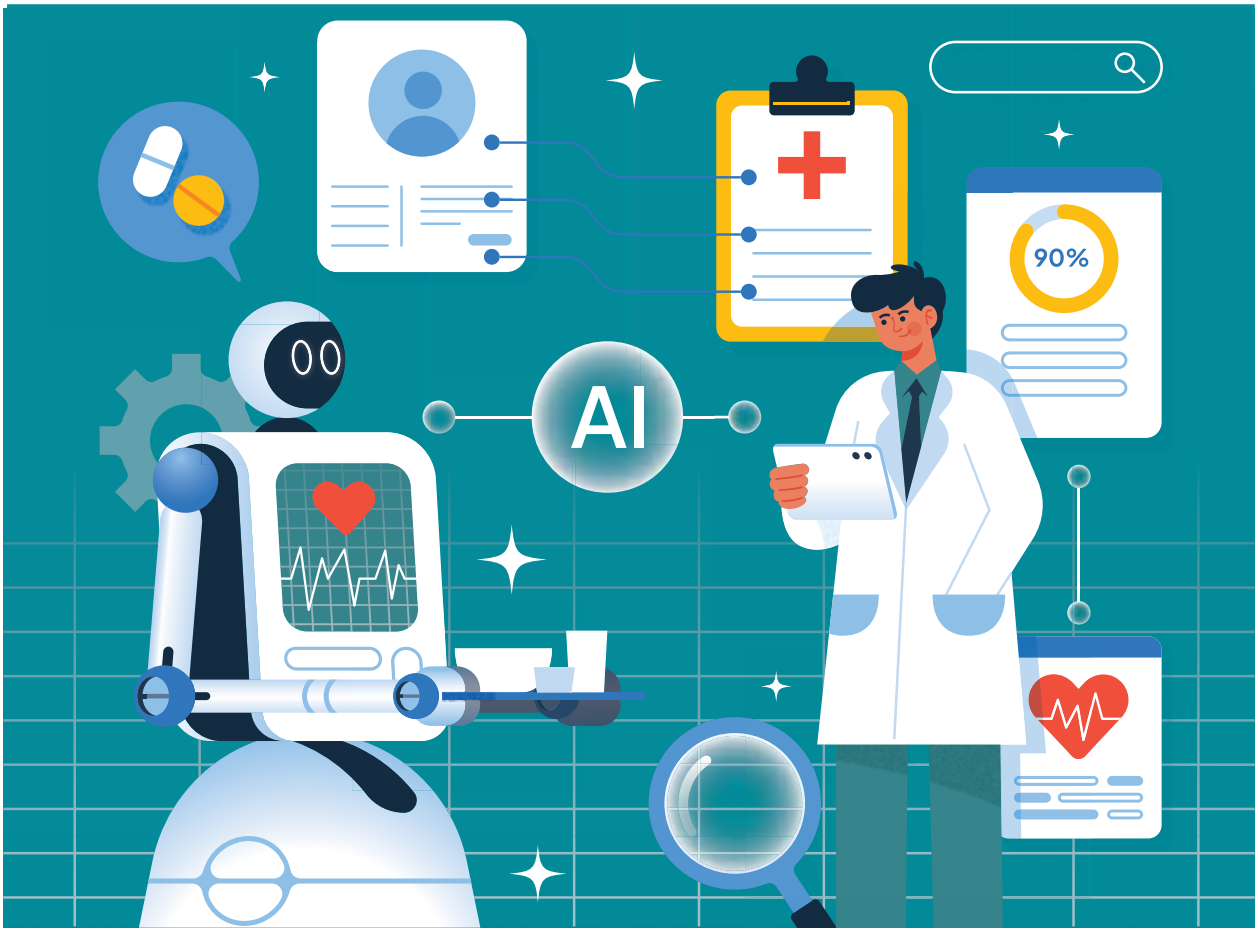
ИННОВАЦИИ В ОНКОЛОГИИ



Искусственный интеллект повышает эффективность скрининга и диагностики злокачественных опухолей

Основная масса исследований эффективности искусственного интеллекта (ИИ) посвящена скрининговым программам с применением методов визуализации, а именно: маммографическому скринингу, скринингу рака легкого (РЛ) с применением низкодозовой компьютерной томографии (НДКТ) и скринингу рака предстательной железы (РПЖ) с применением мультипараметрической магнитно-резонансной томографии (МРТ). Анализ их проводит Д.Г. Заридзе.

Давид Георгиевич ЗАРИДЗЕ,
член-корреспондент РАН, руководитель отдела клинической эпидемиологии НИИ клинической онкологии ФГБУ «НМИЦ онкологии им. Н.Н. Блохина» Минздрава России, президент Противоракового общества России

Эффективность скрининга рака молочной железы (РМЖ), РПЖ, РЛ доказана в рандомизированных клинических исследованиях: скрининг снижает смертность от этих форм рака [1]. Однако результат клинических исследований редко достигается на практике. Скрининг злокачественных новообразований (ЗНО) имеет ряд недостатков.

