[1] Российский государственный социальный университет

[2] Российский университет медицины

[3] Московский клинический научный центр им. А.С. Логинова

Сравнительная оценка возможностей медицинской статистики и интерпретируемого машинного обучения в анализе онкометаболических биомаркеров

К.И. Баирова, к.м.н.[1], А.М. Мкртумян, д.м.н., проф.[2, 3]

Адрес для переписки: Кермен Ивановна Баирова, bairova@list.ru

Для цитирования: Баирова К.И., Мкртумян А.М. Сравнительная оценка возможностей медицинской статистики и интерпретируемого машинного обучения в анализе онкометаболических биомаркеров. Эффективная фармакотерапия. 2026; 22 (38): 28–33.

DOI 10.33978/2307-3586-2026-22-38-28-33

Поиск доступных биомаркеров, позволяющих охарактеризовать метаболические паттерны, ассоциированные с онкологическими заболеваниями, представляется актуальным направлением современной персонализированной медицины. Метаболические, воспалительные и адипокиновые показатели тесно взаимосвязаны, при этом их ассоциация с онкологическим процессом может носить нелинейный характер. Как следствие, изолированный анализ отдельных показателей не всегда отражает онкометаболический профиль пациента.

***Цель исследования** – сравнить возможности традиционной медицинской статистики и интерпретируемого машинного обучения в анализе онкометаболических биомаркеров и формировании комплексного метаболического фенотипа пациента.*

***Материал и методы.** Проведен научно-методический анализ результатов собственных исследований, включавших оценку метаболических биомаркеров у пациентов с впервые выявленными злокачественными новообразованиями. Рассмотрены возможности классического статистического анализа, моделей логистической регрессии, Elastic Net, Decision Tree и CatBoost, а также методов интерпретации SHAP, CatBoost Feature Importance, SHAP Waterfall и анализа коэффициентов регуляризованных моделей.*

***Результаты.** Медицинская статистика позволяет выявлять статистически значимые межгрупповые различия и оценивать отдельные биомаркеры в клиническом контексте. Однако машинное обучение дает возможность анализировать многомерные комбинации признаков, учитывать нелинейные связи и устанавливать их совокупный вклад в классификацию пациентов. При анализе собственных данных наилучшее качество классификации продемонстрировала модель CatBoost (AUROC = 0,9805, AUPRC = 0,9711, F1-score = 0,9302), а независимые методы интерпретации согласованно выделили ведущую роль гипергликемии, глюкозы натощак, индекса HOMA-IR, параметров перорального глюкозотолерантного теста и количества компонентов метаболического синдрома.*

***Заключение.** Традиционная медицинская статистика и интерпретируемое машинное обучение не являются взаимоисключающими подходами. Их сочетанное применение позволяет перейти от изолированной оценки отдельных биомаркеров к системному анализу онкометаболического фенотипа и повышает клиническую интерпретируемость результатов.*

***Ключевые слова:** онкометаболические биомаркеры, медицинская статистика, интерпретируемое машинное обучение, SHAP, CatBoost, инсулинорезистентность*

Введение

Рост распространенности онкологических заболеваний и высокая частота их выявления на поздних стадиях определяют необходимость поиска доступных биомаркеров для ранней стратификации риска и персонализации профилактических подходов [1]. В последние годы особое внимание уделяется метаболическим нарушениям, поскольку висцеральное ожирение, инсулинорезистентность (ИР), компенсаторная гиперинсулинемия, хроническое низкоинтенсивное воспаление и адипоки-

новая дисрегуляция рассматриваются как взаимосвязанные элементы онкометаболического контекста [2–7].

Клиническая сложность решения данной проблемы заключается в том, что метаболические биомаркеры не существуют изолированно. Уровень глюкозы, инсулина, индексы ИР, показатели липидного обмена, адипокины и воспалительные маркеры формируют единую биологическую сеть. Вклад показателя нередко зависит от наличия других признаков, а сами ассоциации могут носить нелинейный характер. Именно поэтому анализ одного отдельно взятого параметра может не отражать полного онкометаболического паттерна.

Обязательной частью клинических исследований остаются традиционные методы медицинской статистики. Они позволяют описывать выборку, сравнивать группы, оценивать статистическую значимость различий и проверять заранее сформулированные гипотезы. Вместе с тем при анализе сложных биомедицинских данных требуются инструменты, способные выявлять скрытые многомерные закономерности. В этом контексте методы машинного обучения, особенно интерпретируемые модели, могут рассматриваться как важное дополнение к классической статистике [8–10].

