روش‌های یادگیری عمیق چند ماهیتی در قطعه‌بندی تصاویر پزشکی

فاطمه نارویی[۱]، احسان اسلامی[۲]

[۱]دانشجوی کارشناسی ارشد هوش مصنوعی، دانشگاه سیستان و بلوچستان، زاهدان، fatemehnarouie3@gmail.com

[۲] عضو هیئت‌علمی، مربی، گروه مهندسی کامپیوتر، دانشگاه ولایت،ایرانشهر، e.eslami@velayat.ac.ir

چکیده

یادگیری عمیق یک رویکرد قدرتمند برای آموزش و تجزیه‌وتحلیل داده‌های پیچیده پزشکی است. اخیراً روش‌های یادگیری عمیق در تحلیل‌های پزشکی بسیار رایج شده‌اند، زیرا به نتایج برجسته‌ای در این زمینه دست‌یافته‌اند. داده‌های چند ماهیتی عملکرد شبکه‌های عصبی را بهبود می‌بخشد.

پیشرفت در صنعت مراقبت‌های پزشکی منجر به حسگرهای تصویربرداری مختلف شده است که تصمیم‌گیری بالینی را بهبود بخشیده است. سناریوهای تشخیص معمولاً نیاز به دیدن اعماق و ساختارهای مختلف بدن انسان دارند که معمولاً تنها با یک روش دیده نمی‌شوند. بنابراین، اطلاعات از حسگرهای مختلف را می‌توان باهم ترکیب کرد تا تصویر جدیدی ساخته شود که اطلاعات تکمیلی بیشتری را به متخصص ارائه دهد. هم‌جوشی تصویر یک موضوع تحقیقاتی بسیار مهم در زمینه بینایی کامپیوتری و تجزیه‌وتحلیل تصویر است. با کاربردهای متعددی ازجمله سنجش‌ازدور [۱]، مادون‌قرمز [۲] و تشخیص بالینی [۳] سروکار دارد. درعین‌حال، نیاز روزافزون به همجوشی تصویر در سیستم‌های پزشکی کنونی، به‌عنوان‌مثال مداخلات هدایت‌شده با تصویر از راه پوست [۴]، روش‌های هدایت‌شده با تصویر [۵]، عمدتاً به دلیل رشد انواع فناوری‌های اکتسابی است. این مقاله مروری جامع از آخرین مطالعات در مورد آموزش و تجزیه‌وتحلیل داده‌های پزشکی چند ماهیتی با استفاده از الگوریتم‌های یادگیری عمیق ارائه می‌دهد.

واژه‌های کلیدی

یادگیری عمیق، داده‌های چند ماهیتی، قطعه‌بندی تصاویر پزشکی، هم‌جوشی تصاویر

مقدمه

همان‌طور که در شکل ۱ نشان داده شده است، تعداد مقالات علمی باهدف ترکیب تصاویر به‌سرعت از سال ۲۰۰۵ و به‌طور خاص در زمینه پزشکی افزایش‌یافته است. تعداد انتشارات در سال ۲۰۱۹ به حدود ۶۰۰ مقاله رسیده است[۶]. ترکیب تصاویر پزشکی چند ماهیتی شامل ترکیب دو یا چند تصویر از روش‌های مشابه یا متفاوت باهدف بهبود محتوای تصویر و حفظ اطلاعات است. پیشرفت سریع در تکنیک‌های تصویربرداری پزشکی توموگرافی کامپیوتری [۱]، گسیل پوزیترون [۲]، تصویربرداری تشدید مغناطیسی[3]، توموگرافی کامپیوتری با انتشار تک فوتون[۴] توجه محققین را به روش‌های همجوشی جلب کرده است[۷].

با استفاده از ترکیب، ادغام یا هم‌جوشی تصاویر می‌توان به نتایج مهمی دست‌یافت که به تقویت سایر وظایف مانند طبقه‌بندی، دسته‌بندی مطالب و تشخیص پزشکی، کمک شایانی می‌کند.

هدف اصلی این کار ارائه یک نمای کلی از روش‌های یادگیری عمیق چند ماهیتی و روش‌های همجوشی تصاویر پزشکی با پیشینه نظری و پیشرفت‌های اخیر است[۸]. هم‌جوشی تصاویر پزشکی در چند بخش موردبررسی قرارگرفته است. در بخش اول به فرایند یادگیری چند ماهیتی پرداخته خواهد شد. در بخش دوم در مورد مجموعه داده‌های چند ماهیتی که در حوزه‌ی پزشکی مورداستفاده قرارگرفته‌اند، صحبت خواهد شد و در بخش سوم روش‌های یادگیری عمیق موردبحث قرار خواهد گرفت در این بخش محبوب‌ترین معماری‌های یادگیری عمیق ارائه‌شده است که شامل CNN، GAN، RNN و شبکه عصبی است. هنگام مواجهه با داده‌های چند ماهیتی ، دانستن نحوه همجوشی اطلاعات از روش‌های مختلف ضروری است. درنتیجه، تکنیک‌های مختلف همجوشی در بخش چهارم توضیح داده‌شده‌اند، در این بخش خلاصه ایی از همجوشی تصاویر پزشکی و روش‌های آن و همچنین مزایا و معایب این روش‌ها موردبحث قرار خواهد گرفت و با چندین رویکرد در هر دسته بررسی می‌شود. درنهایت در بخش ۵ مقایسه روش‌های پیشرفته یادگیری عمیق چند ماهیتی ارائه خواهد شد.

