





دانشگاه اصفهان

دانشکده مهندسی کامپیوتر

## پایان نامه کارشناسی

رشته مهندسی کامپیوتر

بسته اصلی: هوش مصنوعی و رباتیکز

بسته فرعی: نرم افزار

### عنوان

پویشگر و تحلیلگر ارزش مواد غذایی

استاد راهنما:

سرکارخانم دکتر افسانه فاطمی

پژوهشگر:

علی پورقیصری

شهریور ۱۴۰۳





دانشگاه اصفهان

دانشکده مهندسی کامپیوتر

## پروژه کارشناسی رشته‌ی مهندسی کامپیوتر

آقای علی پورقیصری

تحت عنوان

پویشگر و تحلیلگر ارزش مواد غذایی

در تاریخ / / ۱۴ توسط هیأت داوران زیر بررسی و با نمره به تصویب نهایی رسید.

۱- استاد راهنمای پروژه

دکتر

امضا

۲- استاد داور

دکتر

امضا

امضای مدیرگروه



## تشکر و قدردانی

سپاس بیکران از پروردگار که بار دیگر فرصت یادگیری را به من ارزانی داشت. از پدر و مادر عزیز و مهربانم که در طول زندگی با تحمل زحمات فراوان، همواره برای موفقیت و خوشبختی من تلاش کردند، صمیمانه قدردانی می‌کنم. همچنین از استاد ارجمند، دکتر افسانه فاطمی، که با تلاش‌های بی‌دریغ خود در تحقق این پایان‌نامه سهم بسزایی داشتند، نهایت تشکر را دارم. قدردان دوستان عزیزم نیز هستم که با حمایت‌هایشان در به سرانجام رساندن این پایان‌نامه نقش مهمی ایفا کردند.

لازم می‌دانم از تمامی افرادی که در بهبود و توسعه هوش مصنوعی در زبان فارسی سهمی داشته‌اند، تشکر کنم و امیدوارم که با تلاش توسعه‌دهندگان، دیگر هیچ محدودیتی برای انجام پژوهش‌های فارسی در مقایسه با سایر زبان‌ها وجود نداشته باشد.



تقدیم به  
خانواده عزیزم

## چکیده:

در این پروژه، یک برنامه کاربردی توسعه داده شده است که به کاربران امکان می‌دهد با استفاده از تکنیک‌های پیشرفته پردازش تصویر و بینایی ماشین، اطلاعات مهمی را از روی برچسب‌های ارزش غذایی محصولات فروشگاه استخراج و بخوانند. این برنامه با پوشش برچسب‌های موجود بر روی محصولات، اطلاعاتی نظیر میزان انرژی محصول برحسب کیلوکالری، و مقدار مواد مختلف از جمله قند، نمک، چربی، و اسیدهای چرب ترانس برحسب گرم را نمایش می‌دهد. علاوه بر این، برنامه توانایی تشخیص و ارائه مقدار ماده معیار در محصول را برحسب گرم یا میلی‌لیتر دارد، که می‌تواند به کاربران در تحلیل بهتر ارزش غذایی محصول کمک کند.

در مرحله بعد، برنامه با استفاده از اطلاعاتی که به دست آورده، دست به تجزیه و تحلیل می‌زند و مقدار انرژی مصرفی را با مقادیری از فعالیت‌های بدنی روزمره مقایسه می‌کند. نهایتاً، برنامه شماره پروانه بهداشتی ساخت محصول را نیز تشخیص می‌دهد و با استعلام این شماره، کاربران قادر خواهند بود به اطلاعات فنی دقیق‌تری همچون نام تجاری محصول، نام کارخانه تولیدکننده، تاریخ دریافت و انقضای پروانه بهداشتی، و آدرس کارخانه دسترسی پیدا کنند. هدف اصلی این پروژه، افزایش آگاهی مصرف‌کنندگان در انتخاب محصولات سالم‌تر و بهبود کیفیت خرید آنها با ارائه اطلاعات شفاف و قابل اعتماد می‌باشد.



## فهرست مطالب

صفحه

عنوان

۱۰	فصل اول مقدمه و بیان مسئله
۱۰	۱-۱- توضیحات کلی
۱۱	۲-۱- اهداف پروژه
۱۲	۳-۱- دامنه و محدوده پروژه
۱۸	فصل دوم ادبیات پژوهش
۱۸	۱-۲- مقدمه
۲۹	۴-۲- ادبیات ارزیابی
۳۲	۵-۲- تجزیه و تحلیل اطلاعات
۳۴	۶-۲- جمع بندی:
۳۵	فصل سوم پیشینه پژوهش
۳۵	۱-۳- مقدمه
۳۵	۲-۳- تاریخچه و توسعه OCR
۳۶	۳-۳- مروری بر مطالعات مرتبط
۳۸	۴-۳- جمع بندی
۳۹	فصل چهارم ساختار و آموزش مدل‌ها
۳۹	۱-۴- مقدمه
۳۹	۲-۴- مدل تشخیص برچسب
۴۷	۵-۴- جمع بندی
۴۸	فصل پنجم ارزیابی مدل‌ها و استفاده از اطلاعات
۴۸	۱-۵- مقدمه
۴۸	۲-۵- ارزیابی مدل تشخیص برچسب
۵۰	۳-۵- ارزیابی مدل تشخیص کلمات
۵۱	۴-۵- ارزیابی مدل خوانش حروف و ارقام
۵۳	۵-۵- تجزیه و تحلیل انرژی محصول



۵۴	۳-۲- اطلاعات فنی محصول .....
۵۵	۵-۶- جمع بندی .....
۵۶	فصل ششم جمع بندی و پیشنهادهایی برای بهبود پژوهش .....
۵۶	۶-۱- جمع بندی کلی .....
۵۷	۶-۲- بهبودهای قابل اعمال .....
۵۹	منابع .....





## فهرست اشکال

صفحه

عنوان

تصویر ۱- برچسب استاندارد مد نظر در این پروژه.....	۱۴
تصویر ۲- نمونه ای از برچسب غیر استاندارد.....	۱۵
تصویر ۳- نمونه ای از برچسب با عکس برداری نامناسب.....	۱۵
تصویر ۴- نمونه ای از برچسب با اعداد ناخوانا.....	۱۶
تصویر ۵- نمونه ای از برچسب مخدوش.....	۱۶
تصویر ۶- محیط لیبیل استودیو و فرایند برچسب گذاری.....	۲۱
تصویر ۷- محیط وبسایت سایتاک.....	۲۲
تصویر ۸- ساختار کلی مدل یولو.....	۲۶
تصویر ۹- ساختار گرفت.....	۲۹
تصویر ۱۰- مقایسه سرعت و تعداد پارامترهای مدل خوانش متن.....	۲۹
تصویر ۱۱- برچسب ارزش مواد غذایی آمریکایی.....	۳۷
تصویر ۱۲- قالب قابل قبول یولو.....	۴۱
تصویر ۱۳- عملیات تشخیص برچسب.....	۴۱
تصویر ۱۴- یک ردیف از برچسب هر عکس.....	۴۲
تصویر ۱۵- مدل گرفت، قبل از آموزش.....	۴۲
تصویر ۱۶- قالب قابل قبول گرفت.....	۴۳
تصویر ۱۷- تشخیص کلمات و ارقام به وسیله گرفت.....	۴۴
تصویر ۱۸- برچسب های تصویر قبل.....	۴۴
تصویر ۱۹- خروجی مدل اوسی آر قبل از آموزش.....	۴۵
تصویر ۲۰- خروجی مدل اوسی آر بعد از آموزش.....	۴۶
تصویر ۲۱- قالب قابل قبول مدل اوسی آر.....	۴۷
تصویر ۲۲- نمودار سه معیار دقت، صحت و بازخوانی برای مدل یولو.....	۴۹
تصویر ۲۳- نمودار سه معیار دقت، صحت و میانگین هارمونیک برای مدل گرفت.....	۵۰
تصویر ۲۴- نمودار معیارهای دقت و میانگین خطای نرمالیزه شده در OCR.....	۵۱
تصویر ۲۵- تابع زیان داده های آموزش و آزمایش برای مدل OCR.....	۵۲
تصویر ۲۶- خروجی برنامه بعد از تجزیه و تحلیل انرژی هر سهم محصول.....	۵۴
تصویر ۲۷- اطلاعات فنی و نمای وبسایت پس از استعلام محصول.....	۵۵



## فهرست جداول

عنوان

صفحه

جدول ۱- ماتریس درهم‌ریختگی.....	۳۰
جدول ۲- انواع مدل‌های یولو.....	۳۹
جدول ۳- مقایسه آمار اولین و آخرین دوره آموزشی در یولو.....	۴۹
جدول ۴- مقایسه آمار اولین و آخرین دوره آموزشی در گرفت.....	۵۱
جدول ۵- مقایسه آمار دقت و میانگین خطای نرمالیزه‌شده در اولین و آخرین دوره.....	۵۲
جدول ۶- مقایسه آمار مقادیر توابع زیان داده‌های آموزش و آزمایش در اولین و آخرین دوره.....	۵۳



## فصل اول

### مقدمه و بیان مسئله

#### ۱-۱- توضیحات کلی

در دنیای امروز که سرعت زندگی و مشغله‌های روزمره افزایش یافته، اهمیت دسترسی به اطلاعات دقیق و قابل اطمینان درباره محصولات غذایی بیشتر از همیشه احساس می‌شود. مصرف‌کنندگان به دنبال اطلاعاتی هستند که به آنها کمک کند تا با آگاهی کامل‌تر محصولات غذایی مناسب‌تر و سالم‌تری را انتخاب کنند. این موضوع به ویژه با توجه به افزایش نگرانی‌ها درباره تأثیرات مضر مواد غذایی نظیر قند، نمک، چربی‌های اشباع و اسیدهای چرب ترانس بر سلامت انسان، اهمیت بیشتری پیدا کرده است.

یکی از چالش‌های اساسی مصرف‌کنندگان در این زمینه، عدم دسترسی سریع و آسان به اطلاعات دقیق و جامع درباره ارزش غذایی محصولات است. برچسب‌های مواد غذایی که به طور معمول بر روی محصولات درج می‌شوند، به دلیل پیچیدگی اطلاعات یا اندازه کوچک نوشته‌ها، ممکن است برای بسیاری از مصرف‌کنندگان خوانا نباشند یا به راحتی قابل فهم نباشند. علاوه بر این، عدم آگاهی کافی از محتوای دقیق محصولات می‌تواند منجر به انتخاب‌های نادرست و در نهایت تهدید سلامت مصرف‌کنندگان شود.

از سوی دیگر، نیاز به دسترسی به اطلاعات فنی محصولات، نظیر نام تجاری، کارخانه تولیدکننده، و تاریخ انقضا، از اهمیت بسزایی برخوردار است. این اطلاعات نه تنها به مصرف‌کنندگان کمک می‌کند تا از کیفیت و ایمنی محصولات مطمئن شوند، بلکه امکان پیگیری و استعلام شماره پروانه بهداشتی ساخت محصول را نیز فراهم می‌آورد.

