22 th International Conference on
Information Technology, Computer and Telecommunication
Event Place: Tbilisi,Georgia
www.itctconf.ir
بیست و دومین کنفرانس بین المللی
فناوری اطلاعات، کامپیوتر و مخابرات | گرجستان
22th International Conference on Information Technology,Computer and Telecommunication
۲۶ اسفندماه ۱۴۰۲ ——— مجلات معتبر بین المللی

بررسی کارایی شبکه های خصمانه مولد بهبود یافته و یادگیری تقویتی جهت افزایش دقت در شرح نویسی تصاویر پزشکی با سبک بالینی

سجاد زرینی
دانشجوی کارشناسی ارشد،رشته کامپیوتر گرایش هوش مصنوعی و طراحی رباتیکز،دانشگاه آزاد اسلامی واحد تهران مرکز
پروانه اصغری
دکتری ،رشته کامپیوترگرایش نرم افزار،استادیار،دانشکده فنی مهندسی،دانشگاه آزاداسلامی واحد تهران مرکز

چکیده
تجزیه و تحلیل تصاویر پزشکی نقش مهمی در تشخیص و درمان بیماری های مختلف ایفا می کند. با این حال، فرآیند تجزیه و تحلیل این تصاویر و تولید گزارش های دقیق می تواند زمان بر و در معرض خطای انسانی باشد. در این مطالعه، یک رویکرد جدید پیشنهاد می‌شود که شبکه‌های مولد متخاصم و یادگیری تقویتی را برای بهبود کارایی تجزیه و تحلیل تصاویر پزشکی و نوشتن گزارش‌های مربوطه ترکیب می‌کند. این روش، شامل آموزش یک شبکه مولد متخاصم برای تولید تصاویر پزشکی واقعی بر اساس مجموعه معینی از پارامترهای ورودی است. سپس این شبکه با یک چارچوب یادگیری تقویتی برای یادگیری و بهینه سازی فرآیند تجزیه و تحلیل این تصاویر ادغام می شود. از طریق آموزش تکراری، این سیستم قادر به تولید تصاویر پزشکی با کیفیت بالا است که به طور دقیق شرایط مختلف پاتولوژیک را نشان می دهد. با استفاده از قابلیت‌های قدرتمند شبکه‌های متخاصم و یادگیری تقویتی، سیستم ما می‌تواند به طور موثر ویژگی‌ها و ناهنجاری‌های مهم در تصاویر پزشکی را ثبت کند که منجر به گزارش‌های دقیق‌تر و جامع‌تر می‌شود. ارزیابی نتایج کارایی روش پیشنهادی را با صحت ۹۲ درصد و دقت ۹۳ درصد در مقابل سایر روش ها نشان داد.

کلمات کلیدی: شبکه های مولد متخاصم، تحلیل تصویر پزشکی، تولید گزارش، یادگیری تقویتی، کارایی

مقدمه

با پیشرفت مستمر تکنولوژی مدرن پزشکی و توجه روزافزون به سلامت جسمانی، افراد در شرایطی که بدنشان دارای شرایط غیرطبیعی باشد برای تشخیص و درمان بیماری به بیمارستان مراجعه می‌کنند. در این زمان، رادیولوژیست‌ها اغلب نیاز به استفاده از تکنیک های مختلف تصویربرداری پزشکی برای بررسی بدن بیمار دارند تا تصاویر پزشکی[1] مرتبط را به دست آورند و گزارش‌های تشخیص پزشکی ارائه دهند (Ahmad et al., 2022). در محیط پزشکی واقعی، مشکلات مختلفی برای گزارش تشخیص پزشکی وجود دارد که از جمله آنها می توان به این موارد اشاره کرد: ۱) حجم کاری زیاد و راندمان کاری پایین رادیولوژیست ها، در بیمارستان های کوچک تجهیزات و رادیولوژیست کمتری وجود دارد و بیمارستان‌های بزرگ بیش از حد شلوغ هستند، در نتیجه رادیولوژیست‌ها نیاز به خواندن تعداد زیادی تصاویر پزشکی و ارائه گزارش‌های تشخیصی دارند (Ayesha et al., 2021). ۲) کیفیت پایین گزارش‌های تشخیصی، رادیولوژیست ها در بیمارستان های کوچک بی تجربه هستند. در بیمارستان‌های بزرگ، اگرچه رادیولوژیست‌ها دارای تجربه غنی هستند، اما باید کار را در زمان محدودی تکمیل کنند، که ممکن است باعث ارائه گزارش‌های اشتباه شود. برای این منظور، به منظور کاهش حجم کار و بهبود کارایی و کیفیت، می‌توان از علم و فناوری روز برای دستیابی به خواندن خودکار و شرح نویسی[2] تصاویر پزشکی استفاده کرد (Agrawal et al., 2019).

در حال حاضر تصاویر پزشکی با تصاویر طبیعی متفاوت است و کسب آنها با نیاز به نیروی انسانی فراوان، منابع مالی و مادی دشوار است و وضوح آنها کم است، مرزهای بین اندام ها و بافت های مختلف در تصویر محو شده است (Agrawal et al., 2019؛ Allaouzi et al., 2018). علاوه بر این، مجموعه داده های موجود با تصاویر و گزارش های پزشکی استاندارد اندک است. بنابراین تحلیل و تفسیر چنین تصاویر پزشکی با کیفیت بالا بسیار دشوار است. الگوریتم‌های یادگیری ماشین[3] یا یادگیری عمیق[4] موجود عمدتاً از روش‌های سنتی رمزگذاری-رمزگشایی[5] هنگام تولید متن گسسته استفاده می‌کنند و عمدتاً در تصاویر طبیعی و کمتر در تصاویر پزشکی استفاده می شود (Alsharid et al., 2020؛ Djmila et al., 2021). دقت تولید شده گزارش های الگوریتم های سنتی کمتر است. با توسعه یادگیری خصمانه، برخی از شبکه های متخاصم مولد[6] و یادگیری تقویتی[7] برای تولید توصیفات تصاویر طبیعی استفاده می کنند. هنگام تولید گزارش‌ها، ممکن است با مشکلاتی مانند خاص بودن تصاویر پزشکی و مشکلات سوگیری نوردهی مواجه شویم که ممکن است منجر به کیفیت و تنوع پایین‌تر شود (Djmila et al., 2021؛ Beddiar et al., 2022). از این رو در این پژوهش، مدل مولد و متمایز به طور متناوب با روش‌های آموزش خصمانه بر اساس شبکه متخاصم مولد مشروط و یادگیری تقویتی بهبود خواهد یافت. در عین حال، به منظور حفظ سازگاری بین گزارش‌های تولید شده و نتایج واقعی یا اصلی، یک ارزیاب سبک زبان پیشنهاد خواهد شد تا تولید مؤثر گزارش‌های تشخیصی برای سبک بالینی را محقق کند و اطلاعات آسیب‌شناسی بیشتری باقی بماند.

با توجه به مشکلات و چالش‌هایی که برای گزارش‌نویسی تصاویر پزشکی توسط الگوریتم‌های سنتی یادگیری ماشین و دیگر روش‌ها وجود دارد، لذا این پژوهش در تلاش است تا با بکارگیری شبکه‌های مولد متخاصم و یادگیری تقویتی به روشی مناسب برای تحلیل تصاویر پزشکی و گزارش‌نویسی آنها بدست آورد که این روش علاوه بر اینکه سرعت اجرای مناسبی برای گزارش‌نویسی تصاویر پزشکی دارد بتواند تا حد امکان دقت و کارایی را در این حوزه بهبود دهد.

---

[1] Medical images
[2] Caption
[3] Machine Learning (ML)
[4] Deep Learning (DL)
[5] Encryption-Decryption
[6] Generative Adversarial Networks (GANs)
[7] Reinforcement Learning (RL)

این پژوهش دارای اهداف مختلفی است که در نهایت منجر به ارائه روش و راهکاری کارآمد و دقیق جهت شرح نویسی تصاویر پزشکی با سبک بالینی خواهد شد، راهکار پیشنهادی این پژوهش علاوه بر افزایش دقت در شرح‌نویسی و گزارش‌نویسی تصاویر پزشکی، منجر به بهبود کارایی و سرعت پردازش و کاهش سربار محاسباتی خواهد شد. هدف اصلی این پژوهش ارائه روشی مبتنی بر الگوریتم شبکه های خصمانه مولد و یادگیری تقویتی برای شرح نویسی تصاویر پزشکی با سبک بالینی است. با توجه به اینکه اکثر تصاویر پزشکی دارای کیفیت پایین و نامناسب هستند، لذا نیاز به ارائه راهکاری برای شرح نویسی این تصاویر به صورت خودکار و با استفاده از روش‌های مختلف هوش مصنوعی است، بنابراین روش پیشنهادی سعی در ارائه راهکاری برای شرح نویسی تصاویر پزشکی با سبک بالینی دارد. لذا به صورت کلی این مطالعه دارای اهداف مختلفی می‌باشد که به شرح زیر است:

- ارائه روشی بر اساس شبکه های خصمانه مولد بهبود یافته و یادگیری تقویتی جهت شرح نویسی تصاویر پزشکی
- انتخاب معیارها و پارامترهای تأثیرگذار در شرح نویسی تصاویر پزشکی با سبک بالینی
- هدف ویژه در این پژوهش بررسی کارایی شبکه های خصمانه مولد بهبود یافته و یادگیری تقویتی جهت افزایش دقت در شرح نویسی تصاویر پزشکی با سبک بالینی است.

در بررسی‌های صورت‌گرفته از کارهایی که در زمینه پژوهش صورت گرفته است، روش های کمی برای شرح نویسی تصاویر پزشکی با سبک بالینی با در نظر گرفتن تمام معیارها و عوامل تأثیرگذار ارائه شده است و دلیل آن پیچیدگی تحلیل تمامی معیارها و پارامترها است. بر همین اساس در این پژوهش سعی می‌شود روشی بر اساس شبکه های خصمانه مولد و یادگیری تقویتی برای شرح نویسی تصاویر پزشکی ارائه گردد. با توجه به اینکه استفاده از شبکه های خصمانه مولد منجر خواهد شد تا مشکل شرح نویسی تصاویر پزشکی به نحو بهتری بهبود و دقیق تر شود، لذا این پژوهش در تلاش است تا با ترکیب الگوریتم های شبکهه ای خصمانه مولد و یادگیری تقویتی به روش مناسب دست یابد. نوآوری این پژوهش در این است که بتواند راهکاری ارائه دهد که در آن علاوه بر افزایش دقت، سرعت اجرای الگویتم پیشنهادی و کارایی آن را نیز افزایش یابد. به صورت کلی نوآوری این پژوهش شامل موارد زیر است:

- این پژوهش یک مدل کلی براساس شبکه های خصمانه مولد پیشنهاد خواهد کرد که بر اساس یادگیری تقویتی، شبکه عصبی کانولوشن، شبکه عصبی مکرر و مکانیسم توجه ترکیب شده و در یک ماژول بسته بندی می‌شوند تا یک تولید کننده گزارش تشخیص پزشکی بسازند.
- این پژوهش با متمایز کننده برای به دست آوردن اطلاعات پاداش بازخورد و بهبود کیفیت گزارش های تشخیص پزشکی استفاده می شود.

