

تشخیص و درجه‌بندی خودکار رتینوپاتی دیابتی با استفاده از شبکه کانولوشنی سلسله‌مراتبی و مکانیزم توجه دوگانه

مصطفی مجمان سالم الغنامی^۱، منصور فاتح^{۲*}

^۱ دانشجوی دکتری دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شاهرود، شاهرود، ایران

^۲ دانشیار دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شاهرود، شاهرود، ایران

مقاله پژوهشی

چکیده

تاریخ دریافت:

۱۴۰۴/۱۱/۲۱

تاریخ پذیرش:

۱۴۰۵/۰۲/۰۷

کلیدواژه‌ها:

رتینوپاتی دیابتی، یادگیری عمیق، شبکه ResNet، مکانیزم توجه، تحلیل تصاویر پزشکی

نویسنده مسئول:

mansoor_fateh@shahroodut.ac.ir

رتینوپاتی دیابتی یکی از شایع‌ترین عوارض چشمی دیابت و عامل اصلی نابینایی بزرگسالان است. تشخیص زودهنگام آن به دلیل ابعاد کوچک ضایعات اولیه (مانند میکروآنوریزم‌ها) و شباهت با بافت سالم شبکیه، چالشی اساسی است. روش‌های متداول یادگیری عمیق در ایجاد تعادل میان استخراج جزئیات بافتی و درک معنایی کلی محدودیت دارند. در این مقاله برای غلبه بر این محدودیت، شبکه‌ای کانولوشنی با ساختار دوشاخه‌ای پیشنهاد شده که از محو شدن جزئیات ریز ضایعات در لایه‌های عمیق جلوگیری می‌کند. این معماری از شبکه ResNet-50 پیش‌آمोخته به‌دلیل ساختار باقیمانده و توانایی استخراج ویژگی‌های چندسطحی بهره‌برده و اطلاعات را در دو مسیر مدیریت می‌کند. مسیر اول ویژگی‌های سطح میانی (جزئیات محلی) و مسیر دوم ویژگی‌های معنایی عمیق را پردازش می‌کند. ماژول توجه فضایی و کانالی تعبیه‌شده در شبکه، با مدل‌سازی وابستگی‌ها، به‌صورت انطباقی نواحی ضایعه را تقویت و نویز پس‌زمینه را سرکوب می‌کند. همچنین از تابع ضرر ترکیبی (آنتروپی متقاطع و فوکال) برای مدیریت عدم تعادل داده‌ها و تمرکز بر نمونه‌های دشوار استفاده شده است. ارزیابی روی دادگان APTOS 2019 صحت ۹۷/۲۳٪ در تشخیص باینری و ۸۳/۰۶٪ در درجه‌بندی ۵ کلاس را نشان داد. این نتایج در حالت چندکلاس، بیانگر بهبود ۲ درصدی صحت و ۴ درصدی امتیاز F1 نسبت به بهترین روش‌های پیشرفته اخیر است. نتایج تجربی کارایی بالای مدل پیشنهادی را اثبات می‌کند.

doi : 10.22034/ABMIR.2026.24319.1221

E-ISSN: 2821-2037

/The Author 2026. Published by Yazd University This is an open

access article under the CC BY 4.0 License (<https://creativecommons.org/licenses/by/4.0/>).



۱- مقدمه

رتینوپاتی دیابتی^۱ یکی از شایع‌ترین عوارض میکروواسکولار بیماری دیابت است که در صورت عدم تشخیص و درمان به‌موقع، می‌تواند منجر به نابینایی غیرقابل برگشت شود. بر اساس آمارهای جهانی، با توجه به ابتلای صدها میلیون نفر به دیابت، انتظار می‌رود شیوع این عارضه چشمی نیز در دهه‌های آینده به‌شدت افزایش یابد. با توجه به این حجم فزاینده، روش‌های غربالگری سنتی که متکی بر بررسی دستی تصاویر فوندوس توسط چشم‌پزشکان است، با چالش‌هایی نظیر زمان‌بر بودن، هزینه بالا و احتمال خطای انسانی مواجه هستند [۱]. تشخیص زودهنگام ضایعاتی همچون میکروآنوریسم‌ها و خونریزی‌ها، نقش حیاتی در جلوگیری از پیشرفت بیماری ایفا می‌کند [۲]. از این رو، توسعه سیستم‌های تشخیص به کمک کامپیوتر^۲ که قادر به شناسایی دقیق و خودکار علائم این بیماری باشند، به یک ضرورت در حوزه مهندسی پزشکی و هوش مصنوعی تبدیل شده است [۳].

در سال‌های اخیر، با گذار از پردازش تصویر سنتی به سمت الگوریتم‌های یادگیری عمیق، دقت سیستم‌های تشخیص خودکار به‌طور چشمگیری افزایش یافته است. شبکه‌های عصبی کانولوشنی (CNN) به‌عنوان استاندارد طلایی در تحلیل تصاویر شبکه شناخته می‌شوند [۴]. پژوهش‌های متعددی کارایی معماری‌های شناخته‌شده‌ای مانند ResNet و VGG را در طبقه‌بندی تصاویر فوندوس نشان داده‌اند [۵]. به‌عنوان مثال، استفاده از تکنیک یادگیری انتقالی^۳ بر روی مدل‌های پیش‌آمورخته، راهکاری مؤثر برای غلبه بر کمبود داده‌های برجسب‌دار پزشکی بوده و نتایج قابل‌قبولی را در تشخیص درجات مختلف رتینوپاتی ارائه کرده است [۶]. همچنین تلاش‌های اخیر بر ترکیب ویژگی‌های استخراج‌شده از لایه‌های مختلف برای بهبود بازنمایی تصویر متمرکز شده است [۷].

علیرغم پیشرفت‌های حاصل‌شده، مدل‌های فعلی مبتنی بر CNN همچنان با محدودیت‌هایی کمی و کیفی روبرو هستند. یکی از چالش‌های اصلی، افت شدید عملکرد مدل‌ها در تشخیص

کلاس‌های اقلیت و مراحل شدید بیماری است که ناشی از دشواری در ایجاد تعادل میان اطلاعات مکانی جزئی و اطلاعات معنایی کلی است [۸]. به‌طوری‌که در این کلاس‌ها اکثر آمار معیار فراخوانی کمتر از ۵۰ درصد است. در معماری‌های متداول، با عمیق‌تر شدن شبکه، جزئیات دقیق ضایعات کوچک (که در مراحل اولیه بیماری ظاهر می‌شوند) در طی عملیات ادغام^۴ از بین می‌روند. از سوی دیگر، تمرکز صرف بر ویژگی‌های سطح بالا ممکن است منجر به نادیده گرفتن الگوهای ظریف بافتی شود [۹]. درواقع شکاف تحقیقاتی موجود این است که بسیاری از مدل‌ها تمامی سطوح ویژگی را به یک شیوه پردازش می‌کنند، درحالی‌که ماهیت اطلاعات موجود در لایه‌های میانی و انتهایی شبکه متفاوت است و نیازمند مکانیزم‌های توجه^۵ اختصاصی برای برجسته‌سازی نواحی مهم‌تر است.

برای غلبه بر این چالش‌ها، در این پژوهش یک معماری دوشاخه‌ای نوین مبتنی بر شبکه کانولوشنی ResNet-50 پیشنهاد شده است. دلیل انتخاب این شبکه پایه، ساختار اتصالات باقیمانده قوی آن است که استخراج پایدار ویژگی‌های عمیق را بدون خطر محوشدگی گرادیان تضمین می‌کند. نوآوری اصلی این پژوهش در مقایسه با روش‌های رایج یادگیری انتقالی که غالباً به‌صورت تک‌مسیره عمل کرده و تنها از خروجی لایه آخر استفاده می‌کنند، پردازش هم‌زمان و مجزای ویژگی‌های مکانی و سراسری است. برخلاف روش‌های معمول، در مدل پیشنهادی ما، منحصراً از خروجی بلوک‌های میانی (بلوک‌های ۲ و ۳) به‌عنوان منبع ویژگی‌های محلی استفاده شده است؛ زیرا برخلاف بلوک ۱ که حاوی نویزهای روشنایی است و برخلاف بلوک ۴ که رزولوشن مکانی خود را تا حد زیادی ازدست‌داده، این دو بلوک بهترین تعادل را برای حفظ اطلاعات دقیق بافت و لایه‌های ضایعات کوچک دارند. خروجی این دو بلوک با یکدیگر ترکیب شده و در یک مسیر جداگانه پردازش می‌شوند. هم‌زمان، خروجی بلوک نهایی (بلوک ۴) که حاوی ویژگی‌های معنایی انتزاعی است، در مسیری موازی قرار می‌گیرد. هر دو مسیر مجهز به یک ماژول توجه فضایی و

⁴ Pooling

⁵ Attention Mechanism

¹ Diabetic Retinopathy

² Computer-Aided Diagnosis

³ Transfer Learning

۲-۱ روش‌های سنتی و یادگیری ماشین

پیش از ظهور یادگیری عمیق، رویکردهای غالب بر استخراج ویژگی‌های دستی و استفاده از طبقه‌بندهای کلاسیک استوار بودند. اگرچه این روش‌ها امروزه کمتر مورد استفاده قرار می‌گیرند، اما به دلیل تفسیرپذیری بالا و کارایی در مجموعه‌داده‌های کوچک، همچنان در برخی مطالعات اخیر به عنوان روش‌های پایه یا ترکیبی مورد توجه هستند [۱۰]. ژانگ و همکاران [۱۱] یک چارچوب جامع مبتنی بر یادگیری ماشین را توسعه دادند که متغیرهای بالینی را با ویژگی‌های تصویری ترکیب می‌کرد. نتایج آن‌ها نشان داد که اگرچه الگوریتم‌های سنتی به تنهایی دقت محدودی دارند، اما ادغام داده‌های پرونده الکترونیک سلامت با این مدل‌ها می‌تواند پیش‌بینی بیماری را بهبود بخشد.

