دانشنامه علمی چاوش طب آذین · chavoshteb.ir · به‌روزرسانی ۱۹ آبان ۱۴۰۵

اعتبارسنجی هوش مصنوعی پوسیدگی: شواهد کارآزمایی بالینی و محدودیت‌ها

کارآزمایی تصادفی خوشه‌ای مرتنز و همکاران نشان داد دندانپزشکانی که از نرم‌افزار هوش مصنوعی کمک گرفتند، حساسیت بالاتری در تشخیص پوسیدگی پروگزیمال بایت‌وینگ داشتند (۰٫۸۱ در برابر ۰٫۷۲) بی‌آنکه اختصاصیت به‌طور معنادار تغییر کند؛ اما تصمیم‌های درمانی تهاجمی هم بیشتر شد. شواهد کنونی از هوش مصنوعی به‌عنوان دستیار تفسیر پشتیبانی می‌کند، نه جایگزین قضاوت بالینی.

نکات کلیدی این مقاله

- در کارآزمایی تصادفی خوشه‌ای با ۲۲ دندانپزشک، AUC با کمک هوش مصنوعی ۰٫۸۹ و بدون آن ۰٫۸۵ بود.[۳]
- حساسیت با هوش مصنوعی از ۰٫۷۲ به ۰٫۸۱ رسید؛ افزایش فقط در ضایعات مینایی معنادار بود.[۳]
- همان کارآزمایی افزایش تصمیم‌های درمانی غیرتهاجمی و تهاجمی را نشان داد.[۳]
- در اعتبارسنجی ۳۲۳ دندان، حساسیت هوش مصنوعی ۰٫۵۱ تا ۰٫۷۶ و اختصاصیت ۰٫۸۸ تا ۰٫۹۷ گزارش شد.[۴]
- در مرور ۴۲ مطالعه، فقط ۱۱ مطالعه در همه حوزه‌ها خطر سوگیری پایین داشتند.[۱]

مرور سیستماتیک ۲۰۲۲ محمدرحیمی و همکاران ۴۲ مطالعه یادگیری عمیق برای تشخیص پوسیدگی را بررسی کرد و دقت را در دامنه ۶۸ تا ۹۹ درصد یافت؛ اما فقط ۱۱ مطالعه در همه حوزه‌ها خطر سوگیری پایین داشتند.[۱] فاصله میان «دقت گزارش‌شده» و «سود بالینی اثبات‌شده» همان چیزی است که اعتبارسنجی بالینی باید پر کند. این مقاله شواهد این دو سطح را از هم جدا می‌کند.

فهرست مطالب

1. سطوح اعتبارسنجی هوش مصنوعی پوسیدگی
2. کارآزمایی تصادفی تشخیص پوسیدگی
3. اعتبارسنجی چندمرحله‌ای ابزار تجاری
4. محدودیت‌های شواهد کنونی

5. جدول جمع‌بندی نکات
6. پرسش‌های متداول
7. جمع‌بندی
8. مطالعه مرتبط در دانشنامه
9. منابع

سطوح اعتبارسنجی هوش مصنوعی پوسیدگی

شواهد ابزارهای تشخیصی را می‌توان در چند سطح دید: دقت فنی مدل روی داده آزمون، دقت در مقایسه با مرجع مستقل، اثر بر عملکرد پزشک در شرایط واقعی، و در نهایت اثر بر تصمیم درمانی و پیامد بیمار. بیشتر مطالعات هوش مصنوعی پوسیدگی در دو سطح نخست قرار دارند.[۱]

محمدرحیمی و همکاران دقت مدل‌های طبقه‌بندی را روی بایت‌وینگ ۸۷٫۶ تا ۹۵٫۴ درصد و روی پری‌اپیکال ۸۲ تا ۹۹٫۲ درصد گزارش کردند، اما به‌دلیل ناهمگونی روش‌ها فراتحلیل انجام ندادند و کیفیت مطالعه و گزارش‌دهی را پایین دانستند.[۱] فراتحلیل ۲۰۲۵ لوک و رزالله نیز با گنجاندن ۱۴ مطالعه از ۲۱ مطالعه مرور، بر تنوع قابل‌توجه نتایج و نیاز به پژوهش بیشتر درباره عوامل مؤثر بر کارایی تأکید کرد.[۲]

