



Машинное обучение в прогнозировании рецидивов у больных туберкулезом с множественной лекарственной устойчивостью

А. С. АЛЛИЛУЕВ^{1,2}, О. В. ФИЛИНЮК¹, Е. Е. ШНАЙДЕР³, С. В. АКСЕНОВ^{1,3,4}

¹ФГБОУ ВО «Сибирский государственный медицинский университет МЗ РФ», г. Томск, РФ

²ОГАУЗ «Томский фтизиопульмонологический медицинский центр», г. Томск, РФ

³ФГАОУ ВО «Национальный исследовательский Томский политехнический университет», г. Томск, РФ

⁴ФГАОУ ВО «Национальный исследовательский Томский государственный университет», г. Томск, РФ

РЕЗЮМЕ

Цель исследования: оценить возможность применения алгоритмов машинного обучения в прогнозировании рецидива у больных туберкулезом с множественной лекарственной устойчивостью (МЛУ-ТБ).

Материалы и методы. Проведен анализ клинико-эпидемиологических, возрастно-половых, социальных, медико-биологических параметров и особенностей химиотерапии у 346 излеченных пациентов с МЛУ-ТБ. При построении моделей прогнозирования использовались инструменты библиотеки scikit-learn Version 0.24.2 в интерактивной облачной среде работы с программным кодом Google Colaboratory.

Результаты. Анализ характеристик моделей прогнозирования рецидивов у излеченных больных МЛУ-ТБ с помощью алгоритмов машинного обучения, включающих дерево решений, случайный лес, 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). Установлены пять основных предикторов развития рецидивов у излеченных больных МЛУ-ТБ: неоднократные курсы химиотерапии; длительность стажа туберкулеза; деструктивный процесс в легких; общий срок лечения менее 22 мес. и использование в схеме химиотерапии менее пяти эффективных противотуберкулезных препаратов.

Ключевые слова: туберкулез, машинное обучение, множественная лекарственная устойчивость, рецидив, факторы риска, прогноз

Для цитирования: Аллилуев А. С., Филинчук О. В., Шнайдер Е. Е., Аксенов С. В. Машинное обучение в прогнозировании рецидивов у больных туберкулезом с множественной лекарственной устойчивостью // Туберкулез и болезни лёгких. – 2021. – Т. 99, № 11. – С. 27-34. <http://doi.org/10.21292/2075-1230-2021-99-11-27-34>

Machine Learning for Prediction of Relapses in Multiple Drug Resistant Tuberculosis Patients

A. S. ALLILUEV^{1,2}, O. V. FILINYUK¹, E. E. SHNAYDER³, S. V. AKSENOV^{1,3,4}

¹Siberian State Medical University, Tomsk, Russia

²Tomsk Phthisiopulmonology Medical Center, Tomsk, Russia

³Tomsk Polytechnic University, Tomsk, Russia

⁴National Research Tomsk State University, Tomsk, Russia

ABSTRACT

The objective of the study: to evaluate the possibility of using machine learning algorithms for prediction of relapses in multiple drug resistant tuberculosis (MDR TB) patients.

Subjects and Methods. Clinical, epidemiological, gender, sex, social, biomedical parameters and chemotherapy parameters were analyzed in 346 cured MDR TB patients. The tools of the scikit-learn library, Version 0.24.2 in the Google Colaboratory interactive cloud environment were used to build forecasting models.

Results. Analysis of the characteristics of relapse prediction models in cured MDR TB patients using machine learning algorithms including decision tree, random forest, gradient boosting, and logistic regression using K-block stratified validation revealed high sensitivity (0.74 ± 0.167 ; 0.91 ± 0.17 ; 0.91 ± 0.14 ; 0.91 ± 0.16 , respectively) and specificity (0.97 ± 0.03 ; 0.98 ± 0.02 ; 0.98 ± 0.02 ; 0.98 ± 0.02 , respectively).

Five main predictors of relapse in cured MDR-TB patients were identified: repeated courses of chemotherapy; length of history of tuberculosis; destructive process in the lungs; total duration of treatment less than 22 months; and use of less than five effective anti-TB drugs in the regimen of chemotherapy.

