



دانشگاه صنعتی شریف  
دانشکده‌ی مهندسی برق

پایان‌نامه‌ی کارشناسی ارشد  
بیوالکتریک

عنوان:

# قطعه بندی ساختارهای در ریسک در تصاویر پزشکی با استفاده از روش های یادگیری ماشین

نگارش:

رضا کریمزاده

استاد راهنما:

دکتر عمادالدین فاطمی‌زاده

اسفند ۱۴۰۰

سلام افلا

## چکیده

نگارش پایان‌نامه علاوه بر بخش پژوهش و آماده‌سازی محتوا، مستلزم رعایت نکات فنی و نگارشی دقیقی است که در تهیه‌ی یک پایان‌نامه‌ی موفق بسیار کلیدی و مؤثر است. از آن جایی که بسیاری از نکات فنی مانند قالب کلی صفحات، شکل و اندازه‌ی قلم، صفحات عنوان و غیره در تهیه‌ی پایان‌نامه‌ها یکسان است، با استفاده از نرم‌افزار حروف‌چینی زی‌تک و افزونه‌ی زی‌پرشین یک قالب استاندارد برای تهیه‌ی پایان‌نامه‌ها ارائه گردیده است. این قالب می‌تواند برای تهیه‌ی پایان‌نامه‌های کارشناسی و کارشناسی ارشد و نیز رساله‌ی دکتری مورد استفاده قرار گیرد. این نوشتار به طور مختصر نحوه‌ی استفاده از این قالب را نشان می‌دهد.

**کلیدواژه‌ها:** پایان‌نامه، حروف‌چینی، قالب، زی‌پرشین

# فهرست مطالب

۸	۱ مفاهیم اولیه
۸	۱-۱ تصویربرداری پزشکی
۹	۱-۱-۱ تصویربرداری سی تی اسکن
۱۱	۱-۱-۲ تصویربرداری ام آر آی
۱۳	۲-۱ هوش مصنوعی
۱۳	۲-۱-۱ هوش مصنوعی
۱۵	۲-۲-۱ یادگیری ماشین
۱۷	۲-۲-۳ یادگیری عمیق
۲۱	۳-۱ شبکه‌های عصبی کانولوشنی عمیق
۲۳	۳-۱-۱ لایه‌های کانولوشنی
۲۴	۳-۲-۱ توابع فعالیت
۲۵	۳-۳-۱ لایه‌های ادغام
۲۶	۳-۴-۱ لایه‌های تمام متصل
۲۷	۴-۱ روش‌های تعمیم‌پذیری
۲۹	۵-۱ پردازش تصویر و بینایی کامپیوتر
۳۰	۶-۱ جمع‌بندی

## فهرست شکل‌ها

- ۱-۱ دستگاه سی‌تی اسکن [۱] . . . . . ۱۰
- ۱-۲ نمونه‌ی تصویر سی‌تی اسکن از ناحیه‌ی شکم در سه نمای مختلف [۲] . . . . . ۱۱
- ۱-۳ تصویر یک دستگاه ام‌آر‌آی (سمت چپ) و یک لایه از تصویر سه‌بعدی ام‌آر‌آی زانو (سمت راست) [۳] . . . . . ۱۳
- ۱-۴ ارتباط حوزه‌های هوش مصنوعی، یادگیری ماشین، یادگیری عمیق و بینایی کامپیوتر . ۱۵
- ۱-۵ مقایسه‌ی الگوریتم‌های طبقه‌بندی و رگرسیون [۴] . . . . . ۱۶
- ۱-۶ مقایسه‌ی یک نورون طبیعی و مدلسازی ساده‌ی آن [۵] . . . . . ۱۹
- ۱-۷ مقایسه‌ی شبکه‌ی پرسپترون عمیق و کم عمق [۶] . . . . . ۲۰
- ۱-۸ شبکه‌های مولد خصمانه [۷] . . . . . ۲۱
- ۱-۹ ماتریس ویژگی‌های سه بعدی استخراج شده توسط یک شبکه‌ی کانولوشنی [۸] . . ۲۲
- ۱-۱۰ معماری یک شبکه‌ی کانولوشنی برای طبقه‌بندی اعداد دست‌نویس [۹] . . . . . ۲۳
- ۱-۱۱ چگونگی حرکت فیلتر در یک لایه‌ی کانولوشنی [۹] . . . . . ۲۳
- ۱-۱۲ روند محاسبات در یک لایه‌ی کانولوشنی [۹] . . . . . ۲۴
- ۱-۱۳ چهار نمونه از پر استفاده‌ترین توابع فعالیت [۱۰] . . . . . ۲۵
- ۱-۱۴ ادغام میانگین‌گیر و ادغام بیشینه‌گیر [۹] . . . . . ۲۶
- ۱-۱۵ توقف زود هنگام با توجه به خطای داده‌های ارزیابی [۱۱] . . . . . ۲۸

---

۱-۱۶ چگونگی اعمال تکنیک dropout [۱۲] . . . . .	۲۹
--	----

## فهرست جدول‌ها

# فصل ۱

## مفاهیم اولیه

در این فصل به معرفی حوزه‌ها و مفاهیم اولیه‌ی مرتبط با این پژوهش پرداخته می‌شود. در ابتدا، انواع تصویربرداری پزشکی معرفی و بررسی خواهد شد و در ادامه به تعریف حوزه‌های مرتبط با پردازش تصاویر پزشکی و بینایی کامپیوتر پرداخته می‌شود.

### ۱-۱ تصویربرداری پزشکی

زمینه‌ی تصویربرداری پزشکی و سیستم‌های تصویرگر، یک زمینه‌ی بسیار گسترده و به نوبه‌ی خود پیچیده است که با استفاده از انواع پرتوهای ایکس، فراصوت، گاما، امواج الکترومغناطیس و ... با تاباندن به بدن بیمار و بازسازی تصویر از روی پرتوهای دریافتی، صورت می‌گیرد<sup>۱</sup>. تصویربرداری پزشکی کمک بسیار زیادی به متخصصین حوزه‌ی سلامت در روند تشخیص و درمان صورت داده است، به طور کلی می‌توان روش‌های تصویربرداری پزشکی را به صورت زیر خلاصه کرد<sup>[۱۳]</sup>:

- رادیوگرافی<sup>۲</sup>: که شامل تصویربردای‌های بر مبنای پرتوی ایکس است که از مهمترین آن‌ها می‌توان به فلوروسکوپی<sup>۳</sup> و رادیوگرافی‌های پروجکشنال<sup>۴</sup> اشاره کرد.

---

<sup>۱</sup>البته در تصویربرداری هسته‌ای، برعکس سایر روش‌های تصویربرداری، منبع تابش پرتو درون بدن بیمار قرار می‌گیرد و با آشکارسازی پرتوهای دریافتی تصویر نهایی تشکیل می‌گردد.

<sup>۲</sup>Radiography

<sup>۳</sup>Fluoroscopy

<sup>۴</sup>Projectional radiographs



- تصویربرداری هسته‌ای<sup>۵</sup>: که با تزریق ایزوتوپ‌های خاص و ذره‌های پرنرژی گسیل شده از مواد رادیواکتیو به بدن بیمار و آشکارسازی پرتوهای دریافتی، تصویر نهایی ساخته می‌شود. از جمله روش‌های تصویربرداری در این حوزه میتوان به PET<sup>۶</sup> و SPECT<sup>۷</sup> اشاره کرد.
  - تصویربرداری فراصوت<sup>۸</sup>: در این نوع از تصویربرداری از کریستال‌های پیزوالکتریک<sup>۹</sup> برای تولید صوت با فرکانس بالا استفاده می‌شود این صوت به بافت بدن تابانده می‌شود و بازسازی تصویر از روی صوت بازگشتی صورت می‌گیرد.
  - تصویربرداری توموگرافی<sup>۱۰</sup>: در تصویربرداری توموگرافی هدف ساخت یک تصویر سه‌بعدی است برای این منظور از لایه‌های مختلف یک شیء بدون آن که بریده شود تصویربرداری صورت می‌گیرد و این لایه‌ها بر روی یکدیگر انباشت می‌شوند و در نهایت تصویر سه‌بعدی نهایی ساخته می‌شود. از جمله روش‌های موجود در این حوزه می‌توان به تصویربرداری سی‌تی اسکن و تصویربرداری بر مبنای تشدید مغناطیسی (ام‌آرآی)<sup>۱۱</sup> اشاره نمود.
  - انواع دیگر تصویربرداری مانند: تصویربرداری فوتوآکوستیک<sup>۱۲</sup>، تصویربرداری حرارتی<sup>۱۳</sup> و ... نیز وجود دارد که به دلیل کاربردهای کلینیکی کمتر از شرح آن‌ها خودداری می‌شود.
- در عملیات قطعه‌بندی تصاویر پزشکی به دلیل آن که قطعه‌بندی به طور معمول از روی تصاویر توموگرافی سه‌بعدی صورت می‌گیرد، در ادامه به شرح و بسط بیشتر روش‌های تصویربرداری سی‌تی اسکن و ام‌آرآی پرداخته می‌شود.

