



پرديس علوم  
دانشكده رياضي، آمار و علوم كامپيوتر

# بررسی شبکه عصبی Stitchable

نگارنده

ستاره ابهری

استاد راهنما

هدیه ساجدی

پروژه مقطع کارشناسی

رشته علوم کامپیوتر

مرداد ۱۴۰۲

## چکیده

شناسایی به موقع و دقیق بیماری های مرتبط با شبکیه و ماکولار، با توجه به آسیب های جبران ناپذیر ناشی از آنها، از اهمیت بسیاری برخوردار است. AMD یکی از این بیماری های چشمی پیشرفته است که باعث تخریب ناحیه مرکزی شبکه بینایی چشم می شود و می تواند به افت تدریجی دید و کوری منجر شود. شناسایی زودهنگام این بیماری می تواند بهبود قابل توجهی در درمان بیماران داشته باشد. یک روش مرسوم برای تشخیص AMD استفاده از تصاویر OCT (به عنوان یک فناوری عکس برداری غیر تهاجمی) است. در سال های اخیر با وجود پیشرفت چشمگیری که در زمینه پردازش تصویر رخ داده، مدل های جدید زیادی برای دسته بندی تصاویر OCT جهت شناسایی این بیماری ارائه شده اند که عملکرد قابل توجهی داشته اند.

با ارائه مدل VIT، در سال های اخیر مدل های مختلفی برای پردازش تصاویر پزشکی ارائه شده که نتایج این مدل ها به نسبت مدل های دیگر، یعنی شبکه های عصبی پیچشی و شبکه های عصبی عمیق سنتی، بسیار مناسب به نظر می رسد. یکی از مدل های بر پایه ی مدل که به صورت تخصصی برای دسته بندی تصاویر پزشکی ارائه شده است، مدل مبدل بینایی پزشکی است که توانست نسبت به رقبای خود در حوزه های مختلف تصاویر دوبعدی نتایج بهتری کسب کند. در این پروژه روشی بر پایه ی شبکه های عصبی Stitchable و مدل های مبدل بینایی پزشکی برای رسیدن به ساختاری مناسب در مصالحه دقت و کارایی، در طبقه بندی تصاویر OCT شبکیه ی چشم جهت تشخیص AMD ارائه و نشان خواهیم داد که ساختار بدست آمده روی مجموعه دادگان NEH قادر است بدون استفاده از مجموعه دادگان تخصصی دیگری به دقت بهتری نسبت به مدل های ارائه شده قبلی دست یابد.

## پیشگفتار

توسعه سریع علوم کامپیوتر در چند دهه گذشته به پیشرفت بی‌سابقه‌ای در حوزه هوش مصنوعی منجر شده است و کاربرد بسیار گسترده‌ای در چشم پزشکی به ویژه در پردازش تصویر و تجزیه و تحلیل داده داشته. همچنین در سال‌های اخیر، هوش مصنوعی به طرز چشمگیری در بیماری‌های بینایی سنجی به کار گرفته شده و نتایج چشمگیری به دنبال داشته است.

چشمگیرترین پیشرفت‌ها در بیماری مرتبط با سندرم ماکولار مرتبط با سن (AMD) شکل گرفته، که شامل بخش‌های مختلفی از جمله غربالگری، تشخیص، پیش‌بینی و مانیتورینگ می‌باشد. ولی عنصر اصلی در مدیریت AMD همچنان تشخیص مدل‌های پیش‌بینی‌کننده است که اجازه تشخیص سریع بیماران در معرض خطر در مقیاس بزرگ را می‌دهد. شناسایی زودهنگام فنوتیپ‌های AMD با خطر بالا به خصوص مفید است تا جریان خارج شدن مایعات در آینده پیش‌بینی شود که می‌تواند از روش مدیریت به موقع با عوامل ضد رشد عروقی واسکولار (anti-VEGF) بهبود بیشتری داشته باشد و به نتایج بالینی بهتری منجر شود. با وجود اینکه در حال حاضر درمانی برای GA وجود ندارد، درمان‌های امیدوارکننده در افق هستند که نیاز به تغییر دادن مسیر بیماری و پیامدهای پیش‌بینی شده دارند که در آزمایشات بالینی آینده باید استفاده شوند.

با وجود ادبیات گسترده‌ای که در زمینه غربالگری و تشخیص با عملکرد بالا در تشخیص AMD در هر مرحله وجود دارد، یکی از چالش‌های اصلی، ارزیابی خطر تبدیل و پیشرفت بیماری است. بنابراین، نیاز به گسترش دانش در مورد پیش‌بینی‌ها و مدل‌هایی که به پیش‌بینی فنوتیپ‌های AMD در مراحل پیشرفته‌تر کمک می‌کنند، احساس می‌شود. در حوزه نشانگرهای تصویربرداری، به نظر می‌رسد که استفاده از ویژگی‌های مبتنی بر تصویربرداری با استفاده از توموگرافی نوری جمع‌آوری شده (OCT)، چندین مزیت فنی و عملی دارد. علاوه بر به کارگیری آن در مقیاس بزرگ در آزمایشات بالینی روتین برای تشخیص و مدیریت AMD با وجود داشتن وضوح بالای فناوری‌های دامنه طیفی و دیگر تکنولوژی‌ها، فرصت شناسایی تغییرات پیشینی و مراحل زودرس بیماری قبل از ظاهر شدن اجتناب ناپذیر ضایعه‌ها را فراهم می‌کند. بنابراین، اخیراً، طبق تعریف دقیقی که توسط یک گروه توافق‌نامه‌ای از متخصصان برجسته انجام شده است، طبقه‌بندی تعریف شده بر اساس OCT از ویژگی‌های آتروفی و نئوواسکولار در محیط AMD برای این منظور تعریف شده است. همچنین چندین ویژگی کیفی OCT شناسایی شده است که ارزیابی پیشرفت بیماری و پاسخ به درمان را ممکن می‌کند که بر اهمیت شناسایی نشانگرهای OCT به رسیدن به مدیریت بهینه AMD می‌افزاید.

پردازش تصویر در حوزه دسته بندی تصاویر OCT به عنوان یکی از مهم‌ترین روش‌های تشخیص و مدیریت بیماری‌های شبکیه استفاده می‌شود. استفاده از روش‌های پردازش تصویر و هوش مصنوعی برای دسته بندی تصاویر OCT در بیماری‌های شبکیه، بهبود قابل توجهی در تشخیص و مدیریت بیماری‌های شبکیه ایجاد کرده است.

یکی از روش‌های مهم در پردازش تصویر OCT ، استفاده از شبکه‌های عصبی پیچشی (CNN) است. این شبکه‌ها با استفاده از الگوریتم‌های خاص، توانایی تشخیص الگوهای پیچیده را دارند و بهبود قابل توجهی در دقت تشخیص و پیش‌بینی بیماری‌های شبکیه در تصاویر OCT ایجاد کرده‌اند. اما با توجه به رشر روش‌های مبتنی بر ساختار مبدل در چند سال اخیر و پیشی گرفتن این روش‌ها نسبت به شبکه‌های عصبی پیچشی ، نیاز به مطالعات بیشتر در زمینه‌ی OCT با استفاده از ساختارهای مبدل احساس می‌شود.

هدف از این پروژه بررسی عملکرد شبکه عصبی (SNN) Stitchable بر روی دسته‌ای از تصاویر OCT می‌باشد. از آنجایی که مدل ارائه شده جدید مبدل بینایی پزشکی عملکرد چشمگیری بر روی تصاویر پزشکی داشته، در این پروژه نیز مدل‌های آموزش دیده مبدل بینایی پزشکی را به یکدیگر بخیه زده و عملکرد آن را آزمایش می‌کنیم. در این گزارش علاوه بر توضیح کار کرد هر کدام از مدل‌های SNN و مبدل بینایی پزشکی ، مفاهیم اولیه مورد نیاز برای درک بهتر آن‌ها و همچنین روش بخیه زدن آن‌ها را توضیح خواهیم داد.

در فصل ۲ مروری بر مفاهیم اولیه مورد نیاز برای کار کردن با این مدل‌ها را خواهیم داشت و همچنین مزایای استفاده از آن‌ها را بیان می‌کنیم. در فصل ۳ به بررسی مجموعه دادگان NEH که از متشکل از تصاویر OCT می‌باشد می‌پردازیم و همچنین به کارهای پیشین بر روی این نوع تصاویر و نتایج آن‌ها نیز اشاره می‌کنیم. در فصل ۴ به روش SNN و مبدل بینایی پزشکی و روش بخیه زدن مبدل بینایی پزشکی می‌پردازیم. در فصل ۵ نتایج آزمایشات انجام شده بر روی این مدل را بیان می‌کنیم و در نهایت جمع‌بندی و نتیجه‌گیری از کارهای انجام شده را ارائه می‌دهیم.

## فهرست مطالب

۱	مقدمه	۱
۴	مفاهیم مقدماتی	۲
۴	۱.۲ مکانیزم توجه	۱.۲
۵	۲.۲ مبدل‌ها	۲.۲
۵	۱.۲.۲ ساختار کلی و مزایای مبدل‌ها	۱.۲.۲
۷	۲.۲.۲ مبدل‌های بینایی	۲.۲.۲
۹	۳.۲ مبدل‌های دوگانه	۳.۲
۹	۴.۲ تقطیر دانش	۴.۲
۱۱	۵.۲ مخلوط کردن داده‌ها	۵.۲
۱۲	۶.۲ برش مخلوط	۶.۲
۱۳	مجموعه دادگان	۳
۱۴	۱.۳ کارهای مرتبط	۱.۳
۱۷	روش پیشنهادی	۴
۱۷	۱.۴ مروری بر مبدل بینایی پزشکی	۱.۴
۱۷	۱.۱.۴ بلوک پیچشی بهینه	۱.۱.۴
۱۷	۲.۱.۴ بلوک مبدل محلی	۲.۱.۴
۱۹	۳.۱.۴ مبدل بینایی پزشکی میکرو	۳.۱.۴
۱۹	۴.۱.۴ مبدل بینایی پزشکی بسیار کوچک و میکرو همراه با نویز	۴.۱.۴
۱۹	۲.۴ مروری بر شبکه عصبی بخیه‌ای	۲.۴
۲۱	۱.۲.۴ بخیه زدن مبدل بینایی پزشکی	۱.۲.۴
۲۲	آزمایشات و نتایج	۵
۲۲	۱.۵ آزمایشات	۱.۵
۲۳	۲.۵ نتایج	۲.۵
۲۶	جمع بندی و نتیجه گیری	۶

# فصل ۱

## مقدمه

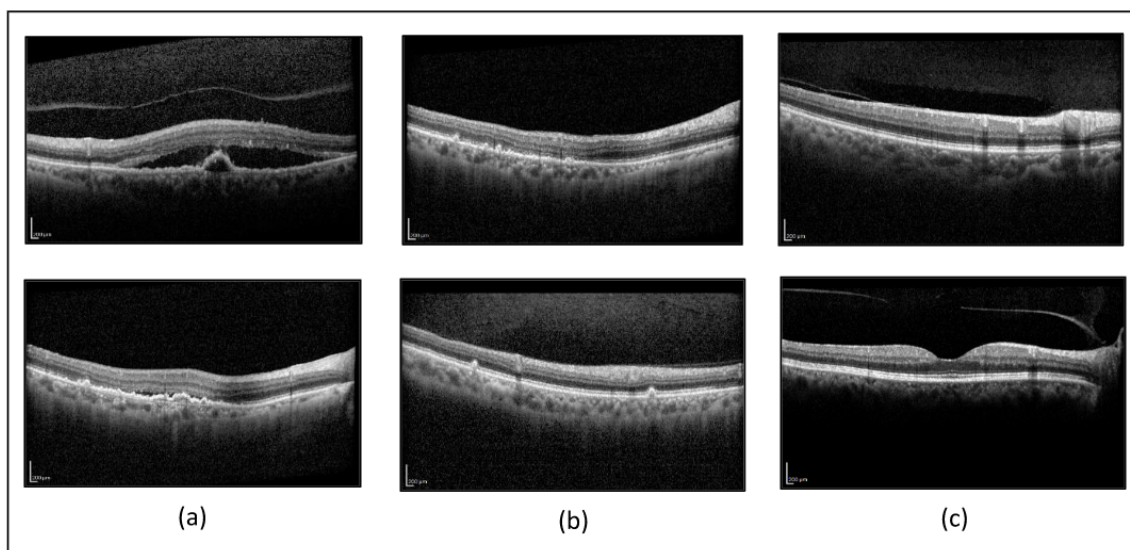
استحاله ماکولار وابسته به سن AMD از بیماری‌های شبکیه است که در افراد بالای ۵۵ سال منجر به نقص بینایی می‌شود. تخمین زده شده است که در سال ۲۰۲۰، این بیماری حدود ۱۹۶ میلیون نفر از جمعیت کره زمین را تحت تأثیر قرار داده و این تعداد، با توجه به روند افزایش جمعیت سالخورده، تا سال ۲۰۴۰ به ۲۸۸ میلیون نفر خواهد رسید [1]. مراحل اولیه بیماری AMD با ظاهر شدن دروزن و رنگ‌رفتگی سلول‌های اپی تلیوم رنگدانه شبکیه (RPE) قابل تشخیص است. با افزایش دروزن، AMD وارد مراحل پیشرفته‌تر می‌شود.