سطوح شواهد برای ابزار هوش مصنوعی پوسیدگی

سطح	پرسش	نمونه شواهد در این مقاله
دقت فنی	مدل روی داده آزمون چقدر درست طبقه‌بندی می‌کند؟	مرور ۴۲ مطالعه با دقت ۶۸ تا ۹۹ درصد
اعتبارسنجی با مرجع	در برابر توافق مشاهده‌گران یا مرجع مستقل چگونه است؟	اعتبارسنجی چهارمرحله‌ای روی ۳۲۳ دندان
اثر بر عملکرد پزشک	آیا تشخیص دندانپزشک با کمک ابزار بهتر می‌شود؟	کارآزمایی تصادفی خوشه‌ای با ۲۲ دندانپزشک
اثر بر درمان و بیمار	آیا تصمیم‌ها و پیامدها بهتر می‌شوند؟	فقط شواهد غیرمستقیم؛ افزایش درمان تهاجمی در همان کارآزمایی

منبع: محمدرحیمی و همکاران (۲۰۲۲)، سابو و همکاران (۲۰۲۴) و مرتنز و همکاران (۲۰۲۱)؛ دسته‌بندی سطوح جمع‌بندی تحریریه است.

کارآزمایی تصادفی تشخیص پوسیدگی

مرتنز و همکاران یک کارآزمایی تصادفی خوشه‌ای متقاطع طراحی کردند: ۲۲ دندانپزشک هر کدام ۲۰ بایت‌وینگ تصادفی از مجموعه ۱۴۰ بایت‌وینگ را ارزیابی کردند که ۱۰ تصویر با کمک نرم‌افزار تجاری مبتنی بر شبکه عصبی کاملاً پیچشی و ۱۰ تصویر بدون آن بود. مرجع را ۴+۱ متخصص مستقل به‌صورت پیکسلی تعیین کرده بودند.[۳]

دندانپزشکان با کمک هوش مصنوعی AUC بالاتری داشتند (۰٫۸۹ در برابر ۰٫۸۵)، عمدتاً به این دلیل که حساسیت آنان به‌طور معنادار بالاتر بود (۰٫۸۱ در برابر ۰٫۷۲)، در حالی که اختصاصیت تغییر معناداری نکرد.[۳] این افزایش حساسیت در ضایعات مینایی دیده شد، نه در ضایعات عاج اولیه یا پیشرفته.[۳]



منبع: Mertens S و همکاران. J Dent 2021;115:103849. اختصاصیت تفاوت معنادار نداشت.

عملکرد دندانپزشکان با و بدون کمک هوش مصنوعی در تشخیص پوسیدگی پروگزیمال بایت‌وینگ؛ نمودار اختصاصی تحریریه بر پایه داده‌های کارآزمایی مرتنز و همکاران.

> «هوش مصنوعی می‌تواند دقت تشخیصی دندانپزشکان را افزایش دهد، اما ممکن است تصمیم‌های درمانی تهاجمی را نیز بیشتر کند.»
>
> “AI can increase dentists' diagnostic accuracy but may also increase invasive treatment decisions.”

> — مرتنز و همکاران، جمله نتیجه‌گیری چکیده کارآزمایی (۲۰۲۱)[۳]

پیام بالینی کارآزمایی

ضایعه مینایی پروگزیمال اغلب با درمان غیرتهاجمی مدیریت می‌شود؛ بنابراین حساسیت بیشتر فقط زمانی سودمند است که به درمان مناسب ترجمه شود. نویسندگان کارآزمایی پیشنهاد کرده‌اند تفاوت اثر هوش مصنوعی میان دندانپزشکان بررسی شود و راهنمایی روشنی درباره انتخاب درمان پس از تشخیص با کمک هوش مصنوعی ارائه شود.[۳]

اعتبارسنجی چندمرحله‌ای ابزار تجاری

سابو و همکاران سطوح پروگزیمال ۳۲۳ دندان را روی رادیوگرافی داخل دهانی در چهار مرحله بررسی کردند: ارزیابی دو مشاهده‌گر مستقل، ارزیابی دوباره با کمک هوش مصنوعی پس از ۴ ماه، ارزیابی مجدد با شبکه پیشرفته‌تر و حذف موارد اختلاف انسانی.[۴] ضریب کاپای کوهن میان مشاهده‌گران و سامانه در مراحل مختلف از ۰٫۴۹ تا ۱ متغیر بود.[۴]