### ۱-۱-۱ تصویربرداری سی‌تی اسکن

تصویربرداری سی‌تی اسکن یک تکنیک تصویربرداری پزشکی است که در رادیولوژی برای استخراج اطلاعات از بدن به صورت غیرتهاجمی استفاده می‌شود و روند تشخیص را سرعت می‌بخشد. برخلاف

<sup>5</sup>Nuclear Imaging

<sup>6</sup>Positron emission tomography

<sup>7</sup>single-photon emission computerized tomography

<sup>8</sup>Ultrasound

<sup>9</sup>Piezoelectric

<sup>10</sup>Tomography

<sup>11</sup>Magnetic Resonance Imaging (MRI)

<sup>12</sup>Photoacoustic imaging

<sup>13</sup>Thermography

دستگاه‌های معمول تصویربرداری پرتوی ایکس، که از یک منبع ثابت پرتوی ایکس استفاده می‌کنند، در سی‌تی اسکن از یک منبع متحرک مجهز به موتور استفاده می‌شود که حول گانتری<sup>۱۴</sup> دستگاه قابلیت چرخش دارد. در طی تصویربرداری، بیمار بر روی یک تخت قرار می‌گیرد و به آهستگی به داخل گانتری وارد می‌شود؛ در همین حین منبع پرتوی ایکس درون گانتری دور بدن بیمار می‌چرخد و باریکه‌ی پرتوهای اشعه‌ی ایکس از بدن بیمار عبور می‌کند. در سی‌تی اسکن از آشکارسازهای دیجیتال پرتوی ایکس استفاده می‌شود که دقیقاً در مقابل منبع پرتوی ایکس قرار گرفته‌اند و با آشکارسازی اشعه‌ی عبوری از بدن بیمار، یک سیگنال به کامپیوتر ارسال می‌شود. شکل ۱-۱ شمای کلی یک دستگاه سی‌تی اسکن را نشان می‌دهد.



شکل ۱-۱: دستگاه سی‌تی اسکن [۱]

هر مرتبه که منبع پرتوی ایکس یک چرخش کامل را انجام می‌دهد دستگاه سی‌تی اسکن از تکنیک‌های پیچیده‌ی ریاضیاتی برای ساخت تصویر دو بعدی برای هر لایه از بدن بیمار از روی سیگنال‌های دریافتی انجام می‌دهد. ضخامت این لایه‌ها بستگی به نوع دستگاه سی‌تی اسکن دارد اما به طور معمول بین یک تا ده میلی‌متر از بافت بدن برای هر لایه است. وقتی بازسازی یک لایه به اتمام رسید بر روی لایه‌های قبلی انباشته می‌شود و در نهایت یک تصویر سه‌بعدی ساخته می‌شود.

هر لایه از تصویر ساخته‌شده امکان نمایش به صورت مجزا و یا به صورت انباشته شده و سه‌بعدی را دارد که قابلیت نمایش، اسکلت، ساختارها و بافت‌های بدن و همچنین ناهنجاری‌های ایجاد شده در بدن را دارا می‌باشد و امکان تشخیص را برای پزشک مهیا می‌سازد. استفاده از تصاویر سی‌تی در روند تشخیص و درمان فواید زیادی دارد از جمله، توانایی چرخش و جابجایی بین لایه‌های مختلف تصویر که امکان مکان‌یابی موقعیت دقیق ناهنجاری را فراهم می‌سازد. شکل ۱-۲ یک نمونه تصویر از ناحیه‌ی شکم را نشان می‌دهد که در آن ساختارها و بافت‌های مختلف به راحتی قابل تفکیک است.

<sup>۱۴</sup>Gantry به محفظه‌ی سیلندری شکل دستگاه سی‌تی اسکن گفته می‌شود که تیوب پرتوی ایکس درون آن قرار می‌گیرد.



شکل ۱-۲: نمونه‌ی تصویر سی‌تی اسکن از ناحیه‌ی شکم در سه نمای مختلف [۲]

از سی‌تی اسکن می‌توان در تشخیص بیماری و آسیب نواحی مختلف بدن استفاده نمود به عنوان مثال از سی‌تی در تشخیص تومور و غدد سرطانی و سایر ناهنجاری‌ها در نواحی شکم، سر و گردن و قفسه‌ی سینه بسیار استفاده می‌شود [۱۴].

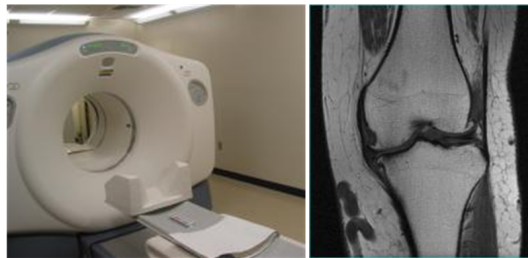
### ۱-۱-۲ تصویربرداری ام‌آر‌آی

ام‌آر‌آی یا روش تصویربرداری با تشدید مغناطیس یکی از روش‌های پیشرفته تصویربرداری پزشکی است. با استفاده از این روش می‌توان تصویر بافت‌های درونی بدن را دید و از آن طریق مشکلات و بیماری‌های اعضای بدن را تشخیص داد. همانطور که می‌دانیم در روش‌های تصویربرداری با اشعه‌ی ایکس مانند رادیوگرافی ساده و یا سی‌تی اسکن بدن تحت تابش مقدار معینی از اشعه یونیزه کننده قرار می‌گیرد که اگر از حد مشخصی بیشتر باشد می‌تواند موجب اشکالاتی در کارکرد سلول‌ها شود. ولی در ام‌آر‌آی از اشعه ایکس استفاده‌ای نمی‌شود و بنابراین نسبت به رادیوگرافی و سی‌تی اسکن بسیار کم ضررتر است. امواج مورد استفاده در ام‌آر‌آی از جنس امواج رادیویی و مغناطیسی هستند که ضرری برای بدن ندارند.

ام آر آی از این واقعیت فیزیکی استفاده می‌کند که پروتون‌هایی که در هسته‌ی اتم‌ها قرار گرفته‌اند مانند کره‌ی زمین در حول محور خود با سرعت زیادی می‌چرخند و در نتیجه یک میدان مغناطیسی در اطراف خود تشکیل می‌دهند. در ام آر آی بیمار در یک میدان مغناطیسی بسیار قوی قرار می‌گیرد. این میدان موجب می‌شود محور چرخش پروتون‌های هسته‌ی اتم‌ها در تمام بافت‌های بدن (بخصوص پروتون‌هایی که در هسته مولکول آب قرار دارند) در امتداد خطوط میدان مغناطیسی ام آر آی قرار گیرند. سپس امواج رادیویی خاصی به سوی بدن بیمار تابانده می‌شود. این امواج که بصورت پالس فرستاده می‌شوند موجب می‌گردند تا محور چرخش پروتون‌ها کمی تغییر کند. با اتمام پالس رادیویی، محور چرخش پروتون دوباره در امتداد خطوط میدان مغناطیس برمیگردد. این برگشت موجب ایجاد یک موج رادیویی (الکترومغناطیسی) جدید می‌شود. سپس این امواج رادیویی ثانویه که از تک تک پروتون‌ها ساطع می‌شوند، توسط گیرنده‌های دستگاه ام آر آی دریافت شده و به کامپیوتر آن ارسال می‌گردند. کامپیوتر ام آر آی بسیار پر قدرت و با توان محاسباتی بالا است. در این کامپیوتر امواج دریافت شده با سرعت تحلیل شده و سپس تصاویری براساس این تحلیل‌ها ساخته می‌شود که پزشک آن‌ها را بر روی مانیتور دستگاه می‌بیند و در صورت لزوم آن‌ها را چاپ می‌کند.

در کامپیوتر ام آر آی مشخص می‌شود که در چه نقاطی از بدن موج رادیویی بیشتری ساطع شده است. هرچه شدت موج دریافتی از نقطه‌ای از بدن بیشتر باشد نشانه تراکم بیشتر پروتون در آن نقطه است و چون فراوان‌ترین اتم بدن که پروتون دارد اتم هیدروژن است که در مولکول آب قرار دارد پس هر جایی که موج رادیویی بیشتری ارسال کرده است در واقع آب بیشتری داشته است. در واقع کاری که ام آر آی انجام می‌دهد این است که نشان دهد در چه نقاطی از بدن آب بیشتری وجود دارد. چون غلظت مولکول آب در بافت‌های بدن متفاوت است و با بیمار شدن بافت‌ها این غلظت باز هم تغییر می‌کند می‌توان با استفاده از اطلاعات دریافتی تصویر بسیار دقیقی از شکل بافت‌های گوناگون بدن ایجاد کرد.

ام آر آی یک روش تصویربرداری دقیق و پر قدرت برای تشخیص مشکلات و بیماری‌های بافت‌های بدن است. یکی از نقاط تمایز این روش با سی تی اسکن در این است که در ام آر آی تصاویر بافت‌های نرم مانند غضروف، تاندون، عصب و رگ‌ها بسیار واضح و دقیق دیده می‌شوند و این روش تصویربرداری بخصوص برای تشخیص بیماری‌های این بافت‌ها مفید است. تصویر حاصل از ام آر آی مانند تصویر سی تی اسکن، سه بعدی است و امکان نمایش لایه‌های مختلف وجود دارد. شکل ۱-۳ در سمت چپ یک دستگاه ام آر آی آر و در سمت راست یک نمونه تصویر دریافت شده از زانو را نشان می‌دهد که با جزئیات بالایی بافت‌ها را از یکدیگر تفکیک کرده است [۳].