Особое значение для клинической интерпретации таких моделей имеют алгоритмы градиентного бустинга, включая CatBoost (Categorical Boosting), и методы объяснимого искусственного интеллекта, в частности SHAP (Shapley Additive Explanations), позволяющие определить вклад признаков на уровне всей выборки и отдельного пациента.

Целью настоящей работы стало сравнение возможностей медицинской статистики и интерпретируемого машинного обучения в анализе онкометаболических биомаркеров, а также разработка модели их совместного применения в клинико-исследовательской практике.

Материал и методы

Работа носила научно-методический характер. В ходе ее проведения нами были сопоставлены два аналитических подхода: традиционная медицинская статистика (ТМС) и интерпретируемое машинное обучение (ИМО). В качестве эмпирической основы взяты результаты собственных исследований, посвященных оценке метаболических биомаркеров у пациентов с впервые выявленными злокачественными новообразованиями.

Первый исследовательский блок включал сравнительный анализ данных 210 пациентов, 110 из которых были без онкологических заболеваний и 100 – с впервые выявленными злокачественными новообразованиями [11]. В данной когорте анализировали клинико-лабораторные и метаболические показатели, включая параметры углеводного обмена, индексы ИР, параметры липидного профиля, индекс висцерального ожирения (ИВО), адипокины и воспалительные маркеры.

Для бинарной классификации применяли следующие модели: Logistic Regression, Elastic Net, Decision Tree и CatBoost. Модели интерпретировали с помощью SHAP, внутренней важности признаков CatBoost, структуры решающего дерева и коэффициентов Elastic Net [11].

Во втором исследовательском блоке для демонстрации возможностей традиционного статистического анализа использовали собственные расчеты, выполненные на выборке из 130 пациентов с впервые диагностированными онкологическими заболеваниями, стратифицированных по индексу массы тела (ИМТ) и возрасту. Такой подход позволил сопоставить информативность ИМТ- и возрастной стратификаций при оценке индексов ИР, метаболических, воспалительных и адипокиновых показателей.

Из индексов ИР были оценены индексы HOMA-IR (Homeostatic Model Assessment for Insulin Resistance), Caro, Matsuda, соотношение триглицеридов и липопротеинов высокой плотности (ТГ/ЛПВП), из расчетных индексов – метаболический индекс (МИ), индекс накопления липидов (ИНЛ), ИВО. Нами также были определены параметры перорального глюкозотолерантного теста (ПГТТ), уровень иммунореактивного инсулина (ИРИ), интерлейкина 10 (IL-10), висфатина и резистина.

Статистический анализ включал сравнение независимых групп, оценку категориальных признаков и интерпретацию значений p. Машинно-обучающий анализ оценивался по метрикам AUROC, AUPRC, Recall, Precision и F1-score. Для клинической интерпретации результатов рассматривались глобальная важность признаков, структура дерева решений, направление ассоциаций в регуляризованных линейных моделях и индивидуальные SHAP Waterfall-графики.

Результаты и их обсуждение

Роль медицинской статистики в анализе онкометаболических биомаркеров

Медицинская статистика является фундаментальной основой доказательной медицины. Она позволяет ответить на базовые вопросы: различаются ли группы между собой, насколько выражены эти различия и являются ли они статистически значимыми? Для оценки онкометаболического профиля обычно используют U-критерий Манна – Уитни, критерий Краскела – Уоллиса, хи-квадрат Пирсона (критерий χ^2), точный критерий Фишера, корреляционный анализ и регрессионные модели.

Преимуществом статистического подхода является прозрачность. Для врача понятны различия медиан, p-значение, доверительный интервал или коэффициент корреляции. Так, статистически значимое повышение значений глюкозы, ИРИ или индекса HOMA-IR у пациентов с онкологическими заболеваниями может быть интерпретировано как более выраженное нарушение углеводного обмена.

Однако классический статистический анализ чаще всего оценивает отдельный параметр или заранее заданную модель, что создает ограничения при изучении сложных биологических систем. Метаболические показатели могут коррелировать между собой, один и тот же биомаркер – иметь разные значения при сочетании с другими признаками, а связь между метаболической дисфункцией и онкологическим процессом – быть нелинейной. Сравнительная характеристика возможностей медицинской статистики и ИМО представлена в табл. 1.