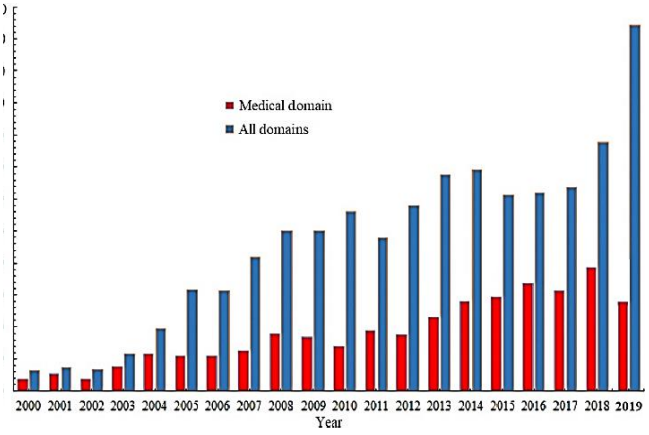
---

۱. CT.

۲. PET.

3 MRI

۴ SPECT



شکل 1: انتشارات علمی با موضوع همجوشی تصویر و تلفیق تصاویر پزشکی

۱. فرایند یادگیری چند ماهیتی

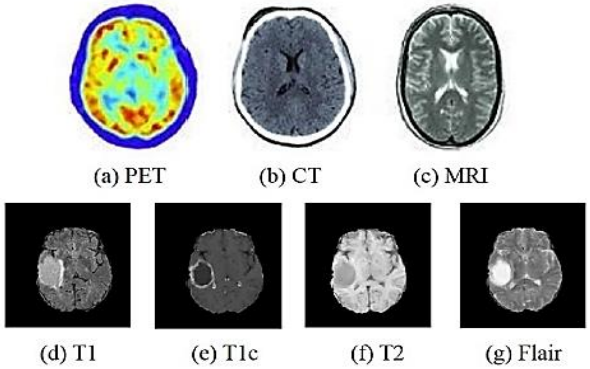
در مرحله اول تجزیه‌وتحلیل داده‌های پزشکی چند ماهیتی، محققان باید در مورد منابع داده، استراتژی ترکیب، استراتژی یادگیری و معماری یادگیری عمیق تصمیم‌گیری کنند. انتخاب ترکیب مناسبی از منابع داده در تحلیل‌های چند ماهیتی بسیار مهم است؛ زیرا ترکیب اشتباه منجر به عملکرد پایین‌تر می‌شود. منابع داده باید اطلاعات تکمیلی را برای بهبود نتایج ارائه دهند. گام بعدی تصمیم‌گیری در مورد نحوه همجوشی ماهیت‌های مختلف است. علاوه بر این، یک استراتژی یادگیری مناسب باید انتخاب شود. درنهایت، محققان باید یک معماری شبکه را انتخاب کنند. دانستن معماری‌های مختلف یادگیری عمیق به یافتن مناسب‌ترین معماری برای تحقیق کمک می‌کند. در بخش‌های بعدی این مفاهیم توضیح داده شده است[۱۱].

۱،۱. داده‌های تصویربرداری

دو دلیل وجود دارد که چرا یکپارچه‌سازی روش‌های تصویربرداری سودمند است. اولاً، همه روش‌های فردی محدودیت‌های خود را دارند. ثانیاً، یک بیماری، اختلال یا ضایعه ممکن است خود را به اشکال، علائم یا علل مختلف نشان دهد. از سوی دیگر، بیماری‌های مختلف ممکن است برخی علائم یا ظاهر مشترک داشته باشند؛ بنابراین، یک روش تصویر فردی ممکن است تصویر کاملی از یک بیماری را نشان ندهد[۱۲][3] . ازآنجایی‌که تصاویر پزشکی می‌توانند دوبعدی یا سه‌بعدی باشند، قبل از آموزش شبکه باید در مورد ابعاد تصاویر تصمیم‌گیری کنیم. رویکرد دوبعدی برش‌های تصویر استخراج‌شده از تصویر سه‌بعدی را می‌گیرد و آن‌ها را به شبکه تغذیه می‌کند. این رویکرد هزینه محاسباتی را کاهش می‌دهد، اما اطلاعات مکانی تصاویر را در جهت z نادیده می‌گیرد[۱۳]. همچنین می‌توانیم از تصاویر سه‌بعدی استفاده کرده و آن‌ها را مستقیماً به شبکه انتقال دهیم که این روش می‌تواند پرهزینه باشد، اما هیچ اطلاعاتی را از دست نمی‌دهد. در جدول ۱مجموعه داده‌های چند ماهیتی که در وظایف مختلف قطعه‌بندی تصویر استفاده‌شده ، ارائه‌شده است.

۱،۲. ماهیت داده‌ها

روش‌های اصلی در تجزیه‌وتحلیل تصاویر پزشکی، توموگرافی کامپیوتری (CT)، تصویربرداری تشدید مغناطیسی (MRI) و توموگرافی گسیل پوزیترون (PET) است. در مقایسه با تصاویر منفرد، تصاویر چندماهیتی به استخراج ویژگی‌ها از نماهای مختلف کمک می‌کنند و اطلاعات تکمیلی را به ارمغان می‌آورند و به نمایش بهتر داده‌ها و قدرت تمایز شبکه کمک می‌کنند. همان‌طور که در [۹] اشاره شده است، تصویر CT می‌تواند اختلالات عضلانی و استخوانی، مانند تومورهای استخوان و شکستگی‌ها را تشخیص دهد، درحالی‌که تصویر MRI می‌تواند کنتراست بافت نرم خوبی را بدون تشعشع ارائه دهد. تصاویر عملکردی، مانند PET، فاقد مشخصات تشریحی هستند، درحالی‌که می‌توانند کمی اطلاعات متابولیک و عملکردی در مورد بیماری‌ها ارائه دهند. روش MRI به دلیل وابستگی به پارامترهای اکتسابی متغیر، مانند تصاویر T1-weighted (T1)، T1-weighted (T1c)، T2-weighted (T2) و بازیابی وارونگی میرایی سیال (Flair) می‌تواند اطلاعات تکمیلی را ارائه دهد. T2 و Flair برای تشخیص تومور با ادم پری تومور مناسب هستند، درحالی‌که T1 و T1c برای تشخیص هسته تومور بدون ادم پری تومورال مناسب هستند. بنابراین، استفاده از تصاویر چند ماهیتی می‌تواند عدم قطعیت اطلاعات را کاهش دهد و تشخیص بالینی و دقت تقسیم‌بندی را بهبود بخشد [۱۰]. چندین تصویر پزشکی چند ماهیتی پرکاربرد در شکل ۲ توضیح داده شده است.