در ادامه، در بخش۲ پیشینه تحقیق، بخش ۳ روش شناسی تحقیق، بخش ۴ یافته های تحقیق و بخش ۵ نتیجه گیری و پیشنهادات، ارائه می‌گردد.

Commented [A1]: يك پاراگراف به آخر مقدمه اضافه كنيد كه ساختار مقاله را شرح ده به اين شكل: ساختار اين مقاله بدين شرح است: در بخش ۲ كارهاي پيشين مرور ميشود و خلاصه اي از اين پژوهشها ارائه ميشود. بخش ۳ به ۰۰۰مي پردازد و....الي آخر

کارهای پیشینه

شرح نویسی تصویر پزشکی به طور خودکار یک توضیح پزشکی برای توصیف محتوای یک تصویر پزشکی خاص ایجاد می کند. مدل‌های شرح تصاویر پزشکی سنتی تنها بر اساس یک ورودی تصویر پزشکی، یک توصیف پزشکی ایجاد می‌کنند. از این رو، ایجاد یک توصیف یا مفهوم پزشکی انتزاعی بر اساس رویکرد سنتی دشوار است. چنین روشی اثربخشی شرح تصاویر پزشکی را محدود می کند. زیرنویس تصاویر پزشکی چند وجهی یکی از روش‌هایی است که برای رفع این مشکل استفاده می شود. در زیرنویس تصاویر پزشکی چند وجهی، ورودی متنی، به عنوان مثال، کلمات کلیدی تعریف شده توسط متخصص، به عنوان یکی از محرک های اصلی تولید توضیحات پزشکی در نظر گرفته می شود. بنابراین، رمزگذاری ورودی متنی و تصویر پزشکی به طور موثر، هر دو برای کار شرح

تصاویر پزشکی چند وجهی مهم هستند. در مقاله Pavlopoulos و همکاران (۲۰۱۹) یک مدل توصیف تصویر پزشکی چند وجهی عمیق پایان به انتها پیشنهاد شده است. بازنمایی کلمات کلیدی متنی، تقویت ویژگی های متنی و توجه به خود پنهان شده برای توسعه رویکرد پیشنهادی استفاده می شود. بر اساس ارزیابی مجموعه داده‌های زیرنویس تصویر پزشکی چند وجهی موجود، نتایج تجربی نشان داده که مدل پیشنهادی با افزایش دقت و کارایی در مقایسه با وضعیت فعلی مؤثر است.

تولید خودکار گزارش‌های پزشکی برای تصاویر شبکیه یکی از راه‌های امیدوارکننده برای کمک به چشم پزشکان برای کاهش حجم کاری و بهبود کارایی کار است. در پژوهش Sing و همکاران (۲۰۲۲) یک شبکه رمزگذاری مبتنی بر زمینه جدید را برای تولید خودکار گزارش‌های پزشکی برای تصاویر شبکیه پیشنهاد شده است. مدل پیشنهادی عمدتاً از یک رمزگذار ورودی چند وجهی و یک رمزگشا با ویژگی ذوبی تشکیل شده است. نتایج تجربی این پژوهش نشان داده که روش پیشنهادی این پژوهش می‌تواند به طور موثر از اطلاعات تعاملی بین تصویر ورودی و زمینه، به عنوان مثال، کلمات کلیدی در مورد استفاده کند. روش پیشنهادی گزارش‌های دقیق‌تر و معنی‌داری را برای تصاویر شبکیه نسبت به مدل‌های پایه ایجاد می‌کند و به عملکردی پیشرفته دست می‌یابد. این عملکرد در چندین معیار معمول مورد استفاده برای وظیفه تولید گزارش پزشکی کارایی مناسبی را نشان داده است.

تولید گزارش تصویر پزشکی با هدف ایجاد توضیحات تشخیص مرتبط با جملات زبان طبیعی از تصاویر پزشکی، که در سیستم تشخیص به کمک رایانه ضروری است. با این وجود، این وظیفه همچنان چالش برانگیز است زیرا تصاویر پزشکی و عبارات زبانی باید به طور مشترک درک شوند که با این حال اختلافات زیادی را در روش نشان می دهند. برای پر کردن این شکاف بصری به معنایی، مطالعه Selivanov و همکاران (۲۰۲۲) یک چارچوب جدید پیشنهاد کرده که از خط لوله رمزگذار-رمزگشا پیروی می‌کند. چارچوب این مطالعه با رمزگذاری تعبیه‌های بصری و معنایی عمیق از طریق یک شبکه سه شاخه در طول مرحله رمزگذاری مشخص می‌شود. شاخه توجه بصری با مکانیسم توجه نرم، در جاسازی‌های بصری از تصاویر پزشکی شرکت می‌کند. شاخه تعبیه گزارش پزشکی تعبیه‌های گزارش معنایی را پیش‌بینی می‌کند. شاخه جاسازی سرفصل های موضوعی پزشکی تعبیه های معنایی برچسب های پزشکی مرتبط را به عنوان اطلاعات تکمیلی به دست می آورد. سپس خروجی های این شاخه ها ذوب شده و به یک رمزگشا برای تولید گزارش وارد می شوند. نتایج تجربی روی دو مجموعه داده معیار عملکرد عالی روش مطالعه را نشان داده است.

رویکردهای مبتنی بر ترانسفورماتور نتایج خوبی در کارهای زیرنویس تصویر نشان داده‌اند. با این حال، رویکردهای فعلی محدودیتی در تولید متن از ویژگی های کلی یک تصویر کامل دارند. بنابراین، مقاله Sing و همکاران (۲۰۲۲) روش‌های جدیدی را برای ایجاد زیرنویس تصویر بهتر به این شرح پیشنهاد کرده است: (۱) استخراج‌کننده بصری جهانی-محلی[1] برای ثبت ویژگی‌های جهانی و ویژگی‌های محلی. (۲) ترانسفورماتور رمزگذار-رمزگشا متقاطع[2] برای تزریق ویژگی‌های رمزگذار چند سطحی به فرآیند رمزگشایی. GLVE نه تنها ویژگی‌های بصری کلی را که می‌توان از کل یک تصویر به دست آورد، مانند اندازه اندام یا ساختار استخوان، بلکه ویژگی‌های بصری محلی را که می‌توان از یک منطقه محلی، مانند ناحیه ضایعه ایجاد کرد، استخراج کرد. با توجه به یک تصویر، CEDT می‌تواند با تزریق خروجی‌های انکودر سطح پایین و سطح بالا به رسیور، توصیف دقیقی از ویژگی‌های کلی ایجاد کند. هر روش به بهبود عملکرد کمک می کند و توصیفی مانند اندازه اندام و ساختار استخوان را ایجاد می کند. مدل پیشنهادی بر روی مجموعه داده اشعه ایکس IU مورد ارزیابی قرار گرفت و عملکرد بهتری نسبت به نتایج خط پایه مبتنی بر ترانسفورماتور به دست آورد.

با پیشرفت در کاربردهای مراقبت الکترونیکی مبتنی بر هوش مصنوعی، نقش تشخیص خودکار بیماری‌های مختلف به سرعت افزایش یافته است. با این حال، اکثر مدل‌های تشخیصی موجود نتایج را به صورت دوتایی ارائه می‌کنند، مانند اینکه آیا بیمار به بیماری

[1] Global-Local Visual Extractor (GLVE)
[2] Cross Encoder-Decoder Transformer (CEDT)

خاصی مبتلا شده است یا خیر. اما موارد زیادی وجود دارد که لازم است اطلاعات توضیحی مناسبی مانند ابتلای بیمار به بیماری خاص همراه با میزان آلودگی ارائه شود. بنابراین، مقاله Tian و همکاران (۲۰۲۰) برای ارائه اطلاعات توضیحی به پزشکان و بیماران، یک شبکه عریض تصویر پزشکی کارآمد (DCNet) پیشنهاد شده است. DCNet سه مدل از پیش آموزش دیده معروف مانند VGG16، ResNet152V2 و DenseNet201 را تشکیل می دهد. ترکیب این مدل ها با جلوگیری از مشکل بیش از حد برازش نتایج بهتری به دست می آورد. با این حال، DCNet به پارامترهای کنترلی خود حساس است. بنابراین، برای تنظیم پارامترهای کنترل، یک DCNet در حال تکامل (EDC-Net) پیشنهاد شد. فرآیند تکامل با استفاده از تکامل دیفرانسیل مبتنی بر کنترل پارامتر خود تطبیقی[1] به دست می آید. نتایج تجربی نشان داده که EDC-Net می‌تواند به طور موثر ویژگی های بالقوه تصاویر زیست پزشکی را استخراج کند. تجزیه و تحلیل مقایسه ای نشان داده که در مجموعه داده Open-i، EDC-Net از مدل های موجود از نظر آماری بهتر عمل می کند.

نوشتن شرح تصاویر پزشکی یک کار بسیار چالش برانگیز است که به ندرت در ادبیات شرح تصاویر طبیعی به آن پرداخته شده است. برخی از تکنیک‌های گزارش نویسی تصویر موجود از اشیاء موجود در تصویر در کنار ویژگی‌های بصری در حین ایجاد توضیحات بهره‌برداری می‌کنند. با این حال، زمانی که نیاز به توضیحات بالینی در توضیحات محتوای تصویر باشد، این امکان برای شرح تصاویر پزشکی وجود ندارد. مطالعه Tumuramve (۲۰۲۳) با الهام از موارد قبلی، استفاده از مفاهیم پزشکی مرتبط با تصاویر، مطابق با ویژگی‌های بصری آنها، برای ایجاد زیرنویس‌های جدید پیشنهاد کرده است. شبکه قابل آموزش سرتاسر این مطالعه از یک رمزگذار ویژگی معنایی بر اساس یک طبقه‌بندی کننده چند برچسبی برای شناسایی مفاهیم پزشکی مرتبط با تصاویر، یک رمزگذار ویژگی بصری و یک مدل حافظه طولانی کوتاه مدت[2] برای تولید متن تشکیل شده است. جستجوی پرتو برای اطمینان از بهترین انتخاب کلمه بعدی برای یک دنباله معین از کلمات بر اساس ویژگی های ادغام شده تصویر پزشکی استفاده می‌شود. این مطالعه پیشنهاد خود را بر روی مجموعه داده توصیف پزشکی ImageCLEF ارزیابی کرده و نتایج اثربخشی و کارایی رویکرد توسعه‌یافته را نشان داده است.