شکل ۱-۳: تصویر یک دستگاه ام‌آر‌آی (سمت چپ) و یک لایه از تصویر سه‌بعدی ام‌آر‌آی زانو (سمت راست) [۳]

## ۲-۱ هوش مصنوعی

در ادامه‌ی این قسمت به بررسی و تعریف حوزه‌های مرتبط با هوش مصنوعی<sup>۱۵</sup> پرداخته خواهد شد.

### ۱-۲-۱ هوش مصنوعی

هوش مصنوعی به عنوان یک رشته‌ی دانشگاهی در دهه‌ی ۵۰ میلادی پایه‌گذاری شد. در واقع اصطلاح هوش مصنوعی توسط جان مک‌کارتی<sup>۱۶</sup>، دانشمند علوم کامپیوتر آمریکایی، ابداع شد که طبق تعریف او، هوش مصنوعی عبارت است از علم ساخت ماشین‌های هوشمند به خصوص برنامه‌های کامپیوتری هوشمند.

اگرچه از تعریف هوش مصنوعی چندین دهه می‌گذرد اما رشد و گسترش آن تنها در دو دهه‌ی اخیر به علت در دسترس بودن داده‌های عظیم<sup>۱۷</sup> و بالا رفتن توان کامپیوتری میسر شده است. هوش مصنوعی بهترین عملکرد خود را با در کنار هم قرار دادن حجم بالای دادگان و الگوریتم‌های هوشمند بدست می‌آورد که باعث یادگیری ویژگی‌ها و الگوهای درون دادگان به صورت خودکار می‌شود. از جمله کاربردهای روز افزون هوش مصنوعی که تقریباً با زندگی همه‌ی ما آمیخته شده است می‌توان به: ماشین‌های خودران، دستیاران صوتی، تشخیص هوشمند چهره برای شناسایی افراد، پروژه‌های وابسته به ژن‌های انسان و ... اشاره نمود.

<sup>15</sup>Artificial Intelligence

<sup>16</sup>John McCarthy

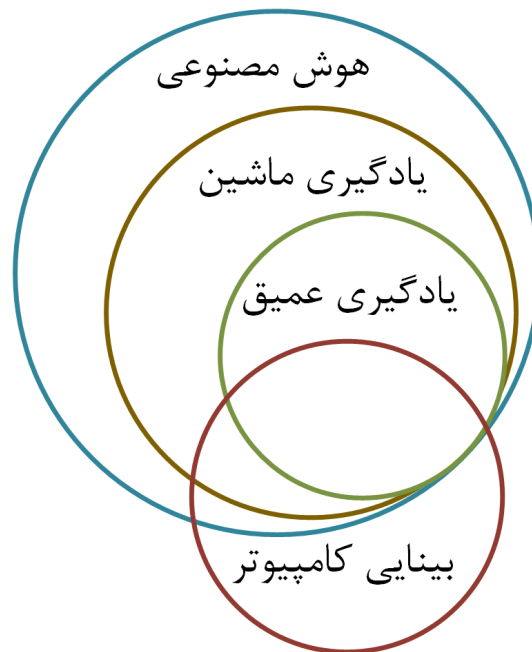
<sup>17</sup>Big Data

اگرچه بسیاری از افراد اصطلاحات هوش مصنوعی، یادگیری ماشین و یادگیری عمیق را به جای یکدیگر استفاده می‌کنند اما در واقع این اصطلاحات تعاریف دقیق‌تر و جداگانه‌ای از یکدیگر دارند. همانطور که پیشتر اشاره شد هوش مصنوعی علم ایجاد ماشین‌ها و برنامه‌های کامپیوتری هوشمند است که همانطور که از این تعریف نیز پیداست حوزه‌ی بسیار وسیعی را دربر می‌گیرد که امروزه تقریباً هر شرکتی ادعای استفاده از هوش مصنوعی در محصولات خود را داراست. بنابراین یادگیری ماشین در زیرمجموعه‌ی هوش مصنوعی قرار می‌گیرد و شامل تکنیک‌های و مدل‌های بسیار پیشرفته‌تر که کامپیوترها را قادر به استخراج ویژگی‌های دقیق‌تر از دادگان می‌سازد و به الگوریتم هوش مصنوعی تحویل می‌دهد. به طور خلاصه می‌توان گفت یادگیری ماشین علم قرار دادن کامپیوترها در عمل است بدون آن که به طور صریح برنامه‌ریزی شده باشند.

در نهایت، یادگیری عمیق یک حوزه‌ی جدید یادگیری ماشین است که از شبکه‌های عصبی عمیق و چند لایه برای استخراج ویژگی و تصمیم‌گیری استفاده می‌کند که این کار موجب استخراج الگوها و ویژگی‌های مرتبه‌ی بالاتر و تفکیک‌گر بهتر می‌شود و در نتیجه دقت الگوریتم افزایش پیدا خواهد کرد. از جمله کاربردهای یادگیری عمیق می‌توان به طبقه‌بندی اشیاء در تصاویر<sup>۱۸</sup>، شناسایی اشیاء در تصاویر<sup>۱۹</sup>، ترجمه از یک زبان به زبان دیگر<sup>۲۰</sup>، تولید تصاویر ساختگی<sup>۲۱</sup> و ... اشاره نمود [۱۵].

به طور کلی می‌توان ارتباط حوزه‌های هوش مصنوعی، یادگیری ماشین و یادگیری عمیق را به صورت شکل ۱-۴ نشان داد. همانطور که ملاحظه می‌شود حوزه‌های یادگیری ماشین و یادگیری عمیق در زیر مجموعه‌ی هوش مصنوعی واقع می‌شود و حوزه‌ی بینایی کامپیوتر علاوه بر داشتن مفاهیم خاص خود در سال‌های اخیر اشتراک نسبتاً زیادی با روش‌های هوش مصنوعی پیدا کرده است (در ادامه به بررسی حوزه‌ی بینایی کامپیوتر پرداخته خواهد شد).

<sup>18</sup>Object Classification<sup>19</sup>Object Recognition<sup>20</sup>Language Translation<sup>21</sup>Fake Image Generation



شکل ۱-۴: ارتباط حوزه‌های هوش مصنوعی، یادگیری ماشین، یادگیری عمیق و بینایی کامپیوتر

## ۱-۲-۲ یادگیری ماشین

یادگیری ماشین شامل سه نوع حوزه‌ی یادگیری است که عبارتند از: یادگیری با سرپرست<sup>۲۲</sup>، یادگیری بدون سرپرست<sup>۲۳</sup> و یادگیری تقویتی<sup>۲۴</sup>، که هدف نهایی این الگوریتم‌ها فراگیری ویژگی‌ها و الگوها از داده‌گان است بدون آنکه به صورت صریح به این ویژگی و الگوها اشاره شده باشد.

### یادگیری با سرپرست

در یادگیری با سرپرست، خروجی مطلوب یا داده‌های Ground Truth در دسترس هستند و در نتیجه با تشکیل یک تابع هزینه بین خروجی مطلوب و خروجی پیش‌بینی شده، با استفاده از روش‌های آموزش مدل‌های یادگیری ماشین الگوریتم آموزش می‌بیند. خروجی مطلوب تعریف شده بسته به وظیفه‌ی مورد نظر می‌تواند گسسته (شماره‌ی مربوط به کلاس‌های مختلف داده‌گان) و یا یک خروجی پیوسته باشد بر

<sup>۲۲</sup>Supervised Learning

<sup>۲۳</sup>Unsupervised Learning

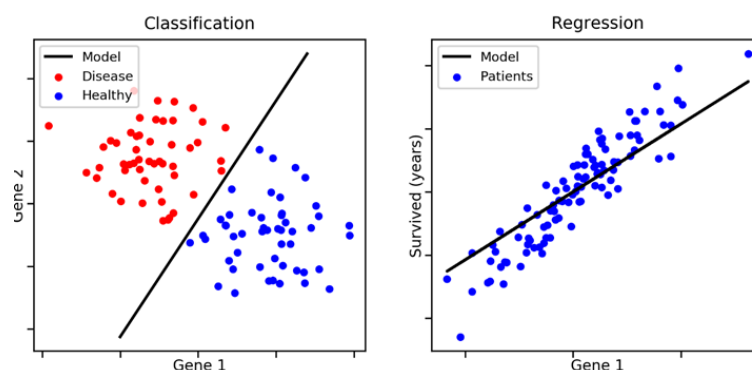
<sup>۲۴</sup>Reinforcement Learning

این اساس یادگیری با سرپرست را می‌توان به دو گروه طبقه‌بندی<sup>۲۵</sup> و رگرسیون<sup>۲۶</sup> تقسیم نمود.

• طبقه‌بندی: در الگوریتم‌های طبقه‌بند هدف نسبت دادن یک مشاهده به کلاس‌های تعریف شده است به عنوان مثال طبقه‌بندی تصاویر تومور پوستی به دو گروه خوش‌خیم و بدخیم یک طبقه‌بندی دو کلاسه است و یا طبقه‌بندی ایمیل‌های دریافتی به دو کلاس ایمیل‌های هرز و ایمیل‌های سالم نیز نمونه‌ای از انجام طبقه‌بندی با روش‌های یادگیری ماشین با سرپرست است.

• رگرسیون: در این نوع مسایل به پیش‌بینی یک مقدار پیوسته نیاز است که در نتیجه خروجی مدل باید یک مقدار پیوسته باشد، به عنوان مثال پیش‌بینی قیمت خانه از روی مترآژ، سال ساخت، منطقه‌ی ساخت و ... یک مساله‌ی رگرسیون است.

