

اصلاح خودکار رانش عملکرد تحت تغییر در فرآیند اخذداده در طبقه بندی تصاویر پزشکی

ملانی روشنیتز^{۱,۲} گالوین خارا، جو برزلن، نیشا شارما^۳،
جاناتان جی. جیمز^۴، او آمروزایی، آدام هروکس^۵، پیتر کچکمنی^۶ و بیاس ریکن^۷
 او بن گلکر^۸

دریافت: ۲۴ فوریه ۲۰۲۳

پذیرفته شده: ۱۰ اکتبر ۲۰۲۳

Published online: 19 October 2023

بررسی به روزرسانی ها

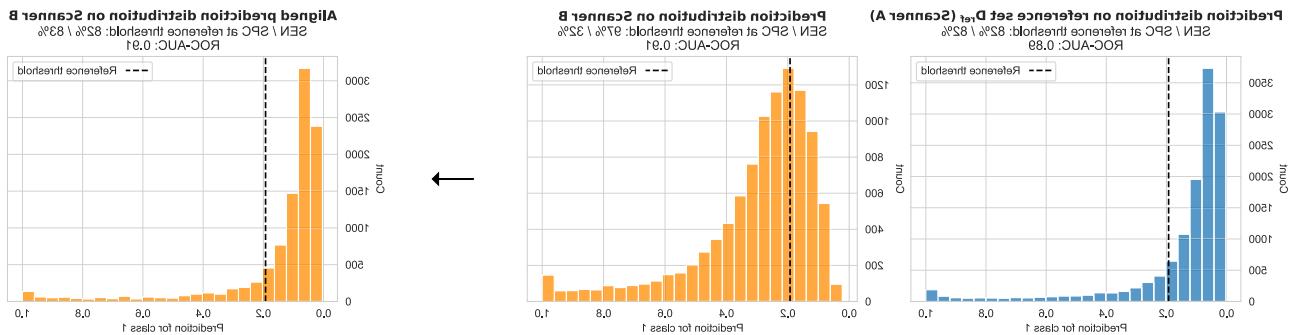
مدل های پیش بینی مبتنی بر تصویر برای تشخیص بیماری، به تغییرات در اکتساب داده ها مانند تعویض سخت افزار اسکنر یا به روزرسانی نرم افزار پردازش تصویر حساس هستند. تفاوت های حاصل در ویژگی های تصویر ممکن است منجر به انحراف در معیارهای عملکرد بالینی مرتب شود که من تواند در تضمیم گیری بالینی، حتی برای مدل هایی که از نظر مساحت زیرمنحی مشخصه عملکرد گیرنده تعیین می یابند، آسیب ایجاد کند. ما هم ترازی پیش بینی بدون نظارت را پیشنهاد می کنیم، یک روش کالیبراسیون مجدد خودکار عمومی که نیازی به حاشیه نویس حقایق زمینی ندارد و فقط به مقادیر محدودی از تصاویر نمونه بدون برچسب از توزیع داده های تغییر یافته نیاز دارد. ما اثربخشی روش پیشنهادی را برای تشخیص و اصلاح انحراف عملکرد در غربالگری سرطان پستان مبتنی بر ماموگرافی و روی داده های هیستوپاتولوژی در دسترس عموم نشان می دهیم. ما نشان می دهیم که روش پیشنهادی من تواند عملکرد مورد انتظار را از نظر حساسیت/اختصاصیت تحت سناریوهای مختلف واقع بینانه تغییر اکتساب تصویر حفظ کند، بنابراین یک محافظ مهم برای استقرار بالینی ارجاعیه من دهد.

نقاطبر اساس یک بدء بستان معنادار بالینی خاص SPC/SEN. در اینجا، یک نقطه عملیاتی با یک آستانه کالیبر شده در امتیاز پیش بینی پیوسته هوش مصنوعی مرتب شده است. آستانه ها اغلب از پیش تعیین شده اند، در حالت ایده آل به عنوان بخشی از مطالعات اعتبارسنجی در مقیاس بزرگ. آستانه ها همچنین ممکن است به عنوان بخشی از بینهای سازی محلی با استفاده از داده های تاریخی از یک سایت استقرار جدید تنظیم شوند.^۹ در اینجا، فرض بر این است که چنین داده های اعتبارسنجی، تماشده داده های جدید ایمیاران نادیده هستند و همچنان باقی مانند نقطه عملیاتی مرتب در محدوده مقایر تأیید شده قرار دارد.

تغییراتی که پس از استقرار رخ می دهنند، چالش اساسی برای هوش مصنوعی تصویربرداری پزشکی ایجاد می کنند.^{۱۰} سیستم های فعلی مستعد تفاوت در ویژگی های تصویربرداری هستند. تغییرات در تصاویر ورودی، به عنوان مثال، به دلیل تعویض سخت افزار اسکنریا به روزرسانی نرم افزار بازسازی و پردازش تصویر، من تواند باعث ایجاد تفاوت در پیش بینی های خروجی هوش مصنوعی شود (شکل ۱ را ببینید).^{۱۱} این تغییرات ممکن است در هر زمانی پس از استقرار رخ دهنند و اغلب خارج از کنترل هستند.

هوش مصنوعی (AI) نویدبخش تجزیه و تحلیل عینی تر، دقیق تر و مقرن به صرفه تر داده های تصویربرداری است و من تواند اساساً گردش کار بالینی را در تشخیص های مبتنی بر تصویر و غربالگری جمعیت متحول کند. استفاده از هوش مصنوعی من تواند به کاهش فشار بر خدمات بهداشتی، به عنوان مثال، از طریق اولویت بندی خودکار موارد بحرانی، کمک کند.^{۱۲} یا با ارجاع نظرات دوم در غربالگری تشخیصی.^{۱۳}

موارد استفاده مختلف، بسته به نقش سیستم هوش مصنوعی در گردش کار بالینی، الزامات متفاوتی در مورد عملکرد هوش مصنوعی خواهند داشت. انتظار می رود یک سیستم هوش مصنوعی که برای اولویت بندی یا اولویت بندی استفاده من شود، حساسیت (SEN) تقريباً کاملی داشته باشد، در حالی که اختصاصیت (SPC) پایین ممکن است قابل قبول باشد. از سوی دیگر، انتظار می رود هوش مصنوعی به عنوان یک خواننده دوم در غربالگری سرطان سینه با خوانش دوگانه، در سطوح SEN/SPC مشابه یک خواننده انسانی عمل کند. بسیاری از سیستم های هوش مصنوعی، مدل های پیش بینی همه کاره ای هستند که من توانند در عملیات مختلف مورد استفاده قرار گیرند.



رویک اسکر کاملاً امدادی، می‌توانیم مشاهده کیم که علیرغم تعمیم ROC-AUC در پیش‌بینی هاو کالیبراسیون مدل وجود دارد، جایزی اعمال روش UPA پیشنهادی بر حسب توزیع پیش‌بینی هاو موزن SPC/SEN. داده‌های منبع به صورت یک فایل داده منبع ارایه شده‌اند.

شکل ۱ / نمایش تغییر در پیش بینی مدل هنگام اعمال یک مدل تشخیص بدخوبی ماموگرافی به یک اسکنر جدید از چپ به راست: الف: توزیع پیش بینی های مشاهده شده در مجموعه مرجع که از یک اسکنر مبادله اکثربت مجموعه توسعه حاصل شده و برای تعیین آستانه طبقه بندی استفاده شده است ب: توزیع پیش بینی های مشاهده شده هنگام اعمال این مدل

یمنی و قابلیت اطمینان هوش مصنوعی تصویربرداری پزشکی، شامل توسعه دهنده‌گان هوش مصنوعی، متخصصان مراقبت‌های بهداشتی، بیماران، تنظیم کنندگان مقررات و سیاست‌گذاران.

تتاح

اماکن چهار مجموعه داده ماموگرافی پستان با برجسب های بدینمی تایید شده توسعه یافته اند.^{۱۰} این اتفاقاً در بیوپسی جانی استفاده می کنیم. برای هر شرکت کننده، چهار تصویر ثبت می شود (دو مجموعه داده آموزشی دی قطاراز OPTIMAM است^{۱۱}) که یک مجموعه داده مباری هر پستان است. در حالی که تمام مجموعه های ارزیابی دیگر، مجموعه داده های داده غنی از بریتانیا است، در حالی که تمام مجموعه های ارزیابی دیگر، مجموعه داده های غربالگری از سه سایت در بریتانیا و چهار سایت در مجارستان هستند که نماینده جمعیت های واقعی در برنامه های ملی غربالگری سرطان پستان مربوطه می باشند. برای همه مجموعه های ارزیابی، اگر بیش از یک مطالعه برای هر شرکت کننده وجود داشت، ما فقط یک رسمی را (به صورت تصادف نمونه گیری شده) نگه داشتیم. ما علاوه بر این، اثربخشی WILDS Camelyon17 UPA را بر روی این مجموعه داده های هیستوپاتولوژی در دسترس مجموعه ارزیابی می کنیم.^{۱۲} این تغییرات مجموعه داده های مطالعه تغییرات مجموعه داده ناشی از تغییرات در پروتکل های رنگ آمیزی از یک بیمارستان به بیمارستان دیگر طراحی شده است. این مجموعه داده شامل تکه هایی از تصاویر کل اسیدی با برجسب های تکه ای است که نشان من دهد بافت سرطانی است یا خیر. مجموعه داده آموزشی از داده های سه بیمارستان تشکیل شده است و مجموعه مرجع از همان مکان ها نمونه برداری شده است. ارزیابی روی دو مجموعه داده دیده نشده از دو مکان جدید بیمارستان انجام می شود. خلاصه ای از وزنگ های مجموعه داده را می توانید در جداول بیاورد.^{۱۳}

از سازنده هوش مصنوعی، اگرچه تغییرات در وزنگی های تصویر ممکن است لزوماً بر قدرت پیش بینی کل (یعنی توانایی تمايز بین مواد مثبت و منفی) تأثیر نگذارد، اما هرگونه تغییر یا اختلاف در توزیع خروجی ممکن است آستانه های کالیبره شده را نامعتبر کند. این امر مستقیماً بر بدء سitan SEN/SPC و در نتیجه، نقطه عملیاتی مورد نظر سیستم هوش مصنوعی تأثیر می گذارد. در صورت عدم تشخیص، چنین اختلاف عملکرد بالینی می تواند منجر به تشخیص کمتر یا بیشتر از حد معمول شود و عواقب شدیدی بر این نتیجه داشته باشد.