پروژه حاضر با هدف ارائه راه‌حلی نوآورانه برای این چالش‌ها طراحی شده است. در این پروژه، یک برنامه کاربردی توسعه داده شده است که به کاربران امکان می‌دهد با پوشش برچسب‌های ارزش غذایی محصولات فروشگاهی، اطلاعات مندرج



بر روی آنها را با استفاده از تکنولوژی پردازش تصویر<sup>۱</sup> و بینایی ماشین<sup>۲</sup> به صورت خودکار بخوانند. این برنامه قادر است میزان انرژی، قند، نمک، چربی‌ها، اسیدهای چرب ترانس و سایر اطلاعات مهم تغذیه‌ای را استخراج کرده و به شکلی ساده و قابل فهم در اختیار کاربران قرار دهد. علاوه بر این، امکان استعمال شماره پروانه بهداشتی ساخت محصول نیز فراهم شده است که به کاربران اجازه می‌دهد اطلاعات فنی محصول را بررسی کرده و از صحت و سلامت آن اطمینان حاصل کنند.

این پروژه با هدف تسهیل دسترسی مصرف‌کنندگان به اطلاعات تغذیه‌ای و فنی محصولات غذایی، افزایش آگاهی عمومی، و ارتقای سلامت جامعه، نقش مهمی در انتخاب‌های آگاهانه‌تر و سالم‌تر ایفا می‌کند.

## ۲-۱- اهداف پروژه

پروژه حاضر با هدف ارتقای سطح آگاهی و توانمندسازی مصرف‌کنندگان در انتخاب محصولات غذایی مناسب و سالم‌تر طراحی شده است. برای دستیابی به این هدف کلی، اهداف جزئی‌تری نیز تعریف شده‌اند که در ادامه به تفصیل به آنها پرداخته می‌شود:

### ۱-۲-۱- افزایش آگاهی مصرف‌کنندگان از محتوای تغذیه‌ای محصولات غذایی

هدف اصلی این پروژه فراهم کردن دسترسی آسان و سریع مصرف‌کنندگان به اطلاعات تغذیه‌ای محصولات غذایی است. با استفاده از این برنامه، کاربران می‌توانند به راحتی میزان انرژی، قند، نمک، چربی‌ها، اسیدهای چرب ترانس و سایر مواد مغذی را مشاهده و تحلیل کنند. این اطلاعات به مصرف‌کنندگان کمک می‌کند تا با شناخت بهتری از محتوای محصولات، تصمیمات بهتری برای سلامتی خود بگیرند.

### ۲-۲-۱- تسهیل فرآیند انتخاب محصولات غذایی سالم و مناسب

یکی دیگر از اهداف مهم این پروژه، ساده‌سازی فرآیند انتخاب محصولات غذایی سالم‌تر است. با استفاده از تکنولوژی پردازش تصویر و بینایی ماشین، برنامه قادر است اطلاعات پیچیده و گاه ناخوانای برچسب‌های مواد غذایی را به سرعت و به صورتی قابل فهم در اختیار کاربران قرار دهد. این امر به کاربران امکان می‌دهد تا در کوتاه‌ترین زمان ممکن، محصولاتی را انتخاب کنند که با نیازها و اهداف سلامتی آنها هماهنگ باشد.

### ۳-۲-۱- ارائه اطلاعات فنی و بهداشتی محصولات

علاوه بر اطلاعات تغذیه‌ای، این پروژه به کاربران اجازه می‌دهد تا با بررسی شماره پروانه بهداشتی ساخت محصول، به اطلاعات فنی مهمی نظیر نام تجاری محصول، نام کارخانه تولیدکننده، تاریخ دریافت و انقضا پروانه بهداشتی، و آدرس کارخانه دسترسی پیدا کنند. این اطلاعات به کاربران کمک می‌کند تا از کیفیت و ایمنی محصولات مطمئن شوند و انتخاب‌های آگاهانه‌تری داشته باشند.

<sup>1</sup> Image processing

<sup>2</sup> Computer vision



#### ۱-۲-۴- افزایش شفافیت در صنعت غذا

یکی دیگر از اهداف این پروژه، افزایش شفافیت در بازار محصولات غذایی است. با ارائه اطلاعات دقیق و معتبر درباره محتوا و مشخصات فنی محصولات، این پروژه به تقویت اعتماد مصرف‌کنندگان نسبت به اطلاعات ارائه‌شده توسط تولیدکنندگان کمک می‌کند و به ایجاد بازاری شفاف‌تر و قابل‌اعتمادتر منجر می‌شود.

#### ۱-۲-۵- حمایت از تولیدکنندگان متعهد به کیفیت

این پروژه می‌تواند به تولیدکنندگان محصولات غذایی که به کیفیت و شفافیت اطلاعات محصولات خود اهمیت می‌دهند، کمک کند تا در بازار رقابتی جایگاه بهتری پیدا کنند. با ارائه ابزارهایی که به مصرف‌کنندگان امکان دسترسی به اطلاعات دقیق و قابل‌اعتماد را می‌دهد، این پروژه به ترویج محصولات باکیفیت و افزایش اعتماد مشتریان به برندهای معتبر کمک می‌کند.

#### ۱-۲-۶- بهبود سلامت عمومی جامعه

هدف نهایی این پروژه ارتقای سلامت عمومی جامعه است. با افزایش دسترسی مصرف‌کنندگان به اطلاعات تغذیه‌ای و فنی دقیق، این پروژه به کاهش مصرف مواد مضر مانند قند، نمک و چربی‌های ناسالم کمک می‌کند. این امر می‌تواند به کاهش بروز بیماری‌های مرتبط با تغذیه نادرست، نظیر چاقی، دیابت، و بیماری‌های قلبی و عروقی منجر شود و در نهایت سطح کلی سلامت جامعه را بهبود بخشد.

این اهداف در مجموع به تحقق یک محیط غذایی سالم‌تر و شفاف‌تر کمک کرده و مصرف‌کنندگان را قادر می‌سازد تا با اطمینان و آگاهی بیشتری محصولات غذایی خود را انتخاب کنند.

#### ۱-۳-۳- دامنه و محدوده پروژه

##### ۱-۳-۱- جمع‌آوری و آماده‌سازی داده‌ها

- **مجموعه داده:** برای آموزش و ارزیابی مدل‌های مورد استفاده، مجموعه‌ای از ۱۰۰۰ تصویر از محصولات غذایی مختلف از چندین فروشگاه جمع‌آوری شده است. این تصاویر به منظور تشخیص برچسب محصول، تشخیص نوشته‌های داخل برچسب، و خواندن متن‌ها، برچسب‌گذاری<sup>۳</sup> و آماده‌سازی شده‌اند. این مجموعه داده برای آموزش سه مدل که در فصل‌های بعد به آنها اشاره خواهیم کرد، استفاده خواهد شد. در اینجا، استفاده از این تصاویر و برچسب‌گذاری آنها به ترتیب آمده است.

<sup>۳</sup> Labeling



۱. جمع آوری تصاویر از فروشگاه‌های مختلف
  ۲. بازبینی تصاویر و تمیز کردن مجموعه داده
  ۳. برچسب‌گذاری مجموعه داده برای مدل تشخیص‌دهنده<sup>۴</sup> برچسب و یافتن آن از داخل تصویر
  ۴. برچسب‌گذاری مجموعه داده برای مدل تشخیص‌دهنده کلمات
  ۵. برچسب‌گذاری مجموعه داده برای مدل خواننده کلمات (OCR)
- برای این برچسب‌گذاری‌ها، از محیط‌های مختلفی استفاده شده که در فصل بعد آنها را معرفی می‌کنیم.

• **برچسب‌گذاری و آماده‌سازی داده‌ها:** تصاویر سه بار و برای سه مدل برچسب‌گذاری شده‌اند:

۱. تشخیص برچسب محصول: شناسایی موقعیت برچسب روی محصول.
  ۲. تشخیص نوشته‌های داخل برچسب: شناسایی و تعیین مناطق متنی داخل برچسب.
  ۳. خواندن نوشته‌ها: استخراج و تفسیر متن‌های موجود در برچسب.
- **آموزش مدل‌ها:** مدل‌های یادگیری ماشین و پردازش تصویر با استفاده از این داده‌ها آموزش دیده‌اند تا به دقت بالا در شناسایی و استخراج اطلاعات برسند. نتیجه این آموزش‌ها و دقت مدل‌ها در فصل‌های بعد به نمایش گذاشته خواهد شد.

### ۱-۳-۲- عملکرد برنامه در محیط واقعی

- **پویش و تحلیل برچسب‌ها:** برنامه طراحی شده است تا در محیط‌های فروشگاه‌ها و با استفاده از دوربین تلفن همراه، برچسب‌های مختلف را پویش کرده و اطلاعات را به صورت دقیق استخراج کند.
- **کاربری و رابط کاربری:** رابط کاربری برنامه فعلاً به گونه‌ای نیست که کاربران بتوانند بر روی تلفن همراه خود و یا از طریق سایت به آن دسترسی داشته باشند.


### ۱-۳-۳- قالب استاندارد برچسب ارزش مواد غذایی

این برنامه توانایی خواندن اکثر برچسب‌های موجود در فروشگاه‌ها را دارد اما قالب استاندارد که اداره کل نظارت و ارزیابی فراوده‌های خوراکی، آرایشی و بهداشتی سازمان غذا و دارو تایین کرده [۱]، به عنوان قالب و چارچوب استاندارد این پروژه می‌باشد که به شرح زیر است.

<sup>۴</sup> Detector

<sup>۵</sup> Models





**سازمان غذا و دارو**  
 شماره پروانه بهداشتی ساخت / ورود / شناسه نظارت  
 ( شماره پروانه در این محل درج شود )

معیار قابل فهم هر سهم+نام محصول + میزان سهم محصول برحسب گرم یا میلی لیتر

انرژی	کیلوکالری
قند	گرم
چربی	گرم
نمک	گرم
اسیدهای چرب ترانس	گرم

اندک ●
متوسط ●
زیاد ●

تصویر ۱- برچسب استاندارد مد نظر در این پروژه

- مواردی که در این برچسب به آن باید اشاره کرد به صورت زیر است.
- در بالاترین قسمت این برچسب، باید معیار قابل فهم هر سهم (برای مثال: یک قاشق غذاخوری از محصول یا هر ۲ عدد از محصول)، نام محصول، و میزان سهم محصول برحسب گرم یا میلی لیتر قرار گیرد. در این قسمت، فقط بخش آخر که میزان سهم محصول است از برچسب خوانده می شود.
  - در قسمت دوم برچسب، انرژی ای که در هر سهم از محصول موجود است و در قسمت اول ذکر شد، نوشته می شود. معیار اندازه گیری این قسمت، همیشه کیلوکالری (Kcal) است و اعداد، قبل از این معیار اندازه گیری قرار می گیرند.
  - بعد از این دو قسمت، چهار قسمت رنگی وجود دارد که هرکدام یک ماده مغزی را نشان می دهد. ترتیبی که این چهار ماده در برچسب دارند به ترتیب: قند، چربی، نمک و اسیدهای چرب ترانس است. معیار اندازه گیری این چهار ماده، گرم است و مانند انرژی، اعداد قبل از معیار قرار می گیرند.
  - شماره پروانه بهداشتی پروانه ساخت نیز یکی دیگر از مواردیست که باید پوشش شود و محل قرار گیری آن، پایین در سمت چپ است.
- سایر مواردی که در برچسب هستند، در این پروژه کاربردی ندارند و پوشش نخواهند شد.

