

Ж.Ж. Смирнова, Д.Ю. Бобров, А.А. Завьялов

ПРОГНОЗИРОВАНИЕ РЕЗУЛЬТАТОВ ДОЗИМЕТРИЧЕСКОЙ ВЕРИФИКАЦИИ ПЛАНОВ ЛУЧЕВОЙ ТЕРАПИИ С ПОМОЩЬЮ МЕТОДОВ МАШИННОГО И ГЛУБОКОГО ОБУЧЕНИЯ: ОБЗОР МЕТОДОЛОГИЧЕСКИХ ПОДХОДОВ

Федеральный медицинский биофизический центр им. А.И. Бурназяна ФМБА России, Москва

Контактное лицо: Жаннета Жамильевна Смирнова, e-mail: smirnova_zhanneta@mail.ru

РЕФЕРАТ

Введение: Современные методы лучевой терапии (IMRT, VMAT, SRS) требуют обязательной дозиметрической верификации каждого плана лечения перед началом лечения. Ключевым критерием оценки является показатель Gamma Passing Rate (GPR). Традиционный процесс верификации, основанный на физических измерениях, является трудоемким и снижает пропускную способность клиники. Это создает потребность в разработке эффективных методов оптимизации процедуры пациент-ориентированного контроля качества.

Цель: Обзор и анализ современных методологических подходов и алгоритмов машинного (МО) и глубокого обучения (ГО), применяемых для прогнозирования результатов GPR, а также определение перспективных направлений для будущих исследований в области виртуального контроля качества.

Заключение: Прогнозирование GPR с помощью методов МО и ГО представляет собой быстро развивающуюся и перспективную область. Обзор демонстрирует эволюцию подходов: от анализа единичных метрик сложности плана к использованию ансамблевых регрессионных моделей и, далее, к сложным архитектурам глубокого обучения. Исследования показывают, что разработка таких моделей позволяет с высокой точностью идентифицировать планы с риском неудачной верификации, что открывает путь к риск-ориентированному подходу и значительному сокращению рутинных измерений. Ключевыми задачами на пути к широкой клинической интеграции остаются обеспечение интерпретируемости моделей, преодоление дисбаланса классов данных, повышение общей применимости моделей и их интеграция в рабочие процессы клиник. Успешное решение этих задач позволит создать интеллектуальные системы поддержки принятия решений, способные повысить эффективность, безопасность и стандартизацию лучевой терапии.

Ключевые слова: лучевая терапия, пациент-ориентированный контроль качества, GPR, гамма-анализ, прогнозирование, машинное обучение, глубокое обучение, виртуальный контроль качества

Для цитирования: Смирнова Ж.Ж., Бобров Д.Ю., Завьялов А.А. Прогнозирование результатов дозиметрической верификации планов лучевой терапии с помощью методов машинного и глубокого обучения: обзор методологических подходов // Медицинская радиология и радиационная безопасность. 2026. Т. 71. № 2. С. 147–152. DOI:10.33266/1024-6177-2026-71-2-147-152

Zh.Zh. Smirnova, D.Yu. Bobrov, A.A. Zavialov

Predicting Gamma Passing Rate for Patient Specific Quality Assurance Using Machine and Deep Learning: A Review of Methodological Approaches

A.I. Burnazyan Federal Medical Biophysical Center, Moscow, Russia

Contact person: Zh.Zh. Smirnova, e-mail: smirnova_zhanneta@mail.ru

ABSTRACT

Introduction: In radiation therapy using advanced techniques such as intensity modulated radiation therapy (IMRT) and volumetric-arc radiation therapy (VMAT), patient specific quality assurance (QA) should be performed before treatment. The measured and planned dose distributions are commonly quantified by means of a gamma analysis (Gamma Passing Rate, GPR). However, patient-specific QA procedures are requiring significant time and effort by the physicists. Various ML and DL models have shown promising prediction accuracy and a high potential as time-efficient virtual QA tool.

Purpose: In this paper, we review the ML and DL based models that were developed for patient specific IMRT and VMAT QA GPR predictions, as well as to identify perspective directions for future research in the field of virtual QA.