برای مقایسه‌ی بهتر این دو گروه، شکل ۱-۵ تصویر کاملی از تفاوت هدف این دو الگوریتم را به نمایش کشیده است. برای طبقه‌بندی، هدف یافتن مرزی است که کلاس‌های مختلف را از هم جدا می‌کند اما در مساله‌ی رگرسیون هدف پیدا کردن یک منحنی است که بهترین تخمین از دادگان را داشته باشد.



شکل ۱-۵: مقایسه‌ی الگوریتم‌های طبقه‌بندی و رگرسیون [۴]

### یادگیری بدون سرپرست

در این نوع از الگوریتم‌های یادگیری خروجی مطلوب در دسترس نیست و یک یادگیری بدون بازخورد است. این نوع یادگیری برای استخراج الگوی و پیوندهای خاص نامعلوم بین دادگان در دسترس، استفاده

<sup>25</sup> Classification

<sup>26</sup> Regression



می‌شود. با تعریف ارایه شده می‌توان این نوع یادگیری را به دو کلاس خوشه‌بندی<sup>۲۷</sup> و تشخیص رابطه<sup>۲۸</sup> تقسیم نمود.

- خوشه‌بندی: در این نوع یادگیری بدون سرپرست هدف پیدا کردن خوشه‌هایی بین دادگان موجود است به طوریکه داده‌های قرار گرفته در این خوشه‌ها دارای ویژگی‌های یکسان باشند. به عنوان مثال با در دست داشتن کل اخبار خبرهای حوزه‌های مختلف از یکدیگر تفکیک داده شوند.
- تشخیص رابطه: هدف در تشخیص رابطه پیدا کردن قانونی است که اکثر دادگان در دسترس، از آن پیروی می‌کنند. به عنوان مثال الگوریتم‌های توصیه کننده در خریدهای اینترنتی محصولاتی که احتمال خرید بیشتری بر اساس جستجوهای قبلی را دارند، نمایش می‌دهند.

### یادگیری تقویتی

در این نوع یادگیری به‌جای استفاده از داده‌های علامت‌گذاری شده و مطلوب از الگوریتم‌های مبتنی بر پاداش و مجازات استفاده می‌شود. به عنوان مثال یک کودک را در نظر بگیرید که برای اولین بار قصد راه رفتن دارد. این کودک اگر به زمین بیفتد ممکن است آسیب ببیند ولی اگر بتواند راه برود با تشویق والدین مواجه می‌شود بنابراین با سعی در بیشینه کردن پاداش خود که همان تشویق و آغوش والدین است و با سعی و خطا اقدام به یادگیری می‌کند. با الگو برداری از چنین روندی الگوریتم‌های یادگیری تقویتی توسعه پیدا کرده‌اند که با سعی و خطا و مشاهدات فراوان، هدف بیشینه کردن پاداش است که در نهایت به یادگیری الگوریتم برای انجام یک عملیات می‌انجامد.

## ۳-۲-۱ یادگیری عمیق

همانند یادگیری ماشین، یادگیری عمیق نیز یک متد یادگیری آماری است که وظیفه‌ی استخراج ویژگی از دادگان خام را دارد. تفاوت اصلی یادگیری عمیق با یادگیری ماشین در استخراج ویژگی توسط شبکات عصبی چند لایه با تعداد زیاد لایه‌های میانی<sup>۲۹</sup> است که پشت سرهم قرار گرفته‌اند. همچنین الگوریتم‌های یادگیری عمیق نسبت به یادگیری ماشین پیچیده‌تر و نیازمند قدرت پردازش بالاتری هستند. بنابراین

<sup>27</sup>Clustering

<sup>28</sup>Association

<sup>29</sup>Hidden Layers

همانطور که پیشتر اشاره گردید، با افزایش تعداد دادگان در دسترس و بالا رفتن قدرت پردازش این حوزه از هوش مصنوعی در دهه‌ی اخیر رشد بسیار چشم‌گیری داشت.

در تعاریف یادگیری عمیق دیدیم که این حوزه زیر مجموعه‌ی یادگیری ماشین قرار می‌گیرد بنابراین تمام متدها و حوزه‌های یادگیری، در یادگیری ماشین، در یادگیری عمیق نیز کاربرد دارد (یادگیری با و بدون سرپرست و یادگیری تقویتی). در ادامه به معرفی چند مدل پر استفاده در حوزه‌ی یادگیری عمیق پرداخته می‌شود.

### شبکه‌های عصبی پرسپترون چند لایه

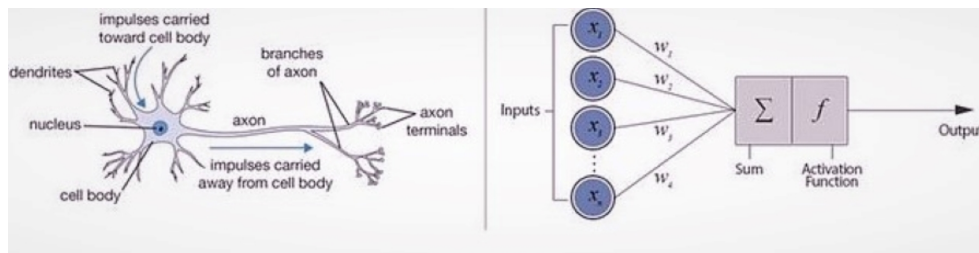
چگونه یک فرد تفاوت میان سگ و گربه را درک می‌کند؟ برای پاسخ به این سوال باید به سراغ بیولوژی شبکه‌ی نورونی داخل مغز برویم. درون مغز انسان به صورت تقریبی حدود ۱۰ میلیارد نورون وجود دارد که هریک از این نورون‌ها به حدود ۱۰ هزار نورون همسایه‌ی خود متصل‌اند. این اتصالات بسیار زیاد و تعداد هنگفت نورون‌ها باعث استخراج ویژگی و یادگیری الگو در داده‌های ورودی حسگرهای بدن می‌شود و بنابراین یک شخص با این ابزار می‌تواند تفاوت میان سگ و گربه را بفهمد و این دو را تفکیک کند.

هر یک از این نورون‌ها سیگنال‌های الکتروشیمیایی را توسط دندریت‌های<sup>۳۰</sup> خود از سایر نورون‌ها دریافت می‌کند و در صورتی که ولتاژ غشای نورون از یک آستانه عبور کند سیگنال‌های دریافتی را توسط آکسون<sup>۳۱</sup> خود به سایر نورون‌ها انتقال می‌دهد. برخلاف تلاش‌های فراوان صورت گرفته برای شناخت عملکرد مغز، هنوز مدل دقیقی برای مطالعه‌ی عملکرد این مجموعه نورونی عظیم ارایه نشده است اما مدل‌سازی یک تک نورون توسط هاچکین-هاکسلی [۱۶] با دقت خوبی ارایه شد. در حوزه‌ی شبکه‌های عصبی مصنوعی از یک مدل بسیار ساده‌تر از مدل هاچکین-هاکسلی استفاده می‌شود که به آن پرسپترون<sup>۳۲</sup> گفته می‌شود. در شکل ۱-۶ یک نورون بیولوژیکی در کنار یک مدل بسیار ساده از نورون نشان داده شده است.

<sup>30</sup>Dendrites

<sup>31</sup>Axon

<sup>32</sup>Perceptron

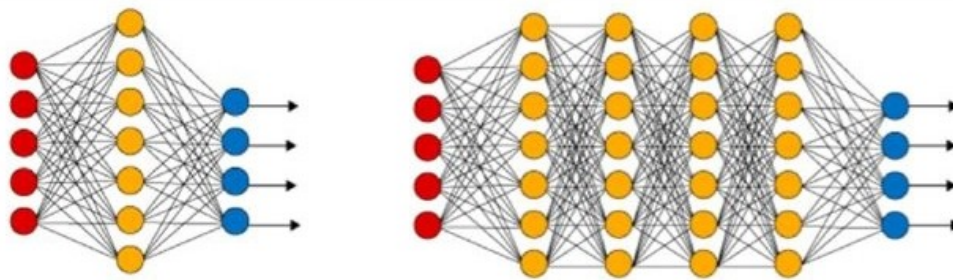


شکل ۱-۶: مقایسه‌ی یک نورون طبیعی و مدل‌سازی ساده‌ی آن [۵]

در این مدل پرسپترون چند نکته حایز اهمیت وجود دارد:

- ورودی: در گره‌های ورودی پرسپترون که به نوعی دندریت‌های نورون هستند مقادیر حقیقی به عنوان میزان عملکرد نورون‌های همسایه به ورودی داده می‌شود. این مقادیر به عنوان مثال می‌تواند شدت روشنایی پیکسل‌های یک تصویر باشد.
  - وزن‌ها: هر ارتباط که از ورودی به این پرسپترون می‌رسد، دارای یک وزن خاص است که می‌تواند هر مقدار حقیقی‌ای باشد. در واقع در روند یادگیری این وزن‌ها هستند که تغییر می‌کنند و نیازمند محاسبات هستند و در نهایت با بدست آمدن وزن‌های بهینه وظیفه‌ی مورد نظر توسط شبکه‌ی عصبی انجام می‌شود.
  - جمع وزن‌دار: پس از آنکه ورودی به پرسپترون داده می‌شود یک جمع وزن‌دار با توجه به وزن ارتباطات نوروونی صورت می‌گیرد و یک مقدار حقیقی بدست می‌آید.
  - تابع فعالیت<sup>۳۳</sup>: همانند یک نورون طبیعی که باید ولتاژ آن از یک آستانه عبور کند تا به اصطلاح آتش کند، یک تابع فعالیت بر سر راه جمع وزن‌دار قرار می‌گیرد تا این مدل‌سازی صورت گیرد.
  - خروجی: در نهایت پس از اعمال تابع فعالیت خروجی نهایی پرسپترون ایجاد می‌شود.
- با قرار دادن تعدادی از این مدل‌های پرسپترون در کنار یکدیگر می‌توان یک شبکه‌ی عصبی ساخت. اگر تعداد لایه‌های زیاد باشد به آن شبکه عصبی عمیق گفته می‌شود. شکل ۱-۷ دو شبکه‌ی تمام متصل کم عمق (سمت چپ) و عمیق (سمت راست) را نشان می‌دهد.