نیازبمیری به یک روش مؤثر برای نظارت بر هوش مصنوعی در استقرارهای دنیای واقعی وجود دارد که امکان اصلاح خودکار انحراف عملکرد ناشی از تغییرات در اخذ تصویر را فراهم کند. موضوع نظارت بر عملکرد، موضوع میرم برای استقرار عملی مدل های یادگیری ماشینی، به ویژه در کاربردهای جیان مانند تشخیص بیماری است^{۱۳}. در گذشته، روش هایی برای تشخیص چنین انحرافاتی در عملکرد و به روزسازی مدل ها ارایه شده است.^{۱۴-۱۵}، اما همه آنها نیاز به حاشیه نوبیس انسانی از نمونه های جدید دارند. با این حال، به دست آوردن حاشیه نوبیس ها در زمان واقعی و پیوند دادن آنها با تصاویر به دست آمدۀ معمول‌گر از عمل بالین امکان پذیر نیست، به ویژه در کاربردهای مانند غربالگری سلطان پستان که در آن تشخیص تأیید شده (مثلًاً بدختیمی اثبات شده با نوبیس) در زمان تجزیه و تحلیل تصویر در دسترس نیست. برخی از روش ها برای تشخیص خودکار تغییرات بر اساس آزمون های آماری که پیش بینی های مدل را مقایسه کنند، پیشنهاد شده اند.^{۱۶} اما همچنان نیاز برآورده نشده ای به روش هایی وجود دارد که امکان انجام خودکار را فراهم می کنند درست را نشیش عملکرد در غیاب هرگونه اطلاعات تشخیصی (مانند برجسته های بیماری). این دقیقاً تمکز این کار است.

تنظیمات آزمایشی

اما ماجندین سناریوی واقع بینانه از تغییر اکتساب را برای ارزیابی اثربخش UPA در زمینه تشخیص سرطان پستان در ماموگرافی دیجیتال و طبقه بندی بافت در هیستوپاتولوژی UPA روشن ساده است که پیش بینی های مدل را از یک دامنه اکتساب نامرئی (مثلث) داده های به دست آمده با یک اسکنر جدید با توزیع پیش بینی مرتع شدت شده در یک دامنه شناخته شده (مثلث) داده های مجموعه اعتبارسنجی که در طول تنوشه روش استفاده شده است) هم تراز می کند. هم تراز کردن پیش بینی با استفاده از نجمخط تکه ای حاصل می شود.

دراین کار، ما یک رویکرد ساده، عمومی و مؤثر برای هم ترازی پیش بینی بدون نظارت (UPA) را پیشنهاد و ارزیابی می کنیم که قادر به تشخیص و اصلاح انحراف عملکرد بالینی مرتبط هوش مصنوعی ناشی از تغییر در اکتساب است. آزمایش های ما اثربخش UPA را در چندین سناریوی دینایی واقعی در زمینه غربالگری پستان و هیستوپانوکلوزی نشان می دهد. در کاربرد غربالگری پستان، ابتدا نشان می دهیم که UPA قادر است خروجی های یک مدل بهینه سازی شده روی یک فروشنده سخت افزار را برای بازیابی SEN/SPCD مورد نظر روی فروشندهگان مختلف در سه مجموعه داده غربالگری پستان در مقایسه بازگرد پرینتیماً تطبیق دهد. ما ماهیت عمومی UPA را با استفاده از WILDS Camelyon17 که در دسترس عموم است، تأیید می کنیم. مجموعه داده های نشان می دهد یک مدل طبقه بندی بهینه شده بر روی یک پروتکل رنگ آمیزی خاص من تواند طور خودکار با یک پروتکل رنگ آمیزی جدید سازگار شود. نکته مهم این است که UPA به گونه ای طراحی شده است که امکان کالیبراسیون مجدد مدادوم را فراهم می کند. زیرا به طور خودکار تغییرات مشاهده شده در توزیع پیش بینی را در طول زمان تشخیص داده و تنظیم می کند. ما این موضوع را با شبیه سازی سناریوهای تغییر چندگانه جمع آوری، از جمله معرفی اسکنرهای جدید و به روزرسانی های نرم افزار پذراش تصویری، نشان می دهیم. ما در مورد الزامات داده ها، فرضیات و محدودیت های روش پیشنهادی بحث می کنیم. ما معقدمند که این کار دای، هر کسی، که نگران ...

میزگردانی کنید؛ این سوابق

مسناریوی اول: استقرار در یک سایت جدید
درین مجموعه اول آزمایش ها، مسناریوی را شیوه سازی می کنیم که در آن من خواهیم
مدل را که بر اساس داده های اعتبارسنجی از یک سایت مرجع بهینه شده است، به سایت های
جدید دیده نشده منتقل کنیم. در اینجا، ممکن است به دلیل استفاده از اسکریپت های مختلف در
سایت های جدید (امانگو اف دیجیتال) یا به طور گسترده تر به دلیل تفاوت در پروتکل
ضمیمه داده را، بقیه در ایندیکاتور را خود دهد.

جدول ۱ | آمار مجموعه داده های غربالگری سرطان پستان

مجموعه داده ها	دستگاه	کشور	نوع	ناموارد	ن تصاویر	درصد تصاویر مثبت
Aسکنتر	هولولوزیک	بریتانیا	مجموعه مرجع دی مرچ	۳۲۲۱	۱۲,۸۸۴	۲.۵%
Bاسکنتر	جوتو	مغارستان	نادیده	۴۱۵۲	۱۶,۶۰۸	۱.۸%
Cاسکنتر	زیمنس	بریتانیا	نادیده	۱۵,۶۱۶	۱۵,۶۱۶	۱.۹%
Dاسکنتر	پداشتو درمان جراحی کنترل	بریتانیا	نادیده	۱۶,۴۶۰	۱۶,۴۶۰	۱.۸%

داده های منبع به صورت یک قابل داده منبع ارائه می شوند.

اسکنتر B به تدریج معرفی می شود، به این معنی که تعداد کل اسکن ها با گذشت زمان افزایش می یابد تا زمانی که سبیت مساوی از هر دو اسکنتر وجود داشته باشد، که در نتیجه تعداد کل موارد دو برابر می شود. آستانه های طبقه بندی برای اسکنتر A بینه شدن. نتایج را می توانید در شکل ۱ بباید.^۴(ردیف پایین)، دوباره UPA قادر است در طول زمان تعادل SEN/SPC را حفظ کند.

سناریو^۴: به روزرسانی پردازش تصویر در این سناریو، ما تأثیر یک به روزرسانی OEM اعمال شده بر الگوریتم پردازش تصویر برای یک اسکنتر واحدرا اداره می کند و در زمان استقرار هوش مصنوعی، آستانه های طبقه بندی برای نسخه اولیه نرم افزار پردازش تصویر بهینه شده اند. در نقطه زمانی بعدی T1، فرض می کنیم که سازنده اسکنتر، به روزرسانی نرم افزاری را بر الگوریتم پردازش تصویر خود اعمال کرده است که منجر به افزایش وضوح تصویر می شود. چنین به روزرسانی های تا حد زیادی خارج از کنترل توسعه دهنگان مدل هوش مصنوعی هستند. نتایج این شیوه سازنی را می توانید در شکل ۱ بباید.^۵ افزایش وضوح تصویر باعث تغییر ناگهانی و قابل توجه در تعادل SEN/SPC می شود. اعمال UPA منجر به سازگاری سریع با ویژگی های جدید تصویر می شود و مدل به سرعت در بدء بستان مورد نظر SEN/SPC عمل می کند.

تأثیر اندازه مجموعه داده ها
برای کسب بیشتر در مورد میزان داده های مورد نیاز برای دستیابی به عملکرد خوب در مرحله هم ترازی UPA، حساسیت روش هم ترازی را نسبت به اندازه مجموعه مرجع (از دامنه منبع) و همچنین اندازه مجموعه هم ترازی (از دامنه جدید) تجزیه و تحلیل می کنیم. نتایج مربوط به اسکنتر ماموگرافی B را می توانید در شکل ۱ بباید.^۶ به همراه نتایج اضافی برای سایر مجموعه داده ها که در یادداشت تکمیلی ۲ ارائه شده است. این نتایج نشان می دهد که یک مجموعه هم ترازی به کوچک ۲۵٪ مورد (... تصویر) برای بازیابی تغییر SEN/SPC در مجموعه داده های ماموگرافی کافی است. برای مجموعه مرجع، با ۵۰٪ مورد در مجموعه هم ترازی، دریافتیم که ۱۰۰٪ مورد برای نتایج پایدار کافی است. در عمل، می توان از همان مجموعه مرجعی که برای اعتبارسنجی مدل و انتخاب آستانه طبقه بندی استفاده شده است، استفاده کرد.

بحث

ما یک روش مؤثر برای اصلاح انحراف عملکرد بالین مرتبط با تغییرات در خط لوله اخذ تصویر(Mتلار) تغییض اسکنترها، به روزرسانی نرم افزار پردازش تصویر، استفاده از پروتکل های رنگ آمیزی مختلف(پیشنهاد کرده ایم، این موضوع در سناریوهای مختلف، از تغییرات یکباره هنگام استقرار در یک سایت جدید گرفته تا انحرافات تدریجی و پیشرونده در ریزگری های اخذ تصویر در طول زمان، نشان داده شده است، آزمایش های ما نشان می دهد که یک ROC-AUC قابل تعیین لزوماً به این معنی نیست که آستانه های طبقه بندی انتخاب شده (یعنی نقاط عملیاتی بالین) به داده های دیده نشده داده از شوند. همانطور که در شکل ها نشان داده شده است^۷، در حالی که تعادل ROC-AUC در تمام مجموعه داده ها تعیین داده شدند، در حالی که تعادل SEN/SPC هنگام اعمال یک آستانه طبقه بندی از پیش تعریف شده، به طور قابل توجهی در بین سایت های متفاوت بود. نتایج ما نشان می دهد که روش هم ترازی پیش بینی بدون نظرات و بیشنهادی می توانند با موفقیت تغییرات آستانه را در مجموعه داده های دیده نشده در دو وظیفه طبقه بندی تصویر بسیار متفاوت کاوش دهد. UPA توائیست بدون استفاده از هیچ برچسبی از دامنه هدف دیده نشده، موازنۀ SEN/SPC مورد نظر را بازیابی کند. نکته مهم این است که این روش ساده اما مؤثر می تواند برای به روزرسانی های مداوم مدل استفاده شود و به ... کمک کند.