مبارزه با سل در منطقه آفریقا و سایر کشورهای با درآمد کم و متوسط عمدتاً توسط تعداد محدود رادیولوژیست های ماهر در آن کشورها به چالش کشیده شده است. جمعیت زیادی وجود دارد، اما تعداد کمی از رادیولوژیست ها وجود دارد که با مشکل رسیدگی به تعداد زیادی از بیماران مواجه می شوند. نوشتن گزارش پزشکی برای هر بیمار زمان نسبتا زیادی می برد، بنابراین تشخیص بسیاری از بیماران و نوشتن دستی گزارش پزشکی برای هر یک از آنها بسیار زمان بر و پر زحمت است. یک سیستم گزارش نویسی تصویر با کمک یادگیری عمیق، پشتیبانی زیادی از رادیولوژیست ها در نوشتن شرح تصویر خودکار CXR ارائه می دهد که می تواند ابزار ارزشمندی برای تشخیص سل و نوشتن گزارش پزشکی برای بیماران باشد و همچنین نتایج سریعتر و دقیق تری ارائه دهد. در مقاله Xue و Huang (۲۰۱۹) کاربرد یادگیری عمیق در عنوان تصویر CXR برای سل مورد بررسی قرار گرفته است. یک مجموعه داده منبع باز از دانشگاه ایندیانا و یک مجموعه داده بالینی محلی از بیمارستان منگو به دست آمد. مجموعه داده دانشگاه ایندیانا شامل ۷۴۷۰ تصویر اشعه ایکس قفسه سینه به همراه ۲۹۵۵ گزارش مرتبط با آنها در قالب xml بود. ۳۱۱ تصویر همراه با گزارش‌هایشان نیز از بیمارستان منگو برای تشکیل مجموعه داده‌های محلی جمع‌آوری شد. دو مدل از پیش آموزش‌دیده‌شده یعنی EfficientNet و CheXNet به‌عنوان استخراج‌کننده ویژگی‌های پایه مورد استفاده قرار گرفتند و برای طراحی دو مدل استفاده شدند که می‌توانند برای تصاویر اشعه ایکس قفسه سینه توضیح ایجاد کنند. مدل ترانسفورماتور Efficient-Net شامل شبکه عصبی پیچشی[3] کارآمد است که به عنوان استخراج کننده ویژگی، رمزگذار ترانسفورماتور وانیلی و رمزگشا برای تولید زیرنویس ها استفاده می شود. مدل

[1] Self-Adaptive Parameter Control-Based Differential Evolution (SAPCDE)
[2] Long Short-Term Memory (LSTM)
[3] Conventional Neural Network (CNN)

CheXnet-LSTM شامل CheXNet CNN است که به عنوان استخراج کننده ویژگی، رمزگذار و LSTM به عنوان رمزگشا برای تولید گزارش استفاده می شود. مدل ها با استفاده از مجموعه داده های دانشگاه ایندیانا آموزش داده شدند و با استفاده از مجموعه داده های دانشگاه ایندیانا و مجموعه داده های محلی ارزیابی شدند. مدل EfficientNet-Transformer بهترین مدل عملکرد را با امتیاز BLEU 0.515 نشان داد که بهتر از نتایج رویکردهای پیشرفته بود. این مدل در یک برنامه وب به کار گرفته شد که به کاربر اجازه می‌داد تصویر اشعه ایکس قفسه سینه را بارگذاری کند و در عرض چند ثانیه یک عنوان پیش‌بینی‌شده را دریافت کند.

گزارش نویسی تصویر توسط چارچوب رمزگذار-رمزگشا پیشرفت فوق العاده ای را در دهه گذشته نشان داده است، جایی که CNN عمدتاً به عنوان رمزگذار و LSTM به عنوان رمزگشا استفاده می شود. با وجود چنین دستاورد چشمگیری از نظر دقت در تصاویر ساده، از نظر پیچیدگی زمانی و کارایی پیچیدگی فضایی فاقد آن است. علاوه بر این، در صورت وجود تصاویر پیچیده با اطلاعات و اشیاء زیاد، عملکرد این جفت CNN-LSTM به دلیل عدم درک معنایی صحنه های ارائه شده در تصاویر، به صورت تصاعدی کاهش می یابد. بنابراین، برای در نظر گرفتن این مسائل، چارچوب رمزگشای رمزگذار شبکه عصبی پیچشی و واحد بازگشتی دروازه ای[1] برای بازسازی عنوان به تصویر در پژوهش Xue و Huang (۲۰۱۹) در نظر گرفته شده است تا زمینه معنایی و همچنین پیچیدگی زمانی را در نظر بگیرد. با در نظر گرفتن حالت‌های پنهان رمزگشا، تصویر ورودی و نمایش‌های معنایی مشابه آن بازسازی می‌شود و نمرات بازسازی از یک بازسازی‌کننده معنایی همراه با احتمال در طول آموزش مدل برای ارزیابی کیفیت عنوان تولید شده استفاده می‌شود. در نتیجه، رمزگشا اطلاعات معنایی بهبود یافته را دریافت می کند و فرآیند تولید عنوان را بهبود می بخشد. در طول آزمایش مدل، ترکیب امتیاز بازسازی و احتمال گزارش نیز برای انتخاب مناسب‌ترین عنوان امکان‌پذیر است. مدل پیشنهادی از نظر پیچیدگی و دقت زمانی نسبت به مدل LSTM-A5 برتری دارد.

تولید خودکار گزارش‌های پزشکی می‌تواند کمک‌های تشخیصی به پزشکان ارائه دهد و حجم کاری آنها را کاهش دهد. برای بهبود کیفیت گزارش های پزشکی تولید شده، تزریق اطلاعات کمکی از طریق نمودارهای دانش یا الگوها به مدل به طور گسترده در روش های قبلی مورد استفاده قرار می گیرد. با این حال، آنها از دو مشکل رنج می برند: ۱) اطلاعات خارجی تزریق شده از نظر مقدار محدود است و پاسخگویی مناسب به نیازهای اطلاعاتی تولید گزارش پزشکی در محتوا دشوار است. ۲) اطلاعات خارجی تزریق شده پیچیدگی مدل را افزایش می دهد و به سختی می توان به طور منطقی در فرآیند تولید گزارش های پزشکی ادغام کرد. بنابراین، مطالعه Zhang و همکاران (۲۰۲۳) یک ترانسفورماتور کالیبره شده اطلاعات[2] را برای رسیدگی به مسائل فوق پیشنهاد کرده است. این مطالعه ابتدا یک ماژول ارتقای اطلاعات پیشرو[3] را طراحی نموده که می‌تواند به طور موثری ویژگی‌های گزارش درونی متعددی را از مجموعه داده‌ها به عنوان اطلاعات کمکی بدون تزریق خارجی استخراج کند و اطلاعات کمکی را می توان به صورت پویا با فرآیند آموزش به روز کرد. در مرحله دوم، یک حالت ترکیبی، که از PEM و ماژول توجه کالیبراسیون اطلاعات پیشنهادی مطالعه تشکیل شده است، طراحی و در ICT تعبیه شده است. در این روش اطلاعات کمکی استخراج شده از PEM به صورت انعطاف پذیر به ICT تزریق می شود و افزایش پارامترهای مدل اندک است. ارزیابی‌های جامع تأیید کرده که ICT نه تنها نسبت به روش‌های قبلی در مجموعه داده‌های اشعه ایکس، IU-X-Ray و MIMIC-CXR برتری دارد، بلکه با موفقیت به مجموعه داده‌های CT COV-CTR COVID-19 نیز گسترش یافته است.

جدول ۱. خلاصه کارهای مرتبط

[1] Gated Recurrent Unit (GRU)
[2] Information Calibrated Transformer (ICT)
[3] Precursor-information Enhancement Module (PEM)