شکل ۲ تصاویر پزشکی چند ماهیتی، (a)-(c) تصاویر پزشکی چند ماهیتی رایج هستند و (d)-(g) توالی‌های مختلف MRI مغز هستند.

۱،۳. تقسیم‌بندی تصویر پزشکی چند ماهیتی

باتوجه‌به متغیر بودن اندازه، شکل و محل بافت هدف، تقسیم‌بندی تصویر پزشکی یکی از چالش‌برانگیزترین وظایف درزمینه تجزیه‌وتحلیل تصویر پزشکی است. علی‌رغم تنوع معماری‌های شبکه تقسیم‌بندی پیشنهادی، مقایسه عملکرد الگوریتم‌های مختلف هنوز دشوار است، زیرا بیشتر الگوریتم‌ها بر روی مجموعه‌های مختلف داده ارزیابی‌شده و در معیارهای مختلف گزارش می‌شوند.

تقسیم‌بندی تصویر پزشکی چند ماهیتی که بر اساس یادگیری عمیق، توصیف می‌شود از سه بخش تشکیل شده است: آماده‌سازی داده، معماری شبکه و استراتژی همجوشی.

در مرحله آماده‌سازی داده‌ها، ابتدا بعد داده انتخاب می‌شود و از پیش‌پردازش برای کاهش تنوع بین تصاویر استفاده می‌شود و همچنین می‌توان از استراتژی تقویت داده‌ها برای افزایش داده‌های آموزشی استفاده کرد تا از مشکل بیش برازش جلوگیری شود. در مراحل معماری شبکه و استراتژی همجوشی، استراتژی‌های همجوشی تصاویر چند ماهیتی اصلی و دقیق شبکه برای آموزش شبکه تقسیم‌بندی ارائه می‌شوند. در وظیفه تقسیم‌بندی تصویر پزشکی چند ماهیتی، همجوشی چندین روش مشکل کلیدی این کار است. باتوجه‌به سطح معماری شبکه که در آن همجوشی انجام می‌شود، استراتژی‌های سه‌گانه همجوشی، همجوشی در سطح ورودی، همجوشی در سطح لایه و همجوشی در سطح تصمیم که در مطالب بعدی موردبحث قرار خواهد گرفت ، قرار دارد.

۱،٤. پردازش داده‌ها

در این بخش پردازش داده‌ها ازجمله انتخاب ابعاد داده، پیش‌پردازش تصویر و تقویت داده‌ها را توضیح داده می‌شود. این مرحله در شبکه تقسیم‌بندی مبتنی بر یادگیری عمیق مهم است.

بعد داده

تصاویر پزشکی معمولاً با تصاویر سه‌بعدی سروکار دارد. برخی از مدل‌ها مستقیماً از تصاویر سه‌بعدی برای آموزش استفاده می‌کنند[۱۴]، درحالی‌که برخی از مدل‌ها تصویر سه‌بعدی را تکه‌تکه پردازش می‌کنند [۱۵]. رویکرد سه‌بعدی، تصویر سه‌بعدی را به‌عنوان ورودی می‌گیرد و هسته کانولوشن سه‌بعدی را برای بهره‌برداری از اطلاعات زمینه فضایی تصویر اعمال می‌کند. اشکال اصلی هزینه محاسباتی پرهزینه است. در مقایسه با استفاده از کل تصویر حجمی برای آموزش مدل، می‌توان از برخی تکه‌های کوچک سه‌بعدی برای کاهش هزینه محاسباتی استفاده کرد. به‌عنوان‌مثال، کامنیتزاس و همکارانش[16] . ده هزار تکه سه‌بعدی تصادفی را در فواصل زمانی منظم برای آموزش تقسیم ضایعه مغزی استخراج می‌کند. رویکرد دوبعدی، تکه یا وصله تصویر استخراج‌شده از تصویر سه‌بعدی را به‌عنوان ورودی می‌گیرد و هسته کانولوشنال دوبعدی را اعمال می‌کند، رویکرد دوبعدی می‌تواند هزینه محاسباتی را به‌طور مؤثر کاهش دهد، درحالی‌که اطلاعات مکانی نادیده می‌گیرد.

جدول 1 . خلاصه‌ای از مجموعه داده‌های تقسیم‌بندی تصویر پزشکی چند ماهیتی

ردیف	دیتاست	آموزش	آزمون	اعتبارسنجی	وظیفه	ماهیت
۱	Brats2012	35	15	N/A	تومور مغزی	T1, T1C, T2, Flair
۲	Brats2013	35	25	N/A	تومور مغزی	T1, T1C, T2, Flair
۳	Brats2014	200	38	N/A	تومور مغزی	T1, T1C, T2, Flair
۴	Brats2015	200	53	N/A	تومور مغزی	T1, T1C, T2, Flair
۵	Brats2016	200	191	N/A	تومور مغزی	T1, T1C, T2, Flair
۶	Brats2017	285	146	46	تومور مغزی	T1, T1C, T2, Flair
۷	Brats2018	285	191	66	تومور مغزی	T1, T1C, T2, Flair
۸	ISLES2015	28	36	N/A	ضایعه سکته مغزی ایسکمیک	T1, T2, TSE, T1c, T2, DWI, CBF, CBV, TTP, Tmax
۹	MRBrainS13	5	15	N/A	بافت مغز	T1, T1_1, mm, T1_IR, Flair
۱۰	NeoBrainS12	20	5	N/A	بافت مغز	T1, T2
۱۱	iSeg-2017	10	13	N/A	بافت مغز	T1,T2
۱۲	CHAOS	20	20	N/A	اندام‌های شکمی	CT, T1-DUAL, T2-SPIR
۱۳	IVD	16	8	N/A	دیسک بین مهره‌ای	In-phase, Opposed-phase, Fat, Water