جدول ۲ | آمار مجموعه داده های غربالگری سرطان پستان

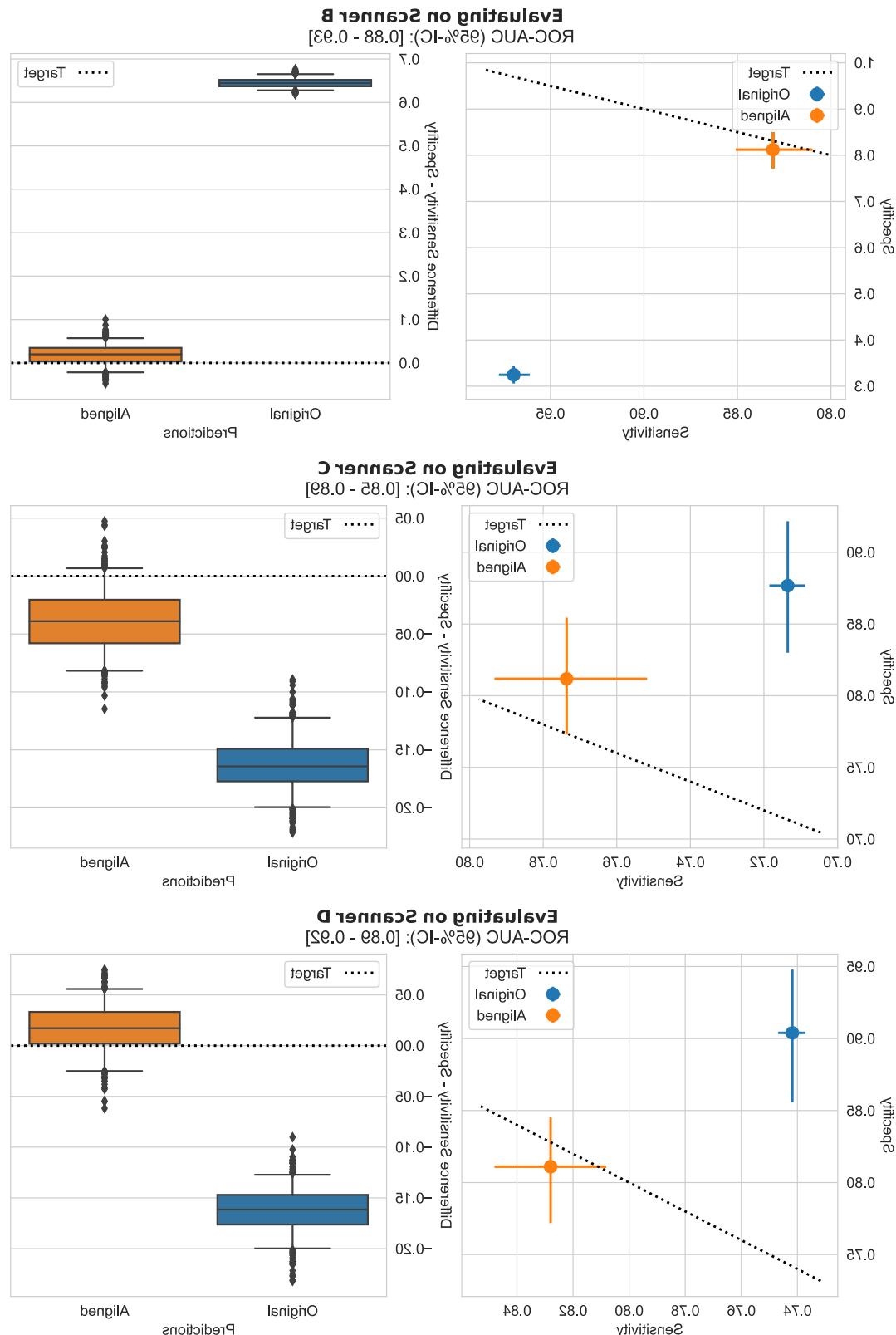
مجموعه داده ها	دستگاه	کشور	نوع	ناموارد	ن تصاویر	درصد مثبت
S3 سایت های نا	WILDS Camelyon17	سایت های نا	مجموعه مرجع دی مرچ	۳۳,۵۶	۳۳,۵۶	۵٪
S4 سایت	نادیده	آستانه های	نادیده	۳۴,۹۴	۳۴,۹۴	۵٪
S5 سایت	نادیده	آستانه های	نادیده	۸۵,۰۴	۸۵,۰۴	۵٪

داده های منبع به صورت یک قابل داده منبع ارائه می شوند.

از جمله استفاده از پروتکل های رنگ آمیزی مختلف (هیستوپاتولوژی)، برای وظیفه تشخیص سرطان پستان، از یک مجموعه داده آزمایشی از یک اسکنتر که در طول توسعه مدل دیده نشده است، استفاده می کنیم. برای وظیفه طبقه بندی بافت هیستوپاتولوژی، داده های ارزیابی از بیمارستان دیگری با استفاده از پروتکل های رنگ آمیزی مختلف من آیند. در هر دو مورد، فرض می کنیم که مدل های طبقه بندی مربوطه بر روی داده های اعتبارسنجی مرجع (قبل از استقرار در مکان های جدید) از قبل کالیبره شده اند. آستانه طبقه بندی در نقطه عملیاتی بالین انتخاب شده است که در آن SEN برابر با SPC در این داده های مرجع است. ما ابتدا عملکرد مدل را از نظر SEN/SPC بر روی داده های ارزیابی ارزیابی می کنیم. معیارها بر اساس مساحت تصویر در زیر منحنی مشخصه عملکرد گیرنده (ROC-AUC) و SEN/SPC در شکل ۱ گزارش شده اند.^۲ برای تشخیص سرطان سینه و در شکل ۱^۳ برای Camelyon17 ماهمهجنین شاخص بودن را گزارش می دهیم^۳، یک معیار ترکیبی که حساسیت و ویژگی را به طور همزمان، قبل و بعد از اعمال UPA در نظر می گیرد. ما دریافتیم که بدون UPA، تغییر قابل توجهی در تعادل SEN/SPC در تمام وظایف و مجموعه داده های ارزیابی وجود دارد. بدون UPA، مدل دیگر در نقطه عملیاتی مورد نظر عمل نمی کند. برعکس، نتایج نشان می دهد که پس از اعمال روش هم ترازی م، تعادل SEN/SPC تا حد زیادی برای داده های دیده نشده بازیابی می شود. علاوه بر این، در یادداشت تکمیلی ۱، نشان می دهیم که UPA مستقل از انتخاب خاص نقطه عملیاتی (مثلث، ویژگی هدف از پیش تعریف شده) نیز عمل می کند.

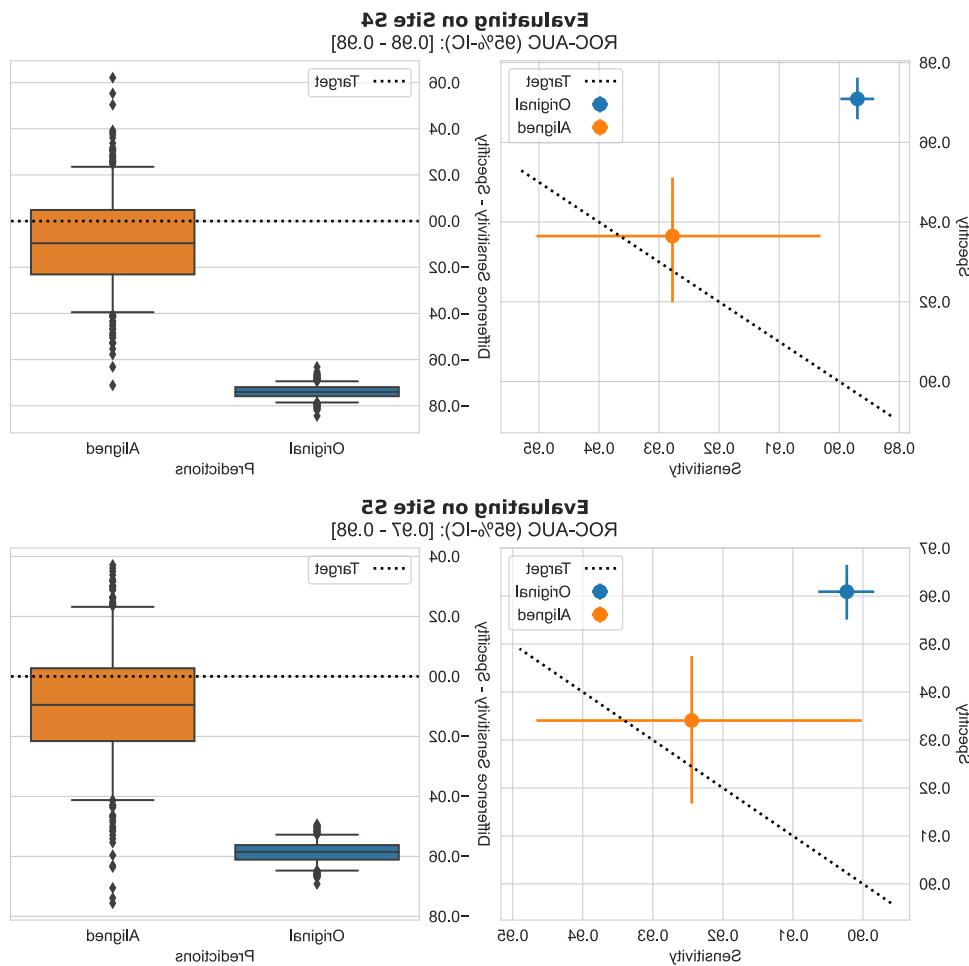
سناریو^۲: انتقال به یک اسکنتر جدید در اینجا، ما سناریوی را شبیه سازی می کنیم که نیاز به به روزرسانی های مداوم مدل دارد. که در آن یک اسکنتر جدید نصب می شود و یک اسکنتر قبیلی به تدریج از رده خارج من شود. یک دوره گذار برنامه ریزی شده وجود دارد که در آن اسکنترهای قدیمی و جدید به صورت موازی با تغییر تدریجی به اسکنتر جدید استفاده می شوند. در ابتدا، فقط داده های از اسکنتر A و در انتهای فقط داده های از اسکنتر B وارد می شوند. آستانه های طبقه بندی مدل هش مصنوعی برای اسکنتر A بهینه شده اند. نتایج را می توانید در شکل ۱ مشاهده کنید.^۴ (بالا)، که در آن سمت چپ تعداد موارد پردازش شده توسط هر اسکنتر از طول زمان برای هر سناریو نشان می دهد و سمت راست نتایج مربوطه را از نظر حساسیت و ویژگی مدل در طول زمان، با و بدون اعمال UPA به طور کل و در سطح اسکنتر نشان می دهد. از شکل، مشاهده می کنیم که معرفی اسکنتر جدید منجر به تغییرات قابل توجهی در SEN/SPC در صورت عدم استفاده از UPA می شود، در حالی که اعمال UPA تعادل مطلوب SEN/SPC را در طول زمان حفظ می کند.