در راستای استخراج ویژگی، تاباکارو و همکاران [۱۲] رویکردی مبتنی بر بافت^۱ را پیشنهاد کردند که از آن‌تروپی شانون و آن‌تروپی فازی برای توصیف ناهم‌واری‌های سطح شبکه استفاده می‌کرد. آزمایش‌های آن‌ها نشان داد که طبقه‌بند Gradient Boosting در تفکیک مراحل اولیه بیماری عملکرد قابل قبولی دارد. همچنین، بهینه‌سازی انتخاب ویژگی یکی دیگر از چالش‌های این حوزه بوده است. لاکرا و همکاران [۱۳] با استفاده از الگوریتم ازدحام ذرات^۲ برای انتخاب ویژگی‌های استخراج شده از تصاویر و اعمال طبقه‌بند ماشین بردار پشتیبان (SVM)، نشان دادند که کاهش ابعاد ویژگی‌ها می‌تواند سربار محاسباتی را بدون افت دقت کاهش دهد. باین‌حال، همان‌طور که رامانایا و همکاران [۱۴] در بررسی جامع خود اشاره کرده‌اند، روش‌های گروهی^۳ مانند جنگل تصادفی^۴ با وجود تفسیرپذیری مناسب، نیازمند مراحل پیش‌پردازش سنگین و مهندسی ویژگی‌های پیچیده هستند که در مقایسه با روش‌های یادگیری عمیق یک ضعف محسوب می‌شود. به بیان دقیق‌تر، محدودیت اصلی روش‌های سنتی، استخراج ویژگی‌ها به صورت ایستا است و مدل توانایی یادگیری سلسله‌مراتبی الگوهای پنهان و پیچیده درجات مختلف بیماری را ندارد.

کانالی سلسله‌مراتبی هستند تا به صورت هوشمند و انطباقی، نواحی مرتبط با ضایعات را تقویت و پس‌زمینه نامرتبط را تضعیف نمایند. مشارکت‌های اصلی این مقاله به شرح زیر است:

- ۱) ارائه یک ساختار انشعابی کارآمد با استفاده از شبکه ResNet-50 پیش‌آمورخته که قادر است ویژگی‌های سطح میانی (حاوی جزئیات مکانی) و سطح بالا (حاوی مفاهیم معنایی) را برخلاف رویکردهای استاندارد تک‌مسیره، به صورت مجزا اما هماهنگ مدیریت کند.
 - ۲) طراحی و پیاده‌سازی ماژول توجه ترکیبی (فضایی و کانالی) اختصاصی در هر دو مسیر که با پالایش سلسله‌مراتبی ویژگی‌ها، نویز پس‌زمینه را سرکوب کرده و دقت مدل را در هر دو سناریوی تشخیص باینری (سالم و بیمار) و درجه‌بندی شدت بیماری بهبود می‌بخشد.
 - ۳) ادغام ویژگی‌های پالایش شده با استفاده از استراتژی ادغام میانگین سراسری و لایه‌های تمام‌متصل که منجر به دستیابی به یک بازنمایی جامع و غنی‌تر از تصویر ورودی می‌شود.
- ادامه این مقاله به این شرح سازمان‌دهی شده است: در بخش دوم، مروری بر روش‌های پیشین در تشخیص رتینوپاتی دیابتی ارائه می‌گردد. در بخش سوم، روش پیشنهادی به تفصیل تشریح می‌شوند. ارزیابی‌های تجربی، معرفی مجموعه داده و مقایسه نتایج در بخش چهارم آورده شده است. در نهایت و در بخش انتهایی به نتیجه‌گیری و بیان جهت‌گیری‌های پژوهشی آینده پرداخته شده است.

۲-۲ روش‌های پیشین

تشخیص خودکار رتینوپاتی دیابتی در سال‌های اخیر سیر تکاملی قابل‌توجهی را طی کرده است. این پیشرفت‌ها از تکنیک‌های پردازش تصویر سنتی آغاز شده و به سمت معماری‌های پیچیده یادگیری عمیق و مدل‌های ترکیبی سوق پیدا کرده است. در این فصل، روش‌های ارائه شده در ادبیات موضوع را در سه دسته کلی روش‌های سنتی مبتنی بر یادگیری ماشین، شبکه‌های عصبی کانولوشنی و یادگیری انتقالی و رویکردهای نوین مبتنی بر مکانیزم‌های توجه بررسی می‌کنیم.

³ Ensemble

⁴ Random Forest

¹ Texture-based

² Particle Swarm Optimization

۲-۲ شبکه‌های عصبی کانولوشنی و یادگیری انتقالی

با ورود شبکه‌های عصبی کانولوشنی، پارادایم تشخیص رتینوپاتی دیابتی تغییر کرد و مدل‌هایی مانند ResNet و EfficientNet به استانداردهای این حوزه تبدیل شدند. یولداس و همکاران [۱۵] در مطالعاتی مقایسه‌ای بین معماری‌های RegNet و EfficientNet نشان دادند که مدل‌های سبک‌تر مانند RegNetX080 در طبقه‌بندی باینری (سالم و بیمار) عملکردی معادل ۹۸/۶٪ دارند، درحالی‌که مدل‌های سنگین‌تر در درجه‌بندی چندکلاسه^۱ موفق‌تر عمل می‌کنند. برای مقابله با چالش عدم تعادل کلاس‌ها که در داده‌های پزشکی رایج است، سانتوس و همکاران [۱۸] یک شبکه عصبی کانولوشنی سیامی^۲ را معرفی کردند. این شبکه با تحلیل جفت تصاویر، قادر است در شرایط یادگیری با نمونه‌های محدود^۳، ویژگی‌های متمایزکننده ضایعات را با دقت بالاتری نسبت به CNN های معمولی استخراج کند.

تلاش‌های دیگری نیز برای بهبود معماری‌های استاندارد صورت گرفته است. جبار و همکاران [۱۷] با ادغام الگوریتم‌های فراابتکاری مانند PSO تطبیقی با معماری‌های GoogleNet و ResNet، موفق شدند دقت تشخیص ضایعات کوچک را در مراحل پیشرفته بیماری افزایش دهند. بااین‌وجود، پژوهشگران همچنان با چالش‌های محاسباتی روبرو هستند. ژو و همکاران [۴] و بولاکشمی و همکاران [۱۰] در مرورهای اخیر خود تاکید کرده‌اند که اگرچه CNN ها دقت بالایی دارند، اما هزینه محاسباتی بالا و وابستگی شدید به کیفیت پیش‌پردازش تصاویر، استقرار آن‌ها را در سیستم‌های غربالگری با منابع محدود دشوار می‌سازد.

یادگیری انتقالی به‌عنوان راهکاری کلیدی برای غلبه بر کمبود داده‌های برچسب‌دار پزشکی، در اکثر پژوهش‌های مدرن مورد استفاده قرار می‌گیرد. موتاوا و همکاران [۱۶] در سال ۲۰۲۳ نشان دادند که استفاده از مدل‌های پیش‌آمورخته مانند MobileNetV2 و DenseNet121، زمانی که با استراتژی‌های ترکیب مجموعه داده‌ها همراه شود، قدرت تعمیم‌پذیری مدل را به‌طور چشمگیری افزایش می‌دهد. این رویکرد به‌ویژه زمانی مؤثر

است که مدل روی تصاویر متنوع از چندین منبع مختلف آموزش دیده باشد. روی و همکاران [۲۰] نیز با معرفی سیستم DR-PanopticNet نشان دادند که هم‌افزایی مدل‌های عمیق پیش‌آمورخته می‌تواند منجر به تشخیص دقیق‌تر مرزهای ضایعات شود. علاوه بر این، استفاده از تکنیک‌های تنظیم دقیق^۴ بر روی لایه‌های آخر شبکه‌هایی مانند ResNet، به مدل اجازه می‌دهد تا ویژگی‌های اختصاصی شبکه را بدون نیاز به آموزش از صفر یاد بگیرد [۱۵].

با این‌وجود، تحلیل مدل‌های مبتنی بر یادگیری انتقالی استاندارد نشان می‌دهد که اکثر این شبکه‌ها منحصراً از خروجی لایه آخر (ویژگی‌های سطح بالا) برای طبقه‌بندی استفاده می‌کنند. این رویکرد تک‌مسیره باعث می‌شود تا جزئیات حیاتی و بافت‌های ریز ضایعات (مانند میکروآنورسم‌ها) که در لایه‌های میانی شبکه وجود دارند، در طی فرآیندهای ادغام از بین بروند و در نتیجه عملکرد مدل در تشخیص کلاس‌های اقلیت با افت شدیدی مواجه شود.

۲-۳ مکانیزم‌های توجه و مدل‌های نوین

یکی از محدودیت‌های اصلی شبکه‌های کانولوشنی، تمرکز ذاتی آن‌ها بر ویژگی‌های محلی و ناتوانی در مدل‌سازی وابستگی‌های دوربرد^۵ است. برای رفع این مشکل، مکانیزم‌های توجه و ترنسفورمرهای بینایی (ViTs) توسعه یافته‌اند [۱۹]. توانی و همکاران [۱۹] مدلی تحت عنوان DRCT (ترنسفورمر کانولوشنی فشرده) ارائه کردند که لایه‌های کانولوشن را برای استخراج ویژگی‌های محلی با بلوک‌های ترنسفورمر برای درک سراسری ترکیب می‌کند. نتایج آن‌ها برتری این رویکرد ترکیبی را بر مدل‌های خالص CNN اثبات کرد.

همچنین، ژو و همکاران [۲۰] یک معماری ترکیبی شامل EfficientNet و Swin Transformer را پیشنهاد دادند که در آن تلفیق ویژگی‌های محلی و سراسری منجر به شناسایی بهتر ضایعات ظریف شد. کاکاده و همکاران [۲۱] نیز با توسعه یک مدل هیبریدی CNN-ViT مجهز به توجه فضایی^۶، نشان دادند که تمرکز دینامیک شبکه بر روی نواحی حیاتی مانند میکروآکسوداها می‌تواند خطای

^۴ Fine-tuning

^۵ Long-range dependencies

^۶ Spatial Attention

^۱ Multi-class grading

^۲ Siamese CNN

^۳ Few-shot learning

تغییرات مقیاس ضایعات (از میکروآنورسم‌های کوچک تا خونریزی‌های گسترده) و نیاز به حفظ اطلاعات مکانی هم‌زمان با استخراج مفاهیم معنایی، ما یک چارچوب یادگیری عمیق دو شاخه و سلسله‌مراتبی را طراحی کرده‌ایم.