<sup>33</sup> Activation Function



شکل ۱-۷: مقایسه‌ی شبکه‌ی پرسپترون عمیق و کم عمق [۶]

### شبکه‌های عصبی کانولوشنی

این نوع از شبکات از جمله محبوب‌ترین مدل‌ها یادگیری عمیق برای پردازش تصویر و ویدیو در حوزه‌های پردازش تصویر و بینایی کامپیوتر هستند. به دلیل آنکه در این پروژه از این نوع شبکه‌ها استفاده‌ی بسیار زیادی شده است در بخش بعدی با جزئیات بررسی خواهد شد.

### شبکه‌های عصبی بازگشتی

به زبان ساده، شبکه‌های عصبی بازگشتی<sup>۳۴</sup> برای پیش‌بینی وقایع و استخراج ویژگی از ورودی‌های قبلی استفاده می‌کنند. به عنوان مثال اگر بخواهیم حرف آخر یک کلمه را پیش‌بینی کنیم منطقی است که حروف‌های ابتدایی آن کلمه را در حافظه قرار دهیم و با استفاده از اطلاعات آن‌ها اقدام به پیش‌بینی نهایی کنیم. بنابراین شبکه‌های عصبی بازگشتی امکان به خاطر سپاری ورودی‌های گذشته را فراهم می‌سازد که این ویژگی این نوع شبکات را از دیگر مدل‌ها متمایز می‌سازد. کاربرد این مدل‌ها برای داده‌های متوالی مانند حوزه‌ی پردازش زبان طبیعی<sup>۳۵</sup>، پیش‌بینی سیگنال‌های بازار سرمایه، تولید موسیقی و ... می‌باشد.

### شبکه‌های مولد خصمانه

شبکه‌های عصبی مولد خصمانه<sup>۳۶</sup> در سال ۲۰۱۴ توسط گودفلو [۱۷] ارایه شد. این نوع مدل‌ها برخلاف مدل‌های دیگر که فقط از دادگان برای یادگیری، ویژگی استخراج می‌کردند، امکان خلق کردن بر اساس

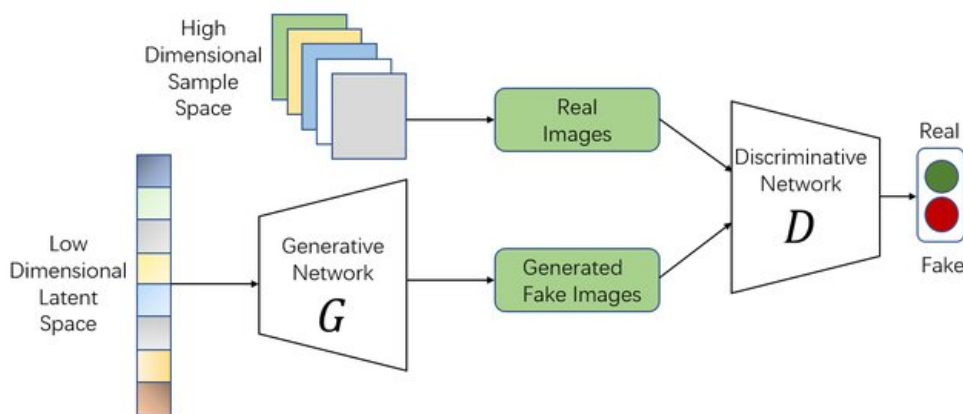
<sup>34</sup>Recurrent Neural Networks

<sup>35</sup>Natural Language Processing

<sup>36</sup>Generative Adversarial Networks (GAN)

دادگان ایجاد گردید. این نوع مدل‌ها از دو شبکه‌ی عصبی تشکیل شده‌اند: شبکه‌ی مولد<sup>۳۷</sup> و شبکه‌ی نقاد<sup>۳۸</sup>. شبکه‌ی مولد یک نویز در ورودی می‌گیرد و با سعی در تخمین توزیع داده‌های آموزش، یک تصویر جعلی ایجاد می‌کند. در دست دیگر شبکه‌ی نقاد سعی در شناسایی داده‌های جعلی از داده‌های حقیقی را دارد. بنابراین یک رقابت بین مولد برای فریب دادن نقاد و نقاد برای تفکیک داده‌های جعلی و حقیقی صورت می‌گیرد.

شکل ۱-۸ چگونگی رقابت بین مولد و نقاد را نشان می‌دهد. همان‌طور که ملاحظه می‌شود، شبکه‌ی نقاد از دو سمت ورودی دریافت می‌کند (داده‌های جعلی و داده‌های حقیقی). در نهایت با تکمیل روند آموزش، شبکه‌ی مولد آنقدر حرفه‌ای می‌شود که با دقت بالایی می‌تواند داده‌هایی تولید کند که نقاد قادر به تشخیص آن از داده‌های حقیقی نباشد. کاربرد این مدل‌ها در خلق دادگان جدید مانند ایجاد تصاویر ساختگی، تولید موسیقی، تولید صدای طبیعی انسان و ... است [۱۸].



شکل ۱-۸: شبکه‌های مولد خصمانه [۷]

### ۳-۱ شبکه‌های عصبی کانوولوشنی عمیق

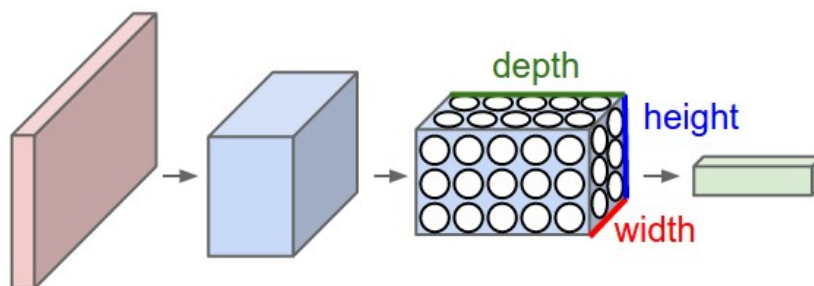
شبکه‌های عصبی کانوولوشنی بسیار مشابه با شبکه‌های تمام متصل و پرسپترون چند لایه هستند که از نورون‌های شامل وزن و بایاس تشکیل شده‌اند. هر نورون یک ورودی دریافت می‌کند و سپس با اعمال یک ضرب داخلی بین ورودی و وزن‌های خود و در ادامه اضافه کردن یک تابع غیرخطی (تابع فعالیت) خروجی خود را تولید می‌کند. در این حالت، تمام شبکه به عنوان یک تابع مشتق‌پذیر عمل می‌کند و در

<sup>37</sup>Generator

<sup>38</sup>Discriminator

نهایت با ایجاد خروجی و تشکیل تابع هزینه با استفاده از الگوریتم Back Propagation شبکه وظیفه‌ی مورد نظر را یاد می‌گیرد.

در شبکه‌های تمام متصل، به تعداد المان‌های ورودی نورون در لایه‌ی ورودی قرار می‌گیرد و در ادامه چندین لایه‌ی مخفی که تمامی نورون‌های هر لایه به لایه‌ی قبل و بعد خود متصل‌اند، بر روی این ورودی فعالیت می‌کنند تا خروجی نهایی تولید گردد. بنابراین با افزایش اندازه‌ی ورودی و نیز تعداد لایه‌ها، تعداد وزن‌هایی که باید در آن‌ها یادگیری صورت گیرد بسیار زیاد خواهند شد. با ظهور شبکه‌های کانوولوشنی علاوه بر رفع مشکل تغییر تعداد پارامترها با تغییر اندازه‌ی ورودی، پیش‌بینی نهایی نسبت به جابجایی و دیگر انتقالات درون تصویر، مقاوم شد. در شبکه‌های کانوولوشنی برخلاف شبکه‌های تمام متصل، نورون‌ها در سه بعد، طول<sup>۳۹</sup> و عرض<sup>۴۰</sup> و عمق<sup>۴۱</sup> چیده شده‌اند و با حرکت بر روی تصویر و استخراج ویژگی از آن خروجی نهایی را ایجاد می‌کنند. شکل ۱-۹ ویژگی‌های استخراج شده از یک تصویر ورودی (حجم قرمز رنگ در تصویر) را نشان می‌دهد که چون نورون‌ها در سه بعد چیده شده‌اند ویژگی‌های استخراج شده نیز سه‌بعدی هستند. با مقایسه‌ی این شکل و شکل ۱-۷ تفاوت دو شبکه‌ی کانوولوشنی و تمام متصل در ابعاد ویژگی‌های استخراج شده مشخص است.