Conclusion: The prediction of Gamma Passing Rates (GPR) using Machine Learning (ML) and Deep Learning (DL) is a rapidly evolving and highly promising field. This review demonstrates the evolution of methodological approaches, from the analysis of individual plan complexity metrics to the application of ensemble regression models, and further to sophisticated deep learning architectures. Research confirms that these predictive models can accurately identify plans at risk of verification failure, paving the way for a risk-based approach and a significant reduction in routine measurements. Key challenges for broader clinical integration remain, including ensuring model interpretability, overcoming class imbalance in datasets, improving model generalizability, and their integration into clinical workflows. Successfully addressing these challenges will enable the creation of intelligent decision-support systems capable of enhancing the efficiency, safety, and standardization of radiotherapy.

Keywords: radiotherapy, PSQA, GPR, gamma analysis, prediction, Machine Learning, Deep Learning, virtual QA

For citation: Smirnova ZhZh, Bobrov DYu, Zavialov AA. Predicting Gamma Passing Rate for Patient Specific Quality Assurance Using Machine and Deep Learning: A Review of Methodological Approaches. Medical Radiology and Radiation Safety. 2026;71(2):147–152. (In Russian). DOI:10.33266/1024-6177-2026-71-2-147-152

Введение

Применение методов планирования конформной лучевой терапии, таких как лучевая терапия модулированной интенсивности IMRT, ротационная лучевая терапия с объемной модуляцией VMAT и стереотаксическая радиохирurgia SRS, обуславливает строгие требования к точности подведения дозы. Между расчитанным дозиметрическим планом в системе планирования (СП) и фактически доставленным распределением дозы могут возникать несоответствия из-за пределов точности динамических характеристик линейного ускорителя и ограничений в алгоритмах СП [1]. Поэтому применение таких методов регламентируется обязательным проведением пациент-ориентированного контроля качества лучевой терапии для дозиметрической верификации каждого плана перед началом лечения пациента [2].

Традиционно дозиметрическая верификация выполняется с помощью измерений в фантоме с использованием различных детекторов: ионизационных камер, диодных массивов, радиохромных пленок или электронных портальных устройств визуализации (EPID) [3, 4]. Для оценки приемлемости планов используется гамма-анализ (GPR) двух дозовых распределений, впервые предложенный в 1998 г. Low et al [5], в котором устанавливаются специальные критерии – разность доз и критерий расстояния, математически реализуемые через расчет так называемого гамма-индекса [6]. Показатель GPR определяется как процент точек измерения, для которых выполняется условие гамма-индекса ≤ 1 . Согласно рекомендациям AAPM Task Group 218, клинически приемлемым считается GPR $\geq 95\%$ для относительно мягких критериев 3%/3 мм и GPR $\geq 90\%$ для более строгих критериев 2%/2 мм.

Дозиметрическая верификация техник инверсного планирования является стандартом пациент-ориентированного контроля качества лучевой терапии, но из-за широкого их применения для обеспечения конформности и снижения дозовой нагрузки на критические органы, увеличивает объем рутинных измерений, снижает пропускную способность линейного ускорителя и накладывает дополнительную нагрузку на медицинских физиков [7]. Таким образом, традиционный подход становится неэффективным по времени и сопряженным со значительными эксплуатационными затратами и риском преждевременного износа оборудования.

В настоящее время развитие методов машинного обучения (МО) и искусственного интеллекта (ИИ) предлагает революционный подход к оптимизации процедур лучевой терапии путем создания виртуальных или прогностических моделей пациент-ориентированного контроля качества [8, 9]. Данные модели предназначены для прогнозирования результатов дозиметрической верификации (GPR) на основе анализа параметров плана лечения, что позволяет заблаговременно выявлять потенциально проблемные планы до проведения физических измерений.

Применение ИИ в радиационной онкологии и медицинской физике подробно описано в публикации МАГАТЭ IAEA-TCS-83, выпущенной в марте 2025 г. [10]. Документ подчеркивает важность ИИ в современной медицине, особенно в радиационной онкологии, диагностической визуализации и ядерной медицине, рассматривает различные виды ИИ, такие как МО, глубокое обучение (ГО) и большие данные. Документ служит важным руководством для медицинских физиков, подчеркивая их ключевую роль в интеграции ИИ в клиническую практику, а также подчеркивает необхо-

димость тщательного рассмотрения рисков и преимуществ ИИ.