همان‌طور که در شکل (۱) نشان داده شده است، این معماری از یک شبکه ResNet-50^۱ پیش‌آمورخته به‌عنوان استخراج‌کننده ویژگی بهره می‌برد. نوآوری اصلی این ساختار در نحوه مدیریت جریان اطلاعات است؛ جایی که ویژگی‌های سطح میانی (حاوی جزئیات بافت و لبه‌ها) و ویژگی‌های سطح بالا (حاوی اطلاعات معنایی و انتزاعی) در دو مسیر مجزا پردازش شده و توسط یک «ماژول توجه فضایی و کانالی سلسله‌مراتبی» پالایش می‌شوند. درنهایت، این ویژگی‌ها برای تصمیم‌گیری نهایی ادغام می‌گردند. در ادامه، جزئیات هر یک از اجزا تشریح می‌شود.

۲-۳ تعریف مسئله

فرض کنید مجموعه داده ورودی شامل N تصویر فوندوس باشد که به‌صورت $D = \{(X_i, y_i)\}_{i=1}^N$ نمایش داده می‌شود. در اینجا، $X_i \in R^{H \times W \times 3}$ تصویر ورودی با ابعاد 512×512 پیکسل و سه کانال رنگی (RGB) است و y_i برچسب کلاس متناظر با آن است (در حالت باینری $\{0, 1\}$). هدف نهایی، یادگیری تابع نگاشت غیرخطی $F(X; \theta)$ است که با بهینه‌سازی پارامترهای θ ، احتمال تعلق تصویر به کلاس بیماری را تخمین بزند.

۳-۳ شبکه استخراج ویژگی

به‌عنوان شبکه پایه در این معماری، از شبکه عصبی کانولوشنی عمیق ResNet-50 بهره گرفته شده است. دلیل انتخاب این معماری، ساختار باقیمانده^۲ آن است که با استفاده از اتصالات پرشی^۳، مشکل محوشدگی گرادیان در شبکه‌های عمیق را مرتفع ساخته و امکان آموزش مؤثر لایه‌های متعدد را فراهم می‌آورد. ما از نسخه پیش‌آمورخته این شبکه بر روی مجموعه داده عظیم ImageNet استفاده می‌کنیم. این استراتژی که تحت عنوان «یادگیری انتقالی» شناخته می‌شود، به مدل اجازه می‌دهد تا به‌جای شروع از وزن‌های تصادفی، با دانش قبلی در مورد ویژگی‌های

تشخیص را کاهش دهد. فراتر از ترنسفورمرها، معماری‌های نوپهوری مانند مدل‌های فضای حالت^۱ نیز موردبررسی قرار گرفته‌اند. لئو و همکاران [۲۲] با استفاده از چارچوب Vision Mamba (VMamba) نشان دادند که مدل‌سازی داده‌های تصویری به‌صورت دنباله می‌تواند پردازش تصاویر فوندوس با وضوح بالا را بهینه‌تر کند. این مطالعات زیربنای نظری مدل پیشنهادی ما را تشکیل می‌دهند که در آن تلاش شده است با استفاده از ماژول‌های توجه سلسله‌مراتبی، مزایای تمرکز بر ویژگی‌های محلی و سراسری در یک بستر کانولوشنی (ResNet) پیاده‌سازی شود.

باوجود موفقیت چشمگیر ترنسفورمرها و مکانیزم‌های توجه در ترکیب ویژگی‌های محلی و سراسری، محدودیت اساسی این مدل‌ها سربار محاسباتی بسیار سنگین، نیاز به داده‌های آموزشی عظیم و پیچیدگی استقرار آن‌ها در سیستم‌های غربالگری واقعی است. علاوه بر این، بسیاری از این مدل‌های هیبریدی از مکانیزم توجه صرفاً در لایه‌های انتهایی استفاده می‌کنند و پتانسیل لایه‌های میانی همچنان نادیده گرفته می‌شود.

مرور ادبیات نشان می‌دهد که اگرچه CNN ها در استخراج ویژگی‌های محلی توانمندند، اما جزئیات ریز را در لایه‌های عمیق از دست می‌دهند. از سوی دیگر، مدل‌های مبتنی بر ترنسفورمر علیرغم درک سراسری عالی، با چالش پیچیدگی محاسباتی و عدم بهینگی برای مجموعه داده‌های کوچک پزشکی مواجه‌اند. شکاف تحقیقاتی بارز در مطالعات پیشین، فقدان یک معماری سبک و کانولوشنی است که بتواند به کمک یک مکانیزم توجه سلسله‌مراتبی، ویژگی‌های بافتی و ویژگی‌های معنایی را به‌صورت هم‌زمان، هدفمند و مجزا استخراج و پالایش کند.

۳-۳ روش پیشنهادی

۳-۱ مقدمه و کلیات

در این بخش، معماری پیشنهادی برای تشخیص خودکار و درجه‌بندی رتینوپاتی دیابتی از تصاویر فوندوس شبکه ارائه می‌شود. با توجه به چالش‌های مطرح‌شده در فصول پیشین، نظیر

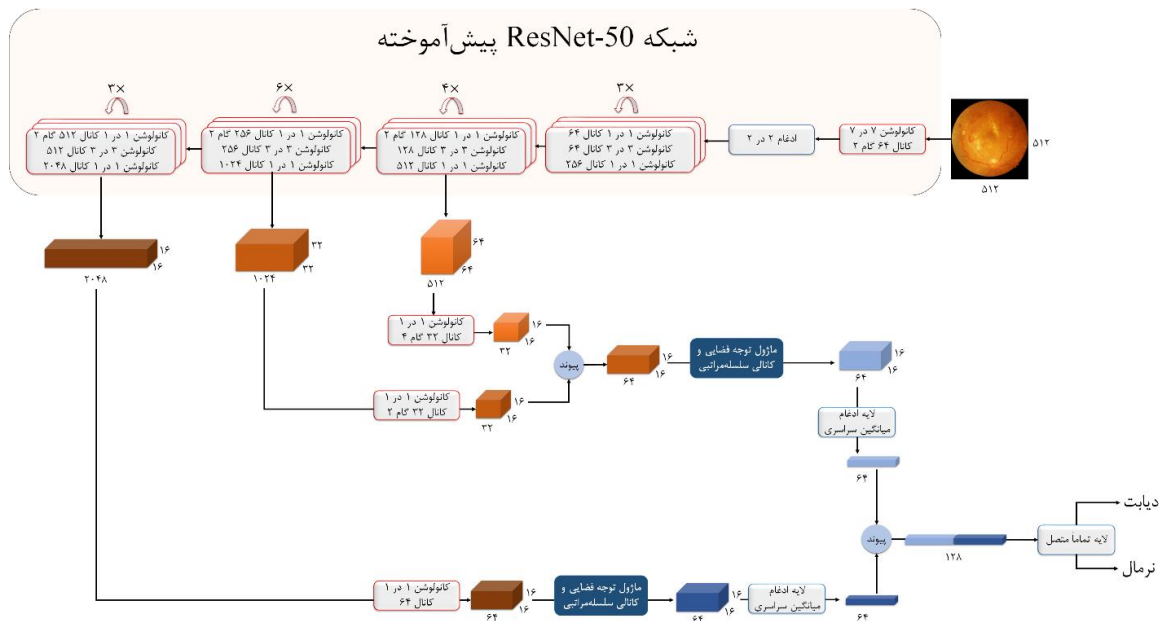
^۳ Skip Connections

^۱ State Space Models

^۲ Residual

داده‌های برچسب‌دار مواجه هستیم، منجر به همگرایی سریع‌تر و افزایش قدرت تعمیم‌پذیری مدل می‌گردد..

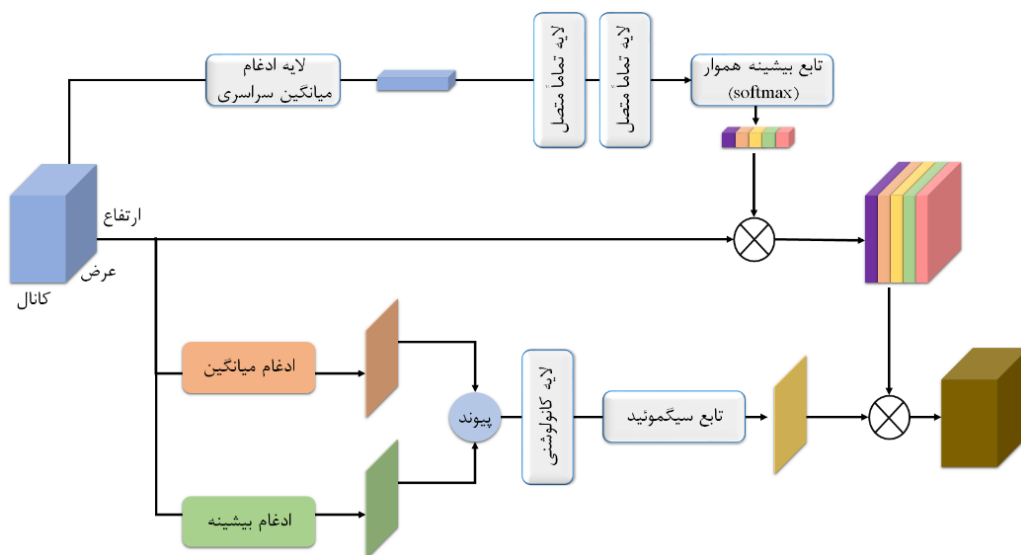
بصری (مانند لبه‌ها، بافت‌ها و اشکال هندسی) فرآیند یادگیری را آغاز کند. این امر به‌ویژه در کاربردهای پزشکی که با محدودیت



شکل (۱): معماری کلی روش پیشنهادی

خروجی بلوک ۲ دارای ۵۱۲ کانال است و ابعاد مکانی آن ۱/۸ تصویر اصلی است. این ویژگی‌ها حاوی جزئیات دقیق بافتی و مرزهای اشیاء هستند.

در روش پیشنهادی، ما از خروجی‌های سه مرحله آخر این شبکه استفاده می‌کنیم که هر کدام نمایانگر سطحی از انتزاع تصویر هستند.