مرجع	روش	مزایا	معایب
Pavlopoulos et al (2019)	یک مدل توصیف تصویر پزشکی چند وجهی عمیق پایان به انتها پیشنهاد شده است.	مدل پیشنهادی با افزایش دقت و کارایی در مقایسه با وضعیت فعلی مؤثر است.	برای مجموعه داده بزرگ نمیتواند موثر باشد.
Singh et al (2022a)	یک شبکه رمزگذاری مبتنی بر زمینه جدید را برای تولید خودکار گزارش‌های پزشکی برای تصاویر شبکیه پیشنهاد شده است.	روش پیشنهادی گزارش‌های دقیق‌تر و معنی‌داری را برای تصاویر شبکیه نسبت به مدل‌های پایه ایجاد می‌کند و به عملکردی پیشرفته دست می‌یابد.	چالش این مقاله، تولید گزارش تصویر پزشکی با هدف ایجاد توضیحات تشخیص مرتبط با جملات زبان طبیعی از تصاویر پزشکی است.
Selivanov et al (2022)	یک چارچوب جدید پیشنهاد کرده که از خط لوله رمزگذار-رمزگشا پیروی می‌کند.	نتایج تجربی روی دو مجموعه داده معیار عملکرد عالی روش مطالعه را نشان داده است.	رویکرد فعلی محدودیتی در تولید متن از ویژگی های کلی یک تصویر کامل دارند.
Singh et al (2022b)	روش‌های جدیدی را برای ایجاد زیرنویس تصویر بهتر به این شرح پیشنهاد کرده است: (۱) استخراج‌کننده بصری جهانی-محلی[1] برای ثبت ویژگی‌های جهانی و ویژگی‌های محلی. (۲) ترانسفورماتور رمزگذار-رمزگشا متقاطع۲ برای تزریق ویژگی‌های رمزگذار چند سطحی به فرآیند رمزگشایی.	مدل پیشنهادی بر روی مجموعه داده اشعه ایکس IU مورد ارزیابی قرار گرفت و عملکرد بهتری نسبت به نتایج خط پایه مبتنی بر ترانسفورماتور به دست آورد.	مدل تشخیصی موجود نتایج را به صورت دوتایی ارائه می‌کنند، مانند اینکه آیا بیمار به بیماری خاصی مبتلا شده است یا خیر. اما موارد زیادی وجود دارد که لازم است اطلاعات توضیحی مناسبی مانند ابتلای بیمار به بیماری خاص همراه با میزان آلودگی ارائه شود.
Tian et al (2020)	یک شبکه عریض تصویر پزشکی کارآمد (DCNet) پیشنهاد شده است. DCNet سه مدل از پیش آموزش دیده معروف مانند VGG16، ResNet152V2 و DenseNet201 را تشکیل می دهد.	نتایج تجربی نشان داده که EDC-Net می‌تواند به طور موثر ویژگی های بالقوه تصاویر زیست پزشکی را استخراج کند.	تکنیک گزارش نویسی تصویر موجود از اشیاء موجود در تصویر در کنار ویژگی‌های بصری در حین ایجاد توضیحات بهره‌برداری می‌کنند. با این حال، زمانی که نیاز به توضیحات بالینی در توضیحات محتوای تصویر باشد، این امکان برای شرح تصاویر پزشکی وجود ندارد.
Tumuramye (2023)	استفاده از مفاهیم پزشکی مرتبط با تصاویر، مطابق با ویژگی‌های بصری آنها، برای ایجاد زیرنویس‌های جدید پیشنهاد کرده است.	این مطالعه پیشنهاد خود را بر روی مجموعه داده توصیف پزشکی ارزیابی کرده و نتایج اثربخشی و کارایی رویکرد توسعه‌یافته را نشان داده است.	روش موجود برای گزارش نویسی برخی بیماری ها کاربرد دارد.
Xue and Huang (2019)	کاربرد یادگیری عمیق در عنوان تصویر CXR برای سل مورد بررسی قرار گرفته است. دو مدل از پیش آموزش‌دیده‌شده یعنی EfficientNet و CheXNet به‌عنوان استخراج‌کننده ویژگی‌های پایه مورد استفاده قرار گرفتند.	مدل EfficientNet-Transformer بهترین مدل عملکرد را با امتیاز BLEU 0.515 نشان داد که بهتر از نتایج رویکردهای پیشرفته بود. این مدل در یک برنامه وب به کار گرفته شد که به کاربر اجازه می‌داد تصویر اشعه ایکس قفسه سینه	با وجود چنین دستاورد چشمگیری از نظر دقت در تصاویر ساده، از نظر پیچیدگی زمانی و کارایی پیچیدگی فضایی فاقد آن است.

[1] Global-Local Visual Extractor (GLVE)
[2] Cross Encoder-Decoder Transformer (CEDT)

مرجع	روش	مزایا	معایب
		را بارگذاری کند و در عرض چند ثانیه یک عنوان پیش‌بینی‌شده را دریافت کند.	
Yang et al (2021)	چارچوب رمزگشای رمزگذار شبکه عصبی پیچشی و واحد بازگشتی دروازه ای[1] برای بازسازی عنوان به تصویر در نظر گرفته شده است تا زمینه معنایی و همچنین پیچیدگی زمانی را در نظر بگیرد.	در طول آزمایش مدل، ترکیب امتیاز بازسازی و احتمال گزارش نیز برای انتخاب مناسب‌ترین عنوان امکان‌پذیر است. مدل پیشنهادی از نظر پیچیدگی و دقت زمانی نسبت به مدل LSTM-A5 برتری دارد.	این روش از دو مشکل رنج می برند: ۱) اطلاعات خارجی تزریق شده از نظر مقدار محدود است و پاسخگویی مناسب به نیازهای اطلاعاتی تولید گزارش پزشکی در محتوا دشوار است. ۲) اطلاعات خارجی تزریق شده پیچیدگی مدل را افزایش می دهد و به سختی می توان به طور منطقی در فرآیند تولید گزارش های پزشکی ادغام کرد.

روش تحقیق

از نظر هدف تحقیق با توجه به اینکه پژوهش حاضر به دنبال ارائه مدل روزآمد توسعه فضایی خانه های منطقه اورامانات کردستان در جهت احیا و بهره وری موثر از میراث بومی منطقه است، در حیطه تحقیقات اکتشافی طبقه‌بندی می‌شود. همچنین، تحقیق حاضر از نظر چگونگی گردآوری داده‌های مورد نیاز، در گروه«تحقیق کیفی» طبقه‌بندی می‌شود. برای این منظور داده‌های کیفی گردآوری شده است که منجر به شناسایی جنبه‌های متعدد پدیده شده و امکان تدوین الگوی مفهومی تحقیق فراهم می‌شود. به طور کلی دلیل انتخاب روش تحقیق کیفی برای تحقیق حاضر عبارتند از:

۱. شناسایی عوامل موثر بر توسعه فضایی خانه های منطقه اورامانات کردستان در جهت احیا و بهره وری موثر از میراث بومی منطقه.
۲. لزوم استفاده از دیدگاه‌های خبرگان دانشگاهی متخصص درباره ارائه مدل روزآمد توسعه فضایی خانه های منطقه اورامانات کردستان در جهت احیا و بهره وری موثر از میراث بومی منطقه.

روش اجرایی مدل پیشنهادی

در این پژوهش، معماری جدیدی طراحی می کنیم که شبکه‌های متخاصم مولد (GANs) و تکنیک‌های یادگیری تقویتی را ترکیب کند. معماری GAN شامل یک مولد و یک تفکیک کننده است که در آن مولد برای تصاویر پزشکی شرح تولید می کند و تفکیک کننده کیفیت زیرنویس های تولید شده را ارزیابی می کند. هدف این معماری جدید بهبود کیفیت و دقت زیرنویس‌های تولید شده است.

معماری GAN از دو جزء اصلی تشکیل شده است: یک مولد و یک تفکیک کننده.

آ) مولد

مولد یک تصویر پزشکی را به عنوان ورودی می گیرد و زیرنویس های مربوطه را ایجاد می کند. آن را با استفاده از یادگیری تقویتی پیاده سازی می کنیم. هدف مولد این است که توزیع متن زیرنویس را با توجه به تصویر پزشکی بیاموزد.

[1] Gated Recurrent Unit (GRU)

شبکه مولد را می توان با استفاده از الگوریتم REINFORCE که یک الگوریتم یادگیری تقویتی استاندارد برای آموزش مدل های مولد است، آموزش داد. مولد برای به حداکثر رساندن پاداش مورد انتظار آموزش دیده است که بر اساس کیفیت و دقت زیرنویس‌های تولید شده تعریف می‌شود.

الگوریتم REINFORCE از روش گرادیان خط مشی برای به روز رسانی پارامترهای شبکه مولد استفاده می کند. گرادیان خط مشی به عنوان حاصلضرب احتمال گزارش عنوان تولید شده و سیگنال پاداش محاسبه می شود. پارامترهای شبکه مولد با استفاده از صعود گرادیان تصادفی برای به حداکثر رساندن پاداش مورد انتظار به روز می شوند.

فرمول به روز رسانی پارامترهای شبکه مولد با استفاده از گرادیان خط مشی به شرح زیر است:

$$\nabla\theta\, J(\theta) = E[\sum t\, \nabla\theta \log P(Ct \mid I) * R(t)] \quad (۱)$$

که:

θ -بردار پارامتر شبکه ژنراتور است

$J(\theta)$ -پاداش مورد انتظار (هدف) است که می خواهیم آن را به حداکثر برسانیم

$\nabla\theta$ -گرادیان نسبت به θ است

Ct -عنوان ایجاد شده در مرحله زمانی t است

-من تصویر پزشکی ورودی است

$P(Ct \mid I)$ -احتمال ایجاد عنوان Ct با توجه به تصویر ورودی I است

$R(t)$ -پاداش در مرحله زمانی t است که معیاری برای کیفیت و دقت عنوان تولید شده است.

فرمول فوق گرادیان پارامترهای شبکه مولد را با توجه به پاداش مورد انتظار محاسبه می کند. از احتمال لاگ عنوان تولید شده $(\log P(Ct \mid I))$ضرب در سیگنال پاداش در هر مرحله زمانی $(R(t))$ استفاده می کند.

برای به‌روزرسانی پارامترها، از شیب تصادفی صعودی استفاده می‌شود که با به‌روزرسانی مکررθ ، پاداش مورد انتظار $(J(\theta))$ را به حداکثر می‌رساند.

معماری مولد بصورت زیر است:

لایه ورودی: ورودی شبکه مولد تصویر پزشکی است. می توان آن را با استفاده از یک لایه کاملاً متصل برای گرفتن ویژگی های تصویر به شبکه تغذیه کرد.

لایه‌های مدل‌سازی زبان: پس از پردازش ویژگی‌های تصویر ورودی، مولد می‌تواند از لایه‌های مدل‌سازی زبان مانند شبکه‌های ترانسفورماتور استفاده کند. این لایه ها با پیش بینی کلمه بعدی در دنباله بر اساس کلمات تولید شده قبلی، زیرنویس های متنی را ایجاد می کنند.

لایه خروجی: لایه خروجی مولد زیرنویس ها را کلمه به کلمه تولید می کند. در هر مرحله زمانی، توزیع احتمال را بر روی واژگان کلمات ممکن تولید می کند.

ب) تفکیک کننده

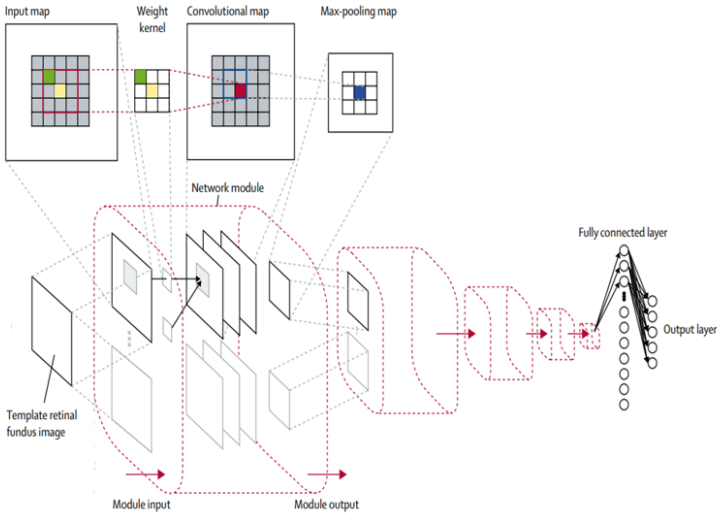
تفکیک کیفیت زیرنویس های ایجاد شده را ارزیابی می کند. تعیین می کند که آیا عنوان داده شده واقعی است (از مجموعه داده آموزشی) یا تولید شده توسط مولد. تمایز با استفاده از یک مدل طبقه بندی، مانند یک شبکه عصبی کانولوشن(CNN) اجرا می شود. برای بهبود کیفیت زیرنویس‌های تولید شده، بازخوردی را به مولد ارائه می‌دهد.

