

ПАРШАКОВА М.С.,

ФГБОУ «СПИИРАН ФИЦ РАН (СПИИРАН)», г. Санкт-Петербург, Россия; e-mail: mary.parshakova@mail.ru

СЕМЕНОВ А.А.,

к.м.н., ФГБОУ ВО «СПбГУ», г. Санкт-Петербург, Россия; e-mail: arseny@thyro.ru

ЛАСКИН М.Б.,к.ф.-м.н., д.э.н., ФГБОУ «СПИИРАН ФИЦ РАН (СПИИРАН)», г. Санкт-Петербург, Россия;
e-mail: laskin.m@iias.spb.su

ПРЕДВАРИТЕЛЬНАЯ ОЦЕНКА КРОВОПОТЕРИ ПРИ ХИРУРГИЧЕСКОМ ЛЕЧЕНИИ ЭНДОКРИННЫХ ЗАБОЛЕВАНИЙ АНСАМБЛЕВЫМИ МЕТОДАМИ РЕШАЮЩИХ ДЕРЕВЬЕВ

DOI: 10.25881/18110193_2026_3_118

Аннотация. В медицинских учреждениях и клиниках накапливается значительный объем информации, который может и должен быть использован при планировании лечебной и организационной деятельности. При этом между влияющими факторами и целевыми KPI имеются интуитивно понятные связи, которые не могут быть описаны функциональными зависимостями. В таких условиях возникает богатый спектр различных задач, для решения которых могут привлекаться методы машинного обучения и искусственного интеллекта.

В настоящей статье рассматривается задача прогнозирования кровопотери при хирургическом лечении заболеваний органов эндокринной системы. Для оценки кровопотерь применяются методы машинного обучения, основанные на решающих деревьях. Данные для исследования предоставлены одной из ведущих мировых клиник.

Ключевые слова: решающие деревья, ансамблевые методы, машинное обучение в медицине, прогнозирование кровопотери, длительность операции.

Для цитирования: Паршакова М.С., Семенов А.А., Ласкин М.Б. Предварительная оценка кровопотери при хирургическом лечении эндокринных заболеваний ансамблевыми методами решающих деревьев. Врач и информационные технологии. 2026; 3: 118-131. DOI: 10.25881/18110193_2026_3_118.

PARSHAKOVA M.S.,

FSBIS "SPC RAS (SPIIRAS) ", St. Petersburg, Russia; e-mail: mary.parshakova@mail.ru

SEMENOV A.A.,

Ph.D., FSBEI HE "SPbSU", St. Petersburg, Russia; e-mail: arseny@thyro.ru

LASKIN M.B.,

DSc., FSBIS "SPC RAS (SPIIRAS) ", St. Petersburg, Russia;

e-mail: laskin.m@ias.spb.su

PRELIMINARY ASSESSMENT OF BLOOD LOSS DURING SURGICAL TREATMENT OF ENDOCRINE DISEASES USING DECISION TREES

DOI: 10.25881/18110193_2026_3_118

Abstract. Medical institutions and clinics accumulate a significant amounts of data that can and should be used in planning clinical and operational activities. There are intuitive connections between influencing factors and target KPIs that cannot be described by functional dependencies. Such conditions give rise to a wide range of tasks that can be addressed using machine learning and artificial intelligence methods. This article examines the task of predicting blood loss during the surgical treatment of endocrine system disorders. Machine learning methods based on decision trees are used to estimate blood loss. The data for the study were provided by one of the world's leading clinics of endocrine surgery from St. Petersburg.

Keywords: decision trees, ensemble methods, machine learning in medicine, surgery duration, blood loss estimation, duration of surgery.

For citation: Parshakova M.S., Semenov A.A., Laskin M.B. Preliminary assessment of blood loss during surgical treatment of endocrine diseases using decision trees. Medical doctor and information technology. 2026; 3: 118-131. DOI: 10.25881/18110193_2026_3_118.

ВВЕДЕНИЕ

В медицинских организациях накапливается значительный объем данных, потенциально применимый для решения широкого спектра клинических и управленческих задач. Однако даже при наличии очевидных причинно-следственных связей между переменными часто оказывается невозможно описать эти зависимости в виде явных функциональных моделей. В подобных ситуациях многие исследователи рассматривают методы машинного обучения как перспективный инструмент, способный обеспечить удовлетворительные результаты моделирования и прогнозирования, см., например, [1]-[12]. В настоящей статье методы машинного обучения рассматриваются как инструмент в системах поддержки принятия решений. В частности, мы рассматриваем задачу оценки кровопотери при хирургическом вмешательстве по накопленным данным профильной клиники, специализирующейся на эндокринных заболеваниях. Оценка возможной кровопотери необходима для предсказания потребности и доступности донорской крови и/или помощи специалистов с большей компетентностью или смежных специальностей. Подобные задачи рассматривались в ряде работ. Предоперационной оценке рисков кровопотерь при кардиохирургических вмешательствах посвящена работа [13], рискам кровопотерь в детской онкологии – работа [14]. В [15] рассматриваются возможности машинного обучения при прогнозировании акушерских кровотечений. Для решения ряда задач медицинской практики, нуждающихся в предварительной (предоперационной) оценке, в том числе прогнозировании кровопотерь, часто рассматриваются методы машинного обучения, см. [15] -[24]. В частности, в работах [25],[26] рассматриваются алгоритмы, основанные на решающих деревьях, в [4],[5] – статистический анализ, логистическая регрессия, случайный лес, нейронные сети, в [7],[16],[18] — нейронные сети. Хороший обзор методов машинного обучения для повышения качества пренатальной диагностики врожденных пороков развития дан в [27]. Исходным материалом для построения таких моделей являются накопленные данные по уже проведенным операциям. Проблема заключается в том, что в алгоритмах машинного обучения модели

обучаются на части данных (обучающих множествах) и проверяются на тестовых множествах, но применяться модель будет на данных, для которых значение целевой переменной заранее неизвестно. В нашей задаче целевых переменных две – длительность операции и объем кровопотери. При этом длительность операции сама является фактором, влияющим на кровопотери, а прогноз надо дать до начала операции. Планирование длительности операции и кровопотери проводится для решения задачи оптимальной загрузки операционного блока клиники, что актуально, так как поток пациентов превышает существующие возможности операционного блока, а работа с большей продолжительностью операционного дня должна учитывать ограничения со стороны обслуживающих подразделений и равномерной загрузки на персонал.

