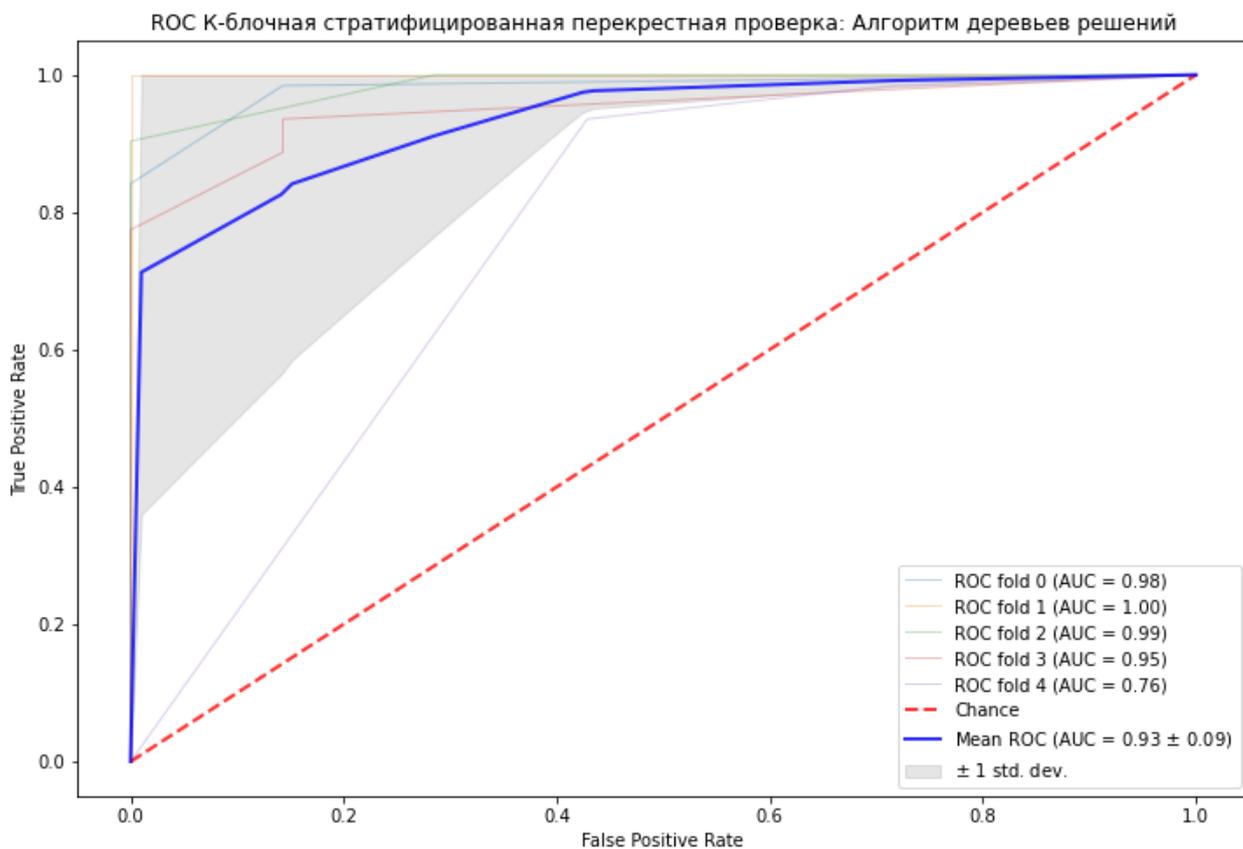


Рисунок 7 – ROC-анализ кривых, построенных для каждого из 5 блоков признаков перекрестной проверки, а также их усредненной (для алгоритма дерево решений)

ПРИЛОЖЕНИЕ Б

(справочное)

Иллюстративный материал этапов работы классификатора на основе
метода случайный лес

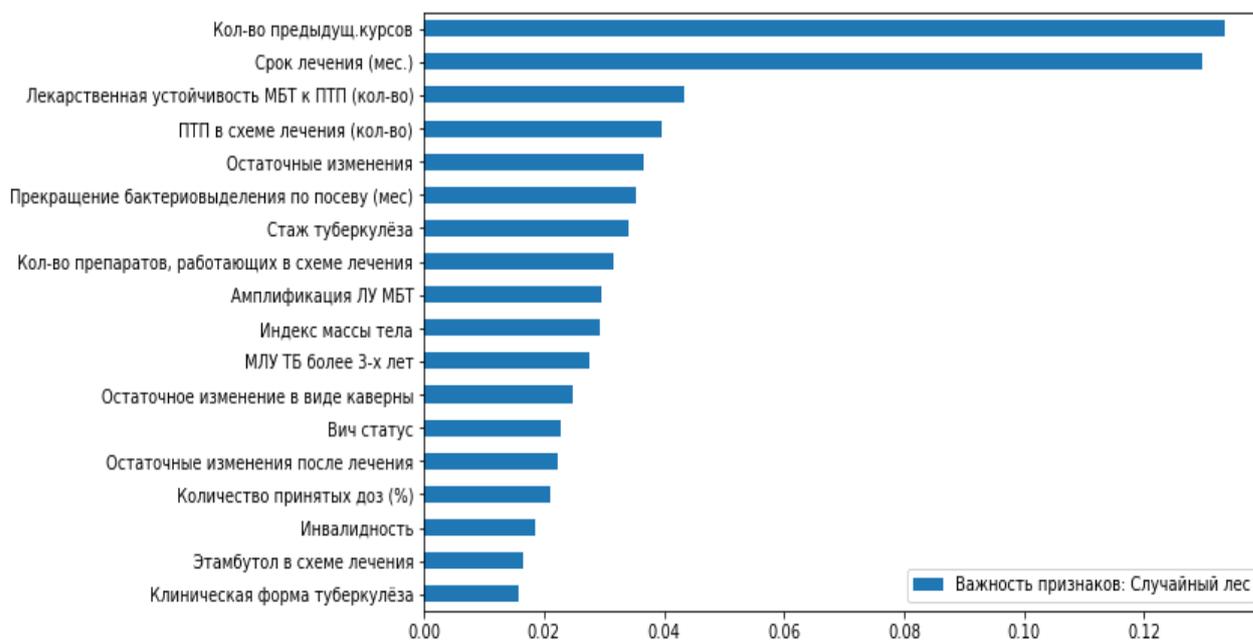


Рисунок 1 – Важность признаков, при использовании метода случайный лес
(блок 1 перекрестной проверки)

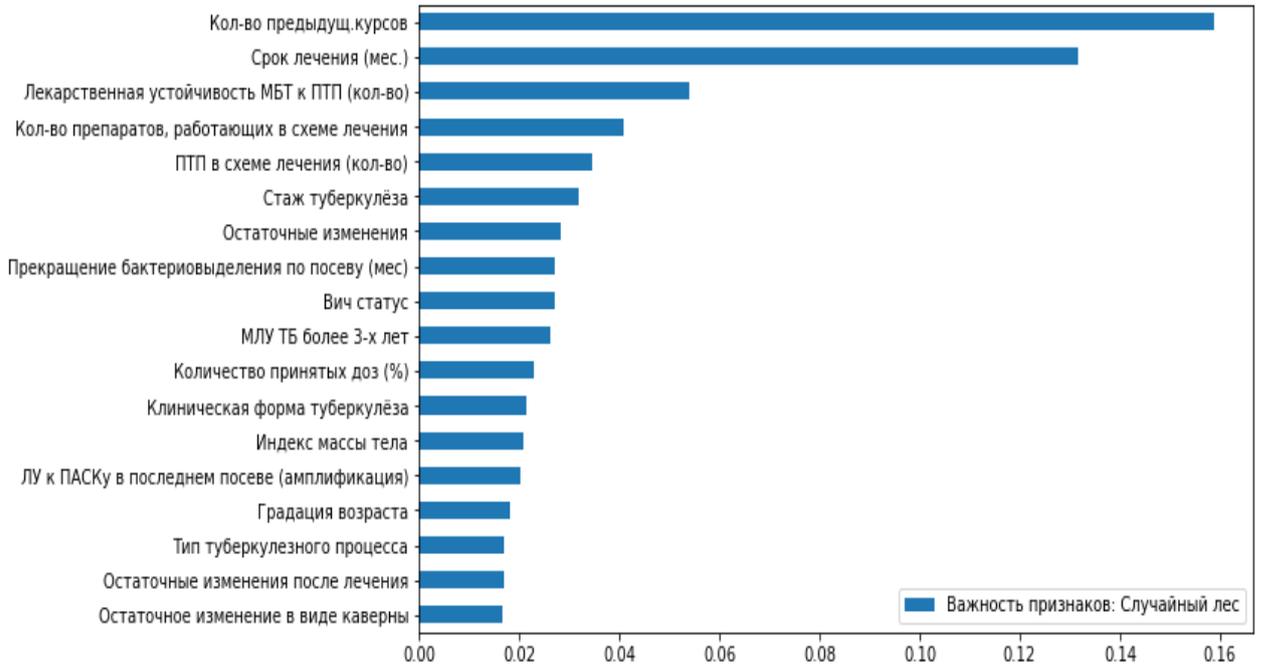


Рисунок 2 – Важность признаков, при использовании метода случайный лес (блок 2 перекрестной проверки)

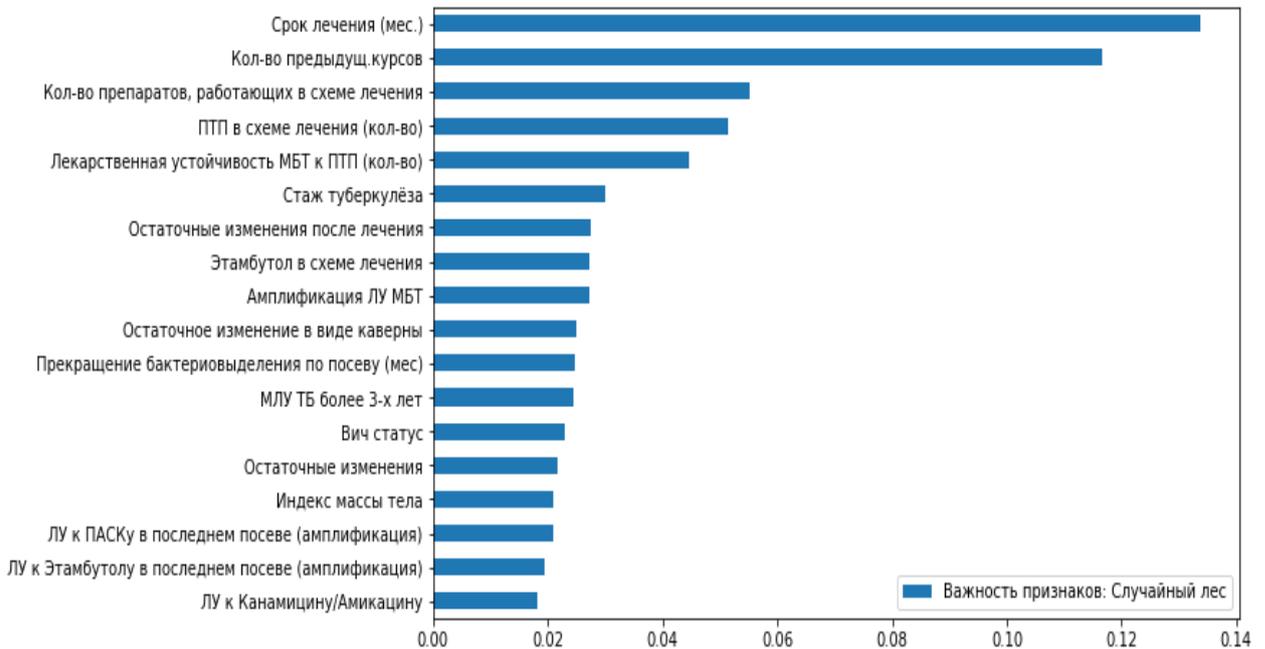


Рисунок 3 – Важность признаков, при использовании метода случайный лес (блок 3 перекрестной проверки)

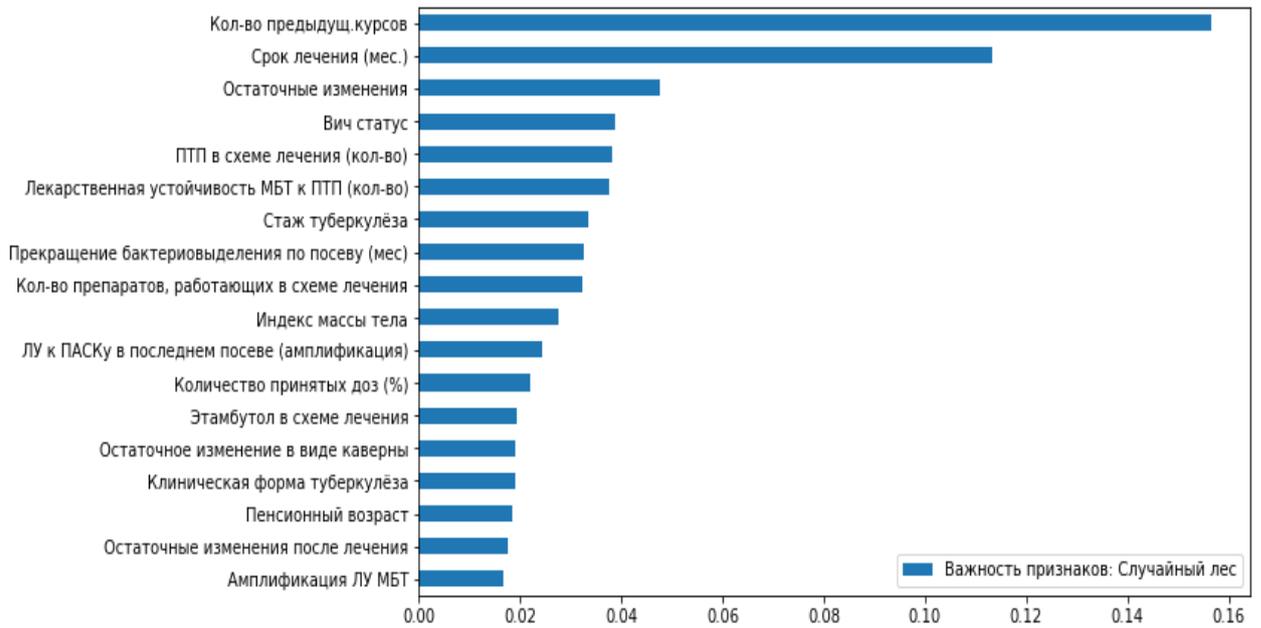


Рисунок 4 – Важность признаков, при использовании метода случайный лес (блок 4 перекрестной проверки)