در روش پیشنهادی، از تکنیک یادگیری تقویتی برای آموزش شبکه مولد برای تولید زیرنویس‌های دقیق برای تصاویر پزشکی استفاده می‌شود. یادگیری تقویتی یک پارادایم از یادگیری ماشینی است که در آن یک عامل یاد می گیرد که اقداماتی را در یک محیط انجام دهد تا سیگنال پاداش را به حداکثر برساند.

شبکه تفکیک کننده در روش پیشنهادی با استفاده از یک CNN پیاده سازی شده است.CNN به دلیل توانایی آنها در ثبت ویژگی های محلی و فضایی به طور گسترده برای کارهای طبقه بندی تصاویر استفاده می شوند.

۱. لایه ورودی: ورودی شبکه تفکیک کننده، زیرنویس های تولید شده است. هر عنوان می‌تواند به‌عنوان دنباله‌ای از جاسازی‌های کلمه یا بردارهای رمزگذاری شده یک‌طرفه نمایش داده شود. لایه ورودی این نمایش ها را پردازش می کند.
۲. لایه های کانولوشن: لایه های کانولوشن در استخراج ویژگی های محلی و فضایی از زیرنویس های ورودی تخصص دارند. این لایه‌ها از چندین فیلتر (همچنین به عنوان هسته شناخته می‌شوند) تشکیل شده‌اند که روی زیرنویس‌های ورودی جمع می‌شوند و نقشه‌های ویژگی را تولید می‌کنند. هر فیلتر ویژگی های مختلف ورودی را ثبت می کند.
۳. لایه ادغام: بعد از لایه های کانولوشن، لایه های حداکثر ادغام معرفی می شوند تا نقشه های ویژگی را پایین بیاورند. حداکثر ادغام ابعاد ویژگی ها را کاهش می دهد و در عین حال برجسته ترین اطلاعات را حفظ می کند.
۴. لایه های کاملا متصل: خروجی حداکثر لایه های ادغام مسطح شده و به لایه های کاملا متصل تغذیه می شود. این لایه ها تمام نورون های لایه قبلی را به لایه بعدی متصل می کنند. لایه های کاملاً متصل مسئول ثبت ویژگی های سطح بالاتر و تصمیم گیری نهایی در طبقه بندی هستند.
۵. لایه خروجی: لایه خروجی از یک نورون منفرد با تابع فعال سازی سیگموئید تشکیل شده است. این نورون یک امتیاز احتمال تولید می‌کند که نشان‌دهنده احتمال واقعی یا ایجاد زیرنویس‌های ایجاد شده است.

شبکه تفکیک کننده با استفاده از یک تابع تلفات متقابل آنتروپی باینری و بهینه سازی نزول گرادیان آموزش داده می شود. افت آنتروپی متقاطع باینری تفاوت بین امتیازهای احتمال پیش‌بینی‌شده و برچسب‌های حقیقت زمینی (واقعی یا تولید شده) را اندازه‌گیری می‌کند. پارامترهای تشخیص دهنده به طور مکرر به روز می شوند تا این تلفات به حداقل برسد.



شکل ۱. معماری تفکیک کننده

شبکه تفکیک کننده با به‌روزرسانی مکرر پارامترها با استفاده از این گرادیان‌ها با تکنیک نزول گرادیان تصادفی (SGD) آموزش داده می‌شود. این تابع فعال سازی به عنوان یک لایه اعمال می شود که روی هر عنصر در شبکه عمل می کند. معماری CNN از ترکیبی از چند لایه CONV-RELU و به دنبال آن لایه های تجمع تشکیل شده است. این طرح تا زمانی که تصویر ورودی به اندازه دلخواه کاهش یابد تکرار می شود. در طول این فرآیند، از لایه‌های مختلفی مانند Average Pooling می‌توان برای نمونه‌برداری بیشتر از نقشه‌های ویژگی استفاده کردیم.

لایه نهایی شبکه حاوی خروجی است که می تواند امتیازات دسته های مختلف طبقه بندی باشد. این لایه آخر ویژگی های سطح بالا استخراج شده توسط لایه های قبلی را می گیرد و بر اساس آنها یک پیش بینی تولید می کند.

یک لایه کانولوشن قبل از هر لایه Maxpooling قرار دادیم. این رویکرد به ویژه برای شبکه‌های بزرگ و عمیق مفید است، زیرا ترکیب لایه‌های کانولوشن چندگانه می‌تواند ویژگی‌های پیچیده‌تری را از داده‌های ورودی قبل از عملیات تجمیع استخراج کند، مانند Maxpooling، اطلاعات مکانی را از طریق نمونه‌برداری پایین حذف می‌کند.

از نظر ریاضی، یک لایه کانولوشن را در نظر بگیریم. با توجه به یک تصویر ورودی یا نقشه ویژگی، بیایید آن را به عنوان X نشان دهیم. هر عنصر در X به صورت X[i, j, c] نشان داده می شود، جایی که i و j نشان دهنده موقعیت مکانی و c نشان دهنده کانال یا عمق است.

یک لایه کانولوشن مجموعه ای از فیلترها یا هسته ها را که با W[l] نشان داده می شود، که در آن l نشان دهنده شماره لایه است، به ورودی X اعمال می کند. هر فیلتر با ورودی در هم می پیچد تا یک نقشه ویژگی تولید کند که با Y[l] نشان داده می شود. این عملیات را می توان به صورت زیر بیان کرد:

Y[l] = W[l] * X که در آن * عملیات کانولوشن را نشان می دهد.

سپس خروجی این لایه کانولوشن از طریق تابع فعال سازی که در این مورد تابع ReLU است، عبور داده می شود. ReLU به صورت زیر تعریف می شود:

$$ReLU(x) = \max(0, x) \quad (۲)$$

پس از اعمال فعال سازی ReLU، نقشه ویژگی به دست آمده به لایه بعدی وارد می شود. در مورد Maxpooling، این لایه با انتخاب حداکثر مقدار در هر منطقه pooling، ابعاد فضایی نقشه ویژگی را کاهش می دهد. فرمول Maxpooling با اندازه pool ، 2*2 و strides ، ۲ به صورت زیر نمایش داده شود:

$$MaxPooling(X) = max\_pool(X, pool\_size=(2, 2), strides=(2, 2)) \quad (۳)$$

این فرآیند برای چندین لایه تکرار می‌شود و هر لایه بر روی خروجی‌های لایه‌های قبلی ساخته می‌شود و ویژگی‌های پیچیده‌تری را استخراج می‌کند و ابعاد فضایی را کاهش می‌دهد تا اندازه مورد نظر به دست آید. در نهایت، لایه خروجی امتیازات پیش‌بینی‌شده را برای دو دسته مختلف در نظر گرفته شده در کار مورد نظر تولید می‌کند. خروجی معماری شبکه تفکیک کننده نمرات کلاس یا احتمالات برای هر کلاس است. لایه کاملاً متصل نهایی مسئول تبدیل ویژگی های آموخته شده به نمرات کلاس است. خروجی را می توان با استفاده از فرمول بدست آورد:

$$output = fully\_connected(input, num\_classes) \quad (۴)$$

پارامتر num_classes تعداد کلاس های مختلف را در کار طبقه بندی نشان می دهد.

در ادامه راهنمای گام به گام طراحی معماری آمده است:

(۱) لایه ورودی: لایه ورودی CNN تصاوی از پیش پردازش شده را می گیرد. ابعاد لایه ورودی به ابعاد تصاویر در مجموعه داده بستگی دارد. (۲) لایه های کانولوشن: لایه های کانولوشن مسئول استخراج ویژگی ها از تصاویر ورودی هستند. در resnet، از بلوک‌های باقیمانده استفاده می‌شود که هر بلوک باقیمانده از چندین لایه کانولوشن تشکیل شده است. (۳) تابع فعال سازی: بعد از هر لایه کانولوشن، یک تابع فعال سازی برای معرفی غیرخطی به شبکه اعمال می شود. از تابع فعال سازی ReLU استفاده شده است. (۴) لایه‌های pooling: لایه‌های pooling برای نمونه‌برداری از نقشه‌های ویژگی به‌دست‌آمده از لایه‌های کانولوشن و در عین حال حفظ اطلاعات مهم استفاده می‌شوند. از Max pooling استفاده کردیم که یک تکنیک تجمع رایج است که حداکثر مقدار را در یک پنجره انتخاب می کند. (۵) بلوک‌های باقیمانده: بلوک‌های باقیمانده نقش مهمی در اجرای شبکه تفکیک کننده دارند. آنها حاوی چندین لایه کانولوشن هستند و خروجی آخرین لایه با ورودی از طریق یک اتصال پرش ترکیب می شود. این شبکه را قادر می‌سازد تا نگاشت‌های باقی‌مانده را بیاموزد و مشکل گرادیان را کاهش دهد. (۶) لایه‌های کاملاً متصل: پس از لایه‌های کانولوشن و pooling، لایه‌های کاملاً متصل برای طبقه‌بندی تصاویر ورودی اضافه می‌شوند. این لایه ها ویژگی های استخراج شده را می گیرند و آنها را به پیش بینی برای کلاس های مختلف تبدیل می کنند. (۷) تابع فعال سازی (لایه خروجی): در لایه خروجی، یک تابع فعال سازی برای طبقه بندی اعمال می شود. (۸) تابع loss: تابع loss تفاوت بین برچسب های پیش بینی شده و برچسب های واقعی را اندازه گیری می کند. از تابع loss ، میانگین خطا برای طبقه بندی استفاده کردیم. پایین بودن loss به منزله بهتر بودن مدل است. اگر مدل بر رویدادهای آموزشی دچار بیش برازش نشده باشد. مقدارloss برای داده‌های آموزش و آزمایش محاسبه می‌شود و میزان

خطای مدل را نشان می‌دهد. در شبکه‌های عصبی هدف اصلی کاهش مقدار loss از طریق اصلاح وزن‌ها است که این اصلاح وزن از طریق تابع بهینه‌ساز انجام می‌گیرد. بیایید نشان دهیم:

D(x) -به عنوان امتیاز احتمال پیش بینی شده تولید شده توسط شبکه تشخیص دهنده برای ورودی x.

y -به عنوان برچسب حقیقت پایه برای ورودی x ، که در آن y=۱ نشان دهنده یک عنوان واقعی و y=۰ نشان دهنده یک عنوان تولید شده است. تابع ضرر برای آموزش تفکیک کننده به وسیله:

$$L(x,y)=-(y\log(D(x))+(1-y)\log(1-D(x))) \quad (5)$$

تابع ضرر بالا، تفکیک کننده را برای هر دو پیش‌بینی نادرست زیرنویس‌های واقعی (y=1) به عنوان زیرنویس‌های تولید شده جریمه می‌کند و بالعکس. هدف این است که با به روز رسانی پارامترهای تشخیص دهنده با استفاده از گرادیان نزول، این تلفات را به حداقل برسانیم. گرادیان برای به روز رسانی پارامترها را می توان به صورت زیر محاسبه کرد:

$$\nabla\theta D\ L(x,y)=-(y/D(x)-(1-y)/(1-D(x)))\ \nabla f(x) \quad (6)$$

جایی که $\nabla\theta D\ L(y,x)$ - گرادیان با توجه به پارامترهای تمایزکننده θD است $\nabla f(x)$ - .گرادیان تابع فعال سازی در ورودی x است. (۹) الگوریتم بهینه سازی: برای آموزش CNN، از الگوریتم بهینه سازی SGD برای به روز رسانی مکرر پارامترهای مدل استفاده شد. (۱۰) نرخ یادگیری: نرخ یادگیری تعیین می کند که مدل با چه سرعتی در طول آموزش سازگار می شود. این یک فراپارامتر مهم است که برای عملکرد بهینه باید تنظیم شود.