Key words: tuberculosis, machine learning, multiple drug resistance, relapse, risk factors, forecast

For citations: Alliluev A.S., Filinyuk O.V., Shnayder E.E., Aksenov S.V. Machine learning for prediction of relapses in multiple drug resistant tuberculosis patients. *Tuberculosis and Lung Diseases*, 2021, Vol. 99, no. 11, P. 27-34. (In Russ.) <http://doi.org/10.21292/2075-1230-2021-99-11-27-34>

Для корреспонденции:

Аллилуев Александр Сергеевич
E-mail: alliluev233@gmail.com

Correspondence:

Aleksandr S. Alliluev
Email: alliluev233@gmail.com

По данным информационного бюллетеня Всемирной организации здравоохранения, в Европе каждый пятый случай туберкулеза (ТБ) – это туберкулез с множественной лекарственной устойчивостью (МЛУ-ТБ), в РФ – каждый третий [11, 15].

При этом, несмотря на доступность противотуберкулезных препаратов (ПТП), результаты лечения в РФ остаются неоптимальными. Сегодня показатели эффективного лечения больных с МЛУ-ТБ составляют только 57% случаев в мире, в РФ –

55% [11]. Эффективное излечение МЛУ-ТБ не гарантирует, что через некоторое время не возникнет рецидив заболевания. За последние несколько лет все больше исследований показывают, что рецидивы МЛУ-ТБ нередки и зарегистрированные показатели рецидивов во всем мире варьируют от 2,5 до 8,5% при сроке наблюдения до 8 лет [12, 13]. Ведение пациентов с рецидивом МЛУ-ТБ представляет собой сложную задачу в связи с тем, что является длительным, осложняется побочными реакциями на принимаемые лекарственные препараты, дорогостоящим по сравнению с лечением лекарственно-чувствительного ТБ [7]. В связи с этим важно знать особенности, связанные с рецидивом МЛУ-ТБ, и факторы, позволяющие предсказать, какие пациенты имеют наибольший риск реактивации заболевания.

Одновременный прогресс в области инфраструктуры информационных технологий и роста мощности компьютеров позволяет сегодня анализировать любые наборы биомедицинских данных для решения задач здравоохранения, используя современные методы анализа, такие как методы машинного обучения (МО) [9]. 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 [5]. В итоге методы МО стали популярным инструментом для медицинских исследователей во всем мире. Уже есть исследования, где их использовали в прогнозировании и диагностике ТБ, а также в предсказании исходов заболевания [6].

Цель исследования: оценить возможность применения алгоритмов МО в прогнозировании рецидива у больных МЛУ-ТБ.

Материалы и методы

В исследовании проведен анализ данных 346 пациентов с МЛУ-ТБ, которые находились на лечении в Томском фтизиопульмонологическом медицинском центре в период с 01.01.2009 г. по 31.12.2011 г. и были эффективно излечены. Критериями включения в исследование явились: возраст более 18 лет; наличие излеченного ТБ органов дыхания с МЛУ микобактерий туберкулеза (МБТ); наличие микробиологического подтверждения устойчивости МБТ минимум к изониазиду и рифампицину (методами посева на плотные питательные среды Левенштейна – Йенсена). Критерии исключения: срок наблюдения после излечения менее 5 лет.

Сформировано две группы пациентов: группа 1 – пациенты, у которых наступил рецидив заболевания во время срока наблюдения (35 пациентов), группа 2 – пациенты с отсутствием рецидива МЛУ ТБ в срок наблюдения после излечения (311 пациентов).

Срок наблюдения для всех пациентов составил не менее 5 лет.

Создана исследовательская база с обезличенными данными пациентов со смешанными признаками – количественными и качественными. Анализировались клиничко-эпидемиологические, социальные, лабораторные, рентгеновские данные пациентов, а также схемы химиотерапии, изменения в лечении при появлении амплификации, приверженность пациентов лечению.

Выделение факторов риска рецидива МЛУ-ТБ для создания прогностических моделей на основе алгоритмов МО начиналось 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 проводились нормализация и стандартизация для численных данных, кодирование для категориальных данных. Данный этап был необходим, так как большинство алгоритмов МО способно работать только с числовыми данными. Изначально было выбрано 106 признаков (исключены дублирующие и несущественные, например гражданство), в итоге в анализ включено 85 признаков.

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

Для построения классифицирующих моделей 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 метода дерева решений можно отнести простоту в применении и интерпретации, что является важным для врачей-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 [1].

Модель случайного леса представляет собой ансамбль оценщиков, которыми являются деревья решений. Каждый из оценщиков настроен на наборе тренировочных образцов, выбранных случайным образом с возвратом из первоначальной выборки [10]. 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 [2, 3]. Для оценки важности признаков чаще используются подходы, основанные на уменьшении неоднородности, как и в предыдущем алгоритме, а также путем вычисления значаний Шепли [4].