شکل (۲): ماژول توجه فضایی و کانالی سلسله‌مراتبی

به افزایش چشمگیر بار محاسباتی در ماژول‌های توجه می‌گردد، ابعاد مکانی بلوک ۲ کاهش یافته و تعداد کانال‌های هر دو بلوک توسط لایه‌های کانولوشن ۱ در ۱ به صورت تهاجمی به ۳۲ کانال فشرده شده است. این کاهش ابعاد به عنوان یک فیلتر تنظیم کننده عمل کرده و شبکه را وادار می‌سازد تا منحصراً برجسته ترین الگوهای بافتی مرتبط با ضایعات را استخراج نماید. مزیت اصلی این طراحی، محدود نگه داشتن شدید تعداد پارامترهای آموزش پذیر و کاهش چشمگیر حجم عملیات ممیز شناور است که استقرار مدل را بر روی سخت افزارهای معمولی امکان پذیر می‌سازد.

ب) شاخه معنایی-سراسری (بلوک ۴)

این شاخه وظیفه درک کلیت تصویر را دارد. خروجی بلوک ۴ (۲۰۴۸ کانال) مستقیماً توسط یک کانولوشن 1×1 فشرده سازی شده و به ۶۴ کانال کاهش می‌یابد ($F_{global} \in R^{16 \times 16 \times 64}$). جداسازی این شاخه تضمین می‌کند که اطلاعات انتزاعی سطح بالا تحت تأثیر جزئیات ریز و نویزدار لایه‌های پایین قرار نگیرند.

۳-۵- ماژول توجه فضایی و کانالی سلسله مراتبی

قلب تپنده معماری پیشنهادی، ماژول توجه نوآورانه‌ای است که در هر دو شاخه شبکه تعبیه شده است. هدف این ماژول، تقلید از سیستم بینایی انسان در تمرکز بر نواحی مهم و نادیده گرفتن پس زمینه است. همان طور که در شکل (۲) نشان داده شده است، این ماژول ویژگی‌های ورودی F را طی دو مرحله توجه کانالی و توجه فضایی پالایش می‌کند.

الف) توجه کانالی

این بخش تعیین می‌کند چه ویژگی‌هایی مهم تر هستند. برخی کانال‌های ویژگی ممکن است حاوی اطلاعات نویز باشند، درحالی که برخی دیگر مستقیماً با ضایعات دیابتی مرتبط اند.

ابتدا ویژگی ورودی از یک لایه ادغام میانگین سراسری عبور می‌کند تا اطلاعات مکانی فشرده شده و بردار توصیف گر کانال $v \in R^{C \times 1 \times 1}$ به دست آید.

این بردار وارد یک شبکه MLP شامل دولایه تماماً متصل می‌شود تا وابستگی‌های غیرخطی بین کانال‌ها مدل سازی شود.

این در حالی است که خروجی بلوک ۳ با ۱۰۲۴ کانال و ابعاد مکانی $1/16$ ، ترکیبی از ویژگی‌های بافتی و معنایی را ارائه می‌دهد. اما خروجی نهایی کانولوشنی با ۲۰۴۸ کانال و ابعاد مکانی $1/32$ ، غنی ترین اطلاعات معنایی و انتزاعی (نظیر وجود کلی ناهنجاری‌ها) را در بردارد و ما از هر سه نوع این اطلاعات در شبکه استفاده می‌کنیم.

۳-۴ پردازش و تلفیق سلسله مراتبی ویژگی‌ها

یکی از چالش های اصلی در تشخیص رتینوپاتی دیابتی، تنوع اندازه ضایعات است؛ میکروآنورسم‌ها بسیار کوچک هستند و نیازمند ویژگی‌های سطح پایین، درحالی که خونریزی‌های وسیع یا اگزوداها نیازمند درک کلی تر تصویر هستند. برای مدیریت این چالش، معماری پیشنهادی جریان اطلاعات را به دو شاخه مجزا تقسیم می‌کند.

الف) شاخه محلی-بافتی (ترکیب بلوک ۲ و ۳)

این شاخه وظیفه استخراج و حفظ جزئیات ریز را بر عهده دارد. از آنجاکه خروجی بلوک ۲ (۵۱۲ کانال) و بلوک ۳ (۱۰۲۴ کانال) از نظر عمق و ابعاد متفاوت هستند، ابتدا یک فرآیند همسان سازی انجام می‌شود.

ابتدا با اعمال لایه‌های کانولوشن 1×1 ، تعداد کانال‌های هر دو خروجی به تعداد ثابت $C_{reduced} = 32$ کاهش می‌یابد. این کاهش بعد، سربار محاسباتی را به طور چشمگیری کم می‌کند.

سپس با استفاده از عملیات نمونه برداری، ابعاد مکانی هر دو نقشه به یک اندازه واحد (16×16) تبدیل می‌شود.

در نهایت، این دو نقشه ویژگی در راستای عمق (Channel-wise) الحاق می‌شوند تا یک تفسور ویژگی ترکیبی با ۶۴ کانال ($F_{local} \in R^{16 \times 16 \times 64}$) ایجاد گردد. این ویژگی ترکیبی، هم زمان دقت مکانی بلوک ۲ و غنای معنایی بلوک ۳ را داراست.

شایان ذکر است که در طراحی مسیر پردازش ویژگی‌های محلی (خروجی بلوک‌های ۲ و ۳)، یک استراتژی مبتنی بر ایجاد گلوگاه اطلاعاتی تعمدی اتخاذ شده است. با توجه به محدودیت‌های منابع محاسباتی در محیط‌های بالینی واقعی، اولویت این پژوهش طراحی مدلی تا حد امکان سبک و کارآمد بوده است. به همین جهت، به جای رویکرد متداول آپ سیمپلینگ بلوک‌های عمیق تر که منجر

اکثریت سوگیری پیدا کند و در تشخیص نمونه‌های دشوار دچار خطا شود.

برای غلبه بر این مشکل و تضمین همگرایی پایدار، ما از یک تابع ضرر ترکیبی استفاده کرده‌ایم که حاصل جمع وزن‌دار تابع ضرر آنتروپی متقاطع و تابع ضرر فوکال است.

این تابع ضرر در دو سناریوی طبقه‌بندی باینری (سالم در مقابل بیمار) و طبقه‌بندی چندکلاسه (درجه‌بندی شدت بیماری) آورده شده است.

سناریوی طبقه‌بندی باینری

در این حالت، هدف تشخیص وجود یا عدم وجود بیماری است. تابع ضرر ترکیبی L_{total} به صورت آورده شده در رابطه (۳) تعریف می‌شود.

$$L_{total} = \lambda L_{BCE} + (1 - \lambda) L_{Focal} \quad (3)$$

که در آن λ ضریب تعادل است.

آنتروپی متقاطع باینری برای حفظ پایداری کلی آموزش استفاده می‌شود و در رابطه (۴) آورده شده است.

$$L_{BCE}(p, y) = -(y \log(p) + (1 - y) \log(1 - p)) \quad (4)$$

ضرر فوکال برای تمرکز بر نمونه‌های سخت و جلوگیری از غلبه کلاس‌های آسان (سالم) بر گرادیان‌ها استفاده می‌شود و در رابطه (۵) آورده شده است.

$$L_{Focal}(p, y) = -(y(1 - p)^y \log(p) + (1 - y)p^y \log(1 - p)) \quad (5)$$

در این روابط، $y \in \{0, 1\}$ برچسب واقعی، p احتمال پیش‌بینی شده توسط مدل برای کلاس مثبت و γ پارامتر تمرکز است که برابر ۲ قرار داده شده است.

۷-۳ سناریوی طبقه‌بندی چندکلاسه

در حالت درجه‌بندی بیماری، فرمولاسیون به صورت تعمیم‌یافته نشان داده شده در روابط (۶) الی (۸) خواهد بود.

$$L_{total} = \lambda L_{CE} + (1 - \lambda) L_{Focal} \quad (6)$$

در نهایت، برخلاف روش‌های متداول که از سیگموئید استفاده می‌کنند، در اینجا طبق معماری از تابع فعال‌ساز بیشینه هموار^۱ استفاده شده است تا توزیع احتمال اهمیت کانال‌ها نرمال‌سازی شود. وزن‌های به دست آمده W_c در ویژگی ورودی ضرب می‌شوند. نحوه محاسبه آن در رابطه (۱) آورده شده است.

$$F' = F \otimes \text{Softmax} \left(\text{MLP}(\text{AvgPool}(F)) \right) \quad (1)$$

ب) توجه فضایی

این بخش تعیین می‌کند کجای تصویر مهم‌تر است. پس از پالایش کانالی، ویژگی F' وارد این مرحله می‌شود.

برای استخراج نقشه‌های اطلاعاتی، دو عملیات ادغام میانگین و ادغام بیشینه در راستای محور کانال اعمال می‌شود. این کار دو نقشه دوبعدی $M_{avg}, M_{max} \in R^{H \times W \times 1}$ تولید می‌کند.

این دو نقشه به یکدیگر الحاق شده و وارد یک لایه کانولوشن می‌شوند تا روابط مکانی همسایگی تحلیل شود.

خروجی از تابع فعال‌ساز سیگموئید عبور می‌کند تا ماسک توجه مکانی M_s (با مقادیر بین ۰ و ۱) تولید شود. در نهایت، این ماسک در ویژگی‌های مرحله قبل ضرب می‌شود تا خروجی نهایی ماژول به دست آید که جزئیات آن در رابطه (۲) آورده شده است.

$$F'' = F' \otimes \sigma \left(\text{Conv}^{7 \times 7}([M_{avg}; M_{max}]) \right) \quad (2)$$

نتیجه این فرآیند، ویژگی‌هایی است که هم از نظر محتوایی (کانال) و هم از نظر مکانی (فضا) پالایش شده و آماده طبقه‌بندی دقیق هستند.

