

Разработка калькулятора на основе модели машинного обучения Random Forest для оценки состава мочевых камней на основании клинических и лабораторно-инструментальных данных.

Научный руководитель – Нестерова Ольга Юрьевна

Терёшина Анастасия Дмитриевна

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова, Москва, Россия

E-mail: tereshinaanastasia51@gmail.com

Введение. Состав мочевых камней является одним из ключевых факторов, определяющих тактику лечения, выбор метода хирургического вмешательства и последующую метафилактику. Несмотря на широкое применение компьютерной томографии, предоперационная неинвазивная дифференциация типов камней ограничена. Методы машинного обучения позволяют интегрировать клинико-лабораторные и визуализационные данные для повышения точности прогнозирования состава конкрементов [1].

Цель: разработка модели машинного обучения для оценки состава мочевых камней и анализ её клинического применения.

Материалы и методы. В исследование включено 158 пациентов, проходивших лечение мочекаменной болезни с удалением конкремента. Анализ химического состава камней выполнялся методом инфракрасной спектроскопии. Модель построена на алгоритме Random Forest с 5-кратной перекрёстной валидацией и рекурсивным отбором признаков (RFE). Качество оценивалось по Accuracy, Balanced Accuracy, Precision, Recall, F1, ROC-AUC, PR-AUC и DCA. Все показатели качества рассчитывались в макро-усреднённом варианте. Интерпретация проводилась с помощью SHAP и Partial Dependence Plots.

Результаты. В финальную модель Random Forest включены 12 клинико-лабораторных предикторов, отобранных методом RFE. Исходная модель показала Accuracy 68,9%, коэффициент Каппа Коэна 0,333, Balanced Accuracy 64,0%, Recall 52,2%, F1 54,9% и ROC-AUC 0,618. Взвешивание классов снизило Accuracy до 65,2%, но увеличило Recall 60,4%, Balanced Accuracy 69,8%, F1 59,8% и ROC-AUC 0,720. При этом оценка Брайера увеличилась с 0,464 до 0,515, указывая на снижение калибровки вероятностных предсказаний. Поклассовый анализ показал рост Recall для фосфатных (с 20,0% до 50,0%) и уратных камней (с 43,8% до 59,4%) при снижении Precision и специфичности. Комплексная оценка ROC-кривых, PR-AUC и DCA продемонстрировала стабильную дискриминационную способность и клиническую полезность исходной модели, что подтвердило её выбор в качестве финальной. На основе модели разработан интерактивный калькулятор для расчёта вероятностей уратного, оксалатного и фосфатного состава камня на основе 12 рутинных показателей.

Заключение. Исходная модель Random Forest показала оптимальное сочетание точности, устойчивости и клинической интерпретируемости и выбрана для прогнозирования состава мочевых камней. Разработанный калькулятор может служить инструментом поддержки принятия клинических решений, дополняя данные компьютерной томографии и лабораторных исследований.

Ключевые слова: мочекаменная болезнь, машинное обучение, Random Forest, состав мочевых камней, прогнозирование.

Источники и литература

- 1) Chmiel J.A., Stuivenberg G.A., Wong J.F.W., Nott L., Burton J.P., Razvi H., Bjazevic J. Predictive Modeling of Urinary Stone Composition Using Machine Learning and Clinical Data: Implications for Treatment Strategies and Pathophysiological Insights // J Endourol. 2024. Vol. 38, No. 8. P. 778-787.