مراحل پیشرفته AMD در دو دسته خشک و تر جای می‌گیرد. در AMD خشک بر اثر از بین رفتن تدریجی سلول‌های RPE، شبکه مویرگی مشیمیه و گیرنده‌های نوری، حساسیت بینایی کاهش می‌یابد. در نوع تر AMD، با ظهور عروق جدید مشیمیه‌ای اختلالات جدی در بینایی رخ می‌دهد. یکی از روش‌های مورد استفاده جهت تشخیص اولیه AMD و پایش افراد در این خصوص، استفاده از تصاویر مقطع‌نگاری همدوس اپتیکی (OCT) است [2]. بر همین اساس در سال‌های گذشته، روش‌های متنوعی برای طبقه‌بندی این تصاویر جهت تشخیص اولیه سریع‌تر AMD ارائه شده است. در شکل ۱.۱ نمونه‌ای از تصاویر OCT با برچسب‌های Normal، Drusen و CMT نشان داده شده است.

هوش مصنوعی در حوزه تصویربرداری، بخصوص در پزشکی چشم‌پزشکی، بسیار موثر و کارآمد است و می‌تواند در تشخیص بیماری‌هایی با شیوع بالا مانند گلوکوم، شبکیه‌پیشانی، رتینوپاتی نوزادان و بیماری ماکولار مرتبط با سن (AMD)، به مراتب بهتر از روش‌های معمولی عمل کند.

به دلیل افزایش تعداد مبتلایان به این بیماری‌ها و به دنبال آن حجم عظیمی از داده‌های پزشکی، نیازی فوری در حوزه ادغام یادگیری ماشین و یادگیری عمیق هوش مصنوعی در محیط‌های چشم‌پزشکی برای تشخیص زودهنگام، پیش‌بینی و درمان به موقع بیماری‌های چشمی خطرناک وجود دارد. در این چارچوب، فناوری هوش مصنوعی توانایی وعده‌دهنده‌ای دارد که خدمات سلامت چشمی را تحول بخشیده و اثر بالینی قابل توجهی داشته باشد و بار پزشکی را کمتر کند.

یادگیری ماشین زیرمجموعه‌ای از فناوری هوش مصنوعی است که در دهه ۱۹۸۰ معرفی شد و شامل یادگیری عمیق و یادگیری ماشین معمولی است. یادگیری ماشین به سیستم کامپیوتری اجازه می‌دهد تا به طور خودکار یک کار را بدون برنامه‌ریزی صریح یاد بگیرد و بهبود یابد. در الگوریتم‌های یادگیری ماشین استفاده شده در هوش مصنوعی، شاخص‌های تصادفی (RF) و ماشین‌های بردار پشتیبانی (SVM) در حوزه چشم‌پزشکی بیشترین کاربرد را دارند.



شکل ۱.۱: نمونه‌ای از تصاویر OCT شبکیه با پرچسب‌های Normal (a)، Drusen (b) و CNV (c):

یادگیری عمیق به عنوان زیرمجموعه‌ای از فناوری هوش مصنوعی در دهه ۲۰۰۰ معرفی شد که با استفاده از ساختار شبکه عصبی مصنوعی از ساختار و عملکرد مغز انسان، ویژگی‌های داده‌ها را یاد می‌گیرد. یادگیری عمیق شامل ورودی‌های محرک چندگانه به لایه‌های مخفی عصبی است که هر کدام می‌توانند ویژگی‌های مختلفی را از محرک‌های ارائه شده یاد بگیرند که امکان انجام کارهای پیچیده را به مدل می‌دهد.

در میان روش‌های مختلف DL، شبکه عصبی پیچشی عمیق به عنوان یکی از مناسب‌ترین روش‌ها برای تشخیص تصویر در پزشکی شناخته شده است. شبکه عصبی پیچشی عمیق از شبکه‌های عصبی مصنوعی سنتی تکامل یافته‌اند و با استفاده از یک الگوی عصبی سه‌بعدی و اعمال یک عملیات فیلترینگ ریاضی ویژه به نام پیچش از ویژگی‌های داده‌ها یاد می‌گیرند. آموزش کامل شبکه عصبی پیچشی عمیق شامل حجم عظیمی از داده‌های آموزشی که قبلاً توسط کارشناسان پزشکی برچسب‌گذاری شده‌اند، منابع محاسباتی و حافظه گسترده و مشکلات همگرایی و بیش‌برازش است که نیاز به تنظیمات تکراری و زمان‌بر دارد. یک جایگزین معتبر برای آموزش شبکه عصبی پیچشی از ابتدا، شامل تنظیم دقیق یک شبکه عصبی پیچشی با استفاده از یادگیری انتقالی است که شامل انتقال ویژگی‌های یادگرفته شده از یک شبکه عصبی پیچشی پیش‌آموزش‌شده برای شروع یک وظیفه جدید است. مدل‌های یادگیری انتقالی مبتنی بر شبکه عصبی پیچشی، در طبقه‌بندی ویژگی‌های مبتنی بر OCT در AMD عملکرد خوبی را نشان داده‌اند. روش‌های یادگیری عمیق مختلفی برای تقسیم بندی تصویر شبکه عصبی شناخته شده‌اند، مانند FCNN، U-Net، Seg-net، Deeplabv3 و AlexNet.

در سال‌های اخیر پس از معرفی مدل‌های بنیایی [3]، نسخه‌های مختلف آن توانستند در بسیاری از مسئله‌های بنیایی کامپیوتر عملکرد بهتری نسبت به شبکه عصبی پیچشی‌ها روی مجموعه دادگان و دامنه‌های مختلف نشان دهند. علیرغم این موضوع، مطالعات کافی برای به کارگیری مدل‌ها روی تصاویر OCT برای تشخیص AMD انجام نگرفته

است. در این پروژه سعی خواهیم کرد مسأله طبقه‌بندی مجموعه دادگان NEH را که یک مجموعه دادگان کوچک محسوب می‌شود، به کمک بخیه زدن مدل‌های مبدل بینایی پزشکی حل کنیم. مبدل بینایی پزشکی یک مدل با ساختار دوگانه متشکل از ماژول‌های توجه و پیچشی است که توسط Manzari et al. برای طبقه‌بندی قدرتمند تصاویر پزشکی طراحی و نشان داده شده که در مجموعه دادگان مورد تست نسبت به سایر شبکه‌های پیچشی و مبدل‌های شناخته شده، دقت بهتری را ارائه می‌دهد [4].

چالش اصلی استفاده از مبدل بینایی پزشکی روی مجموعه دادگان NEH، کوچک بودن این مجموعه دادگان است که منجر به بیش‌برازش شدن مدل می‌شود. برای حل این مشکل مدل کوچک‌تری از مبدل بینایی پزشکی را معرفی کرده و هم‌چنین برای بهبود یادگیری مدل‌های اصلی مبدل بینایی پزشکی از اضافه کردن نویز به مجموعه دادگان آموزشی بهره خواهیم جست. در ادامه برای بهبود نتایج، دو مدل مختلف از مبدل بینایی پزشکی‌های آموزش داده شده را به یکدیگر بخیه خواهیم زد. بخیه‌زدن شبکه‌های عصبی به عنوان یک روش جست و جوی محدود، برای رسیدن به مصالحه مناسبی از اندازه مدل نهایی و دقت آن بر اساس مدل‌های ازپیش‌آموزش داده شده است که توسط Pan et al. در [5] مطرح شده است. در این مرحله نیز چالش اصلی برای بخیه زدن مدل‌های مبدل بینایی پزشکی ناهمگون بودن لایه‌های موجود در هر طبقه و تعداد لایه‌های بالا در طبقه سوم است که تعداد مدل‌های کاندید و مطرح در فضای جست‌و‌جو را بالا می‌برد.

کارهای انجام شده در این مطالعه را می‌توان به صورت زیر خلاصه کرد:

- معرفی و اضافه کردن مدل میکرو به خانواده مدل‌های مبدل بینایی پزشکی کوچک، بسیار کوچک و بزرگ
- استفاده از نویز گاوسی جمع شونده به داده‌های ورودی در زمان آموزش مدل به عنوان روشی برای بهبود نتایج مدل‌های بزرگ‌تر مبدل بینایی پزشکی
- بخیه‌زدن مدل‌های مبدل بینایی پزشکی



## فصل ۲

# مفاهیم مقدماتی

### ۱.۲ مکانیزم توجه

انسان‌ها یک مهارت شناختی پیچیده به نام مکانیزم توجه دارند. وقتی انسان‌ها اطلاعاتی را دریافت می‌کنند، ممکن است انتخاب کنند که به قسمت از اطلاعات که به نظر اصلی و مهم است توجه نکنند و در مقابل دقت بیشتری بر روی اطلاعاتی که به نظر می‌آید از اهمیت کمتری برخوردار است، داشته باشند. توجه (Attention) نامی است که برای این مهارت در نظر گرفته شده است.

استفاده از این تکنیک در شبکه‌های عصبی به شبکه قابلیت توجه به بخش‌های مختلف و مهم دنباله‌ی ورودی را می‌دهد. در واقع، مکانیزم توجه شبکه عصبی کمک می‌کند تا با تمرکز بر بخش‌های مهم و حیاتی دنباله‌ی ورودی، عملکرد بهتری داشته باشد. به جای این که تمام داده‌های دنباله به صورت یکجا به شبکه داده شود، در مکانیزم توجه، هر بخش از دنباله با توجه به اهمیت آن بخش به شبکه معرفی می‌شود. این به این معنی است که شبکه با توجه به اهمیت هر بخش، یک وزن مخصوص به آن اختصاص می‌دهد و به آن بخش توجه بیشتری می‌کند و از آن برای تصمیم‌گیری‌های بعدی استفاده می‌کند.

برای توضیح بهتر، به نمونه‌ای از استفاده از مکانیزم توجه در پردازش تصویر می‌پردازیم: فرض کنید که شما برای تشخیص اشیاء در یک تصویر، یک شبکه عصبی پردازش تصویر (مانند شبکه‌های عصبی پیچشی) آموزش می‌دهید. در این حالت، شبکه ممکن است به طور یکنواخت تمام بخش‌های تصویر را بررسی کند و اطلاعات را از طریق لایه‌های خود به دست‌آورد.

اما با اضافه کردن مکانیزم توجه، شبکه می‌تواند به تصویر با تمرکز بیشتری نگاه کند و به بخش‌های مهم‌تر توجه کند. به عبارت دیگر، مکانیزم توجه به شبکه اجازه می‌دهد تا بخش‌هایی از تصویر که احتمالاً شامل اشیاء هستند یا اطلاعات مهم‌تری دارند، را با دقت بیشتر بررسی کند.

نحوه عملکرد مکانیزم توجه در شبکه‌های پردازش تصویر به این صورت است:

۱. استخراج ویژگی‌ها:

شبکه ابتدا ویژگی‌های اولیه تصویر را استخراج می‌کند

۲. محاسبه امتیاز توجه:

برای هر قسمت از تصویر (معمولاً پیکسل‌ها یا بخش‌های کوچکتر)، یک امتیاز توجه محاسبه می‌شود. این امتیازها نشان‌دهنده اهمیت هر قسمت از تصویر برای وظیفه مدل است.

۳. محاسبه وزن‌های توجه:

امتیازهای توجه به وزن‌های توجه تبدیل می‌شوند. این وزن‌ها نشان‌دهنده میزان تأثیر هر بخش از تصویر بر روی تصمیمات شبکه است.

۴. ادغام توجه:

با استفاده از وزن‌های توجه، ویژگی‌های تصویر به طور وزن‌دار ترکیب می‌شوند. این به شبکه امکان می‌دهد تا تمرکز بیشتری بر روی ویژگی‌های مهم داشته باشد.

۵. استفاده از ترکیب توجه:

ویژگی‌های ترکیب شده توسط مکانیزم توجه به لایه‌های بعدی شبکه داده می‌شوند تا بهترین تصمیمات برای تشخیص اشیاء یا انجام وظیفه مشابه دیگر انجام شود.

در نهایت، با استفاده از مکانیزم توجه، شبکه عصبی قادر به تمرکز بر روی بخش‌های مهم‌تر تصاویر و اطلاعات کلیدی است و این امر می‌تواند عملکرد و دقت شبکه را در وظایف پردازش تصویر بهبود بخشد. مزایای استفاده از مکانیزم توجه در شبکه‌های عصبی شامل بهبود دقت و کارایی شبکه در پردازش داده‌های دنباله‌ای، کاهش نیاز به پارامترهای شبکه و امکان تشخیص و فراگیری روابط بین داده‌های ورودی است. به علاوه، مکانیزم توجه در مواردی که داده‌های ورودی به صورت دنباله‌ای هستند و بخش‌های مختلف دنباله با اهمیت‌های متفاوتی همراه هستند، بسیار موثر و کارآمد است.

## ۲.۲ مبدل‌ها

شبکه‌های مبدل (Transformer) یک نوع معماری از شبکه‌های عصبی هستند که برای پردازش داده‌های توالی‌ای مانند جملات زبانی، متون، تصویر، ویدئو و غیره بسیار موثر هستند. این معماری برای اولین بار در سال ۲۰۱۷ توسط Vaswani و همکاران معرفی شد و باعث توجه بسیاری در زمینه پردازش زبان طبیعی و ترجمه ماشینی شده است.