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 [14]. Для объяснения важности влияния признака на прогноз рецидива имеет также знак, стоящий перед коэффициентом w_i в уравнении регрессии:

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

где $f(\text{Рецидив}|(x_1, x_2, \dots, x_n))$ – 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:

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

где f – признак, использующийся для расщепления признакового пространства, D_p и D_j – набор данных родительского и j -го дочернего узла соответственно, $I(D_p)$ – мера неоднородности в узле p , N_j и N_p – число образцов в родительском и j -м дочернем узле соответственно, $m = 2$ – число регионов, полученных при расщеплении.

Чувствительность и специфичность моделей классификаторов оценивались 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 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 отобраны с помощью сеточного поиска на наборе значений, приведенных в табл. 1.

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 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 методик 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, причем 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 рис. 1 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.

Дерево иллюстрирует ситуацию, когда имеют место наличие у пациентов несколько последовательных признаков, сопряженных с рецидивом МЛУ-ТБ, причем последующие решения включения признака в прогнозную модель основываются на результатах предыдущих. Дерево решений, представленное на рис. 1, изображено в виде графа, вершина – частные признаки, а ветви дерева – различные характеристики пациента, которые могут иметь место в определенном частном состоянии. Каждой ветви присваиваются различные число-

Таблица 1. Гиперпараметры классификаторов, выбранные с помощью сеточного поиска
Table 1. Hyperparameters of classifiers selected with grid search

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
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
Логистическая регрессия	Коэффициент L1-регуляризации α : 0,01; 0,1; 0,0; 1,0; 10,0	0,0
	Максимальная глубина дерева: 3-10	3
	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

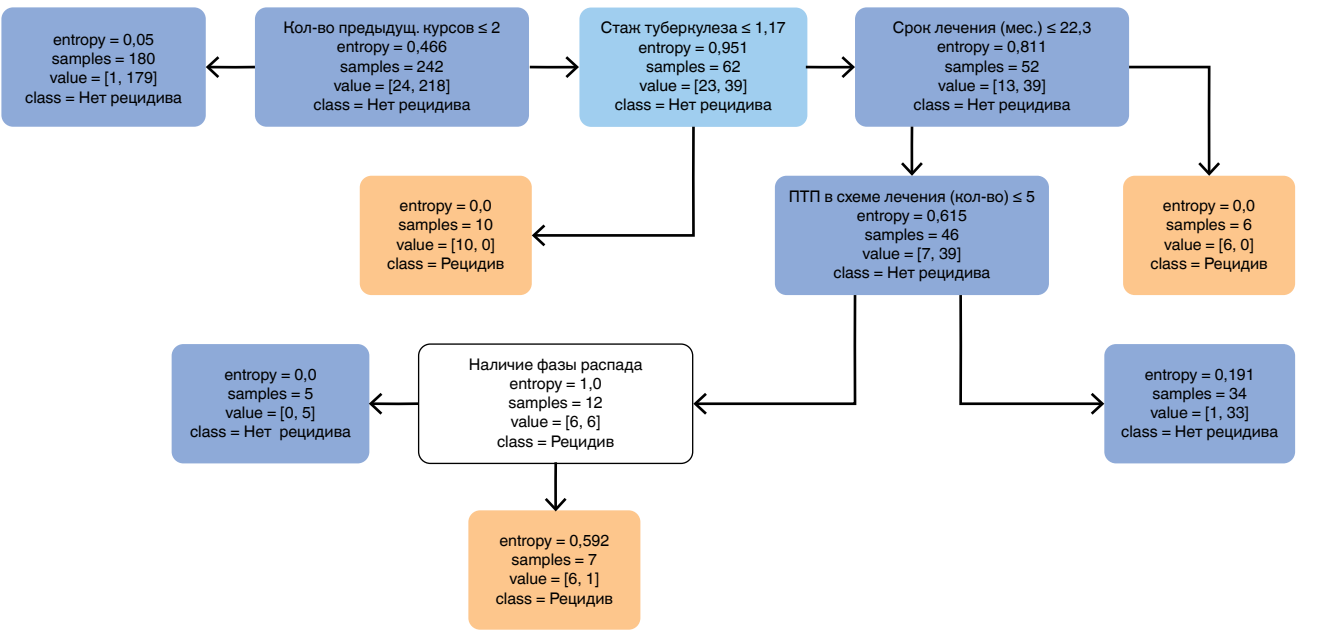


Рис. 1. Дерево решений прогнозирования рецидива МЛУ-ТБ. Entropy – степень неоднородности узла, samples – количество образцов в узле, value – список количества примеров, принадлежащих предсказываемым классам (первое значение – количество пациентов с рецидивом в узле, второе значение – количество пациентов без рецидива), class – предсказанный класс