МАТЕРИАЛЫ И МЕТОДЫ

При планировании загрузки операционного блока необходимо оценить возможную длительность операции и объем возможной кровопотери на основании имеющихся ретроспективных данных, в которых отражены значения влияющих факторов и целевых переменных. Целевые переменные связаны, зависимы: нет операции – нет кровопотери. При переносе алгоритма машинного обучения в систему поддержки принятия решений операции еще нет. Таким образом, чтобы оценить кровопотерю, надо сначала оценить длительность операции. Предметом настоящей статьи является алгоритм предсказания возможной кровопотери по заданному набору влияющих факторов. Все расчеты к настоящей статье выполнены в статистическом пакете R. Исходный набор данных содержит сплошную выборку из 3059 записей об операциях, выполненных в отделении эндокринной хирургии Клиники высоких медицинских технологий им. Н.И. Пирогова, СПбГУ в течение 2 кварталов 2024 года. Данные представляют собой Excel-файл со значениями влияющих факторов и целевых переменных. Факторы, предположительно влияющие на продолжительность операции и кровопотерю: возраст пациента, пол пациента, диагноз, виды медицинской услуги (объем резекции щитовидной железы, объем лимфодиссекции,

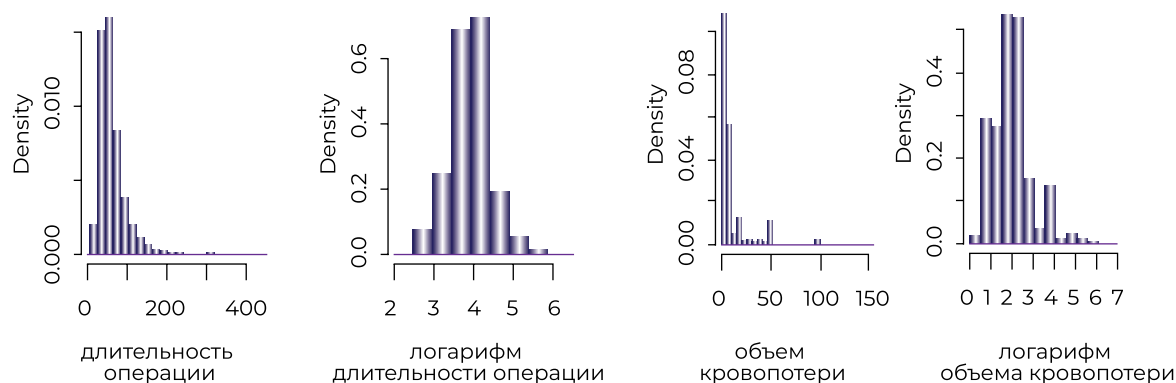


Рисунок 1 – Распределение целевых переменных и их логарифмов.
Слева – длительность операции, справа – объем кровопотери.

удаление околощитовидных желез, операция на надпочечниках), час начала операции, день недели операции, месяц операции, хирург (персонально), объем кровопотери, особенности доступа, прочее. Часть переменных (возраст пациента, час начала операции, длительность операции, объем кровопотери) являются числовыми, часть (оказанная медицинских услуга, диагноз, день недели операции, месяц операции, хирург) – ранговые; факторы пол пациента, особенности доступа, прочее – ранговые, бинарные. Имеющийся набор данных предполагает интуитивно понятную зависимость целевых переменных от влияющих факторов, однако говорить о каких-либо функциональных зависимостях не представляется возможным. Кроме того, очевидно, что целевые переменные «длительность операции» и «объем кровопотери» имеют причинно-следственную связь. Конечная задача исследования – дать прогноз длительности операции и, затем, кровопотери по имеющимся данным. В настоящей статье мы рассматриваем методы машинного обучения, основанные на решающих деревьях (RandomForest, Bagging, Busting). Этот класс алгоритмов уже рассматривался в медицинских приложениях [25],[26]. Прежде всего, отметим особенности целевых переменных «длительность операции» и «объем кровопотери». На рисунке 1 показаны их распределения и распределения логарифмов целевых переменных.

Алгоритмы, в основе которых лежит решающее дерево (см. [28],[29]), присваивают среднее арифметическое значение всем элементам,

оказавшимся на одном листе. Обе целевые переменные имеют правостороннюю асимметрию. Логарифмирование приводит к более симметричным распределениям (логарифмы целевых переменных). Поэтому все построения моделей выполнялись для логарифмов целевых переменных. Традиционно, для методов машинного обучения исходный набор данных разбивался на два множества: обучающее (train, случайная выборка объемом 2000 элементов) и тестовое (test, случайная выборка объемом 1059 элементов). Разбиение выполнялось библиотечной функцией `sample.split()` статистического пакета R (см., [30]).

Для оценки длительности операции (в логарифмах) рассмотрены алгоритмы RandomForest, Bagging (функция `randomForest` из библиотеки `randomForest`), Busting (функция `gbm` из библиотеки `GBM`). Наименьшее значение среднеквадратичного отклонения (СКО, MSE) результатов моделирования от наблюдаемых значений на тестовом множестве дает градиентный бустинг (`gbm`) (таблица 1).

Приведенные в таблице 1 результаты оценки точности трех алгоритмов объясняют наш выбор алгоритма `gbm` для дальнейшего моделирования.

При построении `gbm` модели для оценки будущей кровопотери в состав влияющих факторов входит переменная «длительность операции».

Библиотечная функция `summary()` позволяет оценить важность факторов, участвовавших в построении модели. Результаты показаны в таблице 2.

Таблица 1 – Результаты оценки точности алгоритмов

	Количество деревьев	MSE	RMSE
RandomForrest (6 предикторов)	10 000	0,1353979	0,3679646
Bagging	10 000	0,1406267	0,3750023
Busting (gbm)	10 000	0,1313889	0,3624761

Таблица 2 – Важность факторов. Слева для длительности операции, справа для объема кровопотери

var	rel.inf	var	rel.inf
Код диагноза	59,1973	Код диагноза	57,6174
Хирург	19,8498	Длительность операции	14,7039
Резекция щитовидной железы	5,3187	Возраст	6,0096
Лимфодиссекция	3,7865	Хирург	5,0284
Возраст	3,2389	Месяц	4,4343
Месяц	3,1016	Лимфодиссекция	3,7380
День недели	1,9768	Час начала операции	3,3430
Удаление околощитовидных желез	1,6075	День недели	2,8744
Час начала операции	1,3545	Резекция щитовидной железы	1,6572
Пол	0,3471	Пол	0,3117
Операция на надпочечниках	0,2214	Удаление околощитовидных желез	0,2732
Особенности доступа	0,000001	Операция на надпочечниках	0,0075
прочее	0,000001	Особенности доступа	0,000001
		прочее	0,000001

Наибольшее участие в формировании модели gbm для переменной «длительность операции» принимали факторы: «диагноз», «хирург», «резекция щитовидной железы», «лимфодиссекция». Наибольшее участие в формировании модели gbm для переменной «объем кровопотери» принимали факторы: «диагноз», «длительность операции», «возраст» «хирург». Факторы «особенности доступа», «прочее» на продолжительность операции и объем кровопотери значимо не влияют. На рисунке 2 показано соответствие наблюдаемых значений логарифмов длительностей операции и значений, предсказанных алгоритмом gbm на тестовом множестве, а также ядерная оценка плотности распределения наблюдаемых и предсказанных значений логарифма длительности операции.

