

ARTIFICIAL INTELLIGENCE VS CLINICAL EXPERTISE: A COMPARATIVE ANALYSIS OF DIAGNOSTIC EFFECTIVENESS IN MODERN DENTISTRY**Tulaanbaev Fayzullo**

Assistant of the Department of Therapeutic Dentistry, Kimyo International University

Ravshanbekova Xonzoda,**Nosirova Maftuna**<https://doi.org/10.5281/zenodo.19987257>

Abstract: This article presents a critical analysis of the current state of research comparing the diagnostic effectiveness of artificial intelligence (AI) systems and dental practitioners. Based on data from systematic reviews, meta-analyses, and original clinical studies published between 2024 and 2026, the achievements and limitations of AI are examined in areas such as cariology, oral mucosal pathology, periodontology, and orthodontics. Particular attention is given to the issue of external validity in existing studies, differences in AI performance depending on the clinician's experience, and the depth of pathological processes. The study concludes that AI should be considered as a clinical decision support tool rather than a replacement for dental professionals.

Keywords: artificial intelligence, dentistry, diagnostic accuracy, comparative effectiveness, deep learning, convolutional neural networks, clinical validation.

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ПРОТИВ КЛИНИЧЕСКАЯ ЭКСПЕРТИЗА: СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЭФФЕКТИВНОСТИ ДИАГНОСТИКИ В СОВРЕМЕННОЙ СТОМАТОЛОГИИ.**Туланбаев Файзулло**

Ассистент кафедры терапевтической стоматологии Международного университета

Кимё

Равшанбекова Хонзода**Носирова Мафтуна**

Аннотация: В статье представлен критический анализ современного состояния исследований, посвященных сравнению диагностической эффективности систем искусственного интеллекта (ИИ) и врачей-стоматологов. На основе данных систематических обзоров, мета-анализов и оригинальных клинических исследований, опубликованных в период 2024-2026 гг., рассматриваются достижения и ограничения ИИ в таких областях, как кариесология, патология слизистой оболочки полости рта, пародонтология и ортодонтия. Особое внимание уделяется проблеме внешней валидности существующих исследований, различиям в эффективности ИИ в зависимости от клинического опыта врача и глубины патологического процесса. Делается вывод о целесообразности рассмотрения ИИ в качестве инструмента клинической поддержки принятия решений, а не замены врача-стоматолога.

Ключевые слова: искусственный интеллект, стоматология, диагностическая точность, сравнительная эффективность, глубокое обучение, сверточные нейронные сети, клиническая валидация.

Для цитирования: [ФИО автора]. Искусственный интеллект vs. клиническая экспертиза: сравнительный анализ эффективности диагностики в современной стоматологии. 2026.

1. Введение

Цифровая трансформация здравоохранения в последнее десятилетие приобрела характер парадигмального сдвига, и стоматология не является исключением. Технологии искусственного интеллекта (ИИ), основанные преимущественно на архитектурах глубокого обучения и сверточных нейронных сетях (CNN), все активнее внедряются в клиническую практику, предлагая новые решения для диагностики, планирования лечения и прогнозирования исходов.

Стремительный рост числа публикаций в этой области отражает высокий научный и практический интерес к проблеме. Согласно данным крупномасштабного обзора систематических обзоров, включавшего 116 работ, пол-чувствительность диагностических алгоритмов ИИ в стоматологии составляет 0,85 (95% ДИ: 0,76–0,93), а пол-специфичность — 0,93 (95% ДИ: 0,90–0,95). Современные модели на основе CNN демонстрируют пол-точность до 93,1%.

Однако за этими впечатляющими цифрами скрывается сложная и неоднозначная реальность. Большинство исследований проводятся в контролируемых лабораторных условиях на ретроспективных датасетах, что ставит под вопрос генерализуемость полученных результатов на реальную клиническую практику. Более того, систематический обзор, опубликованный в 2025 году, показал, что 79% исследований в этой области являются ретроспективными и проведены в контролируемых исследовательских условиях, при этом риск систематической ошибки наиболее высок в домене отбора пациентов ввиду частого использования дизайна «случай-контроль» и архивных изображений.