Цель настоящей работы заключается в обзоре ключевых методологических подходов и используемых алгоритмов МО в области прогнозирования результатов пациент-ориентированного контроля качества для планов лучевой терапии, а также в определении перспективных направлений для будущих исследований.

Материал и методы

Проведен аналитический обзор зарубежной научной литературы. Были изучены исследования, посвященные созданию и валидации моделей МО и ГО для задач виртуального контроля качества в лучевой терапии. Критериями отбора служили научная значимость работы, наличие полнотекстового доступа, а также детальное описание методологии, используемых признаков и их влияние на результат гамма-анализа (GPR).

Результаты и обсуждение

Эволюция методологических подходов к прогнозированию результатов GPR

Первоначальные попытки предсказать результаты GPR были основаны на поиске корреляций между различными метриками сложности плана и результатом гамма-анализа, полученным в результате верификации. Идея заключалась в том, что чем сложнее план (выше модуляция и скорость движения лепестков коллиматора, меньше размеры апертур), тем выше вероятность ошибок доставки дозы и, следовательно, ниже GPR [11].

Было предложено множество таких метрик. McNiven et al [12] ввели количественную оценку сложности модуляции (Modulation Complexity Score, MCS), которая учитывает вариабельность позиций лепестков коллиматора и площади апертур между сегментами. Значение оценки сложности модуляции находится в диапазоне от 0 до 1, более низкое значение MCS (близкое к 0) ассоциируется с высокой модуляцией пучка (быстрое и сложное движение лепестков МЛК, малые апертуры), что потенциально может привести к большим расхождениям между рассчитанной и доставленной дозой. Более высокое значение MCS (близкое к 1) указывает на более простой и, как правило, более надежный в реализации план.

Park et al [13] предложили использование индекса модуляции Modulation Index (MI) для оценки скорости движения лепестков МЛК, который рассчитывается как отношение суммарной длины пути всех лепестков к длине дуги облучения. Более высокое значение MI указывает на более интенсивное и быстрое движение МЛК, что характерно для высокомодулированных планов.

В другой своей работе Park et al [14] показали, что наибольшую корреляцию с точностью подведения дозы демонстрирует не столько абсолютная скорость движения лепестков МЛК, сколько величина их ускорения. Объясняется это тем, что высокое ускорение создает значительную инерционную нагрузку на приводы коллиматора, приводя к динамическим ошибкам позиционирования и, как следствие, к ухудшению GPR.

Другие исследования были сфокусированы на таких параметрах, как длина пути лепестков МЛК и общее количество мониторинговых единиц (МЕ), а также на признаках, извлеченных с помощью радиометрических методов и флюоресцент-карт – например, контраста и энтропии [15, 16]. Преимущество таких признаков заключается в том, что они глобально описывают сложность всей картины излучения, которую должна точно воспроизвести механическая система ускорителя.

Хотя эти исследования и выявили статистически значимые корреляции, прогностическая сила отдельных метрик часто оказывалась недостаточной для надежного предсказания результатов дозиметрической верификации в клинической практике, особенно для строгих критериев гамма-анализа (например, 2 %/2 мм). Это связано с тем, что на точность доставки плана влияет совокупное воздействие множества факторов, а не одна отдельно взятая характеристика.

Регрессионные и ансамблевые модели машинного обучения на основе метрик сложности

Следующим логическим шагом стало объединение множества метрик сложности в рамках единой прогностической модели с использованием МО, в частности – ансамблевых алгоритмов. В отличие от одиночных моделей, ансамблевые методы объединяют множество простых классификаторов (например, деревьев решений), что позволяет получить более точный и устойчивый прогноз за счет усреднения их мнений или последовательного улучшения результатов. Это позволило учесть нелинейные взаимосвязи между параметрами и повысить точность предсказаний. Ключевым отличием от предыдущих подходов стало применение именно регрессионных моделей, которые решают задачу поиска количественной зависимости и прогнозирования непрерывного значения (в данном случае – GPR), в отличие от классификации, которая лишь относит объект к категории.