### ۱.۲.۲ ساختار کلی و مزایای مبدل‌ها

کارایی شبکه‌های مبدل نسبت به سایر معماری‌های پردازش توالی مانند RNN و LSTM بسیار بالاست و به دلیل استفاده از مکانیزم توجه به خود و لایه‌های کاملاً متصل از اهمیت ویژه‌ای برخوردار است. حال به طور جزئی‌تر هر کدام از اجزای اصلی شبکه‌های مبدل را توضیح می‌دهیم:

□ توجه به خود چندسری:

لایه توجه به خود چندسری یکی از اجزای کلیدی شبکه‌های مبدل است که اجازه می‌دهد تا شبکه به صورت هوشمندانه توجه به تمام کلمات یا بخش‌های توالی داشته باشد و ارتباطات بین آن‌ها را بررسی کند. این لایه دارای سه مرحله اصلی است:

۱. محاسبه امتیاز توجه:

ابتدا برای هر کلمه یا بخش از توالی ورودی، یک بردار امتیاز توجه محاسبه می‌شود. این بردار امتیاز توجه مشخص می‌کند که هر کلمه به کدام کلمات دیگر توجه بیشتری دارد.

۲. وزن دهی توجه:

سپس، با استفاده از بردار امتیاز توجه، بردار ویژگی‌های ورودی با وزن‌دهی از کلمات دیگر ترکیب می‌شود. این به این معناست که هر کلمه به توالی‌های دیگر با توجه به اهمیت وزن‌دهی، به طور متفاوت توجه می‌کند.

۳. توجه چندسرها:

به منظور افزایش توانایی مدل در تعامل با اطلاعات مختلف، لایه توجه به خود چندسرها معمولاً از چندین سر یا سر برای محاسبه امتیازهای توجه مختلف استفاده می‌کند. هر سر برای محاسبه بردار امتیاز توجه خود را دارد و نتایج آن‌ها به صورت موازی ترکیب می‌شوند.

#### □ شبکه عصبی پیشخور

لایه کاملاً متصل یا شبکه عصبی پیشخور یک لایه عصبی ساده است که پس از لایه توجه به خود در هر بلوک ترانسفورمر اجرا می‌شود. این لایه دارای دو لایه کاملاً متصل است:

۱. لایه اول (لایه خطی):

این لایه بردار ویژگی‌های خروجی از لایه توجه به خود را به فضای ویژگی‌های جدید تبدیل می‌کند. این تبدیل از فضای ویژگی‌های اصلی به فضای ویژگی‌های با ابعاد بزرگتر و با خصوصیات غیرخطی است.

۲. لایه دوم (لایه فعال سازی):

این لایه از یک تابع غیرخطی (مانند ReLU یا GELU) استفاده می‌کند تا خروجی نهایی را تولید کند. این عمل تنظیم‌های غیرخطی به مدل امکان می‌دهد تا قابلیت‌های پیچیده‌تری را از داده‌ها استخراج کند.

استفاده از لایه کاملاً متصل به عنوان لایه پس از توجه به خود، به مدل اجازه می‌دهد تا اطلاعات اضافی و پردازش‌های غیرخطی را به ویژگی‌های استخراج شده از مکانیزم توجه به خود اضافه کند، که باعث بهبود عملکرد مدل و بهترین استفاده از اطلاعات توالی‌ها می‌شود.

این ترکیب موثر از مکانیزم توجه به خود و لایه کاملاً متصل، موجب افزایش کارایی ترانسفورمر و توانایی آن در پردازش داده‌های توالی‌ای با طول‌های مختلف و انجام مسائل مختلف پردازش زبان طبیعی و ترجمه ماشینی می‌شود. معماری ترانسفورمر دارای مزایای زیر است:

- مقیاس‌پذیری: به علت استفاده از توجه به خود، این شبکه‌ها به طور موثری مقیاس‌پذیر هستند و توانایی کار با توالی‌های بلند را دارند.
- بهبود کار با توالی‌های طولانی: در معماری‌های قبلی مانند RNN، مشکل کار با توالی‌های بلند و کاهش گرادینت ممکن بود، اما با استفاده از توجه به خود، این مشکل رفع شده است.
- اطلاعات مکانی: استفاده از مکانیزم انتقال به شبکه این امکان را می‌دهد تا اطلاعات مکانی توالی‌ها را در نظر بگیرد و مکان را در توالی‌ها مشخص کند.
- کارایی: به دلیل استفاده از مکانیزم توجه به خود به جای لایه‌های RNN، محاسبات سریع‌تر انجام می‌شود و معماری کارایی بیشتری دارد.

## ۲.۲.۲ مبدل‌های بینایی

مبدل بینایی (Vision Transformers) یک مدل عمیق از خانواده مدل‌های تبدیل متنی است که به صورت اصلی برای مسائل بینایی ماشین مورد استفاده قرار می‌گیرد. این مدل‌ها در تحلیل و فهم تصاویر استفاده می‌شوند و با استفاده از تبدیل موجود در معماری آنها، توجه به نقاط مختلف تصاویر را انجام می‌دهند. مدل‌های مبدل اصلی برای پردازش متن به صورت یک دنباله از کلمات طراحی شده‌اند. اما برای استفاده آنها در مسائل بینایی، لازم است تصاویر را به شکل دنباله‌ای از نقاط مختلف (پیکسل‌ها یا باکس‌های کوچک‌تر) نمایش داد. این ایده با نام "پیکسل به دنباله" شناخته می‌شود و هدف اصلی این رویکرد این است که از قدرت مدل‌های مبدل در پردازش دنباله‌ها در مسائل بینایی نیز بهره‌برداری کند. معماری مبدل‌های بینایی از دو بخش اصلی تشکیل می‌شود:

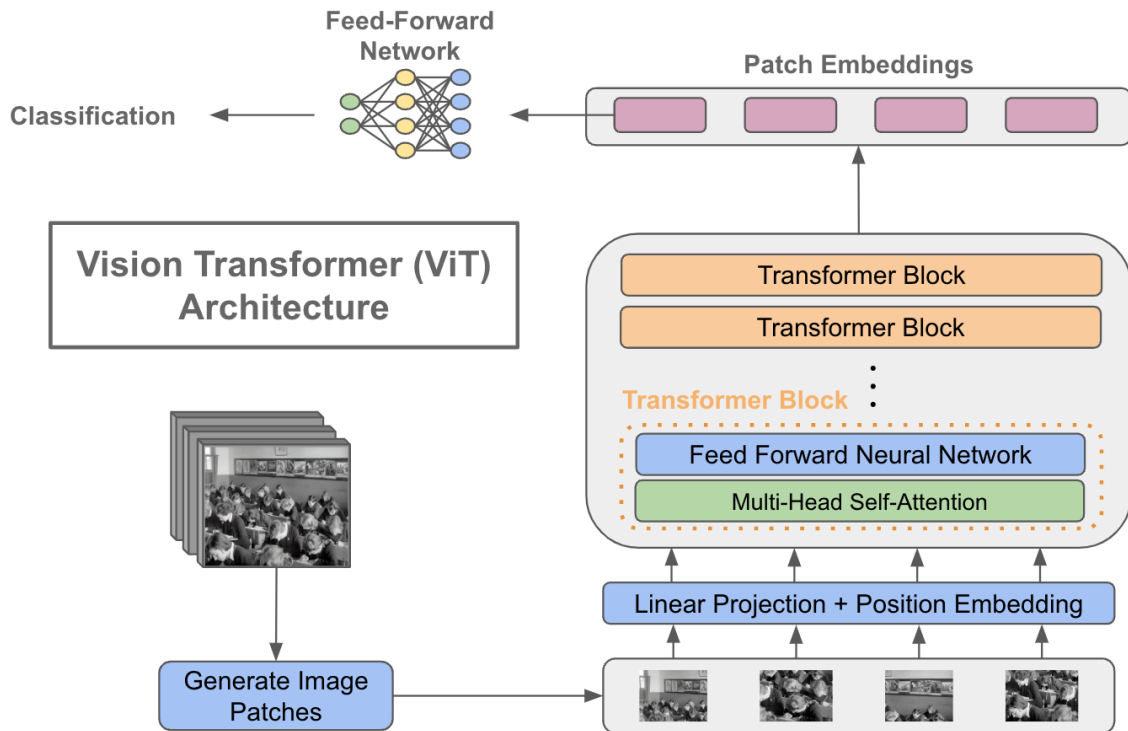
۱. بخش تعبیر: ابتدا تصویر ورودی با استفاده از یک لایه تعبیر به دنباله‌ای از بردارها تبدیل می‌شود. در این لایه، هر پیکسل یا باکس تصویر با یک بردار به طول ثابت نمایش داده می‌شود. این بردارها معمولاً دارای اطلاعاتی از مکان‌های مختلف تصویر هستند.

۲. بخش مبدل:

سپس از معماری مبدل برای پردازش دنباله‌ای از بردارها استفاده می‌شود. معمولاً از یک تعداد مشخصی لایه مبدل استفاده می‌شود تا اطلاعات مربوط به تصویر در جریان پردازش حفظ شود و همچنین وابستگی‌ها و روابط بین نقاط مختلف تصویر استخراج شود. در این مرحله، مدل می‌تواند به طور مستقل به نقاط مختلف تصویر توجه کند و از اطلاعات بینایی بهره‌برداری کند.

مدل مبدل بینایی نتایج قابل توجهی نسبت به شبکه‌های عصبی پیچشی به دست می‌آورد در حالی که در مقایسه با شبکه‌های عصبی پیچشی، منابع محاسباتی کمتری برای پیش آموزش استفاده می‌کند. مدل مبدل بینایی (ViT) در کل دارای یک باور ناقص است که باعث افزایش وابستگی به تنظیمات مدل یا افزایش داده‌ها (AugReg) در هنگام آموزش در مجموعه داده‌های کوچکتر می‌شود.

همچنین مبدل‌های بینایی بایاس استقرایی خاص تصویر بسیار کمتری نسبت به شبکه عصبی پیچشی دارند. در شبکه عصبی پیچشی خاصیت محلی، ساختار همسایگی دو بعدی و هم‌اندازی ترجمه‌ای (تغییر نسبت به ترجمه) در هر لایه در سرتاسر مدل وارد می‌شود. اما در مبدل‌های بینایی، تنها لایه‌های MLP (پرسپترون چندلایه‌ای) به صورت محلی و هم‌اندازی ترجمه‌ای هستند، در حالی که لایه‌های خود توجه در سطح جهانی هستند. همچنین در این مدل ساختار همسایگی دو بعدی بسیار اندکی استفاده می‌شود: در ابتدای مدل با برش تصویر به قطعات و همچنین در زمان تنظیم مجدد دقیق برای تطبیق بردارهای موقعیت برای تصاویر با رزولوشن‌های مختلف. به جز این موارد، در زمان مقدماتی اولیه بردارهای موقعیت هیچ اطلاعاتی درباره موقعیت‌های دو بعدی قطعات ندارند و همه روابط فضایی بین قطعات باید از ابتدا یاد گرفته شوند. این قضیه باعث می‌شود زمانی که روی مجموعه دادگان‌های کوچکتر آموزش می‌بینند، در منظم سازی مدل یا افزایش داده‌ها مورد اعتماد تر باشند.



شکل ۱.۲: معماری یک مدل دسته‌بندی با استفاده از مبدل بینایی

مراحل کلی یک مبدل بینایی که در شکل ۱.۲ نشان داده شده را می‌توان به ترتیب زیر بیان کرد:

۱. تقسیم تصویر به قسمت‌های کوچک (patch) با اندازه‌های ثابت
۲. تخت کردن (flatten) این قطعات تصویر
۳. ایجاد تعبیرهای خطی (linear embeddings) با ابعاد کمتر از این قطعات تصویر

۴. افزودن تعبیرهای موقعیتی (positional embedding)

۵. تغذیه توالی به عنوان ورودی به کدگذار ترانسفورمر

۶. پیش‌آموزش مدل ViT با برچسب‌های تصویر، که سپس در مجموعه دادگان کاملاً نظارت می‌شود

۷. تنظیم مجدد جریان مجموعه دادگان برای طبقه‌بندی تصویر

## ۳.۲ مبدل‌های دوگانه

مبدل دوگانه (Hybrid Transformer)، معماری‌ای است که ترکیبی از لایه‌های مبدل به همراه لایه‌های پیچشی را به کار می‌گیرد. این ترکیب معماری می‌تواند از قابلیت‌های برتر هر دو ساختار بهره‌برداري کند و در برخی مسائل پردازش تصویر که هم‌زمان نیاز به کاربرد مفاهیم مبتنی بر موقعیت (از طریق شبکه پیچشی) و ارتباطات دقیق بین اجزاء تصویر (از طریق مبدل) دارند، عملکرد بهتری ارائه می‌دهد.

مبدل‌ها به ویژه با معماری مبدل دوگانه، به تحقیق و توسعه روش‌های پیشرو در بینایی ماشین، شناسایی اشیاء، دسته‌بندی تصاویر، تولید تصاویر و بسیاری دیگر از وظایف پردازش تصویر کمک کرده‌اند. برخی از اجزای کلیدی معماری مبدل‌های دوگانه عبارتند از:

□ لایه‌های پیچشی: این لایه‌ها برای افزایش استفاده از اطلاعات مبتنی بر موقعیت در تصاویر استفاده می‌شوند. از طرف دیگر، لایه‌های پیچشی می‌توانند الگوهای مشخصی از ویژگی‌ها را استخراج کنند که در کنار مبدل‌ها می‌توانند از اطلاعات معنایی برای تصاویر بهره ببرند.

□ لایه‌های مبدل: این لایه‌ها برای مدل‌کردن ارتباطات بین اجزاء تصویر به کار می‌روند. از طریق مبدل‌ها، مدل می‌تواند اطلاعات دقیق‌تری از ارتباطات بین اجزاء را درک کند و بهینه‌تر ویژگی‌های مورد نیاز را استخراج کند.