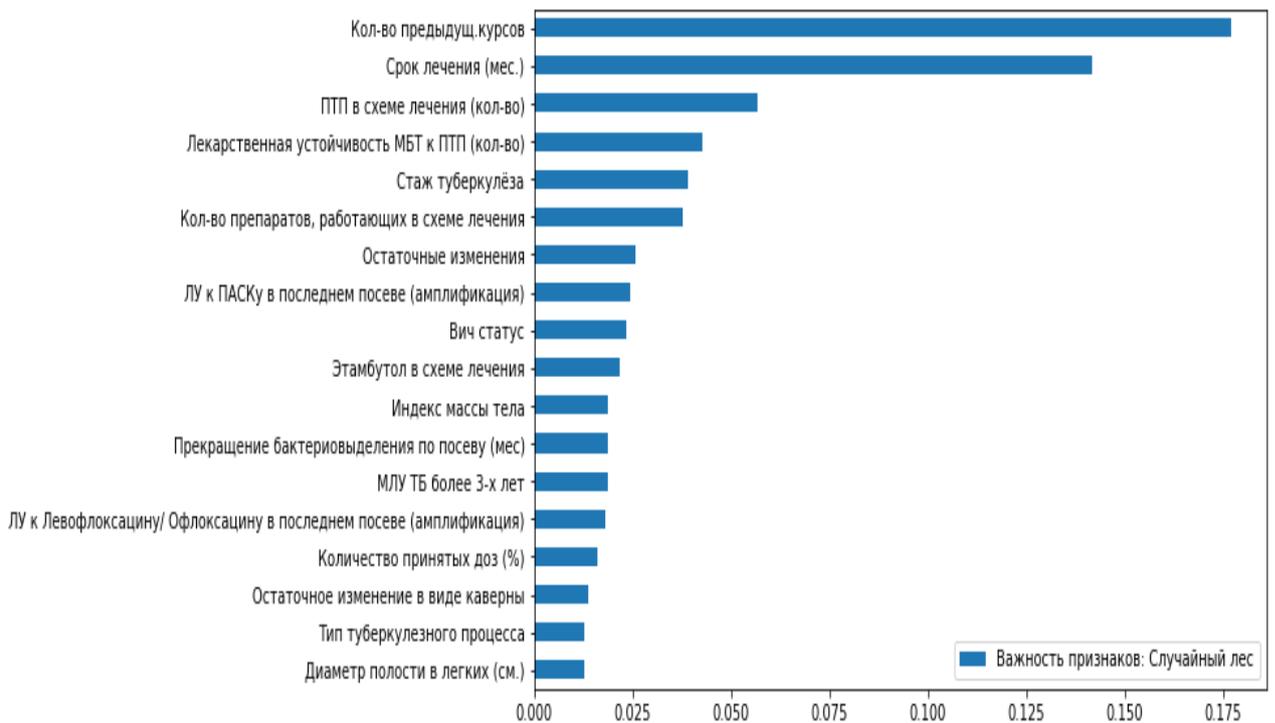


Рисунок 5 – Важность признаков, при использовании метода случайный лес (блок 5 перекрестной проверки)

Важность признаков, усредненная по 5 блокам перекрестной проверки для алгоритма случайного леса

Важность признаков: Случайный лес	0.148425
Кол-во предыдущ. курсов	0.130025
Срок лечения (мес.)	0.044512
Лекарственная устойчивость МБТ к ПТП (кол-во)	0.044085
ПТП в схеме лечения (кол-во)	0.039526
Кол-во препаратов, работающих в схеме лечения	0.033670
Стаж туберкулёза	0.031963
Остаточные изменения	0.027721
Прекращение бактериовыделения по посеву (мес.)	0.027046
Вич статус	0.023531
Индекс массы тела	0.022687
МЛУ ТВ более 3-х лет	0.020936
ЛУ к ПАСКу в последнем посеве (амплификация)	0.019865
Остаточное изменение в виде каверны	0.019810
Этамбутол в схеме лечения	0.019505
Количество принятых доз (%)	0.019309
Остаточные изменения после лечения	0.018559
Амплификация ЛУ МБТ	0.015468
Клиническая форма туберкулёза	0.015040
Тип туберкулезного процесса	0.013617
ЛУ к Этамбутолу в последнем посеве (амплификация)	0.012868
Диаметр полости в легких (см.)	0.012537
Градация возраста	0.012127
ЛУ к Левофлоксацину/ Офлоксацину в последнем по...	0.010910
ЛУ к Канамицину/Амикацину	0.010452
Пенсионный возраст	0.010422
Результат микроскопии на начало курса	0.010354
Инвалидность	0.008915
Результат посева на начало курса	0.008535
Код начального места лечения	0.008465
Наличие психического заболевания	0.008002
Гепатит (В, С, В+С)	0.006172
ЛУ к Капреомицину в последнем посеве (амплифика...	0.006048
ЛУ к Этионамиду в последнем посеве (амплификация)	0.005765
Количество курсов по IV режиму	0.005032
Наличие Капреомицина в начальной схеме лечения	0.005020
Наличие фазы распада	0.004777
Нарушение режима	0.004762
Работающий пациент	0.004663
Наличие ПАСКа в схеме лечения	0.004542
Пациент поступил из мест лишения свободы	0.004480
Бездомный	0.004479
Наличие Капреомицина в схеме лечения	0.004456
ЛУ к Этионамиду на начало лечения	0.004411
Злоупотребление алкоголем	0.004314
Пол пациента	0.004137
Семейное положение	0.004064
Контакт с пациентом с МЛУ ТВ	0.004003
Перерывы в лечении в анамнезе	0.003779
Наличие язвенной болезни желудка	0.003738
Употребление наркотических препаратов	0.003637
ЛУ к Этамбутолу	0.003541
Наличие Этамбутола в схеме лечения	0.003353
ЛУ к Этамбутола на начало лечения	0.003324
ЛУ к Этионамиду на момент обнаружения МЛУ	0.003267
ЛУ к Капреомицину на начало лечения	0.003146
Хронические неспецифические заболевания легких	0.003018
Наличие Пиразинамида в схеме лечения	0.002946
Наличие Протионамида в схеме лечения	0.002503
Наличие Резонизада в схеме лечения	0.002439
ЛУ к ПАСК	0.002390
ЛУ к ПАСКу на начало лечения	0.002286
Табакокурение	0.002242
ЛУ к Амикацину/ Канамицину на начало лечения	0.002096
ЛУ к Левофлоксацину/ Офлоксацину на начало лечения	0.002066
ЛУ к Левофлоксацину/Офлоксацину	

Наличие Циклосерина в схеме лечения	0.002044
Теризидон в схеме лечения	0.001949
Наличие Левофлоксацина/ Офлоксацина в схеме леч...	0.001918
Почечная недостаточность	0.001898
ЛУ к Капреомицину	0.001877
ЛУ к Циклосерину в последнем посеве (амплификация)	0.001765
ЛУ к Канамицину/ Амикацину на момент обнаружени...	0.001713
Наличие диабета	0.001553
ЛУ к Циклосерину	0.001339
Дыхательная недостаточность	0.001278
Начилие Моксифлоксацина в схеме лечения	0.000980
Оперативные вмешательства по поводу ТБ	0.000959
ЛУ к Циклосерину на начало лечения	0.000404
ЛУ к Стрептомицину	0.000179
ЛУ к Стрептомицину на начало лечения	0.000174
Сердечно-сосудистые заболевания	0.000118
ЛУ к Стрептомицину в последнем посеве (амплифик...	0.000067
Пиразинамид в схеме лечения	0.000000

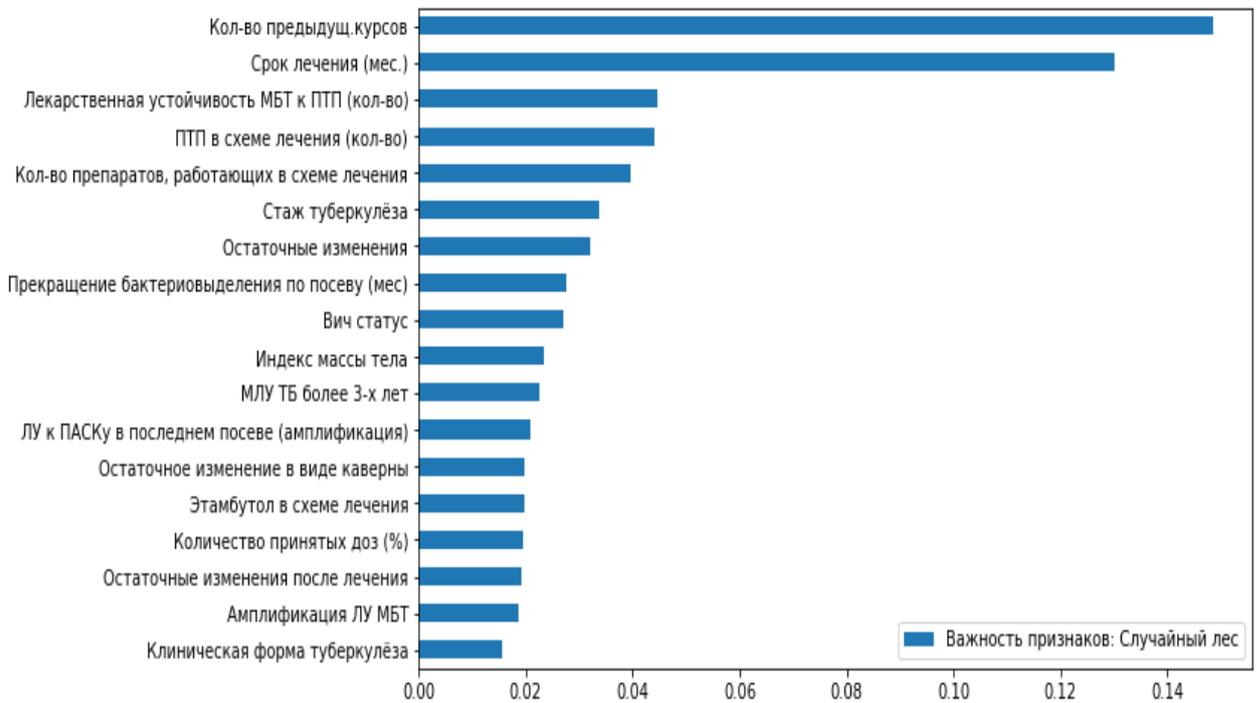


Рисунок 6 – Усредненная важность признаков по 5 блокам проверки, при использовании метода случайный лес

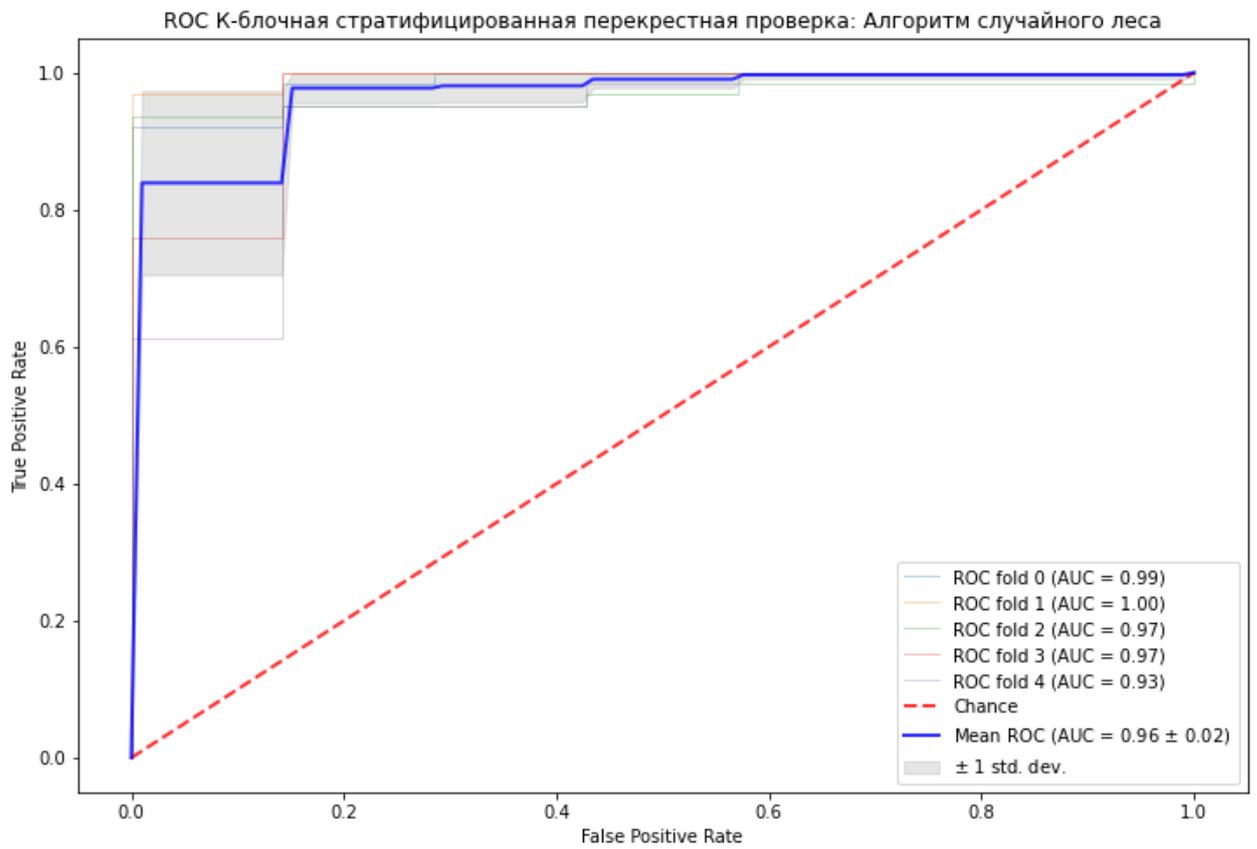


Рисунок 7 – ROC-анализ кривых, построенных для каждого из 5 блоков признаков перекрестной проверки, а также их усредненной (для алгоритма случайный лес)