شکل ۱-۹: ماتریس ویژگی‌های سه بعدی استخراج شده توسط یک شبکه‌ی کانوولوشنی [۸]

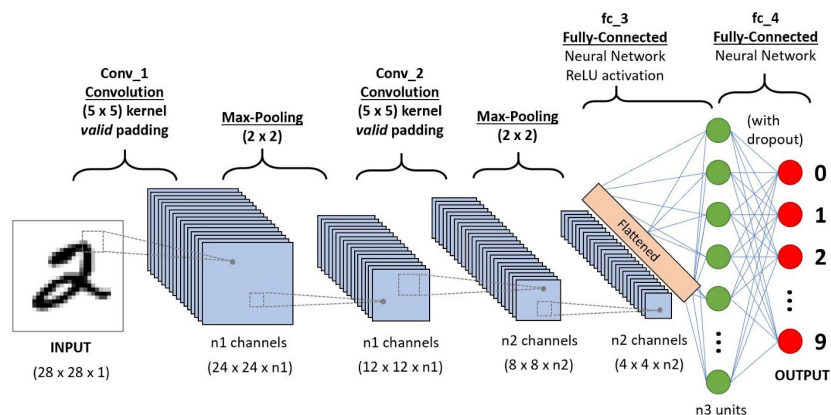
اولین کاربرد شبکه‌های کانوولوشنی در طبقه‌بندی تصاویر صورت گرفت و پر استفاده‌ترین شبکه‌ها در این حوزه از معماری با لایه‌های کانوولوشن، توابع فعالیت، لایه‌های ادغام<sup>۴۲</sup> و لایه‌های تمام متصل برای طبقه‌بندی استفاده می‌شود. در شکل ۱-۱۰ معماری یک شبکه‌ی طبقه‌بند برای طبقه‌بندی اعداد دست‌نویس مشاهده می‌شود که در ادامه به توضیح مختصر هر یک از لایه‌های استفاده شده می‌پردازیم.

<sup>۳۹</sup>Height

<sup>۴۰</sup>Width

<sup>۴۱</sup>Depth

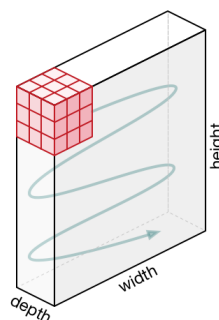
<sup>۴۲</sup>Pooling Layers



شکل ۱-۱۰: معماری یک شبکه‌ی کانولوشنی برای طبقه‌بندی اعداد دست‌نویس [۹]

### ۱-۳-۱ لایه‌های کانولوشنی

لایه‌های کانولوشنی شامل یکسری فیلترهایی هستند که با عمق ماتریس ویژگی برابری می‌کنند اما طول و عرض این فیلترها کوچکتر است و بنابراین با لغزاندن بر روی ماتریس ویژگی و در واقع محاسبه‌ی همبستگی<sup>۴۳</sup> به جای کانولوشن، ماتریس ویژگی جدید محاسبه می‌گردد. به عنوان مثال، اگر ابعاد تصویر ورودی  $3 \times 100 \times 100$  باشد حتماً عمق فیلتر باید با عمق تصویر برابر باشد بنابراین فیلتر می‌تواند ابعاد  $3 \times 5 \times 5$  داشته باشد. شکل ۱-۱۱ یک ماتریس ویژگی و حرکت یک فیلتر بر روی آن را در یک لایه‌ی کانولوشنی را نشان می‌دهد



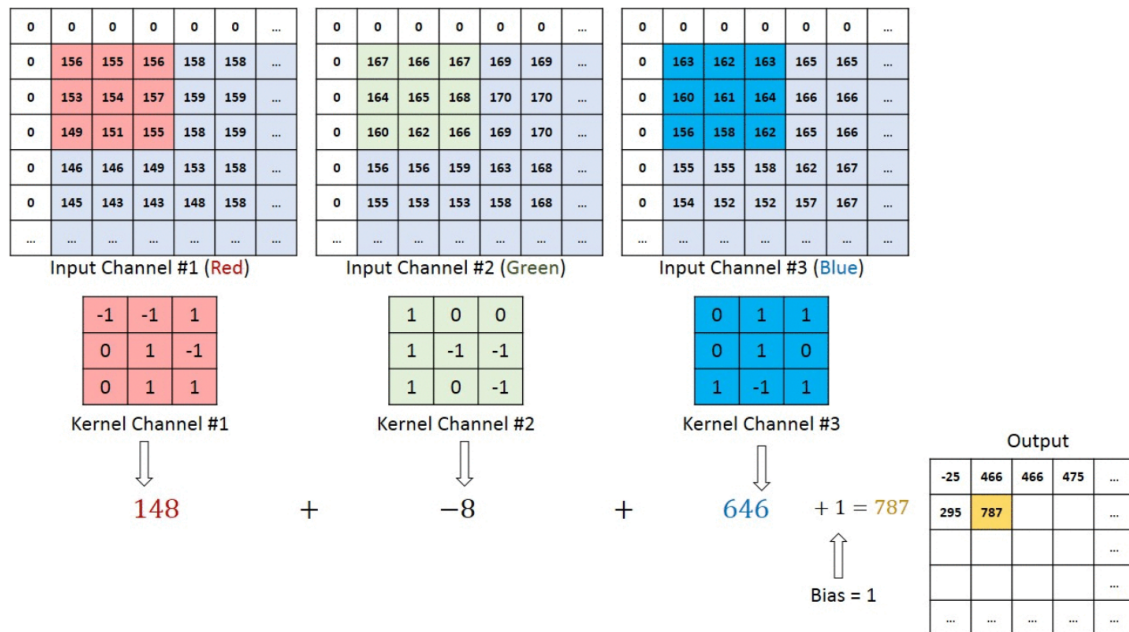
شکل ۱-۱۱: چگونگی حرکت فیلتر در یک لایه‌ی کانولوشنی [۹]

برای درک بهتر چگونگی محاسبات در یک لایه‌ی کانولوشنی، شکل ۱-۱۲ را در نظر بگیرید.

<sup>43</sup>Correlation



همانطور که پیشتر گفته شد، عمق فیلتر و عمق تصویر باید برابر باشد (در اینجا ۳) سپس با لغزیده شدن فیلتر بر روی تصویر مشابه شکل قبل و انجام یک ضرب نقطه‌ای و مجموع‌گیری بین سه کانال و در نهایت اضافه کردن بایاس به مقدار بدست آمده، مقدار خروجی فیلتر محاسبه می‌شود.



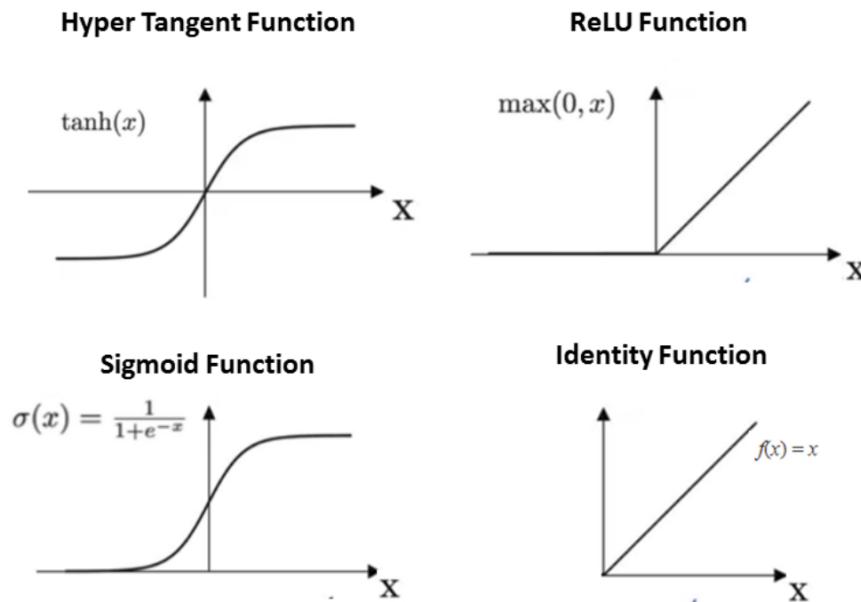
شکل ۱-۱۲: روند محاسبات در یک لایه‌ی کانولوشنی [۹]

## ۱-۳-۲ توابع فعالیت

همانطور که در بخش‌های قبل گفته شد، یک نورون طبیعی با رسیدن به ولتاژ آستانه آتش می‌کند بنابراین برای شبیه‌سازی این رفتار از یک تابع فعالیت استفاده می‌شود. تابع فعالیت در واقع تصمیم می‌گیرد که نورون مدل‌سازی شده با چه مقادیری فعالیت کند و میزان فعالیت آن در شکل‌گیری خروجی نهایی چه مقدار باشد. همچنین تابع فعالیت چون غیر خطی است، امکان تخمین و استخراج ویژگی‌های غیر خطی توسط شبکه فراهم می‌شود. شکل ۱-۱۳ چهار نمونه از پر استفاده‌ترین توابع فعالیت در شبکه‌های عمیق را نشان می‌دهد که عبارتند از: ReLU<sup>۴۴</sup>، Tanh، Sigmoid و Identity. به دلیل آنکه در شبکه‌های عمیق با مشکل محو شدن گرادیان مواجهیم بنابراین استفاده از ReLU برای رفع این مشکل بسیار بیشتر است.