پیش‌پردازش

پیش‌پردازش نقش مهمی در بخش‌بندی چندبعدی دارد، به‌ویژه برای قطعه‌بندی تصویر پزشکی چند ماهیتی زیرا شدت، کنتراست و نویز متفاوتی در تصاویر وجود دارد. بنابراین، برای شبیه‌تر جلوه دادن تصاویر و روان‌سازی و کمی سازی آموزش شبکه، برخی از تکنیک‌های پیش‌پردازش قبل از تغذیه به شبکه تقسیم‌بندی اعمال می‌شوند. تکنیک‌های پیش‌پردازش معمولی شامل ثبت تصویر، تصحیح میدان سوگیری و عادی‌سازی شدت است. برای مجموعه داده BraTS، ثبت تصویر قبلاً در اختیار عموم قرار گرفته است [17]از روش N4ITK برای اصلاح تحریف داده‌های MRI استفاده شد[18] .

تقویت داده‌ها

اغلب اوقات، تعداد زیاد و کافی برچسب برای آموزش به دلایل مختلف در دسترس نیست. برچسب‌گذاری مجموعه داده‌ها به یک متخصص در این زمینه نیاز دارد که پرهزینه و زمان‌بر است. هنگام آموزش شبکه‌های عصبی بزرگ از داده‌های آموزشی محدود، مشکل بیش برازش باید در نظر گرفته شود[۱۹]. افزایش داده‌ها راهی برای کاهش بیش برازش و افزایش حجم داده‌های آموزشی است. با تبدیل (چرخش، ترجمه، مقیاس، برگردان، تحریف و اضافه کردن مقداری نویز مانند نویز گاوسی) تصاویر جدیدی را در مجموعه داده‌های آموزشی ایجاد می‌کند. هم تصویر اصلی و هم تصاویر ایجادشده به شبکه عصبی تغذیه می‌شوند. به‌عنوان‌مثال، آیینسی و همکاران[۲۰] پیشنهاد کردند که با استفاده از انواع زیادی از تکنیک‌های افزایش داده‌ها مانند چرخش‌های تصادفی، مقیاس‌گذاری تصادفی، تغییر شکل‌های الاستیک تصادفی، تقویت تصحیح گاما و آینه‌سازی در حین تمرین، بیش برازش را برطرف کند.

۲. روش‌های مبتنی بر یادگیری عمیق

۲,۱. یادگیری عمیق

یادگیری عمیق به یک شبکه عصبی با چندین لایه از واحدهای پردازش غیرخطی اشاره دارد [21]. هر لایه متوالی از خروجی لایه قبلی به‌عنوان ورودی استفاده می‌کند. شبکه می‌تواند با استفاده از این لایه‌ها ویژگی‌های سلسله‌مراتبی پیچیده را از مقدار زیادی داده استخراج کند.

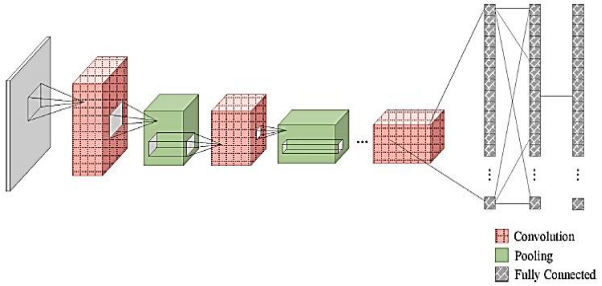
روش‌های یادگیری عمیق نوعی الگوریتم‌های یادگیری بازنمایی هستند که به ماشین اجازه می‌دهند تا به‌طور خودکار نمایش موردنیاز را از داده‌های ورودی خام کشف کند. در سال‌های اخیر، یادگیری عمیق پیشرفت‌های قابل‌توجهی در طبقه‌بندی تصاویر، تشخیص، تشخیص اشیا و تجزیه‌وتحلیل تصویر پزشکی ایجاد کرده است، آن‌ها نتایج عالی قابل‌مقایسه یا گاهی اوقات برتر از متخصصان انسانی تولید کرده‌اند. در میان الگوریتم‌های یادگیری عمیق شناخته‌شده، مانند رمزگذارهای خودکار پشته‌ای [۲۲]، ماشین‌های بولتزمن عمیق [۲۳]و شبکه‌های عصبی کانولوشنال[۲۴] ، موفق‌ترین الگوریتم برای تقسیم‌بندی تصویر، شبکه‌های عصبی کانولوشنال است.

۲,۲. معماری‌های مهم در یادگیری عمیق

این بخش برخی از محبوب‌ترین معماری‌های شبکه عصبی را معرفی می‌کند که تأثیر زیادی بر تحلیل‌های پزشکی گذاشته‌اند. شبکه‌های کاملاً متصل[۵] ، CNN و RNN عمدتاً برای کارهای تحت نظارت استفاده می‌شوند. از سوی دیگر، شبکه‌های عصبی بدون نظارت، ازجمله GAN ها، ماشین‌های محدودشده بولتزمن[۶] و رمزگذارهای خودکار می‌توانند در غیاب داده‌های برچسب‌دار استفاده شوند. علاوه بر این، ترکیبی از این تکنیک‌ها را می‌توان در کارهای نیمه نظارتی استفاده کرد [۱۱].

