Journal of
Babol University of Medical Sciences
مجله علمی دانشگاه علوم پزشکی بابل
e-ISSN: 2251-7170 p-ISSN: 1561-4107

Enhancing the Interpretability of Cervical Cancer Diagnosis: Refining Geometry-Based Heatmaps Using Ricci Flow on the SIPaKMeD Dataset

F. Abbasi Varaki (MD)[*1], A. Ariyaei Motahar (MD)[1], M. Mohammadian Amiri (MD)[2]

1.School of Medicine, Iran University of Medical Sciences, Tehran, I.R.Iran.
2.Shahid Akbarabadi Clinical Research Development Unit (ShACRDU), Iran University of Medical Sciences, Tehran, I.R.Iran.

•**Corresponding Author: F. Abbasi Varaki (MD)**
Address: School of Medicine, Iran University of Medical Sciences, Tehran, I.R.Iran.
Tel: +98 (26) 32205994. **E-mail:** farhanabbasip7@gmail.com

Article Type
Research Paper

Received:
Dec 8th 2024
Revised:
Jan 25th 2025
Accepted:
Feb 4th 2025

ABSTRACT

Background and Objective: Cervical cancer detection is a critical task where accurate and interpretable diagnostic tools can significantly enhance clinical outcomes. Existing explainability methods, such as Grad-CAM, provide visual insights into model predictions but often lack precision in highlighting diagnostically relevant regions. To address this, we integrated Ricci Flow into the gradient computation process, aiming to refine the geometric structure of heatmaps and improve their interpretability.

Methods: In this methodological study, the SIPaKMeD dataset, comprising 4,049 cervical smear images across five diagnostic categories, was preprocessed and augmented to address class imbalance and ensure data uniformity. A ResNet50 model was trained on the dataset for 40 epochs using a balanced set of whole-slide images. Ricci Flow was applied to the gradient matrix, interpreted as a Riemannian tensor, to iteratively smooth and refine the geometry of the gradient space. Heatmap quality was evaluated through supervised comparison with cropped cell images and by computing insertion and deletion metrics to quantify the alignment of heatmaps with diagnostically critical regions.

Findings: The Ricci Flow-enhanced Grad-CAM outperformed the standard Grad-CAM, achieving an insertion AUC of 0.671 and a deletion AUC of 0.153. The refined heatmaps consistently demonstrated a sharper focus on diagnostically important regions, with over 90 percent alignment with expert annotations. Additionally, the Ricci Flow-based method provided a deeper geometric understanding of the feature space, emphasizing the region's most critical for classification.

Conclusion: The results of the study demonstrated that integrating Rich Flow into Grad-CAM, utilizing geometric smoothing, enhances the interpretability of heatmaps and offers a novel approach in the field of Explainable Artificial Intelligence (XAI) for cervical cancer diagnosis. This method not only improves model accuracy but also aligns with the hypothesis that neural networks alter data topology during training.

Keywords: *Ricci Flow, Explainable AI, Uterine Cervical Neoplasms, Heatmap Interpretability, SIPaKMeD Dataset, Deep Learning.*

Cite this article: Abbasi Varaki F, Ariyaei Motahar A, Mohammadian Amiri M. Enhancing the Interpretability of Cervical Cancer Diagnosis: Refining Geometry-Based Heatmaps Using Ricci Flow on the SIPaKMeD Dataset. *Journal of Babol University of Medical Sciences*. 2026; 28: e33.

Journal of
Babol University of Medical Sciences
مجله علمی دانشگاه علوم پزشکی بابل
e-ISSN: 2251-7170 p-ISSN: 1561-4107

بهبود قابلیت تفسیر در تشخیص سرطان دهانه رحم: اصلاح نقشه حرارتی مبتنی بر هندسه از طریق جریان Ricci در مجموعه داده سیپاکمِد

فرحان عباسی ورکی (MD)[1*]، علی آریایی مطهر (MD)[1]، مهدیس محمدیان امیری (MD)[2]

۱. دانشکده پزشکی، دانشگاه علوم پزشکی ایران، تهران، ایران
۲. واحد توسعه تحقیقات بالینی شهید اکبرآبادی (ShACRDU)، دانشگاه علوم پزشکی ایران، تهران، ایران

نوع مقاله
مقاله پژوهشی

دریافت:
۱۴۰۳/۹/۱۸

اصلاح:
۱۴۰۳/۱۱/۶

پذیرش:
۱۴۰۳/۱۱/۱۶

چکیده

سابقه و هدف: تشخیص سرطان دهانه رحم با استفاده از ابزارهای تشخیصی دقیق و قابل تفسیر می‌توانند به طور چشمگیری نتایج بالینی را بهبود بخشند. روش‌های موجود برای تبیین پذیری (مانند Grad-CAM)، بینش‌های بصری درباره پیش بینی‌های مدل ارائه می‌دهند، اما اغلب در برجسته سازی مناطق مرتبط با تشخیص، دقت کافی را ندارند. برای رفع این مسئله، ما جریان ریچی (Ricci Flow) را در فرآیند محاسبه گرادیان ادغام کردیم تا ساختار هندسی نقشه‌های حرارتی (Heatmaps) را اصلاح و قابلیت تفسیر آن‌ها را افزایش دهیم.

مواد و روش‌ها: در این مطالعه روش شناختی، مجموعه داده SIPaKMeD متشکل از پنج دسته تصاویر اسمیر دهانه رحم (۴۰۴۹ تصویر)، پیش پردازش و افزایش داده شد تا عدم تعادل کلاس ها رفع و یکنواختی تصاویر تضمین شود. یک مدل ResNet50 با آموزش روی این مجموعه داده به مدت ۴۰ دوره (epoch) با استفاده از مجموعه‌های متعادل از تصاویر تماماسلاید (whole-slide) توسعه یافت. جریان ریچی بر ماتریس گرادیان اعمال شد که به عنوان یک تانسور ریمانی تفسیر گردید تا هندسه فضای گرادیان را به صورت تکراری هموار و اصلاح کند. کیفیت نقشه‌های حرارتی از طریق مقایسه نظارت شده با تصاویر برش خورده سلولی و محاسبه معیارهای درج (Insertion) و حذف (Deletion) ارزیابی شد تا همترازی نقشه‌های حرارتی با مناطق بحرانی تشخیصی سنجیده شود.

یافته‌ها: روش Grad-CAM بهبود یافته با جریان ریچی، عملکرد بهتری نسبت به Grad-CAM استاندارد نشان داد و به مقادیر AUC 0.671 در معیار درج و AUC 0.153 در معیار حذف دست یافت. نقشه‌های حرارتی اصلاح شده به طور پیوسته تمرکز دقیق‌تری بر مناطق مهم تشخیصی داشتند و بیش از ۹۰٪ همخوانی با حاشیه نویسی‌های متخصصان نشان دادند. علاوه بر این، روش مبتنی بر جریان ریچی، درک هندسی عمیق‌تری از فضای ویژگی ارائه داد و مناطق حیاتی برای طبقه بندی را پررنگ‌تر کرد.