### ۱-۳-۴- خارج از دامنه پروژه

- محصولات غیر بسته بندی:** پروژه محدود به محصولات غذایی بسته بندی شده با برچسب های واضح و قابل خواندن است و شامل مواد غذایی فاقد برچسب، برچسب غیر استاندارد سازمان غذا و دارو، برچسب هایی که به طور کامل و صحیح عکس برداری نشده و یا برچسب های بسیار آسیب دیده نمی شود. چند نمونه از این برچسب ها را می توان در زیر مشاهده کرد.



فرآورده کاکائویی تلخ		
در هر ۱۰۰ گرم	در هر سهم (۳ گرم)	
انرژی	458/28	13/74
قند	42/53	1/27
چربی کل	30	0/9
نمک	0/22	0/006
اسید چرب ترانس	0/15	0/004
کم	متوسط	زیاد

شماره پرونده بهداشتی: 56/13732

تصویر ۲- نمونه ای از برچسب غیر استاندارد

هر کاسه نودل سس گوجه فرنگی تند (۲۰۰ میلی لیتر)	
انرژی	۱۶۹/۸۴ کیلوکالری
چربی کل	۰ گرم
کربوهیدرات	۷/۳۴ گرم
پروتئین	۲/۳۶ گرم
نمک	۰ گرم
اسید چرب ترانس	۰ گرم
کم	متوسط

تصویر ۳- نمونه ای از برچسب با عکس برداری نامناسب





تصویر ۴- نمونه ای از برچسب با اعداد ناخوانا



تصویر ۵- نمونه ای از برچسب مخدوش

- پشتیبانی از ویژگی‌های خاص: برنامه به طور خاص برای استخراج و نمایش اطلاعات تغذیه‌ای و فنی طراحی شده است و ویژگی‌های خاصی مانند تجزیه و تحلیل تغذیه‌ای پیشرفته یا مشاوره‌های بهداشتی را شامل نمی‌شود. همچنین رنگ‌ها در برچسب ارزش مواد غذایی، ارزیابی نمی‌شوند.

#### ۴-۱- چالش‌ها

- سیستم‌های OCR معمولاً بر روی متن‌های معمولی آموزش داده می‌شوند که توسط جملات و پاراگراف‌ها تشکیل می‌شوند، اما برچسب‌های تغذیه‌ای، جدولی هستند و تاکید زیادی بر اعداد دارند. ثانیاً، این نمودارها از نظر رنگ‌ها، روشنایی، فونت و اندازه متن، الگوها، بافت‌ها و غیره بسیار متغیر هستند. که توسعه یک سیستم عمومی که قادر به تعمیم همه این شرایط باشد، بسیار چالش برانگیز است.
- تصاویر گرفته‌شده از برچسب‌ها ممکن است شامل نویزهای پس‌زمینه‌ای مانند بازتاب نور، سایه، یا حتی سایر محصولات مجاور باشند که می‌توانند باعث کاهش دقت در تشخیص متن شوند. فیلتر کردن این نویزها بدون از دست دادن اطلاعات مهم از جمله چالش‌های اساسی است.



- برخی برجسبها دارای اطلاعات بسیار فشرده هستند که در یک فضای کوچک قرار داده شده‌اند. تشخیص دقیق این متون فشرده که ممکن است از فونت‌های بسیار کوچک استفاده کنند، چالشی برای OCR به حساب می‌آید.
  - در صورتی که برنامه باید در زمان واقعی عمل کند، یعنی کاربر بتواند به محض گرفتن عکس از برجسب، اطلاعات را مشاهده کند، چالش‌هایی مانند بهینه‌سازی سرعت پردازش و کاهش تاخیر زمانی مطرح می‌شوند. این موضوع به خصوص در دستگاه‌های موبایل با منابع محدود مانند حافظه و پردازنده، چالش‌برانگیز است.
- این چالش‌ها نه تنها اهمیت پیچیدگی پروژه را نشان می‌دهند، بلکه نیاز به استفاده از رویکردهای نوآورانه و بهینه‌سازی‌های مداوم را برای دستیابی به نتایج مطلوب تأکید می‌کنند.



## فصل دوم

### ادبیات پژوهش

#### ۲-۱- مقدمه

در این فصل، به توضیح مفاهیم کلیدی مطرح شده در این پایان نامه پرداخته و به اختصار هر یک را بررسی می‌کنیم. در انجام این پژوهش، از ابزارها، روش‌ها و الگوریتم‌های مختلفی در حوزه های مختلف استفاده شده است که در این فصل به کاربرد هر یک از آنها در این پروژه اشاره خواهیم کرد.

مطالب این فصل به ۴ بخش تقسیم شده است: سه بخش اول ادبیات داده، ادبیات مدل و ادبیات ارزیابی هستند. در هر بخش، مباحث مرتبط با آن حوزه به طور جداگانه و در ارتباط با پروژه حاضر مورد بررسی قرار می‌گیرند. همچنین بخش آخر، مربوط به بحث کالری مصرفی و به دست آوردن اطلاعات فنی محصول می‌باشد.

#### ۲-۲- ادبیات داده

بخش مهمی از این پایان نامه به ایجاد و توسعه مجموعه داده‌ای برای برچسب‌های محصولات داخلی اختصاص یافته است. در این مسیر، از ابزارها و کتابخانه‌های متعددی استفاده شده که هر یک نقش کلیدی در فرآیند برچسب زدن، استخراج و پیش‌پردازش داده‌ها ایفا کرده‌اند. این ابزارها و کتابخانه‌ها نه تنها برای جمع‌آوری داده‌ها، بلکه برای پاک‌سازی، اصلاح، و آماده‌سازی آنها جهت استفاده در مدل‌های مختلف به کار گرفته شده‌اند.

در ادامه، برخی از این ابزارها و کتابخانه‌ها به تفصیل معرفی خواهند شد و نقش و کاربرد دقیق هر یک از آنها در این پایان نامه مورد بررسی قرار می‌گیرد. همچنین، به نحوه استفاده از این ابزارها در مراحل مختلف پروژه، از جمله استخراج داده‌های خام، پردازش اولیه، و آماده‌سازی نهایی برای تجزیه و تحلیل و مدل‌سازی، پرداخته می‌شود. هدف این بخش ارائه یک دیدگاه جامع از ابزارها و تکنیک‌هایی است که به پیشبرد تحقیق کمک کرده‌اند و نقش آنها در ایجاد یک مجموعه داده معتبر و قابل اعتماد برای زبان فارسی را تبیین می‌کند.



## ۲-۱-۲- فرایند برچسب گذاری تصویر<sup>۶</sup>

فرایند برچسب گذاری تصویر یک مرحله مهم در آماده سازی داده ها برای پروژه های یادگیری ماشین، به ویژه در حوزه بینایی ماشین و تشخیص اشیاء، است. این فرایند شامل شناسایی و برچسب گذاری نواحی خاص در تصاویر است که مدل های یادگیری ماشین بر اساس آن آموزش می بینند. در ادامه، مراحل اصلی این فرایند را توضیح می دهیم:

### ۱. جمع آوری تصاویر

اولین مرحله شامل جمع آوری مجموعه ای از تصاویر مرتبط با هدف پروژه است. این تصاویر می توانند از منابع مختلفی مانند دوربین ها، پایگاه های داده عمومی، اینترنت، و غیره جمع آوری شوند.

### ۲. انتخاب ابزار برچسب گذاری

ابزارهای فراوانی برای برچسب گذاری وجود دارد که چند نمونه در قسمت بعدی توضیح داده شده است.

### ۳. تعریف برچسب ها

قبل از شروع برچسب گذاری، باید دسته بندی ها یا کلاس هایی که قرار است برچسب گذاری شوند، تعریف شود. برای مثال، در این پروژه، برچسب ها شامل برچسب ارزش غذایی، اعداد داخل برچسب، مواد مغذی، واحدهای اندازه گیری و پروانه بهداشتی ساخت می باشند.

### ۴. برچسب گذاری نواحی

در این مرحله، کاربر با استفاده از ابزار برچسب گذاری، نواحی مربوط به هر دسته بندی را در تصاویر انتخاب می کند. این کار می تواند شامل ترسیم کادرهای محدودکننده<sup>۷</sup>، ماسک های تقسیم بندی<sup>۸</sup> یا نقاط کلیدی<sup>۹</sup> باشد که به شناسایی دقیق نواحی مورد نظر کمک می کند.

### ۵. بازبینی و تأیید

پس از برچسب گذاری اولیه، نواحی برچسب گذاری شده بازبینی می شوند تا اطمینان حاصل شود که برچسب ها به درستی و با دقت اعمال شده اند. در پروژه های بزرگتر، این مرحله ممکن است توسط افراد مختلف یا با استفاده از الگوریتم های خودکار انجام شود.

<sup>۶</sup> Image annotation

<sup>۷</sup> Bounding boxes

<sup>۸</sup> Segmentation masks

<sup>۹</sup> key points



## ۶. ذخیره‌سازی داده‌های برچسب‌گذاری شده

پس از تأیید، برچسب‌ها به همراه تصویر در قالب‌های استاندارد مانند ایکس‌ام‌ال<sup>۱۰</sup>، جی‌سان<sup>۱۱</sup>، یا فایل‌های متنی ذخیره می‌شوند. این داده‌ها سپس برای آموزش مدل‌های یادگیری ماشین استفاده می‌شوند.

## ۷. آماده‌سازی برای آموزش مدل

در نهایت، داده‌های برچسب‌گذاری شده برای آموزش مدل آماده می‌شوند. این مرحله شامل تقسیم‌بندی داده‌ها به مجموعه‌های آموزشی، اعتبارسنجی و آزمایش است که طبق ساختاری که برای آموزش هر مدل مشخص شده است، مرتب می‌شوند.

فرایند برچسب‌گذاری تصویر، یک گام اساسی برای اطمینان از دقت و عملکرد بالای مدل‌های یادگیری ماشین در پروژه‌های بینایی ماشین است. کیفیت داده‌های برچسب‌گذاری شده به طور مستقیم بر عملکرد نهایی مدل تأثیر می‌گذارد.

### ۲-۲-۲- پلتفرم<sup>۱۲</sup> روبوفلو<sup>۱۳</sup>

روبوفلو یک پلتفرم جامع برای توسعه پروژه‌های بینایی ماشین است که به ویژه در بخش برچسب‌گذاری تصاویر بسیار قدرتمند است. این پلتفرم ابزارهای متنوعی برای برچسب‌گذاری دستی، نیمه‌خودکار، و خودکار ارائه می‌دهد که کاربران می‌توانند با آن‌ها به سادگی نواحی موردنظر در تصاویر را مشخص کنند. روبوفلو از انواع مدل‌های بینایی ماشین مانند تشخیص اشیاء، تقسیم‌بندی، و طبقه‌بندی پشتیبانی می‌کند.

