



بکارگیری بینایی کامپیوتر (CV) و یادگیری ماشین (ML) در حوزه سلامت و پزشکی

محمد حسین صادقی راد

کارشناس واحد هوش مصنوعی شرکت NOICT

آدرس پست الکترونیک نویسنده (hsadeghierad4@gmail.com)

امیر حسین صدری

مدیر عامل شرکت NOICT

علی زمان

مدیر دپارتمان اینترنت اشیا شرکت NOICT

چکیده

تاکنون صنعت بهداشت از فناوری هوش مصنوعی استفاده بسیاری کرده است. دو حوزه تخصصی و در حال رشد هوش مصنوعی، بینایی کامپیوتر و یادگیری ماشین، در پزشکی است که پتانسیل این را دارد که یکی از کاربردی ترین فناوری ها در جهت نجات جان بیماران باشد. با بینایی ماشین، سیستم های کامپیوتری قادر به درک روندها و کشف ناهنجاری ها در تصاویر هستند. کاربردهای آن در مراقبت های پزشکی روز به روز در حال افزایش است. این فناوری اکنون به بخشی از فرآیندهای معمول در پزشکی تبدیل شده است که برای تشخیص و درمان بیماری ها و یا شرایط خاص استفاده می شود. در دهه ۲۰۰۰ بود که انقلابی در صنعت پزشکی و هوش مصنوعی رخ داد و استفاده از ماشین های یادگیری در زمینه تشخیص بیماری ها از طریق تصاویر پزشکی (مانند MRI و CT scan) افزایش پیدا کرد. این پیشرفت ها سال به سال ادامه داشت و در حال حاضر از تلفیق هوش مصنوعی و پزشکی در تحقیقات بسیاری در حوزه توسعه دارو، تشخیص بیماری ها و مراقبت های بهداشتی و درمانی به کار گرفته می شود. در این مقاله، ابتدا توضیحی در مورد دو مفهوم پردازش تصویر (Image Processing) و بینایی کامپیوتر (Computer Vision) ارائه می شود. در ادامه یادگیری ماشین یا ML (Machine Learning) و یادگیری عمیق (Deep Learning) مرور می گردد. کاربردهای بینایی کامپیوتر در حوزه پزشکی و کاربردهای یادگیری عمیق در حوزه پزشکی به عنوان دو بخش اصلی مقاله در بخش بعدی بررسی و ارائه می شوند.

واژگان کلیدی: بینایی کامپیوتر، یادگیری ماشین، پزشکی، سلامت

۱- مقدمه

بیش از یک سوم مغز انسان به پردازش تصاویر دریافتی از چشم انسان یا همان بینایی اختصاص دارد. بینایی کامپیوتر در دنیای امروز کاربردهای متنوعی از تحلیل تصاویر پزشکی گرفته تا خودروهای خودران دارد. در طی یک دهه آینده، به نظر میرسد که استفاده هر چه بیشتر یادگیری ماشین در طراحی سیستم های دارای تعامل با انسان مشاهده خواهد شد. یادگیری ماشین به این معنا است که ماشین بتواند برنامه، ساختار یا داده هایش را بر اساس ورودیها یا در پاسخ به اطلاعات خارجی، به نحوی تغییر دهد که رفتارشان به آن چه از او انتظار می رود نزدیکتر شود، به عبارت دیگر می توان گفت یعنی قدرت تجزیه و تحلیل داشته باشد. یادگیری ماشین در پزشکی مورد استفاده قرار گرفته است که این موضوع در این مقاله مرور می شود. مفهوم دیگر که در این مقاله مطرح است یادگیری عمیق می باشد. یادگیری عمیق با نتایج تجربی موفق و کاربردهای گسترده، پتانسیل تغییر آینده علم پزشکی را داراست. امروزه استفاده از هوش مصنوعی به طور فزاینده ای رایج شده است و در رشته های مختلفی همچون تشخیص سرطان مورد استفاده قرار می گیرد. یادگیری عمیق همچنین دید رایانه ای، تصویر برداری و تشخیص پزشکی دقیق تر را امکان پذیر می کند. بنابراین تعجبی ندارد که در گزارشی از Report Linker اشاره شده است که انتظار می رود بازار هوش مصنوعی در صنعت پزشکی از ۲/۱ میلیارد دلار در سال ۲۰۱۸ به ۳۶ میلیارد دلار در سال ۲۰۲۵ برسد! [۲۱] در این مقاله، ابتدا توضیحی در مورد دو مفهوم پردازش تصویر (Image Processing) و بینایی کامپیوتر (Computer Vision) ارائه می شود. در ادامه یادگیری ماشین یا ML (Machine Learning) به اختصار تشریح می شود. سپس یادگیری عمیق (Deep Learning) مرور می گردد. دو موضوع اصلی مقاله یعنی کاربردهای بینایی کامپیوتر در حوزه پزشکی و کاربردهای یادگیری عمیق در حوزه پزشکی در ادامه بررسی و ارائه می شوند.

۲- پردازش تصویر (Image Processing) و بینایی کامپیوتر (Computer Vision)

۲-الف) پردازش تصویر (Image Processing)

پردازش تصویر (Image processing) امروزه بیشتر به موضوع پردازش دیجیتالی تصاویر (پردازش تصویر دیجیتال) گفته می شود که شاخه ای از پردازش سیگنال است که با پردازش سیگنال دیجیتال که نماینده تصاویر برداشته شده با دوربین دیجیتال یا اسکن شده توسط اسکنر هستند سر و کار دارد. تصاویر دیجیتالی در قالب ماتریس های عددی دو بعدی و سه بعدی ذخیره می شوند که هر یک از اعداد موجود در ماتریس، پیکسل های تصاویر را نشان می دهند. پردازش تصویر یعنی کارهایی که باید انجام دهید تا رایانه تصاویر و محتوای ویدیویی را بتواند بهتر ببیند و تفسیر کند.

روش های مختلفی برای پردازش تصویر وجود دارد و طبقه بندی تصویر یکی از سه روش اصلی در این زمینه است:

طبقه بندی تصویر یا Image Classification

طبقه بندی تصویر سؤالی که پاسخ می دهد این است که نوع تصویر چیست و خروجی آن یک کلاس است؛ برای مثال، «سگ» یا «گربه».

دو مثال جهت طبقه بندی تصویر عبارتند از [۱۱]:

۱. سیستم تشخیص / طبقه بندی خودکار بیماری های پوستی رنگی با استفاده از تصاویر درماتوسکوپی:



شکل ۱: سیستم تشخیص بیماری پوستی

۲. بندی بیماران به عنوان افراد مبتلا به COVID-19 یا نرمال (سالم) بر اساس یک مجموعه از تصاویر اشعه ایکس قفسه سینه بیماران COVID-19.

البته دو مثال بالا طبقه بندی دودویی می باشد. طبقه بندی دودویی یا Binary Classification هست. همان طور که از نام آن پیداست در این دسته از مسائل، فقط دو طبقه (دو کلاس) موجود است و هر کدام از نمونه ها به یکی از کلاس ها تعلق دارند.

تشخیص اشیا یا Object Detection

سؤالی که در این روش پاسخ داده می شود این است که جسم کجاست؟ این مدل مختصات یک به اصطلاح «جعبه محدود کننده» یا همان bounding box را در اطراف جسم موجود در تصویر خروجی می دهد. اگر از قبل می دانید که به دنبال سگ حاضر در تصویر هستید و می خواهید بدانید که آیا تصویر خاص یک سگ را نشان می دهد یا نه و اگر بله، کجای تصویر است، باید از object detection استفاده کنید.

بخش بندی تصویر یا Image Segmentation

سؤالی که در این روش پاسخ داده می شود این است که شکل جسم چگونه است؟ مدل برای هر شیء در تصویر یک پوشش پیکسلی (pixel-wise mask) ایجاد می کند. بخش بندی تصویر اطلاعات دقیق تری در مورد اندازه و شکل جسم می دهد. اگرچه این مدل ها از نظر محاسباتی گران تر هستند، اغلب برای بهبود کارایی سیستم استفاده می شوند؛ به این شکل که الگوریتم ها می توانند فقط به پردازش بخش های مرتبط تصویر بپردازند؛ برای مثال، شناسایی چهره تلفن های هوشمند فقط باید خطوط و اشکال را در شکل صورت تجزیه و تحلیل کند، نه پس زمینه را.

Object detection vs. image segmentation vs. image classification

Object detection



Image segmentation



Image classification

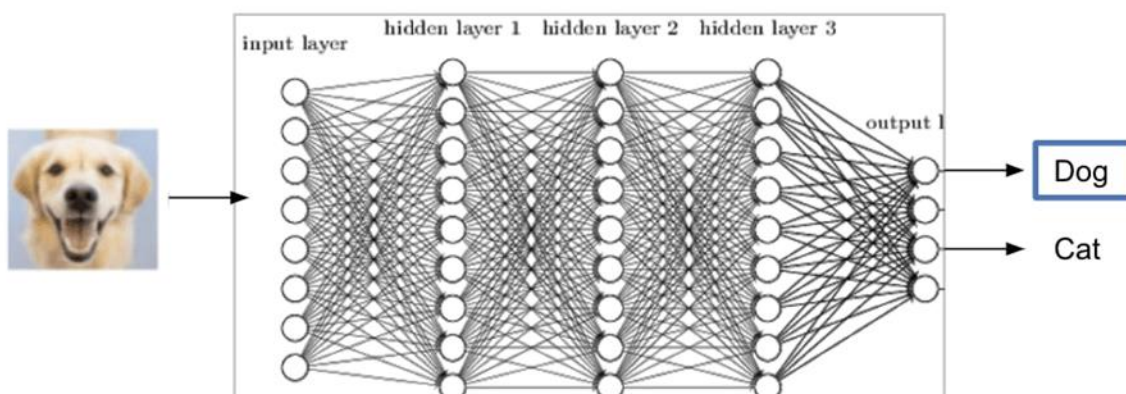


شکل ۲: روش های مختلفی برای پردازش تصویر [۱]

1. Input picture

2. Analysis through deep neural network

3. Output class



شکل ۳: طبقه بندی تصویر یا Image Classification با CNN [۱]

۲-ب) بینایی کامپیوتری (Computer Vision)

بینایی کامپیوتری (Computer Vision) حوزه‌ای از هوش مصنوعی است که کامپیوترها و سیستم‌ها را قادر می‌کند اطلاعات معنی‌داری را از تصاویر دیجیتال و فیلم‌ها و دیگر ورودی‌های بصری استخراج کنند و براساس آن اطلاعات، اقداماتی انجام دهند یا توصیه‌هایی ارائه کنند. بینایی کامپیوتری زیررشته‌ی هوش مصنوعی (Artificial Intelligence) و یادگیری ماشین (Machine Learning) است که فناوری‌ها و ابزارهایی را بررسی می‌کند که امکان آموزش کامپیوترها برای درک و تفسیر اطلاعات بصری از دنیای واقعی را فراهم می‌کند. دیدن جهان، بخش آسان کار است. برای این کار فقط به یک دوربین نیاز است؛ باین‌حال برای اینکه یک ماشین بتواند ببیند، اتصال ساده‌ی دوربین به آن کافی نیست؛ بخش چالش‌برانگیز کار طبقه‌بندی و تفسیر اشیاء در تصاویر و فیلم‌ها، رابطه‌ی میان آن‌ها و مفهوم اصلی آن است. بینایی کامپیوتری حوزه‌ای نوآورانه است که از آخرین فناوری‌های یادگیری ماشین برای ساخت سیستم‌های نرم‌افزاری استفاده می‌کند تا به انسان در زمینه‌های مختلف کمک کند. از خرده‌فروشی گرفته تا حفاظت از حیات وحش، این الگوریتم‌های هوشمند مسائل طبقه‌بندی تصاویر و تشخیص الگو را به راحتی در حد دقت انسانی حل می‌کنند و حتی گاهی بهتر از انسان این کار را انجام می‌دهند. این به معنی آن است که کامپیوتر باید به طور مؤثر این سه کار را انجام دهد:

- a) به طور خودکار بفهمد اشیای موجود در تصویر چه هستند و در کجا قرار دارند؛
- b) این اشیاء را دسته‌بندی و روابط میان آن‌ها را درک کند؛
- c) مفهوم عکس یا فیلم را درک کند.

به عبارت دیگر، هدف کلی در این حوزه این است که اطمینان حاصل شود رایانه یک تصویر را به اندازه‌ی انسان یا بهتر از او درک می‌کند. [۲]

۲-پ) تفاوت بینایی کامپیوتری و پردازش تصویر

بینایی کامپیوتری و پردازش تصویر دو حوزه‌ی نزدیک به هم هستند که شباهت‌هایی با هم دارند، اما یکسان نیستند.

- ❖ پردازش تصویر به تکنیک‌های مورد استفاده برای دستکاری و بهبود تصاویر دیجیتال اشاره دارد. این تکنیک‌ها شامل مواردی مانند فیلتر کردن و تشخیص لبه است. پردازش تصویر در طیف گسترده‌ای از برنامه‌ها از جمله عکاسی دیجیتال، تصویربرداری پزشکی و فشرده سازی ویدئو استفاده می‌شود.
- ❖ از سوی دیگر، بینایی کامپیوتری حوزه وسیع‌تری است که نه تنها شامل پردازش تصویر، بلکه استفاده از الگوریتم‌ها و تکنیک‌های یادگیری ماشین برای تفسیر و درک داده‌های بصری است. سیستم‌های بینایی کامپیوتری به گونه‌ای طراحی شده‌اند که روشی را که انسان‌ها از دنیای اطراف خود می‌بینند و درک می‌کنند، تقلید می‌کنند. این نه تنها شامل شناخت الگوها و اشکال در تصاویر، بلکه درک زمینه‌ای است که آن تصاویر در آن ارائه می‌شوند.
- ❖ در حالی که پردازش تصویر بر بهبود و دستکاری تصاویر متمرکز است، بینایی کامپیوتر بر استخراج معنا و اطلاعات از آن تصاویر متمرکز است. به عنوان مثال، پردازش تصویر ممکن است برای حذف نویز از یک تصویر استفاده شود، در حالی که بینایی کامپیوتر ممکن است برای تشخیص اشیاء در آن تصویر و ردیابی حرکات آن‌ها در طول زمان استفاده شود. [۲]

۲-ت) تفاوت بینایی ماشین و بینایی کامپیوتری

بینایی ماشین و بینایی کامپیوتر موضوعات مرتبطی هستند که شباهت‌هایی با هم دارند، اما دقیقاً یکسان نیستند.

- ❖ بینایی ماشین به استفاده از دوربین‌ها، حسگرها و سایر دستگاه‌های سخت‌افزاری برای ضبط و تجزیه و تحلیل داده‌های بصری اشاره دارد. سیستم‌های بینایی ماشین معمولاً در محیط‌های تولیدی و صنعتی برای خودکارسازی فرآیندهایی مانند کنترل کیفیت، بازرسی و مونتاژ استفاده می‌شوند.
- ❖ از سوی دیگر، بینایی کامپیوتر حوزه وسیع‌تری است که نه تنها استفاده از دستگاه‌های سخت‌افزاری برای گرفتن داده‌های بصری، بلکه استفاده از الگوریتم‌ها و تکنیک‌های یادگیری ماشین برای تفسیر و درک آن داده‌ها را نیز در بر

می‌گیرد. سیستم‌های بینایی کامپیوتری در طیف گسترده‌ای از کاربردها، از جمله تصویربرداری پزشکی، رباتیک و وسایل نقلیه خودران استفاده می‌شوند.

❖ در حالی که سیستم‌های بینایی ماشین برای انجام وظایف خاص، مانند شناسایی عیوب در یک محصول در خط مونتاژ طراحی شده‌اند، سیستم‌های بینایی کامپیوتری به گونه‌ای طراحی شده‌اند که انعطاف‌پذیرتر و سازگارتر باشند. سیستم‌های بینایی کامپیوتری را می‌توان برای تشخیص و شناسایی طیف وسیعی از اشیاء و الگوها آموزش داد و می‌توان آن‌ها را طوری برنامه‌ریزی کرد که با شرایط و محیط‌های متغیر سازگار شوند.

به طور خلاصه، بینایی ماشین زیرمجموعه‌ای از بینایی کامپیوتر است که بر استفاده از دستگاه‌های سخت‌افزاری برای گرفتن داده‌های بصری برای کاربردهای صنعتی خاص تمرکز دارد، در حالی که بینایی کامپیوتر حوزه وسیع‌تری است که شامل استفاده از الگوریتم‌ها و تکنیک‌های یادگیری ماشین برای تفسیر و درک بصری داده‌ها برای طیف گسترده‌ای از برنامه‌ها می‌شود. [۲]

۲-۲) استخراج و تطابق ویژگی‌های تصویر

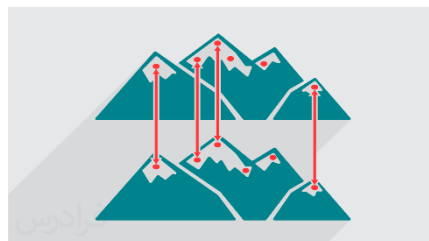
استخراج و تطابق ویژگی‌های محلی همانند الگوریتم SIFT یا Scale Invariance Feature Transform، یک مبحث اساسی در بینایی کامپیوتر می‌باشد که کاربردهای زیادی در بازیابی تصاویر، شناسایی اشیاء، ثبت تصاویر و ... دارد. چند الگوریتم معروف این بخش عبارتند از:

الگوریتم SIFT:

الگوریتم سیفت (SIFT) در سال ۱۹۹۹ توسط David Lowe ابداع گردید. الگوریتم SIFT یک روش برای آشکارسازی، استخراج و توصیف نقطه مشخصه‌های کلیدی در تصاویر است که می‌تواند در کاربردهایی مانند تناظریابی بین تصاویر، شناسایی اشیاء، بازسازی سه‌بعدی صحنه‌ها و غیره مورد استفاده قرار گیرد.

