

Опыт разработки и внедрения программного обеспечения на основе искусственного интеллекта в анализе электронных медицинских карт

Александр Гусев

к.т.н., директор по развитию Webiomed,
эксперт по искусственному интеллекту ЦНИИОИЗ Минздрава России,
ст. научный сотрудник НПКЦ ДиТ ДМЗ

Платформа прогнозной аналитики Webiomed



Анализ обезличенных медицинских данных

Автоматический анализ медицинских данных, включая извлечение информации из неструктурированных врачебных записей с помощью NLP-технологий



Искусственный интеллект

Сбор больших данных и машинное обучение для выявления подозрений на заболевания и глубокого интеллектуального анализа сведений о пациенте



Прогнозная аналитика

Поддержка принятия управленческих и врачебных решений на основе персональной оценки риска и точных прогнозов возможного ухудшения здоровья пациента в будущем



Рекомендации врачу и пациенту

Персональные советы врачу и пациенту по профилактике заболеваний, сформированные на основе утвержденных клинических рекомендаций

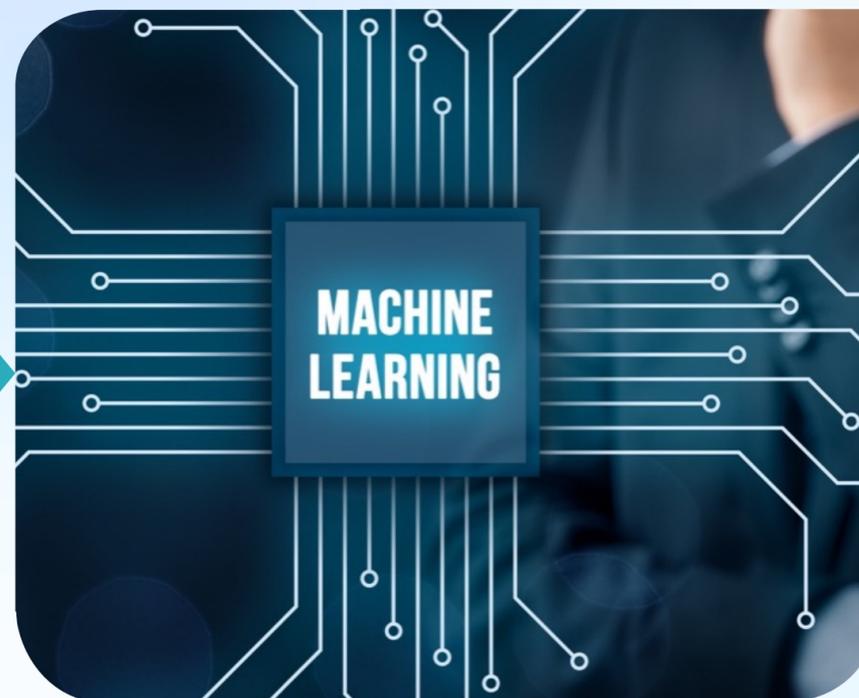


Технологические этапы работы платформы



#1. Извлечение данных

Процесс, используемый для превращения необработанных больших (часто неструктурированных и/или несвязанных) данных в полезную информацию



#2. Применение моделей

Применение технологий машинного обучения для создания моделей анализа имеющихся данных с целью решения диагностической или прогностической задачи



#3. Визуализация

Комплекс методик, практик и программного обеспечения к представлению результатов обработки данных путем максимально эффективного их использования