ПРИЛОЖЕНИЕ В

(справочное)

Иллюстративный материал этапов работы классификатора на основе
метода XGBoost

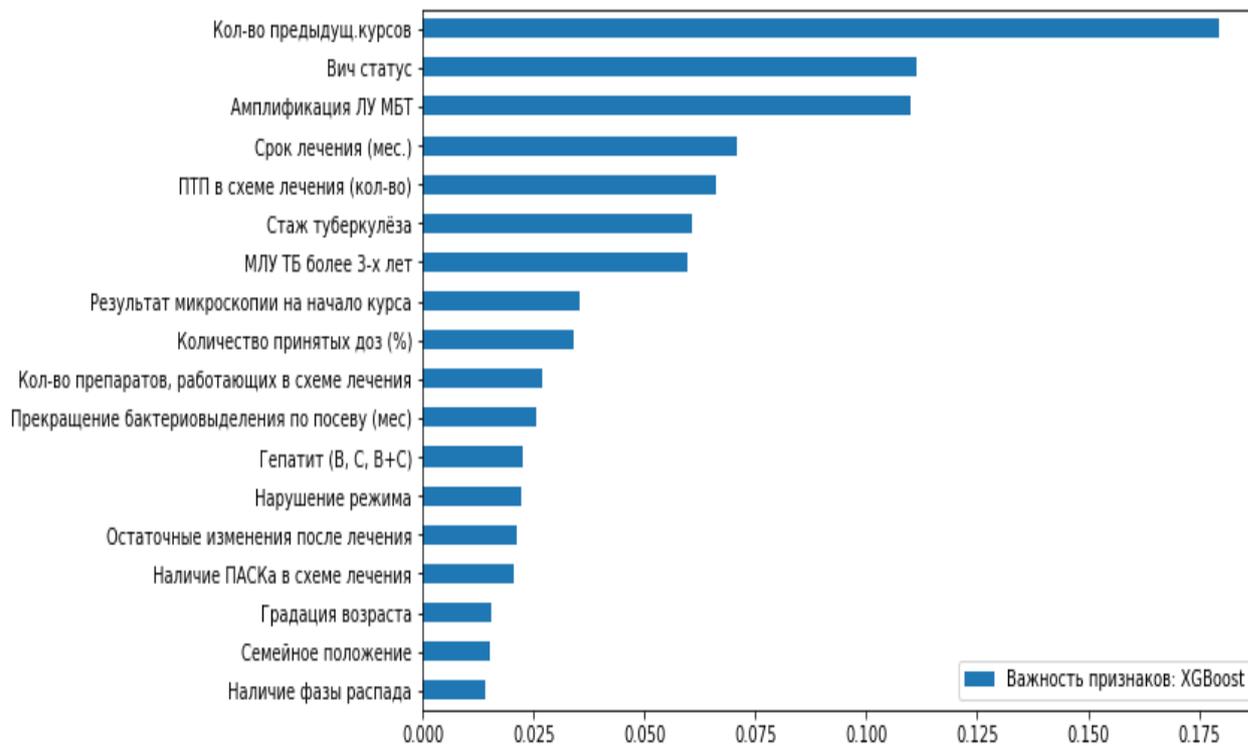


Рисунок 1 – Важность признаков, при использовании метода XGBoost
(блок 1 перекрестной проверки)

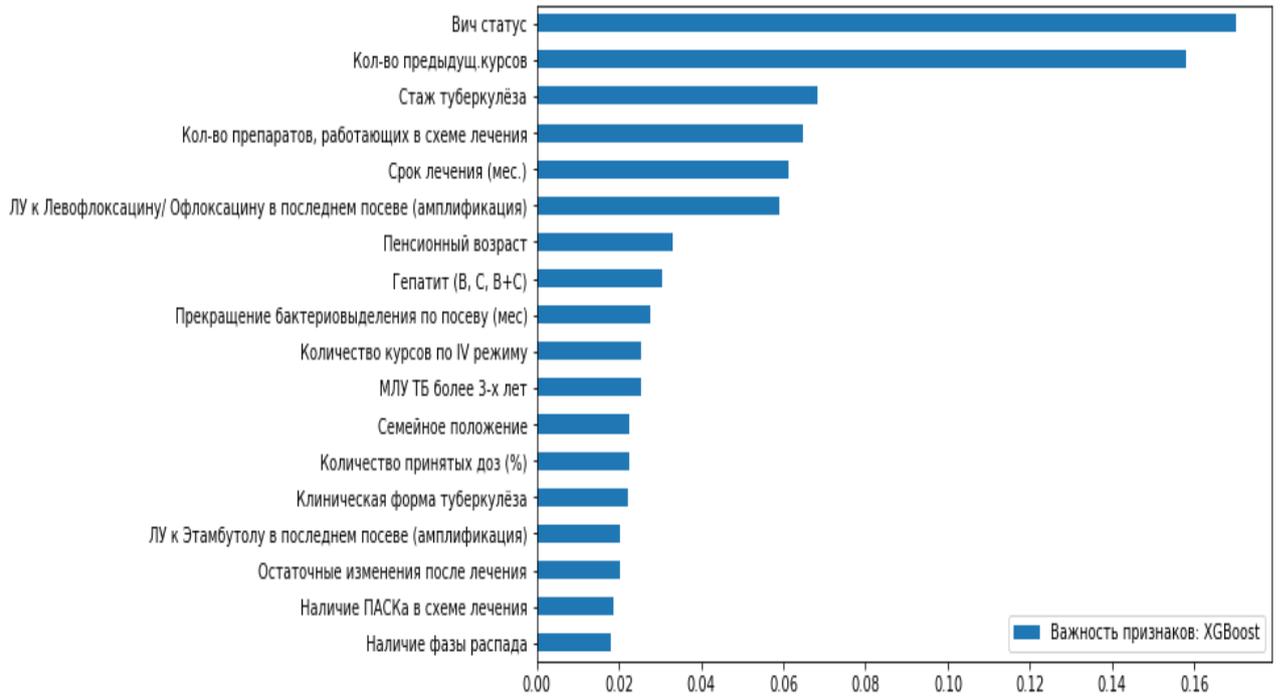


Рисунок 2 – Важность признаков, при использовании метода XGBoost
(блок 2 перекрестной проверки)

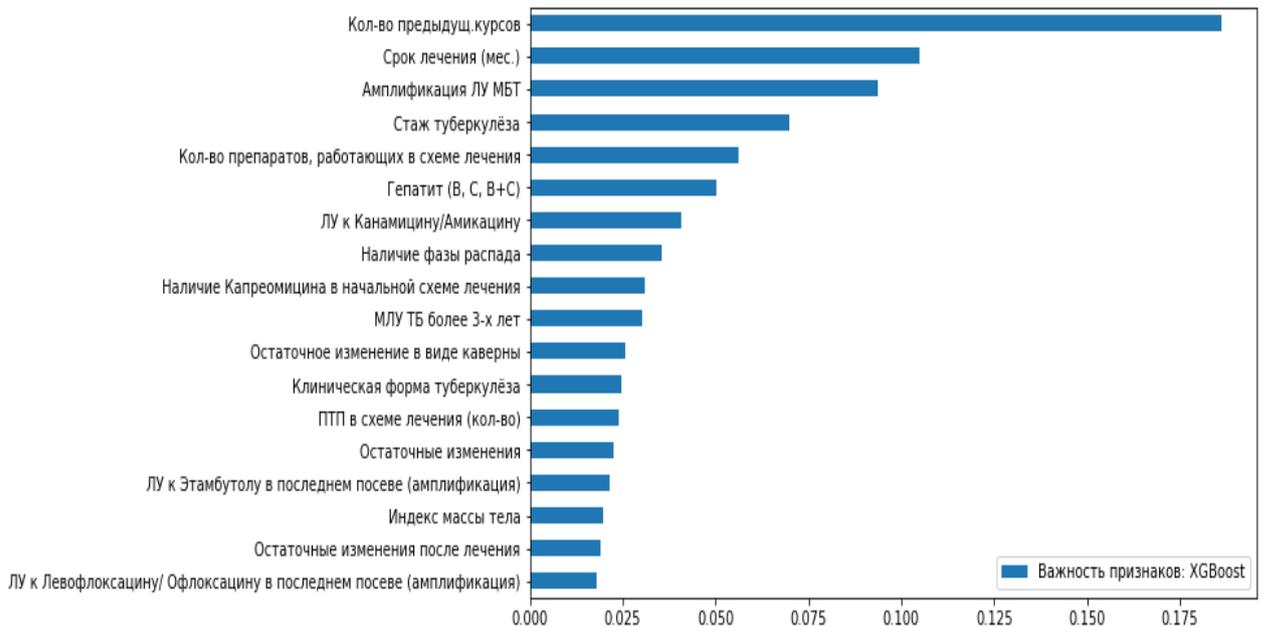


Рисунок 3 – Важность признаков, при использовании метода XGBoost
(блок 3 перекрестной проверки)

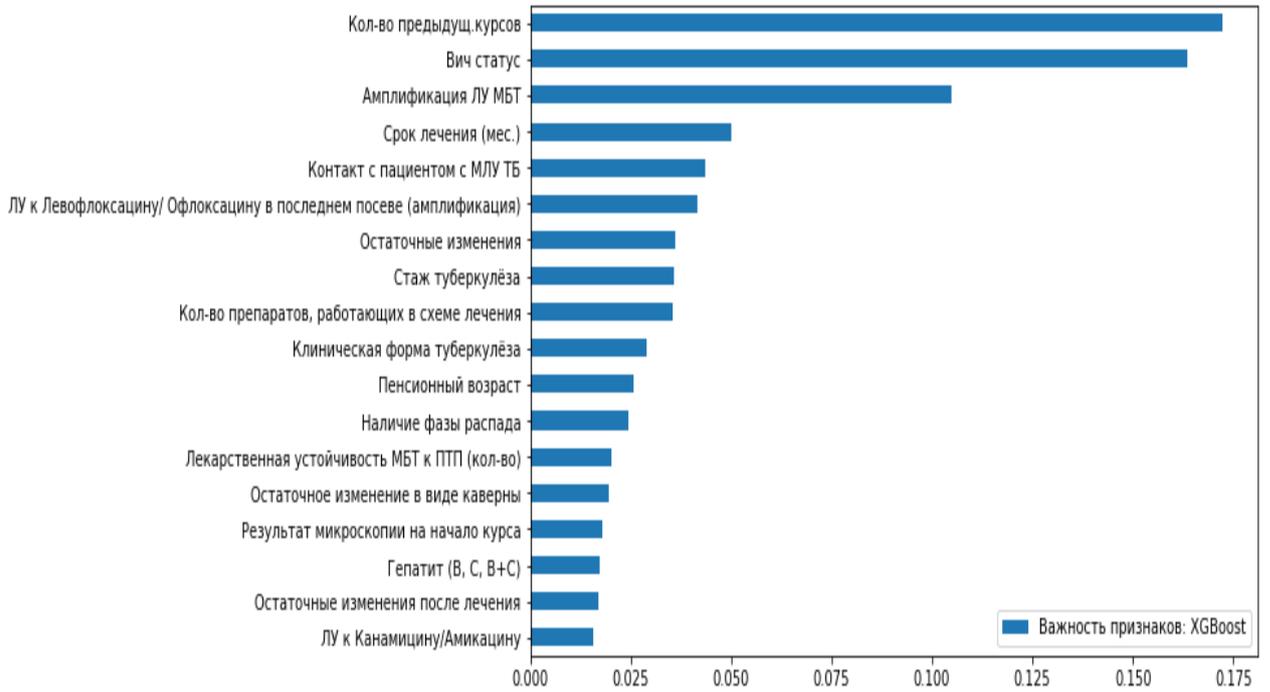


Рисунок 4 – Важность признаков, при использовании метода XGBoost (блок 4 перекрестной проверки)

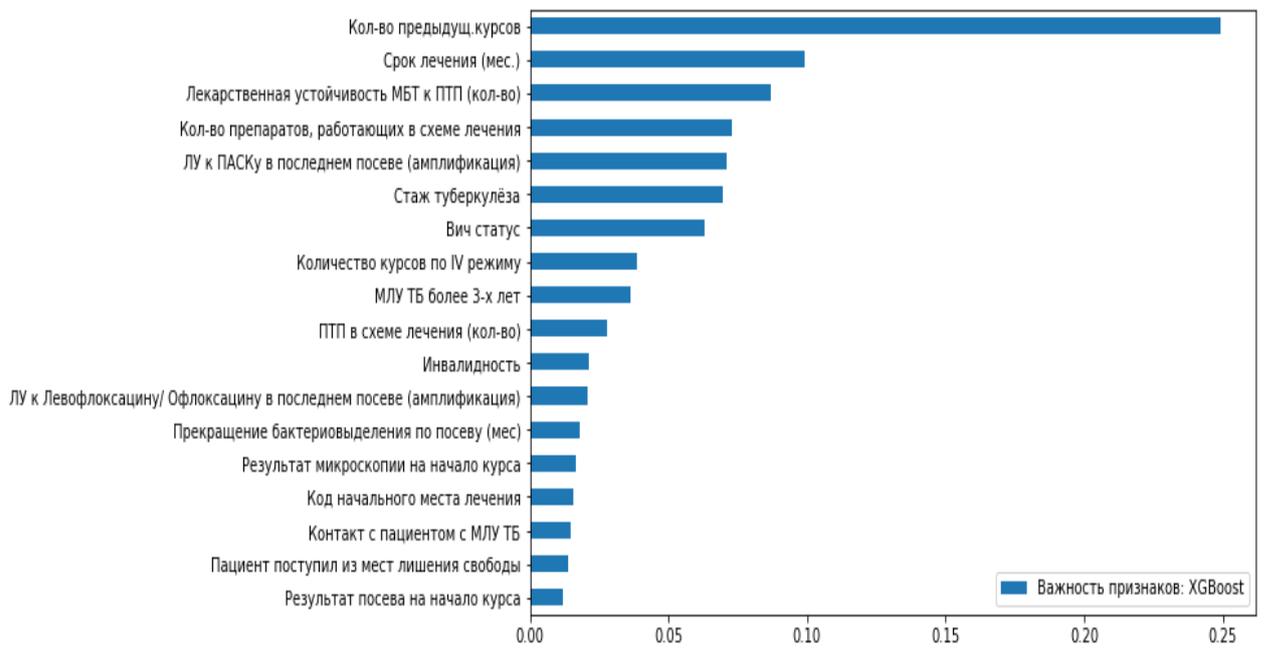


Рисунок 5 – Важность признаков, при использовании метода XGBoost (блок 5 перекрестной проверки)