Значительной проблемой является нехватка квалифицированных рентгенологов, выраженная вариабельность в оценке изображений, чрезмерное количество ложноположительных результатов, что, в свою очередь, увеличивает нагрузку на рентгенологов в связи с необходимостью пересмотра большого количества снимков и ростом повторных приглашений. Например, в США

STATUS QUO

Астрономический рост заболеваемости некоторыми ЗНО, скорее всего за счет пациентов с клинически незначимыми формами рака, сопровождается растущей нагрузкой на здравоохранение страны. Многие тысячи пациентов проходят скрининг ЗНО, при котором выявляются клинически незначимые образования. Последствия – ненужные дополнительное обследование и дорогостоящее лечение, часто чрезмерное. В то же время при профилактических осмотрах не попадают в поле зрения агрессивные формы, так называемые интервальные раки, что оказывает отрицательное влияние на прогноз, показатели выживаемости и смертности.

большинство рентгенологов превышают рекомендованное количество повторных приглашений. У трети пациентов диагноз рака биопсией не подтверждается.
Страдает и охват скринингом населения. Например, в США скрининг РЛ проходят меньше 10% целевой популяции [2].
Серьезнейшей проблемой скрининга остается гипердиагностика, то есть выявление клинически незначимых образований, которое сопровождается ненужным дополнительным обследованием, а иногда и ненужным лечением. Так, в США внедрение программы ПСА-скрининга в 1986 г. без наличия доказательств его эффективности (не было доказано, что раннее выявление снижает смертность) привело к резкому росту заболеваемости в стране. Осознание ошибки и практически отмена ПСА-скрининга в стране сопровождались снижением заболеваемости РПЖ и дальнейшей ее стабилизацией [3].
В то же время скрининг пропускает до 30% агрессивных форм рака (интервальные раки). Компьютерные программы в помощь рентгенологам, такие как программы для интерпретации маммограмм, мультипараметрической МРТ и низкодозовой компьютерной томографии (НДКТ), CAD (computer-aided detection), PI-RADS (Prostate Imaging and Reporting and Data System), lung-RADS (registration, analysis and data system) оказались недостаточно эффективными.

Скрининг рака молочной железы

Остановлюсь на исследованиях, целью которых являлась разработка алгоритма ИИ для повышения эффективности маммографической диагностики, снижения количества ложноположительных результатов и связанных с этим повторных приглашений. Обе работы основаны на хорошо охарактеризованных и прослеженных когортах пациенток – участниц диагностических и скрининговых программ. В южнокорейском исследовании [4] алгоритм ИИ разработан на базе глубокой нейронной сети обработки изображения с использованием 170 230 маммографических исследований, проведенных в разных странах. Эффективность ИИ оценивалась на основании показателя кривой AUROC (area under ROC – площадь под кривой ROC). Диагностическая эффективность при прочтении маммограмм только ИИ (0,940, 95% допустимый интервал (ДИ) 0,91–0,96) была достоверно выше по сравнению с прочтением только рентгенологом (0,810, 95% ДИ 0,77–0,85) и рентгенологом + ИИ (0,881, 95% ДИ 0,850–0,911) ($p < 0{,}0001$). Чувствительность и специфичность оценки маммограмм также достоверно выше при их прочтении только ИИ по сравнению с прочтением рентгенологами. Преимущество ИИ сохраняется во всех подгруппах, независимо от возраста, характеристик ткани молочной железы, наличия кальцификаций, гистологического типа и т.д. Например, при плотной молочной железе чувствительность ИИ на 16% выше, чем рентгенологов ($p < 0{,}0001$), а разница в специфичности составляет 14,7% ($p < 0{,}0001$).
В исследовании, проведенном в Швеции [5] на базе когорты 55 581 женщины, прошедшей маммографический скрининг, использовалась коммерческая система ИИ [4]. Исследование показало, что ИИ не только не уступает результатам стандартного прочтения маммограмм двумя радиологами, но имеет ряд преимуществ. Количество неправильных интерпретаций маммограмм было на 49% ниже (2916 vs. 3854) при прочтении маммограмм только ИИ, чем при применении стандартного метода; повторных приглашений – на 47% (768 vs. 1629) ниже при прочтении маммограмм только ИИ по сравнению со стандартным методом. Соответственно, меньше количество женщин, проходящих дополнительное обследование, включая томосинтез и ультразвуковое исследование (УЗИ), не говоря уже о стрессе, который испытывают женщины в промежутке между первым и дополнительным обследованиями.

Скрининг рака легкого

D. Ardila и соавт. [6] разработали алгоритм глубокого обучения на основе базы данных лиц, прошедших НДКТ-скрининг в рамках Национальной программы скрининга РЛ в США (NLST). При прочтении 6716 изображений, полученных из базы данных NLST, эффективность алгоритма ИИ (94,4%) была выше эффективности рентгенологов, которые оценивали изображения на основании классификационной системы Американского общества радиологов. При валидации на 1130 случаях, полученных из другого источника, эффективность также была высокой – 94,2%. Прочтение ИИ изображений НДКТ привело к снижению количества ложноположительных и ложноотрицательных результатов на 11 и 5% соответственно.
Модель ИИ глубокого обучения (Sybil) разработана с использованием НДКТ от 15 000 участников NLST, США [2]. Модель не нуждается в аннотации демографических, клинических и радиологических данных. Разработка модели проводилась в трех последовательных группах: обучение, разработка и тестирование. Тестирование модели проведено на трех независимых базах данных. Модель предсказывает риск развития рака легкого через один год (AUC = 0,92, 0,86 и 0,95), два (AUC = 0,86, 0,82, 0,87), три (AUC = 0,80, 0,79, 0,87), четыре (AUC = 0,77, 079, 0,71), пять (AUC = 0,75, 0,78, 0,77) и шесть (0,75, 0,74) лет.