На рисунке 2 можно заметить характерное свойство моделей, основанных на решающих деревьях. Поскольку при каждом разбиении

множеств в алгоритме построения решающего дерева элементам одного множества присваивается их среднее значение, в области малых значений результат предсказания завышается, а в области верхних значений – занижается. Может быть построен линейный тренд (показан красной линией), указывающий на возможность использования линейной зависимости между наблюдаемыми и предсказанными значениями (в логарифмах).

Пусть линейная модель, связывающая предсказанные моделью gbm значения логарифмов переменной «длительность операции» с наблюдаемыми значениями логарифмов длительностей операций имеет вид

$$\ln(D_{\text{пр}}) = a_1 \cdot \log(D_{\text{н}}) + b_1 + \varepsilon_1 \quad (1)$$

где $D_{\text{пр}}$ – предсказанная длительность, $D_{\text{н}}$ – наблюдаемая длительность, ε_1 – случайная

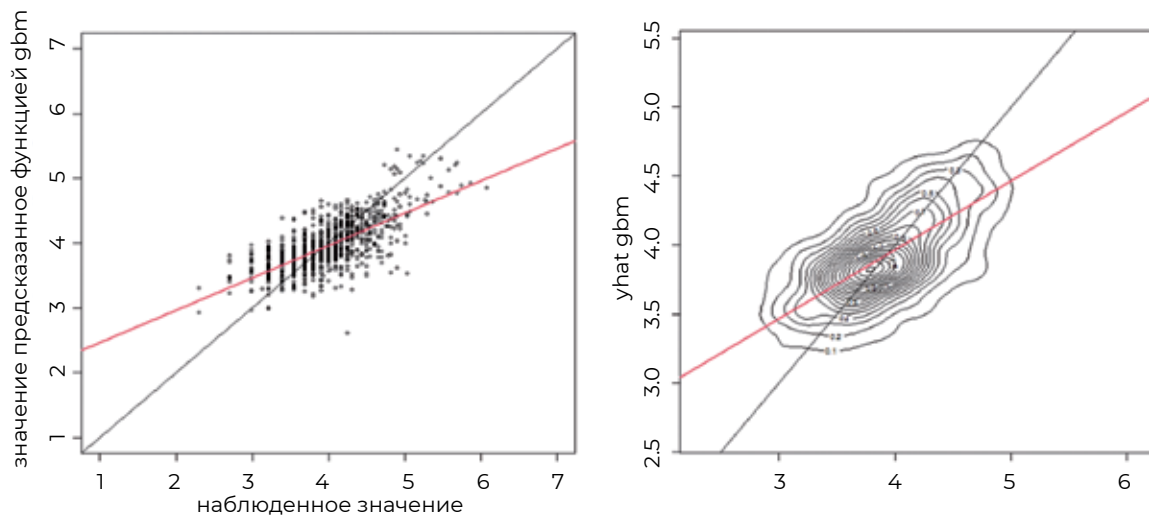


Рисунок 2 – Слева – соответствие наблюдаемых значений логарифмов длительностей операции и значений, предсказанных алгоритмом gbm.

Справа – ядерная оценка плотности и её линии уровня.

По горизонтали – наблюдаемые значения логарифмов длительностей операции, по вертикали – значения, предсказанные алгоритмом gbm.

Черная линия – биссектриса первого координатного угла (линия точных предсказаний), красная линия – линейный тренд.

ошибка. Параметры тренда определялись с применением библиотечной функции $\text{lm}()$ статистического пакета R на тестовом множестве. Коэффициенты равны: $a_1 = 0,524$, $b_1 = 1,86$. Ошибки ε_1 проверены на нормальность параметрическим тестом Колмогорова-Смирнова. Результат теста: $p\text{-value} = 0,1168$ при среднем $0,009$ и стандартном отклонении $0,22$ (значение $p\text{-value} > 0,05$ не дает оснований отклонить нулевую гипотезу о нормальности ошибок). При смене разбиений на обучающее и тестовое множество ε_1 остается устойчиво нормальным со средним около нуля. Теперь обученную модель скорректируем на тестовом множестве. Подробное описание процедуры корректировки нами приведено в статье [31]. При практическом применении обученной и протестированной модели фактор «длительность операции» отсутствует, есть только его прогноз, на основании которого оцениваются возможные кровопотери. Предсказательная модель, построенная алгоритмом gbm и скорректированная на образовавшийся линейный тренд, дает нам модель связи между предсказанными

значениями логарифма длительности операции и наблюдаемыми значениями вида (1), при этом $\varepsilon_1 \in N(0, \sigma_1^2)$. Обозначим неизвестную фактическую длительность операции - D . Оценка будущей длительности операции строится на основании формулы (1): $\ln(D_{\text{пр}}) = a_1 \cdot \ln(D) + b_1 + \varepsilon_1$, откуда следует, что (см. [31])

$$D = D_{\text{пр}}^{\frac{1}{a_1}} \cdot e^{-\frac{b_1}{a_1}} \cdot e^{\varepsilon_1^*} \quad (2)$$

где $\varepsilon_1^* = -\frac{\varepsilon_1}{a_1}$ нормальная ошибка с нулевым средним и стандартным отклонением $\frac{\sigma_1}{a_1}$. Таким образом, формула

$$E(D) = D_{\text{пр}}^{\frac{1}{a_1}} \cdot e^{-\frac{b_1}{a_1}} \quad (3)$$

дает нам корректировку предсказаний обучающих моделей для оценки длительности операции и является медианной оценкой. На рассматриваемых данных были получены значения

$a_1 = 0,524, b_1 = 1,86, \sigma_1 = 0,22$. Тогда формула для оценки длительности операции

$$E(D) = D_{\text{пр}} \cdot e^{\frac{1}{a_1} - \frac{b_1}{a_1}} = D_{\text{пр}} \cdot e^{\frac{1}{0,524} - \frac{1,86}{0,524}} = D_{\text{пр}}^{1,91} \cdot e^{-3,55} \quad (4)$$

(стандартное отклонение $\sigma(D) = \frac{\sigma_1}{a_1} = \frac{0,22}{0,524} = 0,42$).

Теперь рассмотрим результаты построения предсказательной gbm-модели для целевой переменной «объем кровопотери». Сначала строится gbm-модель для целевой переменной «логарифм объема кровопотери». Модель построена с аналогичными настройками библиотечной функции gbm, отличие заключается в том, что теперь к влияющим факторам добавляется фактор «длительность операции» (см. табл. 2).

Введем обозначения: V – объем кровопотери, $V_{\text{пр}}$ – объем кровопотери, предсказанной алгоритмом gbm. Формула линейного тренда и корректировки предсказаний логарифма объема кровопотери аналогичны формулам (2):

$$\ln(V_{\text{пр}}) = a_2 \cdot \ln(V) + b_2 + \varepsilon_2$$

$$\ln(V) = \frac{1}{a_2} \ln(V_{\text{пр}}) - \frac{b_2}{a_2} - \frac{\varepsilon_2}{a_2}$$

При построении линейного тренда были получены значения: $a_2 = 0,3, b_2 = 1,35$.