Цель настоящей статьи — представить сбалансированный, основанный на данных доказательной медицины анализ сравнительной эффективности ИИ и врачей-стоматологов в диагностике основных стоматологических заболеваний, выделив как области, где ИИ демонстрирует несомненные преимущества, так и ситуации, требующие критической интерпретации результатов.

2. Материалы и методы обзора

Настоящая работа представляет собой нарративный обзор, основанный на анализе литературы за период 2024–2026 гг. Поиск проводился в базах данных PubMed, Scopus, Web of Science и IEEE Xplore с использованием комбинации ключевых слов: «artificial intelligence», «deep learning», «dental diagnosis», «comparative study», «clinical validation». В обзор включались систематические обзоры, мета-анализы, рандомизированные клинические исследования и проспективные когортные исследования, оценивавшие диагностическую точность ИИ в сравнении с клинической оценкой стоматологов.

3. Результаты сравнительного анализа

3.1. Общая характеристика диагностической эффективности

Мета-анализ 12 систематических обзоров с низким риском систематической ошибки показал, что диапазон диагностической точности ИИ варьирует от 82% до 95% в зависимости от стоматологической специальности. При этом наблюдалась значительная гетерогенность результатов: I^2 для пол-чувствительности составила 98,26%, для специфичности — 87,49%, для AUC — 86,62%. Высокая гетерогенность указывает на существенные различия в методологии исследований, характеристиках датасетов и пороговых значениях алгоритмов, что затрудняет прямое сравнение результатов разных работ.

3.2. Диагностика кариеса

Кариес остается наиболее распространенным стоматологическим заболеванием, и значительная часть исследований сравнительной эффективности ИИ сосредоточена именно в этой области.

В проспективном клиническом исследовании, проведенном на базе стоматологической клиники Европейского университета Кипра, оценивалась диагностическая точность коммерческой ИИ-системы Diagnocat™ при детекции апроксимального кариеса на прикусных рентгенограммах. Исследование включало 100 цифровых рентгенограмм (1540 апроксимальных поверхностей), референсный стандарт был установлен двумя независимыми экспертами (специалистами в области оперативной стоматологии и рентгенологии) с последующим достижением консенсуса. Согласие между экспертами было почти идеальным (κ Коэна = 0,8096).

Система Diagnocat™ продемонстрировала высокую специфичность — 94,3% (95% ДИ: 92,4–96,0%) и отрицательную прогностическую ценность (NPV) — 96,1% (95% ДИ: 94,7–97,3%), что указывает на высокую надежность алгоритма при исключении кариеса. Общая точность составила 91,6% (95% ДИ: 89,7–93,4%).

Однако чувствительность оказалась умеренной — 73,1% (95% ДИ: 65,9–79,9%), а положительная прогностическая ценность (PPV) — низкой: 64,7% (95% ДИ: 57,7–71,5%). Согласие с референсным стандартом было существенным (κ = 0,638), но нулевая гипотеза об отсутствии различий между ИИ и экспертами была отвергнута ($p < 0,001$). Примечательно, что при анализе в зависимости от глубины поражения PPV для эмалевых и дентинных поражений составляла всего 42,0% и 39,2% соответственно, тогда как NPV превышала 98,0%.

Эти результаты имеют важное клиническое значение: высокая NPV позволяет использовать ИИ для скрининга и надежного исключения кариеса, однако низкая PPV и умеренная чувствительность указывают на то, что система может пропускать значительную долю истинных поражений и давать ложноположительные результаты, требующие верификации врачом.