الگوریتم Sift دو مرحله اصلی دارد: Feature Extraction و Feature Matching

در مرحله اول (Feature Matching) ابتدا الگوریتم در تمام تصاویر می‌گردد و یک سری نقاط خاصی رو پیدا می‌کند، به این نقاط Key Point گفته می‌شود. نکته مهم در این الگوریتم نرم افزار این است که می‌تواند تصاویری که با مقیاس‌های مختلف هستند را نیز تناظریابی کند. این مسئله در الگوریتم Sift حل شده است. یعنی اگر فاصله فرد نسبت به سوژه کم یا زیاد بشود یا حتی وضعیت روشنایی تغییر کند یا حتی جهت تصاویر فرد تغییر کند، الگوریتم سیفت میتواند به راحتی این نقاط مشترک را پیدا کند. به عنوان مثال الگوریتم در عکس اول ۵۰ هزار نقطه پیدا کرده در عکس دوم ۶۰ هزار نقطه و در عکس سوم ۴۰ هزار نقطه تصویری پیدا می‌نماید. هر نقطه تصویری دو مختصات x و y دارد و تعداد (مثلاً ۱۲۸) ویژگی خاص به اون نقطه که توصیف‌گر الگوریتم سیفت می‌تواند مشخص کند. نکته بارز الگوریتم سیفت همین ۱۲۸ ویژگی است که برای هر کدام از این نقاط در نظر می‌گیرد.



شکل ۴: نقاط خاص در روش SIFT

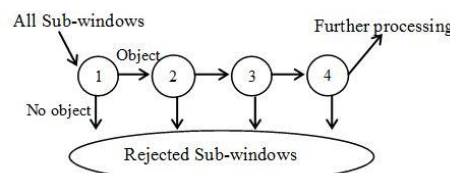
مرحله دوم Feature Matching است. یعنی الگوریتم شروع می‌کند و از بین همه Key point که استخراج کرده می‌گردد تا ببیند برای نقطه اول در تصویر اول چه نقطه متناظری در تصویر دوم وجود دارد. نقطه متناظر نقطه‌ای است که در آن ۱۲۸ ویژگی با نقطه اول برابر باشد. به نقاط گرهی Tie Point گفته می‌شود. الگوریتم این عمل را برای تمام نقاط تکرار می‌کند. در مثال فوق ۲۰ هزار Tie Point بدست آمده است. [۳]

الگوریتم SURF

The Speeded Up Robust Features (SURF) یک روش تشخیص و توصیف ویژگی برای تصاویر است. این یک الگوریتم قوی و سریع است که اغلب در برنامه های بینایی کامپیوتری مانند تشخیص اشیا و ثبت تصویر استفاده می شود. SURF به عنوان یک نسخه "سرعت یافته" از الگوریتم تغییر ویژگی (SIFT) Scale-Invariant در نظر گرفته می شود. کارایی محاسباتی آن باعث می شود که آن را برای برنامه های بلادرنگ مفیدتر از SIFT کند. [۴]

ویولا جونز

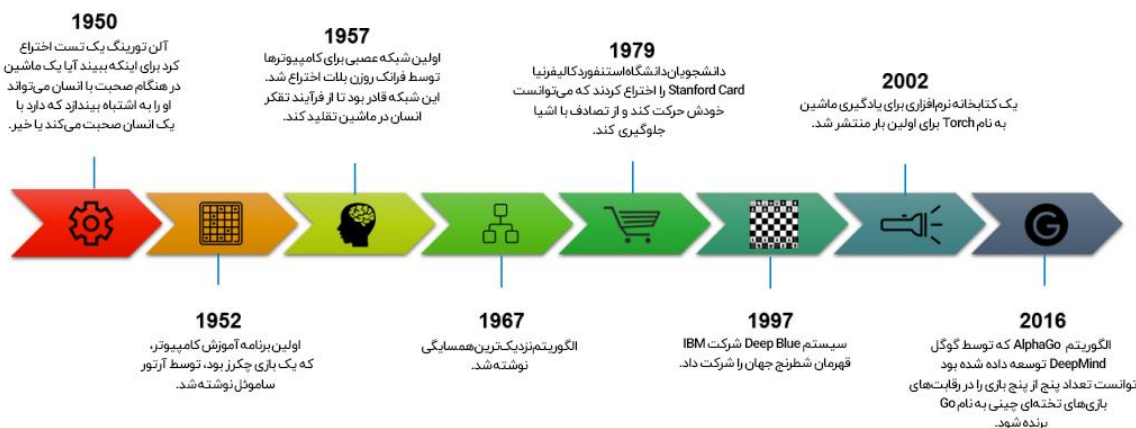
Viola-Jones یک الگوریتم بینایی کامپیوتری برای تشخیص اشیا، به ویژه برای تشخیص چهره در تصاویر است. این الگوریتم توسط پل ویولا و مایکل جونز در سال ۲۰۰۱ توسعه داده شد. این الگوریتم از تکنیکی به نام "تصویر انتگرال" استفاده می کند که امکان محاسبه سریع ویژگی های هار را فراهم می کند، که برای مطابقت با ویژگی های چهره های معمولی انسان استفاده می شود. این الگوریتم همچنین از «طبقه بندی کننده های آبخاری» که گروهی از ویژگی های هار هستند، برای پیش بینی در مورد حضور چهره در یک تصویر استفاده می کند. الگوریتم Viola-Jones بسیار کارآمد است و به طور گسترده در بسیاری از برنامه ها مانند سیستم های امنیتی، برچسب گذاری عکس در جایی که قدرت محاسباتی محدود است استفاده می شود. [۴]



شکل ۴: طبقه بندی کننده های آبخاری در روش ویولا جونز

۳- یادگیری ماشین یا ML (Machine Learning)

یادگیری ماشین (Machine Learning) یکی از زیر مجموعه های هوش مصنوعی است که به سیستم ها این امکان را می دهد تا به صورت خودکار یادگیری و پیشرفت داشته باشند بدون اینکه به برنامه نویسی صریحی برای آن داشته باشند. تمرکز اصلی یادگیری ماشینی بر توسعه برنامه های رایانه ای است که بتوانند به داده ها دسترسی پیدا کنند و از آن برای یادگیری خود استفاده کنند. فرآیند یادگیری با مشاهدات یا داده ها آغاز می شود، مانند مثال ها، تجارب مستقیم و یا دستور العمل ها، تا به یک الگو در داده ها برسند و بر اساس این مثال هایی که ارائه می شود، تصمیمات بهتری بگیرند. هدف اصلی آن است که به کامپیوتر این اجازه داده شود که بدون دخالت و کمک انسان به طور اتوماتیک یادگیری داشته باشند و بتوانند اقدامات خود را بر مطابق با آن تنظیم کنند. یادگیری ماشینی توانایی یادگیری مستقل را برای ماشین ها ایجاد می کند. یعنی به عبارتی یک ماشین می تواند از تجربیات، مشاهدات و الگوهایی که بر اساس یک مجموعه داده تجزیه و تحلیل می کند، آموزش ببیند. نکته مهم آن است که برای این کار لازم نیست به صورت اختصاصی برنامه ریزی شده باشد.

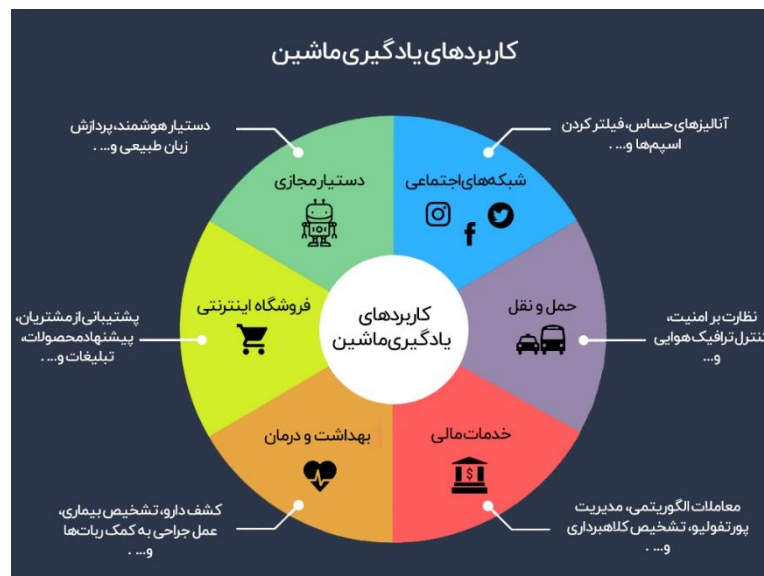


شکل ۵: تاریخچه یادگیری ماشین [۵]

الگوریتم های بسیار مختلفی برای یادگیری ماشین وجود دارد و هر روزه صدها الگوریتم جدید نیز تولید می شوند، و به طور

- معمول توسط سبک یادگیری (learning style) (مانند یادگیری نظارت شده، یادگیری بدون نظارت، یادگیری نیمه نظارت) و یا با توجه به شباهتشان در فرم و عملکرد (مانند طبقه بندی، برگشت، درخت تصمیم گیری، دسته کردن، یادگیری عمیق و...) گروه بندی می شوند. صرف نظر از سبک یادگیری یا عملکرد، تمام الگوریتم‌های یادگیری ماشینی، موضوعات زیر را در بر دارند:
- نمایش: مجموعه ای از طبقه بندی کننده‌ها یا زبانی که کامپیوتر آن را می فهمد.
 - ارزشیابی: همچنین معروف به عملکرد هدف/نمره دهی.
 - بهینه سازی: روش جست و جو؛ اغلب طبقه بندی کننده ای با بالاترین امتیاز.

هدف اساسی الگوریتم‌های یادگیری ماشینی، تعمیم یادگیری‌ها به فراتر از نمونه‌های آموزش داده شده است، یعنی تفسیر موفقیت آمیز داده‌ها. یادگیری ماشین آنالیز مقادیر انبوهی از داده‌ها را امکان پذیر می کند. این یادگیری در شناسایی فرصت‌های سودآور و یا خطرناک معمولاً نتایج سریعتر و دقیق تری ارائه می کند اما برای آموزش آن ممکن است به زمان و منابع اضافی نیاز داشته باشد. تلفیقی از یادگیری ماشینی و هوش مصنوعی و فناوری‌های شناختی می تواند در پردازش حجم زیادی از اطلاعات موثر باشد. [۵]



شکل ۶: برخی کاربردهای یادگیری ماشین [۵]

الگوریتم‌های یادگیری ماشین عمدتاً به سه دسته یادگیری تحت نظارت (Supervised Learning)، یادگیری بدون نظارت (Unsupervised Learning) و یادگیری تقویتی (Reinforcement Learning) تقسیم می شود:

۳-الف) الگوریتم یادگیری ماشین نظارت شده

الگوریتم یادگیری ماشین نظارت شده می تواند از آن چه که در گذشته آموخته اند و همچنین داده‌های جدید برچسب گذاری شده، برای پیشبینی آینده استفاده کند. این کار از آنالیز مجموعه داده‌های آموزشی شروع می شود، الگوریتم یادگیری یک عملکرد استنباطی تولید می کند تا پیش بینی‌های مربوط به مقادیر خروجی را انجام دهد. این نوع سیستم قادر است پس از آموزش‌های کافی برای هر داده جدیدی هدف مشخص کند. [۵]

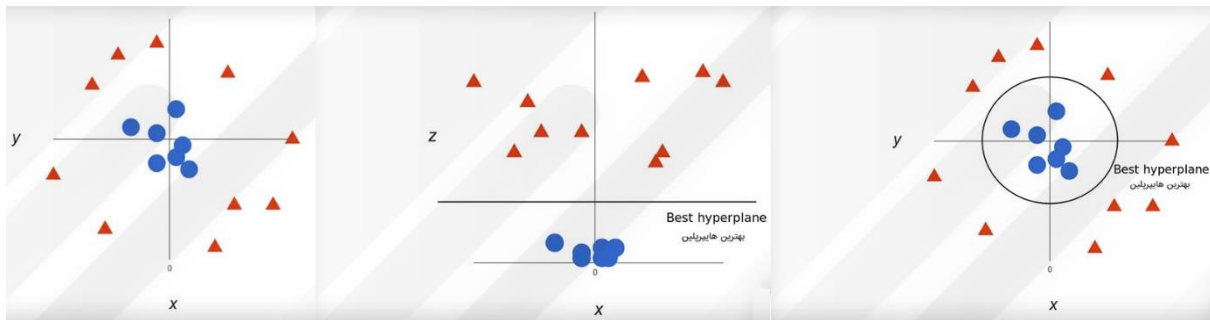
۳-الف-۱) ماشین بردار پشتیبان یا SVM:

ماشین بردار پشتیبان یا Support Vector Machine که به اختصار به آن SVM گفته می‌شود یک الگوریتم یادگیری ماشین با ناظر است که نمونه‌ی داده‌هایی را به صورتی نقاطی در فضا نشان داده شده است، با استفاده از یک خط یا هایپرپلین (Hyperplane)، از هم جدا می کند. این جداسازی به گونه‌ای است که نقاط داده‌ای که در یک طرف خط هستند مشابه هم و در یک گروه قرار می گیرند. نمونه داده‌های جدید هم بعد از اضافه شدن به همان فضا در یکی از دسته‌های موجود قرار خواهند گرفت. در الگوریتم ماشین بردار پشتیبان، هنگام ساختن یک هایپرپلین برای نقاط داده‌ی تفکیک پذیر خطی کار آسانی است، اما وقتی

داده‌ها غیر خطی تفکیک پذیر باشند، کار چالش برانگیز خواهد بود. چند مفهوم رایج در الگوریتم ماشین بردار پشتیبان عبارتند از: حاشیه (Margin)، هایپرپلین (Hyperplane)، فضای ویژگی (Feature Space)

- حاشیه در SVM به فضایی بین دسته‌های داده‌های طبقه‌بندی شده اشاره دارد. حاشیه مانند منطقه‌ای امن بین دو گروه است که باید جدا شود؛ هدف این است که این منطقه را تا جای ممکن گسترده گردد تا حتی اگر داده‌های جدیدی اضافه شوند، با اطمینان بیشتری بتوان آن‌ها را طبقه‌بندی کرد. حاشیه بزرگ‌تر نشان‌دهنده جداسازی قوی‌تر و در نتیجه، مدل دقیق‌تری است.
- هایپرپلین در SVM، خط یا صفحه‌ای است که داده‌ها را به دو دسته جدا می‌کند. در فضای دوبعدی، این یک خط مستقیم است؛ اما در فضاهای با ابعاد بالاتر، به صفحه‌ای بزرگ‌تر تبدیل می‌شود که می‌تواند داده‌ها را در ابعاد مختلف جدا کند. هایپرپلین بهینه آن صفحه‌ای است که بزرگترین حاشیه را بین دسته‌های مختلف داده‌ها ایجاد می‌کند.
- فضای ویژگی به بعد یا دامن‌های گفته می‌شود که در آن داده‌ها نمایش داده می‌شوند و تجزیه و تحلیل می‌شوند. در SVM، تبدیل داده‌ها به فضای ویژگی بالاتر این امکان را فراهم می‌کند که داده‌هایی که در ابعاد اصلی خود قابل جداسازی خطی نیستند، به شکلی تبدیل شوند که بتوان آن‌ها را به طور خطی جدا کرد.

اگر داده‌های اولیه به شکل زیر (سمت چپ) باشند، لازم است که بعد سوم را اضافه نمود. اکنون یک بعد z جدید ایجاد می‌شود که به این شکل محاسبه می‌شود: $z = x^2 + y^2$ (این معادله یک دایره است). این کار فضایی سه‌بعدی ایجاد می‌کند. در فضای z و X نقاط داده‌ی، به راحتی با یک خط تفکیک شدند. آنچه باقی مانده ترسیم مجدد آن به شکل دوبعدی است. به این فرایند که توضیح داده شد حقه‌ی کرنل (Kernel Trick) گفته می‌شود. [۶]



شکل ۷: مفهوم حقه‌ی کرنل در الگوریتم SVM [۶]

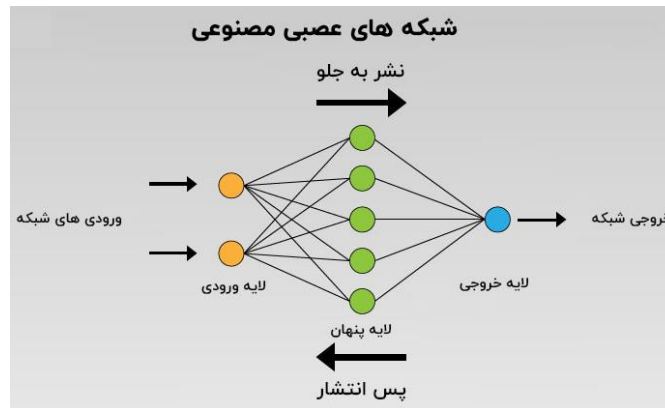
۳-الف-۲) شبکه‌های عصبی مصنوعی (ANN):

شبکه‌های عصبی، زیرشاخه‌ای از علم یادگیری ماشین نظارت شده هستند که از ساختار و عملکرد مغز انسان و شبکه‌های عصبی بیولوژیکی الهام گرفته‌اند. این شبکه‌ها از نورون‌های مصنوعی برای پردازش داده‌ها استفاده و به وسیله وزن‌ها و اتصالات میان نورون‌ها، الگوها را مدل می‌کنند. شبکه عصبی مجموعه‌ای از الگوریتم‌ها است که عملکرد مغز انسان را تقلید می‌کند. این شبکه‌ها بر اساس ساختار و عملکرد مغز انسان مدل‌سازی و از گره‌های به هم پیوسته به نام نورون‌ها تشکیل شده‌اند که برای پردازش و تحلیل داده‌ها مورد استفاده قرار می‌گیرند. در قلب شبکه عصبی، نورون قرار دارد که سیگنال‌های ورودی را دریافت می‌کند، آن‌ها را پردازش و سیگنال خروجی را تولید می‌کند. شبکه‌های عصبی از لایه‌های مختلفی از نورون‌ها تشکیل شده‌اند که هر لایه وظیفه خاصی در فرآیند یادگیری دارد. شبکه عصبی از سه جزء اصلی تشکیل شده است.