Fig. 1. MDR TB relapse prediction decision tree. Entropy – degree of node heterogeneity, samples – number of samples in the node, value – list of the number of samples belonging to the predicted classes (the first value is the number of patients with recurrence in the node, the second value is the number of patients without recurrence), class – predicted class

вые характеристики признака с указанием степени неоднородности распределения признака в ветви дерева (entropy), количества примеров, принадлежащих предсказываемым классам (количество

пациентов с рецидивом или без рецидива в ветви дерева – value), и непосредственно прогнозируемого состояния (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 ТБ чуть более года (случаи впервые выявленного МЛУ-ТБ), а также случаи повторного лечения, но не более двух эффективных курсов химиотерапии ТБ, не имеют рецидива МЛУ-ТБ как минимум в течение 5 лет наблюдения. Однако если у пациентов наблюдались неоднократные курсы химиотерапии (два и более), диагностирован деструктивный ТБ, общий срок лечения не достигал 22,3 мес., а в схеме химиотерапии присутствовало 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 тестирующем 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 4-му режиму. Кроме того, классификатор не выявил глубинной связи между рецидивом МЛУ-ТБ и коморбидностью у пациентов в виде язвенной болезни желудка и двенадцатиперстной кишки, сахарного диабета и сердечно-сосудистых заболеваний, а также зависимости от употребления психоактивных веществ и табакокурения.

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.

Во втором эксперименте для демонстрации надежности выявленных прогностических моделей использована стратифицированная К-блочная перекрестная проверка для каждого метода МО, включенного в исследование [8]. Всего исследуемый набор данных поделен на 5 блоков, используемых при построении набора классификаторов. Для моделей, использующих деревья решений, снова 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.

На рис. 2 показаны гистограммы распределения значимости признаков для каждого из выбранных в исследовании алгоритмов МО. Всего на рис. 2 представлено 10 важных признаков, определенных классификаторами, позволяющих выделить их в ведущие (первые пять) и второстепенные факторы риска развития рецидива у излеченных больных МЛУ-ТБ.

Лидирующим признаком, как уже было отмечено выше, который 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 лидирующую пятерку 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, включено наличие вирусных гепатитов В, С, В + С. Данный фактор определен только алгоритмом XGBoost.

К менее важным из 10 признаков, но введенных в модель прогноза большинством методов, относятся признаки, влияющие на эффективность лечения – количество эффективно работающих ПТП в схеме химиотерапии, сроки абациллирования, амплификация МБТ к левофлоксацину на фоне лечения. Однократно в модели «случайный лес» определен вклад «наличие в легочной ткани остаточных изменений после ранее перенесенного туберкулеза» и в «дерево решений» – индекс массы тела. В улучшенной модели логистической регрессии «высвечиваются» и социальные факторы риска в виде наличия инвалидности по ТБ и пенсионный возраст.

Самым наглядным методом, позволяющим оценить качество построенных моделей прогноза, является построение ROC-кривой на основании рассчитанных показателей специфичности и чувствительности. На рис. 3 приведены ROC-кривые для каждой из разработанных моделей.

Числовые трактовки показателя AUC, представленные в табл. 2, 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.

Заключение

Инструменты МО могут применяться к получению набора значимых признаков и прогностических моделей, характеризующих эффективность химиотерапии у больных МЛУ-ТБ.

Настоящее исследование выделило предикторы развития рецидива у успешно пролеченных больных МЛУ-ТБ, которые включают 5 основных признаков:

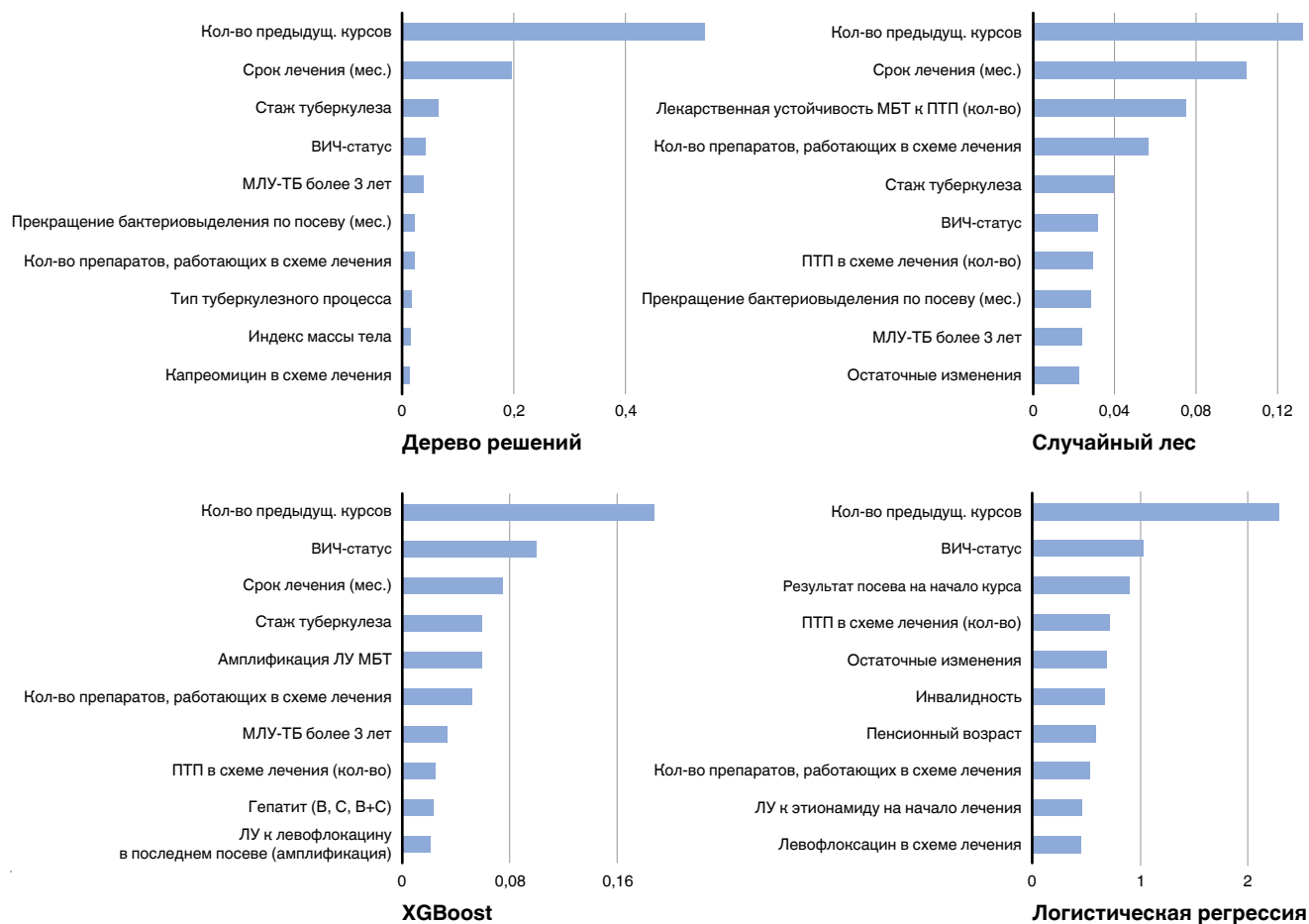


Рис. 2. Значимость признаков в моделях классификаторов, определяющих факторы риска развития рецидива МЛУ-ТБ

Fig. 2. Significance of features in classifier models determining risk factors for MDR-TB relapse

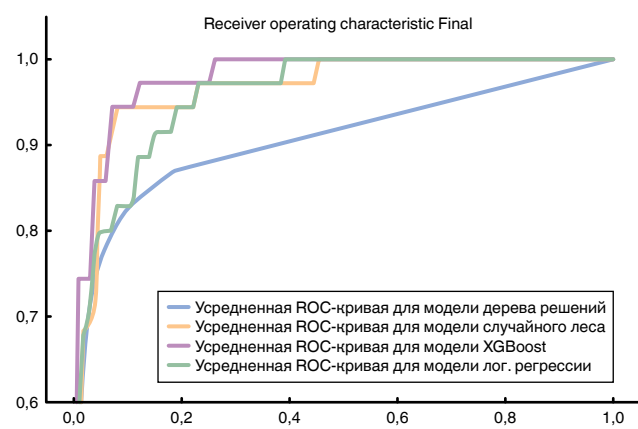


Рис. 3. ROC-кривые для моделей классификаторов, определяющих признаки, сопряженные с рецидивом МЛУ-ТБ

Fig. 3. ROC curves for classifier models defining attributes associated with MDR TB relapse

наличие у пациентов повторных курсов химиотерапии; сопряжение со стажем заболевания; присутствие в легочной ткани полости распада; химиотерапия менее чем 5 ПТП; срок лечения менее 22 мес.