۲,۲,۱. CNN

معماری CNN شامل سه‌لایه اصلی است: لایه کانولوشن، لایه همجوشی و لایه کاملاً متصل. لایه کانولوشن از چندین فیلتر برای استخراج ویژگی‌های سطح بالا استفاده می‌کند. لایه همجوشی اندازه فضایی نقشه‌های ویژگی به‌دست‌آمده از لایه کانولوشن را کاهش می‌دهد که منجر به کاهش قدرت محاسباتی موردنیاز برای پردازش داده‌های تصویر می‌شود. همچنین باعث عدم‌تغییر ترجمه می‌شود که توانایی نادیده گرفتن ترجمه‌های هدف در ورودی است. معمولاً بعد از لایه‌های کانولوشن و تلفیقی، چندلایه کاملاً به هم متصل می‌شوند. اصلاحات مختلفی مانند فرمول‌بندی مجدد ساختاری، منظم سازی و بهینه‌سازی پارامترها در معماری CNN از سال ۱۹۸۹ تا به امروز انجام‌شده است[۲۵]. شکل ۳ نمونه‌ای از معماری پایه CNN را نشان می‌دهد. اگرچه CNNها عمدتاً برای تجزیه‌وتحلیل تصویر استفاده می‌شوند، برخی از مطالعات از آن‌ها برای تجزیه‌وتحلیل داده‌های متوالی استفاده کرده‌اند [26].



شکل 3 نمونه ایی از CNN

۲,۲,۲. RNN

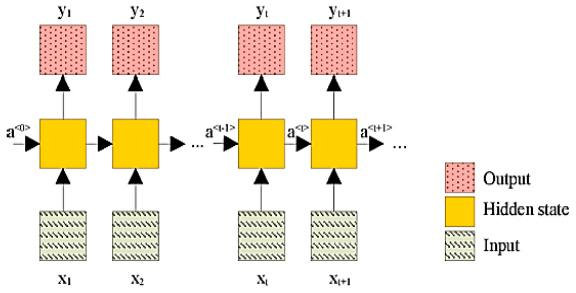
RNNها دسته‌ای از شبکه‌های عصبی هستند که با موفقیت داده‌های توالی را مدل می‌کنند[27]. داده‌های ترتیبی داده‌های مرتب‌شده‌ای هستند که در آن‌ها اطلاعات مرتبط به دنبال یکدیگر می‌آیند، مانند یک توالی DNA. فرض اصلی یک RNN سنتی این است که هر آیتم در یک‌سری ورودی را یکی پس از دیگری تجزیه کند و بردار "حالت پنهان" خود را در هر مرحله به‌روز کند. در پایان هر مرحله، این بردار پنهان یاد می‌گیرد که زمینه همه ورودی‌های قبلی را نشان دهد؛ بنابراین، زمانی که تصمیم می‌گیرد، ورودی فعلی و آنچه از ورودی‌های قبلی آموخته است را در نظر می‌گیرد. این‌یک مزیت مهم RNN است؛

---

[۵] FCN

[۶] RBM

زیرا مجموعه‌ای از داده‌ها حاوی اطلاعات مهمی از آینده است. مزیت دیگر RNN توانایی آن در پردازش ورودی توالی با طول متغیر است. همچنین، اندازه ورودی بر اندازه RNN تأثیر نمی‌گذارد. نمونه‌ای از یک RNN ساده در شکل ۴ نشان‌داده‌شده است. از سوی دیگر، RNN در گرفتن وابستگی‌های طولانی خیلی خوب نیست. یکی دیگر از معایب RNN اصلی، مشکل محو گرادیان است. واحد بازگشتی دروازه‌ای[۷] توسعه‌ای از RNN است که باعث می‌شود RNN اساسی در مواجهه با این دو مشکل بهتر شود.



شکل 4 نمونه ایی از RNN

۲،۲،۳. شبکه عصبی مبتنی بر توجه

مدل‌های توجه اخیراً در جامعه هوش مصنوعی به‌عنوان یک مؤلفه اساسی معماری عصبی بسیار محبوب شده‌اند[15]. شهود پشت مدل‌های توجه را می‌توان با استفاده از سیستم‌های بیولوژیکی انسان توضیح داد. سیستم پردازش بصری ما تمایل دارد به‌طور انتخابی بر روی برخی از بخش‌های تصویر تمرکز کند درحالی‌که اطلاعات نامربوط دیگر را نادیده می‌گیرد، به‌نحوی‌که می‌تواند به درک کمک کند[۲۸]. به همین ترتیب، در چندین مشکل مانند زبان و گفتار، برخی از بخش‌های ورودی اهمیت بیشتری نسبت به سایرین دارند. مکانیسم توجه به یک مدل اجازه می‌دهد تا به‌صورت پویا فقط به بخش‌های خاصی از ورودی توجه کند که به انجام کار به‌طور مؤثر کمک می‌کند. بخش‌های مهم داده‌ها بر اساس زمینه انتخاب می‌شوند و از طریق روش آموزشی با نزول گرادیان آموخته می‌شوند. یکی از دلایلی که مدل‌های توجه بسیار محبوب شده‌اند این است که تفسیرپذیری شبکه‌های عصبی را بهبود می‌بخشند که در غیر این صورت مدل‌های جعبه سیاه در نظر گرفته می‌شوند.