Важность признаков, усредненная по 5 блокам перекрестной проверки для алгоритма экстремального градиентного бустинга XGBoost

Важность признаков: XGBoost	
Кол-во предыдущ. курсов	0.188984
Вич статус	0.101680
Срок лечения (мес.)	0.077305
Амплификация ЛУ МБТ	0.061704
Стаж туберкулёза	0.060841
Кол-во препаратов, работающих в схеме лечения	0.051301
МЛУ ТБ более 3-х лет	0.031258
ЛУ к Левофлоксацину/ Офлоксацину в последнем по...	0.027920
Лекарственная устойчивость МБТ к ПТП (кол-во)	0.025201
ПТП в схеме лечения (кол-во)	0.024673
Гепатит (В, С, В+С)	0.024165
Прекращение бактериовыделения по посеву (мес.)	0.020370
Наличие фазы распада	0.018420
Количество принятых доз (%)	0.017815
Результат микроскопии на начало курса	0.017689
Остаточные изменения после лечения	0.017642
Клиническая форма туберкулёза	0.017432
Остаточные изменения	0.017076
Остаточное изменение в виде каверны	0.014403
ЛУ к ПАСКу в последнем посеве (амплификация)	0.014170
Семейное положение	0.013195
Количество курсов по IV режиму	0.012810
Индекс массы тела	0.012189
Пенсионный возраст	0.011810
Контакт с пациентом с МЛУ ТБ	0.011641
ЛУ к Канамицину/Амикацину	0.011352
Результат посева на начало курса	0.010112
ЛУ к Этамбутолу в последнем посеве (амплификация)	0.010111
Градация возраста	0.009513
Наличие ПАСКа в схеме лечения	0.009232
Инвалидность	0.008904
Нарушение режима	0.008891
Наличие Капреомицина в начальной схеме лечения	0.008223
ЛУ к Этионадиму на начало лечения	0.008087
Диаметр полости в легких (см.)	0.004450
Код начального места лечения	0.003612
Пациент поступил из мест лишения свободы	0.003450
Работающий пациент	0.002807
Пол пациента	0.002062
ЛУ к Этамбутола на начало лечения	0.001821
Наличие язвенной болезни желудка	0.001485
Наличие Капреомицина в схеме лечения	0.000969
Наличие Левофлоксацина/ Офлоксацина в схеме леч...	0.000917
Злоупотребление алкоголем	0.000892
ЛУ к Канамицину/ Амикацину на момент обнаружени...	0.000524
ЛУ к Левофлоксацину/Офлоксацину	0.000428
Наличие психического заболевания	0.000371
Тип туберкулезного процесса	0.000093
Почечная недостаточность	0.000000
Сердечно-сосудистые заболевания	0.000000
Этамбутол в схеме лечения	0.000000
Табакокурение	0.000000
Начилие Моксифлоксацина в схеме лечения	0.000000
Оперативные вмешательства по поводу ТБ	0.000000
Теризидон в схеме лечения	0.000000
ЛУ к Стрептомицину	0.000000
Наличие Пиразинамида в схеме лечения	0.000000
Наличие Резонизада в схеме лечения	0.000000
Наличие Протионамида в схеме лечения	0.000000
Пиразинамид в схеме лечения	0.000000
ЛУ к Этамбутолу	0.000000
Употребление наркотических препаратов	0.000000
Дыхательная недостаточность	0.000000
ЛУ к Стрептомицину на начало лечения	0.000000
Хронические неспецифические заболевания легких	0.000000

ЛУ к Этионамиду на момент обнаружения МЛУ	0.000000
ЛУ к Амикацину/ Канамицину на начало лечения	0.000000
ЛУ к Капреомицину на начало лечения	0.000000
ЛУ к Циклосерину на начало лечения	0.000000
ЛУ к ПАСК	0.000000
ЛУ к Левофлоксацину/ Офлоксацину на начало лечения	0.000000
ЛУ к ПАСКу на начало лечения	0.000000
Перерывы в лечении в анамнезе	0.000000
ЛУ к Стрептомицину в последнем посеве (амплифик...	0.000000
Наличие Циклосерина в схеме лечения	0.000000
Наличие диабета	0.000000
ЛУ к Циклосерину	0.000000
ЛУ к Капреомицину в последнем посеве (амплифика...	0.000000
ЛУ к Циклосерину в последнем посеве (амплификация)	0.000000
ЛУ к Этионамиду в последнем посеве (амплификация)	0.000000
ЛУ к Капреомицину	0.000000
Бездомный	0.000000
Наличие Этамбутола в схеме лечения	0.000000

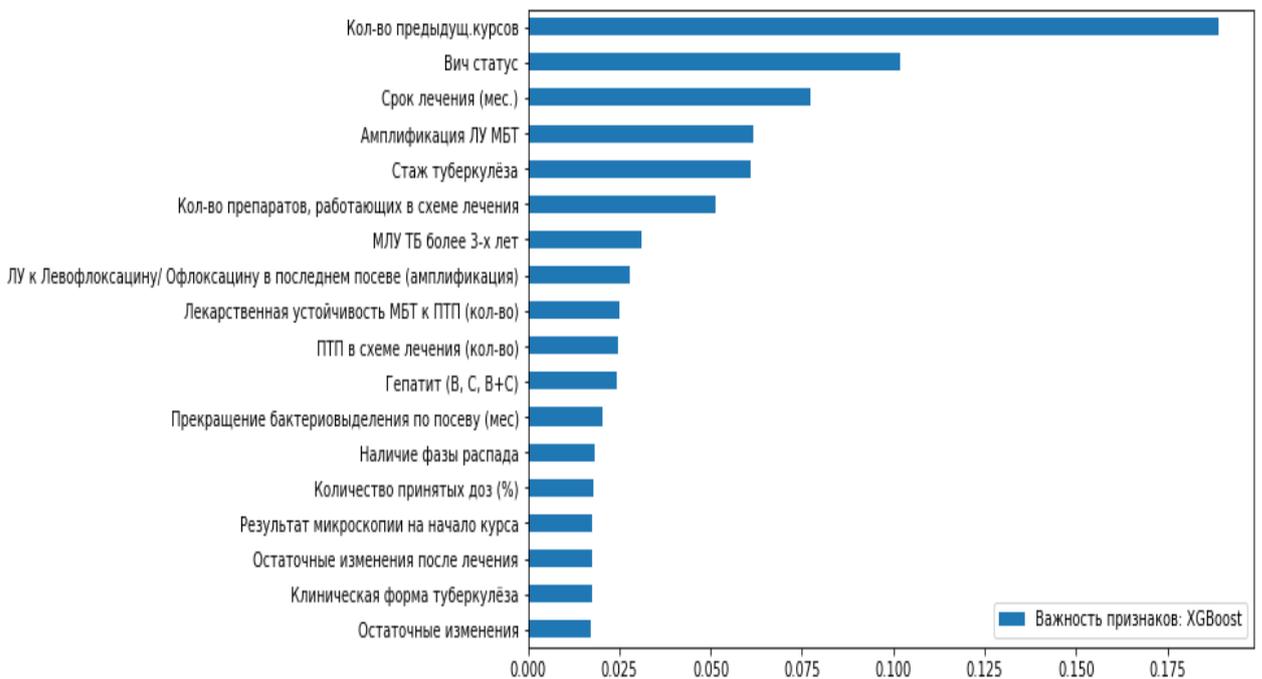


Рисунок 6 – Усредненная важность признаков по 5 блокам проверки, при использовании метода XGBoost

ROC K-блочная стратифицированная перекрестная проверка: Алгоритм экстремального градиентного бустинга XGBoost

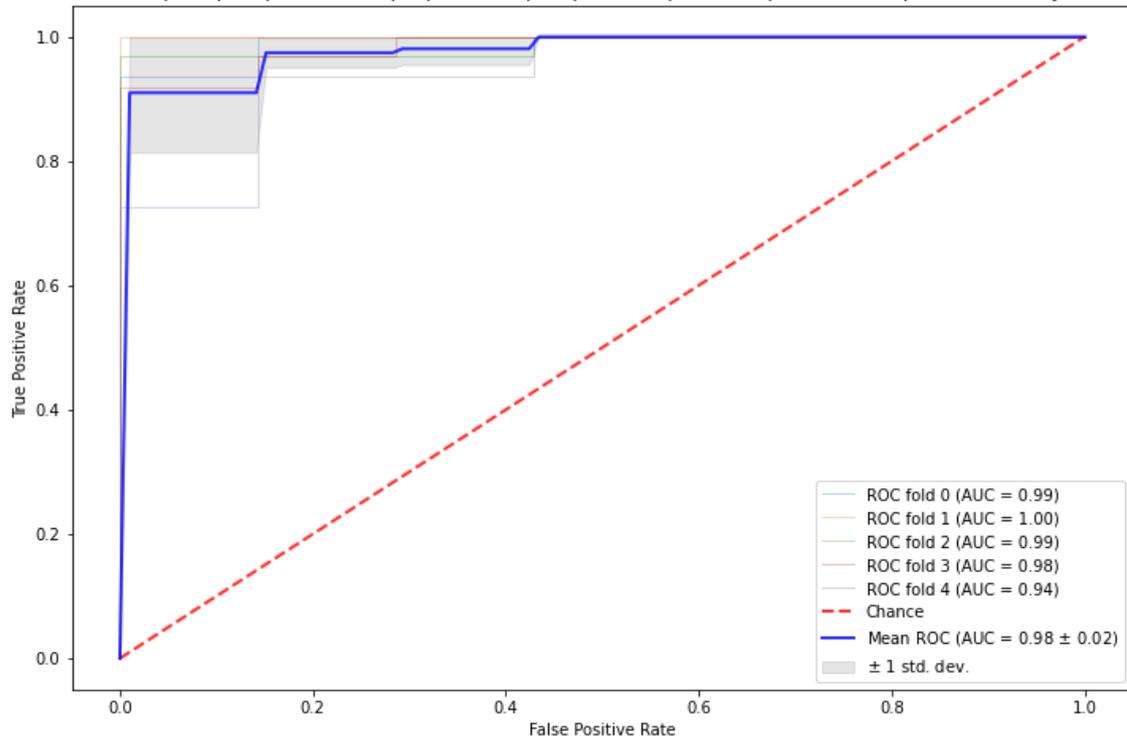


Рисунок 7 – ROC-анализ кривых, построенных для каждого из 5 блоков признаков перекрестной проверки, а также их усредненной (XGBoost)

ПРИЛОЖЕНИЕ Г

(справочное)

Иллюстративный материал этапов работы классификатора на основе
метода логистическая регрессия

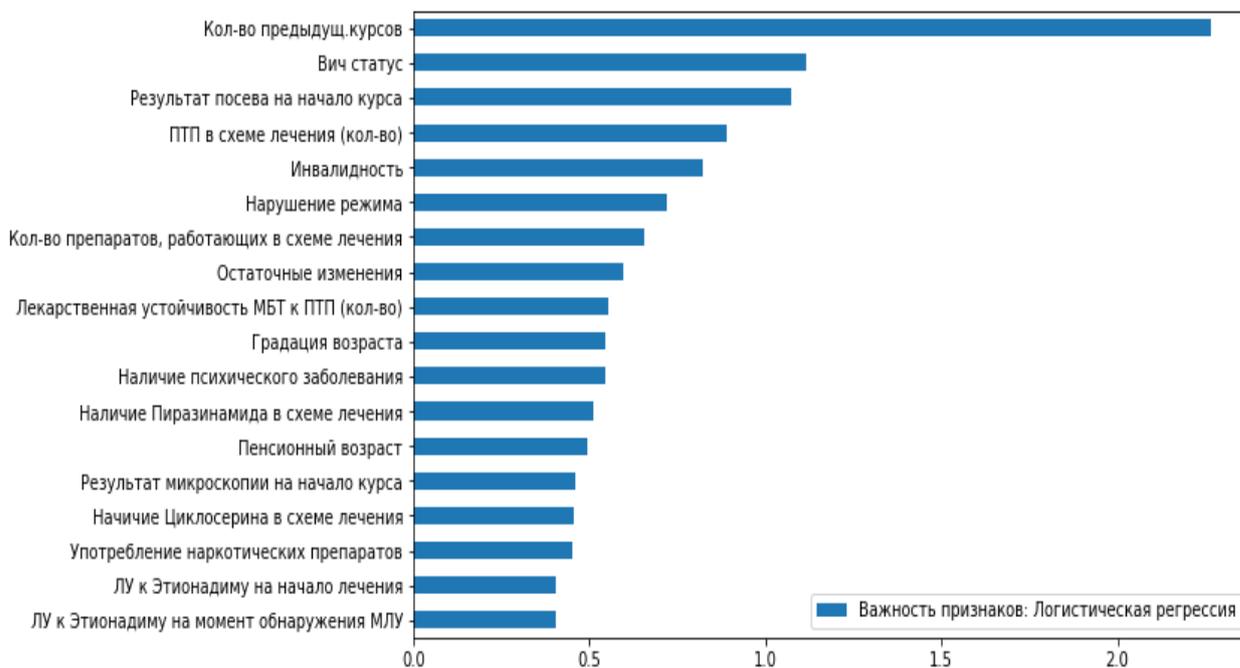


Рисунок 1 – Важность признаков, при использовании метода логистическая регрессия (блок 1 перекрестной проверки)