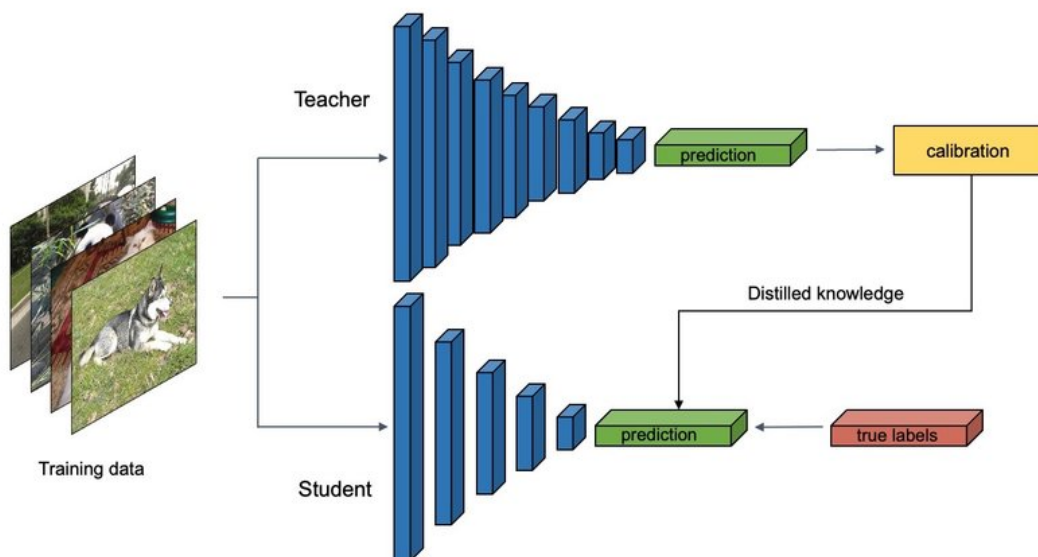
□ ادغام چند مقیاسه: معماری مبدل دوگانه معمولاً از ادغام چند مقیاسه برای بهبود نتایج استفاده می‌کند. این ادغام باعث می‌شود که مدل اطلاعات مربوط به اندازه‌ها و موقعیت‌های مختلف تصویر را به خوبی درک کند و نتایج دقیق‌تری ارائه دهد.

از زمان ارائه معماری مبدل دوگانه، تحقیقات بسیاری در زمینه پردازش تصویر با استفاده از این مدل‌ها انجام شده است و نتایج موفقیت‌آمیزی در وظایف مختلف به دست آمده است. هرچند همچنان مسائل و چالش‌های بسیاری برای ترکیب این دو ساختار وجود دارند و تحقیقات بیشتری برای بهبود معماری مبدل دوگانه در زمینه پردازش تصویر در حال انجام است.

مدل LeViT یک نمونه از این مبدل‌ها می‌باشد که از لایه پیچشی برای جایگزین کردن بلوک تعبیر تکه‌ای استفاده می‌کند که باعث تولید یک رابط دسته‌بندی با سرعت بالا می‌شود.

## ۴.۲ تقطیر دانش

تقطیر دانش (KD) یک تکنیک مهم در حوزه یادگیری عمیق است که برای انتقال دانش یک مدل پیچیده‌تر به یک مدل ساده‌تر مورد استفاده قرار می‌گیرد. هدف اصلی این روش، افزایش کارایی و سرعت مدل‌ها با کاهش پیچیدگی آن‌هاست.



شکل ۲.۲: ساختار کلی تقطیر دانش

این تکنیک به ویژه در مواردی که مدل‌ها پیچیده می‌باشد و امکان استفاده از منابع مورد نیاز را نداریم، مفید است. فرآیند KD به این صورت کار می‌کند که یک مدل قوی‌تر (معمولاً با اندازه بزرگتر و پیچیده‌تر) به نام "مدل معلم" یا "مدل پیش‌دستی" و یک مدل کوچک‌تر و ساده‌تر به نام "مدل دانش‌آموز" تعریف می‌شوند. مدل معلم با استفاده از داده‌های آموزشی آموزش دیده شده و عملکرد خوبی دارد. حالا هدف اصلی این است که اطلاعاتی که مدل معلم از داده‌ها یاد گرفته است، به مدل دانش‌آموز منتقل شود (شکل ۲.۲). این روش می‌تواند به دو شکل انجام شود، روش هدف نرم و هدف سخت:

□ روش هدف نرم: در روش هدف نرم، مدل معلم پاسخ‌های خود را با توزیع احتمالی نرمالیزه شده ارائه می‌دهد که این توزیع احتمالی به عنوان برچسب‌های آموزشی برای مدل دانش‌آموز استفاده می‌شود. به عبارت دیگر، برای هر داده آموزشی، مدل معلم یک بردار خروجی که با احتمالی نرمالیزه شده تولید می‌کند و این بردار به عنوان برچسب آموزشی برای مدل دانش‌آموز در نظر گرفته می‌شود. برای محاسبه خطای آموزش، از تابع میانگین مربعات خطا بین بردار خروجی مدل معلم و توزیع احتمالی خروجی مدل دانش‌آموز استفاده می‌شود.

$$\mathcal{L}_{KD} = \alpha T^2 \sum_{i=1}^N \text{KL} \left( \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)}, \frac{\exp(z_i^T/T)}{\sum_j \exp(z_j^T/T)} \right)$$

در این فرمول،  $z_i$  بردار خروجی مدل دانش‌آموز برای داده آموزشی  $i$ ،  $z_i^T$  بردار خروجی مدل معلم برای همان داده آموزشی،  $T$  دمای softmax برای توزیع احتمالی نرمالیزه شده،  $N$  تعداد داده‌های آموزشی و KL فاصله Kullback-Leibler است.

□ روش هدف سخت: در روش هدف سخت، مدل معلم فقط یک برچسب صحیح برای هر داده آموزشی ارائه می‌دهد. به عبارت دیگر، برای هر داده آموزشی، مدل معلم یک برچسب صحیح تولید می‌کند که به عنوان برچسب آموزشی برای مدل دانش‌آموز در نظر گرفته می‌شود. برای محاسبه خطای آموزش، از تابع میانگین مربعات خطا بین برچسب صحیح مدل معلم و برچسب آموزشی مدل دانش‌آموز استفاده می‌شود.

$$\mathcal{L}_{KD} = \alpha \sum_{i=1}^N (\text{softmax}(z_i/T)) \log(\text{softmax}(z_i/T)) + (1 - \alpha) \sum_{i=1}^N (\text{softmax}(z_i)) \log(\text{softmax}(z_i))$$

در این فرمول،  $z_i$  بردار خروجی مدل دانش‌آموز برای داده آموزشی  $i$ ،  $T$  دمای softmax برای توزیع احتمالی نرمالیزه شده،  $N$  تعداد داده‌های آموزشی و softmax تابع softmax است. همچنین،  $\alpha$  و  $1 - \alpha$  وزن‌های متناسب برای تعیین میزان تأثیر هر دو قسمت از تابع هدف هستند.

همچنین، می‌توان از چند مدل معلم به عنوان یک مجموعه (Ensemble) استفاده کرد. در این روش، توزیع احتمالی خروجی‌های همه مدل‌های معلم، به عنوان برچسب‌های آموزشی برای مدل دانش‌آموز استفاده می‌شود. این روش می‌تواند بهبود عملکرد مدل دانش‌آموز را نسبت به استفاده از یک مدل معلم به صورت تنها فراهم کند. KD در پردازش تصویر می‌تواند به عنوان یک روش کاهش پیچیدگی و افزایش سرعت آموزش مدل‌های عمیق استفاده شود. با این روش، مدل دانش‌آموز با بهره‌گیری از دانش مدل معلم، به‌طور معمول با دقت بیشتری آموزش می‌بیند و در عین حال با تعداد پارامترهای کمتری نسبت به مدل معلم به نتایج مشابهی دست می‌یابد. برای مثال، می‌توان به کاربرد KD در آموزش مدل‌های تشخیص اشیاء از تصاویر اشاره کرد. در این حوزه، مدل‌های عمیق با تعداد پارامترهای بالا و پیچیدگی زیادی استفاده می‌شوند. با استفاده از KD، می‌توان به سادگی این مدل‌ها را به صورت مدل‌های کوچک‌تری با دقت نزدیک به مدل اصلی آموزش داد. یکی از انواع این مدل‌ها مدل EfficientNet می‌باشد که در سال ۲۰۱۹ معرفی شد. این مدل با استفاده از روش‌هایی مانند شبکه‌های عصبی بازگشتی (RNN) و مدل‌های توجه، پارامترهای قابل توجه کمتری نسبت به مدل‌های عمیق پیشین داشت و با دقت بالایی در تشخیص اشیاء از تصاویر بهترین نتایج را در مقایسه با مدل‌های قبلی به دست آورد. با استفاده از KD، می‌توان مدل‌های کوچک‌تری با دقت نزدیک به EfficientNet آموزش داد و همچنین سرعت آموزش و پیش‌بینی را افزایش داد.

## ۵.۲ مخلوط کردن داده‌ها

مخلوط کردن داده‌ها یک روش آموزشی برای شبکه‌های عصبی در پردازش تصویر است که با مخلوط کردن داده‌های آموزشی، بهبود اعتبار سنجی و جلوگیری از بیش‌برازش را هدف می‌گیرد. در این روش، دو تصویر انتخاب شده و با یکدیگر مخلوط می‌شوند و برچسب‌های متناظر با هر دو تصویر نیز با یکدیگر مخلوط می‌شوند. در واقع، با ایجاد یک نوع تصادف در فرآیند آموزش، نوعی از اطلاعات جدید به شبکه عرضه می‌شود که باعث بهبود دقت پیش‌بینی شبکه می‌شود. فرمول محاسبه تصویر مخلوط شده ( $\tilde{x}$ ) و برچسب مخلوط شده ( $\tilde{y}$ ) به صورت زیر است:

$$\tilde{x} = \lambda x_i + (1 - \lambda)x_j$$



$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j$$

در این فرمول،  $x_i$  و  $x_j$  دو تصویر انتخاب شده برای مخلوط کردن هستند،  $y_i$  و  $y_j$  برچسب‌های متناظر با هر دو تصویر هستند و  $\lambda$  یک عامل تصادفی است که مقدار آن بین ۰ و ۱ تعیین می‌شود. استفاده از مخلوط کردن داده‌ها می‌تواند به جلوگیری از بیش برآزش مدل و افزایش دقت آن در مواردی که داده‌های آموزشی کم و نادر هستند کمک کند. این روش در بسیاری از مسابقات مربوط به پردازش تصویر مانند مسابقات ImageNet و COCO به عنوان یکی از روش‌های موفق آموزش مدل‌های عمیق مورد استفاده قرار می‌گیرد.

## ۶.۲ برش مخلوط

برش مخلوط یک روش آموزشی برای شبکه‌های عصبی در پردازش تصویر است که با ترکیب تصاویر در هنگام آموزش، بهبود دقت پیش‌بینی شبکه را هدف می‌گیرد. در این روش، یک بخش مستطیلی از یک تصویر به طور تصادفی انتخاب شده و به تصویر دیگری الصاق می‌شود. سپس برچسب‌های متناظر با هر دو تصویر نیز با یکدیگر مخلوط می‌شوند. این روش به شبکه‌های عصبی کمک می‌کند تا علاوه بر استفاده از اطلاعات بصری، از اطلاعات موجود در برچسب‌ها نیز بهره بگیرند و دقت پیش‌بینی شبکه را افزایش دهند.

فرمول محاسبه تصویر مخلوط شده ( $\tilde{x}$ ) و برچسب مخلوط شده ( $\tilde{y}$ ) به صورت زیر است:

$$\tilde{x} = M \odot x_i + (1 - M) \odot x_j$$

$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j$$

در این فرمول،  $x_i$  و  $x_j$  دو تصویر انتخاب شده برای مخلوط کردن هستند،  $y_i$  و  $y_j$  برچسب‌های متناظر با هر دو تصویر هستند و  $M$  یک ماسک مستطیلی است که بخش متصل شده به تصویر دوم را مشخص می‌کند. همچنین،  $\lambda$  یک عامل تصادفی است که مقدار آن بین ۰ و ۱ تعیین می‌شود.

استفاده از برش مخلوط باعث می‌شود که شبکه عصبی در فرآیند آموزش، اطلاعات بصری و موجود در برچسب‌ها را با هم ترکیب کند و دقت پیش‌بینی شبکه را افزایش دهد. این روش در بسیاری از مسابقات مربوط به پردازش تصویر مانند مسابقات ImageNet و COCO به عنوان یکی از روش‌های موفق آموزش مدل‌های عمیق مورد استفاده قرار می‌گیرد.

## فصل ۳

### مجموعه دادگان

ارزیابی روش این مطالعه، روی مجموعه دادگان NEH انجام خواهد شد. مجموعه دادگان NEH متشکل از ۱۶۸۲۲ تصویر B-scan OCT با برچسب‌های Normal، Drusen، و CNV است که با تصویر برداری از چشم ۴۴۱ نفر از افراد بالای ۵۰ سال و بدون عارضه دیگری غیر از AMD در شبکیه، توسط سیستم تصویربرداری Heidelberg SD-OCT در بیمارستان چشم‌پزشکی نور تهران جمع‌آوری شده است.