Как было отмечено выше, особенность построения модели для оценки объема кровопотери, заключается в том, что обученную модель надо будет применять в условиях, когда все влияющие факторы заданы, а длительности операции нет, т.к. нет самой операции. В этом случае мы вынуждены брать оценку длительности операции, построенную по формуле (1) и скорректированную по формулам (2)-(4).

Целевые переменные D и V образуют двумерную случайную величину (D, V) . На рисунке 3 показано рассеяние точек $(\ln(D), \ln(V_{\text{пр}}))$. Здесь D – предсказанная и скорректированная длительность операции.

Эллиптический характер рассеяния точек $(\ln(D), \ln(V_{\text{пр}}))$ является основанием для проверки распределения компонент на совместную нормальность. Для проверки применялись тесты библиотеки MVN статистического пакета R. Тесты не дали положительного результата. На рисунках 3 и 4 (ядерная оценка плотности) видно, что у двумерного случайного вектора $(\ln(D), \ln(V_{\text{пр}}))$ имеются заметные положительные асимметрии и эксцессы.

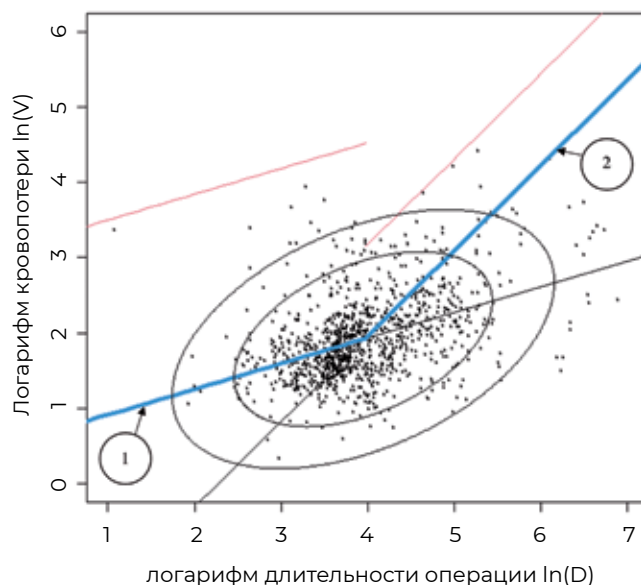


Рисунок 3 – Рассеяние точек $(\ln(D), \ln(V_{\text{пр}}))$, по горизонтали – предсказанный и скорректированный логарифм длительности операции, по вертикали – предсказанный логарифм кровопотери.

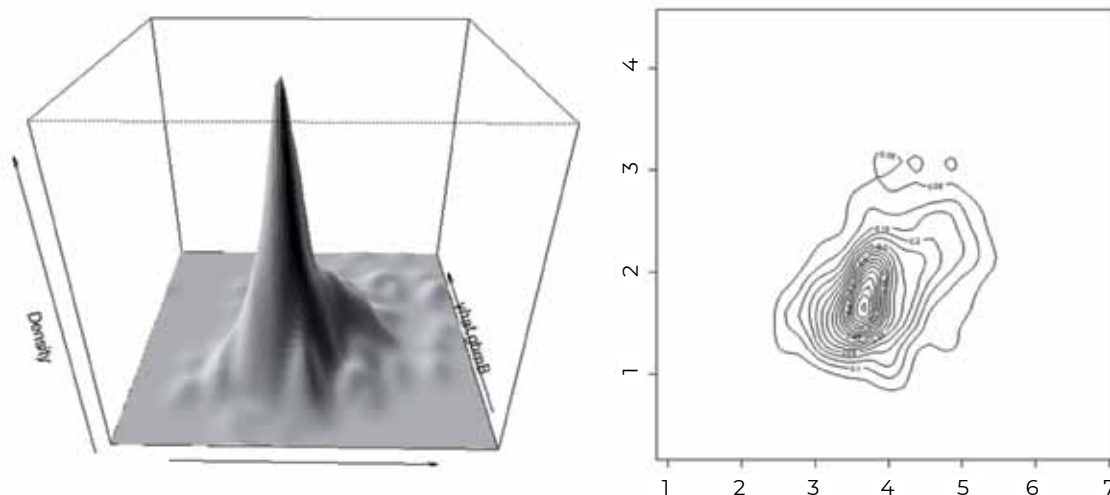


Рисунок 4 – Ядерная оценка плотности двумерной величины $(\ln(D), \ln(V_{пр}))$. Слева в 3d, справа – линии уровня.

На рисунке 3 для сравнения показаны линии уровня нормального распределения (эллиптические области с вероятностью попадания 0,65 и 0,95). Нормальное распределение вектора $(\ln(D), \ln(V_{пр}))$ указывало бы на главную ось эллипса как на линию линейной регрессии. Для наших данных получены:

- уравнение линейной регрессии для предсказанных, но не скорректированных значений $\ln(V)$ (рисунок 3, линия 1):

$$\ln(V) = 0,34058 \cdot \ln(D) + 0,57347 \quad (5)$$

- уравнение линейной регрессии для предсказанных и скорректированных значений $\ln(V)$ (рисунок 3, линия 2):

$$\ln(V) = 1,1353 \cdot \ln(D) - 2,5884 \quad (6)$$

При этом линия 1 имеет угол наклона меньше, а линия 2 больше, чем главная ось эллипсов рассеяния для идеального (нормального) распределения. Точка пересечения указанных прямых имеет координаты $\ln(D) = 3,98$, $\ln(V) = 1,93$, что соответствует в исходных размерностях значениям $D = 53,5$ минуты, $V = 6,9$ мл. При планировании операции важно избежать недооценки возможной выраженной потери крови пациентом для адекватной профилактики развития геморрагического шока, своевременной

организации сбора и переливания аутокрови, подбора донорской крови с учетом фенотипа. Поэтому при предсказанных и скорректированных длительностях операции ниже 53,5 минут для оценки кровопотери лучше использовать уравнение (5), а выше 53,5 минут — уравнение (6).

Уравнения (5) и (6) являются средними оценками логарифмов объема кровопотери. Такая оценка может недооценить риски во время операции для пациентов с редкими группами крови. Рассеяние точек $(\ln(D), \ln(V_{пр}))$ изучено на ошибки уклонения от линий регрессии (5) и (6). Ошибки нормальны, со средними, близкими к нулю, для линии (5) получено стандартное отклонение 0,451, для линии (6) стандартное отклонение равно 0,61. Чтобы оценка кровопотери покрывала уже наблюдавшиеся риски кровопотери в исходных данных, верхнюю границу для оценки логарифма кровопотери в первом случае следует сдвинуть на шесть стандартных отклонений вверх, во втором случае – на два стандартных отклонения вверх. Эти линии имеют уравнения

$$\ln(V) = 0,34058 \cdot \ln(D) + 0,57347 + 6 \cdot 0,451 = 0,34058 \cdot \ln(D) + 3,28 \quad (7)$$

$$\ln(V) = 1,1353 \cdot \ln(D) - 2,5884 + 2 \cdot 0,61 = 1,1353 \cdot \ln(D) - 1,37 \quad (8)$$

и показаны на рисунке 3 красным цветом.