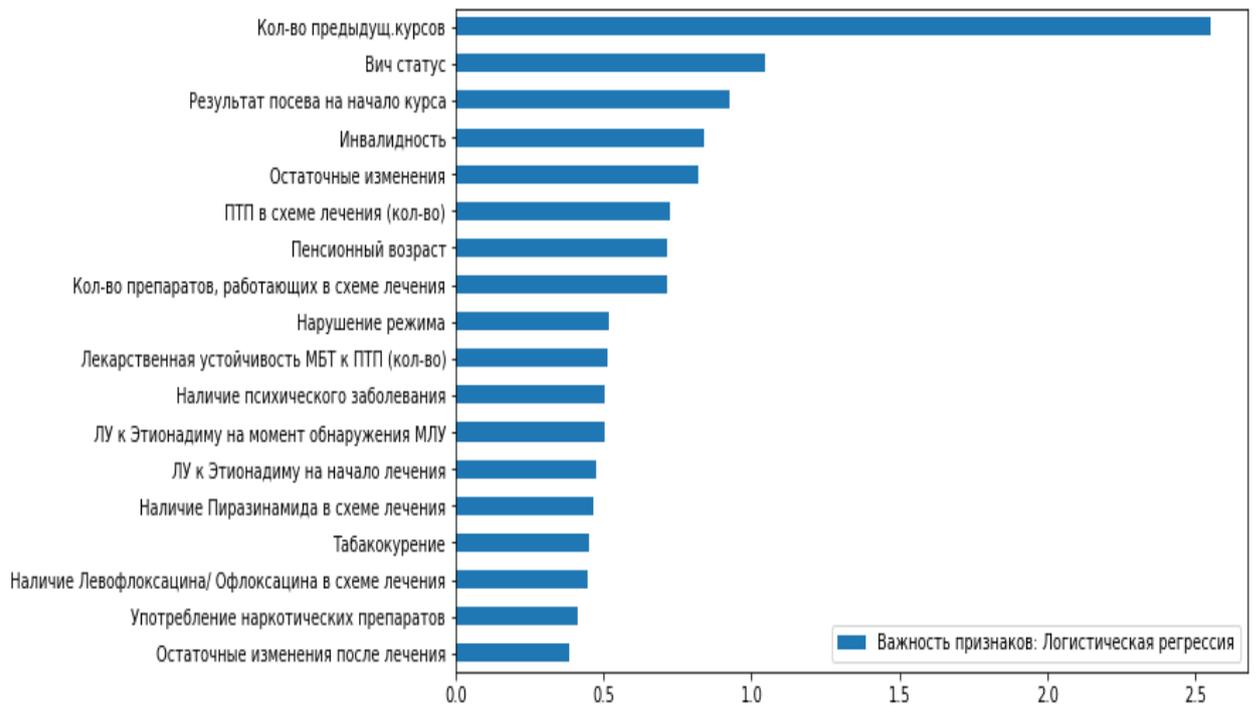


Рисунок 2 – Важность признаков, при использовании метода логистическая регрессия (блок 2 перекрестной проверки)

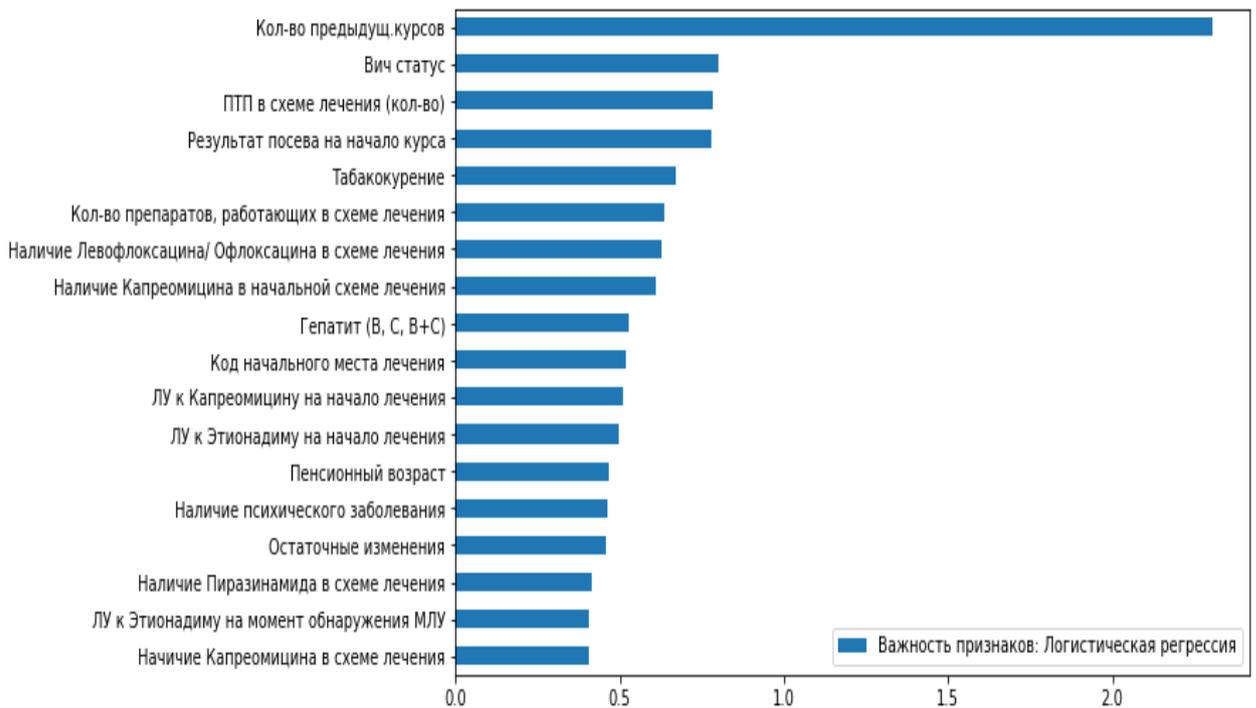


Рисунок 3 – Важность признаков, при использовании метода логистическая регрессия (блок 3 перекрестной проверки)

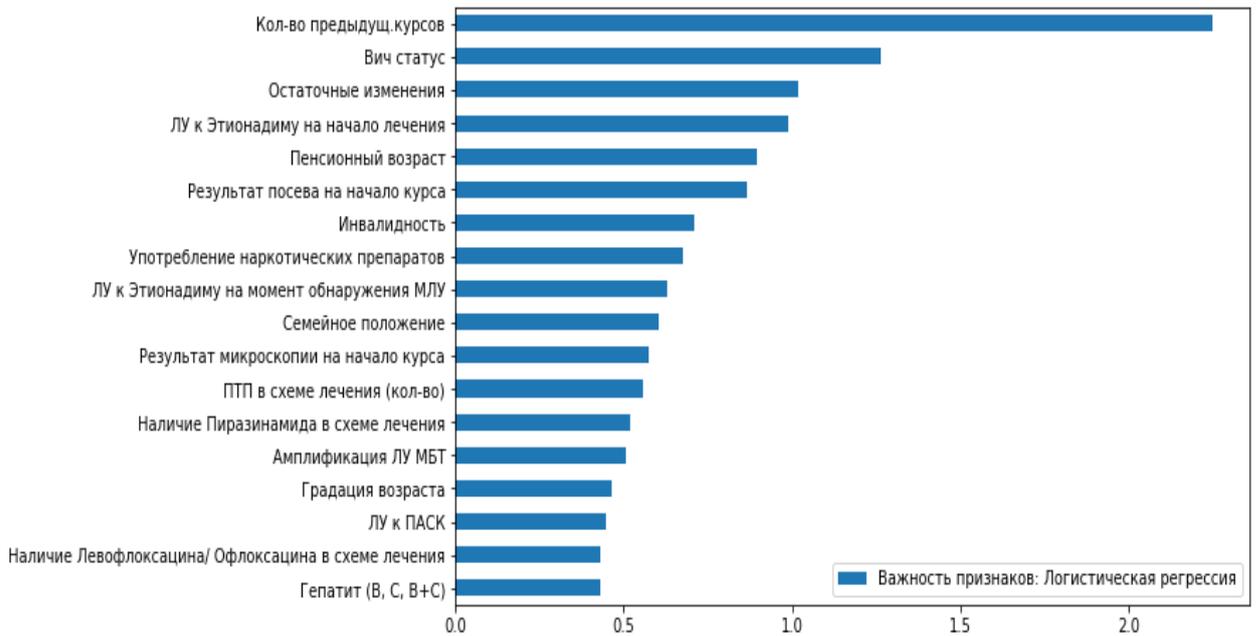


Рисунок 4 – Важность признаков, при использовании метода логистическая регрессия (блок 4 перекрестной проверки)

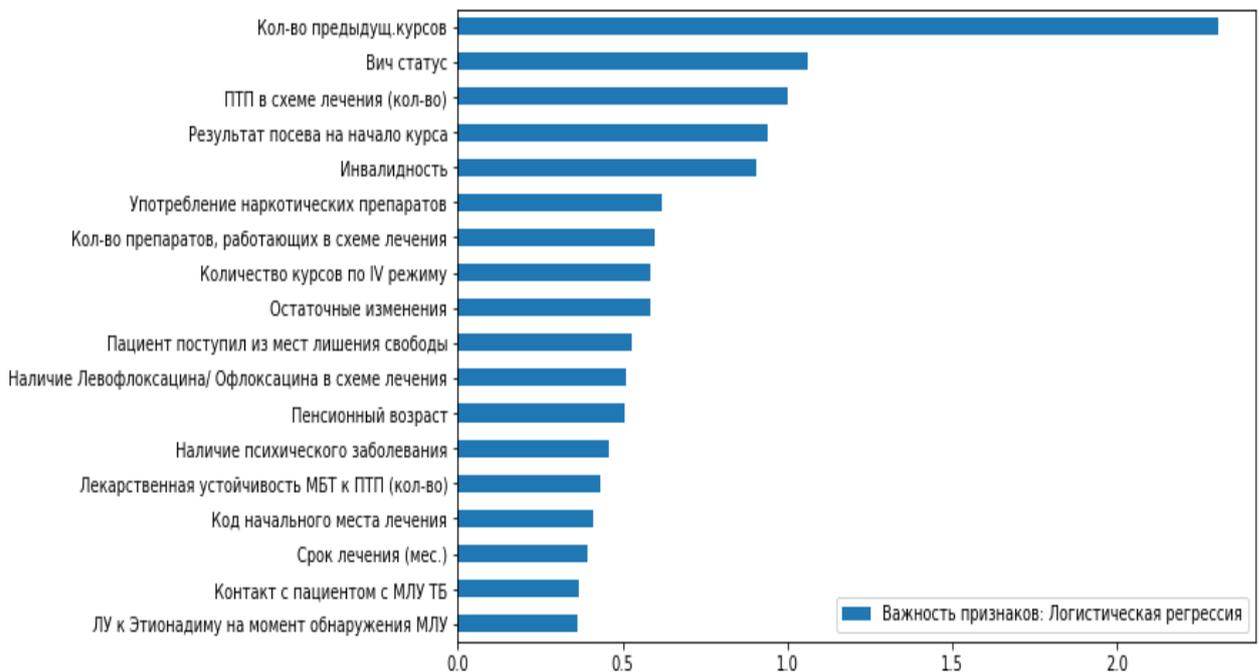


Рисунок 5 – Важность признаков, при использовании метода логистическая регрессия (блок 5 перекрестной проверки)

Важность признаков, усредненная по 5 блокам перекрестной проверки для логистической регрессии

Важность признаков: Логистическая регрессия	2.334125
Кол-во предыдущ. курсов	1.056821
Вич статус	0.916553
Результат посева на начало курса	0.791116
ПТП в схеме лечения (кол-во)	0.705145
Инвалидность	0.695329
Остаточные изменения	0.616157
Пенсионный возраст	0.587286
Кол-во препаратов, работающих в схеме лечения	0.487341
ЛУ к Этионадиму на начало лечения	0.464433
Нарушение режима	0.462975
Наличие Левофлоксацина/ Офлоксацина в схеме леч...	0.461176
ЛУ к Этионадиму на момент обнаружения МЛУ	0.459498
Наличие психического заболевания	0.454311
Употребление наркотических препаратов	0.432155
Наличие Пиразинамида в схеме лечения	0.400542
Лекарственная устойчивость МБТ к ПТП (кол-во)	0.399357
Табакокурение	0.398593
Градация возраста	0.356548
Гепатит (В, С, В+С)	0.327760
Наличие Циклосерина в схеме лечения	0.324431
Наличие Капреомицина в начальной схеме лечения	0.316708
Результат микроскопии на начало курса	0.305700
Количество курсов по IV режиму	0.291542
Семейное положение	0.283390
Пациент поступил из мест лишения свободы	0.278368
ЛУ к ПАСК	0.268273
Наличие фазы распада	0.262847
Наличие Капреомицина в схеме лечения	0.261916
Срок лечения (мес.)	0.259583
Диаметр полости в легких (см.)	0.252855
Код начального места лечения	0.244591
ЛУ к Циклосерину в последнем посеве (амплификация)	0.232288
ЛУ к Канамицину/Амикацину	0.228958
Тип туберкулезного процесса	0.222567
Контакт с пациентом с МЛУ ТБ	0.221867
Остаточные изменения после лечения	0.218083
Пол пациента	0.216862
Работающий пациент	0.206012
Наличие Протионамида в схеме лечения	0.199563
ЛУ к Амикацину/ Канамицину на начало лечения	0.195395
Этамбутол в схеме лечения	0.187548
Амплификация ЛУ МБТ	0.180539
Пиразинамид в схеме лечения	0.177046
ЛУ к Капреомицину на начало лечения	0.175068
Наличие язвенной болезни желудка	0.172095
Наличие Моксифлоксацина в схеме лечения	0.170033
Наличие ПАСКа в схеме лечения	0.161063
ЛУ к Левофлоксацину/ Офлоксацину на начало лечения	0.160680
МЛУ ТБ более 3-х лет	0.159271
ЛУ к Этамбутола на начало лечения	0.159264
Наличие Резонизада в схеме лечения	0.159264
Теризидон в схеме лечения	0.151743
Остаточное изменение в виде каверны	0.147352
Клиническая форма туберкулёза	0.146853
Стаж туберкулёза	0.130805
Бездомный	0.123272
ЛУ к ПАСКу на начало лечения	0.122223
Оперативные вмешательства по поводу ТБ	0.115767
Хронические неспецифические заболевания легких	0.112490
ЛУ к Левофлоксацину/Офлоксацину	0.112208
Индекс массы тела	0.111794
ЛУ к Стрептомицину в последнем посеве (амплифик...	0.111794
ЛУ к Этионадиму в последнем посеве (амплификация)	0.107975
Наличие диабета	0.107112
Прекращение бактериовыделения по посеву (мес.)	0.103499