۲،۲،٤. GAN

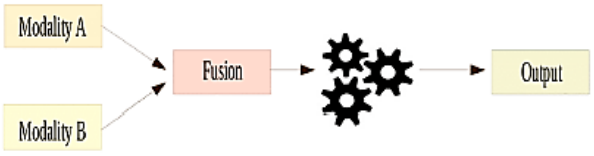
GANها الگوریتم‌های یادگیری بدون نظارت هستند که از تابع هزینه نظارت‌شده به‌عنوان بخشی از فرایند آموزشی خود استفاده می‌کنند. GANها شبکه‌های مولد و متمایز گر را تشکیل می‌دهند [29]. شبکه مولد از یک بردار نویز به‌عنوان ورودی خود برای تولید تصویر استفاده می‌کند، سپس این تصویر تولیدشده به شبکه متمایز گر داده می‌شود. شبکه متمایز گر سعی می‌کند بین تصویر واقعی و تولیدشده تمایز قائل شود. به‌عبارت‌دیگر، متمایز گر طبقه‌بندی‌کننده‌ای است که تعیین می‌کند کدام تصویر واقعی و کدام‌یک جعلی است. در فرایند آموزش، هدف مولد ایجاد تصاویر واقعی است که متمایز گر را گیج می‌کند. بااین‌حال، مهم است که این شبکه‌ها را در یک سطح نگه‌دارید و آن‌ها را باهم بهبود بخشید. در سال‌های اخیر، GANها در تجزیه‌وتحلیل‌های پزشکی رایج شده‌اند و برای کارهای مختلفی مانند تقویت داده‌ها و ترجمه تصویر استفاده می‌شوند. یکی از کاربردهای GAN برای ترجمه تصویر CycleGAN است که امکان تبدیل تصویر از یک دامنه به دامنه دیگر را فراهم می‌کند.

۳. ساختارهای همجوشی

وقتی با داده‌های چند ماهیتی کار می‌کنیم، باید تصمیم بگیریم که چگونه آن‌ها را یکپارچه کنیم. سه‌راه برای ترکیب ماهیت‌های مختلف وجود دارد: همجوشی در سطح ورودی، همجوشی در سطح لایه و همجوشی در سطح تصمیم.

۳،۱. همجوشی در سطح ورودی

این روش همچنین به‌عنوان یکپارچه‌سازی اولیه، یکپارچه‌سازی مبتنی بر ویژگی و یکپارچه‌سازی مبتنی بر داده شناخته می‌شود. در این روش، روش‌های مختلف قبل از انجام تجزیه‌وتحلیل، همان‌طور که در شکل ۵ نشان‌داده‌شده است [۱۱]، ترکیب می‌شوند. یکی از مزایای همجوشی در سطح ورودی این است که رابطه بین روش‌های مختلف را پیدا می‌کند. بااین‌حال، برای یافتن این رابطه، همه روش‌ها باید برای هر نمونه در مجموعه آموزشی موجود باشد که در عمل به‌سختی ارضا می‌شود. یکی دیگر از معایب این روش این است که منجر به یک بردار ویژگی بسیار بزرگ می‌شود که باعث هزینه محاسباتی بالایی می‌شود.

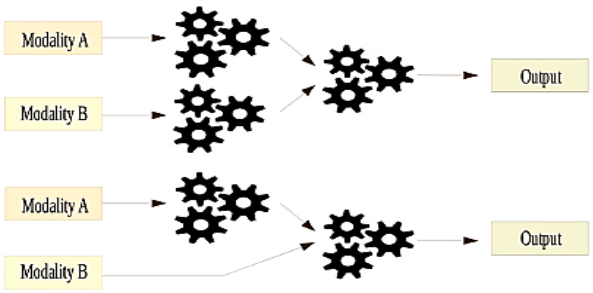


شکل 5. تصویر راهبردهای ترکیبی مختلف برای یادگیری چند ماهیتی. (الف) همجوشی در سطح ورودی

[۷] GRU

۳،۲. همجوشی در سطح لایه

این روش به همجوشی میانی و همجوشی مبتنی بر تحول نیز معروف است. در این روش، یک یا چند ماهیت به‌طور مستقل به یک شبکه داده می‌شود، سپس نمایش‌های میانی آن‌ها در لایه‌ای از شبکه ترکیب می‌شوند. مانند همجوشی در سطح ورودی، این روش رابطه بین ماهیت-های مختلف را پیدا می‌کند و به همه ماهیت‌ها برای هر نمونه در مجموعه آموزشی نیاز دارد. شکل ۶ این روش را نشان می‌دهد.

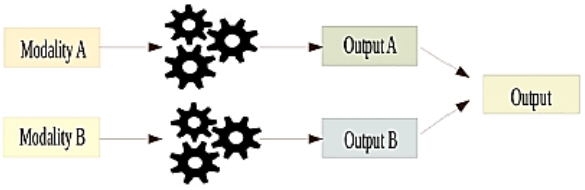


شکل 6. همجوشی در سطح لایه

۳،۳. همجوشی در سطح تصمیم

همجوشی دیررس و همجوشی مبتنی بر مدل نام‌های دیگر این روش است. در این روش، هر ماهیت به‌عنوان یک ورودی واحد برای آموزش یک شبکه عصبی استفاده می‌شود، سپس خروجی‌های مدل‌ها برای تصمیم‌گیری نهایی ترکیب می‌شوند (شکل۷).

این روش رایج است؛ زیرا برای هر نمونه به همه ماهیت‌ها نیاز ندارد. برخلاف سایر تکنیک‌های همجوشی، این تکنیک نمی‌تواند رابطه بین روش‌های مختلف را پیدا کند. مدل‌هایی که از همجوشی در سطح تصمیم استفاده می‌کنند، می‌توانند به عملکرد بالایی دست یابند؛ زیرا فضای جستجو نسبت به سایر تکنیک‌های همجوشی کوچک‌تر است. در این تکنیک، اطلاعات به‌طور مستقل از روش‌های مختلف آموخته می‌شود، بنابراین احتمال بیش برازش کمتر از روش‌های دیگر است [30]. محبوب‌ترین تکنیک مبتنی بر رأی اکثریت است که در آن پس از آموزش هر مدل به‌طور جداگانه، پیش‌بینی نهایی بر اساس اکثر پیش‌بینی‌های شبکه‌های فردی انتخاب می‌شود [31].



شکل 7 . تصویری از استراتژی‌های ترکیبی مختلف برای یادگیری چند ماهیتی،ج) همجوشی در سطح تصمیم.