3.3. Сравнение с врачами разного уровня квалификации

Принципиально важным представляется вопрос о том, как эффективность ИИ соотносится с уровнем клинического опыта врача. Исследование, проведенное Sowmya и соавт. (2024), напрямую сравнивало диагностическую эффективность ИИ и 10 врачей-стоматологов начального уровня (с опытом работы 1–3 года) при детекции кариеса и периапикальных инфекций на панорамных снимках.

ИИ-система продемонстрировала чувствительность 92%, специфичность 89% и точность 90% при детекции кариеса, тогда как начинающие врачи показали средние значения 80%, 85% и 82% соответственно. Для периапикальных инфекций различия были еще более выраженными: чувствительность ИИ — 95%, точность — 92% против 85% и 86% у начинающих врачей ($p < 0,05$).

Однако исследование имеет ряд ограничений. Во-первых, в качестве референсного стандарта использовался консенсус трех опытных рентгенологов, а не гистологическое подтверждение. Во-вторых, ИИ-система обучалась на большом датасете аннотированных изображений, тогда как начинающие врачи, вероятно, имели ограниченный опыт интерпретации панорамных снимков. В-третьих, исследование не включало группу опытных врачей, что не позволяет утверждать, что ИИ превосходит или хотя бы не уступает специалистам высокого уровня.

3.4. Диагностика поражений слизистой оболочки полости рта

Раннее выявление предраковых и злокачественных поражений слизистой оболочки полости рта является критически важной задачей, где традиционный клинический осмотр имеет ограниченную чувствительность, особенно вне специализированных центров.

Проспективное клиническое исследование, проведенное Shah и соавт. (2025), включало 180 пациентов с подозрительными поражениями слизистой оболочки полости рта. ИИ-система OralScanX v2.5 (архитектура CNN, обученная на датасете из 10 000 аннотированных изображений) показала общую диагностическую точность 91,1% ($\pm 4,3$), чувствительность 93,6% и специфичность 89,2%. Согласие между ИИ и гистопатологическим диагнозом было существенным ($\kappa = 0,83$; $p < 0,001$). Наивысшая точность была достигнута для лейкоплакии (95,2%) и красного плоского лишая (92,4%). Напротив, клинический диагноз, поставленный двумя опытными патологами полости рта, показал лишь умеренное согласие с гистологией ($\kappa = 0,72$).

Эти данные демонстрируют, что в специфической области оральной патологии ИИ может не только не уступать, но и превосходить клиническую диагностику, выполняемую даже опытными специалистами. Однако авторы отмечают, что ошибочная классификация происходила преимущественно в случаях слабой дисплазии и неспецифических язв — наиболее сложных для дифференциальной диагностики клинических ситуациях.

3.5. Пародонтология

В области пародонтологии перспективным направлением является использование глубокого обучения для оценки ранних признаков заболевания. Chen и соавт. (2025) разработали систему на основе Mask R-CNN с бэкбоном ResNet-101 для оценки диспропорции маргинальных гребней (Marginal Ridge Discrepancy — MRD) — важного раннего индикатора пародонтита.

Система достигла впечатляющих показателей сегментации: пиксельная точность составила 98,11%, а recall для зубов, коронок и костных структур достиг 97,60%, 97,24% и 97,53% соответственно. Модель классификации MRD показала точность 93,41% со средней угловой ошибкой всего $0,85^\circ$. Эти результаты свидетельствуют о том, что в структурированных задачах анализа рентгенографических изображений с четкими анатомическими ориентирами ИИ может достигать эффективности, сопоставимой или даже превосходящей человеческую.

3.6. Ортодонтия

Ортодонтия представляет собой область, где ИИ демонстрирует особенно высокие показатели благодаря доступности больших размеченных датасетов цефалометрических изображений. Согласно систематическому обзору, ортодонтические приложения достигают 95–98% точности при идентификации цефалометрических ориентиров. Исследования Midlej и соавт. (2024) показали, что модели машинного обучения могут классифицировать скелетные классы II и III малокклюзии с точностью до 0,99. Однако следует отметить, что эти результаты получены преимущественно в исследовательских условиях, и их клиническая валидация остается ограниченной.