سناریو^۳: اضافه کردن یک اسکنتر جدید مشابه سناریوی بالا، در اینجا یک اسکنتر جدید B نصب شده است، با این حال، اسکنتر قدیمی A همچنان به صورت موازی مورد استفاده قرار خواهد گرفت. اسکنتر جدید



هر معیار، ستون سمت راست: تفاوت بین حساسیت و اختصاصیت قبل و بعد از هم ترازی. نمودارهای جعبه ای از ۵۰۰ نمونه برداری مکرر از مجموعه های ارزیابی و هم ترازی ساخته شده اند؛ هر جعبه، صدک های ۷۵٪، ۵۰٪ و ۲۵٪ توزیع بوت استردپ را نشان می دهد؛ ویسکرها نشان دهنده صدک های ۹۵٪ و ۹۰٪ هستند و هر نقطه ای خارج از این محدوده به عنوان یک داده پرت نمایش داده می شود. UPA در بازیابی تعادل حساسیت/اختصاصیت مورد نظر در تمام مجموعه داده های خارج از توزیع مؤثر است. داده های منبع به عنوان یک فایل داده منبع ارائه می شوند.

شکل ۲ | سناریو ۱: استقرار در یک سایت خبری - وظیفه غربالگری سینه. ستون سمت چپ: ویژگی در تابیح حساسیت قبل و بعد از هم ترازی پیش بینی. برای این تحلیل، ما یک مجموعه ارزیابی (از ۲۵۰ مورد) و یک مجموعه هم ترازی مجزا (از ۱۰۰ مورد) را از نظام موادر موجود نمونه برداری می کنیم، این نمونه برداری ۵ بار با جایگزینی تکرار می شود. حساسیت، ویژگی، ROC-AUC روی این ۵۰ نمونه بوت استردپ اندازه گیری می شوند و نتایج بر اساس میانگین نتایج روی نمونه های بوت استردپ گزارش می شوندو میله های خطی، فاصله اطمینان ۹۵٪ بوت استردپ را برای



شکل ۳ | سناریو ۱: استقرار در یک سایت جدید - وظیفه هیستوتولوژی WILDS Camelyon17. ستون سمت‌چپ: وزنی در تابع حساسیت قفل و بعد از هم ترازی. مجموعه‌های ارزیابی (شامل ۱۵۰ وصله) و یک مجموعه هم ترازی مجزا (شامل ۵۰ وصله) را از تمام وصله‌های موجود نمونه برداری می‌کنیم، این نمونه برداری ۵۰ بار با جایگزینی تکرار می‌شود. حساسیت، وزنی، ROC-AUC در ۵۰ نمونه برداری مکرر از مجموعه‌های ارزیابی و هم ترازی اندازه‌گیری شوندو نتایج بر اساس میانگین نتایج روی نمونه‌های بوت استرپ گزارش می‌شوند و میله‌های خطای فاصله اطمینان ۹۵٪ بوت استرپ را برای هر معیار نشان می‌دهند. سمت راست

درمورد رانش عملکرد پیشرونده. تحلیل حساسیت ما نشان می‌دهد که UPA فقط به مقادیر محدودی از داده‌ها از توزیع جابجا شده نیاز دارد و می‌تواند از داده‌های اعتبارسنجی موجود به عنوان مرجع استفاده مجدد کند. فقط بان ۵۰٪ مورد از حوزه‌هدف ون = با ۱۰۰٪ مورد از دامنه مبنی، نتایج تقریباً برابر و پایداری را برای کالیبراسیون مجدد خروجی‌های مدل مشاهده کردیم. این واقعیت که فقط به مقدار محدودی از داده‌ها از دامنه هدف نیاز است، برای پیاده سازی عمل UPA مهم است. UPA که به عنوان ابزار برای نظارت مداوم بر عملکرد و کالیبراسیون مجدد استفاده می‌شود، می‌تواند در یک بازه زمانی نسبتاً کوتاه از داده‌های ورودی جدید اجرا شود. در غربالگری سرطان سینه،

جدول ۳ | شاخص یودن برای سناریوی ۱، استقرار در یک سایت جدید، غربالگری پستان

مجموعه داده‌ها	قبل از اصلاح	پس از اصلاح
استرنبر	(۰..۲۱)-(۰..۴۵)	(۰..۲۹۵)-(۰..۳۹۵)
استرنC	(۰..۲۳)-(۰..۵۹۴)	(۰..۵۹۹)-(۰..۵۷۳)
استرنD	(۰..۶۴۰)-(۰..۶۲۳)	(۰..۶۴۴)-(۰..۶۲۳)

بهصورت میانگین روی ۵ نمونه بوت استرپ گزارش شده است، و انحراف معیار داخل پراانتز قرار دارد. داده‌های مبنی به صورت فاصله داده‌های منبع ارجیه شده اند.

اگر فرض نقص شود، اما شیوع هدف مشخص باشد، با امکان جمع آوری برحسب راه روی مجموعه داده های دیده نشده وجود داشته باشد، UPA همچنان من تواند با اطمینان از اینکه شیوع در مجموعه مرجع انتخاب شده با شیوع هدف مطابقت دارد (مثلاً با استفاده از تکیه های نمونه گیری مجدد) اعمال شود. با این حال، اگر هیچ اطلاعات با اطلاعات کافی در مورد علل تغییر توزیع بین مجموعه مرجع و دامنه هدف جدید در دسترس نباشد، UPA نباید بدون بررسی بیشتر برای کالیبراسیون مجدد خودکار استفاده شود. در عمل، ما پیش بینی من کنیم که UPA در کنار نظرارت جامع برداههای ورویدی، از جمله متن اطلاعات در مورد جمعیت بیمار، استفاده شود. این امر امکان تشخیص و علمت گذاری تغییرات غیرمنتظره در جمعیت بیمار را فراهم می کند. نظرارت بر داده های جمعیت پایه در حال حاضر یک روش استاندارد در برنامه های غربالگری است. این من تواند با ارزهای نظرارت خودکار، به عنوان مثال با استفاده از مدل های کمکی هوش مصنوعی، برای پیش بینی ویژگی های بیمار و داده ها از تصاویر ورویدی (مثلاً سین، تراکم سینه، کیفیت تصویری غیره) و روش هایی که به طور خاص برای تشخیص انحراف توسعه یافته اند، تکمیل شود.

تشخیص تغییرات، جنبه ای مهم از نظرارت مستمر بر عملکرد است. توجه داشته باشید که UPA ها لزوماً باید به صورت بالدرنگ اعمال شوند. به این معنی که برای استقرار پیاده سازی یک تأخیر زمانی (مثلاً چند روز) بین تشخیص و اصلاح تغییر ممکن است منطقی باشد. این امر، بازرسی انسانی در حلقه را در مورد اینکه آیا متابع دیگری ممکن است در تغییر شناسایی شده نقش داشته باشند، تسهیل می کند. در اینجا، توجه داریم که اثرات ترکیبی متابع مختلف با یا پیش بالقوه در مجموعه داده ها من تواند تجزیه و تحلیل علت ریشه ای تغییر توزیع داده ها را پیچیده کند.^{۱۹} با این حال، حتی در چنین شرایطی، UPA من تواند برای تشخیص تغییر توزیع در طول استقراره به کار گرفته شود. همانطور که در مورد تغییر شیوع در یادداشت تکمیلی^۴ نشان داده ایم.

اینکه شکل توزیع پیش بینی حفظ شده است، که ممکن است تفسیر مقدار مطلق پیش بینی را غیرممکن کند. این مهم است، زیرا نشان داده شده است که کالیبراسیون مدل لزوماً هنگام انتقال از اعتبارسنجی به حوزه های آزمایش دیده نشده قابل انتقال نیست.^{۲۰} در شکل ۱، دیدیم که چگونه شکل توزیع قبل و بعد از هم ترازی تغییر می کند و چگونه هم ترازی، شکل مرتع را حفظ می کند. هنگام مقایسه خطای کالیبراسیون مورد انتظار(ECE)^{۲۱}، ما دریافتیم که پس از هم ترازی، ECE = 0.13 بین مرتع (ECE = 0.14) و توزیع هم ترازی شده (ECE = 0.14) (ECE = 0.29) حفظ شکل توزیع پیش بینی ممکن است در کاربردهایی که از مقدار مطلق پیش بینی های مدل امنیت اولویت بندی آگاه از عدم قطعیت استفاده می کند، از اهمیت ویژه ای برخوردار باشد.^{۲۲} یا وقتی در مدل های پیش بینی ریسک استفاده می شوند.^{۲۳}

از نظر محدودیت ها، لازم به ذکر است که UPA فقط بر روی پرداختن به اثرات تغییر در اکتساب تمرکز دارد و برای مقابله با انواع دیگر تغییرات توزیع که ممکن است باعث رانش عملکرد شوند، مانند تغییر در جمعیت، شیوع پا خاشیه نویس، طراحی نشده است.