Одной из первых работ в этой области стало исследование Valdes et al [17], которые использовали модель пуассоновской регрессии с лассо-регуляризацией (Poisson Lasso, PL). Выбор пуассоновской регрессии был обусловлен природой данных: GPR, по сути, является отношением счетчиков (то есть числа точек, удовлетворяющих условию гамма-анализа, к общему их числу), и эта модель идеально подходит для прогнозирования этих величин, в отличие от линейной регрессии, которая может давать заведомо неверные прогнозы (например, за пределами диапазона 0–100 %). Применение лассо-регуляризации позволило избежать переобучения (Overfitting) модели, т.к. работа с 90 признаками на ограниченной выборке из 498 IMRT планов создавала высокий риск того, что модель «запомнит» шум в данных, а не выявит реальные закономерности, что привело бы к плохой работе на новых планах. Она решила и проблему мультиколлинеарности, поскольку многие метрики сложности (например, разные варианты расчета скорости МЛК) были тесно взаимосвязаны, что «зашумляло» модель и делало ее неустойчивой. Лассо-регуляризация целенаправленно «обнуляет» веса менее значимых и избыточных признаков, автоматически выполняя отбор наиболее информативных предикторов. В результате модель стала не только более надежной и способной к обобщению, но и более интерпретируемой. Модель показала среднюю абсолютную погрешность (Mean Absolute Error, MAE) менее 3 % для критерия 3 %/3 мм. На практике это означает, что GPR по прогнозу модели для большинства планов отличался от реального измеренного всего на несколько процентов, что вполне достаточно для клинического использования с целью стратификации рисков, а именно надежного выявления «проблемных» планов (с процентом точек < 90 %) и значительного сокращения объемов рутинных измерений.

Li et al [18] расширили этот подход для планов VMAT в области головы и шеи и в гинекологии. Они не только использовали PL-модель для регрессии, но и применили метод случайного леса (Random Forest, RF) для би-

нарной классификации планов на «прошедшие» и «не прошедшие» контроль качества. Важным выводом их работы стало то, что модель RF продемонстрировала значительно более высокую чувствительность (Sensitivity), или полноту (Recall) – способность обнаруживать «не прошедшие» планы по сравнению с PL-моделью (1):

$$\text{Sensitivity} = \frac{TP}{TP+FN} \times 100 \% \quad (1),$$

$$\text{Sensitivity} = \frac{TP}{TN+FP} \times 100 \% \quad (2),$$

где TP (True Positives) – планы, которые и по прогнозу модели, и по факту измерений прошли контроль качества (модель сработала верно), FP (False Positives) – планы, которые по прогнозу модели должны были пройти, но по факту измерений не прошли (это самая опасная ошибка модели, так как «плохой» план пропускается), TN (True Negatives) – истинно-отрицательные планы, которые и по прогнозу модели, и по факту измерений не прошли контроль качества (модель верно их спронозировала), FN – ложноотрицательные планы, которые по прогнозу модели не прошли, но по факту измерений были «хорошими» (это напрасная тревога).

В условиях технической валидации (TV) модель RF достигла чувствительности 100 % для критерия 3 %/2 мм с порогом 90 %, что означало отсутствие ложноотрицательных результатов. Модель корректно идентифицировала все планы, которые в реальности не прошли бы верификацию. По расчетам авторов, это позволило бы сократить объем необходимых измерений для 81,2 % планов. При этом в условиях независимой клинической валидации (CV) чувствительность RF для того же критерия составила всего 66,7 %, что подчеркивает важность внешнего тестирования моделей.

Работа Bin et al [19] является еще одним независимым подтверждением эффективности алгоритма Random Forest для данной задачи, в которой RF также показал себя как наиболее стабильный и точный алгоритм.

Исследование Thongsawad et al [20] было сфокусировано только на планах VMAT области головы и шеи. Авторы извлекли признаки, связанные со скоростью/ускорением МЛК и текстурным анализом флюенс-карт, и сравнили производительность моделей AdaBoost и Bagged Regression Trees. Оба метода являются ансамблевыми, но используют принципиально разные стратегии, что позволяет провести всестороннее сравнение.

Алгоритм AdaBoost (Adaptive Boosting) строит цепочку из простых моделей (например, неглубоких деревьев), где каждая следующая модель фокусируется на исправлении ошибок, допущенных предыдущей. Это делает AdaBoost чрезвычайно мощным в поиске сложных, нелинейных зависимостей и позволяет достичь очень высокой чувствительности. Способность этого алгоритма «вытягивать» слабые, но значимые закономерности из сложного набора признаков (текстуры флюенс-карт в сочетании с динамикой МЛК) делает его идеальным методом обнаружения даже неочевидных «провальных» планов.