<sup>۴۴</sup>Rectified Linear Unit





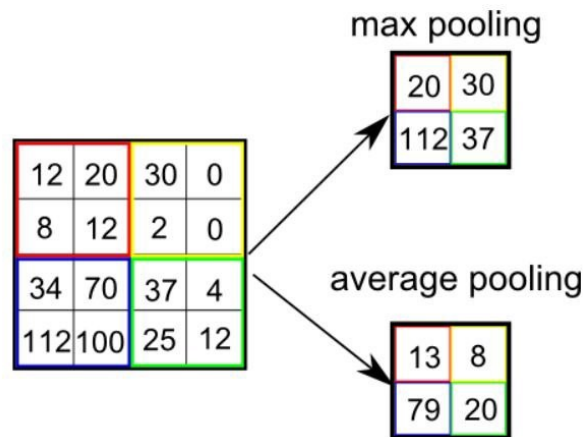
شکل ۱-۱۳: چهار نمونه از پر استفاده‌ترین توابع فعالیت [۱۰]

### ۳-۳-۱ لایه‌های ادغام

لایه‌های ادغام وظیفه‌ی کاهش ابعاد ماتریس‌های ویژگی استخراج شده توسط لایه‌های کانولوشنی را دارند که این کار باعث کاهش حجم محاسبات مورد نیاز و افزایش سرعت خواهد شد. علاوه بر این، لایه‌ها ویژگی‌های غالب را استخراج می‌کنند که نسبت به دوران و جابجایی مقاوم هستند. نحوه‌ی عملکرد لایه‌های ادغام مانند لایه‌های کانولوشنی است و بر روی ماتریس ویژگی لغزانیده می‌شود، با این تفاوت که در لایه‌های ادغام محاسبات نسبت به لایه‌های کانولوشنی متفاوت است. در شکل ۱-۱۴ دو نوع از پر استفاده‌ترین لایه‌های ادغام نشان داده شده است، ادغام میانگین‌گیر<sup>۴۵</sup> و ادغام بیشینه‌گیر<sup>۴۶</sup>. اندازه‌ی فیلترهای ادغام  $2 \times 2$  است و با گام ۲ بر روی ماتریس ویژگی لغزانیده شده‌اند. در ادغام میانگین‌گیر، بین ۴ عدد موجود در فیلتر میانگین‌گیری می‌شود و در ادغام بیشینه‌گیر، بیشینه به خروجی می‌رود.

<sup>۴۵</sup> Average Pooling

<sup>۴۶</sup> Max Pooling



شکل ۱-۱۴: ادغام میانگین‌گیر و ادغام بیشینه‌گیر [۹]

### ۱-۳-۴ لایه‌های تمام متصل

لایه‌های تمام متصل در بخش شبکه‌های پرسپترون چند لایه معرفی شد. به طور معمول این لایه‌ها برای طبقه‌بندی با استفاده از ویژگی‌های استخراج شده از دادگان استفاده می‌شوند. پیش از ظهور شبکه‌های عصبی عمیق که خودشان ویژگی‌های لازم را استخراج می‌کنند، در روش‌های یادگیری ماشین، ویژگی‌های یک مجموعه داده به صورت دستی و یا با روش‌های کلاسیک استخراج می‌شد و با استفاده از شبکه‌های تمام متصل وظیفه طبقه‌بندی تکمیل می‌شد. در شبکه‌های کانوولوشنی این ویژگی‌ها توسط لایه‌های کانوولوشن و ادغام استخراج می‌شود و در نهایت با دادن این ویژگی‌ها به لایه‌های تمام متصل کار طبقه‌بندی انجام می‌شود [۸]. لایه‌های آخر شکل ۱-۱۰ این لایه را نشان می‌دهد.

از جمله شبکه‌های موفق کانوولوشنی در حوزه طبقه‌بندی تصاویر طبیعی می‌توان به AlexNet که در سال ۲۰۱۲ ارایه شد و انقلاب شبکه‌های عمیق را آغاز کرد، اشاره نمود [۱۹]. پس از ارایه این مدل، معماری‌های جدید و روش‌های تعمیم‌پذیری<sup>۴۷</sup> زیادی ارایه گردید که نتایج را در زمینه طبقه‌بندی تصاویر حتی از انسان نیز بهتر کرد. از جمله این معماری‌ها می‌توان به معماری‌های VGGNet، ResNet، GoogleNet و ... اشاره نمود [۲۰].

<sup>47</sup>Regularization Terms

## ۴-۱ روش‌های تعمیم‌پذیری

به هرگونه تکنیکی که قدرت تعمیم مدل را افزایش می‌دهد و خطای تعمیم‌پذیری را کم می‌کند، روش تعمیم‌پذیری گفته می‌شود. از جمله‌ی این روش‌ها می‌توان به موارد زیر اشاره نمود:

- جریمه‌ی وزن‌های شبکه<sup>۴۸</sup>
- توقف زود هنگام<sup>۵۱</sup>
- افزایش مجموعه دادگان<sup>۴۹</sup>
- روش Dropout
- مقاومت در برابر نویز<sup>۵۰</sup>
- روش Batch-Normalization

در ادامه به معرفی مختصر هر یک از این روش‌ها پرداخته خواهد شد.

### جریمه‌ی وزن‌های شبکه

برای جلوگیری از یادگیری بیش‌ازحد<sup>۵۲</sup> بر روی داده‌های آموزش یکی از روش‌ها، جریمه‌ی وزن‌های شبکه است به این صورت که با اضافه کردن یک جمله به تابع هزینه‌ی شبکه، که آن جمله می‌تواند نرم اول یا دوم وزن‌های شبکه باشد؛ از بزرگ شدن بیش از حد وزن‌ها جلوگیری می‌کند و همچنین با کاهش پیچیدگی مدل، قدرت تعمیم‌پذیری آن بالا می‌رود.

### افزایش مجموعه دادگان

در مدل‌های عصبی، هرچه قدر تنوع و تعداد دادگان بیشتر باشد در نهایت مدل بهتر آموزش می‌بیند بنابراین روش‌هایی برای افزایش تعداد دادگان پیشنهاد شده است که به عنوان مثال می‌توان به دوران با زوایای تصادفی، چرخش عمودی و افقی، تغییرات شدت روشنایی و ... اشاره نمود.

<sup>48</sup>Parameter (weights) Penalties

<sup>49</sup>Dataset Augmentation

<sup>50</sup>Noise Robustness (input/output)

<sup>51</sup>Early Stopping

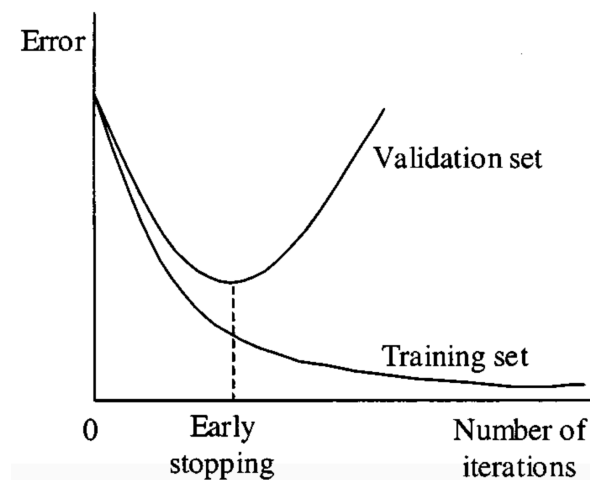
<sup>52</sup>Overfitting

## مقاومت در برابر نویز

با اضافه کردن نویز به ورودی یا خروجی مدل و بالا بردن عدم قطعیت، شبکه سعی در استخراج ویژگی‌های مقاوم‌تری می‌کند و در نتیجه مدل نهایی قدرت تعمیم‌پذیری بالاتری دارد [۲۱].

## توقف زود هنگام

اگر دادگان آموزش را به دو مجموعه داده‌ی آموزش و ارزیابی تقسیم کنیم (بر روی دادگان آموزش وزن‌های شبکه تغییر می‌کنند و دادگان ارزیابی در تغییر وزن‌ها نقشی ندارند) و در انتهای هر  $epoch^{۵۳}$  مقدار خطای شبکه را برای این دو گروه محاسبه کنیم نموداری مشابه شکل زیر بدست می‌آید که نشان می‌دهد پس از مدتی، شبکه در حال حفظ کردن دادگان آموزش است و خطای دادگان آموزش افزایش می‌یابد بنابراین قدرت تعمیم‌پذیری شبکه در حال کم شدن است. به توقف روند آموزش بعد از مشاهده روند افزایشی خطای تعمیم‌پذیری، توقف زود هنگام گفته می‌شود.