حساسیت سامانه ۰٫۵۱ تا ۰٫۷۶، اختصاصیت ۰٫۸۸ تا ۰٫۹۷ و دقت تشخیصی ۰٫۷۶ تا ۰٫۸۶ گزارش شد.[۴] یعنی سامانه در رد پوسیدگی قوی‌تر از شناسایی آن بود. چند نفر از نویسندگان این مطالعه کارمند شرکت سازنده نرم‌افزار بودند؛ تعارض منافعی که در خوانش نتایج باید در نظر گرفت.[۴]

مقایسه دو مطالعه اعتبارسنجی بالینی

ویژگی	مرتنز و همکاران (۲۰۲۱)	سابو و همکاران (۲۰۲۴)
طرح	کارآزمایی تصادفی خوشه‌ای متقاطع	اعتبارسنجی چهارمرحله‌ای
تصویر	بایت‌وینگ	بایت‌وینگ و پری‌اپیکال
نمونه	۲۲ دندانپزشک؛ ۱۴۰ بایت‌وینگ	۳۲۳ دندان؛ دو مشاهده‌گر
مرجع	۴+۱ متخصص مستقل، پیکسلی	توافق مشاهده‌گران انسانی و سامانه
یافته کلیدی	حساسیت ۰٫۸۱ با AI در برابر ۰٫۷۲ بدون AI	حساسیت ۰٫۵۱ تا ۰٫۷۶؛ اختصاصیت ۰٫۸۸ تا ۰٫۹۷
محدودیت	افزایش درمان تهاجمی؛ سود فقط در مینا	مرجع مبتنی بر توافق؛ تعارض منافع

منبع: چکیده مقالات ذکرشده؛ اعداد عیناً نقل شده‌اند.

محدودیت‌های شواهد کنونی

سه محدودیت در این شواهد تکرار می‌شود. نخست، کیفیت روش‌شناختی: در مرور محمدرحیمی و همکاران فقط ۱۳ مطالعه (۳۱٪) نگرانی پایین درباره کاربردپذیری داشتند.[۱] دوم، ناهمگونی معیارها: طیفی از شاخص‌های دقت تصویر، شیء یا پیکسل به کار رفته که ادغام نتایج را دشوار می‌کند.[۱] سوم، مرجع تشخیصی: وقتی مرجع بر توافق مشاهده‌گران استوار است، خطای مشترک انسان و سامانه پنهان می‌ماند.[۴]

لوک و رزالله با وجود نتایج امیدوارکننده، بر چالش‌های پیاده‌سازی و دسترسی به داده تأکید کردند.[۲] برای رادیولوژیست، این یعنی عدد دقت یک محصول را باید همراه با جمعیت آزمون، نوع گیرنده، مرجع و طراحی مطالعه خواند.

ملاحظه برای پیاده‌سازی

برداشت تحریریه: پیش از به‌کارگیری ابزار هوش مصنوعی پوسیدگی در گردش کار، عملکرد آن را روی نمونه‌ای از تصاویر همان مرکز و همان نوع حسگر بسنجید، خروجی را به‌عنوان نظر دوم ثبت کنید و پروتکل درمانی ضایعات مینایی را از قبل مشخص کنید تا حساسیت بیشتر به درمان تهاجمی غیرضروری نینجامد.



مسیر اعتبارسنجی ابزار هوش مصنوعی پوسیدگی پیش از استفاده بالینی؛ طراحی اختصاصی تحریریه چاوش طب آذین.

جدول جمع‌بندی نکات

نکات کاربردی اعتبارسنجی هوش مصنوعی پوسیدگی

محور	نکته کلیدی
شاهد سطح بالا	یک کارآزمایی تصادفی؛ حساسیت ۰٫۸۱ در برابر ۰٫۷۲
محدوده سود	بیشتر در ضایعات مینایی
ریسک	افزایش تصمیم درمانی تهاجمی
اعتبارسنجی مستقل	اختصاصیت بالا؛ حساسیت متغیر
پیاده‌سازی	سنجش محلی و نقش نظر دوم

منبع: جمع‌بندی تحریریه بر پایه منابع فهرست‌شده در همین مقاله.