۶-۳ تابع ضرر

یکی از چالش‌های اساسی در آموزش مدل‌های یادگیری عمیق برای تشخیص پزشکی، مسئله عدم تعادل کلاس‌ها^۲ است. در مجموعه داده‌های رتینوپاتی دیابتی، معمولاً تعداد تصاویر سالم (کلاس اکثریت) بسیار بیشتر از تصاویر دارای ضایعات شدید (کلاس اقلیت) است. استفاده از توابع ضرر معمول مانند آنتروپی متقاطع^۳ به تنهایی، ممکن است باعث شود مدل به سمت کلاس

³ Cross-Entropy

¹ Softmax

² Class Imbalance

الف) طبقه‌بندی باینری (۲ کلاس) که تفکیک تصاویر به دودسته سالم (کلاس صفر) و بیمار (کلاس‌های یک تا چهار) انجام شده است.

ب) درجه‌بندی شدت بیماری (پنج کلاس) که تشخیص دقیق مرحله پیشرفت بیماری در هر پنج کلاس فوق را نشان می‌دهد. همچنین برای نشان دادن قابلیت تعمیم روش پیشنهادی مقایسه روی مجموعه داده IDRiD نیز انجام شده است. مجموعه داده IDRiD شامل ۵۱۶ تصویر فوندوس است که از بیمارستان چشم‌پزشکی شیوا نترالایا در اوتار پرادش هند جمع‌آوری شده است [۲۴]. برخلاف مجموعه داده‌های بزرگ‌تر، این دادگان تمرکز بسیار زیادی بر روی موارد متوسط تا شدید رتینوپاتی دیابتی دارد که آن را برای ارزیابی عملکرد مدل در تشخیص پیشرفت بالینی بیماری به گزینه‌ای بسیار ارزشمند و البته چالش‌برانگیز تبدیل می‌کند. این مجموعه شامل حاشیه‌نویسی‌های دقیق برای درجه‌بندی بیماری و بخش‌بندی ضایعات است.

۴-۲ معیارهای ارزیابی و شرایط آزمایش و مقایسه

برای سنجش کمی عملکرد مدل، از چهار معیار استاندارد شامل صحت (Accuracy)، دقت (Precision)، فراخوانی (Recall) و امتیاز F1 استفاده شده است. این معیارها بر اساس مولفه‌های ماتریس درهم‌ریختگی شامل مثبت صحیح (TP)، منفی صحیح (TN)، مثبت کاذب (FP) و منفی کاذب (FN) تعریف می‌شوند. صحت: نسبت کل پیش‌بینی‌های درست به کل داده‌ها است که در رابطه (۹) نشان داده شده است.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

دقت: نسبت تشخیص‌های درست مثبت به کل موارد پیش‌بینی شده به‌عنوان مثبت است که در رابطه (۱۰) نشان داده شده است.

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

فراخوانی: نسبت تشخیص‌های درست مثبت به کل موارد واقعاً مثبت (حساسیت مدل) است که در رابطه (۱۱) نشان داده شده است.

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

که آنتروپی متقاطع دسته‌ای

$$L_{CE} = - \sum_{c=1}^C y_c \log(p_c) \quad (7)$$

ضرر فوکال چندکلاس

$$L_{Focal} = - \sum_{c=1}^C (1 - p_c)^{\gamma} y_c \log(p_c) \quad (8)$$

که در آن C تعداد کل کلاس‌ها، y_c نشانگر باینری برای کلاس c و p_c احتمال پیش‌بینی شده برای آن کلاس است.

علت استفاده هم‌زمان از این دو تابع این است که L_{CE} گرایان‌های پایداری ایجاد می‌کند و از واگرایی مدل در اوایل آموزش جلوگیری می‌نماید، در حالیکه L_{Focal} در مراحل پایانی آموزش، وزن خطای نمونه‌هایی که مدل در تشخیص آن‌ها مشکل دارد را افزایش می‌دهد.

۴-۳ ارزیابی روش پیشنهادی

۴-۱ مجموعه داده‌گان

برای ارزیابی کارایی مدل پیشنهادی، از مجموعه داده معتبر و استاندارد APTOS 2019 استفاده شده است که توسط بیمارستان چشم‌پزشکی آراویند (Aravind Eye Hospital) در هند جمع‌آوری شده است [۲۳]. این پایگاه داده شامل ۳۶۶۲ تصویر رنگی فوندوس با کیفیت بالا است که از بیماران در مناطق روستایی و تحت شرایط تصویربرداری متنوع اخذ شده‌اند. تمامی تصاویر توسط تیمی از متخصصان بالینی بازبینی شده و بر اساس مقیاس بین‌المللی شدت رتینوپاتی دیابتی (ICDRSS) درجه‌بندی شده‌اند. این مجموعه داده دارای ۵ کلاس شدت بیماری به‌صورت زیر است:

کلاس صفر: بدون رتینوپاتی

کلاس یک: خفیف

کلاس دو: متوسط

کلاس سه: شدید

کلاس چهار: پرولیفراتیو

در آزمایش‌های این پژوهش، به‌منظور سنجش جامع مدل، ارزیابی‌ها در دو سناریوی متفاوت انجام شده است.

مواجه می‌گردد که قدرت تعمیم‌پذیری آن را به میزان قابل توجهی ارتقا می‌دهد.

همچنین لازم به ذکر است که به‌منظور انجام یک مقایسه کاملاً عادلانه و استاندارد، نتایج گزارش‌شده برای تمامی مدل‌های پایه از مقالات دیگر نقل نشده‌اند؛ بلکه تمامی این معماری‌ها توسط نویسندگان و تحت شرایط تنظیمات آموزشی کاملاً یکسان (شامل نرخ یادگیری، اندازه دسته، توابع بهینه‌ساز و استفاده از همین تقسیم‌بندی داده‌های آموزش و آزمون) پیاده‌سازی و بازآموزی شده‌اند تا تفاوت عملکرد نهایی صرفاً ناشی از ساختار شبکه‌ها باشد.

۴-۳ نتایج کمی

در این بخش به مقایسه روش پیشنهادی با روش‌های دیگر می‌پردازیم. روش پیشنهادی در دو حالت باینری و چندکلاسه ارزیابی شده است.

در جدول (۱)، عملکرد روش پیشنهادی در تشخیص وجود یا عدم وجود بیماری در مقایسه با مدل‌های پایه شناخته‌شده آورده شده است. همان‌طور که مشاهده می‌شود، روش پیشنهادی با بهره‌گیری از معماری انشعابی و ماژول توجه سلسله‌مراتبی، به صحت ۹۷/۲۳٪ دست‌یافته است که بهبود قابل‌توجهی نسبت به مدل ResNet50 خام نشان می‌دهد.

این نتایج نشان می‌دهند که ترکیب ویژگی‌های سطح پایین و بالا، توانسته است حساسیت مدل را به ۹۶/۹۰٪ برساند که در کاربردهای غربالگری پزشکی برای اطمینان از عدم تشخیص منفی کاذب (بیمار سالم تلقی نشود) بسیار حیاتی است.

علاوه بر بهبود نسبت به مدل پایه، روش پیشنهادی در مقایسه با سایر معماری‌های عمیق نظیر VGG19 و Xception نیز برتری خود را حفظ کرده است. اگرچه شبکه‌ای مانند VGG19 عملکرد رقابتی نزدیکی از خود نشان داده، اما مدل ما با دستیابی به امتیاز F1 برابر با ۹۷/۱۷٪، تعادل پایدارتری میان معیار دقت و فراخوانی ایجاد کرده است. این دستاورد مؤید آن است که استراتژی پیشنهادی در استخراج ویژگی‌های سلسله‌مراتبی و استفاده از ماژول توجه، ضعف‌های ساختاری مدل‌های کانولوشنی متداول در نادیده

امتیاز F1: میانگین همساز دقت و فراخوانی که معیاری متعادل برای ارزیابی است که در رابطه (۱۲) نشان داده‌شده است.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (12)$$

روابط فوق برای حالت باینری صادق هستند. برای ارزیابی در حالت چند کلاسه، با توجه به عدم تعادل تعداد نمونه‌ها در کلاس‌های مختلف دادگان APTOS، استفاده از روش میانگین‌گیری وزن‌دار از منظر آماری تطابق بیشتری با توزیع داده‌ها دارد. با این وجود، در این پژوهش به‌منظور حفظ پروتکل‌های استاندارد ارزیابی و امکان انجام یک مقایسه کاملاً عادلانه و یک‌به‌یک با مدل‌های پایه و سایر مطالعات پیشین، از استراتژی میانگین‌گیری ماکرو استفاده‌شده است. در این روش، معیارها برای هر کلاس به‌صورت جداگانه محاسبه‌شده و سپس میانگین حسابی آن‌ها بدون در نظر گرفتن وزن کلاس‌ها در نظر گرفته می‌شود تا شرایط آزمایش و گزارش نتایج برای معماری پیشنهادی دقیقاً مشابه با سایر روش‌های مقایسه‌شده در این حوزه باشد.

در راستای انجام آزمایش‌ها، پیش از ورود تصاویر به شبکه، عملیات پیش‌پردازش شامل تغییر اندازه تمامی تصاویر به رزولوشن استاندارد ۵۱۲ در ۵۱۲ پیکسل و نرمال‌سازی مقادیر شدت روشنایی انجام گرفت. مجموعه داده استاندارد APTOS 2019 شامل ۳۶۶۲ تصویر است که به‌صورت تصادفی اما با حفظ تناسب توزیع کلاس‌ها به سه بخش تقسیم گردید. این بخش‌ها شامل ۸۰ درصد از تصاویر (۲۹۳۰ تصویر) برای آموزش، ۱۰ درصد (۳۶۶ تصویر) برای اعتبارسنجی و تنظیم ابرپارامترها و ۱۰ درصد باقی‌مانده (۳۶۶ تصویر) به‌عنوان داده‌های آزمون مجزا برای ارزیابی نهایی مدل هستند.

به‌منظور مقابله با چالش کمبود داده و جلوگیری از بیش‌برازش مدل، از تکنیک‌های افزایش داده شامل چرخش تصادفی، قرینه‌سازی افقی و عمودی، و تغییرات تصادفی در روشنایی و کنتراست استفاده شد. شایان ذکر است که فرآیند افزایش داده به‌صورت پویا و در حین بارگذاری داده‌ها در هر اپاک انجام می‌شود؛ بنابراین، تعداد فیزیکی نمونه‌های آموزشی روی دیسک ثابت باقی می‌ماند، اما شبکه در هر تکرار با نسخه‌های تغییر یافته و جدیدی از تصاویر

گرفتن جزئیات ریزبافتی را جبران کرده و منجر به تفکیک‌پذیری دقیق‌تر در فضای ویژگی شده است.