#1. Извлечение

Анализ неструктурированных ЭМК для
получения из них пригодных данных

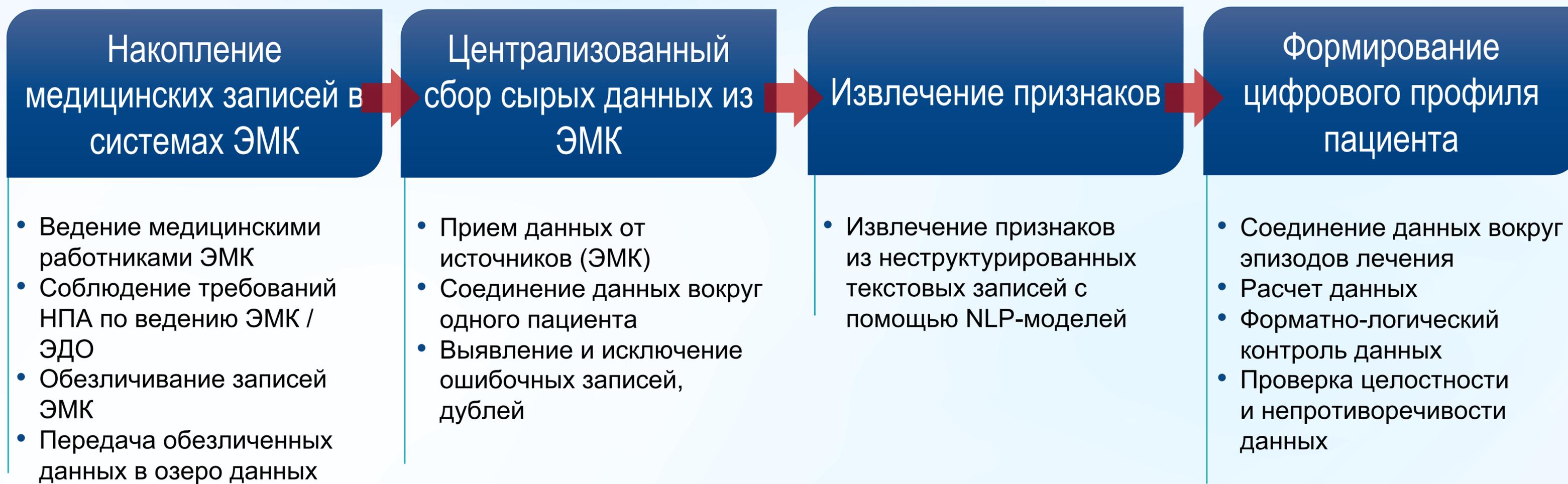
Извлечение данных из обезличенных ЭМК

- ✓ До 80% клинически-значимой информации хранится в ЭМК в неструктурированных текстовых записях
- ✓ Webiomed автоматически извлекает из ЭМК нужные данные (признаки) с помощью NLP-сервиса
- ✓ Врачам не нужно заполнять специальные экранные формы с множеством полей или вести отдельные регистры и мониторинги
- ✓ Всю нагрузку по извлечению информации из ЭМК Webiomed берет на себя
- ✓ Этим мы экономим время врача на приеме, давая возможности больше заниматься пациентом



2923

Поддерживаемых признаков



#2. Модели МО

Пример: прогнозирование выживаемости в течение 1 года при раке легкого

Формирование набора данных

Для разработки модели прогнозирования выживаемости пациента в течение 1 года при раке легкого был создан набор данных реальной клинической практики на основе базы данных обезличенных электронных медицинских карт Webiomed



* Вся информация собирается строго в обезличенном виде
 * Для сбора данных с операторами ЭМК были подписаны соответствующие соглашения, разрешающие проведение научно-исследовательских работ на основе полученных обезличенных ЭМК

* Извлечение данных осуществлялось с помощью технологий NLP, а также программным способом с применением расчетных формул и маппинга значений для лабораторной диагностики

* Критерии включения (формирования) набора:

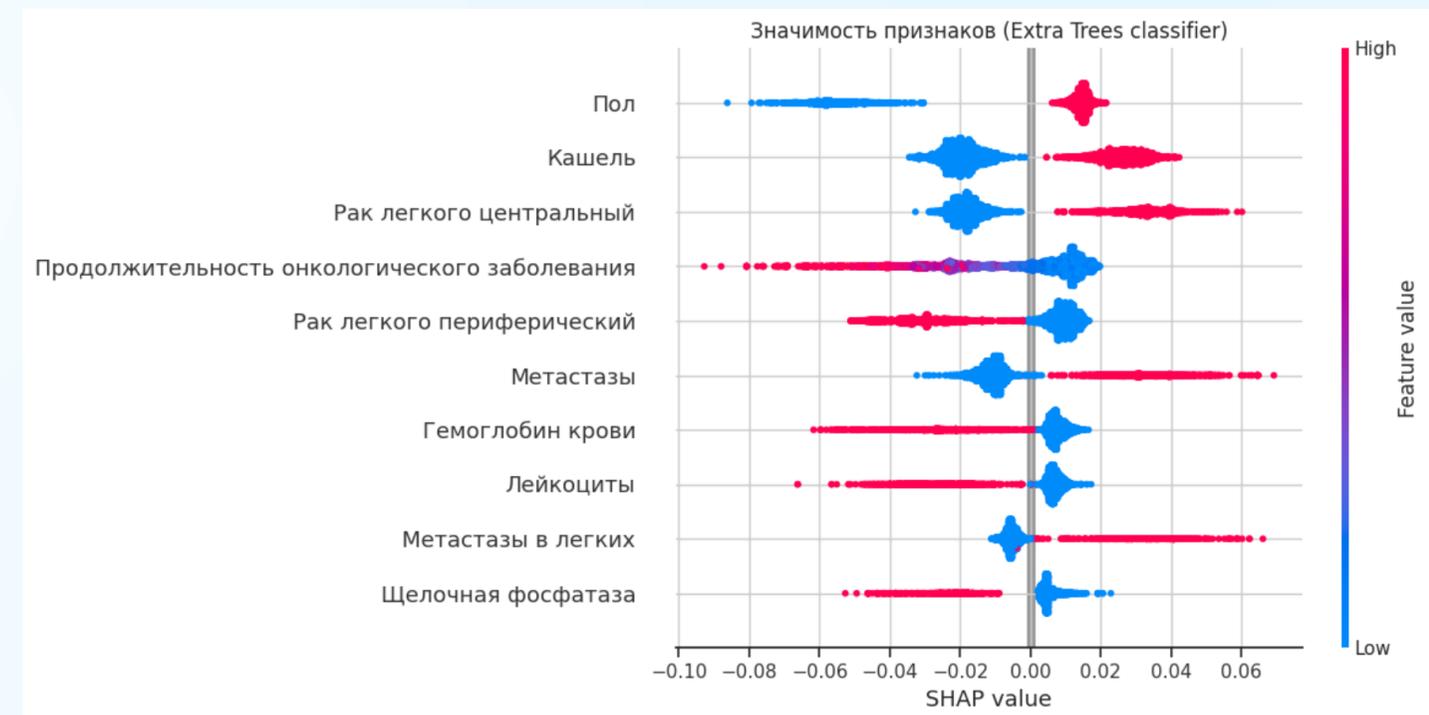
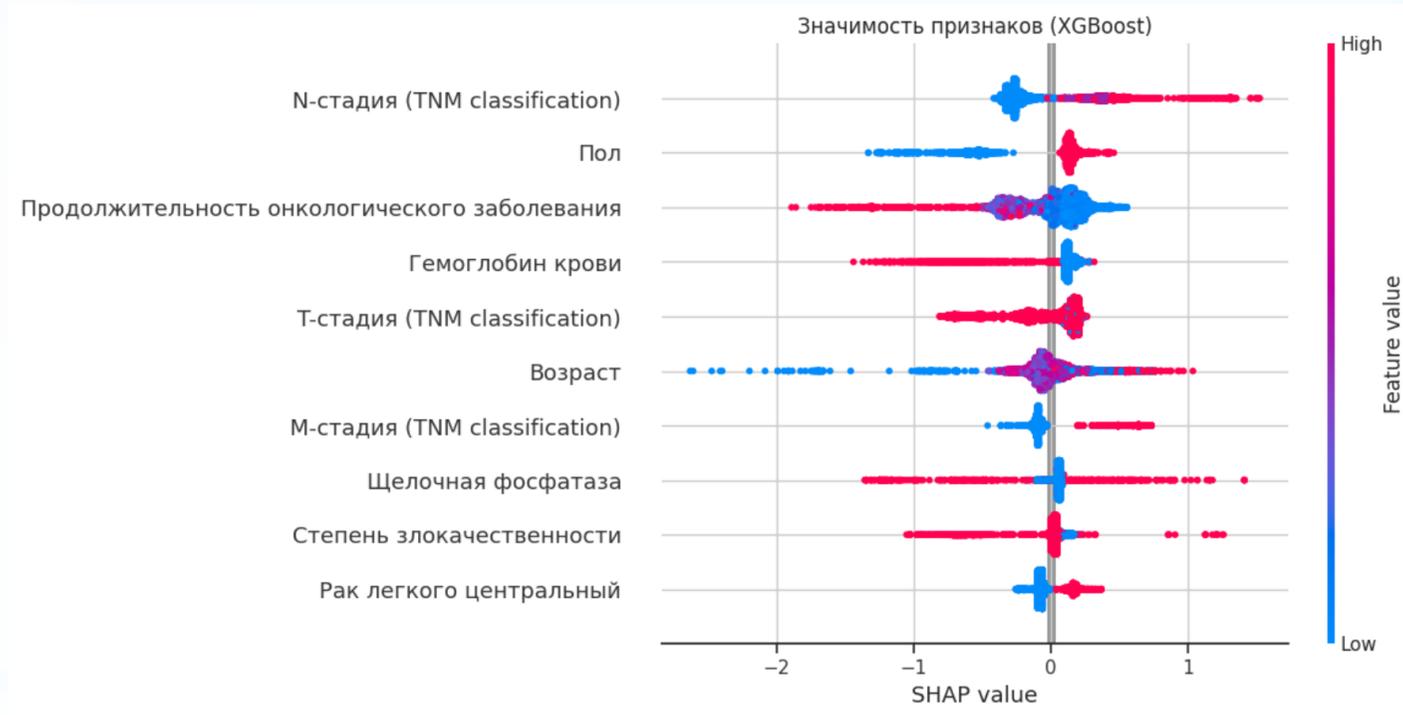