همه تصاویر دارای کیفیت تصویر (اندازه‌گیری شده توسط سیستم تصویربرداری) بوده و توسط متخصصین برچسب‌گذاری و با فرمت‌های jpg و tiff ذخیره شده‌اند. [6] جهت انجام آزمایش، تصاویر به ترتیب با نسبت‌های 20 : 10 : 70 به صورت تصادفی به قسمت‌های آموزش، ارزیابی و آزمون تقسیم شده‌اند. جزئیات نحوه تقسیم مجموعه دادگان در جدول ۱.۳ قابل مشاهده است.

جدول ۱.۳: جزئیات نحوه تقسیم مجموعه دادگان NEH به قسمت‌های آزمون، ارزیابی و آموزش

	جایگاه استفاده از تصویر	تعداد تصاویر	
		برچسب تصویر	
مجموعه دادگان NEH-v2	Train	Normal	5991
		Drusen	3507
		CNV	2278
	Validation	Normal	861
		Drusen	491
		CNV	330
	Test	Normal	1732
		Drusen	1000
		CNV	632

### ۱.۳ کارهای مرتبط

در [7, 8, 9, 10] ویژگی‌های تصاویر OCT پس از پیش‌پردازش استخراج شده و به کمک ویژگی‌های استخراج شده و یک طبقه‌بند کلاسیک، تشخیص تصاویر حاوی نشانه‌های AMD صورت می‌گیرد. به‌طور ویژه [7] از ویژگی‌های مبتنی بر هیستوگرام‌های محلی و یک شبکه طبقه‌بند بیزی استفاده می‌کند. Srinivasan et al. با به‌کارگیری طبقه‌بند ماشین بردار پشتیبانی روی ویژگی‌های بدست آمده از توصیف‌کننده‌های HOG توانسته است تصاویر دارای برچسب‌های Normal، AMD و DME را به ترتیب با دقت‌های 86.67، 100 و 100 درصد تشخیص دهد [8]. هم‌چنین Sun et al نیز براساس آموزش واژه‌نامه و آموزش سه طبقه‌بند ماشین بردار پشتیبانی تصاویر SD-OCT با برچسب‌های Normal، AMD و DME را تفکیک کند.

در [10] نیز الگوریتمی بر پایه آموزش واژه‌نامه و طبقه‌بند شاخص‌های تصادفی ارائه و نشان‌داده شده که این روش عملکردی مشابه انسان روی تصاویر OCT مجموعه دادگان European Genetic Database به نمایش می‌گذارد. با توجه به قدرت مدل‌های مبتنی بر یادگیری عمیق در استخراج خودکار ویژگی‌ها، در سال‌های اخیر از این مدل‌ها برای بهبود نتایج دسته‌بندی روی مجموعه دادگان بزرگ‌تر استفاده شده است. در [11] و [12] به ترتیب از شبکه‌های VGG تغییر یافته و AlexNet برای دسته‌بندی تصاویر OCT استفاده شد. در [13] با ترکیب شبکه عصبی پیچشی چند مقیاسه‌ها و بهره‌گیری از یک تابع هزینه جدید توانست سه کلاس AMD، Normal و DME را در مجموعه دادگان NEH با امتیاز 0.9985 در معیار AUC طبقه‌بندی کند. هم‌چنین Kermany et al. در [14] و Li et al. در [15] به کمک آموزش انتقالی و به‌ترتیب با استفاده از شبکه‌های آموزش‌دیده InceptionV3 و VGG16 روی مجموعه دادگان UCSD توانستند به ترتیب دقت‌های 96.6% و 98.6% را کسب کنند.

Serener et al. نشان داد که برای تشخیص تصاویر دارای برچسب‌های AMD خیس و AMD خشک، شبکه ResNet50 نسبت به AlexNet عملکرد بهتری دارد [16]. در [17] شبکه‌های InceptionV3، ResNet50 و VGG16 مبتنی بر تکنیک آموزش انتقالی برای تشخیص تصاویر دارای برچسب AMD، Normal، AMD خیس و AMD خشک آموزش داده شدند و نشان داده شد که شبکه InceptionV3 عملکرد بهتری نسبت به دو شبکه دیگر از خود نشان می‌دهد.

در [18] مدل IFCNN که جهت بهره‌مندی توأمان از ویژگی‌های سطح پایین و بالا در شبکه‌های عصبی پیچشی، ویژگی‌های هر لایه را با تمام لایه‌های قبلی به صورت تکراری تلفیق می‌کند، برای طبقه‌بندی تصاویر مجموعه دادگان UCSD توسط Fang et al. مطرح شد. Huang et al. با به‌کارگیری ویژگی‌های استخراج شده توسط یک شبکه تقسیم‌بندی در مدل مطرح شده به نام Layer Guided CNN، که منجر به افزایش تمرکز در نواحی بامعنا تر برای تشخیص آسیب‌های ایجاد شده در شبکه چشم می‌شد، توانست عملکرد بهتری نسبت به شبکه IFCNN ثبت کند [19].

Fang et al. نیز در [20] روش جدیدی تحت عنوان شبکه عصبی پیچشی آگاه از آسیب را معرفی می‌کند که با بهره‌گیری از یک شبکه تشخیص آسیب، جهت ایجاد یک جدول توجه برای وزن‌دهی اطلاعات بدست آمده از فیلترهای پیچشی محلی، دقت طبقه‌بندی را در ازای افزایش پیچیدگی محاسباتی (به دلیل استفاده از یک شبکه مستقل از برای تولید جدول توجه) بهبود می‌بخشد. مشابه با [13] در [21] نیز از تلفیق ویژگی‌های بدست آمده از ساختار شبکه عصبی

پیچشی چند مقیاسه استفاده شده است، اما بر خلاف روش پیشنهادی Rasti et al. که نیاز به تنظیم کردن تابع اتلاف داشت، تابع اتلافی حساس به هزینه به کار گرفته شده که نیازی به تنظیم کردن آن نیست و موجب مدل کردن بهتر مجموعه دادگان UCSD نامتناسب می شود.

علیرغم این که تلفیق ویژگی های چند مقیاسه امکان استفاده توأمان مدل از ویژگی های سطح بالا و پایین را فراهم می کند، به دلیل استفاده از چندین ساختار متفاوت پیچشی از سطحی تا عمیق، زمان و هزینه یادگیری و استنتاج مدل نسبت به مدل های کلاسیک شبکه عصبی پیچشی، افزایش می یابد. Das et al. در سال ۲۰۲۰، B-scan attentive CNN را معرفی کرد که از مکانیزم توجه به خود در ساختار شبکه پیچشی استفاده کرده و بر اطلاعات مهم تر تصاویر B-scan 3D OCT برای انجام بهتر طبقه بندی، تأکید می کند [22].

در [23] RAG-FW را معرفی کرد که به عنوان یک چارچوب پیچشی دوگانه از شبکه های طبقه بند و تقسیم بندی توانست میانگین دقت 98.70% را برای تشخیص درجه آسیب های شبکه روی چندین مجموعه دادگان از تصاویر OCT دریافت کند. همچنین Thomas et al. توانست با استفاده از فیلترهایی با سایزهای مختلف در ساختار شبکه عصبی پیچشی چند مقیاسه ساختارهای محلی تصاویر را استخراج کرده و میانگین دقت وزن دار 99.73% را برای طبقه بندی باینری Normal در برابر AMD در مجموعه دادگان UCSD کسب کند [24].

در سال ۲۰۲۲ Sotoudeh-Paima et al. ضمن معرفی نسخه دوم از مجموعه دادگان NEH، توانست به کمک یک ساختار چند مقیاسه برای تلفیق ویژگی های یک شبکه VGG16 آموزش دیده روی مجموعه دادگان UCSD و ImageNet به دقت 93.4% روی نسخه بدترین حالت NEH، دست یابد. در واقع به دلیل این که حجم مجموعه دادگان NEH بسیار کم تر از مجموعه دادگان UCSD بود، استفاده از مدل آموزش داده شده روی مجموعه دادگان UCSD و ImageNet منجر به افزایش 1.2% در میانگین دقت مدل نسبت به، استفاده از مدل آموزش داده شده روی مجموعه دادگان صرفاً ImageNet شد [6]. در [25] با بهره گیری از الگوریتمی بهینه شده برای تقسیم بندی تصاویر شبکه و یک ساختار گروهی از درخت کیسه دار و دو مدل یادگیری عمیق برای تصاویر ۲ و ۳ بعدی یک طبقه بند تصاویر SD-OCT به کلاس های Early-AMD، Non-AMD و Int.-AMD ارائه شده است.

اخیراً Liew et al. روشی دستی به نام MSK MP که نسخه تعمیم یافته روش الگوهای باینری محلی است، برای استخراج ویژگی های تصاویر OCT معرفی کرده و از ویژگی های استخراج شده از این روش برای طبقه بندی تصاویر OCT با طبقه بندی های کلاسیک از جمله Ada Boost، Naïve Bayes، SVM، Random Forest و Rus Boost استفاده کرده و نشان داده است که برای مجموعه دادگان مختلف از تصاویر OCT، طبقه بند ماشین بردار پشتیبانی با کرنل های چند جمله ای عملکرد بهتری نسبت به طبقه بندی های کلاسیک دیگر ارائه می دهد [26]. در [27] شبکه dual guidance برای طبقه بندی و تقسیم بندی تصاویر OCT ارائه شده است. شبکه هدایت دوگانه از اطلاعات مربوط به یک وظیفه برای بهبود نتایج وظیفه دیگر بهره می برد، این ویژگی باعث بهبود نتایج هر دو وظیفه می شود اما در مقابل به دلیل متوالی بودن آموزش و تست در این شبکه ها پیچیدگی شبکه نسبت به شبکه های چند وظیفه ای افزایش می یابد. هم چنین محدودیت دیگر برای این شبکه ها نیاز به مجموعه دادگانی است که همزمان برچسب های کلاس و و پیکسلی را داشته باشد.

در [28] شبکه GABNet برای طبقه بندی تصاویر OCT شبکه معرفی می شود که از ترکیب یک بلوک توجه جهانی و یک شبکه عصبی پیچشی تشکیل می شود. در [29] روش شبکه عصبی پیچشی چند شاخه ای معرفی می شود که

هدف آن ارائه یک طبقه‌بند چند مدلی برای تشخیص بهتر AMD، به کمک ویژگی‌های استخراج شده از تصاویر OCT و CFP و تلفیق آن‌ها بر اساس مکانیزم توجه است. هم‌چنین در این مطالعه برای غلبه بر مشکل کمبود داده‌های چند مدلی یک شبکه مولد تخصصی طراحی و به کار گرفته می‌شود. در [30] Akinniyi et al. با استفاده از یک شبکه مقیاس سازگار ویژگی‌های چند مقیاسی تصاویر ورودی را استخراج کرده و در ادامه به کمک تلفیق ویژگی‌های استخراج شده در یک ساختار سلسله‌مراتبی دقت طبقه‌بندی را روی مجموعه دادگان UCSD به 99.69% (محاسبه شده با روش ارزیابی اعتبارسنجی پنج لایه متقابل) بهبود داده‌است. در [31] با بهره‌گیری از مدل InceptionResNetV2 به عنوان استخراج‌کننده ویژگی‌های تصاویر استفاده کرده و با ترکیب طبقه‌بندهای کلاسیک، عمل طبقه‌بندی را روی مجموعه دادگانی با ۵ کلاس انجام داده است.

## فصل ۴

### روش پیشنهادی

در این قسمت ابتدا مروری بر ساختار مبدل بینایی پزشکی به عنوان مبدل پایه برای بخیه زدن خواهیم داشت و سپس در ادامه ضمن مرور شبکه عصبی بخیه‌ای روش به کار رفته برای بدست آوردن مدل نهایی را شرح خواهیم داد.

#### ۱.۴ مروری بر مبدل بینایی پزشکی

مبدل بینایی پزشکی ساختاری دوگانه است متشکل از بلوک‌های پیچشی و مبدل برای استفاده از نقاط قوت هر دو ساختار جهت به کارگیری اطلاعات محلی و جهانی تصاویر برای بهبود قدرتمندی و دقت طبقه‌بندی تصاویر پزشکی که اخیراً توسط Manzari et al. معرفی شده است.