٤. مقایسه روش‌های ارائه شده:

یادگیری عمیق، از طریق معماری‌ها و مفاهیم مختلف خود، به پیشرفت‌های مداوم در وظایف بینایی دست می‌یابد. علی‌رغم کمبود داده، CNNها در مقایسه با سایر مدل‌های تحت نظارت و بدون نظارت، مناسب‌ترین معماری‌ها برای همجوشی تصویر گزارش شده‌اند. در بخش‌های بالا، مجموعه بزرگی از پیشرفته‌ترین شبکه‌های بخش‌بندی تصویر پزشکی چند ماهیتی مبتنی بر یادگیری عمیق ارائه شد. مقایسه آن‌ها در جدول ۲ خلاصه شده‌اند. برای چالش BraTS، این روش‌ها از سال ۲۰۱۳ به نتیجه رسیده‌اند، زیرا روش‌های یادگیری عمیق از سال ۲۰۱۳ به کار گرفته شده‌اند. مجموعه داده‌های تصویر پزشکی چند ماهیتی در دسترس عموم برای کار تقسیم‌بندی نادر هستند، بیشترین استفاده از مجموعه داده‌های BraTS است. برای تقسیم‌بندی آن‌ها، بهترین روش فعلی در Ref ارائه شده است. آن‌ها از استراتژی همجوشی سطح ورودی برای همجوشی مستقیم ماهیت‌های مختلف در فضای ورودی استفاده می‌کنند، آن‌ها ساختار رمزگذار-رمزگشا CNN را به همراه یک شاخه VAE اضافی (تغییر خودکار) به قسمت رمزگذار اعمال می‌کنند. شاخه VAE می‌تواند تصویر ورودی را بازسازی کند و از ویژگی‌های نقطه پایانی رمزگذار بهتر بهره‌برداری کند. همچنین یک راهنمایی اضافی و یک نظم‌بخشی به بخش رمزگذار ارائه می‌دهد. استراتژی همجوشی تصویر مرسوم، نگاشت مستقیم بین تصاویر منبع و تصاویر هدف را می‌آموزد، استراتژی تلفیق شامل دو مرحله اساسی است: اندازه‌گیری سطح فعالیت و قانون ادغام اندازه‌گیری سطح فعالیت با طراحی فیلترهای محلی برای استخراج جزئیات با فرکانس بالا و اطلاعات شفافیت محاسبه‌شده اجرا می‌شود.

سپس تصاویر منبع مختلف با استفاده از قوانین طراحی‌شده برای به دست آوردن وضوح تصویر مقایسه می‌شوند. برای دستیابی به عملکرد بهتر، این مسائل پیچیده‌تر و پیچیده‌تر می‌شوند، بنابراین پیشنهاد دستی یک استراتژی همجوشی ایده آل که به‌طور کامل به مسائل مهم مربوط می‌شود، دشوار است. برای این منظور، یک شبکه مبتنی بر یادگیری عمیق می‌تواند به‌طور مستقیم نقشه‌برداری را رمزگذاری کند. به‌این‌ترتیب برخی از معماری‌های شبکه یادگیری عمیق بالقوه را می‌توان برای طراحی یک استراتژی ترکیبی تصویر مؤثر بررسی کرد.

جدول ۲: مقایسه روش‌های پیشرفته قطعه‌بندی تصاویر پزشکی چند ماهیتی

مقاله	پیش‌پردازش	داده	شبکه	سطح همجوشی	نتیجه
[49]	Normalization Bias Field Correction	سه‌بعدی	CNN CRF	ورودی	whole/core/enhanced 0.84/0.66/0.63
[20]	Normalization Bias Field Correction	دوبعدی	CNN	ورودی	whole/core/enhanced 0.84/0.71/0.57
[41]	Normalization Data Augmentation	سه‌بعدی	U-Net ResNet	ورودی	whole/core/enhanced 0.85/0.74/0.64
[50]	Normalization Bias Field Correction	سه‌بعدی	FCN RNN ،CRF	ورودی	whole/core/enhanced 0.86/0.73/0.62
[57]	Normalization	سه‌بعدی	U-Net ResNet	ورودی	whole/core/enhanced 0.87/0.75/0.64
[48]	Normalization Bias Field Correction	دوبعدی	U-Net ResNet	ورودی	whole/core/enhanced 0.87/0.77/0.78
[44]	N/A	سه‌بعدی	CNN DenseNet	لایه	CSF/WM/GM 0.95/0.91/0.90t
[78]	Normalization	سه‌بعدی	DenseNet	لایه	CSF/WM/GM 0.96/0.91/0.90
[46]	N/A	دوبعدی	DenseNet ، U-Net	لایه	0.9191 +- 0.0179
[47]	N/A	دوبعدی	FCN	تصمیم	CSF/WM/GM 0.85/0.88/0.87

۵. نتیجه‌گیری

اخیراً روش‌های یادگیری عمیق چند ماهیتی در قطعه‌بندی تصویر به نتایج قابل‌توجهی دست‌یافته‌اند، زیرا داده‌های چند ماهیتی عملکرد شبکه را بهبود می‌بخشد و اطلاعات تکمیلی مفیدی ارائه می‌دهند. این مقاله مروری جامع از آخرین مطالعات در مورد تجزیه‌وتحلیل داده‌های پزشکی چند ماهیتی با استفاده از الگوریتم‌های یادگیری عمیق ارائه می‌دهد. روش‌های مختلف، معماری‌های یادگیری عمیق و استراتژی‌های تلفیقی نیز در این مقاله معرفی شده‌اند. مشکلات رایج و چالش‌های باز در این زمینه شناسایی‌شده‌اند. ما معتقدیم که روش‌های یادگیری عمیق در تجزیه‌وتحلیل داده‌های پزشکی چند ماهیتی در سال‌های آینده یک حوزه تحقیقاتی فعال باقی خواهد ماند. یادگیری عمیق، از طریق معماری‌ها و مفاهیم مختلف خود، به پیشرفت‌های مداوم در وظایف بینایی دست می‌یابد. درواقع، نمایش‌های سلسله‌مراتبی برای همجوشی تصویر پزشکی کارآمد به نظر می‌رسد. CNNها در مقایسه با سایر مدل‌های تحت نظارت و بدون نظارت، مناسب‌ترین معماری‌ها برای همجوشی تصویر گزارش شده‌اند. بااین‌حال، به عملکرد محاسباتی بالا و حجم زیادی از داده‌های آموزشی نیاز دارد. ذکر این نکته مهم است که همجوشی تصویر معمولاً می‌تواند عملکرد طبقه‌بندی و تقسیم‌بندی را از طریق اطلاعات تکمیلی افزایش دهد.