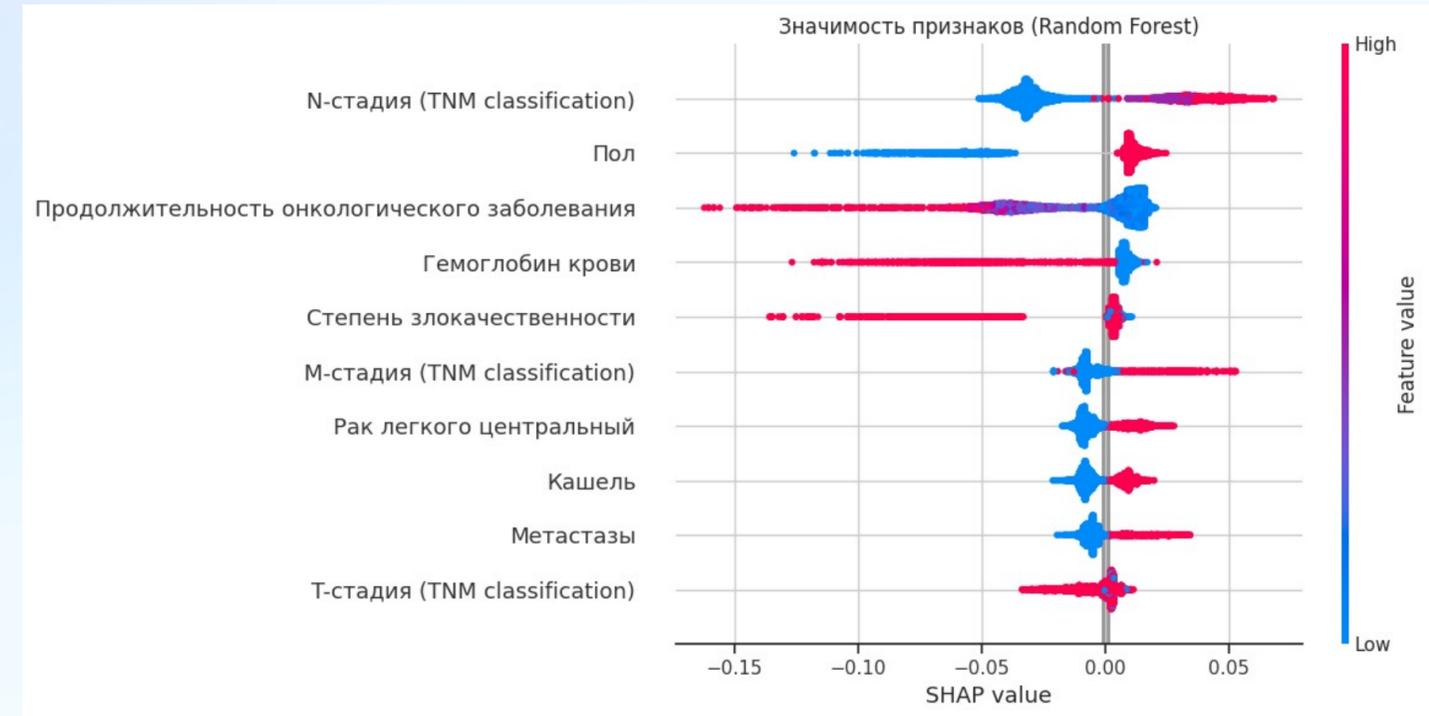
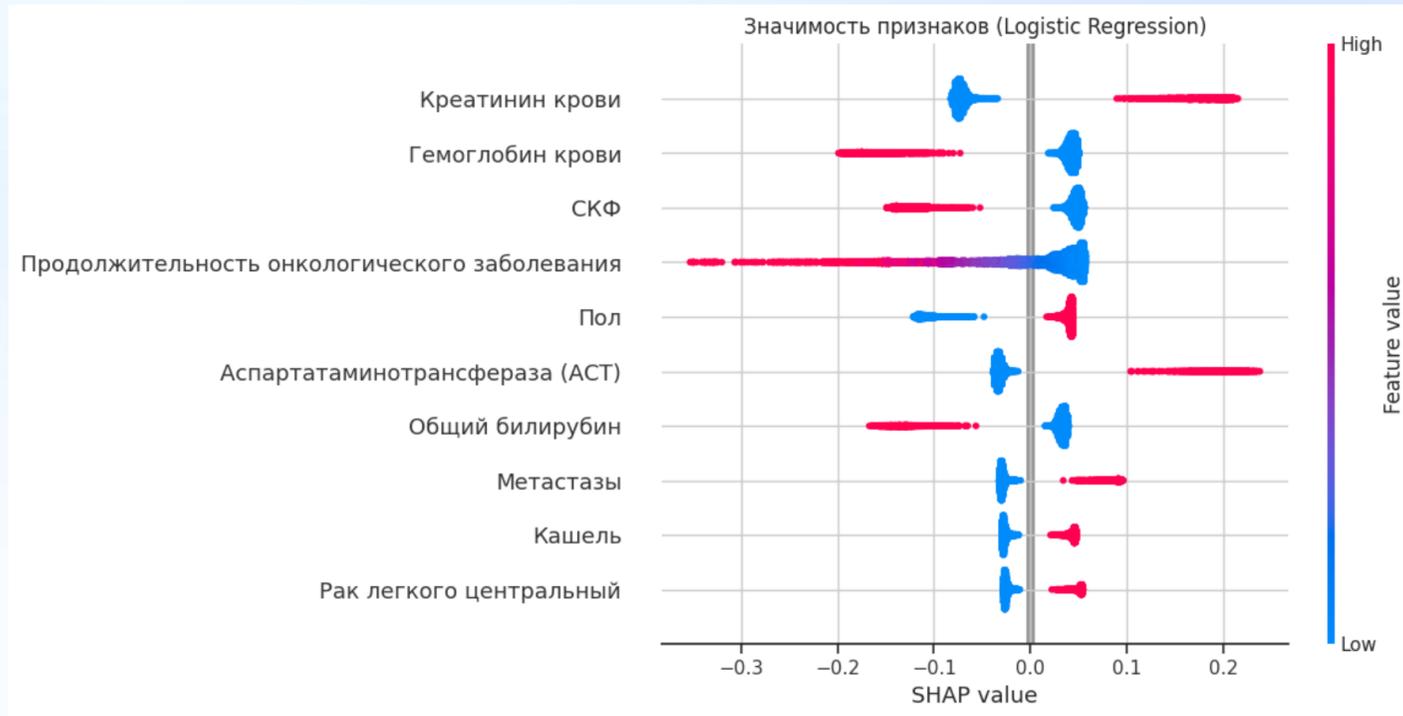
- Возраст от 18 лет
- Наличие медицинского случая с диагнозом «рак легкого»: С34, С34.0, С34.1, С34.2, С34.3, С34.8, С34.9