- لایه ورودی: این لایه وظیفه دریافت داده‌های اولیه ورودی به شبکه را دارد.
- لایه‌های پنهان: این نوع لایه‌ها، بین لایه ورودی و لایه خروجی قرار می‌گیرند و محاسبات اصلی شبکه در همین لایه‌ها انجام می‌شوند.
- لایه خروجی: نتیجه نهایی برای ورودی‌های داده شده توسط کاربر تولید می‌شود.

نورون‌ها بلوک‌های سازنده شبکه‌های عصبی هستند که سیگنال‌های ورودی را دریافت می‌کنند، محاسبات را انجام می‌دهند و خروجی را ارسال می‌کنند. در یک شبکه عصبی مصنوعی، نورون‌ها با توابع ریاضی نشان داده می‌شوند. این توابع میانگین وزن‌دار

ورودی‌ها را می‌گیرند و با استفاده از یک تابع فعال سازی، خروجی هر نورون را تعیین می‌کنند. شبکه عصبی با استفاده از الگوریتم «پس انتشار» (Backpropagation)، وزن‌های اتصالات بین نورون‌ها را تنظیم می‌کند تا خروجی پیش‌بینی شده آن به خروجی واقعی نزدیک‌تر شود. با تکرار این فرآیند با تعداد زیادی مثال، شبکه یاد می‌گیرد که پیش‌بینی‌های دقیق انجام دهد.



شکل ۸: شبکه‌های عصبی مصنوعی (ANN)

لایه‌های پنهان نقش مهمی در فرآیند یادگیری دارند. این لایه‌ها وظیفه استخراج ویژگی‌های پیچیده از داده‌های ورودی را بر عهده دارند. هر لایه ورودی‌های لایه قبلی را دریافت می‌کند، محاسبات را انجام می‌دهد و نتایج را به لایه بعدی منتقل می‌کند. هر چه شبکه عمیق‌تر باشد، ویژگی‌های انتزاعی و سطح بالاتری می‌تواند یاد بگیرد. این یادگیری سلسله مراتبی به شبکه‌های عصبی اجازه می‌دهد تا الگوها را در سطوح مختلف پیچیدگی تشخیص دهند. برخی از انواع شبکه‌های عصبی پرکاربرد عبارتند از:

- «شبکه عصبی پرسپترون چند لایه» (Multi Layer Perceptron | MLP)
- «شبکه عصبی پیچشی» (Convolution Neural Networks | CNN)
- «شبکه عصبی بازگشتی» (Recurrent Neural Networks | RNN)
- «شبکه عصبی تابع پایه شعاعی» (Radial Basis Functional Neural Network | RBF)

شبکه‌های عصبی به جای تکیه بر قوانین و الگوهای از پیش تعریف شده، با پردازش مقادیر زیادی داده و تنظیم وزن اتصالات بین نورون‌ها، فرآیند یادگیری را انجام می‌دهند. یکی از تفاوت‌های اصلی بین یادگیری ماشین کلاسیک و روش‌های شبکه عصبی، در نحوه استخراج ویژگی از داده‌های ورودی است. در روش‌های یادگیری ماشین کلاسیک، معمولاً متخصصان با استفاده از دانش خود از نوع مسئله و روش‌های آماری مناسب، ویژگی‌های مهم را استخراج می‌کنند که این ویژگی‌ها به عنوان ورودی به مدل داده می‌شوند. در مقابل، در شبکه‌های عصبی، می‌توانند به طور تقریباً خودکار ویژگی‌های مهم را از داده‌های ورودی استخراج کنند. یکی از مزیت‌های اصلی شبکه‌های عصبی، توانایی پردازش داده‌های پیچیده و بدون ساختار است که مهم‌ترین نقش شبکه عصبی در یادگیری ماشین است. برخلاف الگوریتم‌های یادگیری ماشین کلاسیک که نیاز به پردازش و ساختاردهی داده‌ها دارند، شبکه‌های عصبی می‌توانند داده‌های خام مانند تصاویر، صدا و متن را به طور مستقیم پردازش کنند. همچنین، شبکه‌های عصبی با توانایی تجزیه و تحلیل داده‌ها به اجزای کوچک‌تر و شناخت الگوها، درک جامعی از مسائل پیچیده پیدا می‌کنند. به عنوان مثال، در تجزیه و تحلیل تصاویر، این شبکه‌ها می‌توانند بهتر از الگوریتم‌های کلاسیک لبه‌ها، شکل‌ها و بافت‌ها را شناسایی کرده و درک بهتری از تفاوت‌های ظریف و پیچیدگی‌های موجود در داده‌های تصویری داشته باشند. [۷]

۳-ب) الگوریتم یادگیری ماشین بدون نظارت

در مقابل، زمانی از الگوریتم‌های یادگیری ماشین بدون نظارت استفاده می‌شود که اطلاعات مورد نیاز برای آموزش نه طبقه بندی شده باشد و نه برچسب زده باشد. یادگیری بدون نظارت چگونگی اینکه سیستم می‌تواند توصیف ساختار پنهان از داده‌های بدون برچسب استنباط کنند، مطالعه می‌کند. این نوع سیستم خروجی مناسب را مشخص نمی‌کند و تنها می‌تواند داده‌ها را

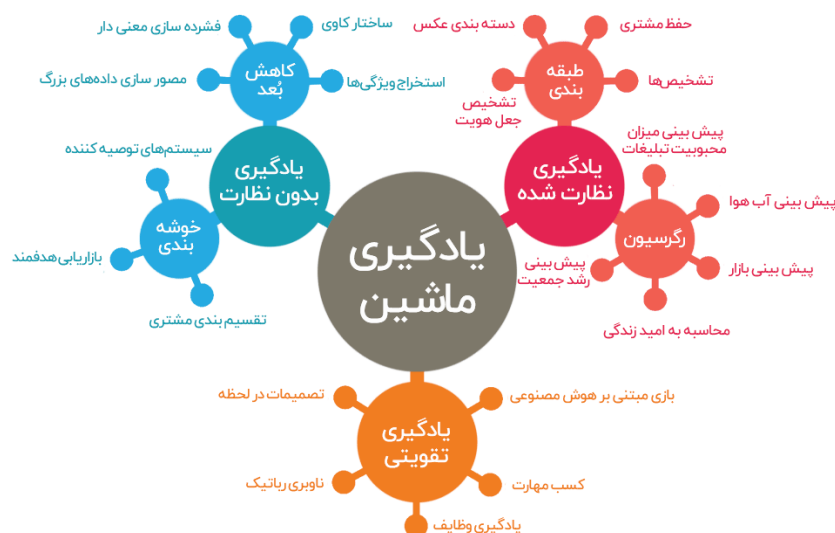
کاوش کند و از داده‌های برجسب زده نشده ساختارهای پنهان را استنتاج می کند. [۵] یکی از الگوریتم های یادگیری بدون نظارت، یادگیری عمیق است که در بخش بعد به آن پرداخته می شود.

۳-پ) الگوریتم یادگیری ماشین نیمه نظارت شده

الگوریتم یادگیری ماشین نیمه نظارت شده بین دو نوع قبلی قرار دارد. این سیستم از هر دو نوع داده برجسب زده شده و برجسب نزده شده برای آموزش استفاده می کند. سیستم‌هایی که از این روش استفاده می کنند، می توانند دقت یادگیری را تا میزان قابل توجهی بهبود ببخشند. معمولاً زمانی این نوع یادگیری انتخاب می شود که داده‌های برجسب زده شده بدست آمده نیاز به منابع ماهر و مرتبط برای آموزش و یادگیری دارند. در غیر این صورت، دستیابی به داده‌های دارای برجسب معمولاً نیازی به منابع اضافی ندارد. [۵]

۳-ت) الگوریتم یادگیری ماشین تقویت کننده

الگوریتم‌های یادگیری ماشین تقویت کننده روشی هستند که به وسیله اقدامات با محیط خود در تعامل هستند و خطاها و پاداش‌ها را کشف می کنند. آزمایش، جست و جوی خطاها و پاداش‌های تأخیری مهم ترین ویژگی‌های یادگیری تقویتی هستند. این نوع یادگیری به ماشین‌ها و عوامل نرم افزار اجازه می دهد تا به طور خودکار، برای به حداکثر رساندن عملکرد خود، رفتار ایده آل خود را مشخص کنند. این سیستم از بازخورد پاداش ساده استفاده می کند تا ببیند کدام عمل بهتر است و این امر به عنوان سیگنال تقویت شناخته شده است. [۵]



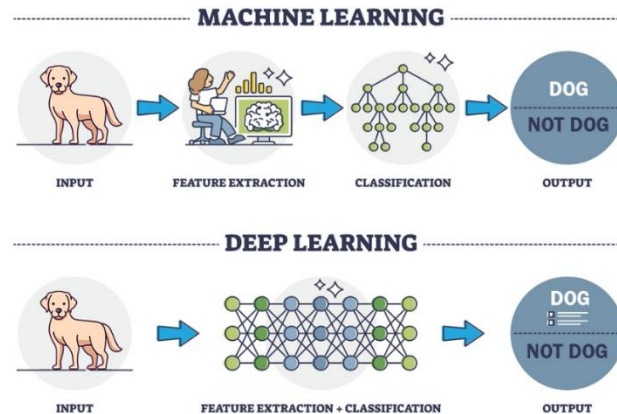
شکل ۹: سه دسته اصلی الگوریتم های یادگیری ماشین [۵]

۴- یادگیری عمیق (Deep Learning)

۴-الف) مفهوم یادگیری عمیق:

یادگیری عمیق زیرمجموعه‌ای از یادگیری ماشین (Machine Learning) و از روش‌های هوش مصنوعی (Artificial Intelligence) است که شیوهی یادگیری انسان‌ها را مورد تقلید قرار می‌دهد. یادگیری عمیق راهی برای اتوماتیک کردن تحلیل‌های پیش‌بینی‌کننده (Predictive Analysis) است. برخلاف الگوریتم‌های خطی یادگیری ماشین، الگوریتم‌های سلسله‌مراتبی یادگیری عمیق در هر مرحله بر میزان پیچیدگی و انتزاعشان افزوده می‌شود. مثال معروف الگوریتم یادگیری عمیق عبارت است از: کودکی نوپا برای اولین بار کلمه‌ی سگ را به زبان می‌آورد، اما معنای پشت این کلمه را نمی‌داند. او هر بار به چیزهای مختلفی اشاره می‌کند و کلمه‌ی سگ را به زبان می‌آورد. پدر و مادر این کودک در پاسخ به هر اشاره‌ی او می‌گویند «بله، این سگ هست» و یا «نه، این سگ نیست». بعد از مدتی کودک به ویژگی‌های مشترک تمامی سگ‌ها پی می‌برد. اتفاقی که در اینجا می‌افتد این است که بدون اطلاع کودک، سلسله‌مراتبی از انتزاعات در ذهن او در حال شکل‌گیری است؛ انتزاعاتی که در هر مرحله مبتنی بر مفاهیم لایه‌های قبلی هستند. از این طریق است که مفاهیمی پیچیده

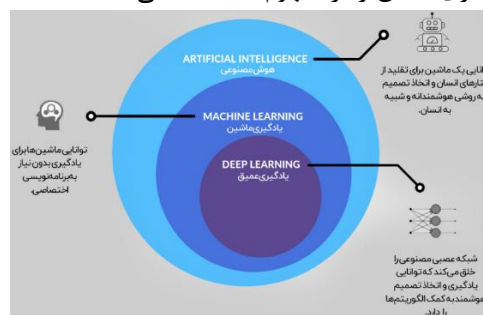
و انتزاعی در ذهن کودک شکل می گیرد.



شکل ۱۰: مثال معروف سگ در الگوریتم یادگیری عمیق [۸]

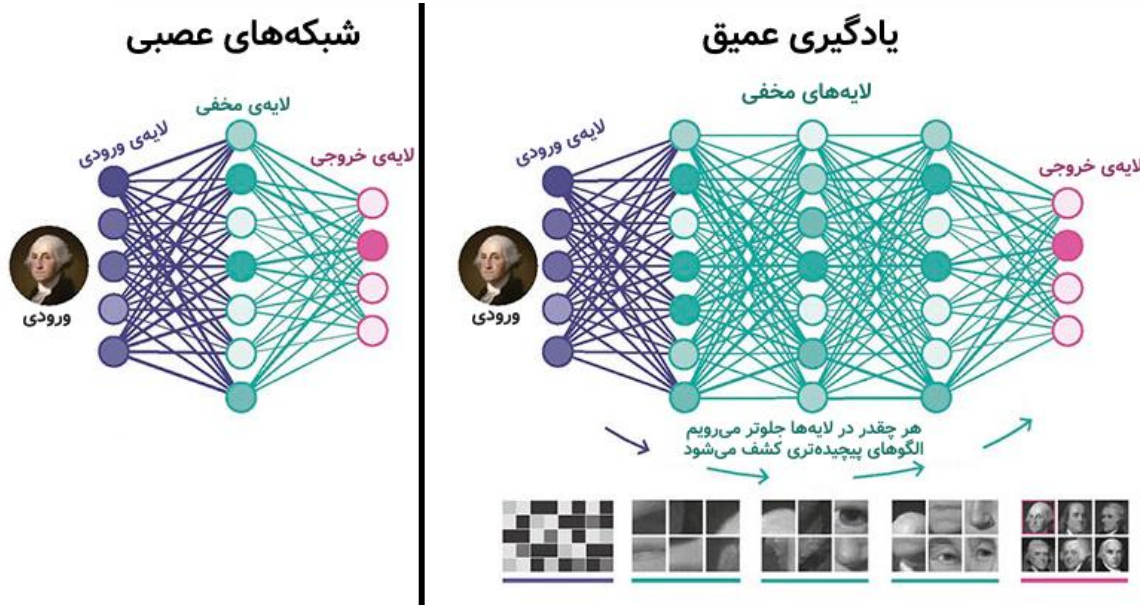
یادگیری عمیق نسبت به فرایند یادگیری نظارت شده (Supervised Learning) تفاوت های جدی دارد. برای مثال، به کامپیوتر آموزش داده می شود تا سگ را در تصاویر شناسایی کند. برای این کار می بایست جزء به جزء برای کامپیوتر توضیح داده شود که در جست و جوی تصاویر باید به دنبال چه ویژگی هایی بگردد. در واقع باید ملاک تصمیم گیری برای وجود یا عدم وجود سگ در تصویر برای کامپیوتر تعریف می گردید. این کار که استخراج ویژگی (Feature Extraction) نامیده می شود، کاری پر زحمت است و موفقیت کامپیوتر در آن کاملاً به تبهر برنامه نویس در تعریف کردن مجموعه ویژگی های سگ بستگی دارد. اما یادگیری عمیق برخلاف یادگیری ماشین نیازی به نظارت ندارد و به خودی خود می تواند مجموعه ویژگی ها را تشخیص دهد. یادگیری بدون نظارت (Unsupervised Learning) نه تنها سریع تر است، بلکه دقت بالاتری هم دارد. برنامه های کامپیوتری برای اینکه تصویر یک سگ را شناسایی کنند، باید مانند همان کودک مثال قبل عمل کنند. هر کدام از الگوریتم های یادگیری عمیق به صورت سلسله مراتبی روش تبدیل خطی را بر روی ورودی های خود اعمال می کنند و از آنچه آموخته اند در ساختن یک مدل آماری به شکل خروجی استفاده می کنند. گفتنی است که این فرایند تا جایی که خروجی به سطح قابل قبولی از دقت برسد، ادامه پیدا خواهد کرد. واژه ی عمیق در یادگیری عمیق برداشتی از همین تعداد لایه هایی است که داده ها باید از آن عبور کنند. برنامه های یادگیری عمیق برای اینکه به سطح دقت قابل قبولی برسند، باید به میزان عظیمی از داده های آموزشی دسترسی داشته و دارای قدرت پردازش بالایی باشند. تمامی این ها در عصر کلان داده و محاسبات ابری در دسترس برنامه نویسان قرار گرفته است. برای کودک ممکن است فهمیدن مفهوم سگ هفته ها و حتی ماه ها زمان ببرد. اما با استفاده از الگوریتم های یادگیری عمیق، برنامه ی کامپیوتر در ظرف چند دقیقه می تواند تصویر سگ را شناسایی کند. فقط کافی است که مجموعه ای از داده های آموزشی را در اختیار برنامه بگذارید تا برنامه فوراً میلیون ها تصویر را جست و جو کند و تصاویر مورد نظر را به درستی تشخیص دهد. [۸]

یادگیری عمیق زیر شاخه ای از یادگیری ماشین است که شبکه عصبی اجزای سازنده آن را تشکیل می دهند. در یادگیری عمیق، شبکه های عصبی دارای تعداد زیادی لایه مخفی هستند. لایه مخفی یک لایه از نورون ها است که بین لایه ورودی و لایه خروجی قرار دارد. به عبارتی دیگر، آنچه که یادگیری عمیق را از مفهوم شبکه عصبی جدا میکند تعداد لایه های بیشتر آن است. [۷]



شکل ۱۱: یادگیری عمیق یکی از الگوریتم های یادگیری ماشین [۵]