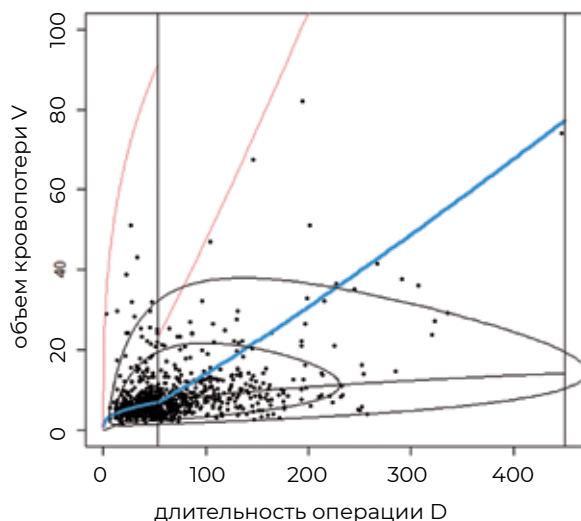


Рисунок 5 – Рассеяние точек ($D, V_{пр}$). По горизонтали предсказанная и скорректированная длительность операции, по вертикали – объем кровопотери.

Результаты и обсуждение. Преобразуем рисунок 3 в исходные размерности задачи, в плоскость (D, V), (см. рисунок 5).

Из формул (5), (6), (7), (8) получаем зависимости оценки объема кровопотери от длительности операции:

$$V = e^{0,57347} \cdot D^{0,34058} = 1,77 \cdot D^{0,34058} \quad (9)$$

$$V = e^{-2,5884} \cdot D^{1,1353} = 0,075 \cdot D^{1,1353} \quad (10)$$

$$V = e^{3,28} \cdot D^{0,34058} = 26,58 \cdot D^{0,34058} \quad (11)$$

$$V = e^{-1,37} \cdot D^{1,1353} = 0,254 \cdot D^{1,1353} \quad (12)$$

Зависимости (9) и (10) показаны на рисунке 5 синим цветом, эти экспоненты имеют общую точку при $D = 53,5$ минуты (показано черной вертикальной линией слева). Зависимости (11) и (12) показаны на рисунке 5 красным цветом. Как видно на рисунке 5 большинство операций имеет продолжительность до 100 минут, но есть и аномально длинные случаи. Максимальное значение длительности операции в исходных данных 450 минут. Модель не может использоваться за пределами данных, на которых она обучалась. При появлении новой статистики, отличающейся от исходной, модель следует переобучить.

Операции аномальной длительности не являются типичными и требуют индивидуального изучения.

Методика оценки объема кровопотери:

1. Построить оценку логарифма длительности операции по модели gbm.
2. Скорректировать оценку логарифма длительности операции по формуле (4).
3. Предсказанные и скорректированные значения логарифмов длительностей операции включить в состав влияющих факторов для оценки объема кровопотери.
4. Построить оценку логарифма кровопотери по модели gbm.
5. Построить линейные тренды зависимостей предсказанного логарифма объема кровопотери и предсказанного и скорректированного логарифма объема кровопотери от длительности операции.
6. Вернуться к исходным размерностям в минутах и мл и построить зависимости вида (9) - (12).

Для оценки кровопотери в зависимости от длительности операции можно использовать формулы вида (9) и (10), так как они являются медианными оценками (ожидается, что 50% кровопотерь будет ниже, а 50% – выше медианного значения). Весьма вероятно, что такая оценка может оказаться недостаточной в некоторых случаях (например, при редкой группе крови). В этом случае лучше применить формулы (11) и (12), т.к. они дают оценку кровопотери, превышающую соответствующие случаи в обучающем множестве (в собранном статистическом материале клиники).

Пример 1. Длительность операции 30 минут (меньше 53,5) – выбираем формулу (9), оценка объема кровопотери $V = 1,77 \cdot 30^{0,34058} = 5,64 \sim 6$ мл. При необходимости используем формулу (11), оценка объема кровопотери $V = 26,58 \cdot 30^{0,34058} = 84,6 \sim 85$ мл.

Пример 2. Длительность операции 120 минут (больше 53,5) – выбираем формулу (10), оценка объема кровопотери $V = 0,075 \cdot 120^{1,1353} = 17,2 \sim 18$ мл. При необходимости используем формулу (12), оценка объема кровопотери $V = 0,254 \cdot 120^{1,1353} = 58,2 \sim 59$ мл.

Кросс-валидация. Основным источником изменчивости результатов моделей машинного обучения является правило разбиения исходных

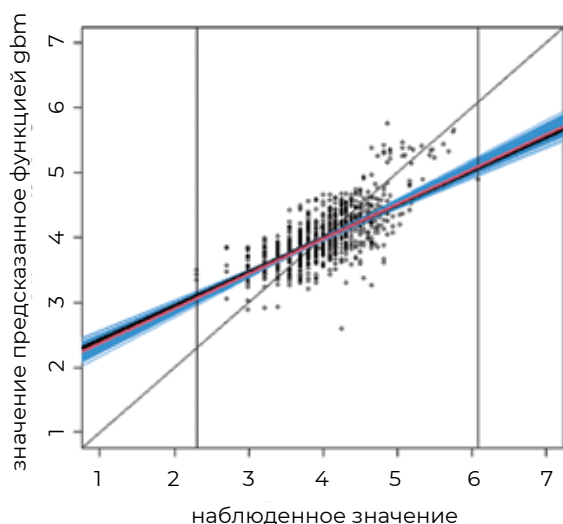


Рисунок 6 – Соответствие наблюдаемых значений логарифмов длительностей операции и значений, предсказанных алгоритмом gbm (для одного разбиения). Голубым цветом показаны линии тренда для 1000 разных разбиений. Красным цветом показана средняя линия, черным – линейный тренд при первом разбиении.

данных на обучающее и тестовое множества. В предложенной схеме изменениям подвержены сами прогнозы, линейные тренды, корректировки (как для длительности операции, так и для объема кровопотери). Для проверки использовался метод Монте-Карло. Исходный набор данных разбивался на обучающее и тестовое множество случайным образом. Число повторений

1000. Пропорция между обучающим и тестовым множеством не изменялась: 2000 записей в обучающем множестве и 1059 в тестовом. На рисунке 6 показано как меняются линии тренда для прогноза длительности операций в зависимости от разбиения исходного набора данных на обучающее и тестовое множества.

Диапазон наблюдаемых значений логарифмов длительностей операций от 2,302 до 6,0867 (показаны на рисунке 6 вертикальными черными линиями). Наибольший разброс значений линий тренда находится на границах указанного диапазона и составляет $\pm 3,5\%$ слева и $\pm 3,0\%$ справа от средней линии (показана красным цветом). На рисунке 7 показаны характеристики ошибок для разных линий трендов (всего их -1000).