جدول (۱): ارزیابی روش پیشنهادی در مقایسه با روش‌های پایه در دیتاست APTOS 2019 حالت باینری

مدل	صحت	دقت	فراخوانی	امتیاز F1
VGG16	٪۹۵/۴۰	٪۹۵/۶۳	٪۹۴/۵۱	٪۹۵/۰۲
VGG19	٪۹۶/۰۰	٪۹۶/۱۳	٪۹۵/۱۶	٪۹۵/۶۲
ResNet50	٪۹۲/۴۰	٪۹۲/۱۷	٪۹۱/۳۵	٪۹۱/۶۸
MobileNet	٪۹۲/۲۰	٪۹۱/۵۶	٪۹۰/۴۹	٪۹۰/۸۷
Xception	٪۹۲/۰۰	٪۹۱/۶۰	٪۹۰/۷۲	٪۹۱/۰۸
Khan و همکاران [۲۵]	٪۸۶/۲۲	-	-	-
Sikder و همکاران [۲۶]	٪۹۴/۲۰	-	-	٪۹۶/۴۷
Folorunsho و همکاران [۲۷]	٪۹۵/۶۳	٪۹۷/۷۹	٪۹۳/۶۴	٪۹۸/۷۹
روش پیشنهادی	٪۹۷/۲۳	٪۹۷/۴۵	٪۹۶/۹۰	٪۹۷/۱۷

در شبکه و تمرکز هدفمند بر نواحی حاوی ضایعات ریز، قدرت تمایزپذیری شبکه را به میزان قابل‌توجهی ارتقا داده و از خطای طبقه‌بندی در کلاس‌های دشوار جلوگیری کرده است. افت کارایی در حالت پنج کلاسه نسبت به حالت باینری امری طبیعی است، زیرا تمایز بین کلاس‌های نزدیک (مثلاً خفیف و متوسط) بسیار دشوار است.

جدول (۲) نتایج چالش‌برانگیزتر تعیین شدت بیماری را نشان می‌دهد. در اینجا روش پیشنهادی با مدل‌های پیشرفته اخیر (State-of-the-art) مقایسه شده است. این برتری عملکرد نسبت به مدل‌های پیشین تصادفی نبوده و مستقیماً ریشه در معماری دوشاخه‌ای پیشنهادی دارد. به‌طور دقیق‌تر، دلیل اصلی این بهبود، به‌کارگیری ماژول توجه فضایی و کانالی در مسیر ویژگی‌های میانی است که با سرکوب انطباقی نویزهای پس‌زمینه (مانند بازتاب نور

جدول (۲): ارزیابی روش پیشنهادی در مقایسه با روش‌های پیشرفته اخیر در دیتاست APTOS 2019 حالت پنج کلاسه

مدل	صحت	دقت	فراخوانی	امتیاز F1
ResNet34 [۲۸]	٪۷۹/۰۰	٪۶۱/۰۰	٪۶۳/۰۰	٪۶۰/۰۰
EfficientNet_B0 [۲۹]	٪۷۹/۰۰	٪۶۳/۰۰	٪۵۹/۰۰	٪۶۱/۰۰
EfficientNet + SVM [۲۹]	٪۸۱/۰۰	٪۶۴/۰۰	٪۵۵/۰۰	٪۵۷/۰۰
DenseNet201 [۳۰]	٪۷۸/۳۳	٪۶۳/۷۷	٪۶۱/۲۶	٪۶۱/۸۵
Quantum CNN [۳۱]	٪۷۲/۴۰	٪۴۹/۵۷	٪۴۹/۹۵	٪۴۹/۵۰
Shi و همکاران [۳۲]	٪۷۵/۶۱	-	-	-
Mohanty و همکاران [۳۳]	٪۷۹/۵۰	-	-	-
Gangwar و همکاران [۳۴]	٪۸۲/۱۸	-	-	-
روش پیشنهادی	٪۸۳/۰۶	٪۷۰/۲۸	٪۶۳/۹۸	٪۶۵/۸۰

افت کارایی در حالت ۵ کلاسه نسبت به حالت باینری امری طبیعی است، زیرا تمایز بین کلاس‌های نزدیک (مثلاً خفیف و متوسط) بسیار دشوار است. باین حال، امتیاز F1 بالاتر مدل پیشنهادی نسبت

روش ما با دستیابی به صحت ٪۸۳/۰۶ و امتیاز F1 برابر با ٪۶۵/۸۰، عملکرد بهتری نسبت به روش‌های تک‌مسیره مانند ResNet34 یا روش‌های ترکیبی مانند EfficientNet+SVM داشته است.

به‌منظور اثبات قدرت تعمیم‌پذیری معماری پیشنهادی و اطمینان از عدم بیش‌برازش روی یک توزیع داده خاص، عملکرد مدل بر روی مجموعه داده چالش‌برانگیز IDRiD نیز مورد ارزیابی قرار گرفت. همان‌طور که در جدول (۳) مشاهده می‌شود، شبکه پیشنهادی با دستیابی به صحت ۵۹/۳۱ درصد، موفق شده است عملکرد بهتری نسبت به مدل پایه ResNet50 و همچنین سایر روش‌های پیشرفته اخیر نظیر Explicd و PCBM ثبت نماید.

جدول (۳): ارزیابی روش پیشنهادی در مقایسه با مدل‌های پیشرفته

روی مجموعه داده IDRiD

مدل	صحت
LaBo [۳۵]	۴۸/۹۷٪
PCBM [۳۶]	۵۲/۴۶٪
ResNet50	۵۷/۸۲٪
Explicd [۳۷]	۵۸/۶۳٪
روش پیشنهادی	۵۹/۳۱٪

افت کلی سطح صحت تمامی مدل‌ها در این مجموعه داده نسبت به دادگان APTOS کاملاً توجیه‌پذیر است؛ زیرا مجموعه داده IDRiD دارای حجم نمونه بسیار محدودتر (تنها ۵۱۶ تصویر) بوده و توزیع آن به‌شدت به سمت کلاس‌های دشوار و مراحل پیشرفته بیماری (متوسط تا شدید) متمایل است که تفکیک آن‌ها از یکدیگر پیچیدگی بالایی دارد. با این وجود، برتری مدل پیشنهادی در این شرایط سخت، نشان‌دهنده اثربخشی ماژول توجه فضایی-کانالی و ساختار استخراج ویژگی سلسله‌مراتبی در شناسایی الگوهای ظریف بافتی است که حتی در شرایط کمبود داده و چالش‌برانگیز بودن توزیع کلاس‌ها نیز پایداری و کارایی خود را به خوبی حفظ می‌کند.

۴-۴ مطالعات فرسایش

به‌منظور صحت‌سنجی فرضیات مطرح‌شده در این پژوهش و اندازه‌گیری دقیق سهم هر یک از اجزای پیشنهادی در ارتقای عملکرد نهایی، مجموعه‌ای از آزمایش‌های کنترل‌شده بر روی مجموعه داده APTOS (در سناریوی طبقه‌بندی باینری) طراحی و اجرا گردید. هدف از این آزمایش‌ها، پاسخ به دو پرسش اساسی است: نخست اینکه آیا استراتژی تلفیق ویژگی‌های لایه‌های میانی (بلوک ۲ و ۳) برتری معناداری نسبت به استفاده تک‌لایه دارد و

به رقبای نشان می‌دهد که روش ما در مدیریت کلاس‌های نامتوازن موفق‌تر عمل کرده است. این موفقیت ناشی از دو عامل است: (۱) استفاده از تابع ضرر ترکیبی فوکال که تمرکز را روی نمونه‌های سخت می‌گذارد و (۲) توجه هم‌زمان به جزئیات.

به‌منظور بررسی دقیق‌تر عملکرد مدل در سناریوی چالش‌برانگیز درجه‌بندی شدت بیماری (۵ کلاس) و ارزیابی توانایی شبکه در تفکیک کلاس‌های اقلیت، ماتریس درهم‌ریختگی مربوط به مجموعه داده آزمون در شکل (۳) رسم شده است. همان‌طور که مشاهده می‌شود، مدل در تشخیص نمونه‌های سالم (کلاس صفر) با دقت بسیار بالایی عمل کرده است. در کلاس‌های میانی، عمده خطاهای طبقه‌بندی مربوط به کلاس‌های مجاور است (به‌عنوان مثال اشتباه گرفتن کلاس یک با کلاس دو) که به دلیل شباهت بسیار زیاد بصری ضایعات در مراحل گذار بیماری امری طبیعی است. نکته حائز اهمیت، عملکرد مدل در کلاس‌های بحرانی و پیشرفته (کلاس‌های شدید و پرولیفراتیو) است؛ علی‌رغم محدودیت شدید تعداد نمونه‌ها در این دو کلاس نسبت به سایر دسته‌ها، شبکه توانسته است با بهره‌گیری از مکانیزم توجه سلسله‌مراتبی و تابع ضرر ترکیبی، ویژگی‌های متمایزکننده این ضایعات را استخراج کرده و از طبقه‌بندی اشتباه آن‌ها به‌عنوان نمونه کاملاً سالم تا حد بسیار بالایی جلوگیری کند. این توزیع خطا در ماتریس درهم‌ریختگی تأیید می‌کند که روش پیشنهادی در مدیریت عدم تعادل کلاس‌ها موفق عمل کرده است.

	سالم	خفیف	کلاس پیش‌بینی شده متوسط	شدید	پرولیفراتیو
سالم	197	2	0	0	0
خفیف	10	12	8	0	0
کلاس واقعی متوسط	0	2	69	11	5
شدید	0	0	6	8	3
پرولیفراتیو	3	0	11	1	18

شکل (۳): ماتریس درهم‌ریختگی

نمونه‌برداری کاهشی، در تشخیص ضایعات کوچک مانند میکروآنوريسم‌های اولیه دچار ضعف است.