* 1 класс НД – смерть в течение года от даты прогноза по причине, связанной с раком легкого
 * 0 класс НД – обращение за медицинской помощью через 1 год после прогноза (факт нахождения в живых)

* Критерии исключения:

- Записи с пропущенными значениями
- Записи с ошибками в данных

Поиск предикторов (feature selection)



Поиск лучшего алгоритма машинного обучения

№	Модель	Чувствительность (Recall)	Специфичность	Прогностическая ценность K1 (Precision)	Прогностическая ценность K0	F1 (положительный)	F1 (отрицательный)	ROC-AUC	Точность (accuracy)	Порог классификации (Индекс Юдена)
0	Logistic Regression	0.76	0.55	0.27	0.91	0.4	0.69	0.7	0.59	0.46
1	Multi Layer Perceptron	1.0	0.0	0.18	nan	0.3	nan	0.66	0.18	0.00
2	Random Forest	0.56	0.8	0.38	0.89	0.45	0.84	0.73	0.76	0.80
3	Extra Trees classifier	0.69	0.64	0.29	0.91	0.41	0.75	0.71	0.65	0.49
4	XGBoost	0.66	0.76	0.38	0.91	0.48	0.83	0.78	0.74	0.22

Лучшие результаты получены у модели XGBoost

Точность 74%, чувствительность 66%, специфичность 76%

Вход модели: 72 признака, в т.ч.: возраст, СКФ, ФВЛЖ, гемоглобин, лейкоциты и т.д.

Выход модели: вероятность смертельного исхода от рака легкого в течение года

Интерпретация: threshold = 0.22, > высокий риск, < низкий

#3. Визуализация

Встраивание моделей в программное обеспечение платформы, вывод результатов

Система поддержки принятия врачебных решений

11



- ✓ **Автоматический анализ** обезличенной электронной медицинской карты
- ✓ **Выявление подозрений** на пропущенные врачом заболевания
- ✓ **Выявление факторов риска**
- ✓ **Прогнозирование возможных негативных событий** в здоровье пациента, включая ухудшение имеющихся заболеваний, госпитализацию или смерть
- ✓ Комплексная **оценка риска** пациента
- ✓ Персональные **клинические рекомендации** для врача и пациента

WebioMed
Номер запроса 191671
Возраст 74 года
Пол Женский
Уровень внимания Высокий
Дата оценки 28.07.2021, 13:20

Полнота обследования (от 0 до 100%)
0%

Исходные данные >
Показать накопленные данные >
Оставить отзыв >

Обязательно нужно обратить внимание
Ничего критичного
Система ничего опасного не выявила
Оценка недоступна

Подозрения 3
• Анемия неуточненная D64.9
• Витамин B12-дефицитная анемия D51
• Фолиеводефицитная анемия D52

Предупреждения 3
• Повышенные цифры АД (выявлено АД больше 130/85 мм рт.ст.)
• Пациент приоритета 3 уровня внимания по диспансеризации (не обращался в медицинскую организацию более 2 лет)

Заболевания 5
• Железodefицитная анемия D50.0-D50.9
• ИБС I20-I25
• ОРВИ J00, J02, J04-J06

Оценка рисков 6
Сердечно-сосудистые заболевания **Очень высокие**
Сахарный диабет **Низкие**
Лёгочные заболевания **Высокие**
Инфекционные заболевания **Высокие**
Акушерские и гинекологические заболевания **Низкие**
Наркологические заболевания **Низкие**

Факторы риска 10
• Увеличенная ЧСС в покое
• Повышенный уровень глюкозы крови
• Кровотечения в анамнезе
• Абдоминальное ожирение

Симптомы 2
• Анемия
• Гиперхромия

Лекарства
Не выявлены

Рекомендации пациенту
В связи с очень высоким риском сердечно-сосудистых заболеваний рекомендовано : Диета: Ограничение жиров менее 30%, насыщенных жиров менее 10% и обогащение диеты зерновыми, овощами, фруктами, рыбой. Физическая активность: 2,5-5 часов умеренной физической нагрузки в неделю, или 30-60 минут в день. Соблюдать вес в диапазоне ИМТ 18-25 кг/м2. Контроль АД, при АД 140/90 мм рт.ст. и выше обратиться к врачу. Контролировать холестерин и другие показатели липидов в крови, целевые уровни: холестерин менее 4,0, ЛПНП менее 1,5 ммоль/л. Контролировать сахар крови.