Ограничения изолированного анализа отдельных биомаркеров

Изолированный анализ биомаркера удобен для клинической интерпретации, но может быть недостаточным для понимания системной метаболической дисфункции. Уровень глюкозы натощак отражает состояние углеводного обмена, но не показывает компенсаторную гиперинсулинемию. Индекс HOMA-IR характеризует базальную ИР, но не динамический ответ на глюкозную нагрузку. Индекс Matsuda оценивает общую чувствительность тканей к инсулину, однако требует проведения ПГТТ и определения уровня инсулина в динамике [12, 13]. Значения адипокинов и воспалительных маркеров отражают функциональное состояние жировой ткани и воспалительный фон, но сами по себе не являются универсальными диагностическими критериями. Умеренное повышение уровня глюкозы может иметь разное клиническое значение у пациента с нормальным уровнем инсулина и у пациента с выраженной гиперинсулинемией, высоким индексом HOMA-IR, сниженным индексом Matsuda и повышенным уровнем резистина. Именно поэтому при анализе онкометаболического риска важно переходить с оценки отдельных показателей к анализу метаболического паттерна, включающего гипергликемию, гиперинсулинемию, повышение индекса HOMA-IR, снижение индексов Matsuda и Caro, увеличение числа компонентов метаболического синдрома, изменение адипокинового профиля и снижение уровня противовоспалительных маркеров, включая IL-10. Результаты статистического анализа онкометаболических биомаркеров представлены в табл. 2.

Возможности машинного обучения в анализе многомерных метаболических паттернов

Машинное обучение позволяет анализировать большое количество признаков одновременно и выявлять сложные комбинации переменных, связанных с принадлежностью пациента к определенной группе. В отличие от классического статистического подхода оно не ограничивается заранее заданными линейными связями и может учитывать нелинейные зависимости, взаимодействия между признаками и скрытые многомерные структуры данных.

В собственном исследовании для бинарной классификации пациентов применялось семейство моделей логистической регрессии, Elastic Net, Decision Tree и CatBoost [11]. На тестовой выборке модель CatBoost продемонстрировала наилучшее качество классификации – AUROC = 0,9805, AUPRC = 0,9711, Recall = 0,9091, Precision = 0,9524 и F1-score = 0,9302 (табл. 3) [11]. При использовании модели Decision Tree были получены более низкие метрики. Однако она обладала высокой клинической интерпретируемостью. Линейные модели имели меньшую дискриминирующую способность.

Принципиальное отличие машинного обучения заключается в том, что оно отвечает не только на вопрос о наличии различий между группами, но и на вопрос о том, какие признаки и их сочетания вносят наибольший вклад в отнесение конкретного пациента к одной из сравниваемых групп.

Клиническая значимость интерпретируемого машинного обучения

Одним из главных ограничений использования машинного обучения в медицине является проблема черного ящика. Высокие показатели классификации сами по себе недостаточны для клинического применения, если непонятно, на основании каких признаков модель приняла решение. В онкологии это особенно важно, поскольку необходимо не только определить принадлежность пациента к группе с впервые выявленными злокачественными новообразованиями или к группе без онкологических заболеваний, но и объяснить, какие именно нарушения сформировали неблагоприятный онкометаболический профиль.

Интерпретируемое машинное обучение позволяет перевести технический результат модели в клинически понятную форму. Так, SHAP-анализ показывает глобальный вклад признаков в решение модели, модель Decision Tree визуализирует последовательность принятия решений, модель Elastic Net позволяет оценить направление ассоциации признаков в регуляризованной линейной модели, модель SHAP Waterfall объясняет решение на уровне конкретного пациента.

При использовании собственных данных авторов продемонстрирована высокая согласованность независимых методов интерпретации. В частности, ведущими предикторами были гипергликемия, глюкоза натощак, индекс HOMA-IR, параметры ПГТТ и количество компонентов метаболического синдрома [11]. Антропометрические и липидные показатели имели меньшую классификационную значимость, чем функциональные маркеры углеводного обмена и ИР. Согласованность методов интерпретации представлена в табл. 4 [11].