Алгоритм Bagged Regression Trees (Bootstrap Aggregating, «Баггинг») строит множество независимых деревьев решений на разных случайных подвыборках исходных данных, а затем усредняет их предсказания. Это не столько увеличивает мощность, сколько стабилизирует модель и резко снижает дисперсию (variance),

предотвращая переобучение. Сравнивая эти два подхода, авторы смогли оценить, что важнее для их данных: максимальная прогностическая сила (AdaBoost) или максимальная стабильность (Bagging). Для обеих моделей были рассчитаны чувствительность и специфичность (Specificity), которая отвечает на вопрос: «Какой процент из всех реально «хороших» планов модель корректно идентифицировала как «хорошие»?».

Сочетание высокой чувствительности (94,12 %) и идеальной специфичности (100 %) у AdaBoost делает его не просто точным алгоритмом, а практичным инструментом для внедрения в клинику, который надежно отсеивает «хорошие» планы и с минимальным риском выявляет «плохие». Клинически это означает, что в данном исследовании не было ни одного случая ложной тревоги. Модель AdaBoost не отправила ни одного заведомо «хорошего» плана на дополнительную оптимизацию.

Глубокое обучение и анализ последовательностей контрольных точек

В то время как в предыдущих подходах использовали предварительно рассчитанные статические метрики, глубокое обучение открыло возможность работать непосредственно с «сырыми» данными плана. Рекуррентные нейронные сети (RNN), в частности сети с долгой краткосрочной памятью (LSTM), идеально подходят для обработки временных рядов, каковыми являются данные управления МЛК и гантри во время ротационных планов VMAT. Это связано с их уникальной архитектурой, включающей механизм «ворот», который позволяет LSTM избирательно сохранять информацию на долгих временных промежутках. Благодаря этому сеть может выявлять сложные причинно-следственные связи в данных, например, понимать, как комбинация параметров МЛК в начале ротационного поля влияет на дозовое распределение в его конце.

McCarthy et al [21] удачно применили LSTM для прогнозирования результатов верификации каждого ротационного поля в стереотаксических планах HyperArc (Varian Medical Systems) для лечения метастазов в головном мозге. Их ключевым нововведением был переход от бинарной классификации («прошел/не прошел») к трехклассовой: Ideal (наилучший, $GPR \geq 95\%$), Investigate (удовлетворительный и требующий детального изучения, $85\% \leq GPR < 95\%$) и Replan (план нуждается в перепланировании, где $GPR < 85\%$). Это более точно отражает клинический процесс, где планы с пограничным GPR часто требуют дополнительного анализа, а не автоматического перепланирования. Их LSTM-модель показала чувствительность 93,2 % и специфичность 96,6 % для самого важного класса Replan, что дает медицинским физикам высокую уверенность в необходимости оптимизации плана.

Современные подходы к прогнозированию GPR на основе глубокого обучения подходят не только для анализа временной динамики плана. Передовые сверточные нейронные сети (CNN) используются для комплексного осмысления пространственной сложности плана и анатомии пациента, автоматической идентификации и выделения наиболее информативных признаков, способствующих созданию надежных и стабильных моделей.

Ключевые методологические проблемы и пути их решения

Совместное использование таких методов с традиционными метриками сложности плана может повысить точность прогнозирования GPR, но такая комбинация ведет к уменьшению интерпретируемости

моделей, создавая «черный ящик», затрудняющий понимание внутренних механизмов работы модели. Одним из вариантов решения данной проблемы является разработка гибридных моделей – интеграция глубоких нейросетей с интерпретируемыми компонентами, например, с логистической регрессией или деревом решений.

Однако прозрачность – не единственное условие для клинического успеха. Даже если модель интерпретируема, она может работать нестабильно в реальных клинических условиях. Модели, обученные на данных одного института, одного типа линейного ускорителя, одной анатомической области и одной системы для пациент-ориентированного контроля качества, могут демонстрировать резкое падение производительности при применении в других условиях. Создание моделей, устойчивых к вариациям в оборудовании и протоколах, требует надежной валидации и возможно объединения данных из множества институтов.