نتیجه‌گیری: نتایج مطالعه نشان داد که ادغام جریان ریچی در Grad-CAM، با استفاده از هموارسازی هندسی، قابلیت تفسیر نقشه‌های حرارتی را افزایش می‌دهد و رویکردی نوین در حوزه هوش مصنوعی تبیین پذیر (XAI) برای تشخیص سرطان دهانه رحم ارائه می‌کند. این روش نه تنها دقت مدل را بهبود می‌بخشد، بلکه با فرضیه تغییر توپولوژی داده‌ها توسط شبکه‌های عصبی در طول آموزش همسو است.

واژه‌های کلیدی: جریان ریچی، هوش مصنوعی توضیح‌پذیر، نئوپلاسم‌های دهانه رحم، تفسیرپذیری نقشه حرارتی، مجموعه داده SIPaKMeD، یادگیری عمیق.

استناد: فرحان عباسی ورکی، علی آریایی مطهر، مهدیس محمدیان امیری. بهبود قابلیت تفسیر در تشخیص سرطان دهانه رحم: اصلاح نقشه حرارتی مبتنی بر هندسه از طریق جریان Ricci در مجموعه داده سیپاکمِد. مجله علمی دانشگاه علوم پزشکی بابل. ۱۴۰۵؛ ۲۸: ۳۳ e.

▣ این مقاله مستخرج از طرح تحقیقاتی دانشگاه علوم پزشکی ایران می‌باشد.

* مسئول مقاله: دکتر فرحان عباسی ورکی
آدرس: تهران، دانشگاه علوم پزشکی ایران دانشکده پزشکی. تلفن: ۰۲۶-۳۲۲۰۵۹۹۴
رایانامه: farhanabbasip7@gmail.com

مقدمه

سرطان دهانه رحم، به عنوان یک نگرانی مهم در حوزه سلامت جهانی، همواره مورد توجه تحقیقات پزشکی بوده است (۱). تشخیص زودهنگام و مداخله به موقع برای بهبود نتایج درمان بیماران حیاتی است. تست پاپ اسمیر، یک روش غربالگری شناخته شده، شامل جمع آوری سلول‌های دهانه رحم و بررسی آن‌ها زیر میکروسکوپ برای شناسایی ناهنجاری‌ها است (۲). اگرچه این روش مؤثر است، اما فرآیند دستی آن زمان‌بر بوده و ممکن است هنگام مواجهه با تغییرات ظریف سلولی دچار خطاهای انسانی شود (۱).

برای رفع این محدودیت‌ها و بهبود دقت و کارایی غربالگری سرطان دهانه رحم، سیستم‌های تشخیص به کمک کامپیوتر (CAD) به عنوان یک راه حل امیدوار کننده مطرح شده‌اند. این سیستم‌ها از تکنیک‌های پیشرفته یادگیری ماشین و یادگیری عمیق برای تحلیل تصاویر دیجیتالی سلول‌های دهانه رحم و شناسایی ناهنجاری‌های احتمالی استفاده می‌کنند. با خودکار سازی فرآیند تحلیل، سیستم‌های CAD می‌توانند بار کاری پاتولوژیست‌ها را کاهش داده و ثبات نتایج تشخیصی را بهبود بخشند (۴و۳).

یکی از چالش‌های کلیدی در توسعه سیستم‌های CAD مؤثر برای غربالگری سرطان دهانه رحم، دسترسی به مجموعه داده‌های بزرگ و باکیفیت است. مجموعه داده SIPaKMeD، به عنوان یک منبع ارزشمند در این حوزه، شامل ۴۰۴۹ تصویر حاوی سلول‌های جدا شده دهانه رحم است که به‌صورت دستی حاشیه‌نویسی شده‌اند (۵).

شبکه عصبی کانولوشنال عمیق ResNet50، که یک شبکه پیشرفته در حوزه یادگیری عمیق است، در کارهای مختلف طبقه بندی تصاویر، از جمله تحلیل تصاویر پزشکی، عملکرد بسیار مؤثری داشته است. معماری باقیمانده (Residual) این شبکه امکان آموزش شبکه‌های عمیق‌تر را فراهم کرده و منجر به بهبود عملکرد می‌شود. با اعمال ResNet50 روی مجموعه داده SIPaKMeD، محققان می‌توانند ویژگی‌های معنی‌دار را از الگوهای پیچیده در تصاویر سلول‌های دهانه رحم استخراج کرده و پیش‌بینی‌های دقیقی انجام دهند (۷و۶).

روش‌های تفسیرپذیری در هوش مصنوعی: روش‌هایی مانند Grad-CAM (نقشه‌های فعال‌سازی کلاس‌وزن‌دار گرادیان) و نقشه‌های سالینسی (Saliency Maps) به‌طور گسترده در هوش مصنوعی تفسیرپذیر (XAI) برای ارائه توضیحات بصری از تصمیمات مدل‌های یادگیری عمیق، به‌ویژه در کارهای بینایی کامپیوتر، استفاده شده‌اند. این تکنیک‌ها مناطق مؤثر در پیش‌بینی‌های مدل را برجسته می‌کنند و درک و تفسیر رفتار مدل را برای کاربران آسان‌تر می‌سازند (۸).

با این حال، Grad-CAM ممکن است به دلیل حساسیت به نویز، نقشه‌های حرارتی (Heatmaps) پراکنده و ناقص تولید کند (۹). Grad-CAM بر مناطق قبل از ورود داده به لایه‌های پنهان تأکید می‌کند، بنابراین نمی‌توان مطمئن بود که نقشه‌های حرارتی تولید شده مناطقی را نشان می‌دهند که مدل بر اساس آن‌ها تصمیم می‌گیرد. این موضوع باعث می‌شود اتکای بالینی به مدل دشوار شود و نتوان مطمئن بود که مدل بیش‌برازش (Overfitting) شده یا بر روی مجموعه داده‌های بالینی مشابه قابل تعمیم است (۱۰).

بر اساس یک فرضیه، شبکه‌های عصبی داده‌ها را با تغییر توپولوژی بردارهای اولیه وارد شده به شبکه دسته‌بندی می‌کنند (۱۲و۱۱). برخی محققان پیشنهاد می‌کنند که تفاوت بین هر لایه در یک شبکه عصبی به عنوان ضریب انحنای ریچی (Ricci Curvature) در نظر گرفته شود (۱۳). جریان ریچی (Ricci Flow) نوعی معادله دیفرانسیل با مشتقات جزئی (PDE) است که یک منیفولد ریمانی (Riemannian Manifold) را به‌صورت پیوسته به شکل هموارتری تغییر می‌دهد.

با در نظر گرفتن فضاهای توپولوژیک در یک شبکه عصبی به عنوان یک منیفولد ریمانی، متریک ریمانی روشی برای تعریف فاصله‌ها روی منیفولد با توجه به شکل آن است (۱۳). برای رفع این چالش، ما جریان ریچی را معرفی می‌کنیم تا گرادیان‌ها را هموار کرده و انسجام را افزایش دهد، که منجر به تصاویر بصری هندسی‌تر و سازگارتر می‌شود. جریان ریچی گرادیان‌ها را به عنوان یک تانسور متریک بر اساس انحنای ریچی تنظیم می‌کند و با پخش کردن مناطق با انحنای بالا و فشرده سازی مناطق صاف‌تر، هندسه را یکنواخت‌تر می‌کند (۱۴).