Таблица 2. Метрики качества работы классификаторов, построенных с помощью К-блочной стратифицированной проверки в прогнозировании рецидива МЛУ-ТБ

Table 2. Quality metrics for classifiers constructed using K-block stratified screening in predicting MDR-TB recurrence

Алгоритм	Чувствительность, отклонение	Специфичность, отклонение	AUC, отклонение
Дерево решений	0,740 ± 0,167	0,97 ± 0,03	0,9 ± 0,1
Случайный лес	0,91 ± 0,17	0,98 ± 0,02	0,960 ± 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

Алгоритмы МО в виде «дерево решений», «случайный лес», XGBoost (градиентный бустинг) и логистическая регрессия позволили с высокой точностью и чувствительностью определить риск развития рецидива МЛУ-ТБ и дополнительно продемонстрировали сложную взаимосвязь развития рецидива МЛУ-ТБ между особенностями возбудителя ТБ, способного изменяться (мутировать), клинической характеристикой пациента (возраст, наличие сопутствующей патологии) и возможностями применяемого лечения.

Конфликт интересов. Авторы заявляют об отсутствии у них конфликта интересов.

Conflict of Interests. The authors state that they have no conflict of interests.

ЛИТЕРАТУРА

1. Ghiasi M. M., Zendehboudi, S. Decision tree-based methodology to select a proper approach for wart treatment // *Comp. Biol. Med.* – 2019. – Vol. 4. – P. 400-409. DOI: 10.1016/j.combiomed.2019.04.001. PMID: 31077954.
2. Hou N., Li M., He L., Xie B., Wang L., Zhang R., Yu Y., Sun X., Pan Z., Wang K. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost // *J. Transl. Med.* – 2020. – Vol. 18, № 1. – P. 462. DOI: 10.1186/s12967-020-02620-5. PMID: 33287854.
3. Inoue T., Ichikawa D., Ueno T., Cheong M., Inoue T., Whetstone W.D., Endo T., Nizuma K., Tominaga T. XGBoost, a machine learning method, predicts neurological recovery in patients with cervical spinal cord injury // *Neurotrauma Reports*. – 2020. – Vol. 1, № 1. – P. 8-16. DOI:10.1089/neur.2020.0009.
4. Ji C., Zou X., Hu Y., Liu S., Lyu L., Zheng X. XG-SF: An XGBoost classifier based on shapelet features for time series classification // *Procedia Comp. Sci.* – 2019. – Vol. 147. – P. 24-28. DOI: 10.1016/j.procs.2019.01.179.
5. Kouchaki S., Yang Y., Walker T. M., Walker S. A., Wilson D. J., Peto T., Crook D. W., CRyPTIC Consortium; Clifton D. A. Application of machine learning techniques to tuberculosis drug resistance analysis // *Bioinformatics*. – 2019. – Vol. 35, № 13. – P. 2276-2282. DOI:10.1093/bioinformatics/bty949. PMID: 30462147.
6. Lakhani P., Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks // *Radiology*. – 2017. – Vol. 284, № 2. – P. 574-582. DOI: 10.1148/radiol.2017162326. PMID: 28436741.
7. Lange C. et al. Management of patients with multidrug-resistant tuberculosis // *Intern. J. Tuberc. Lung Dis.* – 2019. – Vol. 23, № 6. – P. 645-662. DOI: 10.5588/ijtld.18.0622. PMID: 31315696.
8. Pal K., Patel B. V. Data classification with k-fold cross validation and holdout accuracy estimation methods with 5 different machine learning techniques // *Proceedings of the 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)* – 2020. pp. 83-87. DOI: 10.1109/ICCMC48092.2020.ICCMC-00016.
9. Rajkomar A., Dean J., Kohane I. Machine learning in medicine // *N. Engl. J. Med.* – 2019. – Vol. 380, № 14. – P. 1347-1358.
10. Shi Yu, Wong W.-K., Goldin J. G., Brown M. S., Kim G. H. J. Prediction of progression in idiopathic pulmonary fibrosis using CT scans at baseline: a quantum particle swarm optimization - random forest approach // *Artif. Intellig. Med.* – 2019. – Vol. 100, № 7. – P. 101719. DOI: 10.1016/j.artmed.2019.101709. PMID: 31607341.
11. World Health Organization. Tuberculosis surveillance and monitoring in Europe 2019: 2017 data. – 2019.
12. Yim J. J., Koh W. J. MDR-TB recurrence after successful treatment: additional studies using molecular genotyping are needed // *Intern. J. Tuberc. Lung Dis.* – 2015. – Vol. 19, № 4. – P. 371. DOI: 10.5588/ijtld.15.0117. PMID: 25859987.
13. Yoshiyama T., Morimoto K., Okumura M., Sasaki Y., Ogata H., Shiraishi Y., Kudou S. Long term outcome of multidrug-resistant TB patients in Fukujiji Hospital in Japan // *Trans. Roy. Soc. Trop. Med. Hyg.* – 2014. – Vol. 108, № 9. – P. 589-590. DOI: 10.1093/trstmh/tru080. PMID: 24902580.
14. Zhao X., Liu Y., Zhang A., Gao B., Feng Q., Huang H., Zhu X., Sun X., Xu D. Logistic regression analyses of factors affecting fertility of intrauterine adhesions patients // *Ann. Transl. Med.* – 2020. – Vol. 8, № 4. – P. 49. DOI: 10.21037/atm.2019.11.115. PMID: 32175343.
15. Zumla A. et al. World Tuberculosis Day 2021 Theme – ‘The Clock is Ticking’ – and the world is running out of time to deliver the United Nations General Assembly commitments to End TB due to the COVID-19 pandemic // *Intern. J. Infect. Dis.* – 2021. DOI: 10.1016/j.ijid.2021.03.046. PMID: 33746094.