Практическая модель совместного применения статистики и машинного обучения

Наиболее рациональным является не противопоставление медицинской статистики и машинного обучения, а их последовательное объединение в рамках единого исследовательского дизайна. На первом этапе статистический анализ описывает выборку, выявляет межгрупповые различия и формирует клинически значимые гипотезы. На втором этапе машинное обучение оценивает совокупный вклад признаков и строит модель классификации. На третьем этапе методы интерпретации позволяют объяснить, какие биомаркеры определяют результат модели.

Данный подход особенно важен при анализе онкометаболических биомаркеров, поскольку позволяет сохранить клиническую прозрачность классической статистики и одновременно использовать возможности машинного обучения для выявления многомерных нелинейных паттернов. В результате врач получает не только перечень статистически значимых различий, но и объяснимую модель онкометаболического фенотипа пациента.

Таблица 1. Сравнительная характеристика медицинской статистики и интерпретируемого машинного обучения

Критерий	Медицинская статистика	Интерпретируемое машинное обучение
Основной вопрос	Есть ли статистически значимые различия между группами?	Можно ли классифицировать пациента по совокупности признаков?
Единица анализа	Отдельный показатель или заранее заданная модель	Многомерный набор признаков
Тип связей	Связи чаще линейные или заранее определенные	Может учитывать нелинейные связи и взаимодействия
Основной результат	Медианы, p-значения, доверительные интервалы, корреляции	AUROC, AUPRC, Recall, Precision, F1-score, важность признаков
Сильные стороны	Проверка гипотез и описание различий	Выявление многомерных паттернов и комбинаций признаков
Ограничения	Трудность учета сложных взаимодействий между признаками	Риск переобучения и необходимость внешней валидации
Оптимальная роль	Первичный анализ, клиническое описание выборки	Стратификация риска и объяснение многомерных решений

*Таблица 2. Пример результатов статистического анализа онкометаболических биомаркеров**

Аналитический блок	Статистически значимые показатели	Направление изменений	Клиническая интерпретация
Онкологическая группа vs группа сравнения	Глюкоза, ИРИ, индекс HOMA-IR, ИВО, резистин, индексы Matsuda и Caro, IL-10	В группе онкологических больных выше значения глюкозы, ИРИ, индексов HOMA-IR, ИВО и резистина, при этом ниже значения индексов Matsuda и Caro, а также уровня IL-10	Функциональные метаболические нарушения выражены сильнее, чем различия в отношении значений ИМТ и ОТ
Стратификация онкологической группы по ИМТ	Индексы HOMA-IR, Caro и Matsuda, а также МИ, ИНЛ, ПГТТ-глюкоза, ПГТТ-инсулин, ИРИ, IL-10, висфатин, резистин	При возрастании ИМТ повышаются значения индекса HOMA-IR, ИНЛ, ПГТТ-инсулина, ИРИ, висфатина и резистина, снижаются значения индексов Caro и Matsuda, а также IL-10	Увеличение ИМТ связано с усилением ИР, гиперинсулинемии и воспалительно-адипокиновой дисрегуляции
Возрастная стратификация	Индекс Matsuda, ТГ/ЛПВП выше нормы, высокий риск по ИВО, вес, ИМТ, повышенная ОТ, АЛТ, средний инсулин при ПГТТ	Различия менее системные, чем при стратификации по ИМТ	Стратификация по ИМТ лучше отражает онкометаболический фенотип, чем изолированный учет возраста

* Сводная таблица составлена авторами на основании результатов собственных исследований и расчетов.
Примечание. ОТ – окружность талии, АЛТ – аланинаминотрансфераза.

*Таблица 3. Сравнение качества моделей машинного обучения при классификации пациентов**

Модель	AUROC	AUPRC	Recall	Precision	F1-score
CatBoost	0,9805	0,9711	0,9091	0,9524	0,9302
Decision Tree	0,8701	0,8800	0,9091	0,7692	0,8333
Logistic Regression (Elastic Net)	0,7581	0,6604	0,7727	0,6071	0,6800
Logistic Regression (L2, Ridge)	0,7614	0,6626	0,7727	0,6071	0,6800
Logistic Regression (L1, Lasso)	0,7256	0,6158	0,8636	0,5278	0,6552
Logistic Regression (без регуляризации)	0,6761	0,5664	0,7273	0,5926	0,6531

* Метрики рассчитаны на тестовой выборке по данным собственного исследования [11].
Примечание. Recall соответствует чувствительности, Precision – положительной прогностической ценности.