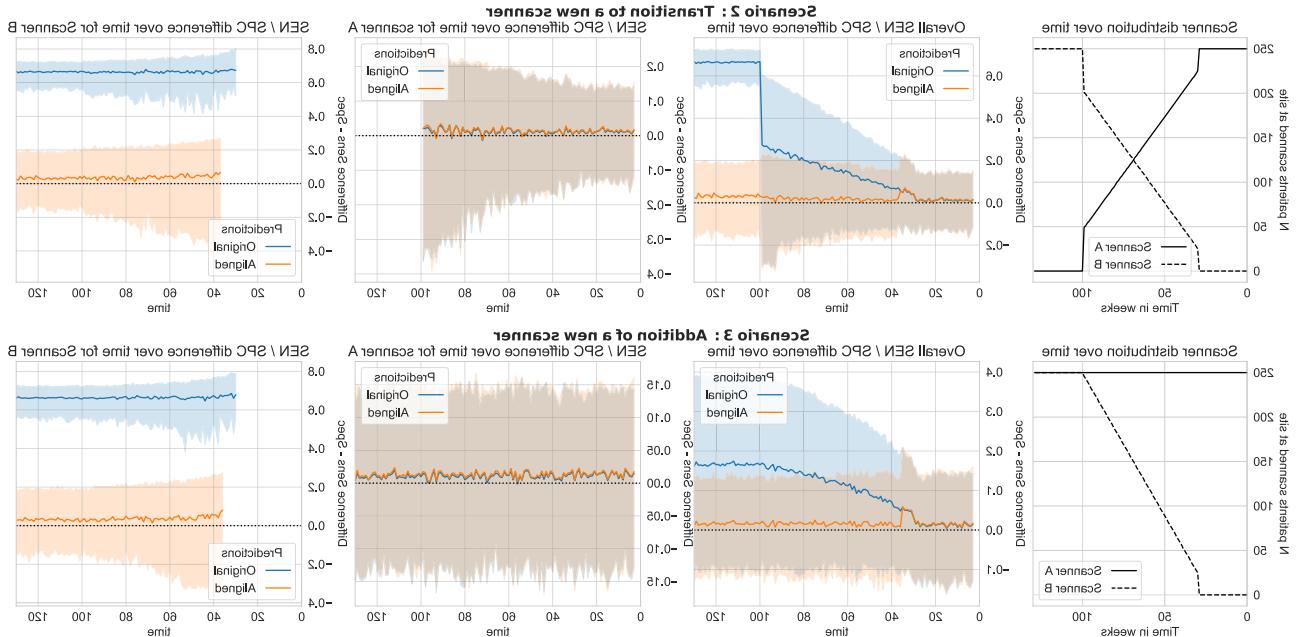
آنکه مهم این است که در UPA فرض من کنیم که شیوع (تفقیباً) بین دامنه های مرجح و هدف حفظ من شود. این یک فرض منطقی در سیاری از سناریوهای بالینی عملی است، به عنوان مثال، در برنامه های غربالگری که شیوع مورد انتظار مشخص است و معمولاً در مکان های با جمعیت های مشابه مشاهده است. توجه داشته باشید که اگر شیوع

جدول ۴ | شاخص یودن برای سناریوی ۱، استقرار در یک سایت جدید، هیستوپاتولوژی

مجموعه داده ها	قبل از اصلاح	پس از اصلاح
S4	(۰.۸۶۸-۰.۸۶۴)	(۰.۸۶۴-۰.۵۰)
S5	(۰.۸۵۹-۰.۸۶۳)	(۰.۸۶۳-۰.۶)

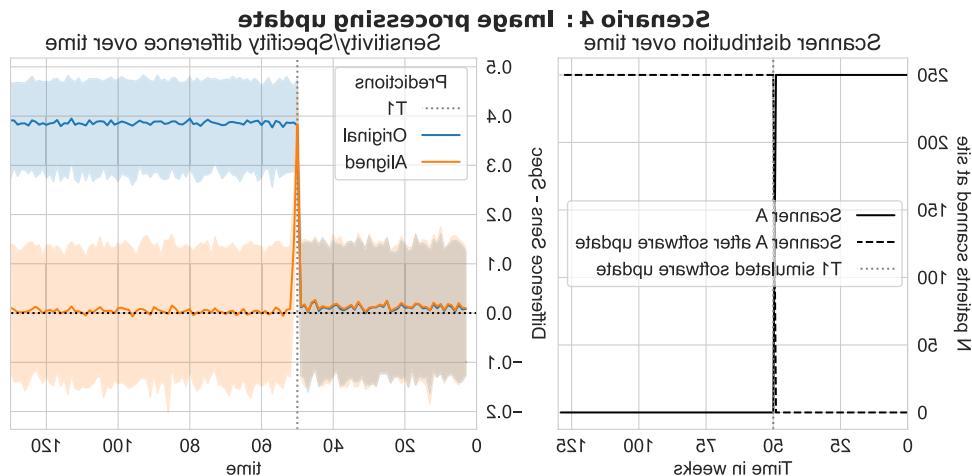
به صورت میانگین روی ۵۰ نمونه بوت استرب گزارش شده است، و انحراف معیار داخل پرانتز قرار دارد. داده های منبع به صورت فایل داده های منع ارایه شده اند.

ثانیاً باید توجه داشته باشیم که UPA برای مقابله با تغییر آستانه طراحی شده است، نه تعمیم کلی مدل. نکته مهم این است که در اینجا فرض می کنیم که تغییر اکتساب در درجه اول باعث تغییر در پیش بینی های مدل می شود که نقطه عملیاتی انتخاب شده را بی اعتبار می کند در حال که ROC-AUC را در دامنه های مختلف حفظ می کند. همه مدل های بررسی شده در این مطالعه، شرایط لازم را برآورده می کنند.



راز نظر اسکنر مقایسه می کند. هدف، جلوگیری از انحراف بین حساسیت و اختصاصیت در حضور تغییراتی در فرآیند اخذ تصویر است. روش پیشنهادی با موقیت اختلاف صفر بین حساسیت و اختصاصیت را در طول زمان حفظ می کند، در حالی که مدل تطبیق تیاقه می تواند منجر به تغییرات چشمگیر در تعادل حساسیت و اختصاصیت شود. داده های منبع به عنوان یک قابل داده ایم از اینها در SEN/SPC در تمام اسکن ها، دو سنتون سمت راست، تعادل UPA

شکل ۳ | سناریوهای ۲ و ۳: هم ترازی پیش بینی در طول زمان تحت تغییر مداوم در جمع آوری داده ها. هر شبیه سازی ۲۵ بار تکرار می شود، خطوط ممتد میانکن تفاوت بین حساسیت و اختصاصیت را در تمام نمونه های بوت استرب نشان می دهند و نواحی سایه دار نشان دهنده فاصله اطمینان بوت استرب حدود ۷۵ تا ۹۵٪ مستند. نمودارهای سنتون سمت چپ تعادل اسکن های پردازش شده توسط اسکنر A و اسکنر B را در طول زمان برای هر سناریو نشان می دهند. نمودارهای سنتون وسط، تعادل حساسیت-اختصاصیت را در طول زمان با و بدون اعمال مقایسه می کنند.



تعادل حساسیت-ویژگی در طول زمان با و بدون اعمال AUPA. هدف، جلوگیری از تغییر بین حساسیت و ویژگی در حضور تغییر تدریجی در اکتساب است. روش پیشنهادی با موفقیت اختلاف صفر SEN/SPC را در طول زمان حفظ کرد، در حالی که مدل تطبیق نیافته می‌تواند منجر به تغییرات چشمگیر در تعادل حساسیت-ویژگی شود. داده‌های منبع به عنوان یک فایل داده منبع ارائه می‌شوند.

اما باید با تنظیم کننده‌ها مذکوره و توافق شود و نظارت انسانی مناسبی را در استقرار لحاظ کند.^{۳۵} اعتبارسنجی کامل فرآیندهای مؤثر بر عملکرد یک دستگاه پزشکی، و همچنین نظارت جامع پس از ورود به بازار و پایش مددامون مدل، برای موفقیت این رویکرد در بک محیط‌فانونمند بسیار مهم است.

روش‌ها

هم ترازی پیش‌بینی بدون نظارت شامل اعمال تطبیق توزیع تجمعی تکه ای خطی بین توزیع پیش‌بینی روی مجموعه داده دیده‌نشده و توزیع پیش‌بینی مرجع است. تطبیق خط تکه ای که به عنوان «تطبیق هیستوگرام» نیز شناخته می‌شود، یک تکنیک شناخته شده در زمینه پردازش تصویر است که در آن به طور سنتی برای استانداردسازی شدت‌های تصویر اعمال می‌شود.^{۳۶} UPA در تنظیمات‌ما، این الگوریتم تطبیق بر پیش‌بینی‌های خروجی مدل اعمال می‌شود. برای برآشگاری الگوریتم تطبیق، به مجموعه‌ای از پیش‌بینی‌ها از دامنه خارج از توزیع، «مجموعه مرجع» (مجموعه مرجع) و همچنین مجموعه‌ای از پیش‌بینی‌ها از دامنه خارج از توزیع، «مجموعه هم ترازی» نیاز داریم.

الگوریتم تطبیق شامل سه مرحله ساده زیر است:

- توزیع تجمعی مشاهده شده ی پیش‌بینی‌ها را روی مجموعه‌های مرجع و هم ترازی محاسبه کنید.
- یک میان یاب خط برآش دیده تا توزیع تجمعی مجموعه هم ترازی را با توزیع تجمعی مجموعه مرجع مطابقت دهد.
- این میان یاب خط را روی هر پیش‌بینی آزمایشی جدیدی که از دامنه‌ی نادیده می‌آید، اعمال کنید.

شبیه کد الگوریتم تطبیق در یادداشت تکمیلی ۳ ارائه شده است، و یک پیاده‌سازی کاملاً کاربردی پایتون در کد تکمیلی ارائه شده است.