REFERENCES

1. Ghiasi M.M., Zendehboudi, S. Decision tree-based methodology to select a proper approach for wart treatment. *Comp. Biol. Med.*, 2019, vol. 4, pp. 400-409. doi: 10.1016/j.combiomed.2019.04.001. PMID: 31077954.
2. Hou N., Li M., He L., Xie B., Wang L., Zhang R., Yu Y., Sun X., Pan Z., Wang K. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. *J. Transl. Med.*, 2020, vol. 18, no. 1, pp. 462. doi: 10.1186/s12967-020-02620-5. PMID: 33287854.
3. Inoue T., Ichikawa D., Ueno T., Cheong M., Inoue T., Whetstone W.D., Endo T., Nizuma K., Tominaga T. XGBoost, a machine learning method, predicts neurological recovery in patients with cervical spinal cord injury. *Neurotrauma Reports*, 2020, vol. 1, no. 1, pp. 8-16. doi:10.1089/neur.2020.0009.
4. Ji C., Zou X., Hu Y., Liu S., Lyu L., Zheng X. XG-SF: An XGBoost classifier based on shapelet features for time series classification. *Procedia Comp. Sci.*, 2019, vol. 147, pp. 24-28. doi: 10.1016/j.procs.2019.01.179.
5. Kouchaki S., Yang Y., Walker T.M., Walker S.A., Wilson D.J., Peto T., Crook D.W., CRyPTIC Consortium; Clifton D. A. Application of machine learning techniques to tuberculosis drug resistance analysis. *Bioinformatics*, 2019, vol. 35, no. 13, pp. 2276-2282. doi:10.1093/bioinformatics/bty949. PMID: 30462147.
6. Lakhani P., Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, 2017, vol. 284, no. 2, pp. 574-582. doi: 10.1148/radiol.2017162326. PMID: 28436741.
7. Lange C. et al. Management of patients with multidrug-resistant tuberculosis. *Intern. J. Tuberc. Lung Dis.*, 2019, vol. 23, no. 6, pp. 645-662. doi: 10.5588/ijtld.18.0622. PMID: 31315696.
8. Pal K., Patel B.V. Data classification with k-fold cross validation and holdout accuracy estimation methods with 5 different machine learning techniques. *Proceedings of the 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*. 2020, pp. 83-87. doi: 10.1109/ICCMC48092.2020.ICCMC-00016.
9. Rajkomar A., Dean J., Kohane I. Machine learning in medicine. *N. Engl. J. Med.*, 2019, vol. 380, no. 14, pp. 1347-1358.
10. Shi Yu, Wong W.K., Goldin J.G., Brown M.S., Kim G.H.J. Prediction of progression in idiopathic pulmonary fibrosis using CT scans at baseline: a quantum particle swarm optimization - random forest approach. *Artif. Intellig. Med.*, 2019, vol. 100, no. 7, pp. 101719. doi: 10.1016/j.artmed.2019.101709. PMID: 31607341.
11. World Health Organization. Tuberculosis surveillance and monitoring in Europe 2019: 2017 data. 2019.
12. Yim J.J., Koh W.J. MDR-TB recurrence after successful treatment: additional studies using molecular genotyping are needed. *Intern. J. Tuberc. Lung Dis.*, 2015, vol. 19, no. 4, pp. 371. doi: 10.5588/ijtld.15.0117. PMID: 25859987.
13. Yoshiyama T., Morimoto K., Okumura M., Sasaki Y., Ogata H., Shiraishi Y., Kudou S. Long term outcome of multidrug-resistant TB patients in Fukujiji Hospital in Japan. *Trans. Roy. Soc. Trop. Med. Hyg.*, 2014, vol. 108, no. 9, pp. 589-590. doi: 10.1093/trstmh/tru080. PMID: 24902580.
14. Zhao X., Liu Y., Zhang A., Gao B., Feng Q., Huang H., Zhu X., Sun X., Xu D. Logistic regression analyses of factors affecting fertility of intrauterine adhesions patients. *Ann. Transl. Med.*, 2020, vol. 8, no. 4, pp. 49. doi: 10.21037/atm.2019.11.115. PMID: 32175343.
15. Zumla A. et al. World Tuberculosis Day 2021 Theme – ‘The Clock is Ticking’ – and the world is running out of time to deliver the United Nations General Assembly commitments to End TB due to the COVID-19 pandemic. *Intern. J. Infect. Dis.*, 2021. doi: 10.1016/j.ijid.2021.03.046. PMID: 33746094.