ЛУ к Капреомицину	0.102062
ЛУ к Капреомицину в последнем посеве (амплифика...	0.098957
Наличие Этамбутола в схеме лечения	0.098868
ЛУ к Стрептомицину на начало лечения	0.097580
Злоупотребление алкоголем	0.093794
ЛУ к Левофлоксацину/ Офлоксацину в последнем по...	0.090821
ЛУ к Канамицину/ Амикацину на момент обнаружени...	0.089970
ЛУ к Стрептомицину	0.089312
ЛУ к ПАСКу в последнем посеве (амплификация)	0.084061
Перерывы в лечении в анамнезе	0.078668
ЛУ к Циклосерину на начало лечения	0.069076
ЛУ к Этамбутолу в последнем посеве (амплификация)	0.068677
ЛУ к Циклосерину	0.057776
Количество принятых доз (%)	0.056048
Сердечно-сосудистые заболевания	0.047627
ЛУ к Этамбутолу	0.047566
Дыхательная недостаточность	0.032117
Почечная недостаточность	0.024510

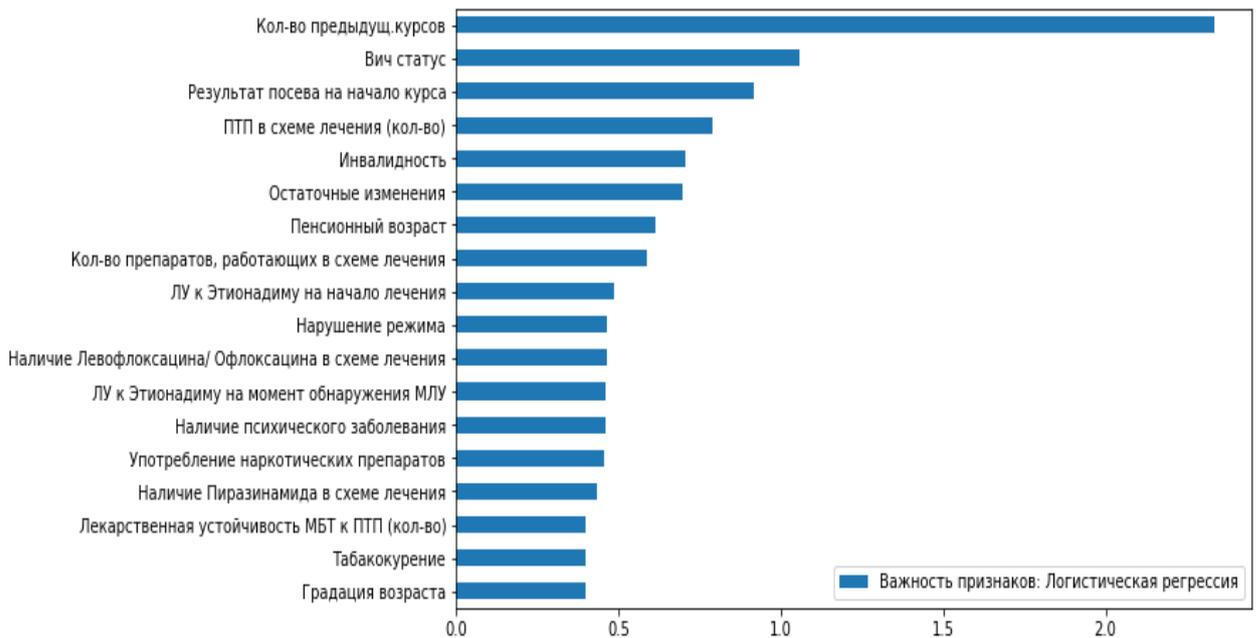


Рисунок 6 – Усредненная важность признаков по 5 блокам проверки, при использовании метода логистическая регрессия

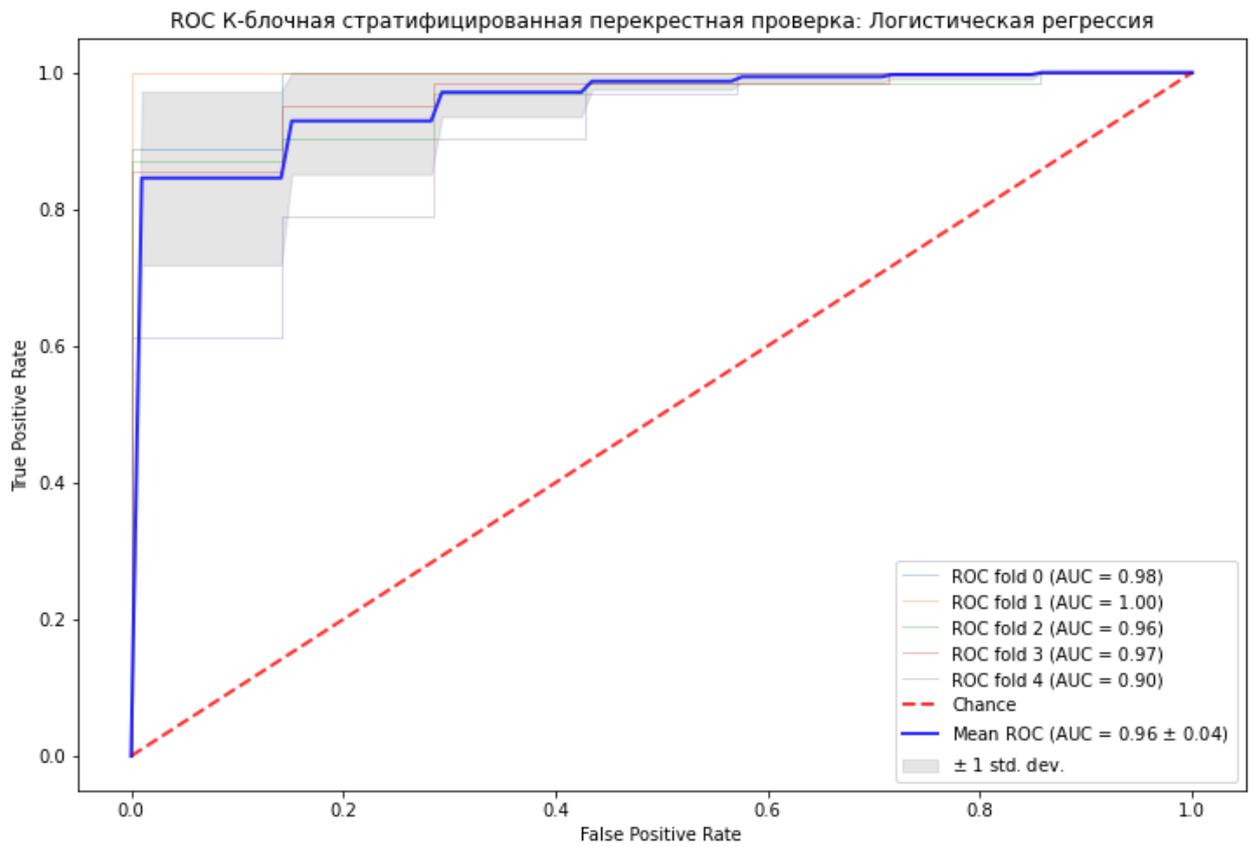


Рисунок 7 – ROC-анализ кривых, построенных для каждого из 5 блоков признаков перекрестной проверки, а также их усредненной (для алгоритма логистическая регрессия)

ПРИЛОЖЕНИЕ Д

(справочное)

Тестируемый набор признаков пациента № 11**Тестирующие признаки пациента № 11**

Количество курсов по IV режиму	0.00
Результат микроскопии на начало курса	0.00
Результат посева на начало курса	1.00
Клиническая форма туберкулёза	4.00
Нарушение режима	1.00
Пол пациента	1.00
Индекс массы тела	21.20
Градация возраста	4.00
Стаж туберкулёза	12.24
МЛУ ТБ более 3-х лет	5.84
Семейное положение	1.00
Контакт с пациентом с МЛУ ТБ	0.00
Работающий пациент	0.00
Бездомный	0.00
Пенсионный возраст	0.00
Инвалидность	1.00
Пациент поступил из мест лишения свободы	0.00
Вич статус	0.00
Хронические неспецифические заболевания легких	0.00
Злоупотребление алкоголем	1.00
Употребление наркотических препаратов	0.00
Наличие диабета	0.00
Наличие психического заболевания	1.00
Остаточные изменения	0.00
Гепатит (В, С, В+С)	0.00
Наличие язвенной болезни желудка	0.00
Почечная недостаточность	0.00
Сердечно-сосудистые заболевания	0.00
Табакокурение	1.00
Дыхательная недостаточность	0.00
Оперативные вмешательства по поводу ТБ	0.00
Кол-во предыдущ. курсов	3.00
ЛУ к Стрептомицину	1.00
ЛУ к Этамбутолу	1.00
ЛУ к Канамицину/ Амикацину на момент обнаружения МЛУ	0.00
ЛУ к Капреомицину	0.00
ЛУ к Циклосерину	0.00
ЛУ к Этионадиму на момент обнаружения МЛУ	0.00
ЛУ к Левофлоксацину/Офлоксацину	0.00
ЛУ к ПАСК	0.00
ЛУ к Стрептомицину на начало лечения	1.00
ЛУ к Этамбутолу на начало лечения	1.00
ЛУ к Амикацину/ Канамицину на начало лечения	0.00
ЛУ к Капреомицину на начало лечения	0.00
ЛУ к Циклосерину на начало лечения	0.00
ЛУ к Этионадиму на начало лечения	1.00
ЛУ к Левофлоксацину/ Офлоксацину на начало лечения	0.00
ЛУ к ПАСКу на начало лечения	0.00
Перерывы в лечении в анамнезе	0.00
Амплификация ЛУ МБТ	0.00
Кол-во препаратов, работающих в схеме лечения	3.00
ЛУ к Стрептомицину в последнем посеве (амплификация)	1.00
ЛУ к Этамбутолу в последнем посеве (амплификация)	1.00

ЛУ к Канамицину/Амикацину	0.00
ЛУ к Капреомицину в последнем посеве (амплификация)	0.00
ЛУ к Циклосерину в последнем посеве (амплификация)	0.00
ЛУ к Этионамиду в последнем посеве (амплификация)	1.00
ЛУ к Левофлоксацину/ Офлоксацину в последнем посеве (амплификация)	0.00
ЛУ к ПАСКу в последнем посеве (амплификация)	0.00
Наличие Капреомицина в схеме лечения	0.00
Наличие Циклосерина в схеме лечения	1.00
Наличие Этамбутола в схеме лечения	1.00
Наличие Капреомицина в начальной схеме лечения	1.00
Наличие Моксифлоксацина в схеме лечения	0.00
Наличие Левофлоксацина/ Офлоксацина в схеме лечения	1.00
Наличие ПАСКа в схеме лечения	0.00
Наличие Протионамида в схеме лечения	1.00
Наличие Резонизада в схеме лечения	0.00
Наличие Пиразинамида в схеме лечения	1.00
Код начального места лечения	1.00
Тип туберкулезного процесса	2.00
Наличие фазы распада	1.00
Количество принятых доз (%)	99.00
Диаметр полости в легких (см.)	4.00
Остаточное изменение в виде каверны	0.00
Остаточные изменения после лечения	2.00
Лекарственная устойчивость МБТ к ПТП (кол-во)	3.00
Прекращение бактериовыделения по посеву (мес.)	16.70
Этамбутол в схеме лечения	0.00
Теризидон в схеме лечения	0.00
Пиразинамид в схеме лечения	1.00
ПТП в схеме лечения (кол-во)	5.00
Срок лечения (мес.)	24.20

ПРИЛОЖЕНИЕ Е

(справочное)

Тестируемый набор признаков пациента № 56**Тестирующие признаки пациента № 56**

Количество курсов по IV режиму	0.00
Результат микроскопии на начало курса	0.00
Результат посева на начало курса	2.00
Клиническая форма туберкулёза	18.00
Нарушение режима	0.00
Пол пациента	1.00
Индекс массы тела	27.76
Градация возраста	3.00
Стаж туберкулёза	0.60
МЛУ ТБ более 3-х лет	0.04
Семейное положение	0.00
Контакт с пациентом с МЛУ ТБ	0.00
Работающий пациент	1.00
Бездомный	0.00
Пенсионный возраст	0.00
Инвалидность	0.00
Пациент поступил из мест лишения свободы	0.00
Вич статус	0.00
Хронические неспецифические заболевания легких	0.00
Злоупотребление алкоголем	0.00
Употребление наркотических препаратов	0.00
Наличие диабета	0.00
Наличие психического заболевания	0.00
Остаточные изменения	0.00
Гепатит (В, С, В+С)	0.00
Наличие язвенной болезни желудка	0.00
Почечная недостаточность	0.00
Сердечно-сосудистые заболевания	0.00
Табакокурение	0.00
Дыхательная недостаточность	0.00
Оперативные вмешательства по поводу ТБ	0.00
Кол-во предыдущ. курсов	2.00
ЛУ к Стрептомицину	1.00
ЛУ к Этамбутолу	0.00
ЛУ к Канамицину/ Амикацину на момент обнаружения МЛУ	0.00
ЛУ к Капреомицину	0.00
ЛУ к Циклосерину	0.00
ЛУ к Этионадиму на момент обнаружения МЛУ	0.00
ЛУ к Левофлоксацину/Офлоксацину	0.00
ЛУ к ПАСК	0.00
ЛУ к Стрептомицину на начало лечения	1.00
ЛУ к Этамбутолу на начало лечения	0.00
ЛУ к Амикацину/ Канамицину на начало лечения	0.00
ЛУ к Капреомицину на начало лечения	0.00
ЛУ к Циклосерину на начало лечения	0.00
ЛУ к Этионадиму на начало лечения	0.00
ЛУ к Левофлоксацину/ Офлоксацину на начало лечения	0.00
ЛУ к ПАСКу на начало лечения	0.00
Перерывы в лечении в анамнезе	0.00
Амплификация ЛУ МБТ	1.00
Кол-во препаратов, работающих в схеме лечения	2.00
ЛУ к Стрептомицину в последнем посеве (амплификация)	1.00
ЛУ к Этамбутолу в последнем посеве (амплификация)	0.00