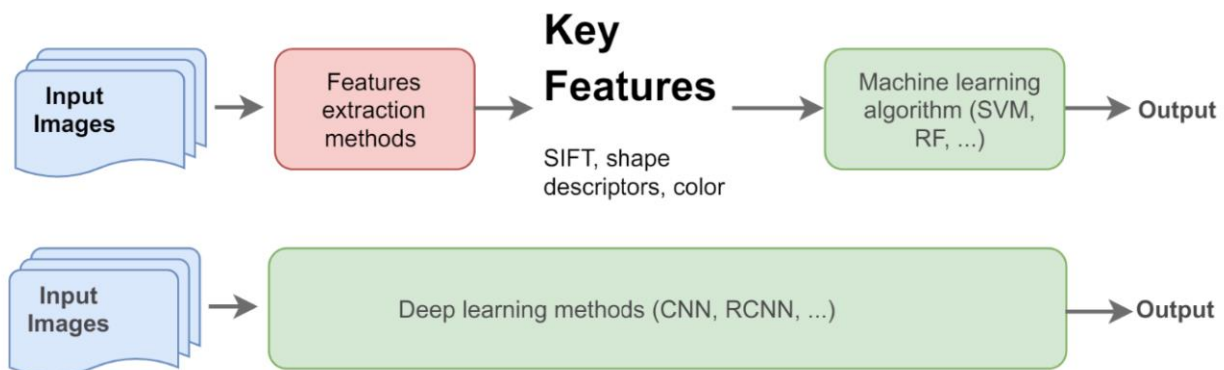
نوعی پیشرفته از الگوریتم‌های یادگیری ماشین به نام شبکه‌های عصبی مصنوعی، پایه و اساس اغلب مدل‌های یادگیری عمیق را تشکیل می‌دهند. به همین دلیل است که گاهی یادگیری عمیق با عناوینی مانند یادگیری عصبی عمیق (Deep Neural Learning) و یا شبکه‌سازی عصبی عمیق (Deep Neural Networking) نیز خوانده می‌شود [۸]. به عبارت دیگر الگوریتم‌های یادگیری عمیق همان شبکه‌های عصبی هستند که تعداد لایه‌های آن‌ها زیاد است. به معنای دقیق‌تر تفاوت بین یادگیری عمیق و شبکه‌های عصبی آن است که، یادگیری عمیق حالت یادگیری بوده در حالی که شبکه‌های عصبی نوعی الگوریتم در بین یادگیری ماشین است. [۹]



شکل ۱۲: الگوریتم شبکه‌ی عصبی عمیق و شبکه عصبی عادی [۹]

در شکل بالا، در قسمت راست، یک الگوریتم شبکه‌ی عصبی عمیق ایجاد شده است که منجر به یادگیری عمیق می‌شود. این شبکه (سمت راست) نسبت به شبکه‌های عصبی عادی (سمت چپ)، تعداد لایه‌های بالاتری داشته و با این کار الگوریتم امید دارد که بتواند الگوهای پیچیده‌تری را از میان داده‌ها کشف کند. البته با گذشت زمان و توسعه الگوریتم‌های جدید، یادگیری عمیق دیگر محدود به شبکه‌هایی مانند شکل بالا نیست و با توجه به کاربردهای مختلف، معماری‌های متفاوتی شناسایی الگو در مسائل با توجه به نوع و شکل داده‌ها ابداع شده است. [۹]

یکی از مزایای یادگیری عمیق، حذف مراحل مهندسی ویژگی و استخراج و تطابق ویژگی‌های کلیدی می‌باشد: [۱۱]



شکل ۱۳: حذف مراحل مهندسی ویژگی و استخراج و تطابق ویژگی‌های کلیدی [۱۱]

۴-ب) داده کاوی در یادگیری عمیق

داده کاوی (data mining) فرآیند استخراج اطلاعات مفید از مجموعه داده‌های بزرگ است. در یادگیری عمیق، داده کاوی برای اطمینان از اینکه مدل، اطلاعات کافی برای یادگیری دارد، ضروری است. پروسه زیر، در داده کاوی انجام می‌گیرد:

- پیش پردازش داده‌ها (data pre-processing)، شامل تمیز کردن، تبدیل و عادی سازی داده‌ها برای سازگار سازی آن‌ها برای مدل است.
- مهندسی ویژگی (feature engineering)، فرآیند انتخاب و تبدیل ویژگی‌های مربوطه در داده‌ها برای بهبود عملکرد مدل است.
- تشدید داده‌ها (data augmentation)، شامل ایجاد داده‌های اضافه از داده‌های موجود، با اعمال تبدیل‌هایی مانند دوران (rotation)، مقیاس‌بندی (scaling) و چرخاندن (flipping) است.

مجموعه داده‌های بزرگ برای یادگیری عمیق یک اصل مهم و ضروری است، زیرا به مدل اجازه می‌دهد تا دقت داده‌ها را بالا ببرد. [۱۰]

۴-پ) الگوریتم‌های یادگیری عمیق YOLO:

روش‌های معمول مبتنی بر یادگیری عمیق برای تشخیص و شناسایی اشیاء از تصاویر شامل شبکه‌های کانونی مبتنی بر منطقه به نام Faster R-CNN، اندازه‌گیری‌های یکضربی (SSD)، شبکه‌های کانونی کاملاً مبتنی بر منطقه (R-FCN) و تکنیک "تنها یک بار نگاه کن" یا YOLO و سایرین می‌شود. این روش‌ها عملکرد برتری را در طیف گسترده‌ای از وظایف پیچیده در حوزه‌ی CV، به خصوص در محل محلی سازی و شناسایی اشیاء، از خود نشان داده‌اند. الگوریتم‌های مبتنی بر طبقه‌بندی در دو مرحله انجام می‌شوند. آن‌ها با شناسایی مناطق مدنظر یا regions of interest در یک تصویر شروع می‌کنند. سپس از شبکه‌های عصبی کانولوشنی یا CNN برای طبقه‌بندی این مناطق استفاده می‌کنند. از آنجا که باید پیش‌بینی‌هایی را برای هر منطقه مشخص شده ارائه نمود، این نوع الگوریتم‌ها ممکن است کند باشند. شبکه عصبی کانولوشنی مبتنی بر منطقه یا Region-based CNN (RCNN) و Faster-RCNN، Mask-RCNN نمونه‌های شناخته‌شده‌ای از این نوع الگوریتم هستند. الگوریتم‌های مبتنی بر رگرسیون کلاس‌ها و کادرهای محدودکننده را برای کل تصویر در یک بار اجرای الگوریتم پیش‌بینی می‌کنند. الگوریتم‌های خانواده YOLO و SSD دو نمونه از شناخته‌شده‌ترین نمونه‌های این گروه هستند. آن‌ها به‌طور گسترده در زمینه تشخیص اشیاء به شکل بلادرنگ استفاده می‌شوند. [۱۹]

You Only Look Once یا YOLO یک الگوریتم پیشرفته تشخیص شیء به شکل بلادرنگ است که در سال ۲۰۱۵ جوزف ردمن (Joseph Redmon)، سانتوش دیوالا (Santosh Divvala)، راس گیرشیک (Ross Girshick) و علی فرهادی (Ali Farhadi) در مقاله تحقیقاتی معروف "You Only Look Once, Unified, Real-Time Object Detection" آن را معرفی کردند. نویسندگان با جداسازی فضایی کادرهای محدودکننده و اختصاص مقدار احتمال به هر یک از تصاویر شناسایی شده با استفاده از یک شبکه عصبی کانولوشنی (CNN)، مسئله‌ی تشخیص شیء را به‌عنوان یک تسک رگرسیونی (Regression) به جای یک تسک طبقه‌بندی (Classification) مطرح می‌کنند.

این الگوریتم براساس این چهار رویکرد کار می‌کند:

- بلوک‌های رسوبی یا Residual blocks: این مرحله با تقسیم تصویر اصلی (A) به سلول‌های شبکه‌ای (کادرها) $N \times N$ به شکل مساوی شروع می‌شود. هر سلول در شبکه مسئول مکان‌یابی و پیش‌بینی کلاس شیئی که پوشش می‌دهد، همراه با مقدار احتمال/اطمینان (Probability / Confidence value) است.
- رگرسیون کادر محدودکننده یا Bounding box regression: مرحله بعدی تعیین کادرهای محدودکننده است که همان مستطیل‌هایی است که اشیای موجود در تصویر را نشان می‌دهند. می‌توان به تعداد اشیای درون یک تصویر ارائه شده است کادرهای محدودکننده در نظر گرفت. YOLO با استفاده از یک مازول رگرسیون منفرد در قالب زیر، ویژگی‌های این کادرهای محدودکننده را تعیین می‌کند
- اشتراک روی اجتماع یا به‌اختصار IOU: اغلب اوقات، یک شیء منفرد در یک تصویر می‌تواند چندین نامزد کادر

محدودکننده برای پیش‌بینی داشته باشد. هدف IOU (که مقدار آن بین ۰ و ۱ است) دورانداختن کادرهای محدودکننده‌ای است که مناسب نیستند و درواقع پیش‌بینی خوبی محسوب نمی‌شوند. به این شکل که:

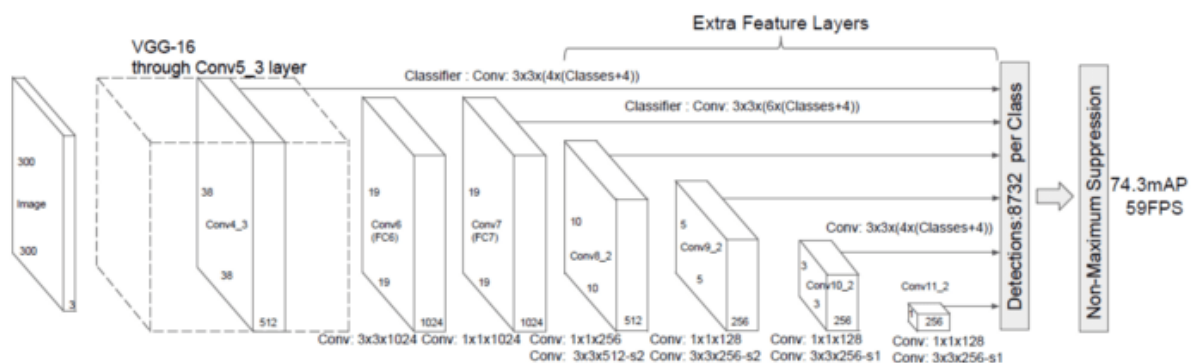
- ✓ کاربر حد آستانه IOU خود را تعریف می‌کند که برای مثال می‌تواند ۰.۵ باشد.
- ✓ سپس YOLO مقدار IOU هر کادر محدودکننده (سلول) را محاسبه می‌کند که مقدار ناحیه اشتراک تقسیم بر ناحیه اجتماع است.
- ✓ درنهایت، مدل کادرهای محدودکننده‌ی دارای IOU کمتر از حد آستانه را نادیده می‌گیرد و آن‌هایی را که IOU بیشتر از حد آستانه دارند در نظر می‌گیرد.

• سرکوب غیرحداکثری یا Non-Maximum Suppression: نکته‌ای که اینجا باید در نظر گرفت این است که تنظیم حد آستانه برای IOU همیشه کافی نیست؛ زیرا یک شیء می‌تواند چندین کادر با IOU بیشتر از حد آستانه داشته باشد و درنظرگرفتن همه آن کادرها ممکن است باعث اضافه کردن نویز شود. اینجاست که می‌توان از NMS استفاده نمود تا فقط کادرهایی را با بالاترین امتیاز احتمال تشخیص نگه داشت. بنابراین به این شکل کادرهای محدودکننده‌ای را در خروجی وجود خواهد داشت که با احتمال بالا اشیای موجود در تصویر را شناسایی می‌کنند.

به‌طور خاص در جراحی، مکان‌یابی (Localization) ارگان‌های بدن به‌شکل بلادرنج، به‌دلیل تنوع بیولوژیکی از یک بیمار به بیمار دیگر، می‌تواند چالش‌برانگیز باشد. تشخیص کلیه در CT اسکن با استفاده از YOLOv3 برای تسهیل مکان‌یابی کلیه‌ها به‌صورت دوبعدی و سه‌بعدی از اسکن‌های توموگرافی کامپیوتری (CT) نمونه‌ای از این استفاده است.

۴-ت) الگوریتم یادگیری عمیق SSD:

آشکارساز چند باکس تک شات (SSD) چند ماه پس از YOLO به عنوان یک جایگزین ارزشمند مطرح شد. به طور مشابه با YOLO، تشخیص شی در یک انتشار رو به جلو شبکه انجام می‌شود. این مدل انتها به انتها CNN تصویر ورودی را از طریق مجموعه‌ای از لایه‌های کانولوشن عبور می‌دهد و کادرهای محصور کننده کاندیدا را از مقیاس‌های مختلف در طول مسیر ایجاد می‌کند. به عنوان حقیقت مبنا برای آموزش، SSD اشیای برچسب گذاری شده را به عنوان نمونه‌های مثبت در نظر می‌گیرد و هر کادر محصور کننده دیگری که با مثبت‌ها همپوشانی ندارد، نمونه‌های منفی هستند. مشخص شد که ساخت مجموعه داده به این روش، آن را بسیار نامتعادل می‌سازد. به همین دلیل، SSD روشی به نام استخراج منفی سخت را درست بعد از انجام NMS اعمال می‌کند. استخراج منفی سخت (hard negative mining) روشی برای انتخاب تنها نمونه‌های منفی با بیش‌ترین اتلاف اطمینان است، به طوری که نسبت بین مثبت و منفی حداکثر ۱:۳ است. این امر منجر به بهینه‌سازی سریع‌تر و پایدار شدن مرحله آموزش می‌شود.

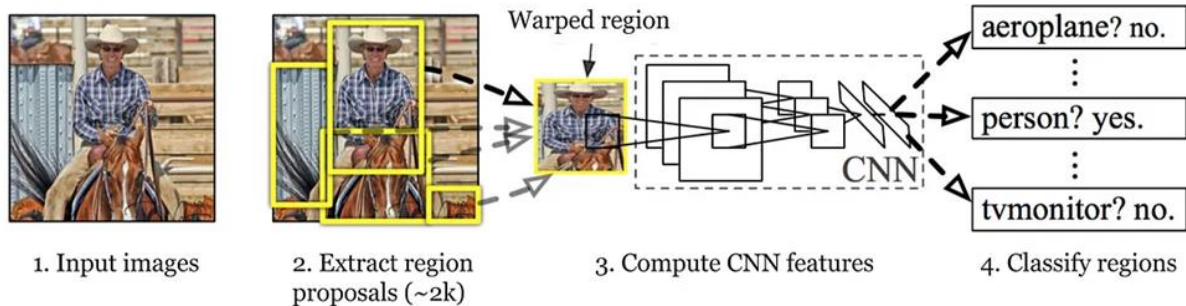


شکل ۱۴: شبکه پایه - ۱۶VGG آشکارساز چند باکس تک شات (SSD)

در تصویر بالا شبکه پایه - ۱۶VGG است. با این حال، امروزه اغلب می‌توان SSD را با شبکه‌های پایه Inception و MobileNet دید. [۲۰]

۴-ث) الگوریتم یادگیری عمیق R-CNN:

در میان الگوریتم‌های مختلف تشخیص اشیا، R-CNN (مخفف Region-based Convolutional Neural Network) و به معنای شبکه‌های عصبی کانولوشنی مبتنی بر ناحیه) نام پرآوازه‌ای است. این الگوریتم در سال ۲۰۱۴ توسط Girshick و همکارانش معرفی شد و با عملکرد فوق‌العاده‌اش، توجه دنیای بینایی ماشین را به خود جلب کرد. می‌توان R-CNN را یکی از پیشگامان حوزه تشخیص اشیا بر پایه یادگیری عمیق به شمار آورد. اما R-CNN دقیقاً چه کاری انجام می‌دهد؟



شکل ۱۵: الگوریتم یادگیری عمیق R-CNN

R-CNN به طور خاص برای تشخیص اشیا در تصاویر طراحی شده است. این الگوریتم با ترکیب شبکه‌های عصبی کانولوشنی با تکنیک‌های شناسایی ناحیه، قابلیت تشخیص و طبقه‌بندی اشیا در تصاویر پیچیده را فراهم می‌کند. با استفاده از R-CNN، می‌توان نواحی مختلف در یک تصویر را به دقت بررسی و اشیا موجود در آن را شناسایی کرد. در ادامه مراحل اجرای الگوریتم R-CNN را به ترتیب بررسی خواهد شد.

تولید ناحیه پیشنهادی: مرحله اول شامل تولید ناحیه پیشنهادی (Region Proposal) است که به معنای جعبه‌های محدودکننده (Bounding Box) یا نواحی است که احتمالاً شامل اشیا مورد نظر در تصویر هستند. در مقاله R-CNN، از روش جستجوی انتخابی (Selective Search) برای این منظور استفاده می‌شود. جستجوی انتخابی از یک الگوریتم بخش‌بندی (Segmentation) برای ایجاد یک مجموعه اولیه از نواحی پیشنهادی استفاده می‌کند. این الگوریتم به‌طور تکراری (Iterative) بخش‌های تصویر را ترکیب می‌کند تا نواحی بزرگ‌تری تشکیل دهد و سپس جعبه‌های محدودکننده‌ای را در اطراف این نواحی ترکیبی رسم می‌کند. جستجوی انتخابی به این شکل کار می‌کند که ابتدا تصاویر به قطعات کوچک‌تر تقسیم می‌شود. سپس این قطعات به‌طور تکراری ترکیب می‌شوند تا نواحی بزرگ‌تری ایجاد شود که می‌توانند اشیا را در بر گیرند. در پایان مرحله جستجوی انتخابی، ۲۰۰۰ ناحیه پیشنهادی برای هر عکس وجود خواهد داشت، که واقعا عدد قابل توجهی است. این نواحی پیشنهادی به عنوان ورودی مراحل بعدی مورد استفاده قرار می‌گیرند.