همچنین، این پلتفرم امکاناتی برای مدیریت نسخه‌ها، همکاری تیمی، و بهینه‌سازی داده‌ها فراهم می‌کند. به لطف یکپارچگی با فریم‌ورک‌های یادگیری ماشین محبوب، داده‌های برچسب‌گذاری شده می‌توانند مستقیماً برای آموزش مدل‌ها استفاده شوند، که باعث افزایش دقت و سرعت فرآیند توسعه مدل‌های بینایی ماشین می‌شود.[۲].

### ۲-۲-۳- پلتفرم لیبل استودیو<sup>۱۴</sup>

لیبل استودیو یک پلتفرم متن‌باز<sup>۱۵</sup> برای برچسب‌گذاری داده‌ها است که از انواع مختلف داده‌ها از جمله تصاویر، ویدئوها، متن‌ها و فایل‌های صوتی پشتیبانی می‌کند. این ابزار به کاربران امکان می‌دهد تا به راحتی پروژه‌های برچسب‌گذاری خود را مدیریت کرده و داده‌هایشان را برای مدل‌های یادگیری ماشین آماده کنند. با قابلیت‌های

<sup>10</sup> XML

<sup>11</sup> JSON

<sup>12</sup> Platform

<sup>13</sup> Roboflow

<sup>14</sup> Label studio

<sup>15</sup> Open source



سفارشی‌سازی بالا، کاربران می‌توانند رابط کاربری و ابزارهای برچسب‌گذاری را متناسب با نیازهای خاص پروژه‌های خود تنظیم کنند.

یکی از ویژگی‌های برجسته این پلتفرم، امکان همکاری همزمان چندین کاربر در برچسب‌گذاری داده‌ها است که باعث افزایش سرعت و دقت کار می‌شود. مدیریت پروژه‌های برچسب‌گذاری نیز با این پلتفرم بسیار ساده‌تر می‌شود، به طوری که کاربران می‌توانند فرآیندهای کاری، تخصیص وظایف و مراحل بازبینی را به طور دقیق کنترل کنند. این ابزار همچنین با فریم‌ورک‌های یادگیری ماشین یکپارچه می‌شود، که داده‌های برچسب‌گذاری شده را مستقیماً برای آموزش مدل‌ها قابل استفاده می‌کند.

علاوه بر این، از افزونه‌ها و ای‌پی‌آی‌های متنوعی پشتیبانی می‌کند که امکان یکپارچگی با سایر سیستم‌ها و گسترش قابلیت‌های پلتفرم را فراهم می‌سازد. این ویژگی‌ها به کاربران اجازه می‌دهد تا این پلتفرم را با ابزارهای موجود خود ترکیب کرده و از آن در محیط‌های مختلف استفاده کنند [۳].



تصویر ۶- محیط لیبل استودیو و فرایند برچسب‌گذاری

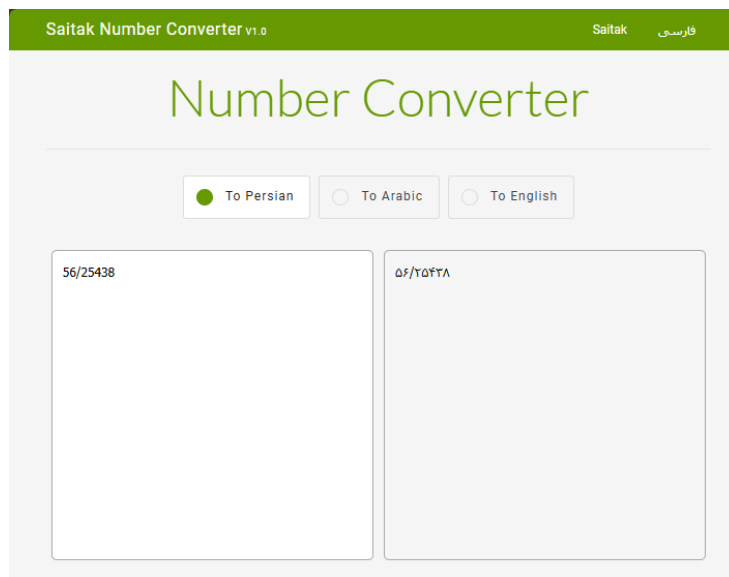
## ۲-۲-۴- وبسایت سایتاک<sup>۱۷</sup>

این وبسایت، بستری برای تبدیل کاراکتر اعداد انگلیسی به کاراکتر اعداد فارسی فراهم می‌کند. این تبدیل از این جهت مفید است که در پلتفرم‌های برچسب‌گذاری، اعداد فارسی با استفاده از صفحه کلید، در دسترس نیستند و برای آموزش مدل‌ها، باید اعداد فارسی بر روی برچسب را، با استفاده از کاراکترهای فارسی نوشت [۴].

<sup>16</sup> API

<sup>17</sup> Saitak





تصویر ۷- محیط وبسایت سائتاک

## ۲-۵- پلتفرم اوپن‌سی‌وی<sup>۱۸</sup>

این کتابخانه یک کتابخانه متن‌باز است که به طور گسترده برای برنامه‌های بینایی ماشین و پردازش تصویر استفاده می‌شود. این کتابخانه ابتدا توسط اینتل<sup>۱۹</sup> توسعه یافت و اکنون توسط یک جامعه کاربری بزرگ نگهداری می‌شود. اوپن‌سی‌وی شامل بیش از ۲۵۰۰ الگوریتم بهینه‌سازی شده است که برای تشخیص و ردیابی اشیاء، شناسایی چهره، تحلیل حرکات، و پردازش تصاویر مورد استفاده قرار می‌گیرند. یکی از ویژگی‌های برجسته این کتابخانه این است که از چندین زبان برنامه‌نویسی پشتیبانی می‌کند و همچنین با سیستم‌عامل‌های مختلفی مانند ویندوز، لینوکس<sup>۲۰</sup>، و مک‌اواس<sup>۲۱</sup> سازگار است.

اوپن‌سی‌وی در طیف وسیعی از کاربردها مانند بینایی ماشین، رباتیک، تحلیل ویدئو، و واقعیت افزوده استفاده می‌شود. از جمله قابلیت‌های این کتابخانه می‌توان به تشخیص لبه‌ها، فیلترگذاری تصاویر، شناسایی ویژگی‌ها، و پردازش ویدئو اشاره کرد. اوپن‌سی‌وی به دلیل سرعت و کارایی بالا در پردازش تصویر، در پروژه‌های تحقیقاتی و تجاری به طور گسترده‌ای مورد استفاده قرار می‌گیرد. این کتابخانه به توسعه‌دهندگان امکان می‌دهد تا با استفاده از ابزارهای قدرتمند و الگوریتم‌های از پیش ساخته شده، برنامه‌های پیچیده بینایی ماشین را به سرعت پیاده‌سازی کنند.

در این پروژه از اوپن‌سی‌وی عموماً برای خواندن عکس‌ها، نمایش اطلاعات خوانده شده بر روی هر عکس، بازبینی دقت و عملکرد مدل‌ها و ذخیره عکس‌ها استفاده شده است.

<sup>18</sup> OpenCV

<sup>19</sup> Intel

<sup>20</sup> Linux

<sup>21</sup> MacOS



## ۲-۲-۶- کتابخانه پیل<sup>۲۲</sup>

پیل یک کتابخانه قدرتمند و کاربردی برای پردازش تصاویر در پایتون است. این کتابخانه امکان باز کردن، تغییر، و ذخیره انواع مختلف تصاویر را فراهم می‌کند و از فرمت‌های تصویری متعددی مانند جی‌پی‌جی<sup>۲۳</sup>، پی‌ان‌جی<sup>۲۴</sup>، گیف<sup>۲۵</sup>، و بی‌ام‌پی<sup>۲۶</sup> پشتیبانی می‌کند. با استفاده از پیل، توسعه‌دهندگان می‌توانند عملیات مختلفی مانند تغییر اندازه، برش، فیلترگذاری، و تبدیل فرمت تصاویر را به سادگی انجام دهند. اگرچه توسعه رسمی پیل متوقف شده است، اما نسخه‌ای به‌روزتر و توسعه‌یافته به نام پیلو<sup>۲۷</sup> به عنوان جایگزین آن مورد استفاده قرار می‌گیرد و به طور گسترده‌ای در پروژه‌های پردازش تصویر استفاده می‌شود.

مشکل اصلی اوپن‌سی‌وی، نمایش کاراکترهای فارسی بر روی عکس است که باعث به هم ریختگی در حرف و ارقام می‌شود. این مشکل با استفاده از کتابخانه پیل برطرف شده و حروف و ارقام به ترتیب درست، بر روی عکس درج می‌شوند.

## ۲-۲-۷- کتابخانه نامپای<sup>۲۸</sup>

نامپای، یک کتابخانه اساسی و محبوب در زبان برنامه‌نویسی پایتون است که برای محاسبات علمی و عددی استفاده می‌شود. نامپای به‌ویژه به دلیل ارائه آرایه‌های چندبعدی<sup>۲۹</sup> با عملکرد بالا شناخته شده است. این آرایه‌ها به توسعه‌دهندگان اجازه می‌دهند تا عملیات‌های پیچیده ریاضی و محاسباتی را به صورت کارآمد و سریع انجام دهند. برخلاف لیست‌های پایتون، آرایه‌های نامپای دارای ابعاد ثابت و نوع داده یکنواخت هستند که این ویژگی‌ها باعث بهبود سرعت و کاهش مصرف حافظه می‌شود. علاوه بر این، نامپیا شامل تعداد زیادی از توابع ریاضی برای انجام محاسبات مختلف مانند جبر خطی، تبدیل فوریه، و تولید اعداد تصادفی است.

نامپای به عنوان یکی از کتابخانه‌های پایه‌ای در عملیات داده‌کاوی<sup>۳۰</sup> و یادگیری ماشین پایتون عمل می‌کند و بسیاری از کتابخانه‌های دیگر به طور مستقیم بر پایه نامپای ساخته شده‌اند. این کتابخانه به دلیل قابلیت‌های گسترده و عملکرد بالا، به‌ویژه در پردازش داده‌های بزرگ و انجام تحلیل‌های عددی در پروژه‌های علمی و مهندسی مورد استفاده قرار می‌گیرد. به عنوان مثال، توسعه‌دهندگان می‌توانند با استفاده از نامپای ماتریس‌ها و بردارها را به سادگی مدیریت کنند و محاسبات پیچیده‌ای مانند ضرب ماتریسی، برون‌یابی، و انتگرال‌گیری عددی را انجام دهند.