Из 1000 разбиений гипотеза о нормальности ошибок не подкрепляется тестом только в 1% случаев (p -value менее 0,05). В 99% случаев ошибки нормальны с практически нулевым средним и стандартным отклонением, сосредоточенным в диапазоне от 0,216 до 0,274 (рисунок 7, в центре). При корректировке ошибка несколько увеличивается, так как $\varepsilon_1^* = -\frac{\varepsilon_1}{a_1}$ и $a_1 < 1$ (см. формулу (2)). Стандартное отклонение $\frac{\sigma_1}{a_1}$ ошибок ε_1^* сосредоточено в диапазоне от 0,405 до 0,517 со средним значением 0,456 (см. рисунок 8). Соответственно, ε_1^* также в 99% случаев нормальны. Их среднее значение находится вблизи нуля.

Таким образом, при смене разбиений исходных данных на обучающее и тестовое множества, изменение оценок логарифмов длительностей операции не превосходит 3-3,5% на границах интервала наблюдений (см. рис.6). При корректировке предсказанных логарифмов

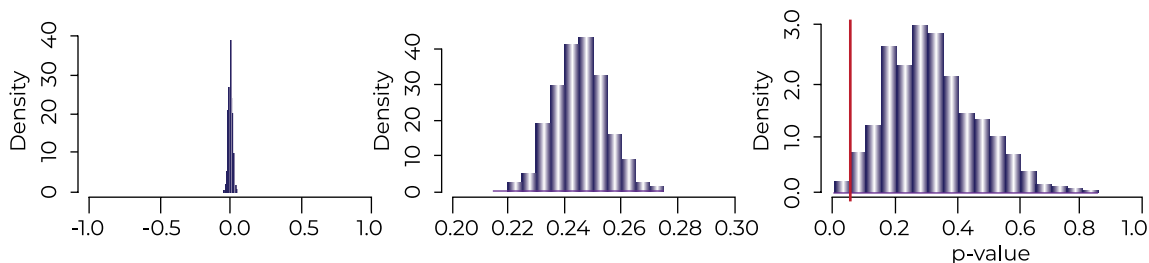


Рисунок 7 – Слева – средние значения ошибок, в центре – стандартные отклонения ошибок, справа значения p -value теста Колмогорова-Смирнова на нормальность ошибок.

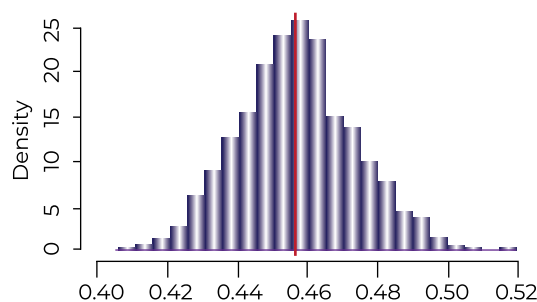


Рисунок 8 – Распределение стандартного отклонения ошибок скорректированных логарифмов длительностей операций ξ_1^* .

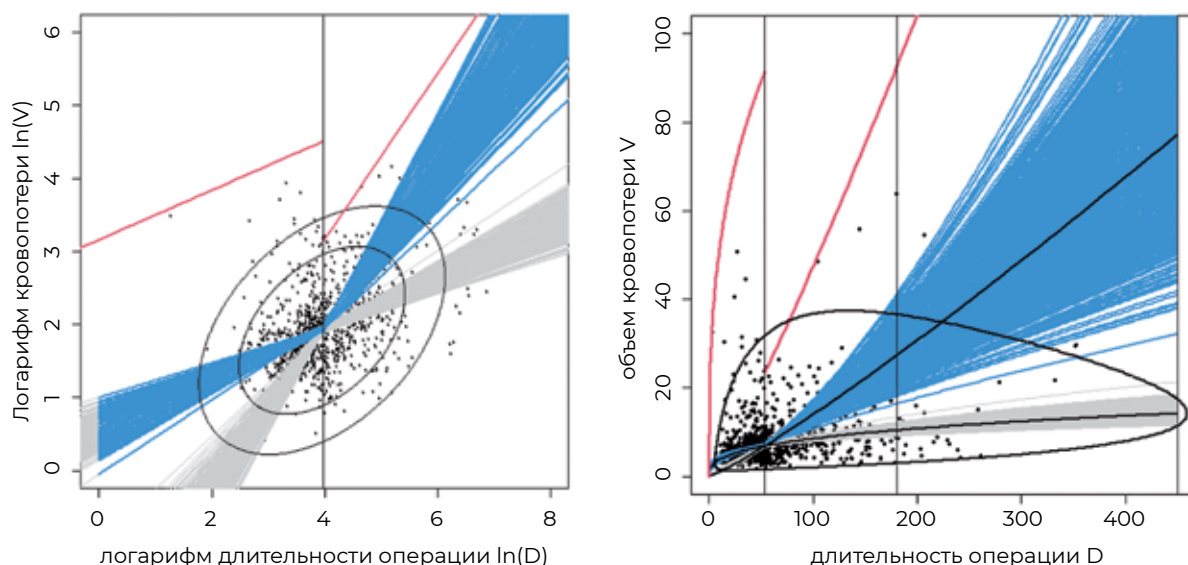


Рисунок 9 – Слева – рассеяние точек $(\ln(D), \ln(V_{np}))$. По горизонтали – предсказанный и скорректированный логарифм длительности операции, по вертикали – предсказанный логарифм кровопотери. Справа – рассеяние точек (D, V_{np}) . По горизонтали предсказанная и скорректированная длительность операции, по вертикали – объем кровопотери. Голубым цветом слева показаны линейные тренды, соответствующие разным разбиениям, справа – соответствующие им экспоненты. Тренды черного цвета соответствуют первому разбиению (см. рис. 3 и рис. 5).

длительностей операции ошибки уклонений нормальны с нулевым средним и стандартным отклонением от 0,405 до 0,517 (см. рис. 8). Размер стандартного отклонения порождается разбиением исходных данных на обучающее и тестовое множества. Мы проводим кросс-валидацию для того, чтобы понять как может меняться оценка кровопотери при смене разбиения

исходных данных на обучающее и тестовое множества. Наложим результаты кросс-валидации на рисунки 3 и 5 (результаты показаны на рисунке 9).

Как видно из рисунка 9 (справа) в большинстве случаев изменения трендов незначительны. Например, при длительности операции до 53,5 минут оценка кровопотери по медианному

значению практически не зависит от разбиения. При длительности операции свыше 53,5 минут разбиение начинает заметно влиять на оценку кровопотери. При $D=180$ минутам (3 часа), оценка кровопотери по медиане может меняться в зависимости от разбиения в диапазоне от 16 до 42 мл при среднем значении 28 мл (см. рис.9 (справа) диапазон голубого цвета при $D=180$). Чем больше длительность операции, тем выше неопределенность в прогнозе кровопотери. В исходных данных была одна запись с длительностью операции 450 минут, но операции такой длительности – редкость. Как видно из рисунка 9 (справа), при прогнозе длительности операции выше 100-120 минут требуется детальное изучение всех входных факторов и особая подготовка к проведению операции – риск высокой кровопотери значительно увеличивается.

ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

Предложена модель оценки будущей кровопотери, основанная на обучаемой модели градиентного бустинга. На формирование

медицинской статистики значительное влияние оказывает человеческий фактор. Его влияние проявляется в отклонениях влияющих факторов от ожидаемых значений, неточности измерений, индивидуальных особенностях пациентов и врачебных заключений и т.д. Данные отдельных пациентов могут значительно отличаться от основной массы наблюдений, однако исключение их из рассмотрения невозможно. Требуются отдельные исследования для выяснения причин аномальных отклонений тех или иных показателей. В любом случае, результаты обработки данных клиники могут служить только вспомогательным аналитическим материалом для лиц, принимающих решения — врачей. Своевременная визуализация повышенных рисков способна уменьшить опасность недооценки ситуации, и, следовательно, принятию более эффективных и безопасных решений. Полученный результат позволяет перейти к решению следующей задачи по оптимальной загрузке операционного блока с учетом всех сопутствующих обстоятельств и ограничений.

ЛИТЕРАТУРА/REFERENCES

1. Кобринский Б.А. Системы искусственного интеллекта в хирургии: возможности, ограничения и перспективы. Обзор литературы // Российский вестник детской хирургии, анестезиологии и реаниматологии. – 2023. – Т.13. – №3. – С.385-404. [Kobriniskii BA. Artificial intelligence systems in surgery: A review of opportunities, limitations, and prospects. Review. Russian Journal of Pediatric Surgery, Anesthesia and Intensive Care. 2023; 13(3): 385-404. (In Russ.)] doi: 10.17816/psaic1547.
2. Болдина Ю.С., Ившин А.А. Возможности машинного обучения в прогнозировании акушерских кровотечений // Акушерство, Гинекология и Репродукция. – 2024. – №18(3). – С.365-381. [Boldina YS, Ivshin AA. Machine learning opportunities to predict obstetric haemorrhages. Akusherstvo, Ginekologiya i Reprodukcia = Obstetrics, Gynecology and Reproduction. 2024; 18(3): 365-381. (In Russ.)] doi: 10.17749/2313-7347/ob.gyn.rep.2024.491.
3. Аль-Юсеф Н., Булатова Л.Р., Гуров О.Н. Искусственный интеллект в сосудистой хирургии: этический философский анализ технологического прогресса // Искусственные общества. – 2025. – Т.20. – Выпуск 2. [Al-Yusef N, Bulatova L, Gurov O. Artificial Intelligence in Vascular Surgery: an Ethico-Philosophical Analysis of Technological Advancement. Artificial societies. 2025; 20(2). (In Russ.)] doi: 10.18254/S207751800034958-2.
4. Гельцер Б.И., Шахгельдян К.И., Рублев В.Ю. и др. Методы машинного обучения в прогнозировании летальных исходов в стационаре у больных ишемической болезнью сердца после коронарного шунтирования // Кардиология. – 2020. – №60(10). – С.38-46. [Geltser BI, Shahgeldyan KJ, Rublev VY, et al. Machine Learning Methods for Prediction of Hospital Mortality in Patients with Coronary Heart Disease after Coronary Artery Bypass Grafting. Kardiologiya. 2020; 60(10): 38-46 (In Russ.)]
5. Овчаренко Е.А., Клышников К.Ю., Кутихин А.Г., Фролов А.В. Машинное обучение в задаче прогнозирования неблагоприятных сердечно-сосудистых событий у пациентов после коронарного шунтирования // Клиническая и экспериментальная хирургия. Журнал имени академика Б.В. Петровского. – 2023. – Т.11. – №3. – С.16-28. [Ovcharenko EA, Klyshnikov KYu, Kutikhin AG, Frolov AV. Machine learning in the problem of adverse cardiovascular events prognosis in patients

- after coronary artery bypass surgery. Clinical and Experimental Surgery. Petrovsky Journal. 2023; 11(3): 16-28. (In Russ.)] doi: 10.33029/2308-1198-2023-11-3-16-28.
6. Баранов А.А., Намазова-Баранова Л.С., Смирнов И.В. и др. Методы и средства комплексного интеллектуального анализа медицинских данных // Труды Института системного анализа Российской академии наук. – 2015. – Т.65. – №2. – С.81-93. [Baranov AA, Namazova-Baranova LS, Smirnov IV, et al. Methods and means of complex intellectual analysis of medical data. Proceedings of the Institute of System Analysis of the Russian Academy of Sciences. 2015; 65(2): 81-93. (In Russ.)]
7. Гусев А.В. Перспективы нейронных сетей и глубокого машинного обучения в создании решений для здравоохранения // Врач и информационные технологии. – 2017. – №3. – С.92-105. [Gusev AV. Perspectives of neural networks and deep machine learning in the creation of solutions for health care. Physician and information technologies. 2017; 3: 92-105.(In Russ.)]
8. Сахибгареева М.В., Заозерский А.Ю. Разработка системы прогнозирования диагнозов заболевания на основе искусственного интеллекта // Вестник Российского государственного медицинского университета. – 2017. – №6. – С.42-46. [Sakhibgareeva MV, Zaozersky AYU. Development of the system of forecasting disease diagnoses on the basis of artificial intelligence. Bulletin of the Russian State Medical University. 2017; 6: 42-46. (In Russ.)]
9. Медников Д.А., Протасова З.У., Крикунова Е.В. Интеллектуальные системы поддержки принятия врачебных решений в кардиологической практике. В сборнике: Юность и знания – гарантия успеха-2021. Сборник научных трудов 8-й Международной молодежной научной конференции. В 3 т. – Курск, 2021. – С.225-245. [Mednikov DA, Protasova ZU, Krikunova EV. Intellectual systems to support medical decision-making in cardiology practice. In Collection: Youth and knowledge – guarantee of success-2021. Collection of scientific papers of the 8th International Youth Scientific Conference. In 3 v. Kursk, 2021. P.225-245.(In Russ.)]
10. Boulitsakis LS, Green D, Holland M, Al Moubayed N. Predicting acute clinical deterioration with interpretable machine learning to support emergency care decision making. Sci. Rep. 2023; 13: 13563.
11. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Comput. Methods Programs Biomed. 2022; 214: 106584.
12. Feretzakis G, Sakagianni A, Anastasiou A, et al. Integrating Shapley Values into Machine Learning Techniques for Enhanced Predictions of Hospital Admissions. Appl. Sci. 2024; 14: 5925.
13. Зюзин В.С. Предоперационная оценка риска кровотечения у кардиохирургических пациентов // Хирургическая практика. – 2024. – №9(2). – С.80-96. [Zyuzin VS. Preoperative bleeding risk assessment in cardiac surgery patients. Surgical practice. 2024; 9(2): 80-96. (In Russ.)] doi: 10.5922/2223-2427-2024-9-2-6.
14. Матинян Н.В., Гьокова Е.В., Иванова Т.Е. и др. Массивная кровопотеря при высокотравматичных операциях в детской онкологии (проспективное рандомизированное исследование) // Анестезиология и реаниматология. – 2024. – №1. – С.31-40. [Matinyan NV, Gyokova EV, Ivanova TE, et al. Massive blood loss in pediatric cancer surgery. Anesteziologiya i Reanimatologiya. 2024; 1: 31-40. (In Russ.)] doi: 10.17116/anaesthesiology202401131.
15. Болдина Ю.С., Ившин А.А. Возможности машинного обучения в прогнозировании акушерских кровотечений // Акушерство, Гинекология и Репродукция. – 2024. – №18(3). – С.365-381. [Boldina YS, Ivshin AA. Machine learning opportunities to predict obstetric haemorrhages. Akusherstvo, Ginekologiya i Reprodukcya. 2024; 18(3): 365-381.(In Russ.)] doi: 10.17749/2313-7347/ob.gyn.rep.2024.491.
16. Абдуалимов Т.П., Обрезан А.Г. Выявление поражения коронарных артерий при помощи алгоритмов глубокого обучения // Кардиология: новости, мнения, обучение. – 2021. – Т.9. – №2(27). – С.9-13.[Abdualimov TP, Obrezan AG. Detection of coronary arterial lesions using deep learning algorithms. Cardiology: news, opinions, training. 2021; 9(2): 9-13. (In Russ.)]
17. Драпкина Ю.С., Макарова Н.П., Васильев Р.А., Амелин В.В., Калинина Е.А. Сравнение прогностических моделей, построенных с помощью разных методов машинного обучения, на примере прогнозирования результатов лечения бесплодия методом вспомогательных репродуктивных технологий // Акушерство и гинекология. – 2024. – №2. – С.97-105. [Drapkina YS, Makarova