استخراج ویژگی از نواحی پیشنهادی: پس از تولید نواحی پیشنهادی، مرحله دوم شامل استخراج ویژگی (Feature Extraction) از هر ناحیه است. یک شبکه عصبی کانولوشنی (CNN) به‌طور مستقل به هر ناحیه پیشنهادی اعمال می‌شود تا Feature map را محاسبه کند. در R-CNN، از روش جستجوی انتخابی برای شناسایی نواحی مورد نظر استفاده می‌شود و در پایان، برای هر عکس حدود ۲۰۰۰ منطقه مختلف به دست می‌آید. این یعنی برای پردازش هر عکس، به ۲۰۰۰ شبکه CNN نیاز است. نویسندگان مقاله اصلی R-CNN، از شبکه AlexNet به عنوان استخراج‌کننده ویژگی استفاده کردند، اما سایر معماری‌های CNN نیز می‌توانند برای این منظور مورد استفاده قرار گیرند. برای استفاده از AlexNet به عنوان استخراج‌کننده ویژگی، کافی است لایه Softmax آخر این معماری را حذف نمود. به این ترتیب آخرین لایه شبکه استخراج‌کننده ویژگی یک لایه متصل با ۴۰۹۶ نورون است. این یعنی ویژگی‌های (Features) ۴۰۹۶ بعدی هستند.

Warped Region: نکته دیگری این است که ابعاد ورودی AlexNet همیشه اندازه ثابتی دارد (۲۲۷ در ۲۲۷ در ۳). اما Region Proposal های اشکال مختلفی دارند. بسیاری از آن‌ها کوچک‌تر یا بزرگ‌تر از اندازه گفته‌شده هستند، بنابراین باید پیش از ورود به شبکه AlexNet، اندازه هر یک از آن‌ها را تغییر داد و به ابعاد مشخص و یکسانی رساند. به نواحی پیشنهادی تغییر اندازه‌یافته Warped Region گفته می‌شود.

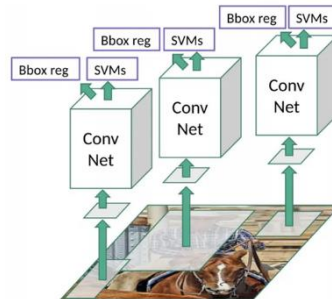
Fine-Tune: همچنین با توجه به اینکه لزوماً تعداد کلاس‌ها برابر با تعداد کلاس‌های مجموعه داده ImageNet که شبکه AlexNet با آن آموزش دیده است نمی‌باشد، لازم است این شبکه طی آموزش مدل R-CNN یک بار روی داده مدنظر نیز Fine-Tune یا تنظیم دقیق شود. برای جدا کردن اشیاء تصویر از پیش‌زمینه (Background) نیز لازم است تعداد کلاس‌های مسئله یک واحد بیش‌تر از تعداد اشیاء مختلف در نظر گرفته شود.

طبقه‌بندی: در مرحله نهایی، ویژگی‌های استخراج شده از پیشنهادات ناحیه وارد یک لایه طبقه‌بند (Classifier) می‌شود. در مقاله اصلی R-CNN، از ماشین بردار پشتیبان (SVM) به عنوان Classifier استفاده شده است. لایه Classifier آموزش داده می‌شود تا هر ناحیه پیشنهادی را در یکی از دسته‌های (Classes) اشیاء از پیش تعریف شده قرار دهد.

ماشین بردار پشتیبان (SVM) یک الگوریتم یادگیری ماشین نظارت شده است که برای طبقه‌بندی و رگرسیون استفاده می‌شود. در R-CNN، SVM به گونه‌ای آموزش داده می‌شود که بتواند ویژگی‌های استخراج شده از هر ناحیه پیشنهادی را به دسته‌های مختلف اشیاء نسبت دهد. این الگوریتم با استفاده از داده‌های آموزشی، مرزهای تصمیم‌گیری را تعیین می‌کند که بتواند اشیاء را با دقت بالا تشخیص دهد.

اصلاح جعبه محدودکننده: برای اینکه جعبه‌های محدودکننده یا همان Bounding Boxها به درستی اطراف شیء مورد نظر قرار بگیرند، از رگرسیون جعبه محدودکننده (BBR یا Bounding Box Regressor) استفاده می‌شود که سعی می‌کند با کاهش تابع هزینه MSE روی مختصات جعبه‌های محدودکننده تعیین شده برای هر شیء، پیش‌بینی دقیقی برای مکان قرارگیری هر شیء در تصویر بدهد.

برای پردازش هر عکس، به ۲۰۰۰ شبکه CNN نیاز است. از آن‌جا که در R-CNN، Regressor Bounding Box هم‌زمان با طبقه‌بندی اشیاء هر جعبه انجام می‌شود، BBR نیز برای هر ناحیه پیشنهادی جداگانه اعمال می‌شود. این مفهوم را در شکل زیر نشان داده شده است:



شکل ۱۶: طبقه بندی اشیاء در R-CNN

همچنین تغییر ابعاد Region Proposalها به یک مقدار ثابت و مشخص، پیش از ورود به یک شبکه عصبی کانولوشنی در شکل مشخص است.

انتخاب نواحی پیشنهادی براساس IoU: تعیین دقیق اینکه هر یک از ۲۰۰۰ جعبه پیشنهادی یک تصویر در مجموعه داده آموزشی (Train dataset)، با کدام یک از جعبه‌های هدف موجود در آن متناظر می‌باشد، کار دشواری است. به صورت شهودی نیز این مشکل قابل فهم است، چرا که اگر یک ناحیه پیشنهادی به طور قابل توجهی از تمامی جعبه‌های هدف که به عنوان Label در مجموعه داده آموزشی قرار گرفته اند دور باشد، تبدیل آن به یک جعبه محدودکننده منطقی نخواهد بود.

به همین دلیل، فقط در شرایطی از یک ناحیه پیشنهادی p استفاده می‌شود که به اندازه کافی به یکی از جعبه‌های هدف g نزدیک باشد. این نزدیکی را با تخصیص p به جعبه هدفی که بیشترین هم‌پوشانی (Intersection over Union) IoU یا اشتراک تقسیم بر اجتماع را با آن دارد تعریف نمود، آن هم به شرطی که میزان هم‌پوشانی بیشتر از یک حد آستانه باشد. این آستانه در مقاله اصلی ۰.۶ تنظیم شده است. تمامی نواحی پیشنهادی‌ای که به هیچ جعبه هدفی تخصیص نیافته‌اند، کنار گذاشته می‌شوند و در فرآیند آموزش استفاده نمی‌شوند. این فرآیند را برای هر کلاس از اشیاء به صورت جداگانه انجام داده می‌شود تا دقت مدل بهبود یابد.

تفکیک کلاس‌ها: برای بهبود دقت مدل در تشخیص این اشیاء، نیاز است که رگرسیون جعبه محدودکننده برای هر کلاس از اشیاء به صورت جداگانه انجام شود. هر کلاس ممکن است خصوصیات و ویژگی‌های خاص خود را داشته باشد. برای مثال، شکل، اندازه، و موقعیت اشیاء مربوط به کلاس «ماشین» ممکن است به طور قابل توجهی با کلاس «انسان» متفاوت باشد. اگر همه جعبه‌های محدود کننده را بدون توجه به کلاس‌های اشیاء با هم ترکیب نمود، ممکن است مدل به درستی نتواند تفاوت‌ها را یاد بگیرد و دقت آن کاهش یابد.

سرکوب غیر حداکثری: در پایان، تکنیک سرکوب غیر حداکثری (Non-Maximum Suppression) برای بدست آوردن نواحی درست استفاده می‌شود. با این تکنیک، جعبه‌های محدودکننده اصلاح شده به کمک BBR که امتیاز همپوشانی (IoU) بیش از ۰.۵ دارند حفظ شده و سایر جعبه‌ها حذف می‌شوند. اگر بیش از یک جعبه محدودکننده با امتیاز بیش از ۰.۵ برای یک شیء بدست آید، جعبه‌ای که بالاترین امتیاز همپوشانی را دارد، پذیرفته می‌شود. سرکوب غیر حداکثری (NMS) یک تکنیک پس پردازشی است که برای کاهش تعداد جعبه‌های محدودکننده استفاده می‌شود. این تکنیک با حذف جعبه‌هایی که همپوشانی زیادی دارند و تنها نگه داشتن جعبه‌ای که بالاترین امتیاز همپوشانی را دارد، به بهبود دقت تشخیص اشیاء کمک می‌کند. NMS با این روش می‌تواند اطمینان حاصل کند که تنها جعبه‌های محدودکننده معتبر و دقیق باقی می‌مانند.

معرفی نسل‌های بعدی الگوریتم R-CNN

موفقیت R-CNN باعث شد محققان به فکر بهبود و ارتقای این الگوریتم بیفتند؛ در نتیجه، ظهور نسل‌های بعدی R-CNN با عملکرد بهتر و سریع‌تر مشاهده می‌شود.

Fast R-CNN: اولین گام برای بهبود R-CNN الگوریتم Fast R-CNN بود. این الگوریتم، با به اشتراک گذاشتن برخی محاسبات، توانست سرعت R-CNN را به شکل چشمگیری افزایش دهد. به زبان ساده، الگوریتم Fast R-CNN به جای آنکه برای هر ناحیه پیشنهادی یک شبکه عصبی کانولوشنال جداگانه اجرا کند، از یک شبکه واحد برای استخراج ویژگی از همه نواحی پیشنهادی به صورت همزمان استفاده کرد. این کار باعث صرفه جویی در زمان و منابع پردازشی شد. از قدرت شبکه‌های عصبی کانولوشنال (CNN) برای استخراج ویژگی استفاده کنید، اما محاسبات اضافی را کاهش دهید. در R-CNN، برای هر ناحیه پیشنهادی، یک گذر جداگانه از CNN انجام می‌شد. این از نظر محاسباتی پرهزینه بود. Fast R-CNN ابتدا با استفاده از یک CNN ویژگی‌های کل تصویر را استخراج می‌کند، سپس بعداً ویژگی‌های مرتبط را برای هر ناحیه پیشنهادی استخراج می‌کند.

معماری Fast R-CNN به شرح زیر می باشد:

- ورودی: ورودی شامل کل تصویر و مجموعه‌ای از نواحی پیشنهادی (جعبه‌های محدودکننده) است که توسط یک الگوریتم خارجی (برای مثال، جستجوی انتخابی) تولید می‌شود.
- استخراج ویژگی کانولوشنال: یک CNN عمیق از پیش آموزش دیده (مانند VGG16) کل تصویر را پردازش می‌کند. این کار یک نقشه ویژگی کانولوشنال ایجاد می‌کند که حاوی اطلاعات غنی در مورد تصویر است.
- استخراج ناحیه مدنظر (RoI Pooling): هر ناحیه پیشنهادی روی نقشه ویژگی اعمال می‌شود سپس یک لایه استخراج ناحیه مدنظر یک بردار ویژگی با اندازه ثابت را از ناحیه مربوط در نقشه ویژگی استخراج می‌کند. این کار اطمینان می‌دهد که همه نواحی برای پردازش بیشتر اندازه ثابتی دارند.
- طبقه‌بندی و رگرسیون جعبه محدودکننده: این لایه‌های کاملاً متصل بردار ویژگی استخراج شده را به عنوان ورودی می‌گیرند. لایه طبقه‌بندی احتمال وجود هر کلاس از اشیاء را در ناحیه پیش‌بینی می‌کند. لایه رگرسیون جعبه محدودکننده جعبه محدودکننده پیشنهادی را برای بهبود دقت آن اصلاح می‌کند.

Fast R-CNN راه را برای الگوریتم‌های تشخیص اشیاء سریع‌تر و کارآمدتر مانند Faster R-CNN هموار کرد. این همچنان یک نقطه عطف چشمگیر در تکامل یادگیری عمیق برای وظایف بینایی کامپیوتر است.

Faster R-CNN: در نهایت، الگوریتم Faster R-CNN پا به عرصه گذاشت که انقلابی دیگر در تشخیص اشیاء به شمار می‌رود. Faster R-CNN یک شبکه عصبی به نام Region Proposal Network (RPN) را معرفی کرد که وظیفه استخراج کننده پیشنهاد

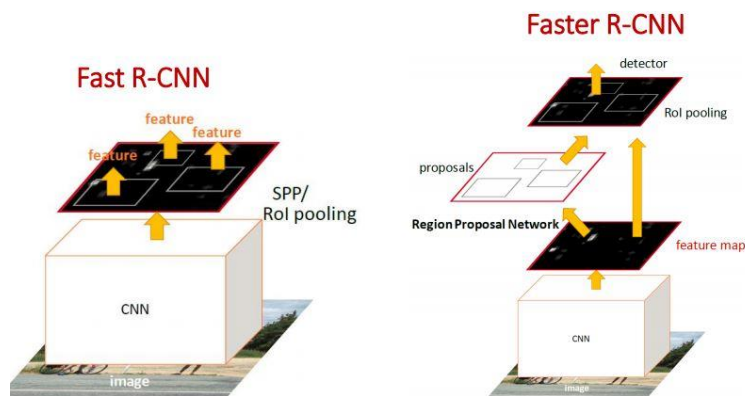
ناحیه را بر عهده گرفت. این شبکه به طور مستقیم به شبکه اصلی متصل است و می تواند به صورت برخط (online) نواحی احتمالی حاوی شیء را پیشنهاد کند. این نوآوری به این می انجامد تا سرعت تشخیص اشیاء با Faster R-CNN به شکل چشمگیری افزایش یابد و در عین حال، دقت تشخیص نیز همچنان در سطح بالایی باقی بماند.

اگرچه Fast R-CNN سرعت تشخیص اشیاء را نسبت به R-CNN به طور قابل توجهی بهبود بخشید، اما همچنان به الگوریتمی مجزا برای پیشنهاد نواحی احتمالی اشیاء (Region Proposal) وابسته بود. Faster R-CNN با ادغام این دو مرحله در یک شبکه واحد انقلابی در زمینه تشخیص اشیاء ایجاد کرد.

ایده اصلی این معماری ادغام پیشنهاد ناحیه و استخراج ویژگی در یک شبکه واحد برای دستیابی به سرعت و دقت بالاست. معماری Faster R-CNN به شرح زیر می باشد:

- شبکه پیشنهاد ناحیه (RPN): یک شبکه عصبی کوچک به نام RPN به ابتدای شبکه اصلی اضافه می شود. RPN روی نقشه ویژگی استخراج شده از تصویر ورودی با CNN پایه (مانند VGG16) اجرا می شود. این شبکه دو خروجی تولید می کند: احتمال اینکه یک ناحیه حاوی شیء باشد (امتیاز شیء) و اصلاحیه (پیشنهاد) برای بهبود دقت جعبه محدودکننده (Bounding Box).
- اشتراک گذاری ویژگی: از خروجی های RPN برای انتخاب نواحی با امتیاز شیء بالا به عنوان نواحی پیشنهادی استفاده می شود. به جای اجرای مجدد CNN پایه برای هر ناحیه پیشنهادی مانند Fast R-CNN، Faster R-CNN از همان نقشه ویژگی استخراج شده توسط RPN استفاده می کند. این کار صرفه جویی قابل توجهی در محاسبات را رقم می زند.
- استخراج ویژگی برای نواحی پیشنهادی: برای هر ناحیه پیشنهادی یک لایه استخر ناحیه مدنظر (RoI Pooling) ناحیه مربوط را از نقشه ویژگی استخراج می کند. این ناحیه به بردار ویژگی با اندازه ثابت تبدیل می شود.
- طبقه بندی و رگرسیون جعبه محدودکننده: مشابه Fast R-CNN، این مرحله ها با استفاده از لایه های کاملاً متصل روی بردارهای ویژگی استخراج شده انجام می شوند. خروجی نهایی فهرستی از اشیای موجود در تصویر به همراه برچسب، امتیاز اطمینان و جعبه محدودکننده دقیق آن هاست.

به طور کلی، Faster R-CNN یک الگوریتم پیشگام در زمینه تشخیص اشیاست که سرعت و دقت را به طور چشمگیری بهبود بخشیده است. این الگوریتم پایه ای برای بسیاری از مدل های تشخیص اشیای پیشرفته امروزی به شمار می رود. [۱۹]



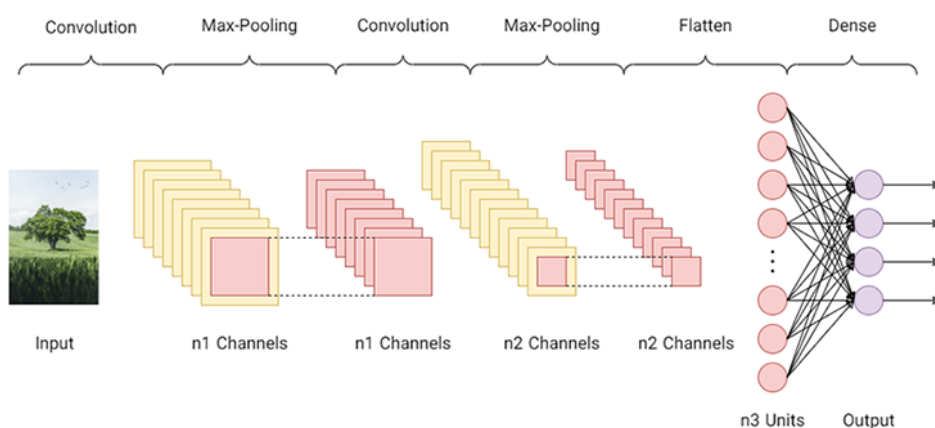
شکل ۱۷: Fast R-CNN و Faster R-CNN

۵- شبکه های عصبی پیچشی

شبکه های عصبی پیچشی یا همگشتی (convolutional neural network) یا ConvNet رده ای از شبکه های عصبی ژرف هستند که معمولاً برای انجام تحلیل های تصویری یا گفتاری در یادگیری ماشین استفاده می شوند. شبکه های عصبی پیچشی به منظور کمینه کردن پیش پردازش ها از گونه ای از پرسپترون های چندلایه استفاده می کنند. ساختار شبکه های پیچشی از فرایندهای زیستی قشر بینایی گربه الهام گرفته شده است. این ساختار به گونه ای است که تک نرون ها تنها در یک ناحیه محدود به تحریک

پاسخ می‌دهند که به آن ناحیه پذیرش گفته می‌شود. نواحی پذیرش نورون‌های مختلف به صورت جزئی با هم همپوشانی دارند به گونه ای که کل میدان دید را پوشش می‌دهند. شبکه‌های عصبی پیچشی نسبت به بقیه رویکردهای دسته‌بندی تصاویر به میزان کمتری از پیش‌پردازش استفاده می‌کنند. این امر به معنی آن است که شبکه معیارهایی را فرامی‌گیرد که در رویکردهای قبلی به صورت دستی فراگرفته می‌شدند. این استقلال از دانش پیشین و دستکاری‌های انسانی در شبکه‌های عصبی پیچشی یک مزیت اساسی است. لایه‌های پیچشی یک عمل پیچش را روی ورودی اعمال می‌کنند، سپس نتیجه را به لایه بعدی می‌دهند. این پیچش در واقع پاسخ یک تک‌نورون را به یک تحریک دیداری شبیه‌سازی می‌کند. هر نورون پیچشی داده‌ها را تنها برای ناحیه پذیرش خودش پردازش می‌کند. مشبک کردن به شبکه‌های پیچشی این اجازه را می‌دهد که انتقال، دوران یا اعوجاج ورودی را تصحیح کنند.