ЛУ к Канамицину/Амикацину	0.00
ЛУ к Капреомицину в последнем посеве (амплификация)	0.00
ЛУ к Циклосерину в последнем посеве (амплификация)	0.00
ЛУ к Этионамиду в последнем посеве (амплификация)	1.00
ЛУ к Левофлоксацину/ Офлоксацину в последнем посеве (амплификация)	0.00
ЛУ к ПАСКу в последнем посеве (амплификация)	0.00
Наличие Капреомицина в схеме лечения	0.00
Наличие Циклосерина в схеме лечения	1.00
Наличие Этамбутола в схеме лечения	0.00
Наличие Капреомицина в начальной схеме лечения	0.00
Наличие Моксифлоксацина в схеме лечения	0.00
Наличие Левофлоксацина/ Офлоксацина в схеме лечения	1.00
Наличие ПАСКа в схеме лечения	0.00
Наличие Протионамида в схеме лечения	1.00
Наличие Резонизада в схеме лечения	0.00
Наличие Пиразинамида в схеме лечения	0.00
Код начального места лечения	3.00
Тип туберкулезного процесса	1.00
Наличие фазы распада	0.00
Количество принятых доз (%)	99.00
Диаметр полости в легких (см.)	0.00
Остаточное изменение в виде каверны	0.00
Остаточные изменения после лечения	0.00
Лекарственная устойчивость МБТ к ПТП (кол-во)	2.00
Прекращение бактериовыделения по посеву (мес.)	8.80
Этамбутол в схеме лечения	1.00
Теризидон в схеме лечения	0.00
Пиразинамид в схеме лечения	1.00
ПТП в схеме лечения (кол-во)	6.00
Срок лечения (мес.)	20.60

ПРИЛОЖЕНИЕ Ж

(справочное)

ПРОГРАММА ДЛЯ ЭВМ

«Программа-калькулятор для оценки риска развития рецидива у больных с
множественно лекарственно устойчивым туберкулезом»

(фрагменты текста программы)

Загрузка базы пациентов

```
originalDf = pd.read_excel('tb_relapse.xlsx', sheet_name='БДТуберкулез')
```

Установка названий признаков на русском языке

```
newColumnNames = {'DAVNTB' : 'Стаж туберкулёза',
  'PRAVOK' : 'Остаточные изменения в правом легком',
  'LEVOK' : 'Остаточные изменения в левом легком',
  'mespospl+2' : 'Прекращение бактериовыделения по посеву (мес.)',
  'POVT' : 'Количество курсов по IV режиму',
  'TrdSK' : 'Теризидон в схеме лечения',
  'POCHNE' : 'Почечная недостаточность',
  'YAZV' : 'Язвенная болезнь желудка',
  'SERDNE' : 'Сердечно-сосудистые заболевания',
  'KOLUST2' : 'Лекарственная устойчивость МБТ к ПТП (кол-во)',
  'tip TB process' : 'Тип туберкулезного процесса',
  'PTPSHEMK' : 'ПТП в схеме лечения (кол-во)',
  'KOLKURS' : 'Кол-во предыдущ. курсов',
  'ETASK' : 'Этамбутол в схеме лечения',
  'KOLABP' : 'Кол-во препаратов, работающих в схеме лечения',
  'SROK2' : 'Срок лечения (мес.)',
  'CVN' : 'Диаметр полости в легких (см.)',
  'DAVNMDR' : 'МЛУ ТБ более 3-х лет',
  'PRAVON' : 'Пат. процесс в правом легком на начало лечения',
  'LEVON' : 'Пат. процесс в левом легком на начало лечения',
  'KODIAPLUS' : 'Клиническая форма туберкулёза',
  'PENS' : 'Пенсионный возраст',
  'SPID' : 'Вич статус',
  'INDTELA' : 'Индекс массы тела',
  'PRIEM' : 'Количество принятых доз (%)',
  'Destruct' : 'Наличие фазы распада',
  'AMPLIFIC' : 'Амплификация ЛУ МБТ',
  'AMT+KANT' : 'ЛУ к Канамицину/Амикацину',
  'ETA' : 'ЛУ к Этамбутолу',
  'KUREN' : 'Табакокурение',
  'DYHNE' : 'Дыхательная недостаточность',
  'OPERTB' : 'Оперативные вмешательства по поводу ТБ',
  'ZSK' : 'Пиразинамид в схеме лечения',
  'OSTTB_ranee' : 'Пневмофиброз',
  'CAP' : 'ЛУ к Капреомицину',
  'STP' : 'ЛУ к Стрептомицину',
  'CUC' : 'ЛУ к Циклосерину',
  'LFX+OFL' : 'ЛУ к Левофлоксацину/Офлоксацину',
  'PAS' : 'ЛУ к ПАСК',
  'YAZV' : 'Наличие язвенной болезни желудка',
  'GEPAT' : 'Гепатит (B, C, B+C)',
  'CVK' : 'Остаточное изменение в виде каверны',
  'OSTTB_ranee' : 'Остаточные изменения',

  'ALK' : 'Злоупотребление алкоголем',
  'BOMJ' : 'Бездомный',
```

```

'CAPP' : 'ЛУ к Капреомицину на начало лечения',
'CAPS' : 'Наличие Капреомицина в схеме лечения',
'CAPT' : 'ЛУ к Капреомицину в последнем посеве (амплификация)',
'CYCP' : 'ЛУ к Циклосерину на начало лечения',
'CYCS' : 'Наличие Циклосерина в схеме лечения',
'CYCT' : 'ЛУ к Циклосерину в последнем посеве (амплификация)',
'ES' : 'Наличие Этамбутола в схеме лечения',
'ETAP' : 'ЛУ к Этамбутолу на начало лечения',
'ETAT' : 'ЛУ к Этамбутолу в последнем посеве (амплификация)',

'ETH' : 'ЛУ к Этионадиму на момент обнаружения МЛУ',
'ETHP' : 'ЛУ к Этионадиму на начало лечения',
'ETHT' : 'ЛУ к Этионамиду в последнем посеве (амплификация)',

'MOXS' : 'Наличие Моксифлоксацина в схеме лечения',
'MSKPL' : 'Результат микроскопии на начало курса',
'MestNach' : 'Код начального места лечения',
'NARK' : 'Употребление наркотических препаратов',
'NARUSH' : 'Нарушение режима',
'HNZL' : 'Хронические неспецифические заболевания легких',
'KS' : 'Наличие Капреомицина в начальной схеме лечения',
'KONMLU' : 'Контакт с пациентом с МЛУ ТВ',
'INVALID' : 'Инвалидность',
'GIT' : 'Проживание в городе',
'GRADVOZ' : 'Градация возраста',

'IZUIN' : 'Пациент поступил из мест лишения свободы',
'ODIN' : 'Семейное положение',
'RAB' : 'Работающий пациент',
'SAHD' : 'Наличие диабета',
'PSIH' : 'Наличие психического заболевания',
'PERER' : 'Перерывы в лечении в анамнезе',
'POL' : 'Пол пациента',

'POSPL' : 'Результат посева на начало курса',
'OSTTB_new' : 'Остаточные изменения после лечения',

'PASP' : 'ЛУ к ПАСКу на начало лечения',
'PASS' : 'Наличие ПАСКа в схеме лечения',
'PAST' : 'ЛУ к ПАСКу в последнем посеве (амплификация)',
'PTNS' : 'Наличие Протионамида в схеме лечения',

'KAN+AM' : 'ЛУ к Канамицину/ Амикацину на момент обнаружения МЛУ',

'LFXT+OFLT' : 'ЛУ к Левофлоксацину/ Офлоксацину в последнем посеве (амплификация)',
'AMP+KANP' : 'ЛУ к Амикацину/ Канамицину на начало лечения',
'LFXP+OFLP' : 'ЛУ к Левофлоксацину/ Офлоксацину на начало лечения',
'STRP' : 'ЛУ к Стрептомицину на начало лечения',

'STRT' : 'ЛУ к Стрептомицину в последнем посеве (амплификация)',
'LFXS+OFLS' : 'Наличие Левофлоксацина/ Офлоксацина в схеме лечения',
'RESS' : 'Наличие Резонизада в схеме лечения',
'ZS' : 'Наличие Пиразинамида в схеме лечения',

'TrdSK' : 'Теризидон в схеме лечения'
}

```

Отбор зависимых переменных и целевой (подтвержден рецидив) переменной

```

dfX = originalDf.drop(['REC'], axis =1)
dfy = originalDf['REC']
y=dfy.values
X=dfX.values

```

Разбиение базы на обучающую и тестовую выборки

```

from sklearn.model_selection import train_test_split

```

```
X_train,X_test,y_train,y_test=train_test_split(X, y, test_size=0.3, random_state=0)
```

Построение классифицирующей модели на основе дерева решений

```
from sklearn import tree
from sklearn.tree import DecisionTreeClassifier
treeModel = DecisionTreeClassifier(criterion='entropy', max_depth=6, min_samples_split=8, min_impurity_split=0.2, random_state=1)
treeModel.fit(X_train,y_train)
```

Оценка метрик модели (матрица ошибок, чувствительность, специфичность, ROC, AUC) для модели на основе дерева решений

```
import sklearn.metrics as metrics
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
from sklearn.metrics import precision_recall_fscore_support as score
from sklearn.metrics import confusion_matrix
y_pred_tree = treeModel.predict(X_test)
confmat_tree=confusion_matrix(y_true=y_test,y_pred=y_pred_tree)
print(confmat_tree)

specificity_tree = confmat_tree[1][1]/(confmat_tree[1][0] + confmat_tree[1][1])
sensitivity_tree = confmat_tree[0][0]/(confmat_tree[0][0] + confmat_tree[0][1])

print(f'Чувствительность (Дерево решений): {round(sensitivity_tree,2)}')
print(f'Специфичность (Дерево решений): {round(specificity_tree,2)}')

probsTree = treeModel.predict_proba(X_test)
predsTree = probsTree[:,1]
fprTree, tprTree, thresholdTree = metrics.roc_curve(y_test, predsTree)
roc_aucTree = metrics.auc(fprTree, tprTree)
print(f'AUC (Дерево решений): {round(roc_aucTree,2)}')
```

Построение классифицирующей модели на основе алгоритма случайного леса

```
from sklearn.ensemble import RandomForestClassifier

parameters = {'n_estimators': [3, 5, 7, 9, 11, 13, 15, 17, 19]}
model = RandomForestClassifier()
gridForest = GridSearchCV(estimator=model, param_grid=parameters)
gridForest.fit(X_train, y_train)
print('Оценка лучшего решения :', gridForest.best_score_)
print('Лучшая модель для дерева :', gridForest.best_estimator_)
```

Оценка метрик модели (матрица ошибок, чувствительность, специфичность, ROC, AUC) для модели на основе алгоритма случайного леса

```
forestModel = gridForest.best_estimator_
y_pred_forest=forestModel.predict(X_test)
confmat_forest=confusion_matrix(y_true=y_test,y_pred=y_pred_forest)
print(confmat_forest)

specificity_forest = confmat_forest[1][1]/(confmat_forest[1][0] + confmat_forest[1][1])
sensitivity_forest = confmat_forest[0][0]/(confmat_forest[0][0] + confmat_forest[0][1])

print(f'Чувствительность (Случайный лес): {round(sensitivity_forest,2)}')
print(f'Специфичность (Случайный лес): {round(specificity_forest,2)}')

probsForest = forestModel.predict_proba(X_test)
predsForest = probsForest[:,1]
fprForest, tprForest, thresholdForest = metrics.roc_curve(y_test, predsForest)
roc_aucForest = metrics.auc(fprForest, tprForest)
print(f'AUC (Случайный лес): {round(roc_aucForest,2)}')
```

Построение классифицирующей модели на основе алгоритма экстремального градиентного бустинга XGBoost

```

from xgboost import XGBClassifier

parameters = {'n_estimators': [3, 5, 7, 9, 11, 13, 15, 17, 19]}
model = XGBClassifier()
gridXGBoost = GridSearchCV(estimator=model, param_grid=parameters)
gridXGBoost.fit(X_train, y_train)
print('Оценка лучшего решения   :', gridXGBoost.best_score_)
print('Лучшая модель для дерева  :', gridXGBoost.best_estimator_)

```

Оценка метрик модели (матрица ошибок, чувствительность, специфичность, ROC, AUC) для модели на основе алгоритма экстремального градиентного бустинга XGBoost

```

XGBModel = gridXGBoost.best_estimator_
y_pred_xgb=XGBModel.predict(X_test)
confmat_XGBModel = confusion_matrix(y_true=y_test, y_pred=y_pred_xgb)
print(confmat_XGBModel)

specificity_XGBModel = confmat_XGBModel[1][1]/(confmat_XGBModel[1][0] + confmat_XGBModel[1][1])
sensitivity_XGBModel= confmat_XGBModel[0][0]/(confmat_XGBModel[0][0] + confmat_XGBModel[0][1])

print(f'Чувствительность (модель XGBoost): {round(sensitivity_XGBModel,2)}')
print(f'Специфичность (модель XGBoost): {round(specificity_XGBModel,2)}')

probsXGBoost = XGBModel.predict_proba(X_test)
predsXGBoost = probsXGBoost[:,1]
fprXGBoost, tprXGBoost, thresholdXGBoost = metrics.roc_curve(y_test, predsXGBoost)
roc_aucXGBoost = metrics.auc(fprXGBoost, tprXGBoost)
print(f'AUC (Дерево решений): {round(roc_aucXGBoost,2)}')

```

Построение классификатора на основе алгоритма Логистической регрессии

```

from sklearn.linear_model import LogisticRegression

parameters = {'C': [0.001, 0.01, 0.1, 1.0, 10.0, 100.0, 1000.0]}
logRegModel = LogisticRegression()
gridLogReg = GridSearchCV(estimator=logRegModel, param_grid=parameters)
gridLogReg.fit(X_train, y_train)
print('Оценка лучшего решения   :', gridLogReg.best_score_)
print('Лучшая модель для лог.регрессии :', gridLogReg.best_estimator_)

logRegModel = gridLogReg.best_estimator_