<sup>22</sup> PIL (Python Imaging Library)

<sup>23</sup> GPJ

<sup>24</sup> PNG

<sup>25</sup> GIF

<sup>26</sup> BMP

<sup>27</sup> Pillow

<sup>28</sup> NumPy

<sup>29</sup> Narray

<sup>30</sup> Data mining





ره کتابخانه‌ای در پایتون است که برای کار با عبارات منظم<sup>۳۲</sup> طراحی شده است. این کتابخانه امکاناتی را برای جستجو، جایگزینی، و تطبیق الگوهای متنی فراهم می‌کند و به توسعه‌دهندگان اجازه می‌دهد تا با استفاده از الگوهای پیشرفته، به سرعت بخش‌های خاصی از متن‌ها را شناسایی و پردازش کنند. با این کتابخانه، می‌توان عملیات‌های مختلفی مانند بررسی وجود الگو، تقسیم متن بر اساس الگو، و جایگزینی متن‌های مطابق با الگو را به سادگی انجام داد. این ابزار به ویژه در پردازش داده‌ها، تحلیل متن، و اعتبارسنجی ورودی‌های کاربر کاربرد زیادی دارد.

## ۲-۲-۹- کتابخانه‌های اصلاح قالب<sup>۳۳</sup>

عریک‌ریشیپر<sup>۳۴</sup> و بیدی<sup>۳۵</sup> دو کتابخانه پایتون هستند که برای پردازش و نمایش صحیح متن‌های عربی و دیگر زبان‌های راست به چپ<sup>۳۶</sup> مانند فارسی استفاده می‌شوند. عریک‌ریشیپر وظیفه تغییر شکل حروف عربی را به گونه‌ای دارد که آن‌ها به درستی به هم متصل شوند، زیرا حروف در این زبان‌ها بسته به موقعیتشان در کلمه، اشکال مختلفی دارند. از سوی دیگر، بیدی برای ترتیب‌گذاری صحیح متن‌های دوزبانه (شامل حروف از راست به چپ و چپ به راست) به کار می‌رود، به‌ویژه وقتی که اعداد یا کلمات لاتین در متن عربی یا فارسی قرار می‌گیرند. این دو کتابخانه معمولاً در کنار هم استفاده می‌شوند تا متن‌های عربی و فارسی به درستی در رابط‌های گرافیکی یا تصاویر نمایش داده شوند. این دو کتابخانه همراه با کتابخانه پیل، برای قرار دادن ارقام و حروف فارسی بر روی عکس، استفاده شده‌اند.

## ۲-۲-۱۰- کتابخانه‌ی جیسان

جیسان کتابخانه‌ای در پایتون است که برای کار با داده‌های فرمت جیسان طراحی شده است. این کتابخانه امکاناتی را برای تبدیل داده‌های پایتون به فرمت جیسان و بالعکس فراهم می‌کند، به طوری که می‌توان به راحتی داده‌ها را بین پایتون و دیگر زبان‌های برنامه‌نویسی یا سیستم‌های مبتنی بر جیسان مبادله کرد. در این پروژه از جیسان برای خواندن داده‌های برچسب‌گذاری شده که در پلتفرم‌های مختلف آماده شده، استفاده می‌شود.

<sup>31</sup> RE

<sup>32</sup> Regular Expressions

<sup>33</sup> Format resaper

<sup>34</sup> Arabic resaper

<sup>35</sup> Bidi (Bidirectional)

<sup>36</sup> RTL



## ۲-۳- ادبیات مدل

در این پایان نامه، برای اینکه بتوانیم حروف، کلمات و اعداد نوشته شده بر روی برچسب را بخوانیم، نیاز بود که مدلهایی را با استفاده از آنچه که گفته شد، آموزش دهیم و از آن استفاده کنیم. برای آموزش این مدل ها از ابزارها و کتابخانه هایی استفاده شد که در این بخش به معرفی آنها می پردازیم. همچنین هر آن چیزی که به سرعت بخشیدن و بهبود این فرایند کمک کرد را نیز معرفی خواهیم کرد.

### ۲-۳-۱- پلتفرم آلترالیتیکس<sup>۳۷</sup>

آلترالیتیکس یک کتابخانه پایتون است که به ویژه برای کار با مدل های یادگیری عمیق در حوزه بینایی ماشین توسعه یافته است. این کتابخانه به خاطر پیاده سازی ساده و کاربردی از مدل های پیشرفته بینایی ماشین مانند یولو<sup>۳۸</sup> مشهور است. این کتابخانه، به توسعه دهندگان این امکان را می دهد که مدل های تشخیص اشیاء و دیگر وظایف بینایی ماشین را با استفاده از پیکربندی های ساده و استاندارد به سرعت آموزش دهند و اجرا کنند.

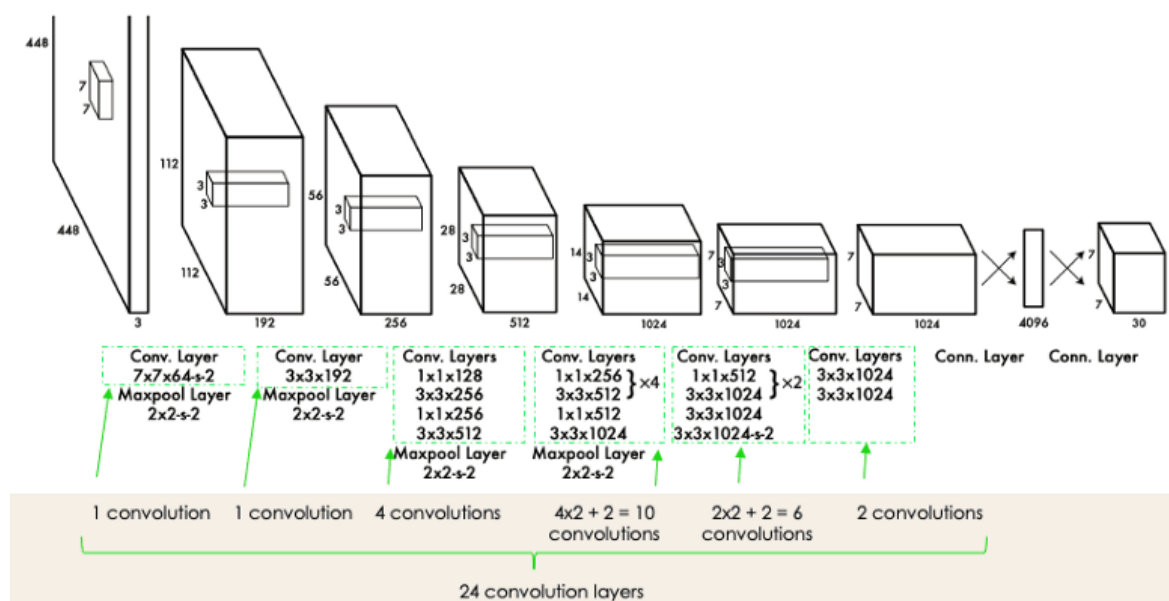
### ۲-۳-۲- الگوریتم یولو

یولو یک الگوریتم پیشرفته برای تشخیص اشیاء در تصاویر و ویدئوها است که به دلیل سرعت و دقت بالای خود شناخته شده است. برخلاف روش های سنتی که تصویر را به بخش های کوچک تر تقسیم کرده و هر بخش را به صورت جداگانه تحلیل می کنند، یولو کل تصویر را به یکباره پردازش کرده و به شناسایی و مکان یابی اشیاء در آن می پردازد. این الگوریتم با استفاده از شبکه های عصبی کانولوشنی به سرعت اشیاء مختلف را در تصویر شناسایی کرده و جعبه های محدود کننده و برچسب های مربوط به هر شیء را پیش بینی می کند. یولو به دلیل عملکرد بالا و توانایی پردازش در زمان واقعی، در کاربردهایی مانند نظارت ویدئویی، رباتیک، و تحلیل تصاویر پزشکی بسیار مورد استفاده قرار می گیرد. ساختار کلی این مدل، در شکل زیر به صورت واضح قابل مشاهده است.

<sup>37</sup> Ultralytics

<sup>38</sup> YOLO (You Only Look Once)





تصویر ۸- ساختار کلی مدل یولو

در این پروژه از یولو ورژن ۸ برای تشخیص برجسب ارزش مواد غذایی، استفاده شده است. این نسخه، به عنوان نسخه جدید و بهبود یافته الگوریتم یولو، به دلیل مزایای متعددی که ارائه می‌دهد، انتخاب مناسبی برای پروژه‌های بینایی ماشین است. این نسخه با بهینه‌سازی‌های الگوریتمی و معماری پیشرفته‌تر، دقت بالاتری در تشخیص اشیاء و سرعت پردازش سریع‌تری نسبت به نسخه‌های قبلی ارائه می‌دهد. به خاطر توانایی‌های بهبود یافته در شناسایی و مکان‌یابی دقیق اشیاء، قابلیت کار در زمان واقعی، و مقیاس‌پذیری برای پردازش داده‌های بزرگ و متنوع، به ویژه در پروژه‌هایی که نیاز به پردازش سریع و دقیق تصاویر دارند، بسیار مناسب است.

## ۲-۳-۳- پلتفرم پایتورچ<sup>۳۹</sup>

پایتورچ یک کتابخانه متن‌باز و محبوب برای یادگیری ماشین و یادگیری عمیق است که توسط فیس‌بوک<sup>۴۰</sup> توسعه یافته است. این کتابخانه به دلیل سهولت استفاده و انعطاف‌پذیری بالا، به ویژه در میان پژوهشگران و توسعه‌دهندگان یادگیری عمیق محبوبیت زیادی دارد. پایتورچ از تانسور<sup>۴۱</sup> برای انجام محاسبات عددی استفاده می‌کند و با گراف<sup>۴۲</sup> محاسباتی پویا، امکان طراحی و آزمایش مدل‌های پیچیده را فراهم می‌سازد. این ویژگی‌ها به توسعه‌دهندگان کمک می‌کند تا به سرعت مدل‌های خود را توسعه داده و آزمایش کنند.

پایتورچ به دلیل سازگاری با ابزارها و کتابخانه‌های دیگر مانند نامپای<sup>۴۳</sup> و سایکیت‌لرن<sup>۴۴</sup> و همچنین پشتیبانی گسترده از سوی جامعه کاربری، یک انتخاب محبوب برای پروژه‌های یادگیری عمیق در حوزه‌هایی مانند پردازش تصویر و

<sup>۳۹</sup> PyTorch

<sup>۴۰</sup> Facebook

<sup>۴۱</sup> Tensor

<sup>۴۲</sup> Graph

<sup>۴۳</sup> Numpy

<sup>۴۴</sup> Scikit-learn



پردازش زبان طبیعی است. این کتابخانه به کاربران در سطوح مختلف اجازه می‌دهد تا به راحتی مدل‌های یادگیری ماشین را پیاده‌سازی و آموزش دهند، و در نتیجه در پروژه‌های تحقیقاتی و تجاری بسیار مورد استفاده قرار می‌گیرد. در این پایان‌نامه نیز برای پیاده‌سازی مدل‌ها و استفاده از آن‌ها، از این کتابخانه استفاده شده است و مدل‌هایی که در این پروژه نام برده خواهند شد، در این فرمت ذخیره شده‌اند.