اگرچه شبکه‌های عصبی پیش‌خور کاملاً همبند می‌توانند برای یادگیری ویژگی‌ها و طبقه‌بندی داده به کار روند، این معماری در کاربرد برای تصاویر به کار نمی‌رود. در این حالت حتی برای یک شبکه کم‌عمق تعداد بسیار زیادی نورون لازم است. عمل پیچش یک راه‌حل برای این شرایط است. [۱۲]



شکل ۱۸: ساختار شبکه عصبی پیچشی [۱۴]

در عملگر کانولوشن یا پیچش (Convolution)، مطابق چهار مولفه مهم وجود دارد که عبارتند از [۱۳]:

- ماتریس یا تصویر ورودی (Input)
- فیلتر یا کرنل کانولوشنی (Convolution Filter)
- عملگر کانولوشن (*)
- ویژگی خروجی کانولوشن (Output)

ورودی																													
<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>۱</td><td>۲</td><td>۳</td><td>۱</td><td>۴</td></tr> <tr><td>۵</td><td>۸</td><td>۹</td><td>۲</td><td>۱</td></tr> <tr><td>۵</td><td>۰</td><td>۱</td><td>۲</td><td>۷</td></tr> <tr><td>۸</td><td>۶</td><td>۴</td><td>۱</td><td>۰</td></tr> <tr><td>۶</td><td>۲</td><td>۰</td><td>۵</td><td>۴</td></tr> </table>	۱	۲	۳	۱	۴	۵	۸	۹	۲	۱	۵	۰	۱	۲	۷	۸	۶	۴	۱	۰	۶	۲	۰	۵	۴	*	فیلتر	=	خروجی
۱	۲	۳	۱	۴																									
۵	۸	۹	۲	۱																									
۵	۰	۱	۲	۷																									
۸	۶	۴	۱	۰																									
۶	۲	۰	۵	۴																									
		<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>۱</td><td>-۱</td><td>۴</td></tr> <tr><td>-۲</td><td>۱</td><td>۰</td></tr> <tr><td>۰</td><td>۲</td><td>۳</td></tr> </table>	۱	-۱	۴	-۲	۱	۰	۰	۲	۳		<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>۱۲</td><td>۴</td><td>۲۷</td></tr> <tr><td>۴۷</td><td>۱۹</td><td>۱۳</td></tr> <tr><td>۱۷</td><td>۱۴</td><td>۴۲</td></tr> </table>	۱۲	۴	۲۷	۴۷	۱۹	۱۳	۱۷	۱۴	۴۲							
۱	-۱	۴																											
-۲	۱	۰																											
۰	۲	۳																											
۱۲	۴	۲۷																											
۴۷	۱۹	۱۳																											
۱۷	۱۴	۴۲																											

شکل ۱۹: عملگر کانولوشن یا پیچش

عملگر کانولوشن (*)، کرنل یا فیلتر کانولوشنی را برمی‌دارد و روی تصویر یا ماتریس ورودی می‌لغزند. به عبارتی دیگر، کرنل یا فیلتر روی تصویر حرکت می‌کند یا تصویر ورودی را اسکن می‌کند. با قرار گرفتن فیلتر روی هر بخش از تصویر، اعداد در فیلتر درایه به درایه در پیکسل‌های متناظر تصویر ضرب می‌شوند. در نهایت همه اعداد باهم جمع می‌شوند. فیلتر به دنبال پیدا کردن نواحی مشابه خود در تصویر است و هر جایی ناحیه مشابه خود را پیدا کرد بلند فریاد می‌زند (عدد بزرگ). پس کانولوشن منجر

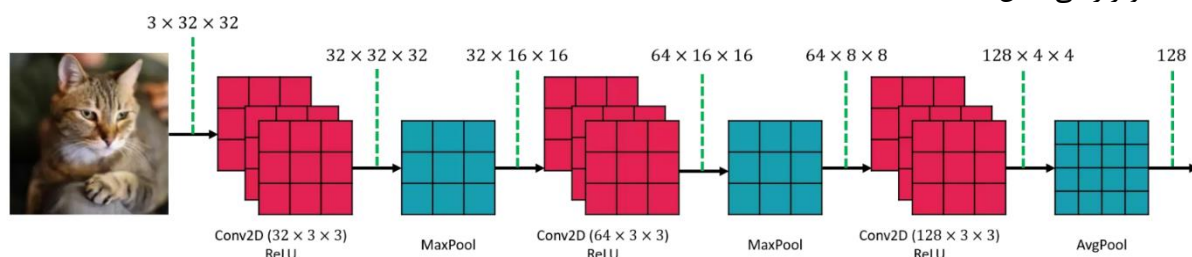
به یافتن الگوهای خاص در تصویر با توجه به فیلتر می‌شود. اعداد موجود در فیلتر بسیار مهم هستند. در نوروں پارامترها متغیر بودند و از طریق فرآیند آموزش بدست می‌آمدند. اینجا هم اعداد موجود در فیلتر از طریق فرآیند آموزش بدست می‌آیند. البته، قبل از پیدایش شبکه عصبی کانولوشن، از کانولوشن در پردازش تصویر و سیگنال بسیار زیاد استفاده می‌شد. اما اعداد فیلترهای کانولوشنی ثابت بودند و توسط یک متخصص این اعداد طراحی می‌شدند.

هسته اصلی شبکه CNN لایه کانولوشنی است که درصد اعظم محاسبات شبکه عصبی کانولوشن را به خود اختصاص داده است. هر لایه کانولوشن در شبکه عصبی کانولوشن شامل مجموعه‌ای فیلتر است و از کانولوشن بین فیلترها و لایه ورودی است که خروجی ساخته می‌شود. به خروجی لایه کانولوشنی، فیچر مپ (Feature Map) گفته می‌شود. تعداد فیلترها، تعداد ویژگی‌های تشخیص داده شده را نشان می‌دهد. این هاپیر پارامتر معمولاً نمایی از ۲ (معمولاً بین ۳۲ تا ۴۰۹۶) انتخاب می‌شود. معمولاً استفاده از فیلترهای بیشتر باعث ایجاد یک شبکه عصبی قدرتمندتر می‌شود. اما این افزایش پارامترها ممکن است باعث overfitting شود. اندازه فیلتر معمولاً به صورت مربعی (برای تصاویر) و 3×3 تعریف می‌شود. اما 5×5 و 7×7 نیز گاهی استفاده می‌شود. با فیلترهای کوچک، تعداد پارامترهای قابل یادگیری بسیار کمتر می‌شود. بعد از فیلتر کردن یک ماتریس 5×5 با فیلتر 3×3 ، سایز خروجی دو سطر و ستون کمتر از ماتریس ورودی داشت. با لایه گذاری یا padding می‌توان به صورت جعلی اندازه ورودی را افزایش داد تا ماتریس خروجی هم‌اندازه ماتریس ورودی بشود. در مورد لایه گذاری صحبت بسیار است و معمولاً در پردازش تصویر بررسی می‌شود. اما یک راه ساده و رایج آن اضافه کردن سطر و ستون صفر به صورت متقارن به دور ماتریس ورودی است. به لایه گذاری صفر zero padding گفته می‌شود.

لایه تجمعی با پولینگ (Pooling layer) یکی دیگر از لایه های مهم در شبکه عصبی کانولوشن است. هدف لایه پولینگ کاهش اندازه مکانی فیچر مپ بدست آمده با استفاده از لایه کانولوشنی است. لایه پولینگ پارامتر قابل آموزش ندارد. صرفاً یک نمونه برداری ساده و موثر انجام می‌دهد. پولینگ عملکردی شبیه کانولوشن دارد و یک پنجره روی تصویر حرکت می‌کند. رایج‌ترین نمونه پولینگ max pooling و average pooling است. تجمع Max Pooling مقدار بیشینه را از قسمتی از تصویر بازمی‌گرداند که توسط کرنل پوشش داده شده است. از سوی دیگر، Average Pooling میانگین همه مقادیر را از قسمتی از تصویر باز می‌گرداند که توسط کرنل پوشش داده شده است.

معمولاً آخرین لایه های یک شبکه عصبی کانولوشن برای طبقه بندی را لایه های فولی کانکتد (Fully Connected Layer) تشکیل می‌دهند. یکی از کاربردهای اصلی لایه فولی کانکتد در شبکه کانولوشن، استفاده به عنوان طبقه بند یا کلاسیفایر (Classifier) است. یعنی مجموعه ویژگی‌های استخراج شده با استفاده از لایه های کانولوشنی، در نهایت تبدیل به یک بردار می‌شوند. [۱۳]

تصویر رنگی از سه صفحه تشکیل شده است. این صفحات عبارتند از صفحه قرمز (R)، صفحه سبز (G) و صفحه آبی (B). مشخصات هر صفحه مشابه همان تصویر سطح خاکستری است و مقادیر اعداد بین ۰ تا ۲۵۵ است. از ترکیب سه عدد، یک رنگ نهایی حاصل می‌شود. به همین خاطر به تصاویر رنگی، تصویر RGB گفته می‌شود. در شکل زیر یک شبکه عصبی کانولوشن ساده برای یک تصویر رنگی نشان داده شده است [۱۳]



شکل ۲۰: شبکه عصبی کانولوشن ساده [۱۳]

مزیت بزرگ استفاده از CNNها این است که نیازی به انجام پیش‌پردازش زیادی روی تصاویر نیست. در بیشتر الگوریتم‌هایی که پردازش تصویر را انجام می‌دهند، فیلترها معمولاً توسط یک مهندس بر اساس روش‌های اکتشافی (heuristic) ایجاد می‌شوند.

CNNها می توانند مهم ترین ویژگی فیلترها را بیاموزند و چون به پارامترهای زیادی احتیاج نیست، صرفه جویی زیادی در وقت و عملیات آزمون و خطا صورت می گیرد. تا زمانی که با تصاویر با ابعاد بالا که هزاران پیکسل دارند کار نکنید، این صرفه جویی چندان به چشم نمی آید. هدف اصلی الگوریتم CNN این است که با حفظ ویژگی هایی که برای فهم آنچه داده ها نشان می دهند مهم هستند، داده ها را به فرم هایی که پردازش آن ها آسان تر است، درآورد. [۱۴]

۶- کاربردهای بینایی کامپیوتر در حوزه پزشکی

در تصویربرداری پزشکی فرد در دستگاه مورد نظر قرار می گیرد و تصاویر استخراج می شود. این تصاویر به کمک پردازش تصویر و یادگیری عمیق می تواند با تصاویر قدیمی و یکسری اجزای مشخص مقایسه شده و بیماری به طور تقریبی تشخیص داده شود. این مورد یکی از بخش هایی است که ممکن است در زمان کوتاهی توسط هوش مصنوعی به طور کامل پوشش داده شود. معمولاً پزشکان به این صورت عمل می کنند که تصویر را با توجه به یک سری اجزای مشخص و تجربه موارد مشابه تجزیه و تحلیل می کنند و نتیجه را به فرد اعلام می شود. حال تصور کنید که یک ماشین با تجربه دیدن چند ده هزار تصویر افراد متفاوت بتواند تصویر فرد را هم بررسی کند. به این ترتیب احتمال تشخیص دقیق تر بسیار بالا می رود [۱۸]. بینایی کامپیوتر، کاربردهای متعددی در حوزه پزشکی پیدا کرده است و روشی را که متخصصان پزشکی در تشخیص، درمان و نظارت بر بیماران استفاده می کنند، متحول کرده است. در این جا برخی از کاربردهای کلیدی بینایی کامپیوتر در حوزه مراقبت های بهداشتی را باهم بررسی می شود.

i. تجزیه و تحلیل تصویربرداری پزشکی: الگوریتم های بینایی کامپیوتری می توانند تصاویر پزشکی مانند اشعه ایکس، سی تی اسکن، MRI و ماموگرافی را برای کمک به تشخیص تجزیه و تحلیل کنند. آن ها می توانند ناهنجاری ها را تشخیص دهند، به تشخیص زودهنگام بیماری هایی مانند سرطان کمک کنند و اندازه گیری های کمی را برای ارزیابی دقیق ارائه دهند.

ii. راهنمای جراحی: سیستم های بینایی کامپیوتری می توانند با ارائه بازخورد و راهنمایی بصری در لحظه به جراحان در طول عمل کمک کنند. آن ها می توانند تصاویر واقعیت افزوده (AR) را در زمینه جراحی پوشش دهند، بافت ها و ساختارهای مهم را برجسته کنند یا اقدامات جراح را هدایت کنند، که منجر به افزایش دقت عمل می شود.



شکل ۲۱: راهنمای جراحی

iii. تشخیص و غربالگری بیماری: بینایی کامپیوتری می تواند فرآیندهای تشخیص و غربالگری بیماری را خودکار کند. این کار می تواند تصاویر یا فیلم های پزشکی را برای شناسایی بیماری هایی مانند رتینوپاتی دیابتی، سرطان پوست یا بیماری های ریه تجزیه و تحلیل کند. این امر امکان تشخیص زودهنگام، تشخیص سریع تر و کاهش بار بر روی متخصصان مراقبت های بهداشتی را فراهم می کند.

iv. مانیتورینگ بیمار از راه دور: تکنیک های بینایی کامپیوتری امکان نظارت از راه دور وضعیت بیماران را با استفاده از دوربین ها یا حسگرها فراهم می کنند. آن ها می توانند علائم حیاتی را ردیابی کنند، حرکات را زیر نظر بگیرند و هرگونه تغییر را تشخیص دهند، و به ارائه دهندگان مراقبت های بهداشتی این امکان را می دهند تا بیماران را از راه دور ارزیابی کنند و مداخلات به موقع، به ویژه برای افراد مسن یا مبتلایان به بیماری های مزمن ارائه کنند.

v. تجزیه و تحلیل رفتاری: الگوریتم های بینایی کامپیوتری می توانند حالات چهره، حرکات بدن و ژست ها را برای ارزیابی



حالات عاطفی، سطح درد یا اختلالات شناختی بیماران تجزیه و تحلیل کنند. این اطلاعات می تواند به تشخیص سلامت روان، مدیریت درد و نظارت بر بیماران مبتلا به اختلالات عصبی کمک کند.

vi. پوشیدنی های نظارت بر سلامت: بینایی رایانه ای می تواند قابلیت های دستگاه های پوشیدنی را افزایش دهد. به عنوان مثال، ساعت های هوشمند با دوربین های داخلی می توانند علائم حیاتی را کنترل کنند، سطح اکسیژن خون را اندازه گیری کنند یا با تجزیه و تحلیل تصاویر یا ویدیو، شرایط پوست را تشخیص دهند.

vii. کشف و توسعه دارو: تکنیک های بینایی رایانه ای می توانند فرآیند کشف دارو را با تجزیه و تحلیل مقادیر وسیعی از داده ها، از جمله ساختارهای مولکولی، برای پیش بینی تداخلات دارو-هدف یا شناسایی گزینه های مناسب دارویی تسریع بخشند. این می تواند منجر به تولید داروی کارآمدتر و هدفمندتر شود. [۲]

viii. کشف های پیشگیری از سقوط: یکی از دستاوردهای بینایی ماشین در پزشکی، ساخت کشف های پیشگیری از سقوط است. از آن رو که سالمندان هنگام به زمین خوردن، آسیب زیادی را به جسم و روح خود وارد می کنند، دانشمندان در تایوان، کشف های پیشگیری از سقوط با استفاده از سیستم تشخیص موانع خطی مبتنی بر دوربین را ساخته اند که یک سیستم تشخیص موانع با لیزر خط جالب را برای جلوگیری از افتادن افراد مسن طراحی می کند.

ix. سونوگرافی متمرکز با شدت بالا: سونوگرافی متمرکز با شدت بالا (HIFU) برای فرسایش ایمن بافت های بدخیم و خوش خیم و به عنوان عاملی برای تحویل دارو پیشنهاد شده، در حالی که MRI برای راهنمایی و نظارت برای درمان در نظر گرفته شده است.

x. سیستم دستیار تعاملی برای بیماران مبتلا به زوال عقل: از آن رو که سالمندان در بیشتر مواقع، فاقد توانایی خودمراقبتی هستند، لذا به منظور توسعه یک سیستم دستیار تعاملی کارآمد و راحت برای پرستاران و بیماران مبتلا به زوال عقل، دانشمندان در چین یک سیستم مراقبت تعاملی بر اساس تصویر عمقی و EEG برای بیماران سالخورده مبتلا به زوال عقل، طراحی کرده اند. این سیستم به عنوان یکی از دستاوردهای کارا در فناوری بینایی ماشین در پزشکی، به گونه ای طراحی شده که مبتنی بر شبکه های عصبی عمیق چندوجهی برای بیمار مبتلا به زوال عقل با نیازهای ویژه باشد.