Включено в реестр российского ПО



Регистрационное удостоверение Росздравнадзора

Комплексная оценка риска пациента

 Оценка рисков
Закреть

 Сердечно-сосудистые заболевания

Очень высокие

 Сахарный диабет

Высокие

 Лёгочные заболевания

Высокие

 Инфекционные заболевания

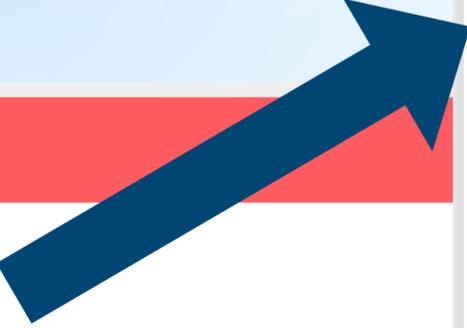
Высокие

 Акушерские и гинекологические заболевания

Низкие

 Наркологические заболевания

Высокие



 Оценка рисков
Закреть

 Сердечно-сосудистые заболевания

Очень высокие

 МЕТОДИКИ ОЦЕНКИ

Сердечно-сосудистый риск у пациентов с АГ 3 стадии: **ОЧЕНЬ ВЫСОКИЙ**
 Клинические рекомендации "Артериальная гипертензия у взрослых", МЗ РФ 2020 г.

AI Риск смерти от ИБС и инсульта в течение 10 лет: **ОЧЕНЬ ВЫСОКИЙ**
 Модель Webiomed WML.CVD.Score - 38.4%

Суммарный сердечно-сосудистый риск: **ОЧЕНЬ ВЫСОКИЙ**
 Клинические рекомендации «Диагностика и коррекция нарушений липидного обмена с целью профилактики и лечения атеросклероза» VI пересмотра

Шкала PROCAM: **ВЫСОКИЙ**
 10 летний риск развития коронарных осложнений - 20-29%

AI Риск развития ССЗ в течение 10 лет: **ВЫСОКИЙ**
 Модель Webiomed WML.CVD.FRS - 27.8%

AI Прогнозирование госпитализации пациентов с заболеваниями сердца в течение ближайших 12-ти месяцев: **ВЫСОКИЙ**
 Модель Webiomed WML.CVD.Hospitalization

Фрамингемская шкала: **НИЗКИЙ**
 10 летний риск развития сердечно-сосудистых заболеваний (инфаркт миокарда, инсульт, фатальное ССЗ) - 6%

Шкала UKPDS Cardiac Risk: **НЕ ОПРЕДЕЛЕН**
 10 летний риск развития ИБС у пациентов с СД 2 типа
 Для определения риска недостаточно данных: Доля гликированного гемоглобина (HbA1c), Микроальбумин в моче

Шкала CART: **НЕ ОПРЕДЕЛЕН**
 Риск остановки сердца у госпитализированных пациентов
 Шкала не была применена: применяется только при нахождении пациента в стационаре

Шкала SCORE (абсолютный риск): **НЕ ОПРЕДЕЛЕН**
 10-летний риск фатального ССЗ
 Шкала не была применена: предназначена для пациентов без ССЗ и сахарного диабета, от 40 до 70 лет

Шкала SCORE (относительный риск): **НЕ ОПРЕДЕЛЕН**
 10-летний относительный риск фатального ССЗ
 Шкала не была применена: предназначена для пациентов от 18 до 39 лет

Шкала "сосудистого возраста": **НЕ ОПРЕДЕЛЕН**
 Шкала не была применена: предназначена для пациентов без ССЗ и сахарного диабета, от 40 до 65 лет

Шкала CHA₂DS₂-VASc: **НЕ ОПРЕДЕЛЕН**
 Риск развития инсульта и тромбэмболических осложнений у пациентов с трепетанием и фибрилляцией предсердий
 Шкала не была применена: предназначена для пациентов старше 18 лет с установленным диагнозом «Фибрилляция предсердий» (коды МКБ-10 I48)

Оценка рисков

Сердечно-сосудистые заболевания **Высокие**

МЕТОДИКИ ОЦЕНКИ

Сердечно-сосудистый риск у пациентов с АГ: ВЫСОКИЙ
Клинические рекомендации "Артериальная гипертензия у взрослых" Минздрава РФ 2020 г.