ИНФОРМАЦИЯ ОБ АВТОРАХ:

ОГАОУЗ «Томский фтизиопульмонологический медицинский центр»,
634009, г. Томск, ул. Розы Люксембург, д. 17.

Алليلев Александр Сергеевич
заместитель главного врача
по организационно-методической работе.
Тел.: 8 (3822) 51-74-15.
E-mail: alliluev233@gmail.com

Шнайдер Екатерина Евгеньевна
врач-статистик организационно-методического отдела.
Тел.: 8 (3822) 60-94-95, доб. 621.
E-mail: nikolaevskaya.e.e@gmail.com

Филинюк Ольга Владимировна
ФГБОУ ВО «Сибирский государственный медицинский университет» МЗ РФ,
доктор медицинских наук, профессор,
заведующая кафедрой фтизиатрии и пульмонологии.
634050, г. Томск, Московский тракт, д. 2.
Тел.: 8 (3822) 90-11-01, доб. 1838.
E-mail: filinyuk.olga@yandex.ru

Аксенов Сергей Владимирович
ФГАОУ ВО «Национальный исследовательский Томский политехнический университет», Инженерная школа информационных технологий и робототехники,
кандидат технических наук,
доцент отделения информационных технологий.
634034, г. Томск, ул. Советская, д. 84, корп. 3.
Тел.: 8 (3822) 60-61-30.
E-mail: axyonov@tpu.ru

INFORMATION ABOUT AUTHORS:

*Tomsk Phthisiopulmonology
Medical Center,
17, R. Luxemburg St., Tomsk, 634009.*

Aleksandr S. Alliluev
*Deputy Head Physician
on Reporting and Statistics.
Phone: +7 (3822) 51-74-15.
Email: alliluev233@gmail.com*

Ekaterina E. Shnayder
*Specialist of Statistics and Reporting Department.
Phone: +7 (3822) 60-94-95, ext. 621.
Email: nikolaevskaya.e.e@gmail.com*

Olga V. Filinyuk
*Siberian State Medical University,
Doctor of Medical Sciences, Professor,
Head of Phthisiology
and Pulmonology Department.
2, Moskovsky Tr., Tomsk. 634050.
Phone: +7 (3822) 90-11-01, ext. 1838.
Email: filinyuk.olga@yandex.ru*

Sergey V. Aksenov
*Tomsk Polytechnic University,
School of Computer Science & Robotics,
Candidate of Technical Sciences,
Associate Professor of Computer Sciences Department.
Build. 3, 84, Sovetskaya St.,
Tomsk, 634034.
Phone: +7 (3822) 60-61-30.
Email: axyonov@tpu.ru*

Поступила 16.08.2021

Submitted as of 16.08.2021