xi. ردیابی از دست دادن خون پس از جراحی: یکی دیگر از کاربردهای بینایی ماشین در پزشکی ردیابی از دست دادن خون پس از جراحی است. در واقع، بینایی ماشین با ردیابی از دست دادن خون بیمار در زمان واقعی و تخمین میزان از دست دادن خون با دقت بالاتری نسبت به پزشکان، پتانسیل کاهش میزان مرگ و میر پس از جراحی را دارد. [۱۵]

xii. پردازش تصویر در مهندسی ژنتیک: تصویربرداری مولکولی روش نوینی برای مشاهده پدیده های سلولی و مولکولی مانند بقای سلول، مهاجرت، تکثیر و حتی تمایز در سطح کل بدن بدون ایجاد آسیب یا جراحی، ارائه می دهد. برای نظارت بر پیوندهای سلولی درون زیستی، از روش ها و رویکردهای مختلف تصویربرداری و پردازش تصویر برای بررسی وضعیت استفاده می شود. [۱۶]

xiii. بررسی حرکات اصلاحی افراد و کمک به انجام تمرینات: بخشی از کارهای درمانی افراد، استفاده از حرکات اصلاحی یا انجام تمرینات ورزشی است. این نکته بسیار مهم است که این حرکات به صورت اصولی انجام شود زیرا در صورت اشتباه بودن حرکت ممکن است به بیمار آسیب وارد کند. از طرفی معمولاً افرادی که به این تمرینات نظارت می کنند نمی توانند وقت کافی برای هر فرد اختصاص دهند. پردازش تصویر می تواند در حرکات مختلف مشکلات کار را به فرد گوشزد کرده و به عنوان یک دستیار برای فرد و مربی عمل کند.

xiv. تشخیص بیماری پوستی: فرد می تواند با استفاده از دوربین مناسب پوست خود را با دقت بالایی اسکن کنید و متوجه بیماری ها یا کمبودهای پوست خود شوید. این مورد در بحث تشخیص بیماری و هم چنین در مقوله ی زیبایی کاربردهای فراوانی دارد. در این مورد هم بهتر است تا این مورد به عنوان دستیار پزشک عمل کند. [۱۷]

۶-الف) پردازش تصویر در پزشکی

پردازش تصویر در پزشکی Medical Image Processing اجازه می دهد بدون نیاز به جراحی، عمیق و دقیق به داخل بدن نگاه

نمود. این فناوری می‌تواند مدل‌های سه‌بعدی از قسمت‌های مختلف بدن را بسازد تا کمک کند درمان‌ها را بهبود بخشد، ابزارهای پزشکی جدیدی طراحی نماید و تشخیص‌های دقیق‌تری ارائه نماید. در واقع پردازش تصویر یکی از ابزارهای اصلی پیشرفت در پزشکی در سال‌های اخیر است. از تکنیک‌های رادیولوژیکی، سونوگرافی، تصویربرداری با رزونانس مغناطیسی (MRI) و توموگرافی کامپیوتری (CT) گرفته تا نوآوری‌های اخیر در یادگیری ماشین و هوش مصنوعی، همگی کاربرد پردازش تصویر در پزشکی هستند.



شکل: بکارگیری بینایی ماشین در پردازش تصویر

موارد زیر مهم‌ترین مزایایی است که medical image processing برای علم پزشکی به ارمغان آورده است:

- با پردازش تصویر مطالعه آناتومی انسان به صورت عمیق و دقیق انجام می‌شود، در حالی که نیازی به استفاده از روش‌های تهاجمی نیست.
- بر پایه داده‌های به دست آمده، مدل‌های سه‌بعدی از ساختارهای آناتومیک ایجاد می‌شوند و برای بهبود نتایج درمان بیمار، بهینه‌سازی فرآیند تشخیص، بهبود دستگاه‌های پزشکی موجود و مکانیزم انتقال دارو به کانون پاتولوژی، بیشتر مورد مطالعه قرار می‌گیرند.
- پردازش تصاویر پزشکی، همراه با بهبود مداوم در کیفیت تصویر، بازنمایی دیجیتال دقیقی از بافت‌ها و ارگان‌ها را فراهم می‌کند. با دانستن هندسه دقیق، اندازه‌گیری‌ها و داده‌های تحلیل آماری می‌توان به درک عمیق و دقیقی از تاثیر دستگاه‌های پزشکی بر آناتومی بیمار دست یافت.
- نتایج پردازش تصویر پزشکی برای برنامه‌ریزی عمل‌های جراحی ضروری است. ارزیابی عملکرد ایمپلنت‌ها اهمیت زیادی برای توسعه بیشتر ایمپلنتولوژی دارد.

فرآیند پردازش تصویر پزشکی با گرفتن داده‌های خام از تصاویر CT یا MRI شروع می‌شود و آن‌ها را به فرمتی بازسازی می‌کند که برای استفاده در نرم‌افزارهای مرتبط مناسب باشد. ورودی معمول برای پردازش تصویر یک نقشه سه‌بعدی از شدت‌های خاکستری که شامل یک شبکه واکسل (voxel) یا همان پیکسل‌های سه‌بعدی است. شدت خاکستری اسکن CT بستگی به جذب اشعه ایکس دارد، در حالی که در MRI، توسط قدرت سیگنال‌ها از ذرات پروتون در طول استراحت‌های پس از اعمال میدان‌های مغناطیسی بسیار قوی، تعیین می‌شود. برای کاربران پزشکی، حجم تصویر بازسازی شده معمولاً برای جدا کردن و ویرایش مناطق مختلف آناتومیک مورد علاقه مانند بافت و استخوان پردازش می‌شود. در نرم‌افزارهای پردازش تصویر پزشکی کاربران می‌توانند عملیات مختلف پردازش تصویر را در سطح ۲ بعدی و ۳ بعدی انجام دهند که شامل موارد زیر می‌شود:

- کاهش و حذف نویزها یا آرتیفکت‌های ناخواسته با فیلترهای تصویر
- برش و نمونه‌برداری مجدد داده‌های ورودی برای پردازش آسان‌تر و سریع‌تر تصاویر
- استفاده از ابزارهای سگمنتیشن برای شناسایی مناطق آناتومیک مختلف، از جمله تکنیک‌های خودکار با استفاده از الگوریتم‌های یادگیری ماشینی مبتنی بر هوش مصنوعی
- اعمال ابزارهای اندازه‌گیری و آمار برای کمی کردن بخش‌های مختلف داده‌های تصویر، به عنوان مثال، مرکز خط‌ها



- وارد کردن مدل های CAD، مانند ایمپلنت ها یا دستگاه های پزشکی، برای مطالعه تعامل آن ها با آناتومی های فردی
- صادر کردن مدل های پردازش شده برای شبیه سازی مبتنی بر فیزیک، کارهای طراحی بیشتر، یا برای چاپ سه بعدی نمونه های فیزیکی از آناتومی بدن [۱۶]

۶-ب) پردازش تصویر در مهندسی پزشکی

مهندسی بیومدیکال یا مهندسی پزشکی یعنی کاربرد ایده ها و مفاهیم طراحی مهندسی در پزشکی و بیولوژی برای اهداف بهداشتی و درمانی. حوزه های تخصصی مهندسی بیومدیکال شامل موارد زیر است:

- بیو اینسترومنتیشن
- بیومتریال ها
- بیومکانیک
- مهندسی سلولی
- بافتی و ژنتیکی
- مهندسی بالینی
- تصویربرداری پزشکی
- ارتوپدی
- مهندسی توانبخشی

در مهندسی بیومدیکال، برای تهیه تصاویر برای کاربردهای تشخیصی و درمانی از حسگرهای پیشرفته و فناوری کامپیوتری استفاده می شود. درواقع مجموعه ای از ساختارهای اطلاعاتی آناتومیک که با بررسی تصاویر بیومدیکال فراهم می شود، به بررسی و مشاهده و در نهایت درمان کمک می کند. برای آماده سازی دستگاه های بیومدیکال، ایمپلنت های دقیق، مفصل و جایگزینی سایر ارگان ها، نیاز به تصاویر باکیفیت است. به عنوان مثال برخی از کاربردهای بیوانفورماتیک تصویری عبارتند از:

- تجزیه و تحلیل پرسرعت فتوتیپ های سلولی
- ساخت اطلس برای مدل های سازمانی بدن
- درک فرآیندهای پویا در سلول ها و موجودات زنده
- پردازش تصویر در مهندسی بیومتریال و بافت

با پیر شدن جمعیت، نیاز به تعمیر، جایگزینی و بازسازی بافت های سخت و نرم بدن با استفاده از بیومتریال ها بیشتر می شود. برای بررسی دقیق اجزای داخلی بدن، تصویربرداری و پردازش تصویر نقش بسیار مهمی در این زمینه پیدا می کند. [۱۶]

۶-پ) پردازش تصویر برای تشخیص سرطان

پردازش تصویر در حوزه انکولوژی (سرطان شناسی) به عنوان یک ابزار قدرتمند برای تشخیص دقیق، برنامه ریزی درمانی موثر و نظارت بر پیشرفت درمان عمل می کند، که همگی به افزایش شانس بهبودی بیماران کمک می کند.

- تصویربرداری دقیق از تومورها: با استفاده از تکنیک های پیشرفته مانند CT اسکن، MRI و ماموگرافی، پزشکان می توانند تصاویر دقیقی از تومورها و ساختارهای اطراف آن ها به دست آورند. پردازش این تصاویر به شناسایی و تعیین موقعیت دقیق تومورها کمک می کند.

- مکان یابی تومور: شایع ترین انواع تومور در انکولوژی، تومورهای خوش خیم، پیش سرطانی و بدخیم هستند. تکنیک های اولیه تصویربرداری به کاهش مرگ و میر ناشی از سرطان کمک می کند. پیش پردازش، تقسیم بندی و عملیات مورفولوژیک سه مرحله پردازش تصویر تومور بوده و هدف از آن تعیین مکان تومور است. تقسیم بندی، شناسایی و استخراج ناحیه تومور از تصاویر حاصل از روش های پردازش تصویر راحت تر است که به رادیولوژیست ها یا متخصصان بالینی برای برنامه ریزی درمان کمک می کند. واضح است که بدون یک تصویر با کیفیت بالا، تمام درمان های پرتوی به صورت نادرست انجام می شود و ممکن است مرگ و میر ناشی از سرطان را افزایش دهد. با استفاده از الگوریتم های پردازش



تصویر، می‌توان موقعیت، اندازه و شکل تومورها را با دقت بالا تعیین کرد، که این اطلاعات برای برنامه‌ریزی دقیق‌تر درمان‌هایی مانند جراحی یا رادیوتراپی ضروری است. در روش‌های درمانی مانند رادیوتراپی، پردازش تصویر به متخصصین اجازه می‌دهد تا میزان و مکان دقیق دوز اشعه‌ای که باید به بافت‌های بیمار وارد شود را تعیین کنند. همچنین امکان پیگیری تغییرات تومور در طول زمان را فراهم می‌کند.

- تشخیص زودهنگام: تکنیک‌های پردازش تصویر می‌توانند در شناسایی تومورهای بسیار کوچک که با روش‌های سنتی قابل تشخیص نیستند، به پزشکان کمک کنند. این امر به تشخیص زودهنگام و افزایش شانس درمان موفقیت‌آمیز کمک می‌کند.
- تجزیه و تحلیل پیشرفته: با استفاده از نرم‌افزارهای پیشرفته پردازش تصویر، پزشکان می‌توانند اطلاعات بیشتری از تصاویر به دست آورند، مانند تخمین حجم تومور، تعیین مرزهای دقیق تومور و تشخیص نوع بافت.
- برنامه‌ریزی درمانی دقیق: پردازش تصویر در روش‌های درمانی مانند رادیوتراپی، به پزشکان کمک می‌کند تا دوز دقیق اشعه‌ای که به تومور وارد می‌شود را تعیین کنند، به طوری که به بافت‌های سالم اطراف آسیب کمتری برسد.
- نظارت بر پیشرفت درمان: پردازش تصویر به پزشکان امکان می‌دهد تا اثربخشی درمان‌ها را با مقایسه تصاویر قبل و بعد از درمان ارزیابی کنند. این امر به آنها کمک می‌کند تا در صورت لزوم تغییراتی در برنامه درمانی ایجاد کنند.
- راهنمایی در جراحی: در جراحی‌های حذف تومور، پردازش تصویر می‌تواند به جراحان کمک کند تا محل دقیق تومور را شناسایی کرده و اطمینان حاصل کنند که تمامی تومور برداشته شده است، همچنین به حفظ بافت‌های سالم اطراف کمک می‌کند. [۱۶]

۷- کاربردهای یادگیری عمیق در حوزه پزشکی

تکنیک‌های یادگیری عمیق از داده‌های ذخیره شده در سوابق سیستم‌های سلامت الکترونیک برای رفع بسیاری از نگرانی‌های مورد نیاز مراقبت‌های پزشکی، مانند کاهش میزان تشخیص نادرست و پیش‌بینی نتیجه مراحل استفاده می‌کنند. با پردازش مقادیر زیادی از منابع مختلف مانند تصویربرداری پزشکی، شبکه‌های عصبی مصنوعی می‌توانند به پزشکان در تجزیه و تحلیل اطلاعات و تشخیص چندین بیماری کمک کنند:

- بررسی سطح گلوکز در بیماران دیابتی: بیماران دیابتی را می‌توان از نظر سطح گلوکز کنترل کرد. یک مدل یادگیری عمیق می‌تواند برای پیش‌بینی زمان افزایش و کاهش سطح گلوکز خون بیماران استفاده کرده و به آن‌ها اجازه می‌دهد تا با خوردن یک میان‌وعده پر قند یا تزریق انسولین، عکس‌العمل نشان دهند. [۲۱]
- تشخیص مشکلات قلبی: در سال ۲۰۰۶ هزینه بستری افرادی که به بیماری قابل پیش‌گیری دچار شده بودند در آمریکا به ۳۰ میلیارد دلار رسید. نیمی از بیماران بستری شده از دو بیماری رنج می‌برند: مشکلات قلبی و دیابت. از یادگیری عمیق می‌توان برای بهبود میزان تشخیص و مدت زمان لازم برای ایجاد پیش‌آگاهی استفاده کرد. این امر می‌تواند تعداد بستری شدگان را به شدت کاهش دهد. [۲۱]
- تشخیص سلول‌های سرطانی و تشخیص سرطان: انکولوژیست‌ها سالهاست که از روشهای تصویربرداری پزشکی مانند توموگرافی کامپیوتری، تصویربرداری با رزونانس مغناطیسی و اشعه ایکس استفاده می‌کنند. در حالی که ثابت شده است که این سیستم‌ها برای بسیاری از انواع سرطان مؤثر هستند؛ تعداد زیادی از بیماران از سرطان‌هایی رنج می‌برند که با این دستگاه‌ها قابل تشخیص نیستند. شبکه‌های عصبی مانند شبکه‌های عصبی پیچشی، وعده‌های مربوط به آینده تشخیص سرطان را محقق می‌کنند. براساس همان تصاویر پزشکی، شبکه‌های عصبی مصنوعی می‌توانند سرطان را در مراحل اولیه با درصد خطای کمتر تشخیص داده و نتایج بهتری را برای بیماران فراهم کنند. به تازگی، دانشمندان موفق به آموزش مدل‌های مختلفی از یادگیری عمیق برای تشخیص انواع مختلف سرطان با دقت بالا شده‌اند. یک مدل یادگیری عمیق توانسته است ۸۵ درصد از میزان تشخیص نادرست سرطان سینه را کاهش دهد.



دانشمندان گوگل یک مدل از شبکه عصبی پیچشی ایجاد کرده اند که سرطان پستان متاستاز شده را از تصاویر آسیب شناسی سریعتر و با دقت بیشتری تشخیص می دهد. این مدل به موفقیت ۹۹ درصدی دست یافت. [۲۱]

- استفاده از یادگیری عمیق در سوابق سلامت الکترونیکی: سیستم های سلامت الکترونیکی، داده های بیمار از قبیل اطلاعات دموگرافیک، سوابق پزشکی و نتایج آزمایش ها را ذخیره می کنند. این سیستم ها با استفاده از الگوریتم های یادگیری عمیق میزان تشخیص صحیح و مدت زمان لازم برای تشخیص بیماری را بهبود می بخشند. این الگوریتم ها از داده های ذخیره شده در سیستم های سلامت الکترونیک استفاده می کنند تا الگوهای روند سلامت و عوامل خطر را تشخیص داده و براساس الگوهای شناسایی شده، نتیجه بگیرند. همچنین محققان می توانند از داده های موجود در سیستم های سلامت الکترونیک در راستای ایجاد مدل هایی در یادگیری عمیق استفاده کنند که احتمال بروز برخی از نتایج مرتبط با سلامتی را پیش بینی می کند. [۲۱]

- ویروس نقص سیستم ایمنی انسان (HIV) : بیش از ۳۶ میلیون نفر در سراسر جهان از ویروس نقص سیستم ایمنی بدن رنج می برند. این افراد برای درمان وضعیت خود نیاز به دریافت دوز روزانه داروهای ضد ویروس دارند. HIV می تواند به سرعت جهش یابد. بنابراین، برای ادامه درمان HIV، باید داروهای تجویز شده برای بیماران را تغییر داد. استفاده از یک الگوی یادگیری عمیق به نام یادگیری تقویتی می تواند در مقابله با این نوع ویروس ها کمک کند. در این روش، مدل پیچشی می تواند بسیاری از نشانگرهای زیستی را با استفاده از هر دوز دارو ردیابی کرده و بهترین مسیر عملی را برای درمان مداوم فراهم کند. [۲۱]

- تشخیص بیماری و پیش بینی درمان: یادگیری عمیق در پزشکی، برای توسعه الگوریتم هایی ایفای نقش می کند که می توانند بیماری ها را تشخیص داده و اثربخشی درمان ها را پیش بینی کنند. کاربردهای این زمینه شامل تشخیص سرطان و پزشکی شخصی است [۱۰]. تیمی از دانشمندان، مدلی از شبکه های عصبی را آموزش دادند تا ۱۷ بیماری مختلف را براساس بوی تنفس بیماران با دقت ۸۵ درصد شناسایی کنند [۲۱]. تشخیص آرتروز از MRI قبل از شروع آسیب نیز توسط یادگیری عمیق انجام می شود [۲۱].