مجموعه داده ای از تصاویر پزشکی را به همراه زیرنویس های مربوطه آنها جمع آوری می کنیم. تصاویر را از قبل پردازش می کنیم، مانند تغییر اندازه، نرمال سازی، و افزایش، برای اطمینان از ثبات و بهبود عملکرد آموزش.

پیش پردازش تصویر

- تغییر اندازه: برای اطمینان از یکنواختی داده ها، اندازه تصاویر را به ابعاد ثابت تغییر دهید.
- اندازه هر تصویر را با استفاده از روش های درون یابی نزدیکترین همسایه به ارتفاع h و عرض w مورد نظر تغییر دهید:

```
resized_image = cv2.resize(image, (w, h), interpolation=cv2.INTER_LINEAR)
```
(۷)

- نرمال سازی: مقادیر پیکسل تصاویر را به یک محدوده مشترک، معمولاً [۰، ۱] نرمالسازی می کنیم تا آنها را با مدل های مختلف سازگار کند و مسائل مربوط به مقیاس های مختلف را کاهش دهد.
- هر مقدار پیکسل را بر حداکثر مقدار تقسیم می کنیم (به عنوان مثال، ۲۵۵ برای تصاویر ۸ بیتی):

```
normalized_image = image / 255.0
```
(۸)

- داده افزایی: انجام تکنیک های مختلف افزایش داده ها برای افزایش مصنوعی اندازه و تنوع داده های آموزشی. افزایش به ویژه زمانی مفید است که مجموعه داده کوچک باشد یا تنوع کافی نداشته باشد.

- تکنیک ها: برخی از تکنیک های رایج داده افزایی تصاویر پزشکی عبارتند از چرخش، پوسته پوسته شدن، چرخاندن و اضافه کردن نویز.

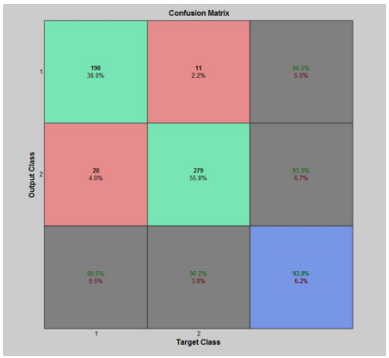
پیش پردازش زیرنویس

- توکن سازی: زیرنویس های متنی را به دنباله ای از نشانه ها یا کلمات تبدیل می کنیم، که بعداً می تواند به مدل داده شود.
- ایجاد واژگان: با جمع آوری تمام نشانه های منحصر به فرد از زیرنویس ها، یک واژگان ایجاد می کنیم. شناسه های عدد صحیح منحصر به فرد را به هر توکن اختصاص می دهیم، که به عنوان ورودی و هدف برای مدل عمل می کند.
- Padding: برای اطمینان از اندازه ورودی ثابت در طول آموزش، دنباله‌های زیرنویس را به یک طول ثابت بکشید یا کوتاه می کنیم.
- اگر طول عنوان کمتر از طول تعیین شده است، آن را با یک نشانه خاص (به عنوان مثال، «<pad>») قرار دهید تا به طول مورد نظر برسد. اگر از طول تعیین شده بیشتر شد، آن را کوتاه می کنیم.

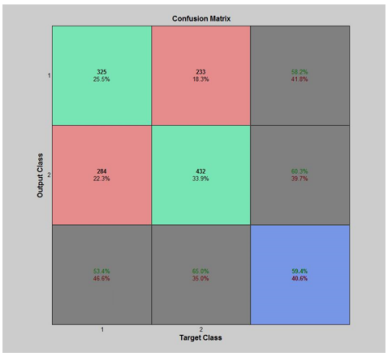
```
padded_caption = caption[:max_length] + ['] * (max_length - len(caption))
```
(۸)

یافته ها

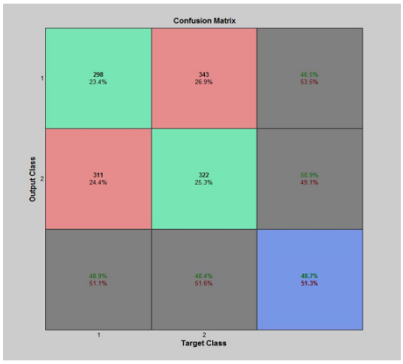
ماتریس confusion برای روش پیشنهادی و مقالات (Ahmad و همکاران (۲۰۲۳)، Ayesha و همکاران (۲۰۲۱) و Agrawal و همکاران (۲۰۱۹) برای کلاس۱: شرح‌هایی است که درست در نظر گرفته می‌شوند. و کلاس۲: عدم شرح درست بصورت شکل های ۲ تا ۵ می باشد.



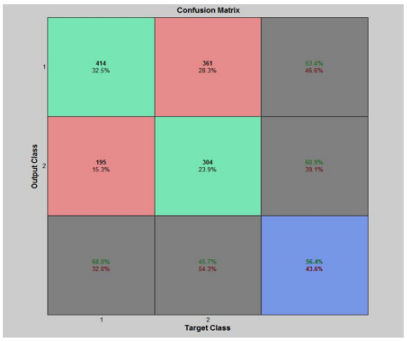
شکل۲. روش پیشنهادی



شکل ۳. مقاله **Ahmad** و همکاران (۲۰۲۳)

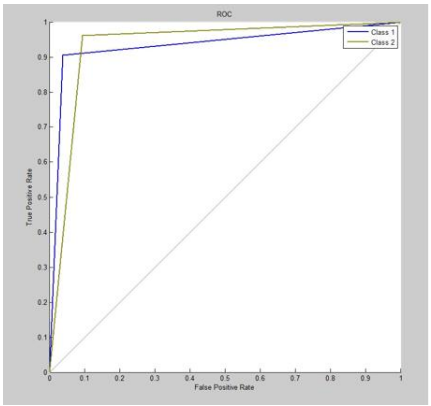


شکل۴. مقاله **Ayesha** و همکاران (۲۰۲۱)

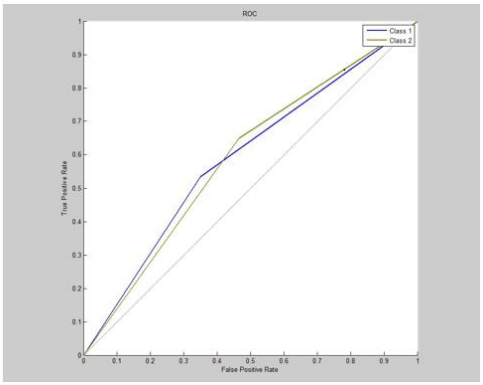


شکل۵. مقاله **Agrawal** و همکاران (۲۰۱۹)

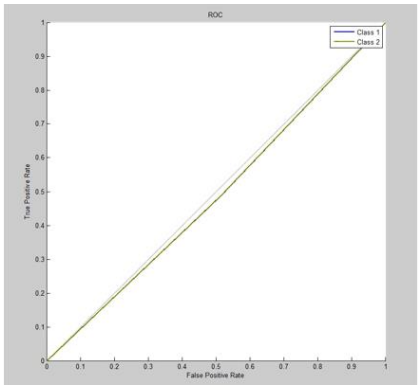
نمودار ROC برای روش پیشنهادی را با مقالات دیگر بصورت شکل های ۶ تا ۱۰ می باشد.



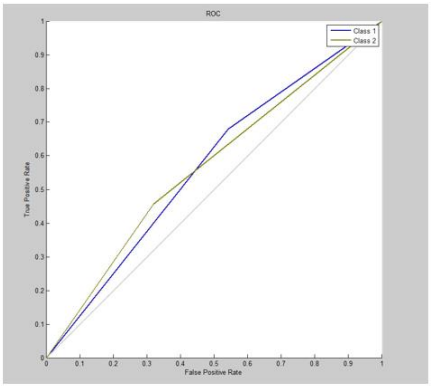
شکل ۶. روش پیشنهادی



شکل ۷. مقاله **Ahmad** و همکاران (۲۰۲۳)



شکل ۸. مقاله **Ayesha** و همکاران (۲۰۲۱)



شکل ۹. مقاله **Agrawal** و همکاران (۲۰۱۹)

نتایج بدست آمده از مقایسات کلی برای روش پیشنهادی با مقالات Ahmad و همکاران (۲۰۲۳)، Ayesha و همکاران (۲۰۲۱) و Agrawal و همکاران (۲۰۱۹) در جدول های ۲ تا ۵ آورده شده است.