پرسش‌های متداول

آیا هوش مصنوعی دقت تشخیص پوسیدگی دندانپزشک را بالا می‌برد؟

در کارآزمایی تصادفی، AUC دندانپزشکان با کمک هوش مصنوعی ۰٫۸۹ و بدون آن ۰٫۸۵ بود، عمدتاً به‌دلیل افزایش حساسیت.[۳]

افزایش حساسیت در کدام ضایعات دیده شد؟

فقط در ضایعات مینایی؛ در ضایعات عاج اولیه یا پیشرفته افزایش معنادار نبود.[۳]

آیا هوش مصنوعی درمان تهاجمی را افزایش می‌دهد؟

در همان کارآزمایی، حساسیت بیشتر با افزایش تصمیم‌های درمانی غیرتهاجمی و تهاجمی همراه بود.[۳]

دقت ابزارهای تجاری در اعتبارسنجی مستقل چقدر است؟

در مطالعه ۳۲۳ دندان، حساسیت ۰٫۵۱ تا ۰٫۷۶ و اختصاصیت ۰٫۸۸ تا ۰٫۹۷ گزارش شد.[۴]

چرا دقت بالای مطالعات اولیه کافی نیست؟

در مرور ۴۲ مطالعه، فقط ۱۱ مطالعه در همه حوزه‌ها خطر سوگیری پایین داشتند و کیفیت گزارش‌دهی پایین بود.[1]

هوش مصنوعی پوسیدگی جایگزین رادیولوژیست می‌شود؟

شواهد فعلی آن را دستیار تصمیم درباره وجود یا نبود پوسیدگی می‌داند، نه جایگزین قضاوت بالینی.[1]

جمع‌بندی

اعتبارسنجی هوش مصنوعی پوسیدگی از مرحله دقت فنی فراتر رفته و دست‌کم یک کارآزمایی تصادفی افزایش حساسیت دندانپزشکان را نشان داده است. اما همان شواهد نشان می‌دهد سود بیشتر در ضایعات مینایی است و ممکن است درمان تهاجمی را افزایش دهد. تا زمانی که کارآزمایی‌های پیامدمحور بیشتری منتشر شود، استفاده مسئولانه یعنی سنجش محلی، نقش نظر دوم و پروتکل درمانی روشن.

مطالعه مرتبط در دانشنامه

- هوش مصنوعی در رادیولوژی دندان و فک
- آشکارساز دیجیتال داخل دهانی: CMOS، CCD و PSP
- سنسور CMOS داخل دهانی و سینتیلاتور
- MTF سنسور دندانی و رزولوشن
- توجیه تصویربرداری در دهان و فک
- تصویربرداری اندودانتیکس از پری‌اپیکال تا CBCT

منابع

شماره‌های بالانویس در متن به همین فهرست پیوند دارند؛ هر منبع با شناسه DOI یا نشانی رسمی ناشر قابل بررسی است.

Mohammad-Rahimi H, Motamedian SR, Rohban MH, et al. Deep learning for caries detection: a systematic review. J Dent. 2022;122:104115. DOI PubMed

Luke AM, Rezallah NNF. Accuracy of artificial intelligence in caries detection: a systematic review and meta-analysis. Head Face Med. 2025;21(1):24. DOI PubMed

Mertens S, Krois J, Cantu AG, Arsiwala LT, Schwendicke F. Artificial intelligence for caries detection: randomized trial. J Dent. 2021;115:103849. DOI PubMed

Szabó V, Szabó BT, Orhan K, et al. Validation of artificial intelligence application for dental caries diagnosis on intraoral bitewing and periapical radiographs. J Dent. 2024;147:105105. DOI PubMed

---

درباره نویسنده: تیم تحریریه علمی چاوش طب آذین

این مقاله را تیم تحریریه علمی چاوش طب آذین بر پایه منابع داوری‌شده و بیانیه‌های رسمی فهرست‌شده تهیه کرده است؛ ادعای بازبینی بالینی مستقل ندارد و جایگزین قضاوت متخصص نیست. آخرین به‌روزرسانی: ۱۹ آبان ۱۴۰۵.