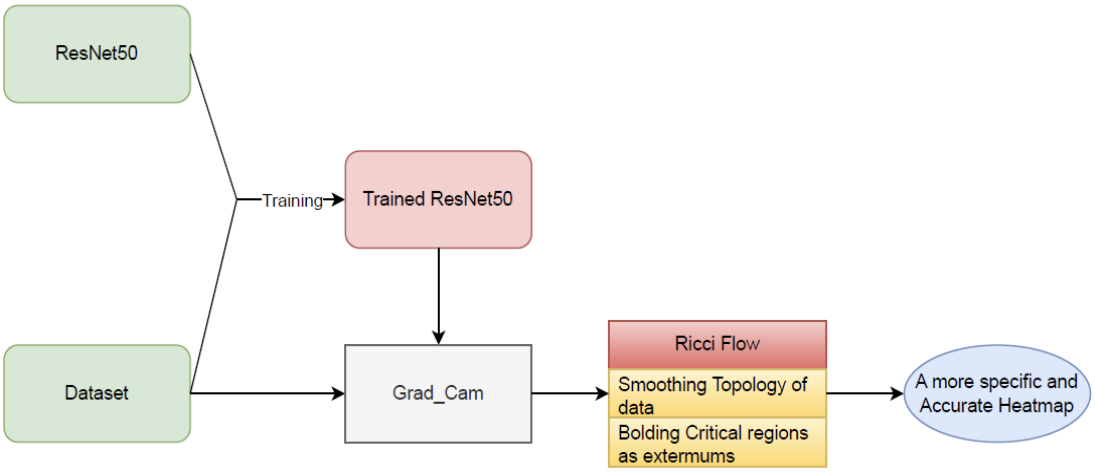
ادغام جریان ریچی در Grad-CAM یک نوآوری پیشگامانه در حوزه هوش مصنوعی تفسیرپذیر، به‌ویژه در تصویربرداری پزشکی است. این روش نه تنها دقت و تفسیرپذیری نقشه‌های حرارتی را بهبود می‌بخشد، بلکه یک چارچوب هندسی برای درک فرآیند تصمیم‌گیری هوش مصنوعی ارائه می‌دهد.

در این تحقیق، هدف ما بهبود وضوح و قابلیت اطمینان تصاویر بصری Grad-CAM است تا آن‌ها را با مناطق بالینی مرتبط در کارهای تصویربرداری پزشکی، مانند تشخیص سرطان دهانه رحم، هم‌تراز کنیم.

مواد و روش ها

این مطالعه تحلیلی گذشته‌نگر بر روی داده‌های تصویربرداری پزشکی از مجموعه داده SIPaKMeD انجام شده است که شامل تصاویر سلول‌های دهانه رحم است و بر اساس ویژگی‌های مورفولوژیکی آن‌ها دسته‌بندی شده‌اند. تصاویر در قالب .bmp ذخیره شده‌اند که یک فرمت تصویری بدون اتلاف است و جزئیات ویژگی‌های سلولی را به خوبی حفظ می‌کند. این مجموعه داده شامل ۴۰۴۹ تصویر از سلول‌های جدا شده است که به‌صورت دستی از ۹۶۶ تصویر خوشه‌ای سلولی مربوط به اسلایدهای پاپ اسمیر برش داده شده‌اند. تصاویر برش شده بر روی انواع خاصی از سلول‌ها متمرکز هستند که امکان طبقه‌بندی دقیق‌تر را فراهم می‌کنند. از این تصاویر برش شده در تحلیل‌ها برای آموزش یک مدل جهت طبقه‌بندی سلول‌های دهانه رحم بر اساس ظاهر و ساختار آن‌ها استفاده شده است (۵). در این مطالعه، تصاویر در وضوح اصلی خود حفظ شدند که برای تحلیل دقیق مورفولوژی سلول‌های دهانه رحم کافی است.

ما از پایتون ۱۰.۳ در محیط Google Colab به عنوان پلتفرم و از پردازنده GPU T4 با ۱۵ گیگابایت رم استفاده کردیم. نمودار جریان کار این مطالعه در شکل ۱ قابل مشاهده است.



شکل ۱. نمودار جریان کار مطالعه

این نمودار جریان کار، روش‌شناسی end-to-end را نشان می‌دهد که از پیاده سازی مجموعه داده SIPaKMeD و ادغام جریان ریچی (Ricci Flow) در محاسبه گرادیان Grad-CAM شروع شده و تا تولید نقشه‌های حرارتی اصلاح شده ادامه می‌یابد.

پیش‌پردازش و افزایش داده‌ها: تصاویر با استفاده از نرمال سازی پیش‌پردازش شدند و به اندازه (۲۲۴، ۲۲۴) تغییر اندازه داده شدند. برای مقابله با مشکل عدم تعادل کلاس‌ها، از تکنیک‌های افزایش داده مانند چرخش، معکوس‌کردن و تغییر مقیاس استفاده شد تا اطمینان حاصل شود که هر یک از پنج دسته (Koilocytotic، Parabasal، Dyskeratotic، Superficial-Intermediate و Metaplastic) تعداد تصاویر یکسانی داشته باشند. تصاویری که دارای محو شدگی، آرتیفکت یا وضوح ناکافی بودند، حذف شدند تا از ایجاد نویز جلوگیری شود و دقت آموزش و ارزیابی مدل تضمین گردد.

معماری ResNet50: ResNet50 یک شبکه عصبی کانولوشنال عمیق (CNN) است که برای حل مشکل گرادیان‌های ناپدید شونده در شبکه‌های بسیار عمیق طراحی شده است. این شبکه از اتصالات باقیمانده (Residual Connections) استفاده می‌کند که به شبکه اجازه می‌دهد تا نگاشت‌های باقیمانده را به جای نگاشت‌های مستقیم یاد بگیرد و این امر آموزش شبکه‌های عمیق‌تر را تسهیل می‌کند.

Grad-CAM و الگوریتم آن: Grad-CAM (نقشه‌های فعال‌سازی کلاس‌وزن‌دار گرادیان) یک تکنیک برای تجسم مناطق یک تصویر است که بیشترین تأثیر را در تصمیم‌گیری یک شبکه عصبی کانولوشنال (CNN) دارند. این روش به‌ویژه برای توضیح تصمیمات مدل‌های یادگیری عمیق، به‌طور خاص در کارهای طبقه‌بندی تصاویر، مفید است. Grad-CAM یک نقشه حرارتی تولید می‌کند که مناطق مؤثر در پیش‌بینی یک کلاس خاص را برجسته می‌کند (۱۵).

ایده اصلی Grad-CAM این است که گرادیان نمره کلاس هدف را نسبت به نقشه‌های ویژگی لایه کانولوشنال نهایی محاسبه کرده و از این اطلاعات برای ایجاد یک نقشه موقعیت‌یابی تفکیک‌پذیر بر اساس کلاس استفاده کند. این نقشه نشان می‌دهد که مدل هنگام پیش‌بینی خود بر کدام بخش‌های تصویر تمرکز کرده است (۱۶).