```

Оценка метрик модели (матрица ошибок, чувствительность, специфичность, ROC, AUC) для модели логистической регрессии

```

y_pred_logReg = logRegModel.predict(X_test)
confmat_model_logReg = confusion_matrix(y_true = y_test, y_pred = y_pred_logReg)
print(confmat_model_logReg)

specificity_logReg = confmat_model_logReg[1][1]/(confmat_model_logReg[1][0] + confmat_model_logReg[1][1])
sensitivity_logReg = confmat_model_logReg[0][0]/(confmat_model_logReg[0][0] + confmat_model_logReg[0][1])

print(f'Чувствительность (Логистич.регрессия): {round(sensitivity_logReg,2)}')
print(f'Специфичность (Логистич.регрессия): {round(specificity_logReg,2)}')

probsLogReg = logRegModel.predict_proba(X_test)
predsLogReg = probsLogReg[:,1]
fprLogReg, tprLogReg, thresholdLogReg = metrics.roc_curve(y_test, predsLogReg)
roc_aucLogReg = metrics.auc(fprLogReg, tprLogReg)
print(f'AUC (Логистич.регрессия): {round(roc_aucLogReg,2)}')

```

Отображение ROC кривых классифицирующих моделей

```

plot.rc('figure',figsize=(12,8))

plot.title('ROC')
plot.plot(fprLogReg, tprLogReg, 'b', color = 'indigo', linestyle = '-
.', linewidth = 2, label = 'Модель логистической регрессии AUC = %0.2f' % roc_aucLogReg)
plot.plot(fprXGBoost, tprXGBoost, 'b', color = 'red', linestyle = '--
', linewidth = 2, label = 'Модель градиентного бустинга AUC = %0.2f' % roc_aucXGBoost)
plot.plot(fprTree, tprTree, 'b', color = 'blue', linestyle = '-
', linewidth = 2, label = 'Модель дерева решений AUC = %0.2f' % roc_aucTree)
plot.plot(fprForest, tprForest, 'b', color = 'green', linestyle = ':', linewidth = 2, label
 = 'Модель случайного леса AUC = %0.2f' % roc_aucForest)

plot.legend(loc = 'lower right')
plot.xlim([0, 1])
plot.ylim([0.75, 1])
plot.ylabel('Доля истинно-положительных')
plot.xlabel('Доля ложно-положительных')
plot.show()

```

Отображение гистограмм важности признаков классифицирующих моделей

```

feature_importances_treeDf = pd.DataFrame(treeModel.feature_importances_,index = dfX.column
s,
columns=['Важность признаков: Дерево решений']).sort_values('Важность признаков: Дерево реш
ений',ascending=True)
feature_importances_treeDf.plot.barh(figsize=(10, 6))

feature_importancesForestDf = pd.DataFrame(forestModel.feature_importances_,index = dfX.col
umns,
columns=['Важность признаков: Случайный лес']).sort_values('Важность признаков: Случайный л
ес',ascending=True)
feature_importancesForestDf.plot.barh(figsize=(15, 40))

feature_importancesXGBoostDf = pd.DataFrame(XGBModel.feature_importances_,index = dfX.column
s,
columns=['Важность признаков: XGBoost']).sort_values('Важность признаков: XGBoost',ascendin
g=True)
feature_importancesXGBoostDf.plot.barh(figsize=(10, 6))

feature_importance_logReg_ver1 = abs(logRegModel.coef_[0])
feature_importances_logRegDf_ver1 = pd.DataFrame(feature_importance_logReg_ver1,index = dfX
.columns,
columns = ['Важность признаков: лог.регрессия']).sort_values('Важность признаков: лог.регре
ссия',ascending=True)
feature_importances_logRegDf.plot.barh(figsize=(10, 6))

```

Выполнение K-блочной стратифицированной перекрестной проверки для модели дерева решений, визуализация ROC-кривых

```

from sklearn.metrics import auc
from sklearn.metrics import plot_roc_curve

tree_kfoldModels = []
tree_kfoldModel = DecisionTreeClassifier(max_depth=5, random_state=1)

from sklearn.model_selection import StratifiedKFold
kfold_tree = StratifiedKFold(n_splits=5, random_state=1, shuffle=True).split(df_X, df_y)

scores = []
matricesTree = []
meanTreeImportances = np.zeros(91)

tprs = []
aucs = []
mean_fpr = np.linspace(0, 1, 100)

```

```

mean_tprTree = np.linspace(0, 1, 100)
mean_fprTree = np.linspace(0, 1, 100)

fig, ax = plot.subplots()
for k, (train, test) in enumerate(kfold_tree):

    tree_kfoldModel.fit(dfX.iloc[train], dfy.iloc[train])
    viz = plot_roc_curve(tree_kfoldModel, dfX.iloc[test], dfy.iloc[test],
                        name='ROC fold {}'.format(k), alpha=0.3, lw=1, ax=ax)

    interp_tpr = np.interp(mean_fpr, viz.fpr, viz.tpr)
    interp_tpr[0] = 0.0
    tprs.append(interp_tpr)
    aucs.append(viz.roc_auc)

ax.plot([0, 1], [0, 1], linestyle='--', lw=2, color='r',
        label='Chance', alpha=.8)

mean_tpr = np.mean(tprs, axis=0)
mean_tpr[-1] = 1.0
mean_auc = auc(mean_fpr, mean_tpr)
std_auc = np.std(aucs)
ax.plot(mean_fpr, mean_tpr, color='b',
        label=r'Mean ROC (AUC = %0.2f  $\pm$  %0.2f)' % (mean_auc, std_auc),
        lw=2, alpha=.8)

mean_fprTree = mean_fpr
mean_tprTree = mean_tpr

std_tpr = np.std(tprs, axis=0)
tprs_upper = np.minimum(mean_tpr + std_tpr, 1)
tprs_lower = np.maximum(mean_tpr - std_tpr, 0)
ax.fill_between(mean_fpr, tprs_lower, tprs_upper, color='grey', alpha=.2,
                label=r' $\pm$  1 std. dev.')

ax.set(xlim=[-0.05, 1.05], ylim=[-0.05, 1.05],
        title="ROC K-
блочная стратифицированная перекрестная проверка: Алгоритм деревьев решений")
ax.legend(loc="lower right")
plot.show()

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import BaggingClassifier
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.model_selection import cross_val_score
from sklearn.ensemble import VotingClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import GridSearchCV
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import precision_recall_curve
from sklearn.metrics import plot_precision_recall_curve
from sklearn.metrics import average_precision_score
from sklearn.model_selection import StratifiedKFold
from sklearn.metrics import plot_roc_curve
from sklearn.metrics import auc
from sklearn.metrics import recall_score
from sklearn.metrics import precision_score

```

```

from mlxtend.feature_selection import SequentialFeatureSelector as sfs
from mlxtend.plotting import plot_sequential_feature_selection as plot_sfs
from joblib import dump, load

def evaluate_sp_sn(confusion_matrix):
    conf_mat = np.ravel(confusion_matrix)
    specificity = conf_mat[3] / (conf_mat[1] + conf_mat[3])
    sensitivity = conf_mat[0] / (conf_mat[0] + conf_mat[2])
    return sensitivity, specificity

#####
###
#Read data from the source
#####
###
# df_x = pd.read_excel ('data_noage.xlsx', sheet_name='X')
# df_y = pd.read_excel ('data_noage.xlsx', sheet_name='y')
# y= df_y["race"].values

df_x = pd.read_excel ('tb_data_new.xlsx', sheet_name='X')
df_y = pd.read_excel ('tb_data_new.xlsx', sheet_name='y')
y= df_y["RELAPSE"].values
df_x.fillna(0, inplace=True)

X_tmp = df_x.values
names = list(df_x)

#X = X_tmp.copy()
scaler = StandardScaler()
X = scaler.fit_transform(X_tmp)

num_features = 10
#####
###
#Prepare train and test datasets
#####
###

#X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)

#####
###
#Train bagging classifier
#####
###
if 0:
    #clf = BaggingClassifier(base_estimator=SVC(kernel='rbf'),
    #                       n_estimators=10, max_samples=0.7, random_state=0).fit(X_test,
y_test)
    clf = BaggingClassifier(base_estimator=SVC(C=0.5, kernel='linear',
class_weight="balanced", gamma='scale'),
                           n_estimators=20, max_samples=1.0, random_state=0)
    random_state = np.random.RandomState(0)
    n_splits=5
    cv = StratifiedKFold(n_splits=n_splits)
    classifier = clf

    tprs = []
    aucs = []
    precs = []
    recs = []
    confusion_matrices = []
    c_reps = []
    classification_reports = []
    mean_fpr = np.linspace(0, 1, 100)

    fig, ax = plt.subplots()
    for i, (train, test) in enumerate(cv.split(X, y)):
        classifier.fit(X[train], y[train])
        viz = plot_roc_curve(classifier, X[test], y[test],

```

```

        name='ROC fold {}'.format(i),
        name='_nolegend_',
        alpha=0.3, lw=1, ax=ax)
    interp_tpr = np.interp(mean_fpr, viz.fpr, viz.tpr)
    interp_tpr[0] = 0.0
    tprs.append(interp_tpr)
    aucs.append(viz.roc_auc)
    precs.append(precision_score(y[test], classifier.predict(X[test]),
average='weighted'))
    recs.append(recall_score(y[test], classifier.predict(X[test]),
average='weighted'))
    confusion_matrices.append(confusion_matrix(y[test], classifier.predict(X[test]),
labels=[0, 1]))
    c_reps.append(classification_report(y[test], classifier.predict(X[test]),
target_names=["Control", "Relapse"]))

    ax.plot([0, 1], [0, 1], linestyle='--', lw=2, color='r',
            label='Chance', alpha=.8)
    print("Average Recall", np.mean(recs, axis=0))
    print("Average Precision", np.mean(precs, axis=0))
    c_sum=np.zeros_like(confusion_matrices[0])
    for i in range (n_splits):
        c_sum=c_sum+confusion_matrices[i]
    sn, sp = evaluate_sp_sn(c_sum)
    print(f'sensitivity= {sn:.2f}, specificity= {sp:.2f}')
    mean_tpr = np.mean(tprs, axis=0)
    mean_tpr[-1] = 1.0
    mean_auc = auc(mean_fpr, mean_tpr)
    std_auc = np.std(aucs)
    ax.plot(mean_fpr, mean_tpr, color='b',
            label=r'Mean ROC (AUC = %0.2f $\pm$ %0.2f)' % (mean_auc, std_auc),
            lw=2, alpha=.8)

    std_tpr = np.std(tprs, axis=0)
    tprs_upper = np.minimum(mean_tpr + std_tpr, 1)
    tprs_lower = np.maximum(mean_tpr - std_tpr, 0)
    ax.fill_between(mean_fpr, tprs_lower, tprs_upper, color='grey', alpha=.2,
                    label=r'$\pm$ 1 std. dev.')

    ax.set(xlim=[-0.05, 1.05], ylim=[-0.05, 1.05],
           title="Receiver operating characteristic example")
    ax.legend(loc="lower right")
    plt.show()
    plt.close()
    print(np.mean(confusion_matrices,axis=0))
    print(np.std(confusion_matrices,axis=0))
# Build step forward feature selection
    sfs_forward = sfs(clf,
                      k_features=num_features,
                      forward=True,
                      floating=False,
                      verbose=2,
                      scoring='accuracy',
                      cv=5,
                      n_jobs=-1)

# Perform SFFS
    sfs_forward = sfs_forward.fit(X_train, y_train, custom_feature_names=names)
# Which features?
    feat_cols = list(sfs_forward.k_feature_idx_)
    print(feat_cols)
    print(sfs_forward.k_feature_names_)
    df1 = pd.DataFrame.from_dict(sfs_forward.get_metric_dict()).T
    df1.to_excel("output_sfs_forward.xlsx")

# Build step backward feature selection
    sfs_backward = sfs(clf,
                      k_features=num_features,
                      forward=False,
                      floating=False,
                      verbose=2,

```

```

        scoring='accuracy',
        cv=5,
        n_jobs=-1)

# # Perform SFFS
#     sfs_backward = sfs_backward.fit(X_train, y_train, custom_feature_names=names)
# # Which features?
#     feat_cols = list(sfs_backward.k_feature_idx_)
#     print(feat_cols)
#     print(sfs_backward.k_feature_names_)
#     df1 = pd.DataFrame.from_dict(sfs_backward.get_metric_dict()).T
#     df1.to_excel("output_sfs_backward.xlsx")

#     clf.fit(X_train, y_train)
#     print('bagging classifier accuracy:', clf.score(X_test,y_test))

if 1:
#     dump(clf, 'BaggingClassifier.joblib')

#     clf1 = DecisionTreeClassifier(max_depth=4)
#     clf2 = SVC(kernel='rbf', probability=True)
#     eclf = VotingClassifier(
#         estimators=[('dt', clf1), ('svm', clf2)],
#         voting='soft', weights=[1, 2])

#     for clf, label in zip([clf1, clf2, eclf], ['DT', 'SVM', 'Ensemble']):
#         scores = cross_val_score(clf, X, y, scoring='accuracy', cv=5)
#         print("Accuracy: %0.2f (+/- %0.2f) [%s]" % (scores.mean(), scores.std(),
label))
        clf3 = SVC(kernel='linear', class_weight="balanced").fit(X_train, y_train)
        #dump(clf3, 'svm.joblib')
        scores = cross_val_score(clf3, X, y, scoring='accuracy', cv=5)
        print("Accuracy: %0.2f (+/- %0.2f)" % (scores.mean(), scores.std()))
        print("Score: %0.2f" % clf3.score(X_test,y_test))
        features = names
        importances = abs(clf3.coef_[0])
        sum_importances = np.sum(importances)
        importances /= sum_importances
        indices = (importances).argsort()[::-1]
        y_pred = clf3.predict(X_test)
        print(confusion_matrix(y_test, y_pred, labels=[0, 1]))
        sn, sp = evaluate_sp_sn(confusion_matrix(y_test, y_pred, labels=[0, 1]))
        print(f'sensitivity= {sn}, specificity= {sp}')
        print(classification_report(y_test, y_pred, target_names=["Control", "Relapse"]))

        plt.bar(range(len(indices[:num_features])), importances[indices[:num_features]],
color='b', align='center')
        plt.xticks(range(len(indices[:num_features])), [features[i] for i in
indices[:num_features]])
plt.xlabel('Признаковые переменные')
        plt.ylabel('Относительная важность')
plt.show()
        plt.close()
        #plt.title('')
        #disp = plot_precision_recall_curve(clf3, X_test, y_test)
        #disp.ax_.set_title('2-class Precision-Recall curve: '
#         'AP={0:0.2f}'.format(average_precision))

print("done")

```