## ۲-۳-۴- پلتفرم محاسبات موازی کودا<sup>۴۵</sup>

کودا یک پلتفرم محاسباتی موازی و رابط برنامه‌نویسی کاربردی است که توسط انویدیا<sup>۴۶</sup> توسعه یافته است. این فناوری به توسعه‌دهندگان اجازه می‌دهد تا از قدرت پردازشی بالای پردازنده‌های گرافیکی<sup>۴۷</sup> برای انجام محاسبات موازی و تسریع پردازش‌ها استفاده کنند. با بهره‌گیری از هزاران هسته پردازشی موجود در کارت گرافیک، کودا می‌تواند زمان محاسبات پیچیده را به طور قابل توجهی کاهش دهد، و این امر به ویژه در زمینه‌هایی مانند یادگیری ماشین و شبیه‌سازی‌های علمی کاربرد دارد.

کودا با زبان‌های برنامه‌نویسی مانند سی<sup>۴۸</sup>، سی‌پلاس‌پلاس<sup>۴۹</sup> و پایتون یکپارچه شده است و بسیاری از کتابخانه‌های یادگیری عمیق مانند تنسورفلو<sup>۵۰</sup> و پایتورچ از آن برای تسریع آموزش مدل‌ها استفاده می‌کنند. این تکنولوژی به توسعه‌دهندگان امکان می‌دهد تا از کارت گرافیک برای انجام محاسبات پیچیده و زمان‌بر به صورت کارآمدتر استفاده کنند، که در نهایت به بهبود عملکرد در کاربردهای مختلفی مانند تحلیل داده‌های بزرگ، رندرینگ<sup>۵۱</sup> گرافیکی و یادگیری عمیق منجر می‌شود.

## ۲-۳-۵- پلتفرم کگل<sup>۵۲</sup>

کگل یک پلتفرم جامع برای تحلیل داده‌ها و مسابقات یادگیری ماشین است که به محققان، دانشمندان داده و توسعه‌دهندگان این امکان را می‌دهد که به صورت آنلاین به مجموعه‌های داده، ابزارهای تحلیلی، و جوامع فعال دسترسی پیدا کنند. کاربران کگل می‌توانند در مسابقات داده‌ای شرکت کنند، مدل‌های یادگیری ماشین را پیاده‌سازی کرده و با دیگران به اشتراک بگذارند. این پلتفرم به ویژه برای آموزش و آزمون مدل‌های مختلف مفید است و از طریق ارائه داده‌های معتبر و ابزارهای قدرتمند، به پیشبرد تحقیق و توسعه در زمینه‌های مختلف کمک می‌کند.

<sup>45</sup> CUDA (Compute Unified Device Architecture)

<sup>46</sup> Nvidia

<sup>47</sup> GPU

<sup>48</sup> C

<sup>49</sup> C++

<sup>50</sup> Tensorflow

<sup>51</sup> Rendering

<sup>52</sup> Kaggle



علاوه بر مسابقات، کگل یک محیط تعاملی برای جستجو و تحلیل داده‌ها ارائه می‌دهد که شامل نوت‌بوک<sup>۵۳</sup>ها و یک پایگاه داده بزرگ از مجموعه‌های داده عمومی است. کاربران می‌توانند با استفاده از زبان‌های برنامه‌نویسی مختلف مانند پایتون<sup>۵۴</sup> و آر<sup>۵۵</sup> به تحلیل داده‌ها بپردازند و کدها و مدل‌های خود را بر روی سخت‌افزارهای مختلف از جمله کارت گرافیک‌ها و پردازنده‌ها اجرا کنند. همچنین می‌توانند نتایج خود را با جامعه کگل به اشتراک بگذارند. این ویژگی‌ها به محققان و توسعه‌دهندگان این امکان را می‌دهد که به سرعت به داده‌های واقعی دسترسی پیدا کنند، مدل‌های خود را ارزیابی کنند و از بازخورد جامعه برای بهبود عملکرد مدل‌های خود بهره‌برداری کنند[۵].

## ۲-۳-۶- پلتفرم ایزی‌اوسی آر<sup>۵۶</sup>

ایزی‌اوسی آر یک کتابخانه منبع باز برای تشخیص و شناسایی متن از تصاویر است که با هدف ارائه یک راه‌حل ساده و سریع برای پردازش متن تصویری طراحی شده است. این کتابخانه به ویژه برای پروژه‌های بینایی ماشین که نیاز به استخراج متن از تصاویر دارند، مناسب است. ایزی‌اوسی آر از مدل‌های پیش‌آموزش‌دیده استفاده می‌کند و به کاربران این امکان را می‌دهد که بدون نیاز به آموزش مدل‌های جدید، به راحتی متن‌ها را از تصاویر مختلف استخراج کنند. با پشتیبانی از زبان‌های مختلف و رابط کاربری ساده، این ابزار به توسعه‌دهندگان و محققان کمک می‌کند تا به سرعت و با دقت بالا متن‌های تصویری را شناسایی کنند.

این پلتفرم شامل دو مدل اصلی است. مدل تشخیص<sup>۵۷</sup> و مدل شناسایی<sup>۵۸</sup>. مدل تشخیص مسئول شناسایی نواحی متنی در تصویر است و به تعیین مکان‌های دقیق متن‌ها کمک می‌کند. این مدل قادر است نواحی مختلف متن را به درستی تشخیص دهد، حتی در شرایط پیچیده مانند تغییرات در اندازه، فونت و زاویه متن. مدل شناسایی به دنبال آن، وظیفه شناسایی و استخراج محتوای متنی از نواحی مشخص‌شده را بر عهده دارد. این مدل متن‌های شناسایی شده را به صورت خوانا و قابل استفاده استخراج می‌کند، و به این ترتیب، مراحل پردازش تصویر و استخراج متن به طور کامل و دقیق انجام می‌شود. در ادامه به معرفی دو مدلی که در این پلتفرم استفاده شده می‌پردازیم[۶].

### ۱. گرفت<sup>۵۹</sup>

مدل تشخیص متن در ایزی‌اوسی آر گرفت است. گرفت یک مدل پیشرفته برای شناسایی مناطق متنی در تصاویر است که به ویژه برای تشخیص متن در تصاویری با پس‌زمینه پیچیده و نویزهای زیاد طراحی شده است. این مدل با استفاده از تکنیک‌های یادگیری عمیق، نواحی متنی را با دقت بالا شناسایی می‌کند و قادر است متن‌های کوچک و نازک را که توسط مدل‌های دیگر ممکن است نادیده گرفته شوند، شناسایی کند[۷].

<sup>53</sup> Notebook

<sup>54</sup> Python

<sup>55</sup> R

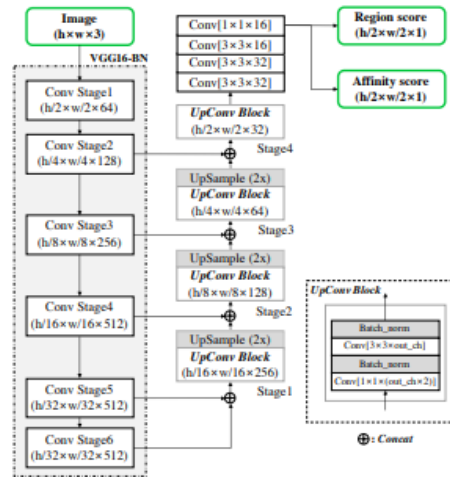
<sup>56</sup> EasyOCR

<sup>57</sup> Detection

<sup>58</sup> Recognition

<sup>59</sup> CRAFT (Character Region Awareness for Text detection)



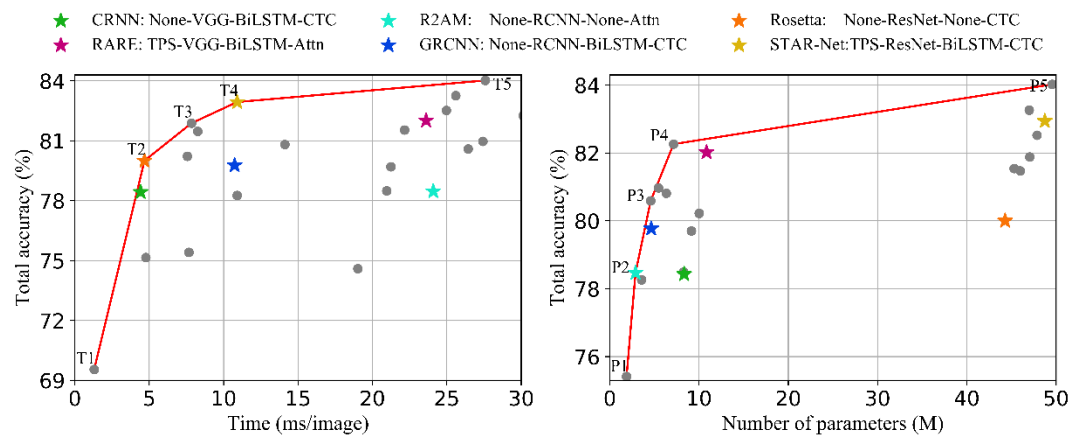


تصویر ۹- ساختار گرفت

## ۲. شناسایی عمیق متن<sup>۶۰</sup>

یک مجموعه از مدل‌ها و ابزارها برای شناسایی و خواندن متن از تصاویر است که به منظور مقایسه عملکرد الگوریتم‌های مختلف شناسایی متن طراحی شده است. این شناساگر (TPS-ResNet-BiLSTM-Attn) نام دارد و یکی از مدل‌های مناسب جهت خواندن نوشته‌ها می‌باشد. عملکرد این مدل با سایر مدل‌ها در تصویر زیر مقایسه شده است [۸].

Previously proposed combinations



تصویر ۱۰- مقایسه سرعت و تعداد پارامترهای مدل خوانش متن

## ۲-۴- ادبیات ارزیابی

پس از اتمام آموزش مدل با داده‌های آموزشی، مرحله ارزیابی مدل با استفاده از معیارهای مختلف آغاز می‌شود. در این بخش، ابتدا به توضیح هر یک از این معیارها پرداخته می‌شود و سپس ابزارهایی که برای تهیه گزارش از عملکرد مدل به کار گرفته شده‌اند، معرفی خواهند شد.

<sup>60</sup> Deep text recognition

## ۲-۴-۱- ماتریس درهم‌ریختگی<sup>۶۱</sup>

این ماتریس، یکی از ابزارهای اصلی برای ارزیابی عملکرد مدل‌های طبقه‌بندی است. ماتریس درهم‌ریختگی چهار ترکیب مختلف از پیش‌بینی‌های مدل و مقادیر واقعی را در نظر می‌گیرد که در زیر توضیح داده شده و در ادامه مورد استفاده قرار می‌گیرد.

جدول ۱- ماتریس درهم‌ریختگی

		Actual	
		Positive	Negative
Predicted	Positive	<b>True Positive</b>	<b>False Positive</b>
	Negative	<b>False Negative</b>	<b>True Negative</b>

۱. **True Positive (TP)**: تعداد نمونه‌هایی که واقعاً مثبت هستند و مدل به درستی آن‌ها را مثبت پیش‌بینی کرده است. این سلول در تصویر با رنگ سبز مشخص شده است.

۲. **False Positive (FP)**: تعداد نمونه‌هایی که واقعاً منفی هستند اما مدل به اشتباه آن‌ها را مثبت پیش‌بینی کرده است. این وضعیت به عنوان **خطای نوع اول** نیز شناخته می‌شود و در تصویر با رنگ قرمز نمایش داده شده است.