Таблица 4. Согласованность методов интерпретации машинного обучения в выявлении ключевых онкометаболических биомаркеров

Метод интерпретации	Что показывает	Ведущие признаки	Клинический смысл
SHAP Global Importance	Глобальный вклад признаков в классификацию	Гипергликемия, глюкоза натощак, индекс HOMA-IR, ПГТТ-глюкоза/инсулин, num_crit	Основной вклад связан с углеводным обменом и ИР
CatBoost Feature Importance	Внутренняя важность признаков в модели CatBoost	Гипергликемия, глюкоза, индекс HOMA-IR, ПГТТ-показатели, масса тела, num_crit	Подтверждает глюкозоцентрированный метаболический паттерн
Decision Tree	Логика клинических правил	Первичный split – гипергликемия, далее – индекс HOMA-IR, возраст, индекс Caro, IL-10	Углеводный обмен находится в корне классификации
Elastic Net Coefficients	Направление ассоциации признаков	Num_crit, ПГТТ-инсулин, ИВО, глюкоза, индекс HOMA-IR	Показывает направление связи признаков в линейной модели
SHAP Waterfall	Индивидуальное объяснение решения	Признаки, смещающие решение к классу 0 или 1 у конкретного пациента	Позволяет понять решение модели на уровне пациента

Примечание. Num_crit – количество компонентов метаболического синдрома.



Статистика проверяет различия, машинное обучение выявляет паттерны, интерпретируемость делает результат понятным врачу

Интегральная модель анализа онкометаболических биомаркеров

В клинической практике совместное применение статистики и машинного обучения может быть использовано для выявления пациентов с неблагоприятным метаболическим профилем. Речь, в частности, идет о больных, у которых традиционные антропометрические показатели могут быть невыраженными, однако уже имеют место функциональные нарушения углеводного обмена, гиперинсулинемия, снижение чувствительности к инсулину и адипокиново-воспалительная дисрегуляция.

Ограничения

Несмотря на методологические преимущества машинного обучения, его применение в медицинских исследованиях требует осторожности. Качество модели зависит от качества исходных данных; его снижают малый объем выборки, пропущенные значения, неоднородность пациентов и отсутствие внешней валидации.

Высокая классификационная эффективность не доказывает причинно-следственную связь. Если SHAP показывает высокий вклад глюкозы или индекса HOMA-IR в классификацию, это означает вклад в решение модели, но не доказывает причинную роль данного показателя в развитии онкологического заболевания. Даже интерпретируемая модель должна быть использована врачом с учетом анамнеза, локализации опухоли, стадии заболевания, сопутствующей патологии и лабораторного контекста.

Перспективы дальнейших исследований

Перспективным направлением является создание интегральных моделей онкометаболического фенотипирования, включающих клинические, антропометрические, лабораторные, воспалительные и адипокиновые показатели. Такие модели должны решать не задачу замены врача, а задачу поддержки принятия решений и выделения пациентов с неблагоприятным метаболическим профилем.

Будущие исследования должны включать проспективный дизайн, многоцентровые выборки, внешнюю валидацию моделей, анализ отдельных локализаций опухолей, сравнение разных алгоритмов машинного обучения и оценку клинической применимости объяснимых моделей. Особое значение имеет разработка визуальных инструментов, позволяющих врачу быстро интерпретировать вклад каждого биомаркера. Обобщенная схема совместного применения ТМС и ИМО при анализе онкометаболических биомаркеров представлена на рисунке.

Заключение

В анализе онкометаболических биомаркеров ТМС и ИМО выполняют разные, но взаимодополняющие функции. Статистические методы позволяют строго описать различия между группами и проверить клинические гипотезы, машинное обучение – проанализировать многомерные данные, выявить нелинейные паттерны и оценить совокупный вклад взаимосвязанных признаков.

Наиболее перспективным представляется комбинированный подход, при котором ТМС используется для первичного анализа и подтверждения различий, а ИМО – для выявления ключевых биомаркеров, анализа их взаимодействия и объяснения решений модели. Такой подход может способствовать переходу от изолированной оценки отдельных показателей

к системному пониманию онкометаболического фенотипа пациента.