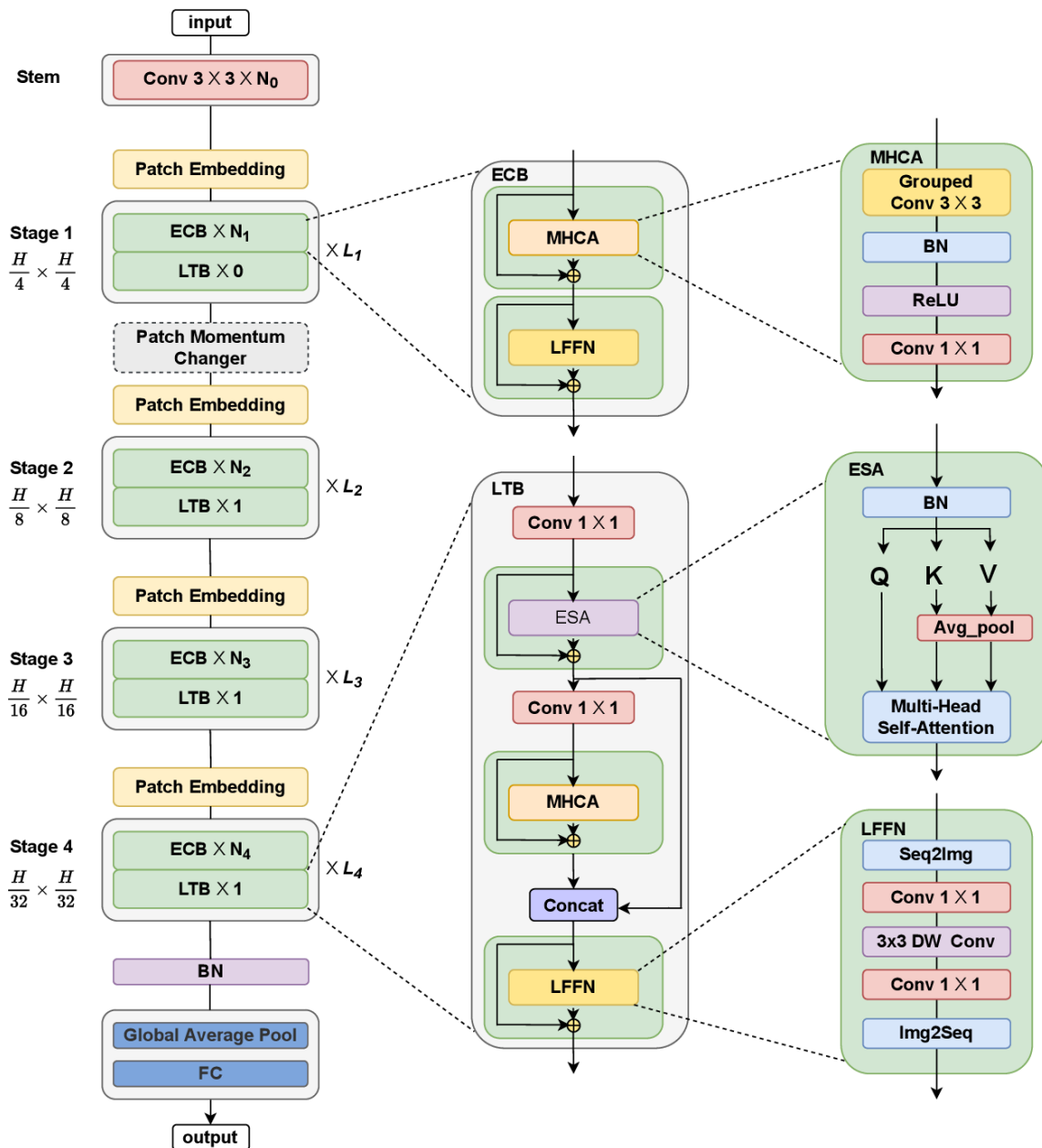
به طور خاص مبدل بینایی پزشکی، از بلوک پیچشی بهینه (ECB) و بلوک مبدل محلی (LTB) بهره می‌برد، هر چند که هر دو این بلوک‌ها از ساختار دوگانه بهره می‌برند اما تمرکز هر یک به ترتیب بیشتر بر روی اطلاعات محلی و دوربرد است. همچنین در [4] از یک لایه افزایش داده‌ها جدید به نام تغییر دهنده حرکت دسته‌ای (PMC) که عملکردی مشابه روش‌های برش مخلوط و مخلوط کردن داده‌ها دارند و منجر به افزایش قدرتمندی مدل می‌شود، استفاده شده است. در ادامه هر یک را مختصراً مرور خواهیم کرد. معماری کلی مبدل بینایی پزشکی در شکل ۱.۴ نشان داده شده است.

##### ۱.۱.۴ بلوک پیچشی بهینه

بلوک پیچشی بهینه به منظور یادگیری بهتر بافت‌های پیچیده پزشکی که در پس زمینه تصاویر قرار دارند ارائه شده و منجر به افزایش دقت مدل می‌شود. این بلوک حاوی ساختاری بر پایه توجه بوده و خود از دو زیر لایه توجه پیچشی چندسری (MHCA) و شبکه عصبی پیشخور محلی (LFFN) تشکیل شده است. لایه MHCA به عنوان یک لایه توجه باعث توجه مدل به نقاط حاوی اطلاعات مهم‌تر شده و لایه LFFN با به کارگیری زیرلایه پیچیدگی عمقی ساختارهای محلی را یاد می‌گیرد.

##### ۲.۱.۴ بلوک مبدل محلی

بلوک مبدل محلی نیز ساختاری مشابه ECB را دنبال می‌کند با این تفاوت که در ابتدا یک مازول توجه به خود بهینه



شکل ۱.۴: ساختار کلی مبدل بینایی پزشکی و نمایش زیرلایه‌های درونی هر بلوک

(ESA) اضافه شده است تا از ویژگی مبدل در دنبال کردن اطلاعات جهانی بهره ببرد. نکته قابل توجه در ماژول ESA استفاده از Average Pool به منظور کاهش هزینه محاسباتی است.

### ۳.۱.۴ مبدل بینایی پزشکی میکرو

مدل‌های مبدل بینایی پزشکی که در [4] معرفی شدند شامل مدل‌های بسیار کوچک، کوچک و بزرگ هستند که هر یک به ترتیب دارای ۱۴.۳۱، ۴۱.۴۴ و ۶۸.۵۷ میلیون پارامتر هستند. علیرغم این که در مجموعه دادگان MedMNIST (که خود شامل چندین زیرمجموعه دادگان در حوزه‌های مختلف پزشکی است) مدل‌های معرفی شده عملکرد مناسبی از خود نشان دادند، ما در این مطالعه دریافتیم که این مدل‌ها روی مجموعه دادگان کوچک NEH دچار بیش‌برازش می‌شوند. به همین علت مدل مبدل بینایی پزشکی میکرو را با تعداد پارامترهای ۵۰.۲۴ میلیون به خانواده مدل‌های مبدل بینایی پزشکی اضافه کردیم.

تنها تفاوت این مدل با سایر مدل‌ها در تعداد لایه‌های طبقه سوم و مقدار نرخ حذف تصادفی مسیر است. مقدار این نرخ برای مدل بسیار کوچک برابر ۱۰.۰ بوده و برای مدل میکرو ۰.۵۰ در نظر گرفته شده است. در جدول ۱.۴ مشخصات کامل‌تر معماری مدل‌های بسیار کوچک و میکرو آورده شده است.

### ۴.۱.۴ مبدل بینایی پزشکی بسیار کوچک و میکرو همراه با نویز

مدل‌های نویزی مبدل بینایی پزشکی به لحاظ ساختاری کاملاً مشابه مدل‌های عادی متناظر مبدل بینایی پزشکی هستند و تنها تفاوت مربوط به آموزش مدل‌های نویزی است. همانطور که پیش‌تر گفته شد به دلیل بیش‌برازش مدل‌های مبدل بینایی پزشکی روی مجموعه دادگان‌های کوچک، مدل‌های نویزی را به مجموعه مدل‌های مبدل بینایی پزشکی اضافه کردیم تا با اضافه کردن نویزی گاوسی با میانگین ۰ و انحراف معیاری به اندازه ۱/۱۰ انحراف معیار مجموعه دادگان آموزشی به ورودی مدل، از بیش‌برازش شدن سریع آن جلوگیری کنیم.

## ۲.۴ مروری بر شبکه عصبی بخیه‌ای

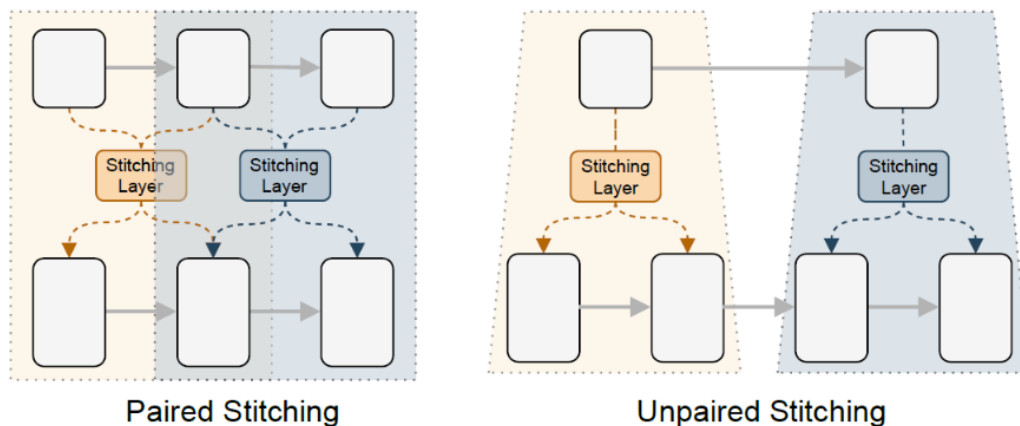
Pan et al. در [5] به منظور رسیدن به یک مصالحه مطلوب بین عملکرد و دقت در شبکه‌های عصبی مصنوعی، روشی را مطرح کرد که با در اختیار داشتن دو یا چند شبکه عصبی از پیش آموزش داده شده (لنگر) جست و جویی با هزینه محاسباتی کم (نسبت به روش‌های مشابه)، برای پیدا کردن ساختاری بهینه با دقت برابر یا بهتر نسبت به لنگرها انجام می‌دهد. که نام آن SNN می‌باشد. نکته قابل توجه در انتخاب لایه‌ها برای بخیه‌زدن استراتژی‌های جفت و بخیه بدون جفت که در روش پیشنهادی به ترتیب برای بخیه‌زدن طبقه‌های هم‌طول و با طول‌های مختلف استفاده شده است. در استراتژی جفت یک پنجره متحرک روی مدلها با گام  $s = 1$  و طول  $k = 2$  حرکت کرده و پیکربندی‌های مختلف برای جست و جو را استخراج می‌کند. هم‌چنین در استراتژی بدون جفت نیز لنگر با طول کم‌تر معیار قرار گرفته و هر لایه از آن متناسب با طول لنگر بزرگ‌تر با بیش از یکی از لایه‌های لنگر بزرگ‌تر تشکیل یک کاندید برای لایه بخیه‌ای می‌دهد (شکل ۲-۴).

بر اساس الگوریتم پیشنهاد شده، میان هر زوج از پیکربندی‌های استخراج شده به صورت (لایه  $m$  ام از لنگر  $i + 1$  ام، لایه  $n$  ام از لنگر  $i$  ام) یک شبکه خطی (بخیه) برای تبدیل نقشه فعال سازی‌های لایه  $n$  ام از لنگر  $i$  ام و لایه  $m - 1$  ام از لنگر  $i + 1$  ام تعریف می‌شوند. در ادامه روی یک دسته کوچک، هر یک از بخیه‌ها به بر اساس پاسخ حالت بسته برای یافتن تبدیل با کمترین میانگین خطا، مقاداردهی اولیه می‌شوند. در چرخه‌های آتی، الگوریتم یک بخیه را به صورت تصادفی انتخاب کرده و پارامترهای آن را به روز رسانی می‌کند. هم‌چنین برای بهبود دقت مدل نهایی، در فرآیند



جدول ۱.۴: مشخصات کامل معماری مدل‌های میکرو و بسیار کوچک از مبدل بینایی پزشکی.  $C$  و  $S$  به ترتیب بیانگر تعداد کانال‌ها و  $stride$  در لایه‌های پیچشی است. همچنین بالانویس بلوک‌های  $ECB$  و  $LTB$  نشانگر ابعاد خروجی هر بلوک است.

Stages	Layer type	MedViT-Tiny	MedViT-Micro
Stem	Convolution Layers	$C = 64, S = 2, 3 \times Conv3$	
		$C = 32, S = 1, 3 \times Conv3$	
		$C = 64, S = 1, 3 \times Conv3$	
		$C = 64, S = 2, 3 \times Conv3$	
Stage 1	Patch Embedding	$C = 96, 1 \times Conv1$	
	MedViT Block	$1 \times [3 \times ECB^{96}]$	
Stage 2	Patch Embedding	$Avg_{pool}, S = 2$	
		$C = 192, 1 \times Conv1$	
	MedViT Block	$1 \times [ECB^{192} \times 3 LTB^{256} \times 1]$	
Stage 3	Patch Embedding	$Avg_{pool}, S = 2$	
		$C = 384, 1 \times Conv1$	
	MedViT Block	$1 \times [ECB^{384} \times 4 LTB^{512} \times 1]$	$2 \times [ECB^{384} \times 4 LTB^{512} \times 1]$
Stage 4	Patch Embedding	$Avg_{pool}, S = 2$	
		$C = 768, 1 \times Conv1$	
	MedViT Block	$1 \times [ECB^{768} \times 2 LTB^{1024} \times 1]$	
Path dropout		0.1	0.05
Number of parameters(M)		31.14	24.50



شکل ۲.۴: استراتژی‌های بخیه زدن جفتی و بخیه زدن بدون جفت

آموزش از تقطیر دانش بر پایه معماری معلم و دانش‌آموز به روش هدف سخت [32] و افزایش داده‌های مخلوط کردن داده‌ها و برش مخلوط استفاده می‌شود.

#### ۱.۲.۴ بخیه زدن مبدل بینایی پزشکی

در این بخش روش استفاده شده برای بدست آوردن مدل بهینه برای بخیه زدن مبدل بینایی پزشکی را شرح خواهیم داد. همانطور که پیشتر گفته شد برای بخیه زدن نیاز به مدل‌های از پیش‌آموزش داده شده است. همچنین جهت بهره‌مندی از تقطیر دانش نیز، به یک مدل معلم احتیاج خواهیم داشت. با آن که می‌توان همزمان تعداد بیش از دو لنگر را بخیه زد، اما ما به بخیه زدن تنها دو لنگر به صورت یکطرفه (کوچک به بزرگ) در هر جست و جو کفایت خواهیم کرد، از طرف دیگر برای کاهش فضای جست و جو از هر دو پیکربندی کاندید مشابه (شماره لایه‌های نزدیک به هم) فقط یکی از پیکربندی‌ها را حفظ خواهیم کرد. همچنین برای مدل معلم نیز از یک مدل مبدل بینایی پزشکی استفاده خواهیم کرد.