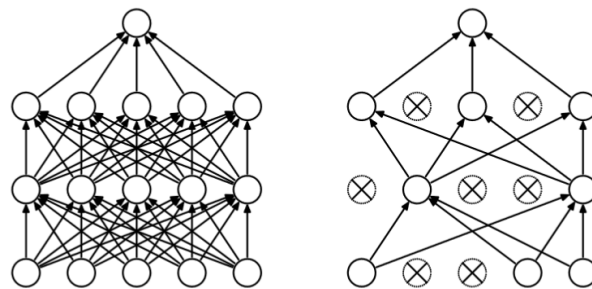


شکل ۱-۱۵: توقف زود هنگام با توجه به خطای داده‌های ارزیابی [۱۱]

## روش Dropout

اگر در یک مدل عصبی دو نورون ویژگی‌هایی که استخراج می‌کنند مشابه باشند و یا همبستگی بالایی داشته باشند این شبکه به صورت کارآمد عمل نمی‌کند. برای رفع این مشکل تکنیک dropout پیشنهاد  $^{۵۳}$  به هر بار آموزش یک مدل بر روی کل مجموعه دادگان یک epoch گفته می‌شود.

شد که با یک درصد احتمال در هر تکرار آموزش، تعدادی از نورون‌ها را غیر فعال می‌کند بنابراین نورون‌های دخیل در روند آموزش مجبور به یادگیری ویژگی‌های مستق از هم و کارآمدتر هستند و در نتیجه قدرت تعمیم‌پذیر مدل افزایش پیدا می‌کند. در شکل ۱-۱۶ یک شبکه با و بدون استفاده از dropout نشان داده شده است که در استفاده از dropout تعدادی از نورون‌ها حذف شده‌اند و نورون‌های باقی مانده مجبور به استخراج ویژگی‌های کارآمد برای پیش‌بینی نتیجه‌ی مطلوب هستند.



شکل ۱-۱۶: چگونگی اعمال تکنیک dropout [۱۲]

### روش Batch-Normalization

این متد مشکل ناهمگونی تابع چگالی احتمال داده‌ها در لایه‌های میانی را حل می‌کند. این مشکل به این صورت پدید می‌آید که در هر batch از داده‌ها، در روند آموزش یک توزیع احتمال پدید می‌آید که این توزیع‌های احتمال با یکدیگر متفاوت‌اند. مشکل مهم‌تر آنکه، این ناهمگونی توزیع‌ها در خروجی لایه‌های شبکه نیز پدید می‌آید. برای رفع این مشکل تکنیک Batch-Normalization معرفی گردید که ابتدا تمام توزیع‌ها را با توجه به میانگین و انحراف معیارشان نرمال‌سازی می‌کنند و سپس با یک سری پارامترهای قابل یادگیری این توزیع‌ها را تغییر می‌دهند تا یکسان‌سازی نهایی صورت پذیرد [۲۲].

## ۵-۱ پردازش تصویر و بینایی کامپیوتر

چشم انسان از تعداد بسیار زیادی سلول حساس به نور تشکیل شده‌است که با دریافت نور محیطی و پردازش آن باعث ایجاد یک درک از محیط و اشیای اطراف می‌شود. این روند علی‌رغم اینکه برای انسان بسیار ساده است اما برای یک ماشین تا بتواند درک از تصویر دریافت شده داشته باشد، بسیار پیچیده است. بنابراین زمینه‌های مختلفی برای پردازش و ایجاد درک از تصویر دیجیتال برای ماشین‌ها ایجاد

شده است.

در بینایی کامپیوتر مشابه قدت بینایی انسان، هدف ایجاد درک و فهم عمیق یک تصویر یا یک ویدیو است اما پردازش تصاویر دیجیتال، پردازش تصاویر با استفاده از یک کامپیوتر است. این پردازش‌ها می‌تواند شامل بهبود کیفیت تصویر و حذف نویز با فیلتر کردن تصویر و دستکاری شدت‌های آن باشد. بنابراین زمینه‌ی پردازش تصویر زیر مجموعه‌ی بینایی کامپیوتر قرار می‌گیرد.

حوزه‌ای بینایی کامپیوتر مرتبط با مبحث هوش مصنوعی و یادگیری ماشین است که با هدف ایجاد چارچوب‌های لازم برای پیاده‌سازی قابلیت بینایی در کامپیوترها و سیستم‌های کامپیوتری شکل گرفته است. دانشمندان فعال در حوزه بینایی کامپیوتر، سعی در تولید تکنیک‌ها و روش‌هایی دارند که مفهوم دیدن را برای یک کامپیوتر تعریف می‌کنند. از این طریق، کامپیوترها توانایی شناسایی و درک محتوای موجود در تصاویر دیجیتال و ویدیوها را پیدا می‌کنند [۲۳].

## ۱-۶ جمع‌بندی

در این فصل با مفاهیم اولیه‌ی این مطالعه آشنایی پیدا شد. در ابتدا مفاهیم تصویربرداری پزشکی معرفی گردید و دو نوع تصویربرداری پرکاربرد (ام‌آزای و سی‌تی اسکن) برای تشخیص و درمان به صورت دقیق‌تر بررسی گردید. در ادامه مفاهیم هوش مصنوعی، یادگیری ماشین و یادگیری عمیق معرفی گردید که با توجه به رشد روز افزون این حوزه‌ها و مدل‌های پرکاربرد، استفاده‌ی این مدل‌ها در بینایی کامپیوتر و پردازش تصویر بسیار رشد نموده است. بنابراین در این پروژه در نقطه‌ی مشترک حوزه‌های معرفی شده قرار گرفته‌ایم و قصد بر آن است با استفاده از مدل‌های یادگیری عمیق در حوزه‌ی پردازش تصاویر پزشکی، قطعه‌بندی تومور و ساختارهای در ریسک را با دقت و سرعت بالاتر نسبت به الگوریتم‌های پیشین، انجام دهیم. در فصل بعد به بررسی و مرور سوابق کارهای انجام شده در حوزه‌ی قطعه‌بندی تصاویر پزشکی پرداخته خواهد شد.

## مراجع

- [1] Pgi gets state-of-the-art ct scan machine, <https://www.tribuneindia.com/news/punjab/pgi-gets-state-of-the-art-ct-scan-machine-179634>. Accessed: 2022-01-21.
- [2] Ct of a normal abdomen and pelvis, thumbnail, <https://commons.wikimedia.org/wiki/>. Accessed: 2022-01-21.
- [3] Magnetic resonance imaging (mri), <https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri>. Accessed: 2022-01-22.
- [4] How to use data science to understand what makes wine taste good, <https://www.freecodecamp.org/news/using-data-science-to-understand-what-makes-wine-taste-good-669b496c67ee>. Accessed: 2022-03-29.
- [5] B. M. Njogholo. *Investigating keystroke dynamics as a two-factor biometric security*. PhD thesis, Strathmore University, 2018.
- [6] J. M. Johnson and T. M. Khoshgoftaar. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1):1–54, 2019.
- [7] L. Cai, Y. Chen, N. Cai, W. Cheng, and H. Wang. Utilizing amari-alpha divergence to stabilize the training of generative adversarial networks. *Entropy*, 22(4):410, 2020.
- [8] Convolutional neural networks, <https://cs231n.github.io/convolutional-networks/convert>. Accessed: 2022-04-02.
- [9] A comprehensive guide to convolutional neural networks — the eli5 way, <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>. Accessed: 2022-04-03.

- [10] Activation functions, <https://www.datacamp.com/community/tutorials/neural-network-models-r>. Accessed: 2022-04-03.
- [11] R. Gençay and M. Qi. Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging. *IEEE Transactions on Neural Networks*, 12(4):726–734, 2001.
- [12] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.
- [13] A. Elangovan and T. Jeyaseelan. Medical imaging modalities: a survey. In *2016 International Conference on emerging trends in engineering, technology and science (ICETETS)*, pages 1–4. iee, 2016.
- [14] Computed tomography (ct), <https://www.nibib.nih.gov/science-education/science-topics/computed-tomography-ct>. Accessed: 2022-01-21.
- [15] J. M. Helm, A. M. Swiergosz, H. S. Haeberle, J. M. Karnuta, J. L. Schaffer, V. E. Krebs, A. I. Spitzer, and P. N. Ramkumar. Machine learning and artificial intelligence: definitions, applications, and future directions. *Current reviews in musculoskeletal medicine*, 13(1):69–76, 2020.
- [16] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500, 1952.
- [17] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.
- [18] Notes on artificial intelligence, machine learning and deep learning for curious people, <https://towardsdatascience.com/notes-on-artificial-intelligence-ai-machine-learning-ml-and-deep-learning-dl-for-56e51a2071c2>. Accessed: 2022-03-28.
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.



- 
- [20] W. Wang, Y. Yang, X. Wang, W. Wang, and J. Li. Development of convolutional neural network and its application in image classification: a survey. *Optical Engineering*, 58(4):040901, 2019.
  - [21] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–115, 2021.
  - [22] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.
  - [23] Difference between image processing and computer vision, <https://www.geeksforgeeks.org/difference-between-image-processing-and-computer-vision/>: :text=image Accessed: 2022-04-05.

## **Abstract**

We present a standard template for typesetting theses in Persian. The template is based on the X<sub>Y</sub>TeX Persian package for the L<sup>A</sup>T<sub>E</sub>X typesetting system. This write-up shows a sample usage of this template.

**Keywords:** Thesis, Typesetting, Template, X<sub>Y</sub>TeX Persian



Sharif University of Technology  
Department of Electrical Engineering

M.Sc. Thesis

# **Organs at Risk (OAR) segmentation using machine learning methods**

By:

**Reza Karimzadeh**

Supervisor:

**Dr. Emad Fatemizadeh**

February 2022