۶. مراجع و منابع

1. Zhou, H., et al., *Unified gradient-and intensity-discriminator generative adversarial network for image fusion.* Information Fusion, 2022. **88**: p. 184-201.
2. Ma, J., Y. Ma, and C. Li, *Infrared and visible image fusion methods and applications: A survey.* Information Fusion, 2019. **45**: p. 153-178.
3. Zhang, Y.-D., et al., *Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation.* Information Fusion, 2020. **64**: p. 149-187.
4. Solbiati, M., et al., *Thermal ablation of liver tumors guided by augmented reality: an initial clinical experience.* Cancers, 2022. **14**(5): p. 1312.
5. Huber, T.C., et al., *Percutaneous therapies for hepatocellular carcinoma: evolution of liver directed therapies.* Journal of Hepatocellular Carcinoma, 2021: p. 1181-1193.
6. Hermessi, H., O. Mourali, and E. Zagrouba, *Multimodal medical image fusion review: Theoretical background and recent advances.* Signal Processing, 2021. **183**: p. 108036.
7. Azam, M.A., et al., *A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics.* Computers in biology and medicine, 2022. **144**: p. 105253.
8. Li, J. and Q. Wang, *Multi-modal bioelectrical signal fusion analysis based on different acquisition devices and scene settings: Overview, challenges, and novel orientation.* Information Fusion, 2022. **79**: p. 229-247.
9. Bhatnagar, G., Q.J. Wu, and Z. Liu, *A new contrast based multimodal medical image fusion framework.* Neurocomputing, 2015. **157**: p. 143-152.

10. Guo, Z., et al., *Deep learning-based image segmentation on multimodal medical imaging.* IEEE Transactions on Radiation and Plasma Medical Sciences, 2019. **3**(2): p. 162-169.
11. Behrad, F. and M.S. Abadeh, *An overview of deep learning methods for multimodal medical data mining.* Expert Systems with Applications, 2022: p. 117006.
12. Zhu, J., et al., *A Heterogeneously Integrated Spiking Neuron Array for Multimode-Fused Perception and Object Classification.* Advanced Materials, 2022. **34**(24): p. 2200481.
13. Qaidi, S.M., et al., *Recycling of mine tailings for the geopolymers production: A systematic review*. Case Studies in Construction Materials, 2022: p. e00933.
14. Sun, Y., et al., *Multi-site infant brain segmentation algorithms: the iSeg-2019 challenge.* IEEE Transactions on Medical Imaging, 2021. **40**(5): p. 1363-1376.
15. Alzubaidi, L., et al., *Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions.* Journal of big Data, 2021. **8**: p. 1-74.
16. Kamnitsas, K., et al., *Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation.* Medical image analysis, 2017. **36**: p. 61-78.
17. Jiang, Z., et al. *Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task*. in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part I 5*. 2020. Springer.
18. Chaitanya, K., et al., *Contrastive learning of global and local features for medical image segmentation with limited annotations.* Advances in Neural Information Processing Systems, 2020. **33**: p. 12546-12558.
19. Pierson, E., et al., *An algorithmic approach to reducing unexplained pain disparities in underserved populations.* Nature Medicine, 2021. **27**(1): p. 136-140.
20. Singh, S.P., et al., *3D deep learning on medical images: a review.* Sensors, 2020. **20**(18): p. 5097.
21. Greener, J.G., et al., *A guide to machine learning for biologists.* Nature Reviews Molecular Cell Biology, 2022. **23**(1): p. 40-55.
22. Ren, P., et al., *A survey of deep active learning.* ACM computing surveys (CSUR), 2021. **54**(9): p. 1-40.
23. Sharma, R., et al., *A systematic literature review on machine learning applications for sustainable agriculture supply chain performance.* Computers & Operations Research, 2020. **119**: p. 104926.
24. Khan, S., et al., *Transformers in vision: A survey.* ACM computing surveys (CSUR), 2022. **54**(10s): p. 1-41.
25. Zhou, Y., et al., *Radiomic Features of 18 F-FDG PET in Hodgkin Lymphoma Are Predictive of Outcomes.* Contrast media & molecular imaging, 2021. **2021**.
26. Liu, P., et al., *Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing.* ACM Computing Surveys, 2023. **55**(9): p. 1-35.
27. Zhang, Y. and L. Zhou, *Novel global polynomial stability criteria of impulsive complex-valued neural networks with multi-proportional delays.* Neural Computing and Applications, 2022: p. 1-12.
28. Chaudhari, S., et al., *An attentive survey of attention models.* ACM Transactions on Intelligent Systems and Technology (TIST), 2021. **12**(5): p. 1-32.
29. Goodfellow, I., et al., *Generative adversarial networks.* Communications of the ACM, 2020. **63**(11): p. 139-144.
30. Chakraborty, K., et al., *Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media.* Applied Soft Computing, 2020. **97**: p. 106754.
31. Shikalgar, A. and S. Sonavane. *Hybrid deep learning approach for classifying alzheimer disease based on multimodal data*. in *Computing in Engineering and Technology: Proceedings of ICCET 2019*. 2020. Springer.