جدول ۲. نتایج روش پیشنهادی

MEAN		۱k	۲k	۳k	۴k
۰/۰۷۴۵	نوع کلاس بندی اشتباه	۰/۱۳۸	۰/۰۴۴	۰/۰۵۴	۰/۰۶۲
۰/۱۰۰۶۲۲	میزان هزینه کلاس بندی اشتباه	۰/۱۶۰۴۰۱۵۷	۰/۰۶۷۵۸۰۲	۰/۰۸۰۲۵۵	۰/۰۹۴۲۵۳
۰/۰۵۰۳۱۱	نرمال شده میزان هزینه کلاس بندی اشتباه	۰/۰۸۰۲۰۰۷	۰/۰۳۳۷۹۰۱	۰/۰۴۰۱۲۷	۰/۰۴۷۱۲۶
۰/۹۰۰۲۳۳	حساسیت	۰/۷۴۷۸۹۹۱۶	۰/۹۶۵۱۷۴۱	۰/۹۴۲۵۸۴	۰/۹۴۵۲۷۴
۰/۹۴۹۲۶۱	نرخ اختصاصی	۰/۹۶۵۶۴۸۸۵	۰/۹۴۹۸۳۲۸	۰/۹۴۸۴۵۴	۰/۹۳۳۱۱
۰/۹۲۵۵	صحت	۰/۸۶۲	۰/۹۵۶	۰/۹۴۶	۰/۹۳۸
۰/۹۲۸۵۲۷	دقت	۰/۹۵۱۸۷۱۶۶	۰/۹۲۸۲۲۹۷	۰/۹۲۹۲۴۵	۰/۹۰۴۷۶۲
۰/۹۰۰۲۳۳	فراخوانی مجدد	۰/۷۴۷۸۹۹۱۶	۰/۹۶۵۱۷۴۱	۰/۹۴۲۵۸۴	۰/۹۴۵۲۷۴
۰/۹۱۱۱۰۷	معیار ترکیبی F	۰/۸۳۷۶۴۷۰	۰/۹۴۶۳۴۱۵	۰/۹۳۵۸۶۷	۰/۹۲۴۵۷۴
۰/۸۱۷۶۲۱	سازگاری	۰/۵۱۸۸۹۱۵۳	۰/۹۴۱۷۶۲۸	۰/۹۰۱۳۴۷	۰/۹۰۸۴۸۴
۰/۹۲۴۷۴۷					
	AUC	۰/۸۵۶۷۷۴۰۱	۰/۹۵۷۵۰۳۵	۰/۹۴۵۵۱۹	۰/۹۳۹۱۹۲
۰/۹۱۵۳۰۹	معیار توازن	۰/۸۲۰۰۹۰۵۳	۰/۹۵۶۸۱۶۷	۰/۹۴۵۴۴	۰/۹۳۸۸۸۹
۲/۵۱۳۳۳۳	زمان اجرا	۲/۶۲۰۴۷۰۳۴	۲/۷۵۰۳۱۴۳	۲/۳۹۲۰۶۱	۲/۲۹۰۴۸۷

جدول ۳. نتایج مقاله (Ahmad و همکاران،۲۰۲۳)

MEAN		k۱	k۲	k۳	k۴
۰/۴۳۲۵۴۴	نرخ کلاس بندی اشتباه	۰/۴۳۴۲۱۰۵۳	۰/۳۹۴۷۳۶۸	۰/۴۰۷۸۹۵	۰/۴۹۳۳۳۳
۰/۵۹۵۰۸۶	میزان هزینه کلاس بندی اشتباه	۰/۶۲۱۳۱۱۴۸	۰/۶۴۶۸۵۳۱	۰/۴۵۸۳۳۳	۰/۶۵۳۸۴۶
۰/۲۹۷۵۴۳	نرمال شده میزان هزینه کلاس بندی اشتباه	۰/۳۱۰۶۵۵۷۴	۰/۳۲۳۴۲۶۶	۰/۲۲۹۱۶۷	۰/۳۲۶۹۲۳
۰/۵۶۲۱۰۱	حساسیت	۰/۵۵۷۳۷۷۰۵	۰/۶۱۵۳۸۴۶	۰/۵۸۳۳۳۳	۰/۴۹۲۳۰۸
۰/۶۲۳۸۶۴	نرخ اختصاصی	۰/۶	۰/۵۴۵۴۵۴۵	۰/۷۵	۰/۶
۰/۵۶۷۴۵۶	دقت	۰/۵۶۵۷۸۹۴۷	۰/۶۰۵۲۶۳۲	۰/۵۹۲۱۰۵	۰/۵۰۶۶۶۷
۰/۹۰۱۱۳	صحت	۰/۸۵	۰/۸۸۸۸۸۸۹	۰/۹۷۶۷۴۴	۰/۸۸۸۸۸۹
۰/۵۶۲۱۰۱	فراخوانی مجدد	۰/۵۵۷۳۷۷۰۵	۰/۶۱۵۳۸۴۶	۰/۵۸۳۳۳۳	۰/۴۹۲۳۰۸
۰/۶۹۱۱۶	معیار ترکیبی F	۰/۶۷۳۲۶۷۳۳	۰/۷۲۷۲۷۲۷	۰/۷۳۰۴۳۵	۰/۶۳۳۶۶۳
-۳/۱۵۶۰۸	سازگاری	-۱/۲۴۲۶۲۳	-۱/۶۵۷۳۴۲۷	-۶/۹۱۶۶۷	-۲/۸۰۷۶۹
۰/۵۹۲۹۸۲	AUC	۰/۵۷۸۶۸۸۵۲	۰/۵۸۰۴۱۹۶	۰/۶۶۶۶۶۷	۰/۵۴۶۱۵۴
۰/۵۸۹۱۲۳	معیار توازن	۰/۵۷۸۱۴۹۸۶	۰/۵۷۸۹۶۵۲	۰/۶۵۶۴۰۸	۰/۵۴۲۹۷۱
۲۲/۲۱۱۹۸	زمان اجرا	۲۷/۸۵۲۸۰۶۸	۲۳/۶۳۰۴۳۹	۲۴/۰۵۷۲۷	۱۳/۳۰۷۳۹

جدول۴. نتایج مقاله(Ayesha و همکاران،۲۰۲۱)

MEAN		k۱	k۲	k۳	k۴
۰/۲۵۴۱۶۷	نرخ کلاس بندی اشتباه	۰/۲۲۳۶۸۴۲۱	۰/۱۱۸۴۲۱۱	۰/۴۰۷۸۹۵	۰/۲۶۶۶۶۷
۰/۳۱۴۷۳۸	میزان هزینه کلاس بندی اشتباه	۰/۳۲۸۴۰۰۲۸	۰/۲۱۲۱۸۴۹	۰/۴۵۸۳۳۳	۰/۲۶۰۰۳۳
۰/۱۵۷۳۶۹	نرمال شده میزان هزینه کلاس بندی اشتباه	۰/۱۶۴۲۰۰۱۴	۰/۱۰۶۰۹۲۴	۰/۲۲۹۱۶۷	۰/۱۳۰۰۱۷
۰/۷۳۳۲۹۸	حساسیت	۰/۷۶۷۴۴۱۸۶	۰/۹۲۸۵۷۱۴	۰/۵۸۳۳۳۳	۰/۶۵۳۸۴۶
۰/۸۱۸۶۱۳	نرخ اختصاصی	۰/۷۸۷۸۷۸۷۹	۰/۸۲۳۵۲۹۴	۰/۷۵	۰/۹۱۳۰۴۳
۰/۷۴۵۸۳۳	صحت	۰/۷۷۶۳۱۵۷۹	۰/۸۸۱۵۷۸۹	۰/۵۹۲۱۰۵	۰/۷۳۳۳۳۳
۰/۹۰۳۲۱۴	دقت	۰/۸۲۵	۰/۸۶۶۶۶۶۷	۰/۹۷۶۷۴۴	۰/۹۴۴۴۴۴
۰/۷۳۳۲۹۸	فراخوانی مجدد	۰/۷۶۷۴۴۱۸۶	۰/۹۲۸۵۷۱۴	۰/۵۸۳۳۳۳	۰/۶۵۳۸۴۶
۰/۷۹۸۷۲۴	معیار ترکیبی F	۰/۷۹۵۱۸۰۷۲	۰/۸۹۶۵۵۱۷	۰/۷۳۰۴۳۵	۰/۷۷۲۷۲۷
-۱/۴۳۵۱۷	سازگاری	۰/۴۶۴۴۱۱۵۶	۰/۸۴۰۳۳۶۱	-۶/۹۱۶۶۷	-۰/۱۲۸۷۶
۰/۷۷۵۹۵۶	AUC	۰/۷۷۷۶۶۰۳۲	۰/۸۷۶۰۵۰۴	۰/۶۶۶۶۶۷	۰/۷۸۳۴۴۵
۰/۷۶۱۷۱۱	معیار توازن	۰/۷۷۷۴۲۵۶۳	۰/۸۶۵۳۸۲۲	۰/۶۵۶۴۰۸	۰/۷۴۷۶۲۷
۹/۷۷۳۶۹۶	زمان اجرا	۴/۴۴۶۳۷۸۵۳	۴/۴۱۳۵۷۹۶	۲۶/۰۳۷۳۷	۴/۱۹۷۴۵۶

جدول ۵. نتایج مقاله [Agrawal و همکاران،۲۰۱۹)

MEAN		k۱	k۲	k۳	k۴
۰/۲۰۴۶۹۳	نرخ کلاس بندی اشتباه	۰/۲۵	۰/۱۳۱۵۷۸۹	۰/۲۱۰۵۲۶	۰/۲۲۶۶۶۷
۰/۲۸۷۱۲	میزان هزینه کلاس بندی اشتباه	۰/۳۵۹۱۳۹۷۸	۰/۲۱۶۸۴۵۹	۰/۳۴۲۵۰۹	۰/۲۲۹۹۸۴
۰/۱۴۳۵۶	نرمال شده میزان هزینه کلاس بندی اشتباه	۰/۱۷۹۵۶۹۸۹	۰/۱۰۸۴۲۲۹	۰/۱۷۱۲۵۴	۰/۱۱۴۹۹۲
۰/۷۸۶۳۴۲	حساسیت	۰/۷۳۳۳۳۳۳۳	۰/۸۸۸۸۸۸۹	۰/۸۲۹۲۶۸	۰/۶۹۳۸۷۸
۰/۸۱۹۷۰۹	نرخ اختصاصی	۰/۷۷۴۱۹۳۵۵	۰/۸۳۸۷۰۹۷	۰/۷۴۲۸۵۷	۰/۹۲۳۰۷۷
۰/۷۰۹۱۲۵	صحت	۰/۷۱	۰/۶۹	۰/۷۲	۰/۷۱۶۵
۰/۸۶۲۲۵۸	دقت	۰/۸۲۵	۰/۸۸۸۸۸۸۹	۰/۷۹۰۶۹۸	۰/۹۴۴۴۴۴
۰/۷۸۶۳۴۲	فراخوانی مجدد	۰/۷۳۳۳۳۳۳۳	۰/۸۸۸۸۸۸۹	۰/۸۲۹۲۶۸	۰/۶۹۳۸۷۸
۰/۸۱۸۷۲۱	معیار ترکیبی F	۰/۷۷۶۴۷۰۵۹	۰/۸۸۸۸۸۸۹	۰/۸۰۹۵۲۴	۰/۸
۰/۴۵۵۰۱۴	سازگاری	۰/۳۴۶۲۳۶۵۶	۰/۷۲۷۵۹۸۶	۰/۶۲۹۲۶۸	۰/۱۱۶۹۵۴
۰/۸۰۳۰۲۶	AUC	۰/۷۵۳۷۶۳۴۴	۰/۸۶۳۷۹۹۳	۰/۷۸۶۰۶۳	۰/۸۰۸۴۷۷
۰/۷۹۳۲۴۴	معیار توازن	۰/۷۵۲۹۱۷۳۶	۰/۸۶۱۵۰۷۷	۰/۷۸۱۷۴۴	۰/۷۷۶۸۰۹
۴/۰۸۰۳۳۱	زمان اجرا	۰/۲۰۲۶۹۷۸۶	۴/۳۱۸۴۴۲۸	۳/۵۹۱۴۴۱	۴/۲۰۸۷۴۲

در این فصل با توجه به موضوع مورد پژوهش، در ابتدا انواع شاخص‌های ارزیابی شرح داده شد و در ادامه به مرور عملکرد سه روش پیشنهادی با ۳ مقاله دیگر پرداختیم در نهایت بهترین ترکیب در این بین روش پیشنهادی با صحت نزدیک ۹۲ درصد و دقت ۹۳ درصد است.