3.7. Проблема объяснимости и прозрачности

Критически важным аспектом внедрения ИИ в клиническую практику является проблема «черного ящика» — неспособности большинства моделей глубокого обучения объяснить логику своего диагностического решения. Rana и Akhtar (2026) предложили гибридную модель, комбинирующую CNN со статистической детекцией аномалий и инструментами объяснимого ИИ (XAI), включая Grad-CAM, SHAP и карты значимости. Авторы продемонстрировали, что такой подход не только повышает диагностическую точность (43,26% против 35,12% для базовой CNN), но и обеспечивает визуализацию критических для принятия решения областей, повышая доверие со стороны клиницистов.

4. Обсуждение

4.1. Где ИИ эффективнее врача?

Анализ доступных данных позволяет выделить области, в которых ИИ демонстрирует статистически значимые преимущества:

1. Обнаружение окклюзионных и апроксимальных поражений на ранних стадиях — ИИ может выявлять изменения плотности тканей, неразличимые при визуальном осмотре .

2. Количественная оценка и стандартизация — в отличие от человека, ИИ обеспечивает воспроизводимые результаты, не подверженные усталости, эмоциональному состоянию или межсубъектной вариабельности .

3. Анализ больших объемов данных — системы ИИ способны обрабатывать сотни рентгенограмм или внутриротовых фотографий за время, недоступное человеку .

4. Скрининг в условиях ограниченного доступа к специалистам — в удаленных или ресурсно-ограниченных регионах ИИ может служить инструментом первичного скрининга и триажа пациентов .

4.2. Где врач сохраняет преимущества?

Несмотря на впечатляющие результаты ИИ, существует ряд областей и аспектов, где клиническая экспертиза врача остается незаменимой:

1. Дифференциальная диагностика редких и атипичных поражений — ИИ-системы обучаются на распространенных паттернах и имеют низкую «энтропию обобщения» для редких заболеваний .

2. Интеграция многомодальных данных — врач учитывает не только данные инструментальной диагностики, но и анамнез, жалобы пациента, общее соматическое состояние, факторы риска — информацию, которую современные ИИ-системы не способны интегрировать эффективно.

3. Интерпретация низкокачественных изображений — большинство ИИ-систем демонстрируют резкое снижение эффективности при работе с изображениями, имеющими артефакты, неправильную проекцию или низкое разрешение .

4. Клиническое принятие решений в условиях неопределенности — при низкой PPV ИИ (42,0% для эмалевых поражений) врач должен критически оценивать и верифицировать каждый положительный результат .

4.3. Методологические ограничения существующих исследований

При интерпретации результатов сравнительных исследований необходимо учитывать ряд системных методологических проблем. Систематический обзор Araidy и соавт. (2025) показал, что большинство исследований оценивались по шкале QUADAS-2 как имеющие высокий риск систематической ошибки, особенно в домене отбора пациентов . Наиболее частые проблемы включают: использование ретроспективных дизайнов, отсутствие внешней валидации на независимых датасетах, дисбаланс классов в обучающих выборках, непрозрачность процесса аннотации и отсутствие проспективных клинических испытаний.

Кроме того, значительная часть исследований сравнивает ИИ с «усредненным» врачом или с врачами начального уровня, что создает предпосылки для завышения относительной эффективности ИИ. Исследований, сравнивающих ИИ с экспертами высокого уровня в условиях реальной клинической практики, крайне мало.

4.4. Клинические рекомендации

На основании анализа данных можно сформулировать следующие предварительные рекомендации:

1. ИИ-системы следует рассматривать как инструмент «второго мнения» и клинической поддержки принятия решений, а не как замену врачу.

2. Высокая специфичность и NPV позволяют использовать ИИ для скрининга (исключения заболевания), однако положительные результаты требуют обязательной верификации врачом.