با افزودن ویژگی‌های بلوک ۲ و ۳ به‌صورت جداگانه (ردیف‌های دوم تا چهارم جدول (۴))، شاهد بهبود تدریجی در عملکرد هستیم. این امر نشان می‌دهد که بازیابی اطلاعات مکانی از لایه‌های کم‌عمق‌تر، به مدل کمک می‌کند تا الگوهای بافتی را بهتر شناسایی کند. با این حال، استفاده هم‌زمان از هر سه بلوک به‌صورت مستقل باوجود رسیدن به صحت ۹۵/۶۰٪، سربار محاسباتی را افزایش می‌دهد.

دوم، آیا تابع ضرر ترکیبی پیشنهادی در مدیریت عدم تعادل داده‌ها موفق‌تر از توابع استاندارد عمل می‌کند.

در جدول (۴)، سیر تکاملی عملکرد مدل با افزودن تدریجی ویژگی‌های خروجی از بلوک‌های مختلف شبکه ResNet50 و اعمال مکانیزم‌های پیشنهادی نمایش داده شده است.

همان‌طور که در ردیف اول جدول (۴) مشاهده می‌شود، استفاده انحصاری از خروجی بلوک ۴ که رویکرد استاندارد در اکثر روش‌های یادگیری انتقالی است، منجر به صحتی معادل ۹۲/۴۰٪ می‌گردد. اگرچه این لایه حاوی غنی‌ترین اطلاعات معنایی است، اما به دلیل از دست دادن رزولوشن مکانی در طی فرآیندهای

جدول (۴): مطالعه فرسایش بر روی معماری شبکه و تأثیر ماژول‌های پیشنهادی

ردیف	خروجی بلوک ۲	خروجی بلوک ۳	خروجی بلوک ۴	ماژول توجه	صحت	دقت	فراخوانی	امتیاز F1	پارامتر قابل آموزش	حجم محاسبات (GFLOPs)	زمان استنتاج برای هر تصویر (ms)
۱			✓		۹۲/۴۰ ± ۰/۴۲	۹۲/۱۵ ± ۰/۴۵	۹۱/۳ ± ۰/۴۸	۹۱/۷۲ ± ۰/۴۶	۰/۶۵	۸۶/۵	۴۷/۲
۲		✓	✓		۹۴/۱۵ ± ۰/۳۸	۹۳/۸ ± ۰/۴۰	۹۳/۵ ± ۰/۴۱	۹۳/۶۵ ± ۰/۳۹	۰/۷۵	۸۶/۹	۴۷/۳
۳		✓		✓	۹۴/۸ ± ۰/۳۵	۹۴/۲ ± ۰/۳۸	۹۴/۵۵ ± ۰/۳۶	۹۴/۳۷ ± ۰/۳۷	۰/۸۲	۸۷/۲	۴۷/۳
۴	✓	✓	✓		۹۵/۶ ± ۰/۳۱	۹۵/۱ ± ۰/۳۳	۹۵/۳ ± ۰/۳۴	۹۵/۲ ± ۰/۳۲	۱/۱	۸۷/۹	۴۸/۱
۵	ترکیب دو بلوک		✓		۹۶/۱ ± ۰/۲۸	۹۶/۰۵ ± ۰/۲۶	۹۵/۸ ± ۰/۲۹	۹۵/۹۲ ± ۰/۲۷	۱/۰۵	۸۷/۶	۴۸/۲
۶	ترکیب دو بلوک		✓	✓	۹۷/۲۳ ± ۰/۱۷	۹۷/۴۵ ± ۰/۱۵	۹۷/۹ ± ۰/۲۱	۹۷/۱۷ ± ۰/۱۷	۱/۲	۸۸/۴	۴۸/۹

(۴) عمل کرده و شکاف معنایی بین ویژگی‌های سطح پایین و بالا را پر می‌کند.

در نهایت، با افزودن ماژول توجه فضایی و کانالی سلسله‌مراتبی به ساختار فوق، تمامی معیارهای ارزیابی به اوج خود می‌رسند. صحت مدل با یک جهش ۱/۱ درصدی به ۹۷/۲۳٪ و فراخوانی به ۹۶/۹۰٪ افزایش می‌یابد. این بهبود چشم‌گیر نشان‌دهنده توانایی ماژول توجه در سرکوب نویزهای پس‌زمینه (مانند بازتاب نور یا عروق سالم) و تمرکز هوشمند بر نواحی حاوی ضایعه است. درواقع، ماژول توجه

نوآوری معماری ما در ردیف پنجم مشهود است؛ جایی که خروجی‌های بلوک ۲ و ۳ ابتدا هم‌مقیاس شده و سپس در یک شاخه واحد ادغام می‌شوند. این رویکرد که منجر به صحت ۹۶/۱۰٪ شده است، اثبات می‌کند که ترکیب جزئیات لبه‌ها (از بلوک ۲) با الگوهای بافتی پیچیده‌تر (از بلوک ۳) در یک فضای ویژگی مشترک، بازنمایی قدرتمندتری نسبت به استفاده جداگانه آن‌ها ایجاد می‌کند. این شاخه ترکیبی به‌عنوان مکمل شاخه معنایی (بلوک

۵- نتیجه‌گیری

در این پژوهش، یک معماری نوین مبتنی بر یادگیری عمیق برای تشخیص و درجه‌بندی خودکار رتینوپاتی دیابتی ارائه شد که هدف اصلی آن غلبه بر چالش برقراری تعادل میان استخراج ویژگی‌های محلی و درک زمینه سراسری تصویر بود. ساختار پیشنهادی با بهره‌گیری از شبکه ResNet-50 به‌عنوان ستون فقرات و تفکیک جریان اطلاعات در دو مسیر مجزا، توانست جزئیات دقیق بافتی (مانند میکروآنورسم‌ها) را از لایه‌های میانی بازیابی کرده و با اطلاعات معنایی عمیق تلفیق نماید. تعبیه «ماژول توجه فضایی و کانالی سلسله‌مراتبی» در این مسیرها، نقش کلیدی در پالایش ویژگی‌ها و تمرکز هوشمند مدل بر نواحی ضایعه ایفا کرد. نتایج ارزیابی بر روی مجموعه داده APTOS 2019 نشان داد که روش پیشنهادی با دستیابی به صحت ۹۷/۲۳٪ در تشخیص باینری و ۸۳/۰۶٪ در درجه‌بندی چندکلاسه، عملکردی برتر نسبت به مدل‌های پیشرفته اخیر داشته و استفاده از تابع ضرر ترکیبی (آنتروپی متقاطع و فوکال)، پایداری مدل را در برابر عدم تعادل کلاس‌ها تضمین کرده است.

باوجود دستاوردهای پژوهش حاضر، این مطالعه با محدودیت‌هایی نیز روبرو است. نخست آنکه، آموزش و ارزیابی مدل منحصراً بر روی تصاویر فوندوس دوبعدی انجام شده و از اطلاعات بالینی تکمیلی بیمار (نظیر سطح قند خون یا سوابق پزشکی) استفاده نشده است. محدودیت دوم، چالش استقرار مستقیم این معماری دوشاخه‌ای بر روی تجهیزات لبه با منابع پردازشی بسیار محدود است. درنهایت، وابستگی مدل به برجسب‌های سطح تصویر باعث می‌شود که سیستم فعلی، مکان دقیق و پیکسلی ضایعات را مشخص نکند. با توجه به این محدودیت‌ها، به‌عنوان مسیرهای پژوهشی آتی، توسعه مدل با رویکردهای چندمدالیت و تلفیق تصاویر فوندوس با داده‌های توموگرافی انسجام نوری (OCT) و سوابق بالینی پیشنهاد می‌گردد. همچنین، فشرده‌سازی و سبک‌سازی شبکه جهت استقرار بر روی پلتفرم‌های موبایل و بهره‌گیری از تکنیک‌های یادگیری با نظارت ضعیف برای مکان‌یابی دقیق ضایعات بدون نیاز به برجسب‌گذاری پرهزینه، از اولویت‌های پژوهشی آینده محسوب می‌شود.

باعث می‌شود ویژگی‌های استخراج‌شده از شاخه ترکیبی، خالص‌سازی شده و تنها اطلاعات مرتبط با بیماری به طبقه‌بند نهایی ارسال گردد.

از منظر تحلیل پیچیدگی محاسباتی، مشاهده می‌شود که اگرچه مدل نهایی (ردیف ششم جدول (۴)) دارای بیشترین تعداد پارامترهای آموزش‌پذیر (۱/۲ میلیون)، بالاترین حجم محاسبات (۸۸/۴ گیگا فلوپس) و زمان استنتاج (۴۹/۸ میلی‌ثانیه) است، اما این افزایش سربار نسبت به مدل‌های پایه بسیار ناچیز و کاملاً قابل چشم‌پوشی است. حفظ زمان استنتاج در محدوده کمتر از ۵۰ میلی‌ثانیه در کنار ارتقای چشمگیر دقت، اثبات می‌کند که معماری پیشنهادی تعادل بسیار مطلوبی میان کارایی تشخیص و هزینه محاسباتی برقرار کرده است.

یکی دیگر از چالش‌های مهم در آموزش مدل‌های تشخیص پزشکی، عدم تعادل ذاتی داده‌هاست که می‌تواند منجر به سوگیری مدل به سمت کلاس سالم (اکثریت) شود. در جدول (۵)، تأثیر استفاده از تابع ضرر ترکیبی در مقایسه با توابع رایج بررسی شده است.

جدول (۵): مطالعه فرسایش بر روی تأثیر توابع ضرر مختلف

تابع ضرر	صحت	دقت	فراخوانی	امتیاز F1
آنتروپی متقاطع	۹۶/۴۵٪	۹۶/۴۰٪	۹۵/۸۰٪	۹۶/۱۰٪
فوکال	۹۶/۸۰٪	۹۶/۳۰٪	۹۶/۸۰٪	۹۶/۵۵٪
ترکیب دو تابع	۹۷/۲۳٪	۹۷/۴۵٪	۹۶/۹۰٪	۹۷/۱۷٪

استفاده از آنتروپی متقاطع به‌نهایی، اگرچه همگرایی مناسبی دارد، اما در معیار فراخوانی که شاخص حساسیت به بیماری است، عملکرد پایین‌تری را نشان می‌دهد. در مقابل، تابع ضرر فوکال با تمرکز بر نمونه‌های دشوار، فراخوانی را بهبود می‌بخشد اما ممکن است در مراحل اولیه آموزش ناپایدار باشد. نتایج نشان می‌دهد که رویکرد ترکیبی ما تعادل بهینه‌ای را ایجاد کرده است. در این حالت، مدل هم از پایداری گرادین‌های آنتروپی متقاطع بهره می‌برد و هم با مکانیزم تمرکز فوکال، حساسیت خود را نسبت به نمونه‌های بیمار حفظ می‌کند. دستیابی به امتیاز F1 برابر با ۹۷/۱۷٪ در حالت ترکیبی، اثبات می‌کند که این تابع هزینه توانسته است نرخ مثبت‌های کاذب و منفی‌های کاذب را به‌طور هم‌زمان کاهش دهد.