- دارو سازی: یادگیری عمیق در کشف دارو نیز موثر ظاهر شده و می تواند خواص دارو ها را پیش بینی و کاندیدهای دارویی بالقوه را شناسایی کند. کاربردهای این زمینه شامل کشف دارو و تغییر کاربری دارو است [۱۰]. تیمی از محققان در دانشگاه تورنتو ابزاری به نام DeepBind ایجاد کرده اند. طوری که یک مدل شبکه عصبی پیچشی که داده های ژنومی را می گیرد و توالی پروتئین های اتصال دهنده DNA و RNA را پیش بینی می کند. محققان می توانند از DeepBind برای ایجاد مدل های رایانه ای استفاده کنند که اثرات تغییرات در توالی DNA را نشان دهد. آنها می توانند از این اطلاعات برای توسعه ابزارها و داروهای تشخیصی پیشرفته تر استفاده کنند. [۲۱] یادگیری عمیق در این بخش با پیش بینی فعالیت های زیستی سلول ها، طراحی مولکولی جدید، پیش بینی پیوندهای مولکولی و تجزیه و تحلیل تصاویر بیولوژیکی به کمک داروسازان می آید و کشف داروهای جدید را امکان پذیر می کند [۲۲].

- تجزیه و تحلیل تصویربرداری پزشکی: یادگیری عمیق برای با توسعه برخی الگوریتم ها می تواند تصاویر پزشکی مانند سی تی اسکن و MRI را تجزیه و تحلیل کند. کاربردهای این زمینه شامل تشخیص و تقسیم تومور است. [۱۰] آری، شرایط زیادی مانند CT یا MRI در زمینه های پزشکی وجود دارد که به کار پردازش بینایی نیاز دارند؛ در سال های اخیر، با تکنیک های یادگیری عمیق، بینایی ماشین به طور چشمگیری توانایی خود را بهبود بخشیده و نقش حیاتی فزاینده ای در حوزه پزشکی و سلامت ایفا کرده است.

- بیماری آلزایمر: آلزایمر یکی از چالش های مهم صنعت پزشکی است. از تکنیک یادگیری عمیق برای تشخیص بیماری آلزایمر در مراحل اولیه استفاده می شود [۲۲].



- جراحی‌ها: جراحی با کمک هوش مصنوعی، متداول‌ترین کاربرد یادگیری عمیق در پزشکی است. یادگیری عمیق در این بخش به کمک پزشکان می‌آید و با تحلیل و بررسی الگوهای جراحی‌های پیشین، جراح را در حین جراحی همراهی می‌کند. معمولاً راهکارهای جراحی ارائه شده توسط یادگیری عمیق، مدت بستری بیمار را تا ۲۱٪ کاهش می‌دهد؛ چرا که این راهکارها تا حد ممکن غیر تهاجمی یا با میزان تهاجمی بسیار کمتر هستند و بدن بیمار برای بهبودی از این نوع جراحی به زمان کمتری نیاز دارد [۲۲].
- ساخت اندام‌های مصنوعی: یکی دیگر از کاربردهای یادگیری عمیق در پزشکی، ساخت اندام‌ها و ارگان‌های مصنوعی است. قطعاً ساخت چنین مصنوعاتی نیازمند بررسی دقیق عملکرد بدن انسان و شیوه پاسخگویی آن به محرک‌های گوناگون است. دانشمندان و مهندسان پزشکی برای ساخت این اندام‌ها از هوش مصنوعی کمک می‌گیرند و با استفاده از الگوریتم‌های یادگیری عمیق حجم عظیمی از داده‌های پزشکی را مورد بررسی قرار می‌دهند و با تشخیص الگوهای گوناگون طراحی اولیه اندام‌ها را انجام می‌دهند. استفاده از این روش در طراحی اندام‌ها سرعت و دقت تولید را تا حد زیادی افزایش می‌دهد [۲۲].
- ژنتیک: DNA انسان، تمامی اطلاعات ژنتیکی او را در خود جای داده است. گسترش کاربردهای یادگیری عمیق در پزشکی این امکان را برای متخصصین ژنتیک فراهم کرده که با بررسی ژنوم‌های هر فرد به ناشناخته‌های بدن انسان دست پیدا کنند. درواقع، یادگیری عمیق با تحلیل داده‌های ژنتیکی و دسته‌بندی آن‌ها و همچنین شناسایی الگوهای همسان، درک و شناخت از ژنوم‌ها را افزایش داده است. این موضوع می‌تواند در پیش‌بینی و طبقه‌بندی جهش‌های ژنتیکی، اتصال ژنوتایپ به فنوتایپ و رفع مشکلات ژنتیکی به پزشکان یاری برساند [۲۲].
- یادگیری عمیق در MRI: هرچند، MRI مزایای بیشتری نسبت به سایر تکنیک‌های تصویربرداری پزشکی دارد اما سرعت کم آن منجر به افزایش هزینه‌ها، مضطرب شدن بیماران و کاهش کیفیت تصویر می‌شود. علاوه بر این، گاهی اجمار ایجاد می‌شود به کودکان داروی آرام بخش داده شود. مؤسسه ملی بهداشت (NIH) کمک‌هزینه‌ای به مبلغ ۲/۳ میلیون دلار (به مدت چهار سال) به گروهی بین رشته‌ای، متشکل از مهندسان، پزشکان و دانشمندان اهدا کرده است تا تکنیک‌های سریع‌تر و دقیق‌تری برای MRI ارائه دهند. رهبری این گروه را دانشیار مهندسی پزشکی، رضوان احمد و استاد مهندسی کامپیوتر، فیلیپ اشنیتزر بر عهده گرفته‌اند. در MRI بدن بیمار چندین بار اسکن می‌شود و این روند ممکن است یک یا بیشتر از یک ساعت به طول بینجامد. در طول اسکن از بیمار خواسته می‌شود چند دقیقه آرام و بی‌حرکت باقی بماند، زیرا کوچک‌ترین حرکتی بر کیفیت تصویر تأثیر می‌گذارد. علاوه بر این، تقاضا برای استفاده از MRI در مواردی که زمان اهمیت زیادی دارد، رو به افزایش است؛ تصویربرداری از قلب در حال تپش یا رویدادهای گذرا که فقط چند ثانیه به طول می‌انجامند، نمونه‌هایی از این دست هستند. حال کاربرد یادگیری عمیق در MRI قصد دارد عملکرد آن را بهبود بخشد. اشنیتزر معتقد است یکی از اصلی‌ترین محرک‌های کاهش مدت زمان MRI است. تلاش می‌شود از تمامی متغیرهایی که در روش‌های تصویربرداری سنتی اندازه‌گیری می‌شوند استفاده نشود، اما در چنین حالتی تمامی اطلاعات کسب‌شده از این متغیرها از دست می‌رود. نحوه کار یادگیری عمیق در MRI به این صورت است که این پژوهشگران به منظور جایگزینی اطلاعات از دست رفته و بازیابی تصویر یک چارچوب یادگیری عمیق قدرتمند و جامع را توسعه داده و صحت آن را ارزیابی می‌کنند. این چارچوب سرعت و کیفیت بازسازی تصویر MRI را افزایش می‌دهد. چارچوب پیشنهادی این پژوهشگران با استفاده از الگوریتم‌های Plug-and-Play (PnP) مدل‌های جمع‌آوری داده را با مدل‌های تصویری آموزش داده شده تلفیق می‌کند. الگوریتم‌های PnP سریع‌تر از روش‌های سنتی تصویربرداری پزشکی، تصاویر را بازسازی می‌کنند و این درحالی است که کیفیت و ارزش تشخیصی این تصاویر بالاتر است. همچنین، این پژوهشگران برای ارتقای کیفیت تصاویر از متدهای کاهش نویز مبتنی بر یادگیری عمیق استفاده می‌کنند. وجه تمایز روش PnP دوگانه بودن آن است؛ به عبارت دیگر در این روش از رویکردهای مبتنی بر یادگیری ماشین و



رویکردهای فیزیک مهندسی استفاده شده و برای ارائه تصویر نهایی متناوباً میان آن‌ها تکرار می‌شود. چنانچه روش پیشنهادی این گروه برای استفاده از یادگیری عمیق در MRI به صورت کامل پیاده‌سازی و اجرا شود، مدت زمان اسکن MRI به نصف کاهش می‌یابد. در این روش نیازی به اصلاح سخت‌افزاری نیست؛ پایانه کامپیوتری که به اسکنر MRI متصل شده است، محاسبات را انجام می‌دهد. پزشکان ترجیح می‌دهند تصاویری که اسکنر MRI ثبت می‌کند را بدون فوت وقت ببینند و بررسی کنند و در صورتی که این الگوریتم توسعه یابد می‌توان امیدوار بود که در عرض چند ثانیه تصویر را بازیابی کند. این تیم برای نمایش کاربردهای گسترده الگوریتم خود، چارچوب آن را بر روی داده‌های MRI بیمارانی که در رده سنی کودکان و بزرگسالان قرار داشتند آزمایش خواهند کرد. این چارچوب به طور خاص می‌تواند در تصویربرداری متحرک قلب و مغز کاربرد داشته باشد. به عقیده احمد، در صورت موفقیت این چارچوب، ارتقای کیفیت تصاویر و تسریع فرایند MRI می‌تواند تقریباً در تمامی کاربردهای MRI مفید باشد. اهمیت کاهش مدت زمان MRI به ویژه در هنگام اسکن بدن کودکان مشخص می‌شود. تصویربرداری کودکان به مراتب دشوارتر است، زیرا کودکان کم سن و سال نمی‌توانند آرام و بی‌حرکت در اسکنر MRI قرار بگیرند و به همین دلیل باید به آن‌ها داروی آرام‌بخش داد. علاوه بر این «طبق نتایج تحقیقات مختلف، استفاده از داروی آرام‌بخش در طولانی مدت عوارض خاص خود را دارد و به همین دلیل مصرف آن را باید به حداقل رساند. از این رو، در تصویربرداری کودکان مهم است که بدن کودک را هر چه سریع‌تر اسکن نمود. همچنین، این تیم در نتیجه استفاده از برخی روش‌های خاص کاهش نویز در راهکاری که برای تصویربرداری از بدن ارائه داده بودند، به نتایج بهتری دست پیدا کرده‌اند. این الگوریتم‌ها با حذف آرتیفکت‌های غیرضروری از نمونه‌ها، کیفیت تصاویر را ارتقا می‌دهند. هدف آن‌ها این است که این روش را به گونه‌ای توسعه دهند که در موارد خاص استفاده شود. تصاویر مغز با تصاویر قلب تفاوت زیادی دارند و به همین دلیل می‌توان روش‌های کاهش نویز را برای موارد خاص آموزش داد. با توجه به اینکه این روش کاهش نویز بسیار خاص است، عملکرد بهتری نیز خواهد داشت. در طول فرایند آموزش، این گروه از هزاران دیتاست تصویری، از جمله دیتاست تصویری fastMRI دانشگاه نیویورک برای آموزش متد کاهش نویز خود استفاده کردند. روش‌های کاهش نویز، شبکه‌های عصبی محاسباتی هستند و ساختار پردازشی آن‌ها بسیار شبیه به مغز پستانداران است. برای آموزش آن‌ها از تعداد زیادی تصویر پزشکی استفاده می‌شود و در نهایت آن‌ها ساختار و نحوه حذف نویز از تصاویر پزشکی را می‌آموزند. این پژوهشگران به عنوان بخشی از پروژه خود و برای کمک به پیشرفت بیشتر این حوزه، منبعی (Repository) متن باز از هزاران دیتاست تصویری قلب ایجاد کرده‌اند. کاری که فلوریان نول با دیتاست fastMRI انجام داد الهام‌بخش این کار بوده است. ایجاد دیتاست به نوعی نقش کاتالیزور را برای پیشرفت‌های این حوزه ایفا می‌کند. بسیاری از گروه‌ها می‌توانند این داده‌های پزشکی را به کار گیرند و برای ارتقای عملکرد این متد با یکدیگر رقابت کنند. پژوهشگرانی از دانشگاه پزشکی ایالتی اوهایو، کارولینا زاربا، دانشیار بیماری‌های قلبی-عروقی، اورلاندو سیمونتی، استاد رادیولوژی و بیماری‌های قلبی-عروقی، گای بروک، دانشیار پژوهشی انفورماتیک زیست‌پزشکی پژوهشگران همکار در این پژوهش هستند. دکتر مای لان هو از بیمارستان ملی کودکان و فلوریان نول، دانشیار رادیولوژی از دانشگاه نیویورک نیز پژوهشگران همکار در این پژوهش هستند. این پژوهش تحت حمایت مؤسسه ملی تصویربرداری زیست‌پزشکی و مهندسی زیست‌شناسی (NIBIB) قرار دارد [۲۲].

بحث و نتیجه‌گیری

در این مقاله دو حوزه تخصصی و در حال رشد هوش مصنوعی، بینایی کامپیوتر و یادگیری ماشین، و کاربرد آن‌ها در سلامت و پزشکی مورد توجه قرار گرفت. بیان شد که در حال حاضر از تلفیق هوش مصنوعی و پزشکی در تحقیقات بسیاری در حوزه توسعه دارو، تشخیص بیماری‌ها و مراقبت‌های بهداشتی و درمانی به کار گرفته می‌شود. در این مقاله، ابتدا توضیحی در مورد دو مفهوم پردازش تصویر و بینایی کامپیوتر ارائه گردید. در ادامه یادگیری ماشین و یادگیری عمیق مرور شد. کاربردهای بینایی

کامپیوتر در حوزه پزشکی و کاربردهای یادگیری عمیق در حوزه پزشکی در دو بخش بررسی و ارائه شدند. تجزیه و تحلیل تصویربرداری پزشکی، پردازش تصویر برای تشخیص سرطان، تشخیص بیماری و پیش‌بینی درمان با یادگیری عمیق، بکارگیری یادگیری عمیق در MRI، بکارگیری بینایی کامپیوتر و یادگیری عمیق در جراحی‌ها از موضوعاتی است که در این دو بخش ارائه شدند.

مراجع

- ۱- "طبقه بندی تصویر یا Image Classification چیست"، مهسا مژدهی، <https://cafetadris.com>
- ۲- "بینایی کامپیوتری (Computer Vision) چیست و چه کاربردهایی دارد؟"، مهسا مژدهی، <https://cafetadris.com>
- ۳- "الگوریتم Sift - چگونگی تناظریابی تصاویر در نرم افزارهای پردازش تصویر"، میثم رفیعی، <https://maan.ir>
- ۴- "الگوریتم های بینایی کامپیوتر"، <https://kaizendatagroup.ir>
- ۵- یادگیری ماشین چیست؟ | الگوریتم‌های ماشین لرنینگ چه کاربردی دارند؟ <https://amerandish.com>
- ۶- "الگوریتم ماشین بردار پشتیبان (Support Vector Machine) چیست؟" مهسا مژدهی، <https://cafetadris.com>
- ۷- "نقش شبکه عصبی در یادگیری ماشین چیست؟" بیتا جلالی، <https://blog.faradars.org>
- ۸- "یادگیری عمیق چیست؟" الهام احمدی ۲ مرداد ۱۴۰۲، <https://quera.org>
- ۹- تفاوت شبکه‌های عصبی (Neural Networks) با یادگیری عمیق (Deep Learning) چیست؟ مسعود کاویانی، <https://chistio.ir>
- ۱۰- یادگیری عمیق (Deep Learning) چیست؟ <https://7learn.com>
- 11- "Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward", ELYAN, E., VUTTIPITTAYAMONGKOL, P., JOHNSTON, P., MARTIN, K., MCPHERSON, K., MORENO-GARCIA, C.F., JAYNE, C. and SARKER, M.M.K.; 2022; Artificial Intelligence Surgery
- ۱۲- شبکه عصبی پیچشی، <https://fa.wikipedia.org>
- ۱۳- شبکه عصبی کانولوشن، سید سجاد اشرفی، <https://howsam.org/convolutional-neural-network>
- ۱۴- شبکه عصبی کانولوشن (CNN) چیست؟ <https://quera.org/blog/what-is-a-convolutional-neural-network>
- ۱۵- هر آنچه که درباره بینایی ماشین در پزشکی باید بدانید، ۱۴۰۱/۲/۲۸، <https://hooshio.com>
- ۱۶- پردازش تصویر در پزشکی چیست؟ ۱۴۰۳/۳/۲۲، <https://hamrah.academy/blog/image-processing-medicine>
- ۱۷- شش کاربرد پردازش تصویر در پزشکی + مزیت‌های استفاده، علیرضا همتی، ۱۴۰۲/۶/۲۵، <https://didbaan.com/mag/image-processing-in-medicine>
- ۱۸- الگوریتم R-CNN چیست و چه نقشی در بینایی ماشین دارد؟، مینا پولایی، <https://cafetadris.com/blog>
- ۱۹- الگوریتم YOLO چیست؟ چه کاربردی در تشخیص اشیا دارد؟ مهسا مژدهی، <https://cafetadris.com/blog>



۲۰- بهترین روش های تشخیص اشیا و ردیابی اشیا با بینایی کامپیوتر، بهنام اوجاقی،

[/https://hamruyesh.com/object-detection-and-tracking-in-2020-part-one](https://hamruyesh.com/object-detection-and-tracking-in-2020-part-one)

۲۱- یادگیری عمیق چگونه به پزشکی کمک می کند؟ عارفه ضیائی ، ۱۸ اسفند ۱۳۹۸،

<https://amanjacademy.com/how-does-deep-learning-help-medicine>

۲۲- کاربردهای متنوع یادگیری عمیق (Deep Learning) در دنیای امروز، <https://iranbmemag.com>