در شبکه‌های عصبی پیچشی، ماتریس گرادیان $G_{ij}^{k} = \frac{\partial y_c}{\partial A_{ij}^{k}}$، نشان دهنده میزان حساسیت امتیاز خروجی ∂y_c به مارتیس‌های فعال سازی A_{ij}^{k} در موقعیت فضایی $(i\ ,\ j)$ در نقشه تعریف می‌باشد. این ماتریس گرادیان را می‌توان به عنوان یک تنسور ریمانی تعبیر کرد، جایی که گرادیان‌ها یک نمایش هندسی از اهمیت نواحی مکانی در فضای ویژگی‌ها را تشکیل می‌دهند.

با در نظر گرفتن G_{ij}^{k} به عنوان یک تنسور ریمانی، این امکان فراهم می‌شود که از ابزارهای هندسه دیفرانسیل، مانند جریان ریتچی (Ricci Flow)، برای تغییر ساختار این گرادیان‌ها استفاده شود. جریان ریتچی یک معادله دیفرانسیل جزئی (PDE) در هندسه است که برای هموارسازی تنسور متریک g_{ij} از یک خمینه به کار می رود:

$$\frac{\partial g_{ij}}{\partial t} = -2 \cdot Ric_{ij}$$

در این فرمول، Ric_{ij} نمایانگر تنسور انحنای ریچی است که میزان انحنا در هر نقطه از خمینه را اندازه‌گیری می‌کند. با تنظیم تکراری تنسور متریک به منظور کاهش نواحی با انحنای بالا، جریان ریتچی هندسه فضا را هموار می‌کند و نمایشی یکنواخت‌تر ایجاد می‌نماید (۱۴-۱۲).

ما جریان ریچی را بر G_{ij}^{k} قبل از محاسبه وزن‌ها برای Grad_CAM استفاده می‌کنیم. بعد از تغییر $G_{ij}^{k}(t)$ در چند t_{max} ماتریس گرادیان برای تبیین اهمیت وزن‌های α^{k} برای ایجاد Grad_CAM مورد استفاده قرار می‌گیرد (۱۸و۱۷).

$$\alpha^{k} = \frac{1}{H.W}\sum_{i=1}^{H}\sum_{j=1}^{W} G_{ij}^{k}(t_{max})$$

در اینجا H و W ابعاد فضایی نقشه‌های ویژگی‌ها α^{k} هستند که اهمیت ویژگی k- ام را نشان می‌دهند.

در روش ما میزان Ric_{ij}^{k} با استفاده از لاپلاس G_{ij}^{k} تقریب زدیم که به عنوان یک روش تقریب برای خم ریچی عمل می‌کند:

$$Ric_{ij}^{k} \approx \Delta G_{ij}^{k} = G_{i+1,j}^{k} + G_{i-1,j}^{k} + G_{i,j+1}^{k} + G_{i,j-1}^{k} - 4G_{i,j}^{k}$$

با استفاده از تخمین لاپلاس جریان ریچی G_{ij}^{k} در هر قدم t به شکل زیر تغییر می‌کند:

$$G_{ij}^{k}(t + \Delta t) = G_{ij}^{k}(t) - 2 \cdot \Delta t \cdot \Delta G_{ij}^{k}(t)$$

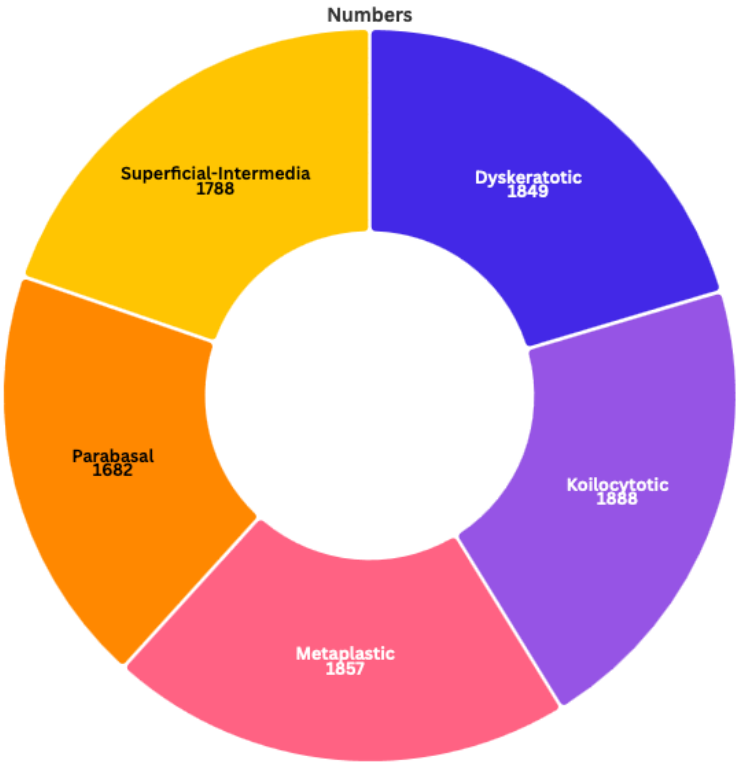
که Δt اندازه زمان هر مرحله برای تغییر ماتریس اولیه می‌باشد.

با تغییر گرادیان‌ها با استفاده از انحنای ریتچی، نواحی هندسی مهم در فضای ویژگی‌ها برجسته می‌شوند، که منجر به نقشه‌های حرارتی هموارتر و هماهنگ‌تری می‌شود. این فرآیند با شهود این که نواحی با انحنای بالا مربوط به ویژگی‌هایی هستند که در فرآیند تصمیم‌گیری مدل اهمیت بیشتری دارند، هم‌راستا است (۱۹). برای ارزیابی دقت و ارتباط بالینی نقشه‌های حرارتی تقویت شده با جریان ریتچی، ارزیابی نظارت شده‌ای با استفاده از تصاویر برش خورده سلول‌ها از مجموعه داده SIPaKMeD انجام دادیم. برای این تحلیل، ۱۰۰ تصویر از هر یک از پنج دسته (Koilocytotic، Parabasal، Dyskeratotic، Superficial-Intermediate و Metaplastic) انتخاب شد. این تصاویر تحت راهنمایی یک متخصص زنان و زایمان بازبینی شدند تا از اعتبار بالینی آن‌ها اطمینان حاصل شود. در این مطالعه، روش وارد سازی و حذف را برای اندازه‌گیری کمی تبیین پذیری نقشه‌های حرارتی تولید شده با استفاده از Grad-CAM و Grad-CAM تقویت شده با جریان ریتچی به کار گرفتیم. منحنی‌های وارد سازی و حذف برای محاسبه متریک مساحت زیر منحنی (AUC) استفاده می‌شوند تا نشان دهند که چند پیکسل از نقشه حساسیت، امتیازهای نقشه‌های تفکیک شده حاصل را اضافه یا کاهش خواهند داد (۲۱و ۲۰). با استفاده از تصاویر برش‌خورده سلول‌ها به‌عنوان حقیقت پایه، نواحی برجسته شده در نقشه‌های حرارتی Grad-CAM تولید شده از تصاویر اسلایدهای کامل را مقایسه کردیم. این مقایسه به صورت نظارت شده انجام شد و ارزیابی کردیم که آیا نواحی شناسایی شده توسط نقشه‌های حرارتی با نواحی مهم از نظر تشخیصی در تصاویر برش خورده هم‌راستا هستند یا خیر.