Суммарный сердечно-сосудистый риск: ВЫСОКИЙ
Клинические рекомендации «Диагностика и коррекция нарушений липидного обмена при атеросклероза» VI пересмотра

AI Риск развития фибрилляции предсердий в течение года: ВЫСОКИЙ
Модель WML.Prognosis.General.Fibrillation_One_Year - 79%

Шкала CART: НИЗКИЙ
Риск остановки сердца у госпитализированных пациентов

AI Риск госпитализации в течение ближайших 12 месяцев у пациентов с сердечно-сосудистыми заболеваниями: НИЗКИЙ
Модель Webiomed WML.CVD.Hospitalization - 1%

AI Риск наличия тромбоза легочной артерии (ТЭЛА): НИЗКИЙ
Модель WML.Diagnosis.General.Pulmonary_Embolism - 15%

Шкала SCORE (относительный риск): НЕ ОПРЕДЕЛЕН
10-летний относительный риск фатального ССЗ
Для определения риска недостаточно данных: Общий холестерин

Фрамингемская шкала: НЕ ОПРЕДЕЛЕН
Шкала не была применена: предназначена для пациентов от 20 до 79 лет

Шкала PROCAM: НЕ ОПРЕДЕЛЕН
10 летний риск развития коронарных осложнений
Шкала не была применена: предназначена для пациентов от 20 до 80 лет

Шкала UKPDS Cardiac Risk: НЕ ОПРЕДЕЛЕН
Шкала не была применена: предназначена для пациентов с установленным сахарным диабетом (коды МКБ-10 E11-E11.9)

Шкала HAS-BLED: НЕ ОПРЕДЕЛЕН
Риск развития кровотечений у пациентов с фибрилляцией предсердий и получающих антикоагулянты
Шкала не была применена: предназначена для пациентов с фибрилляцией предсердий или принимающих антикоагулянты.

AI Риск развития ССЗ в течение 10 лет: НЕ ОПРЕДЕЛЕН
Модель Webiomed WML.CVD.FRS
Модель не была применена: предназначена для пациентов от 44 до 81 года

AI Риск развития фибрилляции предсердий в течение года: ВЫСОКИЙ
Модель WML.Prognosis.General.Fibrillation_One_Year - 79%

AI Риск наличия тромбоза легочной артерии (ТЭЛА): НИЗКИЙ
Модель WML.Diagnosis.General.Pulmonary_Embolism - 15%

Рекомендации врачу

Рекомендовать отказаться от курения.

В связи с очень высоким риском сердечно-сосудистых заболеваний рекомендовано : Немедикаментозные методы профилактики. Назначить: 1. Дезагреганты 2. Статины, целевые уровни: холестерин менее 4,0, ЛПНП менее 1,5 ммоль/л 3. Ингибитор АПФ (сартан при непереносимости) В сложной клинической ситуации консультация кардиолога, липидолога, гематолога и пр.

В связи с высоким риском потенциально тяжелого течения COVID рекомендовано: При необходимости - вакцинирование в первую очередь. Разъяснение правил социального дистанцирования, использования СИЗ, проверка соблюдения рекомендаций.

В связи с высоким риском инфекционно-воспалительного синдрома на основании анализа крови необходим поиск источника инфекционного процесса, дополнительное тестирование на инфекционно-воспалительный процесс.

В связи с очень высоким риском синдрома системной воспалительной реакции взрослых (SIRS) рекомендовано: Обследуйте пациента для исключения инфекционной этиологии: оценка гемодинамики, скрининг на уровень лактата, бактериологический посев крови. При выявлении сепсиса или септического шока (гемодинамическая нестабильность, лактатацидоз, полиорганная дисфункция) немедленно начните антибиотики широкого спектра действия, внутривенными жидкостями и вазопрессорами.

В связи с высоким риском возникновения осложнений родоразрешения рекомендована «третья А группа акушерских стационаров» для проведения родоразрешения: акушерские стационары, имеющие в своем составе отделение анестезиологии-реаниматологии для женщин, отделение реанимации и интенсивной терапии для новорожденных, отделение патологии новорожденных и недоношенных детей (II этап выхаживания).

В связи с высоким общим риском по наркологической патологии совет по ограничению употребления алкоголя и короткое консультирование, направленное на ограничение употребления алкоголя, наблюдение.

Есть вопросы? Пожалуйста, обращайтесь по контактам ниже, постараюсь ответить

Мои контакты

Гусев Александр,
к.т.н., директор по развитию
ООО «К-Скай»

 agusev@webiomed.ru

 +7 (911) 402-35-00



Контакты компании

 **Сайт**
<https://webiomed.ru>

 **ВКонтакте**
<https://vk.com/webiomed>

 **Telegram**
<https://t.me/webiomed>

 **YouTube**
<https://www.youtube.com/>



W E B I O M E D