Скрининг рака предстательной железы

В исследовании T. Chen и соавт. [7] мультипараметрическая МРТ в комбинации с ИИ машинного обучения показали высокую чувствительность и специфичность в дифференциальной диагностике и оценке агрессивности РПЖ по сравнению с PI-RADS. При дифференциальной диагностике РПЖ (РПЖ+/РПЖ-) эффективность модели ИИ (AUC = 0,985, 0,982, 0,999), в зависимости от режима изображения МРТ (T2WI и ADC), достоверно выше эффективности PI-RADS v2 (AUC = 0,867). Эффективность модели ИИ также была выше при дифференциальной диагностике между GS6 и GS7 (AUC модели ИИ = 0,888, 0,865, 0,930 и AUC PI-RADS v2 = 0,763). Таким образом, модель ИИ показала высокую диагностическую эффективность и превосходит PI-RADS v2 в диагностике РПЖ и дифференциальной диагностике между Глисон 6 и Глисон 7.

Представляет интерес еще одна работа, выполненная в Китае [8]. В исследование было включено 4747 пациентов с диагнозом РПЖ. Система диагностики ИИ рака простаты (Prostate Cancer Artificial Intelligence Diagnostic System (PCAIDS)) разработана на основании демографических, клинических, лабораторных данных и УЗИ этих пациентов. Эффективность PCAIDS выше (AUC = 0,82–0,85%), чем эффективность ПСА тестов (0,58–0,73%). При тестировании PCAIDS в трех когортах количество ненужных биопсий снизилось на 32,2, 17,6, 26,3%; в то же время было пропущено 4,9, 4,5, 4,1% случаев клинически значимого РПЖ.

Систематический обзор и метаанализ 12 работ по оценке эффективности ИИ машинного обучения в МРТ-диагностике клинически значимого РПЖ показал высокую суммарную эффективность (AUC = 0,86, 95% ДИ 0,81–0,91). Показатель AUC варьировал от 0,73 до 0,99% [9].

Заключение

1. Эффективность разрабатываемой модели ИИ зависит от правильной постановки конкретных, научно обоснованных задач. Обучение и тестирование алгоритма ИИ следует проводить с использованием детально охарактеризованных по демографическим, клиническим и другим ключевым параметрам когорт онкологических больных и контрольных лиц с адекватным сроком прослеживания.

2. Применение ИИ для скрининга ЗНО имеет не только клиническое, но и экономическое значение: снизится нагрузка на онкологическую службу и здравоохранение в целом. В результате использования ИИ на практике будет решена проблема дифференциальной диагностики и прогнозирования клинически незначимых, клинически значимых и агрессивных форм рака и обоснованного выбора консервативной или агрессивной лечебной тактики для больных с ранними формами рака. НОП [2030]

Литература

1. Заридзе Д.Г., Максимович Д.М., Стилиди И.С. Новая парадигма скрининга и ранней диагностики: оценка пользы и вреда. Вопросы онкологии. 2020; 66 (6): 589–602.
2. Mikhael P.G., Wohlwend J., Yala A., et al. Sybil: a validated deep learning model to predict future lung cancer risk from a single low-dose chest computed tomography. J. Clin. Oncol. 2023; 41 (12): 2191–2200.
3. Pinsky P.F., Parnes H. Screening for prostate cancer. N. Engl. J. Med. 2023; 388 (15): 1405–1414.
4. Kim H.E., Kim H.H., Han B.K., et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. Lancet Digit. Health. 2020; 2 (3): e138–e148.
5. Dembrower K., Crippa A., Colón E., et al. Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. Lancet Digit. Health. 2023; 5 (10): e703–e711.
6. Ardila D., Kiraly A.P., Bharadwaj S., et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat. Med. 2019; 25 (6): 954–961.
7. Chen T., Li M., Gu Y., et al. Prostate cancer differentiation and aggressiveness: assessment with a radiomic-based model vs. PI-RADS v2. J. Magn. Reson. Imaging. 2019; 49 (3): 875–884.
8. Zhang H., Ji J., Liu Z., et al. Artificial intelligence for the diagnosis of clinically significant prostate cancer based on multimodal data: a multicenter study. BMC Med. 2023; 21 (1): 270.
9. Cuocolo R., Cipullo M.B., Stanzione A., et al. Machine learning for the identification of clinically significant prostate cancer on MRI: a meta-analysis. Eur. Radiol. 2020; 30 (12): 6877–6887.