یافته ها

مجموعه داده Sipakmed شامل ۵ دسته است که در شکل ۲ نشان داده شده‌اند. تصاویر این مجموعه داده پیش‌پردازش و تقویت شده‌اند تا یکنواختی داده‌ها تضمین شده و عملکرد مدل بهبود یابد. مراحل اولیه پیش‌پردازش شامل تغییر اندازه، نرمال سازی رنگ و کاهش نویز بود تا کیفیت تصاویر در سراسر

مجموعه داده استاندارد شود. برای مقابله با عدم تعادل کلاس‌ها، تکنیک‌های تقویت داده مانند چرخش، معکوس کردن و مقیاس‌بندی اعمال شد تا اطمینان حاصل شود که هر یک از پنج دسته (Koilocytotic، Parabasal، Dyskeratotic، Superficial-Intermediate و Metaplastic) تعداد برابری از تصاویر داشته باشد. این مجموعه داده متعادل پایه‌ای قوی برای آموزش مدل فراهم کرد.

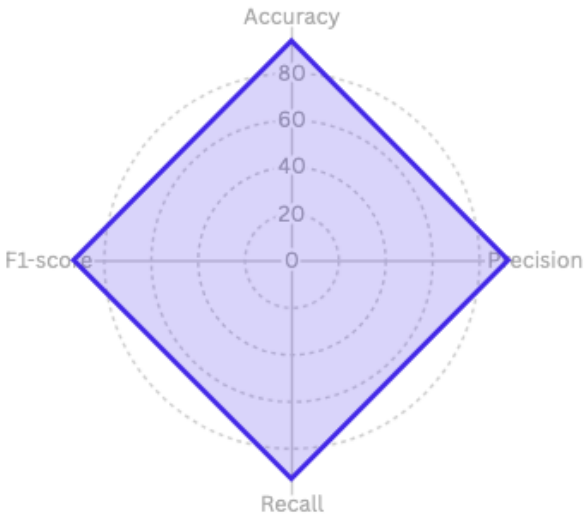


شکل ۲. نمودار دونات تعداد تصاویر در هر دسته

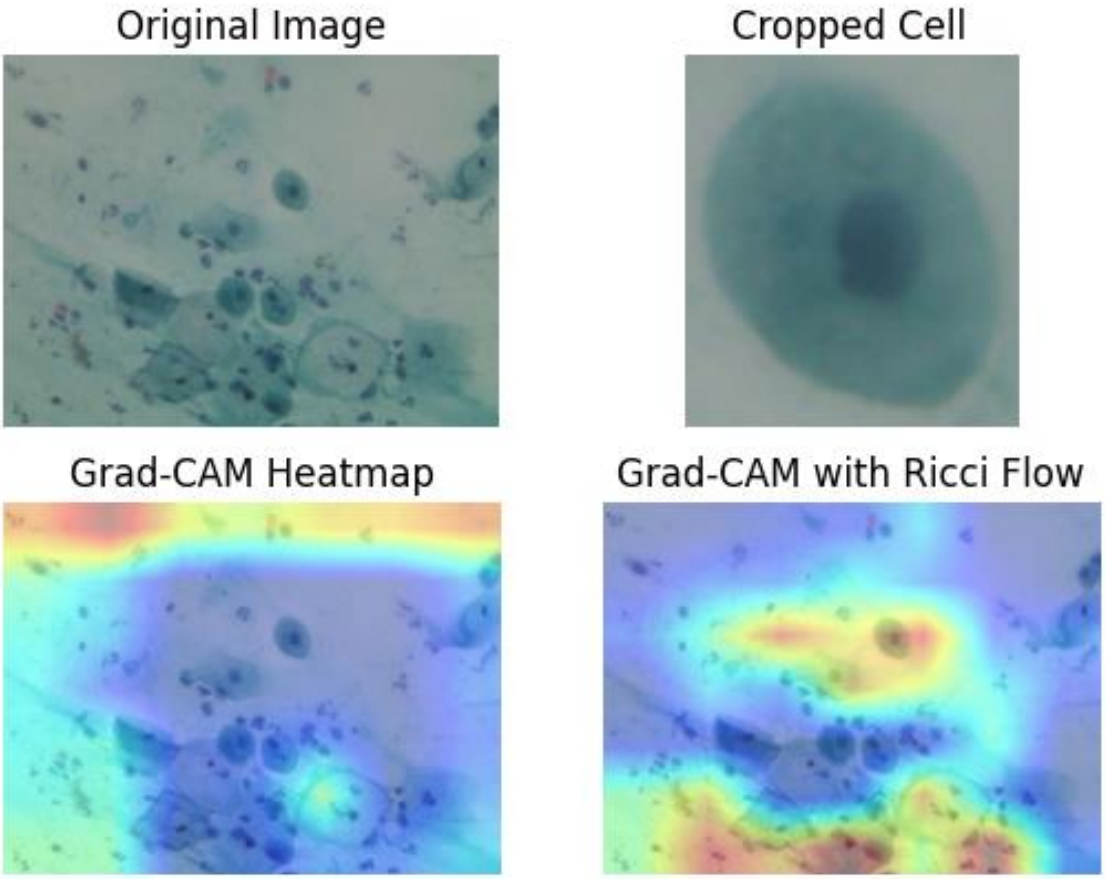
مدل ResNet50، که با لایه چگال نهایی برای طبقه‌بندی ۵ دسته تنظیم شده بود، بر روی تصاویر اسلاید کامل برای ۴۰ دوره آموزشی آموزش داده شد. در طول آموزش، مدل با نرخ یادگیری ۰/۰۰۰۱ به همگرایی رسید. با استفاده از بهینه ساز Adam و تابع ضرر انتروپی متقاطع، مدل آموزش داده شد. توقف زودهنگام برای جلوگیری از بیش‌برازش (overfitting) اعمال شد و بهترین وزن‌های مدل برای تحلیل‌های بعدی ذخیره گردید. در مجموعه تست، مدل به دقت ۹۳/۵٪، دقت میانگین ماکرو ۹۲/۷٪، بازخوانی میانگین ماکرو ۹۳/۱٪ و امتیاز F1 میانگین ماکرو ۹۲/۹٪ دست یافت. نمودار عنکبوتی در شکل ۳ نشان داده شده است.

ادغام جریان ریتچی در چارچوب Grad-CAM به طور قابل توجهی دقت و تمرکز نقشه‌های حرارتی تولید شده را بهبود بخشید. در بیش از ۹۰٪ موارد، نقشه‌های حرارتی تقویت شده با جریان ریتچی با نواحی مهم تشخیصی نسبت به Grad-CAM استاندارد هم‌راستاتر بودند. علاوه بر این، نقشه‌های حرارتی به طور مداوم دقت و تمرکز بیشتری را نشان دادند و در ۱۰۰٪ موارد، تصاویر واضح و بالینی معنی‌داری را به نمایش گذاشتند.