В реальной клинической практике доля «не прошедших» планов в результате дозиметрической верификации обычно не превышает 5–10 %, что создает значительный дисбаланс классов и серьезные проблемы для обучения моделей. Для балансировки классов применяются различные подходы, направленные на искусственное сокращение (Undersampling) выборки мажоритарного класса («хороших» планов) или на искусственное увеличение (Oversampling) количества примеров минорного класса («плохих» планов), где одной из полезных техник является SMOTE (Synthetic Minority Over-sampling Technique), которая создает новые, синтетические примеры «плохих» планов на основе существующих, а не просто дублирует их [23, 24].

Для корректной оценки эффективности предиктивных моделей в условиях дисбаланса классов следует выбирать специализированные метрики, устойчивые к преобладанию мажоритарного класса. Классическая точность (Accuracy), рассчитываемая как отношение количества правильных прогнозов к общему числу наблюдений, становится неинформативной, поскольку может достигать высоких значений (>95 %) за счет корректного прогнозирования многочисленных «хороших» планов, одновременно с полным пропуском клинически значимых «плохих» планов. Рассмотренные научные публикации подчеркивают необходимость расширения набора используемых критериев за пределы стандартной чувствительности и специфичности. Авторы также используют точность (Precision) и F_1 -score (F-мера) – среднее гармоническое между точностью (Precision) и полнотой (Recall), обеспечивающее сбалансированную оценку компромисса между количеством ложных тревог и пропущенных дефектов. Дополнительно применяется критерий площади под кривой precision-recall (PR-AUC), хорошо описывающую эффективность модели обнаруживать объекты минорного класса независимо от распределения классов в выборке.

Заключение

Прогнозирование результатов пациент-ориентированного контроля качества с помощью методов машинного и глубокого обучения является быстро развивающейся и чрезвычайно перспективной областью медицинской физики. Существующие работы убедительно демонстрируют, что модели на основе метрик сложности, ансамблевых методов, сверточных и рекуррентных сетей способны с высокой точностью предсказывать GPR, идентифицируя планы с риском низкого качества верификации.

Современный тренд заключается в разработке многомодальных моделей, объединяющих различные источники данных, таких как комбинацию DICOM-RT параметров планирования с лог-файлами линейных ускорителей, использование данных электронной портальной визуализации в реальном времени, включение информации о техническом состоянии оборудования, а также отбор признаков для обучения и методы их извлечения. Эволюция методов извлечения признаков прошла путь от ручного конструирования к автоматическому анализу. Сегодня доминируют три подхода: ручное проектирование метрик сложности плана (параметры МЛК, МЕ), радиомический анализ изображений и автоматическое извлечение признаков с помощью сверточных нейронных сетей. Ключевыми проблемами остаются дисбаланс классов, интерпретируемость, обобщаемость моделей, а также их интеграция в клинический процесс.

Для интерпретируемости и общей применимости, чтобы быть действительно полезным инструментом, модели должны быть «бесшовно» интегрированы в

клиническую практику, что требует разработки скриптов (например, с использованием ESAPI для Eclipse) и удобных интерфейсов для пользователей. Для повышения доверия клиницистов предпочтительно создание интеллектуальных систем, которые не только прогнозируют GPR, но и предлагают конкретные практические рекомендации по автоматической корректировке параметров оптимизации на основе предсказаний, интеллектуальные рекомендации по изменению геометрии полей, предсказание оптимальных критериев верификации для конкретного плана.

Хотя успешное решение этих задач и откроет путь к созданию интеллектуальных систем пациент-ориентированного контроля качества, основанных на оценке рисков и способных предвидеть проблемы, на данный момент оно не заменяет общепринятую процедуру дозиметрической верификации планов в рамках программы гарантии качества, но существенно способно повысить эффективность, безопасность и стандартизацию лучевой терапии.