## فصل ۵

# آزمایشات و نتایج

### ۱.۵ آزمایشات

□ پیش پردازش داده

برای انجام آزمایش‌ها در ابتدا تمامی تصاویر آزمون، ارزیابی و آزمایش به ابعاد  $224 \times 224$  تغییر داده شده و غیر از این پیش پردازش خاصی روی تصاویر انجام نشده است.

□ افزایش داده‌ها

برای افزایش داده‌ها مطابق [6] از تبدیل‌های جدول ۱.۵ استفاده شده است. قابل ذکر است که پیاده‌سازی تبدیل‌ها ترکیبی از `torchvision.transforms` بوده و لزوماً همه تبدیل‌ها روی یک تصویر اعمال نمی‌شود. همچنین در پایان این مرحله همه تصاویر به `Tensor` تبدیل شده و برای آموزش مدل‌های نوپزی، نویز گاوسی در این مرحله به `Tensor` ها افزوده می‌شود.

جدول ۱.۵: تبدیل‌های انجام شده برای افزایش داده‌ها

نام تبدیل در <code>torchvision.transforms</code>	پارامتر	مقدار یا بازه
RandomAffine	degrees	(-15, 15)
	shear	(-5, 5)
	scale	(0.8, 1.2)
ColorJitter	brightness	(0.8, 1.2)
RandomHorizontalFlip	probability	0.5

## □ مدل‌ها

برای رسیدن به مدل نهایی بخیه‌ای نیاز است ابتدا مدل‌های مبدل بینایی میکرو، بسیار کوچک، میکرو همراه با نويز و بسیار کوچک همراه با نويز آموزش داده شوند. هم‌چنین برای ارزیابی تأثیر استفاده از مدل مبدل بینایی پزشکی به عنوان مدل معلم، یک مدل DenseNet201 نیز آموزش داده و در یکی از پیکربندی‌ها از آن به عنوان معلم استفاده خواهد شد.

## □ تنظیمات آزمایش‌ها

کلیه آزمایش‌ها با سایز دسته‌بندی برابر 32 و بهینه‌ساز AdamW روی 100 epoch انجام شده است. برای آموزش مبدل بینایی پزشکی‌ها و DenseNet201 از آنتروپی متقابل به عنوان معیار تلفات استفاده شده و نیز مقدار ابتدایی نرخ یادگیری برابر 0.005 تنظیم شده است که در epoch‌های 50 و 75 مقدار آن 0.1 برابر می‌شود. برای بخیه‌زدن لنگرها از تنظیمات و پیکره‌بندی پیش‌فرض پیاده‌سازی رسمی SNet که برای بخیه‌زدن LeViT‌ها ارائه شده، استفاده شده است.

## □ ماتریس‌های ارزیابی

برای ارزیابی بهتر روش ارائه شده و مقایسه نتایج بدست‌آمده با نتایج مطالعات قبلی معیارهای دقت، حساسیت و اختصاص و ماتریس در هم ریختگی را به همراه تعداد پارامترهای نهایی مدل و تعداد عملیات‌های ممیز شناور محاسبه و ارائه خواهیم کرد.

## ۲.۵ نتایج

در جدول ۲.۵ دقت آزمون و تعداد پارامترهای چند مدل مختلف مبدل بینایی پزشکی گزارش شده است. ملاحظه می‌شود که نسخه میکرو معرفی شده در این کار، با تعداد پارامترهای کم‌تر نتیجه بهتری نسبت به نسخه کوچک روی NEH از خود نشان داده است. هم‌چنین مشاهده می‌شود که اضافه کردن نويز در آموزش نسخه بسیار کوچک دقت آن را بهبود داده است اما در نسخه میکرو چنین نیست.

در جدول ۳.۵ دقت آزمون برای دو پیکره‌بندی مختلف با تفاوت در مدل معلم‌ها را برای بخیه‌زدن دو مدل مبدل بینایی پزشکی میکرو مختلف (با مقداردهی‌های اولیه متفاوت) ارائه شده است که نشان می‌دهد دقت پیکره‌بندی با مدل معلم مبدل بینایی پزشکی میکرو همراه با نويز بیشتر است.

در جدول ۴.۵ نیز نتایج روش پیشنهادی و سایر مطالعات قبلی آورده شده است. پژوهش‌های اشاره شده در آنها به این ترتیب می‌باشد:

### ۱. پژوهش اول

در پژوهش اول از مدل kernels چند سایزی استفاده شده و ارزیابی به روش اعتبارسنجی پنج لایه متقابل می‌باشد همچنین از نسخه worst-case dataset استفاده شده است. [26]

۲. پژوهش دوم

در این مورد مدل‌ها روی مجموعه دادگان ImageNet از پیش آموزش داده شده و ارزیابی به روش اعتبارسنجی پنج لایه متقابل می‌باشد و از نسخه worst-case dataset استفاده شده. [6]

۳. پژوهش سوم

در این مورد مدل‌ها روی مجموعه دادگان‌های ImageNet و UCSD از پیش آموزش داده شده‌اند، و ارزیابی به روش اعتبارسنجی پنج لایه متقابل می‌باشد. همچنین از نسخه worst-case dataset استفاده شده [6]

۴. پژوهش چهارم

در این پژوهش نهایی مدل استفاده شده، مدل SNN می‌باشد که در آن مدل‌های مبدل بینایی پزشکی میکرو و مبدل بینایی پزشکی بسیار کوچک همراه با نویز بخیه زده شدند و همچنین از مبدل بینایی پزشکی میکرو همراه با نویز برای مدل معلم استفاده می‌شود. و نتیجه بر روی مجموعه دادگان آزمون اعلان شده است.

ملاحظه می‌شود که روش پیشنهادی دقت بهتری نسبت به مدلی که تنها بر روی مجموعه دادگان NEHv2 آموزش داده شده است، دریافت نموده و همچنین نتیجه آن با مدل SOTA که از مجموعه دادگان بزرگ و تخصصی UCSD استفاده کرده قابل رقابت است.

جدول ۲.۵: تبدیل‌های انجام شده برای افزایش داده‌ها

نام مدل	دقت آزمون (%)	تعداد پارامترها (M)
MedViT-Tiny	89.95	31.14
MedViT-Tiny-Noisy	91.11	31.14
MedViT-Micro	<b>93.34</b>	24.50
MedViT-Micro-Noisy	93.16	24.50
DenseNet201	91.79	<b>18.10</b>

جدول ۳.۵: مقایسه دقت آزمون مدل بخیه‌زده شده با استفاده از مدل معلم‌های DenseNet و مبدل بینایی پزشکی

لنگرها	Teacher Model	دقت آزمون
دو مدل MedViT-Micro با مقادری‌های اولیه متفاوت	DenseNet201	%93.19
	MedViT-Micro-Noisy	<b>%93.46</b>

جدول ۴.۵: مقایسه دقت آزمون مدل بخیه‌زده شده با استفاده از مدل معلم های DenseNet و مبدل بینایی پزشکی

نام مدل	معیارهای مقایسه					پژوهش
	FLOPs(G)	Params(M)	Specificity	Sensitivity	Accuracy	
Multi-size kernels	89.2%	88.51%	94.31%	-	-	پژوهش اول
FPN-EfficientNetB0	87.8%	86.6%	93.3%	12.7	-	پژوهش دوم
FPN-ResNet50	90.1%	89.8%	94.8%	31.1	-	
FPN-DenseNet121	90.9%	90.5%	95.2%	14.3	-	
FPN-VGG16	92.0%	91.8%	95.8%	21.6	-	
FPN-VGG16*	93.4%	93.1%	<b>96.5%</b>	21.6	-	پژوهش سوم
روش پیشنهادی	<b>93.49%</b>	<b>93.49%</b>	94.71%	24.50	4.65	پژوهش چهارم

## فصل ۶

### جمع بندی و نتیجه گیری

در این پروژه ما به بررسی روش SNN بر روی مجموعه دادگان NEH پرداختیم. از آنجایی که مدل مبدل بینایی پزشکی در مجموعه دادگان های پزشکی عملکرد خوبی داشته و به نتایج بهتری دست یافته، در آزمایش ها مدل های متفاوت مبدل بینایی پزشکی را آموزش داده و آنها را بخیه زدیم. در این روند متوجه شدیم که مدلهایی که برای مبدل بینایی پزشکی تعریف شده بود، برای مجموعه دادگان NEH که تعداد تصاویر محدودی دارد، بزرگ می باشد. در نتیجه مدل مبدل بینایی پزشکی میکرو را با تعداد پارامتر های کمتر معرفی کردیم و مشاهده کردیم که نتیجه بسیار بهتری نسبت به مدل های مبدل بینایی پزشکی اولیه داشت. همچنین برای بهبود نتایج مدل های اصلی هم اضافه کردن نویز گاوسی به ورودی مدل را آزمایش کردیم و نشان دادیم که روش مؤثری می باشد. در این آزمایش ها مشاهده کردیم که دقت نهایی این روش به مقدار قابل قبولی رسیده و نسبت به روش های فعلی نتایج نسبتاً بهینه ای در موازنه دقت و کارایی داشته است.

البته در این آزمایش ها کمبودهایی نیز داشتیم، برای مثال از آنجایی که مجموعه دادگان مورد مطالعه نامتعادل می باشد می توان به جای معیار آنتروپی متقابل از focal loss استفاده کرد که انتظار داریم عملکرد بهتری داشته باشد و به نتایج دقیق تری دست یابد. از طرفی محدودیت هایی نیز در این مسیر وجود داشت که باعث کند شدن سرعت یا عدم امکان بکارگیری از شیوه های متفاوت شد. برای مثال از آنجایی که در روش ما چند مدل به هم بخیه شده اند، برای استفاده از اعتبارسنجی متقابل نیاز به قدرت محاسباتی زیادی داشتیم که امکان پذیر نبود و به همین دلیل نتوانستیم از این روش بهره ببریم.

در آینده ما قصد داریم که به این محدودیت ها و نواقص بپردازیم و سعی در بهبود آنها داشته باشیم تا به مدلی بهینه و با دقت بالا تر از کارهای فعلی دست یابیم.

مسئله بیماری AMD از اهمیت بالایی برخوردار است و هنوز نیاز به پیشرفت و توسعه روی آن وجود دارد. برای مثال با توسعه مجموعه دادگان های دارای تصاویر بدست آمده از فناوری های مختلف تصویر برداری از شبکه، می توان مدل های چند مدله برای بالابردن قدرت پیشبینی بسازیم و همچنین مدل های چند وظیفه ای ای که جهت انجام دسته بندی تقسیم بندی نواحی درگیر Drusen را نیز انجام دهد تا ردیابی بهبود یا وخامت بیماری آسان تر شود.

## واژه‌نامه فارسی به انگلیسی

Transfer Learning	آموزش انتقالی
Cross Entropy	آنترپی متقابل
Retinal Pigment Epithelium (RPE)	اپی تلیوم رنگدانه شبکیه
Specificity	اختصاص
Multi-Scale Fusion	ادغام چند مقیاسه
Age-related Macular Degeneration(AMD)	استحاله ماکولار وابسته به سن
Five-Fold Cross Validation	اعتبارسنجی پنج لایه متقابل
Unpaired Stitching	بخیه زدن بدون جفت
Paired Stitching	بخیه زدن جفتی
CutMix	برش مخلوط
Efficient Convolutional Block(ECB)	بلوک پیچشی بهینه
Global Attention Block	بلوک توجه جهانی
Local Transformer Block(LTB)	بلوک میدل محلی
Pixel to Sequence	پیکسل به دنباله
Embedding	تعبیر
Patch Momentum Changer(PMC)	تغییر دهنده حرکت دسته‌ای
Knowledge Distillation(KD)	تقطیر دانش
Attention	توجه
Efficient Self-Attention	توجه به خود بهینه
Multi-Head Convolutional Attention(MHCA)	توجه پیچشی چندسر
Multi-Head Attention	توجه چند سر
Multilayer Perceptron	پرسپترون چندلایه‌ای
Depth-wise Convolution	پیچیدگی عمقی
Sensitivity	حساسیت
Bagged Tree	درخت کیسه دار
Drusen	دروزن
Mini-Batch	دسته کوچک
Accuracy	دقت
Bayesian Network Classifier	شبکه طبقه‌بند بیزی



Stitchable Neural Network(SNN)	شبکه عصبی بخیه‌ای
Convolutional Neural Network	شبکه عصبی پیچشی
Lesion-Aware Convolutional Neural Network	شبکه عصبی پیچشی آگاه از آسیب
Multi-Scale CNN	شبکه عصبی پیچشی چند مقیاسه
Feed-Forward Neural Network	شبکه عصبی پیشخور
Generative Adversarial Networks(GAN)	شبکه مولد تخصصی
Choriocapillaris	شبکه مویرگی مشیمیه
Least Square Error	کمترین میانگین خطا
Photoreceptors	گیرنده‌های نوری
Activation Layer	لایه فعال سازی
Fully Connected Layer	لایه کاملاً متصل
Confusion Matrix	ماتریس در هم ریختگی
Support Vector Machine	ماشین بردار پشتیبانی
Vision Transformer	مبدل بینایی
MedViT	مبدل بینایی پزشکی
MedViT-Large	مبدل بینایی پزشکی بزرگ
MedViT-Tiny	مبدل بینایی پزشکی بسیار کوچک
MedViT-Small	مبدل بینایی پزشکی کوچک
MedViT-Micro	مبدل بینایی پزشکی میکرو
Hybrid Transformer	مبدل دوگانه
MixUp	مخلوط کردن داده‌ها
Teacher Model	مدل معلم
Student Model	مدل دانش آموز
Optical Coherence Tomography (OCT)	مقطع‌نگاری هم‌دوسی اپتیکی
Scale-Adaptive	مقیاس سازگار
Path-Dropout	نرخ حذف تصادفی مسیر
Learning-Rate	نرخ یادگیری
Activation Map	نقشه فعال سازی
Dual Guidance	هدایت دوگانه
Hard-Target	هدف سخت
Soft-Target	هدف نرم
Artificial intelligence	هوش مصنوعی
Deep Learning	یادگیری عمیق
Machine Learning	یادگیری ماشین