В условиях роста онкозаболеваемости и необходимости поиска доступных биомаркеров интерпретируемое машинное обучение может стать важным инструментом клинической аналитики, персонализированной стратификации риска и разработки новых направлений в области профилактической онкологии.

Финансирование. *Работа выполнена без внешнего финансирования.*

Конфликт интересов. *Авторы заявляют об отсутствии конфликта интересов.*

Литература

1. Bray F., Laversanne M., Sung H., et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 2024; 74 (3): 229–263.
2. Després J.P., Lemieux I. Abdominal obesity and metabolic syndrome. Nature. 2006; 444 (7121): 881–887.
3. Gregor M.F., Hotamisligil G.S. Inflammatory mechanisms in obesity. Annu. Rev. Immunol. 2011; 29: 415–445.
4. Chiefari E., Mirabelli M., La Vignera S., et al. Insulin resistance and cancer: in search for a causal link. Int. J. Mol. Sci. 2021; 22 (20): 11137.
5. Pollak M. Insulin and insulin-like growth factor signalling in neoplasia. Nat. Rev. Cancer. 2008; 8 (12): 915–928.
6. Ahima R.S., Flier J.S. Adipose tissue as an endocrine organ. Trends Endocrinol. Metab. 2000; 11 (8): 327–332.
7. Caruso A., Gelsomino L., Panza S., et al. Leptin: a heavyweight player in obesity-related cancers. Biomolecules. 2023; 13 (7): 1084.
8. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer; 2009.
9. Lundberg S.M., Lee S.I. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems. 2017; 31: 4765–4774.
10. Prokhorenkova L., Gusev G., Vorobev A., et al. CatBoost: unbiased boosting with categorical features. Advances in Neural Information Processing Systems. 2018; 32: 6639–6649.
11. Bairova K.I., Mkrtumyan A.M. Interpretable machine learning for the identification of key metabolic biomarkers associated with newly diagnosed malignancies. Am. J. Clin. Exp. Med. 2026; 14 (5): 119–128.
12. Matthews D.R., Hosker J.P., Rudenski A.S., et al. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985; 28 (7): 412–419.
13. Matsuda M., DeFronzo R.A. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. Diabetes Care. 1999; 22 (9): 1462–1470.

Comparative Potential of Medical Statistics and Interpretable Machine Learning in the Analysis of Oncometabolic Biomarkers

K.I. Bairova, PhD[1], A.M. Mkrtumyan, PhD, Prof.[2, 3]

[1] *Russian State Social University*
[2] *Russian University of Medicine*
[3] *A.S. Loginov Moscow Clinical Scientific Center*

Contact person: Kermen I. Bairova, bairova@list.ru

The identification of accessible biomarkers that characterize metabolic patterns associated with malignancies is relevant to personalized medicine. Metabolic, inflammatory and adipokine markers are interrelated, and their associations with cancer may be nonlinear; therefore, isolated assessment of individual parameters may not reflect the overall oncometabolic phenotype.

The aim of the study *– to compare conventional medical statistics and interpretable machine learning in the analysis of oncometabolic biomarkers and integrated metabolic phenotype formation.*

Material and methods. *A comparative methodological analysis was performed using the authors' own data on metabolic biomarkers in patients with newly diagnosed malignancies. The possibilities of conventional statistical analysis, Logistic Regression, Elastic Net, Decision Tree and CatBoost were assessed. Model interpretation was based on SHAP analysis, CatBoost feature importance, SHAP waterfall plots and analysis of coefficients in regularized models.*

Results. *Medical statistics identified significant between-group differences and enabled clinical evaluation of individual biomarkers. Machine learning analyzed multidimensional feature combinations, nonlinear relationships and their joint contribution to patient classification. In the authors' dataset, CatBoost showed the best classification performance (AUROC = 0.9805, AUPRC = 0.9711, F1-score = 0.9302). Independent interpretation methods consistently highlighted hyperglycemia, fasting glucose, HOMA-IR, oral glucose tolerance test parameters and the number of metabolic syndrome components.*

Conclusion. *Conventional statistics and interpretable machine learning are complementary approaches. Their combined use supports a transition from isolated biomarkers to systems-level analysis of the oncometabolic phenotype.*

Keywords: *oncometabolic biomarkers, medical statistics, interpretable machine learning, SHAP, CatBoost, insulin resistance*