Иллюстрации

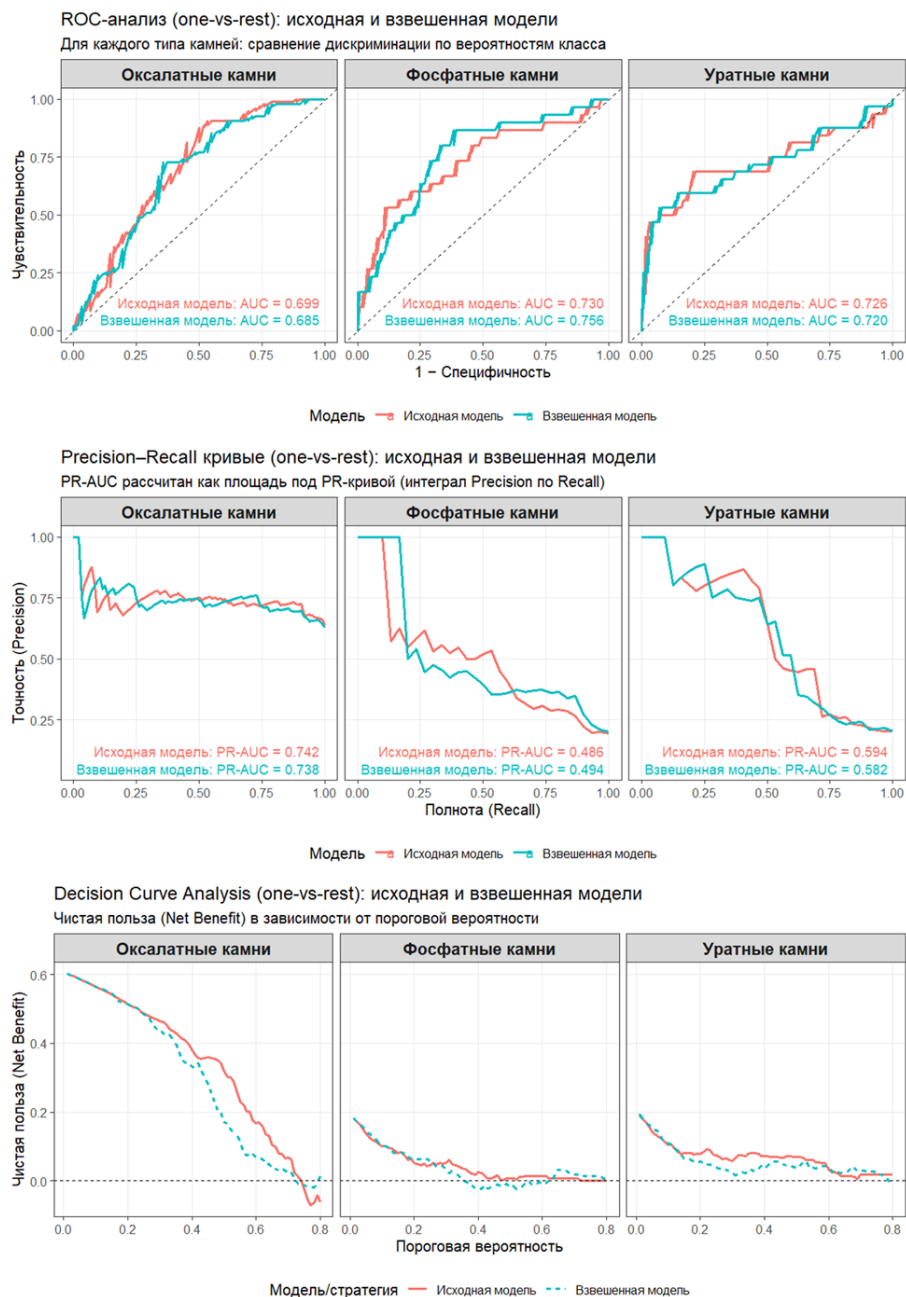


Рис. : Комплексная оценка качества классификации состава мочевых камней для исходной и взвешенной моделей Random Forest: в верхнем блоке представлены ROC-кривые; средний блок отображает Precision–Recall кривые; в нижнем блоке показаны кривые принятия решений (Decision Curve Analysis).

Калькулятор: вероятность состава мочевого камня (Random Forest)

Введите параметры пациента

Плотность камня (НУ)	Креатинин
<input type="text" value="800"/>	<input type="text" value="96,4"/>
Мочевина	Возраст
<input type="text" value="5,5"/>	<input type="text" value="45"/>
ИМТ	Ожирение (0/1)
<input type="text" value="28,7"/>	<input type="text" value="0"/>
Объём камня (мм³)	Натрий
<input type="text" value="210"/>	<input type="text" value="142,5"/>
Гемоглобин	рН мочи
<input type="text" value="139"/>	<input type="text" value="6,0"/>
Протромбин по Квику	Билирубин прямой
<input type="text" value="92"/>	<input type="text" value="1,6"/>

Результат — вероятности (0–1) для трёх типов камней.

Результат

Наиболее вероятно: Оксалаты
Вероятность: 0,771

Тип.камня	Вероятность
Оксалаты	0.77
Ураты	0.15
Фосфаты	0.08

Пояснение

- Вероятности отражают уверенность модели для каждого типа камней.
- Рекомендуется интерпретировать результат совместно с клиникой и данными КТ.

Рис. : Калькулятор для оценки вероятности состава мочевых камней на основании стандартных клинико-лабораторных и инструментальных данных.