بحث و نتیجه‌گیری

در نتیجه، این تحقیق روشی را پیشنهاد می‌کند که یک شبکه مولد آموزش‌دیده با استفاده از یادگیری تقویتی را با یک شبکه تفکیک کننده مبتنی بر CNN برای شرح تصاویر پزشکی ترکیب می‌کند. هدف شبکه مولد ایجاد زیرنویس‌های دقیق برای تصاویر پزشکی است، در حالی که شبکه تفکیک کننده کیفیت زیرنویس‌های تولید شده را ارزیابی می‌کند. این تحقیق نشان می‌دهد که این رویکرد ترکیبی می‌تواند در بهبود دقت و صحت زیرنویس‌های تولید شده مؤثر باشد. به طور کلی، این تحقیق با استفاده از قدرت شبکه های مولد و متمایز به پیشرفت تکنیک های شرح تصاویر پزشکی کمک می کند.

در اینجا برخی از وظایف بالقوه آینده وجود دارد که می توان آنها را بررسی کرد:

بهینه سازی عملکرد: روش های بهبود سرعت و کارایی روش پیشنهادی را بررسی کنید. این می تواند شامل کاوش در معماری های جایگزین یا تکنیک های آموزشی باشد که می تواند الزامات محاسباتی را کاهش دهد و در عین حال دقت زیرنویس های تولید شده را حفظ یا بهبود بخشد.

ارزیابی انسانی و مطالعات کاربر: انجام مطالعات ارزیابی انسانی کامل برای ارزیابی کیفیت و سودمندی زیرنویس‌های ایجاد شده. این می‌تواند شامل جمع‌آوری بازخورد از متخصصان پزشکی یا متخصصان حوزه باشد تا صحت و ارتباط زیرنویس‌ها را تأیید کند. مطالعات کاربر همچنین می تواند بینشی در مورد قابلیت استفاده عملی و پذیرش روش پیشنهادی در سناریوهای دنیای واقعی ارائه دهد.

یکپارچه سازی با سیستم های بالینی: ادغام روش پیشنهادی را در سیستم های بالینی یا پلت فرم های تصویربرداری پزشکی موجود بررسی کنید. این شامل تطبیق روش برای کار یکپارچه با گردش کار مراقبت های بهداشتی و ارزیابی تأثیر آن بر بهبود تصمیم گیری بالینی یا کمک به متخصصان پزشکی در وظایفشان است.

با تمرکز بر این وظایف آینده، این تحقیق می‌تواند زمینه شرح تصاویر پزشکی را بیشتر پیش ببرد و به توسعه سیستم‌های دقیق‌تر و قابل اعتمادتر برای کمک به متخصصان پزشکی در کارشان کمک کند.

منابع

1. Agrawal, Harsh, Desai, Karan, Wang, Yufei, Chen, Xinlei, Jain, Rishabh, Johnson, Mark, Batra, Dhruv, Parikh, Devi, Lee, Stefan. and Anderson, Peter. (2019). Nocaps: novel object captioning at scale. In: Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, pp 8948–8957.
2. Ahmad, Rana Adnan, Azhar, Muhammad. and Sattar, Hina. (2022). An Image captioning algorithm based on the Hybrid Deep Learning Technique (CNN+ GRU). In 2022 International Conference on Frontiers of Information Technology (FIT) (pp. 124-129). IEEE.
3. Allaouzi, Imane, Ben Ahmed, M, Benamrou, Badr. and Mustapha, Ouardouz (2018) Automatic caption generation for medical images. In: Proceedings of the 3rd international conference on smart city applications (SCA'18)
4. Alsharid, Mohammad, El-Bouri, Rasheed, Sharma, Harshita, Drukker, Lior, Papageorghiou, Aris T., and Noble, J. Alison. (2020) A curriculum learning based approach to captioning ultrasound images. In: Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis.
5. Ayesha, Hareem, Iqbal, Sajid, Tariq, Mehreen, Abrar, Muhammad, Sanaullah, Muhammad, Abbas, Ishaq, Rehman, Amjad, Farooq Khan Niazi, Muhammad. and Hussain, Shafiq. (2021). Automatic medical image interpretation: State of the art and future directions. Pattern Recognition. 114 (2021). 107856.
6. Beddiar, Djamila Romaissa, Oussalah, Mourad, Seppänen, Tapio. and Jennane, Rachid. (2022). ACapMed: Automatic Captioning for Medical Imaging. Applied Sciences. Vol. 12, No. 21. 11092.
7. Djamila-Romaissa, Beddiar, Oussalah, Mourad. and Seppänen, Tapio. (2021). Attention-based CNN-GRU model for automatic medical images captioning: Imageclef 2021. In: Proceedings of the working notes of CLEF 2021—conference and labs of the evaluation forum, Bucharest, Romania, pp. 1160–1173
8. Pavlopoulos, John, Kougia, Vasiliki, Androutsopoulos, Ion. (2019). A survey on biomedical image captioning. In: Proceedings of the Second Workshop on Shortcomings in Vision and Language, pp 26–36
9. Selivanov, Alexander, Rogov, Oleg, Y., Chesakov, Daniil, Shelmanov, Artem, Fedulova, Irina. and Dylov. Dmitry V. (2022). Medical image captioning via generative pretrained transformers. Scientific Reports. Vol. 13, No. 1. 4171.
10. Singh, Arjun, Krishna Raguru, Jaya, Prasad, Gaurav, Chauhan, Surbhi, Kumar Tiwari, Pradeep, Zaguia, Atef. and Ullah Mohammad, Aman. (2022a). Medical image captioning using optimized deep learning model. Computational Intelligence and Neuroscience 2022.
11. Singh, Dilbag,, Kaur, Manjit, Alanazi, Jazem Mutared, AlZubi, Ahmad Ali, and Lee, Heung-No. (2022b). Efficient evolving deep ensemble medical image captioning network. IEEE Journal of Biomedical and Health Informatics. Vol. 27, No. 2 1016-1025.
12. Tian, Jiang, Zhong, Cheng, Shi, Zhongchao. and Xu, Feiyu. (2020). Towards automatic diagnosis from multi-modal medical data. terpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support: Second International Workshop, iMIMIC 2019, and 9th International Workshop, ML-CDS 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Proceedings 9 (pp. 67-74). Springer International Publishing.
13. Tumuramye, Ronald. (2023). Deep learning-aided image captioning in chest X-rays for TB screening. PhD dissertation., Makerere University.
14. Xue, Yuan, Huang, Xiaolei. (2019). Improved disease classification in chest x-rays with transferred features from report generation. In: 26th international conference on information processing in medical imaging, pp. 125–138, Springer International Publishing.
15. Yang, Yan, Yu, Jun, Zhang, Jian, Han, Weidong, Jiang, Hanliang. and Huang, Qingming. (2021). Joint embedding of deep visual and semantic features for medical image report generation. IEEE Transactions on Multimedia, Vol. 25. 167-178.
16. Zeng, Xianhua, Wen, Li, Xu, Yang. and Ji, Conghui. (2020). Generating diagnostic report for medical image by high-middle-level visual information incorporation on double deep learning models. Computer methods and programs in biomedicine, Vol. 197. 105700.
17. Zhang, Junsan, Shen, Xiuxuan, Wan, Shaohua, Goudos, Sotirios K., Wu, Jie, Cheng, Ming. and Zhang, Weishan. (2023). A novel deep learning model for medical report generation by inter-intra information calibration. IEEE Journal of Biomedical and Health Informatics.

22 th International Conference on
Information Technology, Computer and Telecommunication
Event Place: Tbilisi,Georgia
www.itctconf.ir
بیست و دومین کنفرانس بین المللی
فناوری اطلاعات، کامپیوتر و مخابرات | گرجستان
22th International Conference on Information Technology,Computer and Telecommunication
۲۶ اسفندماه ۱۴۰۲ ——— مجلات معتبر بین المللی

Investigating the efficiency of improved generative adversarial networks and reinforcement learning to increase the accuracy in describing medical images with clinical style

sajjad Zarini

Masters student computer major artificial intelligence and robotics design Islamic Azad University Tehran branch

Parvane Asghari[1]

PhD Computer Software Orientation Assistant Professor Faculty of Engineering Islamic Azad University Tehran Branch

Abstract

Analysis of medical images plays an important role in the diagnosis and treatment of various diseases. However, the process of analyzing these images and generating detailed reports can be time-consuming and subject to human error. In this study, a new approach is proposed that combines adversarial generative networks and reinforcement learning to improve the efficiency of medical image analysis and corresponding report writing. This method involves training an adversarial generative network to generate real medical images based on a certain set of input parameters. This network is then integrated with a reinforcement learning framework to learn and optimize the analysis process of these images. Through iterative training, the system is able to produce high-quality medical images that accurately represent various pathological conditions. Using the powerful capabilities of adversarial networks and reinforcement learning, our system can effectively capture important features and anomalies in medical images, leading to more accurate and comprehensive reports. The evaluation of the results showed the effectiveness of the proposed method with 92% accuracy and 93% accuracy compared to other methods.

Keywords: Adversarial generative networks, Medical image analysis, Report generation, Reinforcement learning, Efficiency

[1] Corresponding Author