Refrences

- [1] N. Jagadisha, A. P. Prabhu, S. Abhay, C. Nayak, and C. S. Honavarker, "Retinaguard: An optimal cnn solution for diabetic retinopathy," in 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), vol. 1. IEEE, 2024, pp. 1-7.
- [2] Kumar, L. Nelson, and D. Arumugam, "Advanced cnn-based approach for early diabetic retinopathy detection using retinal images," in 2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit). IEEE, 2024, pp. 399-404.
- [3] Mushtaq, Gazala, and Farheen Siddiqui. "Detection of diabetic retinopathy using deep learning methodology." IOP conference series: materials science and engineering. Vol. 1070. No. 1. IOP Publishing, 2021.
- [4] S. Zhu, C. Xiong, Q. Zhong, and Y. Yao, "Diabetic retinopathy classification with deep learning via fundus images: A short survey," IEEE Access, vol. 12, pp. 20540-20558, 2024.
- [5] Lin, Chun-Ling, and Kun-Chi Wu. "Development of revised ResNet-50 for diabetic retinopathy detection." BMC bioinformatics 24.1 (2023): 157.
- [6] Mutawa, S. Alnajdi, and S. Sruthi, "Transfer learning for diabetic retinopathy detection: A study of dataset combination and model performance," Applied Sciences, vol. 13, no. 9, p. 5685, 2023.
- [7] M. Youldash, et al., "Early detection and classification of diabetic retinopathy: A deep learning approach," AI, vol. 5, no. 4, 2024.
- [8] M. V. Kumar, et al., "Diabetic retinopathy classification using quantum-assisted deep learning," in Evolution in Computational Intelligence: FICTA 2024, Springer Nature, 2025.
- [9] Wang, Hao, et al. "Many-MobileNet: Multi-model Augmentation for Robust Retinal Disease Classification." MICCAI Challenge on Ultra-Widefield Fundus Imaging for Diabetic Retinopathy. Cham: Springer Nature Switzerland, 2024. 144-154.
- [10] R. Bala, A. Sharma, and N. Goel, "Comparative analysis of diabetic retinopathy classification approaches using machine learning and deep learning techniques," Archives of Computational Methods in Engineering, vol. 31, no. 2, pp. 919-955, 2024.
- [11] Zhang, Yiye, et al. "Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women." Journal of affective disorders 279 (2021): 1-8.
- [12] G. Tăbăcaru, et al., "A robust machine learning model for diabetic retinopathy classification," Journal of Imaging, vol. 10, no. 1, p. 8, 2023.
- [13] S. Lakhera and A. Garg, "Diabetic retinopathy classification using psosvm based deep learning model," in 2023 Seventh International Conference on Image Information Processing (ICIIP). IEEE, 2023.
- [14] P. K. Ramanaiah, et al., "Diabetic retinopathy classification using various machine learning techniques," TPM-Testing, Psychometrics, Methodology in Applied Psychology, vol. 32, no. S5, pp. 1175-1183, 2025.
- [15] M. Youldash, et al., "Early detection and classification of diabetic retinopathy: A deep learning approach," AI, vol. 5, no. 4, 2024.
- [16] A. Mutawa, S. Alnajdi, and S. Sruthi, "Transfer learning for diabetic retinopathy detection: A study of dataset combination and model performance," Applied Sciences, vol. 13, no. 9, p. 5685, 2023.
- [17] A. Jabbar, et al., "A lesion-based diabetic retinopathy detection through hybrid deep learning model," IEEE Access, vol. 12, pp. 40019-40036, 2024.
- [18] M. Santos, et al., "Predicting diabetic retinopathy stage using siamese convolutional neural network," Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, vol. 12, no. 1, p. 2297017, 2024.
- [19] M. Touati, et al., "Drect: Enhancing diabetic retinopathy classification with a compact convolutional transformer," Big Data and Cognitive Computing, vol. 9, no. 1, p. 9, 2025.
- [20] H. Xu, et al., "A hybrid neural network approach for classifying diabetic retinopathy subtypes," Frontiers in Medicine, vol. 10, p. 1293019, 2024.
- [21] Kakade, Gaurang, and Chaitanya Kakade. "Optimal Detection of Diabetic Retinopathy Severity Levels Using Attention-Based CNN and Vision Transformers (ViT)." 2024 International Conference on Modeling, Simulation & Intelligent Computing (MoSiCom). IEEE, 2024.
- [22] Z. Liu, et al., "Identification of diabetic retinopathy lesions in fundus images by integrating cnn and vision mamba models," PLoS One, vol. 20, no. 1, p. e0318264
- [23] Asia Pacific Tele-Ophthalmology Society, "Aptos 2019 blindness detection," <https://www.kaggle.com/c/aptos2019-blindness-detection>, 2019, accessed: 2024.

- [24] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, F. Meriaudeau, Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research, *Data*, 3(3) (2018) 25.
- [25] Khan, Zubair, et al. "Diabetic retinopathy detection using VGG-NIN a deep learning architecture." *IEEE Access* 9 (2021): 61408-61416.
- [26] Sikder, Niloy, et al. "Severity classification of diabetic retinopathy using an ensemble learning algorithm through analyzing retinal images." *Symmetry* 13.4 (2021): 670.
- [27] Folorunsho, O., et al. "Explainable ensemble deep learning model for predicting diabetic retinopathy based on APTOS 2019 eye pack dataset." *LAUTECH Journal of Engineering and Technology* 19.1 (2025): 1-14.
- [28] P. Chen, "Jetson nano v1: Resnet34 implementation," Kaggle Notebook, 2024. [Online]. Available: <https://www.kaggle.com/code/peiyuchen412/jets-on-nano-v1>.
- [29] V. Palan, "Final project f3: Efficientnet experiments," Kaggle Notebook, 2024. [Online]. Available: <https://www.kaggle.com/code/vatsalpalan99/final-proj-f3>.
- [30] A. Ikram, A. I. Mudassir, and Abdullah, "Diabeticretinopathy classification with densenet201," Kaggle Notebook, 2024. [Online]. Available: <https://www.kaggle.com/code/liveinthemoment/notebook5f0e0b81de>.
- [31] D. Malhotra, "Qcnn-dr: Quantum convolutional neural network," Kaggle Notebook, 2024. [Online]. Available: <https://www.kaggle.com/code/dakshm/qcnn-dr>
- [32] Shi, Bojia, et al. "GoogLeNet-based Diabetic-retinopathy-detection." 2022 14th International Conference on Advanced Computational Intelligence (ICACI). IEEE, 2022.
- [33] Mohanty, Cheena, et al. "Using deep learning architectures for detection and classification of diabetic retinopathy." *Sensors* 23.12 (2023): 5726.
- [34] Gangwar, Akhilesh Kumar, and Vadlamani Ravi. "Diabetic retinopathy detection using transfer learning and deep learning." *Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020)*, Volume 1. Singapore: Springer Singapore, 2020. 679-689.
- [35] Y. Yang, A. Panagopoulou, S. Zhou, D. Jin, C. Callison-Burch, M. Yatskar, Language in a bottle: Language model guided concept bottlenecks for interpretable image classification, in, 2023, pp. 19187--19197.
- [36] M. Yuksekgonul, M. Wang, J. Zou, Post-hoc concept bottleneck models, *arXiv preprint arXiv:2205.15480*, (2022).
- [37] Y. Gao, D. Gu, M. Zhou, D. Metaxas, Aligning human knowledge with visual concepts towards explainable medical image classification, in, 2024, pp. 46-56.

Automatic Detection and Grading of Diabetic Retinopathy Using a Hierarchical Convolutional Network and Dual Attention Mechanism

Mustapha Mejman Salem Alghanami¹, Mansoor Fateh^{2*}

¹ PHD Candidate, Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran

² Associate Professor, Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran

Article Information

Original Research Paper

Received:

2026 February 10

Accepted:

2026 April 27

Keywords:

Diabetic Retinopathy, Deep Learning, ResNet, Attention Mechanism, Medical Image Analysis

Corresponding Author*:

mansoor_fateh@shahroodut.ac.ir

Abstract

Diabetic Retinopathy is one of the most common ocular complications of diabetes and a leading cause of blindness among adults. Early detection of this disease poses a significant challenge in medical image processing due to the minute size of early-stage lesions, such as microaneurysms, and their visual similarity to healthy retinal tissue. Conventional deep learning methods often face limitations in balancing the extraction of fine-grained textural details with the comprehension of global semantic information. In this paper, we propose a Hierarchical Convolutional Neural Network with a Dual Attention Mechanism for the detection and grading of Diabetic Retinopathy. The proposed architecture utilizes a pre-trained ResNet-50 backbone for feature extraction and manages information flow through two distinct pathways. The first pathway focuses on extracting local details by integrating mid-level features, while the second pathway processes deep semantic features. To enhance feature discriminability, both pathways are equipped with a Hierarchical Spatial and Channel Attention Module that adaptively highlights lesion-related regions and suppresses background noise. Furthermore, to address the class imbalance problem, a hybrid loss function combining Cross-Entropy and Focal Loss is employed. Evaluation on the standard APTOS 2019 dataset demonstrates that the proposed method achieves an accuracy of %97.23 in binary classification (Healthy vs. DR) and %83.06 in multi-class severity grading (5-class). Experimental results and ablation studies confirm the efficacy of the dual-branch architecture and attention modules compared to recent state-of-the-art methods.



: 10.22034/ABMIR.2026.24319.1221

E-ISSN: [2821-2037](#)

/The Author 2026. Published by Yazd University This is an open access article under the CC BY 4.0 License <https://creativecommons.org/licenses/by/4.0/>.