تصویر ارائه شده عملکرد برتر Grad-CAM تقویت شده با جریان ریتچی را نشان می‌دهد. به‌طور خاص، نقشه حرارتی تولید شده با روش هدایت شده توسط جریان ریتچی، سلول متاپلاستیک را با تمرکز و دقت بیشتری برجسته می‌کند. برخلاف نقشه حرارتی Grad-CAM استاندارد که توزیع فعال سازی وسیع‌تر و کمتر هدف‌دار نشان می‌دهد، رویکرد جریان ریچی ناحیه تشخیصی مهم را با تمرکز بر ویژگی‌های مرتبط‌تر، برجسته می‌کند (شکل ۴).

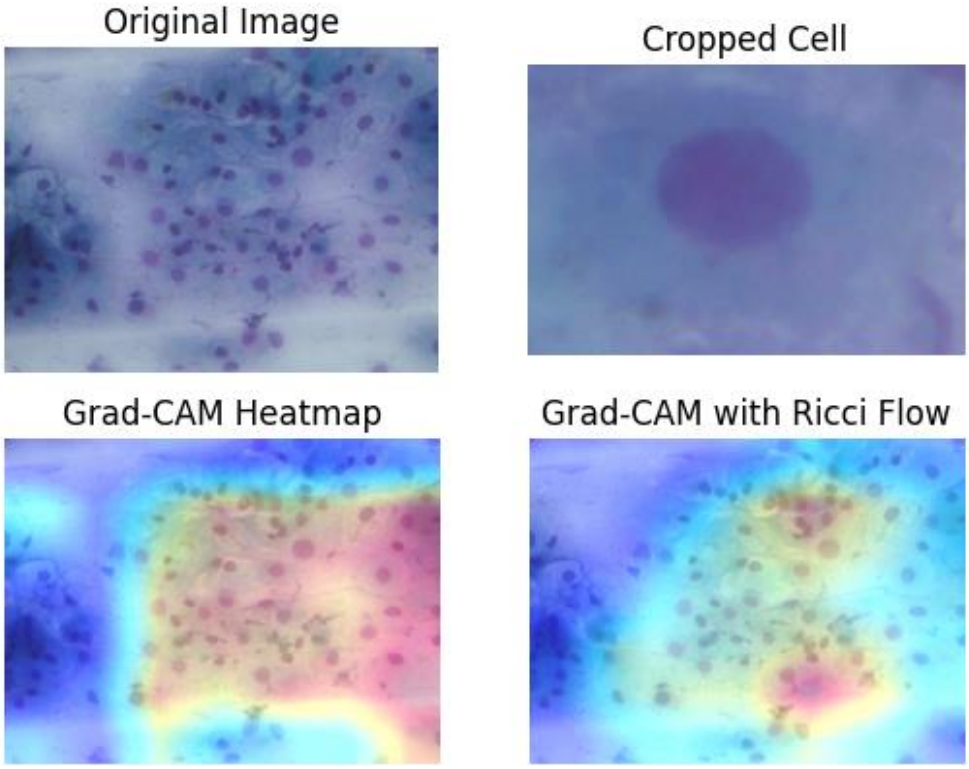


شکل ۳. نمودار عنکبوتی که ارزیابی متوازن و دقیقی از مدل را نشان می‌دهد.



شکل ٤. نقشه حرارتی تقویت شده توسط تنظیمات گرادیان هدایت شده با جریان ریتچی. این تنظیمات نقشه حرارتی را تیزتر کرده و تمرکز را به نواحی بحرانی تشخیصی هدایت می‌کند و فعال سازی‌های پس‌زمینه نامرتبط را حذف می‌کند.

همان‌طور که در شکل ۵ نشان داده شده است، انحنای ریتچی به طور ذاتی میزان انحراف یک منیفلد از مسطح بودن را اندازه‌گیری می‌کند. با تنظیم وزن‌ها و گرادیان‌ها با استفاده از جریان ریتچی، این روش نواحی را که در آن‌ها هندسه (و در نتیجه اهمیت) بیشتر نمایان است، تقویت می‌کند. این منجر به نقشه حرارتی تیزتر و دقیق‌تر می‌شود.



شکل ۵. این تصویر به طور خاص نشان می‌دهد که چگونه Grad-CAM مبتنی بر جریان ریتچی به طور بهینه تمرکز را به نواحی تشخیصی مرتبط بازتوزیع می‌کند و منجر به بهبود قابلیت توضیح و تفسیر می‌شود، به‌ویژه در وظایف تصویربرداری پزشکی که در آن‌ها محلی‌سازی دقیق بسیار حیاتی است.

نتایج ارزیابی مقایسه‌ای AUC (مساحت زیر منحنی) برای روش‌های درج و حذف بر روی ۵۰۰ تصویر انتخاب شده به صورت تصادفی از هر دسته، اثربخشی ادغام جریان ریتچی در Grad-CAM را نشان می‌دهد. نتایج ارزیابی مقایسه‌ای AUC (مساحت زیر منحنی) برای روش‌های درج و حذف بر روی ۵۰۰ تصویر که به طور تصادفی از هر دسته انتخاب شده‌اند، اثربخشی گنجاندن جریان ریتچی در Grad-CAM را نشان می‌دهد. روش درج با یک تصویر خالی آغاز می‌شود و به طور تدریجی "مناطق" ورودی را بر اساس اهمیت رتبه بندی شده توسط نقشه حرارتی وارد می‌کند. با اضافه شدن مناطق مهم‌تر، اعتماد مدل به پیش‌بینی برای کلاس هدف باید افزایش یابد. AUC درج نمایانگر امتیاز اعتماد تجمعی به عنوان تابعی از کسری از پیکسل‌های مهمی است که دوباره اضافه می‌شوند.

AUC درج برای Grad-CAM برابر با ۰/۵۹۲ بود، در حالی که Grad-CAM تقویت شده با جریان ریتچی AUC بالاتری معادل ۰/۶۷۱ به دست آورد. این نشان می‌دهد که روش تقویت شده با جریان ریتچی بهتر نواحی مهم در تصویر را برجسته می‌کند که منجر به توضیح دقیق‌تر و هدفمندتری می‌شود، زیرا نواحی به طور تدریجی افزوده می‌شوند.

از سوی دیگر، برای AUC حذف، Grad-CAM امتیاز ۰/۲۱۳ را کسب کرد، در حالی که Grad-CAM با جریان ریتچی مقداری کمی پایین‌تر معادل ۰/۱۵۳ نشان داد. AUC حذف پایین‌تر نشان دهنده این است که روش تقویت شده با جریان ریتچی به طور مؤثرتری نواحی بحرانی را شناسایی و حذف می‌کند، که منجر به کاهش تیزتر عملکرد مدل هنگام حذف این نواحی مهم می‌شود (جدول ۱).