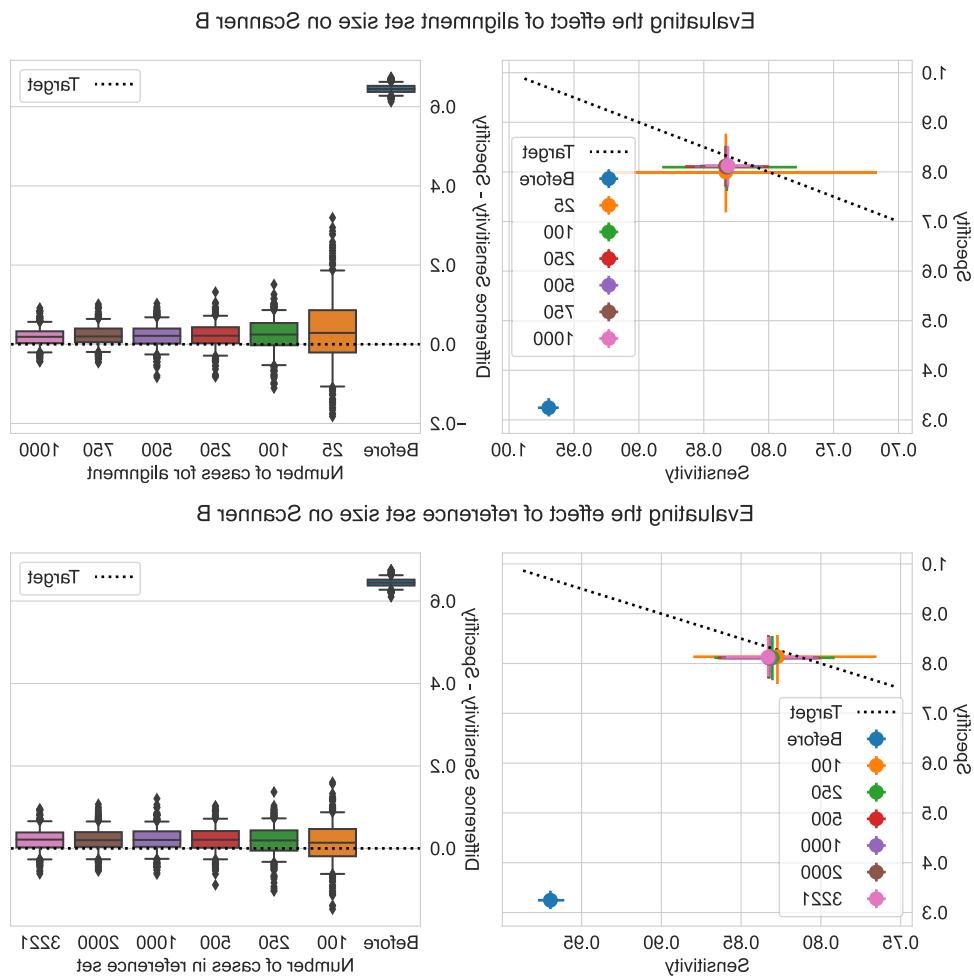
مدل هوش مصنوعی

برای اهداف این مطالعه، ما یک ResNet-50 ساده را آموزش دادیم.^{۳۷} شبکه عصبی کانولوشنی دیقطاری (با لایه مراجعه کنید) با یک زبان آنکروپی مقطعی برای هر وظیفه. برای کاربردهای سرطان سینه، قبل از پردازش تصاویر با مدل هوش مصنوعی، از ماسک کردن سینه و نرم‌مال سازی شدت استفاده می‌کیم. این مدل بر روی یک مجموعه داده غربالگری سینه‌غش شده با ۱۲۸۵ مورد (۷۲.۷٪ تصاویر مثبت) آموزش داده می‌شود. برای مدل هیستوپاتولوژی، ما بر روی یک پیش‌آموزش رسمی از مجموعه داده Camelyon17 با ۴۳۶ تصویر (۵۰٪ مثبت) آموزش می‌دهیم. پس از

شکل ۵ | سناریو^۴: هم ترازی پیش‌بینی در طول زمان در مورد به روزرسانی نرم افزار. هر شبیه سازی ۲۵ بار تکرار می‌شود، خطوط ممتد میانگین تفاوت بین حساسیت و ویژگی را در تمام نمونه‌های بوت استارتر-نشان می‌دهند و نواحی سایه دار نشان دهنده فاصله اطمینان بوت استارتر درصدی ۷۵٪ تا ۹۵٪ هستند. نمودارهای ستون سمت چپ تعداد اسکن‌های پردازش شده توسط اسکنر A و اسکنر B را در طول زمان برای هر سناریو نشان می‌دهند. نمودارهای ستون سمت راست، سیر تکاملی را مقایسه می‌کنند.

این الزام را به خوبی تعمیم می‌دهند و از نظر ROC-AUC به خوبی تعمیم می‌دهند، یعنی توانایی آنها در جداسازی کلاس‌ها حتی زمانی که ویژگی‌های تصویر ورودی تغییر می‌کند، تاحد زیادی حفظ می‌شود. همانطور که درون یاب خطی تکه ای را به خروجی‌های مدل اعمال می‌کیم، UPA تضمین می‌کند که ROC-AUC قابل و بعد از ترازبندی یکسان باشد (Zirar et al., ۲۰۲۱). بازیابی تعادل SEN/SPC مورد انتظار، تا زمانی که ROC-AUC در دامنه‌ها حفظ شود، به عملکرد طبقه‌بندی وابسته نیست. اگر ROC-AUC یک مدل از یک دامنه به دامنه به دیگر به طور قابل توجه کاهش یابد، انتظار می‌رود UPA عملکرد غیربهینه داشته باشد و قادر به بازیابی عملکرد ROC-AUC در داده‌های نشانه نخواهد بود. رسیدگی به مسائل مربوطه افت عملکرد از نظر جداسازی کلی کلاس، یعنی اندازه گیری شده با افت AUC-ROC. تمرکز بخش بزرگی از کارها در تطبیق دامنه (DA) است.^{۳۸} و تعمیم دامنه (DG) (Zirar et al., ۲۰۲۱). این خطوط کاری مکمل یکدیگر هستند، زیرا روش‌های DA/DG بر ساخت مدل‌های DA/DG ROC-AUC تمرکز بر اساس ROC-AUC در دامنه‌ها معمولاً توجه کمی به تغییر آستانه دارند. در اینجا، UPA با تمرکز بر تعمیم پذیری کالیبراسیون مدل و نقاط عملیاتی بالینی از پیش تعریف شده‌اند، به یک نیاز برآورده نشده می‌پردازد. روش پیشنهادی به هیچ وجه به عنوان جایگزینی برای تکنیک‌های سنتی DA/DG در نظر گرفته شده است، بلکه باید علاوه‌بر آن در نظر گرفته شود. در حالی که استفاده از روش‌های DA/DG در طول توسعه مدل ممکن است منجر به افزایش مقاومت در برابر تغییر اکتساب شود، هیچ تضمینی برای عملکرد در حوزه‌های جدید دیده نشده، به ویژه از نظر معیارهای بالینی، وجود ندارد. در اینجا، UPA یک ضمانت مهم برای شناسایی رانش عملکرد در استقرارهای جدید اضافه می‌کند. علاوه بر این، ما معتقدیم که اعتبارسنجی محلی بر روی داده‌های نماینده، ممیزی‌های منظم هوش مصنوعی^{۳۹} و تجزیه و تحلیل دقیق موارد ناسازگار، ممچنان اجزای حیاتی برای استقرار بالینی اینم و اطمینان از این بودن هوش مصنوعی در طول زمان هستند.

علاوه بر این، در مواردی که UPA را در یک مورد استفاده از تصویربرداری پزشکی پیشنهادیم کنیم، باید مقررات جهانی تجهیزات پزشکی در نظر گرفته شود. کنترل دقیق یک دستگاه پزشکی تایید شده برای اطمینان از اینمی و اثربخشی مدارم پس از عرضه به بازار، به ویژه هنگام در نظر گرفتن تغییر بدون نظارت پارامترها، ضروری است. به این ترتیب، مادعا نمی‌کیم که اعمال UPA کاربران را از رویه‌های استاندارد حسابرسی شامل مفایضه با ارزیابی انسانی بی نیاز می‌کند. با این حال، ادغام آن در گردش کار می‌تواند امکان اصلاح خودکار را قبل از دور بعدی حسابرسی فراهم کند، زیرا به روزرسانی‌های جمجم آوری داده‌ها ممکن است در هر زمانی اتفاق بیفتد. نرم افزارهای مدرن هوش مصنوعی به عنوان یک دستگاه پزشکی، شامل گزینه «تغییر پروتکل‌ها» هستندتا به دستگاه‌های هوش مصنوعی اجازه دهنده با توجه به معیارهای بالینی مرتبط، به مرور زمان بهبود یابند.



هرکار، صد کهای ۷۵٪ و ۵٪ توزیع بوت استرب را نشان می‌دهد؛ ویسکرها نشان دهنده صد کهای ۹۵٪ و ۹٪ هستند و هر نقطه ای خارج از این محدوده به عنوان یک داده پرت نمایش داده می‌شود. این تحلیل نشان می‌دهد که با تعداد کم حدود ۲۵۰ مورد در مجموعه هم ترازی، ما در حال حاضر نتایج بسیار خوبی به دست می‌آوریم که در آزمایش‌های نمونه برداری مکرر پایدار است. در تحلیل زیر، عملکرد روش‌ها را برای اندازه‌های مختلف مجموعه مرجع اندازه‌گیری کردیم، از ۵۰ مورد برای هم ترازی استفاده کردیم و اندازه مجموعه مرجع را تغییر دادیم. من توانیم ببینیم که این روش به اندازه مرجع خیلی حساس نیست، حتی اگر توزیع مرجع شامل داده‌های بیشتری باشد. نتایج بهتری حاصل می‌شود. داده‌های منبع به عنوان یک فایل داده منع ارائه می‌شوند.

شکل 6 | تحلیل حساسیت روی تأثیر اندازه های مجموعه های هم ترازی و مرعج بر عملکرد روش هم ترازی(A) و تحلیل حساسیت روی اندازه های مجموعه هم ترازی (B)، از مجموعه مرعج کامل (۳۲۲۱ مورد) استفاده کردند. نتایج روی ۵۰ نمونه بوت استرب گزارش شده اند. برای تحلیل اندازه هم ترازی، هر نمونه بوت استرب با نمونه برداری از یک مجموعه هم ترازی با اندازه مورد نظر از تمام مواد موجود و همچنین یک مجموعه ارزیابی ایجاد می شود (n= ۲۵۰ مورد). برای تحلیل اندازه مرعج، هر نمونه بوت استرب با نمونه گیری از یک مجموعه مرعج با اندازه مورد نظر از تمام مواد موجود و همچنین یک مجموعه ارزیابی ایجاد می شود (n= ۲۵۰ مورد). در سمت چپ، نقاط میانگین SPC/SEN را روی نمونه ها نشان می دهند و میله های خط نشان دهنده فاصله اطمینان بوت استرب ۹۵٪ هستند. در سمت راست،

عملکرد مدل روی مجموعه ارزیابی پس از هم ترازی برای سنجش اثربخشی روش پیشنهادی در کاهش رانش حساسیت-ویژگی. نمونه برداری از مجموعه هم ترازی و ارزیابی پایار تکرار می شود تا تخمین های عدم قطعیت برای معیارهای گزارش شده به دست آید. توجه داشته باشید که در مجموعه داده های غربالگری سینه، برای هر مورد، ۴ تصویر در رله مطالعه داریم.

جزئیات مربوط به «سناریوهای ۲ تا ۴: به روزرسانی های مداوم مدل» دراین سناریوها، ما من خواهیم اثربخشی روش خود را برای به روزرسانی های مداوم مدل رحضور تغییر تدریجی زمانی برای کار ماموگرافی بررسی کنیم. برای این منظور، ما یک جریان داده را با تعریف یک تابع نگاشت زمان به یک توزیع اسکنتر معین که هر یک از سناریوهای فوق، را توصیف می کند، شبیه سازی می کنیم. برای بدست آوردن مجموعه مراجع و مجموعه های شبیه سازی برای اسکنتر A، داده های اصل را تقسیم می کنیم. دی مرجع به دو زیرمجموعه مجزا تقسیم می شود: یک زیرمجموعه تصادفی نمونه گیری شده از 1200 مورد به عنوان مجموعه مرجع (برای تعیین آستانه طبقه بندی شروع و به عنوان هدف هم ترازی در طول شبیه سازی) استفاده می شود؛ سپس بخش باقیمانده شامل 2221 مورد به عنوان منبع داده برای خود شبیه سازی استفاده می شود. در اینجا فرض می کنیم که زمان

درآموزش، آستانه طبقه بنده را در نقطه عملیاتی متعادل (که در آن حساسیت برابر با ویزگ است) روی مجموعه انتبارستنجی درون توزیعی انتخاب می کنیم. بدین مرحله، که ما آن را با مجموعه مرچع نشان می دهیم (به حداوی مراجعه کنید) (۱) و (۲). توجه داشته باشید که مدل های آموزش دیده از نظر ROC-AUC به خوبی به تنظیمات خارج از توزیع خود تعمیم داده هم شوند، اما از نظر حساسیت/ اختصاصیت (همانطور که در بخش نتایج نشان داده شده است) این طور نیستند.