3. Наиболее клинически полезным представляется применение ИИ начинающими врачами, которые получают наибольший относительный прирост диагностической точности.

4. При внедрении ИИ в клиническую практику необходимо обеспечить обучение персонала интерпретации результатов и пониманию ограничений системы.

5. Заключение

Искусственный интеллект демонстрирует значительный потенциал для повышения диагностической эффективности в стоматологии, особенно в таких областях, как детекция кариеса, идентификация цефалометрических ориентиров и скрининг поражений слизистой оболочки полости рта. Мета-анализы подтверждают пул-чувствительность около 85% и специфичность около 93% при использовании современных CNN-архитектур.

Однако разрыв между лабораторной эффективностью и клинической валидацией остается критической проблемой. Большинство существующих исследований имеют ретроспективный дизайн, выполнены в контролируемых исследовательских условиях и не проходили внешней валидации на независимых, репрезентативных клинических выборках. При сравнении с врачами начального уровня ИИ демонстрирует статистически значимые преимущества, однако данные о сравнении с экспертами высокого уровня ограничены и неоднозначны.

Клиническое внедрение ИИ в стоматологию должно основываться на принципах доказательной медицины, с обязательным проведением проспективных многоцентровых рандомизированных исследований, оценивающих не только диагностическую точность, но и влияние на исходы лечения и экономическую эффективность. До получения таких данных ИИ-системы следует рассматривать как вспомогательные инструменты, требующие клинической верификации каждого диагностического заключения.

Благодарности: Автор выражает признательность коллегам, участвовавшим в обсуждении методологии и интерпретации результатов.

Конфликт интересов: Автор заявляет об отсутствии конфликта интересов.

Список литературы

1. Kwiatek J, Leśna M, Piskórz W, Kaczewiak J. Comparison of the Diagnostic Accuracy of an AI-Based System for Dental Caries Detection and Clinical Evaluation Conducted by Dentists. *Journal of Clinical Medicine*. 2025;14(5):1566.

2. Saikia A, et al. Artificial Intelligence in dentistry: an overview of systematic reviews and meta-analysis. *Evid Based Dent*. 2025. DOI: 10.1038/s41432-025-10160-8

3. Sowmya V, Reddy GS, Sharma A, et al. Comparison of artificial intelligence vs. junior dentists' diagnostic performance based on caries and periapical infection detection on panoramic images. *African Journal of Biological Sciences*. 2024;6(12):4561-4565.

4. Chen YJ, et al. Auxiliary evaluation of marginal ridge discrepancy in periodontal disease using deep learning on periapical radiographs. *Journal of Dental Sciences*. 2025. DOI: 10.1016/j.jds.2025.09.015

5. Shah K, Dhimole A, Pandey N, et al. Evaluation of Artificial Intelligence-Based Diagnostic Tools for Accurate Detection and Classification of Oral Lesions: A Clinical Study. *Journal of Pharmacy and Bioallied Sciences*. 2025;17(Suppl 4):S3138-S3140.

6. Araidy S, Batshon G, Mirochnik R. Artificial Intelligence Applications in Dentistry: A Systematic Review. *Oral — Health, Diseases, Therapies, and Technologies*. 2025;5(4):90.

7. Oranger A, et al. Diagnostic Support in Dentistry Through Artificial Intelligence: A Systematic Review. *Bioengineering*. 2025;12(11):1244.

8. Anonymous. Evaluating AI diagnostic accuracy in approximal dental caries detection on bitewing radiographs. *Clinical Oral Investigations*. 2026;30:207.

9. Rana A, Akhtar N. A Hybrid Deep Learning and Anomaly Detection Model for Interpretable Analysis of Dental OPG. *European Journal of Prosthodontics and Restorative Dentistry*. 2026;34(1):25-34.

10. de Lima Pereira SA, et al. Two-step pipeline for oral diseases detection and classification: a deep learning approach. *Frontiers in Oral Health*. 2025;6:1659323.