СПИСОК ИСТОЧНИКОВ / REFERENCES

- Otto K. Volumetric Modulated Arc Therapy: IMRT in a Single Gantry Arc. *Med Phys.* 2008 Jan;35;1:310-7. Doi: 10.1118/1.2818738.
- Miften M., Olch A., Mihailidis D., Moran J., Pawlicki T., Molineu A., Li H., Wijesooriya K., Shi J., Xia P., Papanikolaou N., A Low D. Tolerance Limits and Methodologies for IMRT Measurement-Based Verification QA: Recommendations of AAPM Task Group No. 218. *Med Phys.* 2018 Apr; 45;4:e53-e83. Doi: 10.1002/mp.12810.
- Olch A.J. Dosimetric Performance of an Enhanced Dose Range Radiographic Film for Intensity-Modulated Radiation Therapy Quality Assurance. *Med Phys.* 2002 Sep;29;9:2159-68. Doi: 10.1118/1.1500398.
- Wouter van Elmpt, McDermott L., Nijsten S., Wendling M., Lambin P., Mijnheer B. A Literature Review of Electronic Portal Imaging for Radiotherapy Dosimetry. *Radiother Oncol.* 2008 Sep;88;3:289-309. Doi: 10.1016/j.radonc.2008.07.008.
- Low D.A., Harms W.B., Mutic S., Purdy J.A. A Technique for the Quantitative Evaluation of Dose Distributions. *Med Phys.* 1998 May;25;5:656-61. Doi: 10.1118/1.598248.
- Depuydt T., Van Esch A., Pierre Huyskens D. A Quantitative Evaluation of IMRT Dose Distributions: Refinement and Clinical Assessment of the Gamma Evaluation. *Radiother Oncol.* 2002 Mar;62;3:309-19. Doi: 10.1016/s0167-8140(01)00497-2.
- Ford E.C., Terezakis S., Souranis A., Harris K., Gay H., Mutic S. Quality Control Quantification (QCQ): a Tool to Measure the Value of Quality Control Checks in Radiation Oncology. *Int J Radiat Oncol Biol Phys.* 2012 Nov 1;84;3:e263-9. Doi: 10.1016/j.ijrobp.2012.04.036.
- Chan M.F., Witzum A., Valdes G. Integration of AI and Machine Learning in Radiotherapy QA. *Front Artif Intell.* 2020 Sep 29;3:577620. Doi: 10.3389/frai.2020.577620.
- Valdes G., Chan M.F., Boh Lim Seng, Scheuermann R., O Deasy J., D Solberg T. IMRT QA Using Machine Learning: a Multi-Institutional Validation. *J Appl Clin Med Phys.* 2017 Sep;18;5:279-284. Doi: 10.1002/acm2.12161.
- Международное агентство по атомной энергии. Искусственный интеллект в медицинской физике. Функции, обязанности, образование и подготовка медицинских физиков клинической квалификации: Серия учебных курсов №83. Вена: МАГАТЭ, 2025. 54 с. [*Mezhdunarodnoye Agentstvo po Atomnoy Energii. Iskusstvennyy Intellekt v Meditsinskoj Fizike. Funktsii, Obyazannosti, Obrazovaniye i Podgotovka Meditsinskikh Fizikov Klinicheskoy Kvalifikatsii* = Artificial Intelligence in Medical Physics. Roles, Responsibilities, Education, and Training of Clinically Qualified Medical Physicists. Training Course Series No. 83. Vienna, IAEA Publ., 2025. 54 p. (In Russ.)].
- Younge K.C., Roberts D., Janes L.A., Anderson C., Moran J.M., Matuszak M.M. Predicting Deliverability of Volumetric-Modulated arc Therapy (VMAT) Plans Using Aperture Complexity Analysis. *J Appl Clin Med Phys.* 2016 Jul 8;17;4:124-31. Doi: 10.1120/jacmp.v17i4.6241.
- McNiven A.L., Sharpe M.B., Purdie T.G. A New Metric for Assessing IMRT Modulation Complexity and Plan Deliverability. *Med Phys.* 2010 Feb;37;2:505-15. Doi: 10.1118/1.3276775.
- Park J.M., Park S.-Y., Kim H. Modulation Index for VMAT Considering both Mechanical and Dose Calculation Uncertainties. *Physics in Medicine & Biology.* 2015;60;18:7101-7125. Doi: 10.1088/0031-9155/60/18/7101.
- Park J.M., Wu H.G., Kim J.H., Carlson J.N., Kim K. The Effect of MLC Speed and Acceleration on the Plan Delivery Accuracy of VMAT. *Br J Radiol.* 2015 May; 88;1049:20140698. Doi: 10.1259/bjr.20140698.
- Nyflot M.J., Thammasorn P., Wootton L.S., Ford E.C., Chalovithongse W.A. Deep Learning for Patient-Specific Quality Assurance: Identifying Errors in Radiotherapy Delivery by Radiomic Analysis of Gamma Images with Convolutional Neural Networks. *Med Phys.* 2019 Feb;46;2:456-464. Doi: 10.1002/mp.13338.
- Hideaki Hirashima, Tomohiro Ono, Mitsuhiro Nakamura, Yuki Miyabe, Nobutaka Mukumoto, Hiraku Iramina, Takashi Mizowaki. Improvement of Prediction and Classification Performance for Gamma Passing Rate by Using Plan Complexity and Dosimetric Features. *Radiother Oncol.* 2020 Dec;153:250-257. Doi: 10.1016/j.radonc.2020.07.031.
- Valdes G., Scheuermann R., Hung C.Y., Olszanski A., Bellerive M., Solberg T.D. A Mathematical Framework for Virtual IMRT QA Using Machine Learning. *Med Phys.* 2016 Jul;43;7:4323. Doi: 10.1118/1.4953835.
- Jiaqi Li, Le Wang, Xile Zhang, Lu Liu, Jun Li, Maria F Chan, Jing Sui, Ruijie Yang. Machine Learning for Patient-Specific Quality Assurance of VMAT: Prediction and Classification Accuracy. *Int J Radiat Oncol Biol Phys.* 2019 Nov 15;105;4:893-902. Doi: 10.1016/j.ijrobp.2019.07.049.
- Bin S., Zhang J., Shen L., Zhang Jand Wang Q. Study of the Prediction of Gamma Passing Rate in Dosimetric Verification of Intensity-Modulated Radiotherapy Using Machine Learning Models Based on Plan Complexity. *Front Oncol.* 2023 Jul;21;13:1094927. Doi: 10.3389/fonc.2023.1094927.
- Sangutid Thongsawad, Somyot Srisatit, Todsaporn Fuangrod. Predicting Gamma Evaluation Results of Patient-Specific Head and Neck Volumetric-Modulated Arc Therapy Quality Assurance Based on Multileaf Collimator Patterns and Fluence Map Features: a Feasibility Study. *J Appl Clin Med Phys.* 2022 Jul;23;7:e13622. Doi: 10.1002/acm2.13622.