جدول ۱. ارزیابی مقایسه‌ای AUC درج (مقادیر بالاتر بهتر هستند) و حذف (مقادیر پایین‌تر بهتر هستند) مدل‌ها

	Grad_Cam	Grad_Cam همراه جریان ریچی
درج	۰/۵۹۲	۰/۶۷۱
حذف	۰/۲۱۳	۰/۱۵۳

بحث و نتیجه گیری

نتایج مطالعه ما با شواهد فزآینده‌ای که نشان می‌دهند شبکه‌های عصبی به طور ضمنی ساختارهای هندسی را در محاسبات خود دستکاری می‌کنند، هم‌راستا است، همان طور که در یافته‌های اخیر در ادبیات علمی ذکر شده است. به طور خاص، بهبودهایی که در تبیین پذیری با استفاده از Grad-CAM تقویت شده با جریان ریتچی مشاهده شده است، زیرساخت‌های هندسی فرآیندهای یادگیری عمیق را برجسته می‌کند (۲۲و۱۲). لایه‌های چگالی که پس از محاسبات گرادیان قرار دارند، می‌توانند به عنوان انجام تبدیلات مشابه با جریان ریتچی تفسیر شوند. این با پیشنهادات موجود هم‌خوانی دارد که چنین لایه‌هایی هندسه فضای ویژگی‌ها را از طریق یک فرآیند تکراری تکامل می‌دهند، که به‌طور تقریبی به صورت $g^{l+1} - g^l \approx -\alpha \cdot Ric(g^l)$ فرموله شده است. در اینجا، انحنای ریکی به عنوان یک عملگر صاف کننده است، که از طریق توزیع مجدد و تقویت تمرکز بر روی نواحی بحرانی منیفولد ویژگی عمل می‌کند. به وسیله ادغام جریان ریکی در محاسبه گرادیان، از این خاصیت استفاده می‌شود تا نقشه‌های حرارتی تیزتر و از نظر تشخیصی مرتبط‌تر تولید شود.

موفقیت روش ما نشان می‌دهد که چگونه استفاده از این بینش‌های هندسی می‌تواند تفسیرپذیری و عملکرد مدل‌های یادگیری عمیق را به ویژه در زمینه‌هایی مانند تصویربرداری پزشکی که دقت در آن‌ها بسیار مهم است، بهبود بخشد (۱۲). این با ایده‌ای که مکانیزم‌هایی مانند فعال سازی‌های اصلاح شده (ReLU) و معماری‌های لایه‌ای عمیق در enabling این تغییرات توپولوژیکی نقش اساسی دارند، هم‌راستا است. به طور خاص، فعال سازی‌های ReLU به عنوان نقشه‌های چند به یک عمل می‌کنند که داده‌ها را "تا می‌زنند"، در حالی که عمق و عرض شبکه ابعاد و توالی تغییرات لازم برای گسستن و بازسازی فضای ویژگی‌ها را فراهم می‌کنند (۲۳).

در این راستا، Grad-CAM تقویت شده با جریان ریکی از این ایده‌ها پشتیبانی تجربی ارائه می‌دهد و نشان می‌دهد که چگونه انحنا، که به طور ذاتی با عملیات هندسی شبکه در ارتباط است، در تصفیه و تمرکز بر نواحی بحرانی کمک می‌کند. جریان ریکی هندسه زیرین را به طور تدریجی تنظیم می‌کند، به عنوان یک عملیات صاف کننده عمل می‌کند که ویژگی‌های تشخیصی مرتبط را تقویت کرده و ویژگی‌های بی‌ربط را سرکوب می‌کند. این فرآیند بازتابی از ظرفیت شبکه‌های عصبی برای "تا زدن" و بازسازی فضای ویژگی‌ها است، همان طور که در چارچوب‌های نظری توصیف شده است (۲۳).

تمثیل جدا کردن دایره‌های هم‌مرکز به فضایی قابل تفکیک توسط ابرصفحه به ویژه در این زمینه قابل توجه است. همانطور که شبکه‌های عصبی از لایه‌های متوالی و فعال‌سازی‌ها برای دستیابی به تغییرات توپولوژیکی استفاده می‌کنند، ادغام جریان ریکی در Grad-CAM نشان می‌دهد که چگونه وارد کردن اصول هندسی تفسیرپذیری این تغییرات را بهبود می‌بخشد و نواحی مرتبط‌تر برای طبقه بندی را برجسته می‌کند. این بینش‌ها تعامل پیچیده هندسه در یادگیری عمیق را برجسته می‌کند و راه را برای تحقیقات آینده در مورد مکانیزم‌هایی که شبکه‌های عصبی فضای ورودی خود را تغییر شکل می‌دهند، هموار می‌کند. این رویکرد تأثیر قابل توجهی بر تصویربرداری پزشکی و سایر زمینه‌ها خواهد داشت و به مدل‌ها این امکان را می‌دهد که به طور ذاتی بر روی نواحی بحرانی تشخیصی تمرکز کنند بدون اینکه نیاز به نظارت دستی گسترده‌ای داشته باشند. برای مثال، در زمینه تشخیص سرطان یا تحلیل سلولی، چنین تابع‌های خسارت می‌توانند مدل‌ها را بهینه سازی کنند تا نشانگرهای مورفولوژیک بیماری را اولویت بندی کنند و منجر به خروجی‌های دقیق‌تر و قابل تفسیرتر شوند (۲۵و۲۴).

روش ما بر این که چگونه شبکه عصبی فکر می‌کند، با تجسم فرآیندهای داخلی و تصمیمات در فضای ویژگی تمرکز دارد. در مقابل، بیشتر مکانیزم‌های توضیح‌دهی دیگر مانند مدل‌های زبان بزرگ (LLM) و مکانیزم‌های توجه مبتنی بر توکن آن‌ها می‌توانند به عنوان نشان دادن چگونگی فکر کردن شبکه‌های عصبی دیده شوند، زیرا آن‌ها در ایجاد توضیحات ساختارمند، منسجم و شبیه انسان برتری دارند. مدل‌های زبان بزرگ داستان‌ها و توجیه‌هایی تولید می‌کنند که با انتظارات انسانی هم‌راستا است، حتی زمانی که این توجیه‌ها ممکن است مستقیماً با فرآیندهای واقعی شبکه عصبی مطابقت نداشته باشند. ترکیب این چارچوب با مدل‌های زبان بزرگ (LLMs) این پتانسیل را دارد که به یک سیستم جامع برای تشخیص و تحقیق تبدیل شود. مدل‌های زبان بزرگ می‌توانند به عنوان مفسر عمل کنند و بینش‌های مدل‌های تقویت شده با جریان ریکی را به داستان‌های انسانی قابل درک و دقیق ترجمه کنند. این ترکیب می‌تواند به پزشکان و

محققان توضیحات و پیش‌بینی‌های دقیق و علمی تایید شده ارائه دهد، جریان‌های کاری را ساده‌سازی کند و بار شناختی تفسیر داده‌های پیچیده را کاهش دهد (۲۶). علاوه بر این، چنین مدل‌هایی می‌توانند نقش مهمی در پیشبرد تحقیقات پزشکی ایفا کنند و الگوها و روابط درون داده‌ها را که قبلاً نادیده گرفته شده بودند، کشف کنند. تمرکز و تفسیرپذیری بهبود یافته توسط جریان ریکی می‌تواند به محققان کمک کند تا نشانگرهای جدید زیستی را شناسایی کنند، پیشرفت بیماری را مطالعه کنند و فرضیات را با اطمینان بیشتری آزمایش کنند.