- NP, Vasiliev RA, Amelin VV, Kalinina EA. Comparison of prognostic models built using different methods of machine learning, by the example of predicting the results of infertility treatment by assisted reproductive technologies. *Obstetrics and Gynecology*. 2024; 2: 97-105. (In Russ.)]
18. Корчагин М.П., Говоров А.В., Васильев А.О., Грицков И.О., Пушкарь Д.Ю. Нейросети в онкоурологии // *Вестник урологии*. – 2024. – Т.12. – №4. – С.91-101. [Korchagin MP, Govorov AV, Vasiliev AO, Gritskov IO, Pushkar DYU. Neural networks in oncology. *Vestnik urologii*. 2024; 12(4): 91-101. (In Russ.)]
 19. Драпкина Ю.С., Макарова Н.П., Васильев Р.А. и др. Изучение аналитической обработки клинико-анамнестических и эмбриологических данных пациентов в программе вспомогательных репродуктивных технологий различными методами машинного обучения // *Акушерство и гинекология*. – 2024. – №3. – С.96-107. [Drapkina YS, Makarova NP, Vasiliev RA, et al. Study of analytical processing of clinical-anamnestic and embryological data of patients in the program of assisted reproductive technologies by different methods of machine learning. *Obstetrics and Gynecology*. 2024; 3: 96-107. (In Russ.)]
 20. Huang T, Le D, Yuan L, Xu S, Peng X. Machine learning for prediction of in-hospital mortality in lung cancer patients admitted to intensive care unit. *PLoS ONE*. 2023; 18: e0280606.
 21. Xie F, Zhou J, Lee JW, et al. Benchmarking emergency department prediction models with machine learning and public electronic health records. *Sci. Data*. 2022; 9: 658.
 22. Tschoellitsch T, Seidl P, Böck C, et al. Using emergency department triage for machine learning-based admission and mortality prediction. *Eur. J. Emerg. Med*. 2023; 30: 408-416.
 23. Goto T, Camargo CA, Faridi MK, et al. Machine learning-based prediction of clinical outcomes for children during emergency department triage. *JAMA Netw. Open*. 2019; 2: e186937.
 24. Feretzakis G, Sakagianni A, Loupelis E, et al. Predicting Hospital Admission for Emergency Department Patients: A Machine Learning Approach. *Stud. Health Technol. Inform*. 2022; 289: 297-300.
 25. Гафанович Е.Я., Соколов И.М., Конобеева Е.В., Каширина И.Л., Фирюлина М.А. Исследование основных факторов влияния приверженности лечению статинами с использованием «деревьев решений» // *Моделирование, оптимизация и информационные технологии*. – 2021. – Т.9. – №3(34). [Gafanovich EY, Sokolov IM, Konobeeva EV, Kashirina IL, Firyulina MA. Study of the main factors of influence of adherence to treatment with statins using “decision trees”. *Modeling, optimization and information technologies*. 2021; 9(3). (In Russ.)]
 26. Голубков А.В., Гаврилова М.П. Классификация эпидемиологических прогнозов и применение градиентного бустинга в профилактической медицине // *Medline.ru. Российский биомедицинский журнал*. – 2021. – Т.22. – С.328-343. [Golubkov AV, Gavrilova MP. Classification of epidemiologic forecasts and application of gradient boosting in prophylactic medicine. *Medline.ru. Russian Biomedical Journal*. 2021; 22: 328-343. (In Russ.)]
 27. Миронов Д.С., Спирин И.А., Усынина А.А., Постоев В.А. Возможности применения методов машинного обучения для повышения качества пренатальной диагностики врожденных пороков развития: обзор предметного поля // *Врач и информационные технологии*. – 2025. – №3. – С.22-35. [Mironov DS, Spirin IA, Usynina AA, Postoev VA. Possibilities of applying machine learning methods to improve the quality of prenatal diagnosis of congenital malformations: scoping review. *Medical Doctor and Information Technologies*. 2025; 3: 22-35. (In Russ.)] doi: 10.25881/18110193_2025_3_22.
 28. Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat*. 2001; 29: 1189-1232.
 29. Джеймс Г, Уиттон Д, Хасты Т, Тибширани Р. Введение в статистическое обучение с примерами на языке R. Москва: ДМК-пресс, 2018. [James G, Whitton D, Hastie T, Tibshirani R. Introduction to statistical learning with examples in R. Moscow: DMK-press, 2018. (In Russ.)]
 30. Кабаков Р.И. R в действии. Анализ и визуализация данных на языке R. Москва: ДМК-пресс, 2014. [Kabakov RI. R in action. Data analysis and visualization in the R language. Moscow: DMK-press, 2014. (In Russ.)]
 31. Паршакова М.С., Семенов А.А., Ласкин М.Б. Оценка длительности операции ансамблевыми методами решающих деревьев // *Информатизация и связь*. – 2025. – №3. – С.89-97. [Parshakova MS, Semenov AA, Laskin MB. Operation duration estimation by ensemble methods of decision trees. *Informatization and communication*. 2025; 3: 89-97. (In Russ.)] doi: 10.34219/2078-8320-2025-16-3-89-97.