۳. **False Negative (FN)**: تعداد نمونه‌هایی که واقعاً مثبت هستند اما مدل به اشتباه آن‌ها را منفی پیش‌بینی کرده است. این وضعیت به عنوان **خطای نوع دوم** نیز شناخته می‌شود و در تصویر با رنگ قرمز نمایش داده شده است.

۴. **True Negative (TN)**: تعداد نمونه‌هایی که واقعاً منفی هستند و مدل به درستی آن‌ها را منفی پیش‌بینی کرده است. این سلول نیز با رنگ سبز مشخص شده است.

## ۲-۴-۲- معیار دقت<sup>۶۲</sup>

دقت کلی نشان می‌دهد که مدل به چه میزان در کل پیش‌بینی‌هایش درست عمل کرده است. این معیار نسبت تعداد پیش‌بینی‌های صحیح (چه مثبت و چه منفی) را به کل پیش‌بینی‌ها محاسبه می‌کند. دقت کلی معیار مناسبی زمانی است که داده‌ها دارای توازن خوبی بین کلاس‌ها هستند. این معیار در فرمول زیر نشان داده شده.

<sup>61</sup> Confusion matrix

<sup>62</sup> Accuracy



$$Accuracy = \frac{TP + FN}{All\ data}$$

#### ۲-۴-۳- معیار صحت<sup>۶۳</sup>

صحت نشان می‌دهد که از میان تمام نمونه‌هایی که به عنوان مثبت پیش‌بینی شده‌اند، چه درصدی واقعاً مثبت بوده‌اند. این معیار زمانی مهم است که هزینه‌ی بالایی برای FP وجود دارد. این معیار در فرمول زیر نشان داده شده.

$$Precision = \frac{TP}{TP + FP}$$

#### ۲-۴-۴- معیار بازخوانی یا حساسیت<sup>۶۴</sup>

بازخوانی نشان می‌دهد که مدل از میان تمامی نمونه‌های مثبت واقعی، چه تعداد را به درستی شناسایی کرده است. این معیار زمانی اهمیت دارد که هزینه‌ی بالایی برای از دست دادن نمونه‌های مثبت وجود دارد. فرمول ریاضیاتی این معیار در زیر آمده است.

$$Recall = \frac{TP}{TP + FN}$$

#### ۲-۴-۵- معیار F1

F1 به عنوان میانگین هارمونیک<sup>۶۵</sup> صحت و بازخوانی تعریف می‌شود. این معیار تلاش می‌کند تا یک توازن بین صحت و دقت برقرار کند. اگر دقت و بازخوانی هر دو بالا باشند، F1 نیز بالا خواهد بود، و اگر یکی از این دو پایین باشد، F1 نیز پایین خواهد بود.

به عبارت دیگر، F1 برای مواقعی مناسب است که می‌خواهیم یک معیار واحد داشته باشیم که هم به درستی پیش‌بینی‌های مثبت و هم به میزان پوشش صحیح نمونه‌های مثبت اهمیت دهد.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

<sup>63</sup> Precision

<sup>64</sup> Recall

<sup>65</sup> Harmonic mean





## ۲-۴-۶- میانگین خطای نرمالیزه شده<sup>۶۶</sup>

میانگین خطا، معیاری است که برای تعیین کمیت تفاوت دو دنباله (مانند رشته ها) از یکدیگر با شمارش حداقل تعداد عملیات مورد نیاز برای تبدیل یک رشته به رشته دیگر استفاده می شود. این عملیات معمولاً شامل درج، حذف و جایگزینی است. با نرمال کردن این فرایند، این معیار مقداری بین ۰ و ۱ پیدا می کند. این نرمال سازی کمک می کند که مقایسه بین معیارها راحت تر باشد.

## ۲-۴-۷- کتابخانه مت پلات لیب<sup>۶۷</sup>

مت پلات لیب یک کتابخانه قدرتمند در زبان برنامه نویسی پایتون است که برای مصورسازی داده ها و ایجاد نمودارهای متنوع به کار می رود. این کتابخانه به برنامه نویسان این امکان را می دهد که به راحتی داده ها را تحلیل کنند، الگوها و تغییرات را مشاهده نمایند و اطلاعات را به صورت بصری و قابل فهم نمایش دهند. با استفاده از مت پلات لیب، می توان انواع نمودارهای خطی<sup>۶۸</sup>، دایره ای<sup>۶۹</sup>، میله ای<sup>۷۰</sup>، نقطه ای<sup>۷۱</sup> و سایر نمودارها را ایجاد کرد و آنها را با استفاده از تنظیمات ظاهری و قابلیت های تعاملی مختلف بهینه سازی نمود. این کتابخانه به دلیل کاربرد گسترده ای که در علوم داده، مهندسی، آمار و بسیاری از حوزه های دیگر دارد، به یک ابزار ضروری برای تجزیه و تحلیل داده ها و ارائه نتایج به صورت گرافیکی تبدیل شده است.

در این پایان نامه نیز از مت پلات لیب در ارزیابی مدل ها استفاده شده است. برای بسیاری از تصمیم گیری ها نیاز به رسم نمودارهایی با استفاده از این کتابخانه بود و تمامی نمودارهایی که در فصل های بعدی به نمایش درآمده اند، با استفاده از مت پلات لیب ایجاد شده اند.

## ۲-۵- تجزیه و تحلیل اطلاعات

پس از جمع آوری داده های درج شده بر روی برچسب ارزش غذایی محصولات، نوبت به تجزیه و تحلیل محصول می رسد. در این بخش، ابتدا مفهوم اصلی در این بخش را توضیح داده و سپس ابزارهایی که در این پروسه به ما کمک کردند، معرفی خواهند شد.

<sup>۶۶</sup> Normalized Edit Distance

<sup>۶۷</sup> Matplotlib

<sup>۶۸</sup> Line chart

<sup>۶۹</sup> Pie chart

<sup>۷۰</sup> Bar chart

<sup>۷۱</sup> Scatter plot



## ۲-۵-۱- محاسبه کالری مصرفی

مهم ترین مفهوم در این بخش، معادل متابولیک<sup>۷۲</sup> می باشد. معادل متابولیک، یک معیار استاندارد است که برای اندازه گیری شدت فعالیت های فیزیکی به کار می رود. یک MET به میزان مصرف انرژی در حالت استراحت بدن اشاره دارد. این معیار به شما کمک می کند تا بفهمید که یک فعالیت چقدر انرژی نسبت به حالت استراحت مصرف می کند [۹، ۱۰].

### مفهوم MET

- **یک MET:** برابر با مصرف انرژی در حالت استراحت است که به طور تقریبی معادل ۱ کیلوکالری (کالری) در هر کیلوگرم از وزن بدن در هر ساعت است.
  - **دو METs:** به این معنا است که شما در حال انجام فعالیتی هستید که دو برابر انرژی حالت استراحت مصرف می کند.
- به عنوان مثال، راه رفتن با سرعت متوسط ممکن است ۴ MET داشته باشد، به این معنا که چهار برابر انرژی حالت استراحت مصرف می کند.
- مقدار کالری مصرف شده در یک دقیقه فعالیت، با توجه به معده زیر به دست می آید.

$$Total\ burned\ calories = \frac{3.5 \times Body\ weight \times MET}{200}$$

## ۲-۵-۲- کتابخانه جینجا<sup>۷۳</sup>

جینجا یک موتور قالب سازی قدرتمند و محبوب برای پایتون است که به توسعه دهندگان امکان ایجاد قالب های پویا برای تولید اچ تی ام ال<sup>۷۴</sup>، ایکس ام ال<sup>۷۵</sup> و دیگر فرمت های متنی را می دهد. با استفاده از جینجا، می توان داده ها را به سادگی درون قالب ها جاگذاری کرد و صفحات وب پویا ایجاد نمود. این کتابخانه با داشتن سینتکسی<sup>۷۶</sup> شبیه به پایتون، فیلترها و توابع از پیش تعریف شده، و پشتیبانی از دستورات کنترلی مانند حلقه ها و شرط ها، انعطاف پذیری بالایی در طراحی قالب ها فراهم می کند.

جینجا اغلب در فریم ورک های وب مانند فلسک<sup>۷۷</sup> برای رندر کردن صفحات وب استفاده می شود. ویژگی های امنیتی آن، مانند خودکارسازی فرار برای جلوگیری از حملات XSS، آن را به گزینه ای امن برای توسعه صفحات وب تبدیل

<sup>72</sup> Metabolic Equivalent

<sup>73</sup> Jinja2

<sup>74</sup> HTML

<sup>75</sup> XML

<sup>76</sup> Syntax

<sup>77</sup> Flask



می‌کند. به‌طور کلی، جینجا با سادگی و قدرتی که ارائه می‌دهد، یک ابزار اساسی برای توسعه‌دهندگان پایتون در پروژه‌های وب محسوب می‌شود.

## ۲-۵-۳- کتابخانه سلیوم<sup>۷۸</sup>

سلیوم یک ابزار قدرتمند برای خودکارسازی مرورگرهای وب است که به‌طور گسترده توسط توسعه‌دهندگان و تست‌کنندگان نرم‌افزار استفاده می‌شود. این کتابخانه به شما امکان می‌دهد تعاملات مختلف با مرورگرها را شبیه‌سازی کنید، مانند کلیک کردن بر روی دکمه‌ها، پر کردن فرم‌ها، جابه‌جا شدن بین صفحات، و استخراج داده‌ها از وب‌سایت‌ها. سلیوم از مرورگرهای مختلفی مانند کروم، فایرفاکس، سافاری و اج پشتیبانی می‌کند و می‌تواند با استفاده از زبان‌های برنامه‌نویسی مختلف از جمله پایتون، جاوا، سی‌شارپ و روبی کنترل شود.

سلیوم به‌ویژه برای تست خودکار وب‌سایت‌ها بسیار مفید است. با استفاده از این ابزار، تست‌کنندگان می‌توانند سناریوهای پیچیده‌ای را که کاربران واقعی ممکن است تجربه کنند، شبیه‌سازی کرده و صحت عملکرد سایت را بررسی کنند. به‌طور کلی، سلیوم ابزاری اساسی برای هر کسی است که نیاز به خودکارسازی تعاملات با مرورگرهای وب دارد.

## ۲-۶- جمع بندی:

همان‌طور که پیش‌تر بیان شد، این پایان‌نامه به چهار بخش مجزا با عناوین داده، مدل، ارزیابی و تجزیه و تحلیل اطلاعات تقسیم شده است. در این فصل، برخی از مفاهیم مرتبط با این بخش‌ها معرفی و بررسی شدند. در فصل‌های بعد، با استفاده از این مبانی به تشریح فرآیند انجام کار در این پایان‌نامه خواهیم پرداخت.