محدودیت‌ها و پیشنهادات: در حالی که مدل ما یک ساختار مبتنی بر ResNet بود و لایه پنهان یک لایه پس از استخراج ویژگی‌ها بود، تقریب ما از ضریب −2 برای انحنای ریکی یک تقریب بود و تعداد تکرارها به یک محدود شده بود. با این حال، زمانی که لایه‌های بیشتری در وضعیت پنهان استفاده می‌شوند، باید تعداد تکرارها افزایش یابد و پیدا کردن یک تقریب دقیق‌تر به ما کمک خواهد کرد تا نقشه حرارتی دقیق‌تر و صحیح‌تری تعریف کنیم. یکی از چالش‌های اصلی در تقریب یک ضریب، طبیعت غیرخطی مدل‌های یادگیری عمیق است. بنابراین، تحقیقات در یافتن ضرایب و تکرارها بر اساس تعداد لایه‌های پنهان و وزن‌های استفاده شده در هر لایه ضروری است تا کارآیی و دقت نقشه‌های حرارتی افزایش یابد.

ادغام جریان ریکی در Grad-CAM تفسیرپذیری نقشه‌های حرارتی را با استفاده از صاف سازی هندسی بهبود می‌بخشد و رویکرد جدیدی برای هوش مصنوعی قابل توضیح در تشخیص سرطان دهانه رحم ارائه می‌دهد. این روش نه تنها دقت مدل را بهبود می‌بخشد بلکه با فرضیه‌ای که بیان می‌کند شبکه‌های عصبی در طول آموزش توپولوژی داده‌ها را تغییر می‌دهند نیز هم‌راستا است.

تقدیر و تشکر

بدینوسیله از تمام کسانی که در پیشبرد این مطالعه سهم داشته‌اند، از کارشناسان و پزشکان برای بینش‌ها و راهنمایی‌های بی‌قیمت‌شان، به ویژه در زمینه تشخیص سرطان گردن رحم و همچنین از تیم دیتاست Sipakmed هستند، که برای آموزش و ارزیابی مدل ما ضروری بود، قدردانی می‌گردد.

References

1.Dahiya N, Aggarwal K, Singh MC, Garg S, Kumar R. Knowledge, attitude, and practice regarding the screening of cervical cancer among women in New Delhi, India. Tzu Chi Med J. 2019;31(4):240-3.

2.Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. Lancet. 2019;393(10167):169-82.

3.Shanthi PB, Hareesha KS. Automated detection and classification of cervical cancer using pap smear microscopic images: a comprehensive review and future perspectives. Eng Sci. 2022;19(4):20-41.

4.William W, Ware A, Basaza-Ejiri AH, Obungoloch J. A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images. Comput Methods Programs Biomed. 2018;164:15-22.

5.Plissiti ME, Dimitrakopoulos P, Sfikas G, Nikou C, Krikoni O, Charchanti A. Sipakmed :A New Dataset for Feature and Image Based Classification of Normal and Pathological Cervical Cells in Pap Smear Images. 2018 25th IEEE International Conference on Image Processing (ICIP); 2018. p. 3144-8.

6.He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. p. 770-8.

7.Zhang H, Chen Y, Zhu H, Li W. Modulation recognition technology based on ResNet50. Proceedings of the 2023 6th International Conference on Signal Processing and Machine Learning; 2023. p. 14-20.

8.Saporta A, Gui X, Agrawal A, Pareek A, Truong SQ, Nguyen CD, et al. Benchmarking saliency methods for chest X-ray interpretation. Nature Machine Intelligence. 2022;4(10):867-78.

9.Chattopadhay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. 2018 IEEE winter conference on applications of computer vision (WACV). 2018. p. 839-47.

10.Naitzat G, Zhitnikov A, Lim LH. Topology of deep neural networks. J Machine Learn Res. 2020;21(184):1-40.

11.Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Net. 1989;2(5):359-66.

12.Baptista A, Barp A, Chakraborti T, Harbron C, MacArthur BD, Banerji CRS. Deep learning as Ricci flow. Sci Rep. 2024;14(1):23383.

13.Sheridan N, Hodgson C, Rubinstein H. Hamilton's Ricci Flow [Honours Thesis]. The University of Melbourne, Department of Mathematics and Statistics; 2006. Available from: https://web.math.princeton.edu/~nsher/ricciflow.pdf

14.Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM :Visual Explanations from Deep Networks via Gradient-Based Localization. Int J Comput Vis. 2020;128:336-59.

15.Selvaraju RR, Das A, Vedantam R, Cogswell M, Parikh D, Batra D. Grad-CAM :Why did you say that?. arXiv preprint arXiv:161107450. 2016.

16.Zeng W, Samaras D, Gu D. Ricci Flow for 3D Shape Analysis. IEEE Trans Pattern Anal Mach Intell. 2010;32(4):662-77.

17.Jin M, Kim J, Luo F, Gu X .Discrete Surface Ricci Flow. IEEE Trans Vis Comput Graph. 2008;14(5):1030-43.

18.Yang YL, Guo R, Luo F, Hu SM, Gu X. Generalized Discrete Ricci Flow. Computer Graphics Forum. 2009;28(7):2005-14.

19.Carson T, Isenberg J, Knopf D, Šešum N. Singularity formation of complete Ricci flow solutions. Adv Math. 2022;403:108326.

20.Hama N, Mase M, Owen AB. Deletion and insertion tests in regression models. J Mach Learn Res. 2023;24(290):1-38.
21.Naidu R, Ghosh A, Maurya Y, Kundu SS. Is-cam: Integrated score-cam for axiomatic-based explanations. arXiv preprint arXiv:2010.03023. 2020:1-8.
22.Chen J, Chen H, Wang M, Dai G, Tsang IW, Liu Y. Learning discretized neural networks under ricci flow. J Mach Learn Res. 2024;25(386):1-44.
23.Miikkulainen R. Topology of a Neural Network. In: Sammut C, Webb GI, editors. Encyclopedia of Machine Learning, 1st ed. Boston, MA: Springer US; 2010. p .988-9.
24.Nasarian E, Alizadehsani R, Acharya UR, Tsui KL. Designing interpretable ML system to enhance trust in healthcare: A systematic review to proposed responsible clinician-AI-collaboration framework. Inf Fusion. 2024;108:102412.
25.Xu Q, Xie W, Liao B, Hu C, Qin L, Yang Z, et al. Interpretability of Clinical Decision Support Systems Based on Artificial Intelligence from Technological and Medical Perspective: A Systematic Review. J Healthc Eng. 2023;2023:9919269.
26.Cambria E, Malandri L, Mercorio F, Nobani N, Seveso A. Xai meets llms: A survey of the relation between explainable ai and large language models. arXiv preprint arXiv:2407.15248. 2024.