جزئیات مربوط به «سناریو ۱: استقرار در یک سایت جدید» مالز مدل هوش مصنوعی به همراه یک آستانه طبقه بندی از پیش تعیین شده که روی ... بهینه سازی شده است، استفاده از کنیم.دی مرجع.ما ابتدا عملکرد این مدل را هنگام استقرار در یک سایت با نمونه برداری از یک مجموعه ارزیابی شامل ۲۵۰ مورد از مجموعه داده های خارج از توزیع (به طور معمول ۱۵۰ و صله برای کاربرد هیستوپاتولوژی) اندازه گیری می کنیم و عملکرد مدل را از نظر حساسیت و اختصاصیت روی این مجموعه ارزیابی می کنیم. علاوه بر این، ما یک مجموعه مجزا از $N = 5$ مورد (به طور معمول ۵ و صله) را برای تطبیق با الگوریتم هم ترازی پیش بین نمونه برداری از کنیم. برای کاربرد ماموگرافی، از مجموعه مرجع کامل استفاده می کنیم.دی مرجع به عنوان توزیع مرجع ما (3221 مورد). برای کار هیستوپاتولوژی، 5000 تصویر از دی مرجعاً به عنوان توزیع مرجع خود استفاده کنیم. سپس دوباره ارزیابی می کنیم

واحد در این شبیه سازی ها بر حسب هفتة است. برای هر نقطه زمانی T ، داده ها را طبق توزیع اسکنتر مورد انتظار در نقطه T - همانطور که توسط هر سناریو تعریف شده است - نمونه برداری می کنیم. در هر نقطه زمانی معنی T ، از یک پنجه اجرایی مشکل از داده های دور از $T-1$ و $T-2$ برای برازش هر دو الگوریتم هم ترازی (به طور جداگانه برای هر اسکنتر) استفاده می کنیم. سپس عملکرد مدل را قبل و بعد از هم ترازی روی مجموعه نمونه برداری شده برای زمان T ، هم از نظر اسکنتر و هم به طور کل، ارزیابی می کنیم. علاوه بر این، اطمینان حاصل می کنیم که - در هر نقطه زمانی معنی T ، مجموعه پنجه اجرایی (نمونه های $T-1$ و $T-2$) برای مجموعه ارزیابی (نمونه های T) از هم جدا هستند. در سناریوهای 2 و 4 فرض می کنیم که تعداد کل موارد در هر نقطه زمانی (هفتة) در $= 25$ مورد. در سناریوی 3 ، ما با $= 5$ مورد در ابتدای شبیه سازی و به تدریج تعداد موارد غربالگری شده با افزایش اسکنتر دوم افزایش می یابد تا به ... برسیم. در سناریوی 4 ، ما به روزرسانی نرم افزار را با افزایش وضوح تصاویر و رویدی شبیه سازی کردیم. در این سناریو نیز از تصاویر اسکنتر A استفاده کردیم.

منابع

۱۰. کیم، سی کی و همکاران. یک خط لوله تریاژ خودکار کووید-۱۹ با استفاده از هوش مصنوعی بر اساس رادیوگرافی قفسه سینه و داده های بالینی. *Magazine of Biomedical Engineering* ۵، ۵، N.P.J. (۲۰۲۲).

۱۱. لبیگ، سی. و همکاران. ترکیب نقاط قوت رادیولوژیست ها و هوش مصنوعی برای غربالگری سرطان سینه: یک تحلیل گذشته نگر، لنسنست دیجیت. *Salamat* ۴، e507-e519 (۲۰۲۲).

۱۲. مک کین، اس ام و همکاران. ارزیابی بین المللی یک سیستم هوش مصنوعی برای غربالگری سرطان سینه. *طبیعت* ۵۷، ۸۹-۹۴ (۲۰۲۰).

۱۳. لارسن، م.، آگلن، سی اف، هاف، اس آر، لوند-هانسن، اچ. و هافویند، اس. استراتژی های ممکن برای استفاده از هوش مصنوعی در خواندن ماموگرافی ها بر اساس داده های گذشته نگر از آزمایش غربالگری پورا رادیول. *Journal of Radiology* ۲۲، ۸۲۴۶-۸۲۳۸.۳۳ (۲۰۲۲).

۱۴. پریس، سی. و همکاران. تأثیر سیستم های مختلف ماموگرافی بر عملکرد هوش مصنوعی در غربالگری سرطان سینه. *Radiology* ۳۵، ۵۰-۵۶ (۲۰۲۳).

۱۵. فینلیسون، اس. جی. و همکاران. تغییر پیشک و مجموعه داده ها در هوش مصنوعی. *Journal of Medical Imaging* 18، ۳۸۵-۳۸۶ (۲۰۲۱).

۱۶. Lundstrom, C. IEEE Biomed. Health Inf 7. Stacke, K., Eilertsen, G., Unger, J. & Zivkovic, B. *IEEE Transactions on Biomedical Engineering* 60, ۳۳۴۶-۳۳۵۸.۲۵ (۲۰۲۱).

۱۷. زو، کلیو، ز.، کیایی، ای. و شیانگ، ت. و لوی، CC. تعمیم دامنه: یک بررسی. <https://arxiv.org/abs/2103.02503> (۲۰۲۱).

۱۸. گوش، دی. و اکتا گوش، دی. یک تحقیق چند مرکزی در مقیاس بزرگ در مورد تعمیم دامنه در تشخیص توده مبتنی بر یادگیری عمیق در ماموگرافی: یک بررسی. *Dr. Journal of Diagnostic Radiology* ۹-۱۰. زیست شناسی آکتا-۱۰. (۲۰۲۲).

۱۹. بربن، جی.، زوکر، ک.، اورسی، ان ام و راویکومار، ن. ثبت تصاویر زیست پیشک، تعمیم دامنه و تحلیل خارج از توزع ۱۴-۳۳ (انتشارات بین المللی اشپرینگر، ۲۰۲۲).

۲۰. جنکیز، دی. ای و همکاران. به روزرسانی و نظارت مداوم بر مدل های پیش بینی بالینی: آیا زمان سیستم های پیش بینی پویا فرا رسیده است؟ تشخیص پیش بینی. *Journal of Clinical Medicine* 10, ۱-۱۵ (۲۰۲۱).

۲۱. بینویسا، ام ای ای، انگلهارت، ای جن، شاتس، دبلیو، اشمیت، ام کی و استایربرگ، ای دبلیو. راهنمای روش شناختی برای ارزیابی و به روزرسانی مدل های پیش بینی بالینی: یک بررسی سیستماتیک. *Journal of Diagnostic Radiology* ۳۳۶، ۲۲. *BMJ* (۲۰۲۲).

۲۲. مک کورمیک، تی اچ، رفتري، ای ای، مادیگان، دی. و برد، آر اس. رگرسیون لجستیک پویا میانگین گیری مدل پویا برای طبقه بنده دودوبی. *Biometrika* ۳۰-۳۳، ۶۸ (۲۰۲۲).

۲۳. هیک، جی ای و همکاران. رویکردهای مدل سازی پیش بینی پویا برای جراحی قلب. *Journal of Thoracic and Cardiovascular Surgery* ۱۳۳, ۶۴۹-۶۵۸ (۲۰۲۲).

۲۴. هرگوو، واي. و همکاران. یک رویه آزمایش بسته برای انتخاب روش مناسب برای به روزرسانی مدل های پیش بینی آمار، پیشک. *Journal of Statistical Computation and Simulation* ۱۵, ۴۵۳۹-۴۵۲۹ (۲۰۱۷).

۲۵. کالسترن، ب. و همکاران. اعتبارسنجی و به روزرسانی مدل های ریسک مبتنی بر رگرسیون لجستیک چندجمله ای. *Journal of Clinical Medicine* ۱۷, ۲۱ (۲۰۲۰).

۲۶. آشتلينگک، ای ام، بانگ، دبلیو. و کیمل، اس ای. مقایسه استراتژی های بر روزرسانی پویابرای مدل های پیش بینی بالینی. *Journal of Clinical Medicine* ۲۰, ۵ (۲۰۲۱).

۲۷. لوى، تى جي و همکاران. توسعه و اعتبارسنجی مدل های پیش آگهی خودپایشی و به روزرسانی خودکار بقا برای بیماران بستری شده مبتلا به کووید-۱۹. *Journal of Clinical Medicine* ۱۱, ۶۸۱۳ (۲۰۲۲).

۲۸. رابانسر، س.، گونمن، س. و لیپتون، ز. شکست با صدای بلند: مطالعه تجزیی روش های تشخیص تغییر مجموعه داده ها. *Journal of Clinical Medicine* ۱۹, ۳۹۶-۳۹۸ (۲۰۲۰).

۲۹. رابانسر، س.، گونمن، س. و لیپتون، ز. شکست با صدای بلند: مطالعه تجزیی روش های تشخیص تغییر مجموعه داده ها. *Journal of Clinical Medicine* ۱۹, ۳۹۶-۳۹۸ (۲۰۲۰).