21. Shane McCarthy, Brent Harrison, Damodar Pokhrel. A Predictive Quality Assurance Model for Patient-Specific Gamma Passing Rate of Hyperarc-Based Stereotactic Radiotherapy and Radiosurgery of Brain Metastases. *J Appl Clin Med Phys*. 2025 Sep; 26;9:e70225. Doi: 10.1002/acm2.70225.
22. Tomohiro Kajikawa, Noriyuki Kadoya, Kengo Ito, Yoshiki Takayama, Takahito Chiba, Seiji Tomori, Hikaru Nemoto, Suguru Dobashi, Ken Takeda, Keiichi Jingu. A Convolutional Neural Network Approach for IMRT Dose Distribution Prediction in Prostate Cancer Patients. *J Radiat Res*. 2019 Oct 23;60;5:685-693. Doi: 10.1093/jrr/rrz051.
23. Haibo He, Yang Bai, Edwardo A. Garcia, Shutao Li. ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning In. *IEEE International Joint Conference on Neural Networks*. 2008:1322-1328. Doi: 10.1109/IJCNN.2008.4633969.
24. Chawla N.V., et al. SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*. 2002;16;1:321-357. Doi: 10.1613/jair.953.
25. Alexander F I Osman, Nabil M Maalej. Applications of Machine and Deep Learning to Patient-Specific IMRT/VMAT Quality Assurance. *Appl Clin Med Phys*. 2021 Aug 3;22;9:20-36. Doi: 10.1002/acm2.13375.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.
Финансирование. Исследование не имело спонсорской поддержки.
Участие авторов. Статья подготовлена с равным участием авторов.
Поступила: 20.01.2026. Принята к публикации: 25.02.2026.

Conflict of interest. The authors declare no conflict of interest.
Financing. The study had no sponsorship.
Contribution. Article was prepared with equal participation of the authors.
Article received: 20.01.2026. Accepted for publication: 25.02.2026.