## واژه‌نامه انگلیسی به فارسی

Activation Layer	لایه فعال سازی
Activation Map	نقشه فعال سازی
Accuracy	دقت
Age-related Macular Degeneration(AMD)	استحاله ماکولار وابسته به سن
Artificial intelligence	هوش مصنوعی
Attention	توجه
Bagged Tree	درخت کیسه دار
Bayesian Network Classifier	شبکه طبقه‌بند بیزی
Choriocapillaris	شبکه مویرگی مشیمیه
Confusion Matrix	ماتریس در هم ریختگی
Convolutional Neural Network	شبکه عصبی پیچشی
Cross Entropy	آنترپی متقابل
CutMix	برش مخلوط
Dual Guidance	هدایت دوگانه
Deep Learning	یادگیری عمیق
Depth-wise Convolution	پیچیدگی عمقی
Drusen	دروزن
Efficient Convolutional Block(ECB)	بلوک پیچشی بهینه
Efficient Self-Attention	توجه به خود بهینه
Embedding	تعبیر
Feed-Forward Neural Network	شبکه عصبی پیشخور
Five-Fold Cross Validation	اعتبارسنجی پنج لایه متقابل
Fully Connected Layer	لایه کاملاً متصل
Generative Adversarial Networks(GAN)	شبکه مولد تخصصی
Global Attention Block	بلوک توجه جهانی
Hard-Target	هدف سخت
Hybrid Transformer	مبدل دوگانه
Knowledge Distillation(KD)	تقطیر دانش
Least Square Error	کمترین میانگین خطا

Learning-Rate	نرخ یادگیری
Lesion-Aware Convolutional Neural Network	شبکه عصبی پیچشی آگاه از آسیب
Local Transformer Block(LTB)	بلوک مبدل محلی
Machine Learning	یادگیری ماشین
MedViT	مبدل بینایی پزشکی
MedViT-Large	مبدل بینایی پزشکی بزرگ
MedViT-Micro	مبدل بینایی پزشکی میکرو
MedViT-Small	مبدل بینایی پزشکی کوچک
MedViT-Tiny	مبدل بینایی پزشکی بسیار کوچک
Mini-Batch	دسته کوچک
MixUp	مخلوط کردن داده‌ها
Multilayer Perceptron	پرسپترون چندلایه‌ای
Multi-Head Attention	توجه چند سر
Multi-Head Convolutional Attention(MHCA)	توجه پیچشی چندسر
Multi-Scale CNN	شبکه عصبی پیچشی چند مقیاسه
Multi-Scale Fusion	ادغام چند مقیاسه
Optical Coherence Tomography (OCT)	مقطع‌نگاری همدوسی اپتیکی
Patch Momentum Changer(PMC)	تغییر دهنده حرکت دسته‌ای
Path-Dropout	نرخ حذف تصادفی مسیر
Paired Stitching	بخیه زدن جفتی
Photoreceptors	گیرنده‌های نوری
Pixel to Sequence	پیکسل به دنباله
Retinal Pigment Epithelium (RPE)	اپی تلیوم رنگدانه شبکیه
Scale-Adaptive	مقیاس سازگار
Sensitivity	حساسیت
Soft-Target	هدف نرم
Specificity	اختصاص
Stitchable Neural Network(SNN)	شبکه عصبی بخیه‌ای
Student Model	مدل دانش آموز
Support Vector Machine	ماشین بردار پشتیبانی
Teacher Model	مدل معلم
Transfer Learning	آموزش انتقالی
Vision Transformer	مبدل بینایی
Unpaired Stitching	بخیه زدن بدون جفت

# Bibliography

- [1] Robyn H Guymer and Thomas G Campbell. Age-related macular degeneration. *The Lancet*, 2023.
- [2] Nasim Salimiaghdam, Mohammad Riazi-Esfahani, Paula S Fukuhara, Kevin Schneider, and M Cristina Kenney. Age-related macular degeneration (amd): A review on its epidemiology and risk factors. *The Open Ophthalmology Journal*, 13(1), 2019.
- [3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [4] Omid Nejati Manzari, Hamid Ahmadabadi, Hossein Kashiani, Shahriar B Shokouhi, and Ahmad Ayatollahi. Medvit: a robust vision transformer for generalized medical image classification. *Computers in Biology and Medicine*, 157:106791, 2023.
- [5] Zizheng Pan, Jianfei Cai, and Bohan Zhuang. Stitchable neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16102–16112, 2023.
- [6] Saman Sotoudeh-Paima, Ata Jodeiri, Fedra Hajizadeh, and Hamid Soltanian-Zadeh. Multi-scale convolutional neural network for automated amd classification using retinal oct images. *Computers in biology and medicine*, 144:105368, 2022.
- [7] Abdulrahman Albarrak, Frans Coenen, Yalin Zheng, et al. Age-related macular degeneration identification in volumetric optical coherence tomography

- using decomposition and local feature extraction. In *Proceedings of 2013 international conference on medical image, understanding and analysis*, pages 59–64, 2013.
- [8] Pratul P Srinivasan, Leo A Kim, Priyatham S Mettu, Scott W Cousins, Grant M Comer, Joseph A Izatt, and Sina Farsiu. Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomedical optics express*, 5(10):3568–3577, 2014.
- [9] Yankui Sun, Shan Li, and Zhongyang Sun. Fully automated macular pathology detection in retina optical coherence tomography images using sparse coding and dictionary learning. *Journal of biomedical optics*, 22(1):016012–016012, 2017.
- [10] Freerk G Venhuizen, Bram van Ginneken, Freekje van Asten, Mark JJP van Grinsven, Sascha Fauser, Carel B Hoyng, Thomas Theelen, and Clara I Sánchez. Automated staging of age-related macular degeneration using optical coherence tomography. *Investigative ophthalmology & visual science*, 58(4):2318–2328, 2017.
- [11] Cecilia S Lee, Doug M Baughman, and Aaron Y Lee. Deep learning is effective for classifying normal versus age-related macular degeneration oct images. *Ophthalmology Retina*, 1(4):322–327, 2017.
- [12] Sertan Kaymak and Ali Serener. Automated age-related macular degeneration and diabetic macular edema detection on oct images using deep learning. In *2018 IEEE 14th international conference on intelligent computer communication and processing (ICCP)*, pages 265–269. IEEE, 2018.
- [13] Reza Rasti, Hossein Rabbani, Alireza Mehridehnavi, and Fedra Hajizadeh. Macular oct classification using a multi-scale convolutional neural network ensemble. *IEEE transactions on medical imaging*, 37(4):1024–1034, 2017.
- [14] Daniel S Kermany, Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *cell*, 172(5):1122–1131, 2018.
- [15] Feng Li, Hua Chen, Zheng Liu, Xuedian Zhang, and Zhizheng Wu. Fully automated detection of retinal disorders by image-based deep learning. *Graefe’s Archive for Clinical and Experimental Ophthalmology*, 257:495–505, 2019.

- [16] Ali Serener and Sertan Serte. Dry and wet age-related macular degeneration classification using oct images and deep learning. In *2019 Scientific meeting on electrical-electronics & biomedical engineering and computer science (EBBT)*, pages 1–4. IEEE, 2019.
- [17] De-Kuang Hwang, Chih-Chien Hsu, Kao-Jung Chang, Daniel Chao, Chuan-Hu Sun, Ying-Chun Jheng, Aliaksandr A Yarmishyn, Jau-Ching Wu, Ching-Yao Tsai, Mong-Lien Wang, et al. Artificial intelligence-based decision-making for age-related macular degeneration. *Theranostics*, 9(1):232, 2019.
- [18] Leyuan Fang, Yuxuan Jin, Laifeng Huang, Siyu Guo, Guangzhe Zhao, and Xiangdong Chen. Iterative fusion convolutional neural networks for classification of optical coherence tomography images. *Journal of Visual Communication and Image Representation*, 59:327–333, 2019.
- [19] Laifeng Huang, Xingxin He, Leyuan Fang, Hossein Rabbani, and Xiangdong Chen. Automatic classification of retinal optical coherence tomography images with layer guided convolutional neural network. *IEEE Signal Processing Letters*, 26(7):1026–1030, 2019.
- [20] Leyuan Fang, Chong Wang, Shutao Li, Hossein Rabbani, Xiangdong Chen, and Zhimin Liu. Attention to lesion: Lesion-aware convolutional neural network for retinal optical coherence tomography image classification. *IEEE transactions on medical imaging*, 38(8):1959–1970, 2019.
- [21] Vineeta Das, Samarendra Dandapat, and Prabin Kumar Bora. Multi-scale deep feature fusion for automated classification of macular pathologies from oct images. *Biomedical signal processing and Control*, 54:101605, 2019.
- [22] Vineeta Das, Eedara Prabhakararao, Samarendra Dandapat, and Prabin Kumar Bora. B-scan attentive cnn for the classification of retinal optical coherence tomography volumes. *IEEE Signal Processing Letters*, 27:1025–1029, 2020.
- [23] Taimur Hassan, Muhammad Usman Akram, Naoufel Werghi, and Muhammad Noman Nazir. Rag-fw: A hybrid convolutional framework for the automated extraction of retinal lesions and lesion-influenced grading of human retinal pathology. *IEEE journal of biomedical and health informatics*, 25(1):108–120, 2020.

- [24] Anju Thomas, PM Harikrishnan, Adithya K Krishna, P Palanisamy, and Varun P Gopi. A novel multiscale convolutional neural network based age-related macular degeneration detection using oct images. *Biomedical Signal Processing and Control*, 67:102538, 2021.
- [25] Mousa Moradi, Yu Chen, Xian Du, and Johanna M Seddon. Deep ensemble learning for automated non-advanced amd classification using optimized retinal layer segmentation and sd-oct scans. *Computers in Biology and Medicine*, 154:106512, 2023.
- [26] Alex Liew, Sos Aгаian, and Samir Benbelkacem. Distinctions between choroidal neovascularization and age macular degeneration in ocular disease predictions via multi-size kernels  $\xi$ cho-weighted median patterns. *Diagnostics*, 13(4):729, 2023.
- [27] Shengyong Diao, Jinzhu Su, Changqing Yang, Weifang Zhu, Dehui Xiang, Xinjian Chen, Qing Peng, and Fei Shi. Classification and segmentation of oct images for age-related macular degeneration based on dual guidance networks. *Biomedical Signal Processing and Control*, 84:104810, 2023.
- [28] Xuan Huang, Zhuang Ai, Hui Wang, Chongyang She, Jing Feng, Qihao Wei, Baohai Hao, Yong Tao, Yaping Lu, and Fanxin Zeng. Gabnet: global attention block for retinal oct disease classification. *Frontiers in Neuroscience*, 17:1143422, 2023.
- [29] Qinshuo Wang, Qing Guo, Xianchang Liu, and Ruichun Tang. Tri-branch cnn for age-related macular degeneration categorization with incomplete multi-modality ophthalmology images. In *2023 8th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, pages 436–442. IEEE, 2023.
- [30] Oluwatunmise Akinniyi, Md Mahmudur Rahman, Harpal Singh Sandhu, Ayman El-Baz, and Fahmi Khalifa. Multi-stage classification of retinal oct using multi-scale ensemble deep architecture. *Bioengineering*, 10(7):823, 2023.
- [31] Kuntha Pin, Jung Woo Han, and Yunyoung Nam. Retinal diseases classification based on hybrid ensemble deep learning and optical coherence tomography images. *Electronic Research Archive*, 31(8):4843–4861, 2023.
- [32] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers &

distillation through attention. In *International conference on machine learning*, pages 10347–10357. PMLR, 2021.



## **Abstract**

Timely and accurate identification of retinal and macular diseases is of great importance due to irreparable damage. AMD is one of these advanced eye diseases that destroys the central region of the visual network, leading to gradual vision loss and blindness. Early identification of this disease can lead to significant improvement in patient treatment. In recent years, with the remarkable advancements in image processing, numerous new models have been proposed for classification, which have shown remarkable performance in identifying this disease.

Therefore, in recent years, the VIT model has been used to diagnose AMD, and the results of this method seem to be much better compared to other methods such as convolutional neural networks and traditional deep neural networks. The application of VIT in the diagnosis of AMD shows that this novel approach can also be used in the field of eye disease diagnosis and can lead to a significant improvement in the accuracy of AMD diagnosis.

In this project, we proposed a method using the Stitchable neural network, which has achieved better accuracy on the NEH dataset without the need for additional data.



University of Tehran

University College of Science

School of Mathematics, Statistics, and  
Computer Science

# A study in Stitchable Neural Networks

Author name:

Setareh Abhari

Supervisor:

Hedieh Sajedi

A thesis submitted in partial fulfillment of the requirements for  
the degree of B.Sc. in Computer Science

Jul 2023