اطلاعات اخلاقی

مجموعه داده های داخلی سرطان سینه از بریتانیا و مجارستان قابلً تأیید اخلاقی از سازمان تحقیقات بهداشت خدمات ملی بهداشت بریتانیا (NHS) (HRA) (مرجع: HRA / 0376/19) و شورای تحقیقات پزشکی، کمیته اخلاقی علمی و پژوهشی در مجارستان (ETT-TUKEB) (مرجع: OGYÉI/46651-4/2020) (مرجع: ETT-TUKEB) (مرجع: این مطالعه اصلی مطابق با اصول مندرج در اعلامیه هلسینکی برای همه تحقیقات تحریب انسانی انجام شد. لزوم رضایت آگاهانه برای شرکت در مطالعه توسط HRA و TUKEB- ارزیابی شد که نیازی به آن نیست زیرا این مطالعه شامل استفاده دانویه از ETT بررسی و تأیید شد که داده های این مطالعه شامل ناشناس این داده های گذشته نگر و با نام مستعار بود. مطالعه حاضر از نسخه کامل‌ناشناس این داده های جمع آوری شده قبلي استفاده دانویه کرد. برای مجموعه داده های سرطان سینه در ۱۹۰۰ دوره ۵ ساله OPTIMAM در زوئیه ۲۰۱۹ تأییدیه HRA و نظر اخلاقی مساعد را برایک دوره ۵ ساله قابل تمدید (مرجع: SC/0284/19) برای جمع آوری تصاویر و داده ها از سایت های شرکت کننده جهت ایجاد یک پایگاه داده تحقیقاتی و افزوندن سایت های جمع آوری جدید، دریافت کرد. رضایت بیمار توسط کمیته اخلاق تحقیقاتی Oxford C-

خلاصه گزارش

اطلاعات پیشتر در مورد طراحی تحقیق در خلاصه گزارش نمونه کارهای طبیعت که به این مقاولانیک شده است، محمد است.

دردسترس بودن داده ها

تمام خروجی های مدل ناشناس که از یافته های شرح داده شده در این نسخه خطی پشتیبانی می کنند، به صورت عمومی در مخزن کد ما به آدرس در دسترس هستند.
<https://github.com/biomedia-mira>

داده های مذکور در این مقاله ارایه شده اند که شامل داده های رسم نمودارهای تشان داده شده در شکل ها و همچنین جداول است. دسترسی به مجموعه داده های سرطان سینه OPTIMAM را می توان در وب سایت پژوهه درخواست کرد:
omidb.org

مشارکت های نویسنده
 نسخه اولیه را نوشتند. همه نویسنده‌گان نسخه اولیه را ویرایش و بررسی کردند و نسخه نهایی را تأیید کردند BG و GK. R. نظارت نظارتی را بر عهده داشت AH. پشتیبانی اداری و نظارت بر پروژه را از اینه دارند PK و TR. داده ها را جمع آوری و تخصص پژوهشی ارائه دادند و لیل NS. نتایج را تفسیری داده های اساسی را تأیید کردند TR و BG، JY، MR. GK. جمع آوری داده ها و پیش پردازش را انجام دادند JY و MR. GK. الگوریتم ها را پیاده سازی کرد. آزمایش ها را انجام داد و تجزیه و تحلیل آماری را انجام داد MR. این مطالعه را مفهوم سازی و طراحی کردند BG و GK.

منافع رقابتی

چن کن، چن واي، اي اچ، چن کن، تي آر، بى جى كارمندان شركت فناوري هاي پژوهش خيرون هستندكه حق خريد سهام به عنوان بخشی از بسته استاندارد جبران خدمات آنها در نظر گرفته شده است. ام ار زمان انجام اين کار، كارآموز حقوق بگير خيرون بود. ساير نویسنده‌گان همچ گونه تضاد منافع را اعلام نمی کنند.

اطلاعات تکمیلی

اطلاعات تکمیلی نسخه آنلاین شامل مطالعه تکمیلی است که در آدرس زیر موجود است.
<https://doi.org/10.1038/s41467-023-42396-y>

مکاتبات و درخواست های مربوط به مطالعه باید به ملانی روشنویزی بن گلاکر ارسال شود.

اطلاعات بررسی همتا روابط طبیعت از آن تاکر و دیگر داور (داوران) ناشناس برای مشارکت‌شان در داوری همتا به همتا این اثر تشکر می‌کنم. فایل داوری همتا موجود است.

اطلاعات مربوط به چاپ مجدد و مجوزهای موجود است در <http://www.nature.com/reprints>

پادا شناس اشپرینگر نیجر در مورد ادعاهای مربوط به صلاحیت قضایی در نقشه های منتشر شده و واپسی های سازمانی بن طرف باقی ماند.

Creative Commons Attribution 4.0 دسترسی آزاد این مقاله تحت مجوز بین المللی منتشر شده است که استفاده، اشتراک گذاری، اقتباس، توزیع و تکثیر در هر رسانه یا قالبی را مجاز می‌داند، مادرام که به نویسنده‌گان) اصلی و منبع، اعتبار کافی داده شود، پیوندی به مجوز Creative Commons ارائه شود و در صورت ایجاد تغییرات، مشخص شود که آیا تغییرات ایجاد شده است یا خیر. تصاویر یا سایر مطالعه شخص ثالث در این مقاله در مجوز Creative Commons مقاله گنجانده شده اند، مگر اینکه در خط اعتباری مطلب، خلاف آن ذکر شده باشد. اگر مطالعه در مجوز Creative Commons مقاله گنجانده شده باشد و استفاده مورد نظر شما طبق مقررات قانونی مجاز باشد یا از حد مجاز تجاوز کند، باید مستقیماً از دارنده حق چاپ اجازه بگیرید. برای مشاهده نسخه ای از این مجوز، به <http://creativecommons.org/licenses/by/4.0/>

© نویسنده‌گان (۲۰۲۳)

۲.۱. کو، پی دبلیو و همکاران. WILDS: معیاری برای تغییرات توزیع در طبیعت. در مجموعه مقالات سی و هشتین کنفرانس بین المللی یادگیری ماشین (ویراسته، میلا، ام. و ژانگ، تی). جلد ۱۳۹، صفحات ۵۶۳۷-۵۶۴۴ (PMLR، ۲۰۲۱).

۲.۲. دکتر هالینگ-براون و همکاران. پایگاه داده تصاویر ماموگرافی OPTIMAM: معنی در مقیاس بزرگ از تصاویر ماموگرافی و داده های بالینی. رادیو، مصنوعی. هوشمند. ۳، e200103 (2021).

۲.۳. باندی، پ. و همکاران. از تشخیص متابازهای منفرد تا طبقه بندی وضعیت غدد لنفاوی در سطح بیمار: چالش CAMELYON17 تصویربرداری پژوهشی ترانس IEEE (۲۰۱۹) ۵۶-۵۵، ۳۸-۳۲.

۲.۴. چن، جی جی و همکاران. تنظیم آستانه تصمیم گیری در پیش بینی کلاس. محیط زیست SAR. تحقیقات. ۱۷، ۳۳۷-۳۵۲ (۲۰۰۶).

۲.۵. اوادیا، واي. و همکاران. آيا من توانيد به عدم قطعیت مدل خود اعتماد کنید؟ ارزیابی عدم قطعیت پیش بینی کننده تحت تغییر مجموعه داده ها. /abs/1906.02530 (۲۰۱۹) <https://arxiv.org>

۲.۶. گوئی، س.، پلیس، جن.، سان، واي. و واینبرگر، کن. کیو. درباره کالیبراسیون شبکه های عصبی مدرن. در مجموعه مقالات سی و چهارمین کنفرانس بین المللی یادگیری ماشین (ویراسته، پریکاپ، دی. و ته، وای. دبلیو). جلد ۷۰، صفحات ۱۳۲۱-۱۳۳۰ (PMLR، آگوست 2017).

۲.۷. ون کالستر، بن. و ویکرز، ای. جن. کالیبراسیون مدل های پیش بینی ریسک: تأثیر بر عملکرد تحلیلی تصمیم گیری پژوهشی. تصمیم، ساخت. ۳۵، ۱۶۹-۱۶۳ (۲۰۱۵).

۲.۸. کاسترو، دی س، واکر، آی. و گلاکر، بن. اهمیت علیت در تصویربرداری پژوهشی. نات. کمون. ۱۱، ۳۶۷۳-۳۶۷۸ (۲۰۲۰).

۲.۹. برنهارت، م.، جونز، س. و گلاکر، ب. منابع بالقوه سوگیری مجموعه داده ها، بررسی تشخیص کمتر از حد لازم توسط الگوریتم های یادگیری ماشین را پیچیده می کند. نات. مد. ۲۸، ۱۱۵۸-۱۱۵۷ (۱۱۵۸-۱۱۵۷) (۲۰۲۲).

۲.۱۰. بن-دیوید، س. و همکاران. نظریه ای در مورد یادگیری از حوزه های مختلف. ماخ. پادگیر. ۷۹، ۱۵۱-۱۵۷ (۲۰۲۱).

۲.۱۱. Venkateswara, H. & Panchanathan, S. تطبیق دامنه در بینایی کامپیوترا با یادگیری عمیق. (اشرپینگر نیجر، ۲۰۲۰).

۲.۱۲. لی، ای.، پن، اس. جنی، وانگ، اس. و کوت، تعیین دامنه AC با یادگیری ویزگی تخاصمی. در مجموعه مقالات کنفرانس IEEE در مورد بینایی کامپیوترا و تشخیص الگو. (IEEE، ۲۰۱۸) ۵۴-۵۴۰.

۲.۱۳. Glockner, B. تعیین دامنه از طریق یادگیری مدل-آنکوستیک ویژگی های معنایی. ۳۳. Dou, Q., Coelho de Castro, D., Kamnitsas, K. & فرآیند اطلاعات عصبی پیشرفته. سیستم. ۳۲، ۶۴۶۱-۶۴۵۰ (۲۰۱۹).

۲.۱۴. لیو، ایکس. و همکاران. ممیزی الگوریتمی پژوهشی. سلامت لنت دیجیت. ۴، e384-e397 (2022).

۲.۱۵. ماهلر، م.، آئورزا، س.، آلبسا، ر.، ملوس، س. و وو، ج. ای. جنبه های نظریاتی هوش مصنوعی و نرم افزارهای مبتنی بر یادگیری ماشین به عنوان دستگاه های پژوهشی (SaMD) دقیق. پژوهشی. مصنوعی. هوشمندانه. ۲۳۷-۲۶۵ (۲۰۲۱).

۲.۱۶. گونزالس، آر سی و وینتر، پ. پردازش تصویر دیجیتال (ویرایش دوم). (شرکت انتشاراتی آدیسون-وسلی لانگمن، ۱۹۸۷).

۲.۱۷. کی، ژانگ ایکس، رن اس، سان جن. یادگیری عمیق باقیمانده برای تشخیص تصویربرداری مجموعه مقالات کنفرانس IEEE در مورد بینایی کامپیوترا و تشخیص الگو. (IEEE، ۲۰۱۶) ۷۷-۷۷۸.

تقدیرنامه ها

بن جنی از حمایت اکادمی سلطنتی مهندسی به عنوان بخشی از کرسی تحقیقاتی خود در زمینه فناوری های پژوهش خیرون/RAEng در زمینه استقرار ایمن هوش مصنوعی تصویربرداری پژوهشی سپاسگزار است.