---

<sup>78</sup> Selenium



## فصل سوم

### پیشینه پژوهش

#### ۳-۱- مقدمه

این فصل به بررسی تحقیقات و پروژه‌های پیشین در زمینه پردازش تصویر و تشخیص نوری حروف<sup>79</sup> می‌پردازد. در این زمینه، تحقیقات زیادی صورت نگرفته است، از دلایل آن می‌توان به تفاوت بین چاپ برجسب ارزش مواد غذایی بر روی محصولات مختلف و استفاده از رنگ‌ها با کدهای رنگی مختلف نام برد. چند نمونه تحقیق در این زمینه بر روی برجسب محصولات آمریکایی<sup>۸۰</sup> انجام شده و نمونه این پویشگر بر روی محصولات ایرانی، وجود ندارد. این فصل، چارچوبی را فراهم می‌کند تا نشان دهد چگونه پروژه فعلی می‌تواند به پیشرفت دانش موجود در این زمینه و گسترش آن به زبان فارسی کمک کند و راهکارهایی نوآورانه برای حل مشکلات و بهبود تجربه کاربری ارائه دهد.

#### ۳-۲- تاریخچه و توسعه OCR

تشخیص نوری حروف یکی از تکنولوژی‌های کلیدی در پردازش تصویر است که برای استخراج متون از تصاویر استفاده می‌شود. تکنولوژی OCR از دهه ۱۹۵۰ توسعه یافته و به مرور زمان با بهبودهای مداوم در دقت و کارایی روبرو شده است. از جمله پیشرفت‌های اخیر می‌توان به استفاده از الگوریتم‌های یادگیری عمیق و شبکه‌های عصبی مصنوعی اشاره کرد که دقت تشخیص حروف را به طور قابل توجهی افزایش داده‌اند. به عنوان مثال، الگوریتم‌هایی نظیر Tesseract که به صورت متن باز توسعه یافته‌اند، قادر به تشخیص متون در چندین زبان و با دقت بالایی هستند.

مدل‌های یادگیری عمیق، مانند شبکه‌های عصبی کانولوشنی<sup>۸۱</sup> و شبکه‌های بازگشتی<sup>۸۲</sup>، به دلیل دقت بالا در تشخیص و طبقه‌بندی تصاویر، به طور گسترده‌ای در OCR مورد استفاده قرار گرفته‌اند. مدل‌های مدرنی نظیر MaskOCR و TransOCR از جدیدترین الگوریتم‌ها در این حوزه هستند که دقت‌های بسیار بالایی در تشخیص متن از تصاویر دارند.

<sup>79</sup> Optical character recognition

<sup>80</sup> USDA nutrition label

<sup>81</sup> Convolutional neural network

<sup>82</sup> Recurrent neural network



این مدل‌ها با استفاده از مجموعه داده‌های بزرگ و متنوع، قادر به تشخیص دقیق حروف در شرایط مختلف، نظیر نویز بالا و کیفیت پایین تصاویر، هستند.

### ۳-۳- مروری بر مطالعات مرتبط

مطالعات و کارهای نسبتاً کمی در زمینه پردازش تصویر و بینایی ماشین بر روی برچسب ارزش محصولات غذایی انجام گرفته است که همه این مطالعات بر روی برچسب‌هایی غیر از برچسبی فارسی پویش می‌شود بوده. بعضی از این محصولات به صورت نرم افزار اندرویدی<sup>۸۳</sup> بوده [۱۱]، و برخی دیگر مقالات علمی [۱۲-۱۴].

این برنامه‌ها راه‌های مختلفی برای به دست آوردن ورودی کاربر از اطلاعات تغذیه‌ای دارند. یکی از راه حل‌های این برنامه‌ها، استفاده از اطلاعاتی است که از قبل توسط مجموعه‌ای از کاربران جمع آوری شده است. مصرف‌کننده می‌تواند محصول خود را با نام جستجو کند تا آمار مربوطه را بیابد، اما داده‌های جمع‌آوری شده ممکن است کمبود داشته باشند.

اطلاعات، ممکن است برای هر محصول تکراری باشد و این اطلاعات بایکدیگر تطابق نداشته باشند که باعث گیج شدن کاربر می‌شود. همچنین اطلاعات ممکن است کاملاً نادرست باشد زیرا کاربران به صورت دستی آنها را وارد کرده و هیچ تعهدی نسبت به وارد کردن آنها ندارند. راه‌حل‌های دیگری که پیاده‌سازی شده، حول ساخت یک پایگاه داده از محصولات شناخته شده می‌باشد.

داده‌های تغذیه‌ای پایگاه داده محور، مستلزم جمع آوری تمام داده‌ها در یک مکان واحد است. داده‌های در حال تغییر باید دائماً به روز شوند. این رویکرد به حجم زیادی از زیرساخت، همکاری و نگهداری نیاز دارد. اگر داده‌ها حفظ نشوند یا محصولات جدیدی به آن اضافه نشود، داده‌هایی که کاربر به دنبال آن است ممکن است وجود نداشته باشد یا قدیمی باشد.

آن دسته از مقالاتی که تمرکز بر توسعه برنامه‌ای بر پایه استخراج اطلاعات از برچسب ارزش مواد غذایی داشتند، روندی مانند روند نوشته شده در زیر را دنبال کردند.

۱. الگوریتم<sup>۸۴</sup>، یک عکس را به عنوان ورودی می‌گیرد.
۲. برچسب ارزش مواد غذایی در عکس یافت شده و بریده می‌شود.
۳. عکس به بخش‌های مختلفی تقسیم می‌شود که هر بخش نشانگر یک کلمه می‌باشد.
۴. بخش‌های مختلف که عکس به آنها شکسته شد، خوانده شده تا عدد یا کلمه از آن استخراج شود.
۵. خروجی به دست می‌آید و به صورت لیستی از اطلاعات نمایش داده می‌شود.

داده‌های استفاده شده در این مقالات، بسیار محدود تر از داده‌های موجود در الگوریتم معرفی شده در این مقاله می‌باشد. یکی از مقالات [۱۲] به جای استفاده از OCR برای خواندن کلمات، از کلاس‌بند<sup>۸۵</sup> استفاده کرده به این دلیل که کلمات محدود بوده و نیاز به خواندن ندارند. ولی این کار باعث می‌شود که اگر یک کلاس‌بندی به اشتباه انجام شد، آن ماده

<sup>83</sup> Android

<sup>84</sup> Algorithm

<sup>85</sup> Classifier



مغذی به اشتباه مقدار بگیرد اما با استفاده از OCR، اگر مشکلی در خواندن پیش بیاید، یک یا دو حرف عوض شده و می‌توان از سایر حروف، کلمه را حدس زد. البته اگر تعداد عکس‌های مجموعه داده زیاد باشد، مشکل OCR تا حدودی رفع خواهد شد.

این تحقیقات عمدتاً بروی برچسب محصولات آمریکایی انجام شده است که به شکل زیر است.

<b>Nutrition Facts</b>	
8 servings per container	
<b>Serving size</b>	<b>2/3 cup (55g)</b>
<b>Amount per serving</b>	
<b>Calories</b>	<b>230</b>
% Daily Value*	
<b>Total Fat</b> 8g	<b>10%</b>
Saturated Fat 1g	<b>5%</b>
Trans Fat 0g	
<b>Cholesterol</b> 0mg	<b>0%</b>
<b>Sodium</b> 160mg	<b>7%</b>
<b>Total Carbohydrate</b> 37g	<b>13%</b>
Dietary Fiber 4g	<b>14%</b>
Total Sugars 12g	
Includes 10g Added Sugars	<b>20%</b>
<b>Protein</b> 3g	
Vitamin D 2mcg	10%
Calcium 260mg	20%
Iron 8mg	45%
Potassium 235mg	6%
* The % Daily Value (DV) tells you how much a nutrient in a serving of food contributes to a daily diet. 2,000 calories a day is used for general nutrition advice.	

تصویر ۱۱- برچسب ارزش مواد غذایی آمریکایی

این برچسب شامل اطلاعات بیشتری نسبت به اطلاعات بر روی محصولات غذایی ایرانی دارد. این اطلاعات شامل، کالری هر سهم از محصول، تعداد وعده‌ها به ازای هر محصول، انواع چربی‌ها، کلسترول<sup>۸۶</sup>، سدیم<sup>۸۷</sup>، کربوهیدرات کل<sup>۸۸</sup>، پروتئین<sup>۸۹</sup>، ویتامین د<sup>۹۰</sup>، کلسیم<sup>۹۱</sup>، آهن<sup>۹۲</sup> و پتاسیم<sup>۹۳</sup> می‌باشد. همچنین مقادیر این مواد مغذی هم به میزان وزنی و هم به درصد نوشته شده است.

- <sup>86</sup> Cholesterol
- <sup>87</sup> Sodium
- <sup>88</sup> Total carbohydrate
- <sup>89</sup> Protein
- <sup>90</sup> Vitamin D
- <sup>91</sup> Calcium
- <sup>92</sup> Iron
- <sup>93</sup> Potassium



یکی از تفاوت های اصلی برچسب های آمریکایی و ایرانی در رنگی بودن پس زمینه مواد مغذی می باشد. این رنگی بودن، کار خواندن را کمی دشوار تر از برچسب های آمریکایی کرده به این دلیل که کد رنگی<sup>۹۴</sup> هر رنگ ممکن است با کد رنگی متن نوشته شده به سختی قابل تمایز بخشیدن باشد و مدل<sup>۹۵</sup> باید داده های زیادی را ببیند تا بتواند بین رنگ ها و کدهای رنگی تمایز قائل شود.

در برخی مطالعات انجام شده [۱۴]، تمرکز بر روی خواندن برچسب با چولگی تا ۳۰-۴۵ درجه می باشد. همچنین قابلیت خواندن بارکد نیز با همین چولگی نیز میسر است. اما در این پروژه، تمرکز فقط بر روی خود برچسب بدون چولگی بیش از حد می باشد.

### ۳-۴- جمع بندی

این فصل به بررسی تحقیقات پیشین در زمینه پردازش تصویر و تشخیص نوری حروف پرداخته و چالش های مرتبط با برچسب های ارزش غذایی محصولات را مورد بررسی قرار می دهد. با توجه به کمبود تحقیقات در این زمینه، به ویژه برای محصولات با برچسب های فارسی، تلاش های انجام شده در این پروژه می تواند به پیشرفت دانش موجود و گسترش کاربردهای اوسی آر در زبان فارسی کمک کند. همچنین، مرور تحقیقات مرتبط نشان می دهد که سیستم های فعلی دارای محدودیت هایی هستند، از جمله عدم تطابق داده ها و نیاز به نگهداری پایگاه داده ها. این پروژه با استفاده از الگوریتم های جدید و مجموعه داده های متنوع، به دنبال ارائه راهکارهایی نوآورانه برای بهبود دقت و کارایی سیستم های اوسی آر در تشخیص برچسب های ارزش غذایی است.

---

<sup>94</sup> RGB

<sup>95</sup> Model



## فصل چهارم

### ساختار و آموزش مدل‌ها

#### ۴-۱- مقدمه

در این فصل، به بررسی ساختار و معماری مدل‌های مورد استفاده در این پروژه می‌پردازیم و نحوه انتخاب و تنظیم پارامترهای این مدل‌ها را شرح می‌دهیم. سپس، به چگونگی آماده‌سازی داده‌ها برای آموزش مدل‌ها اشاره خواهیم کرد. در ادامه، فرآیند آموزش مدل‌ها، شامل تنظیمات اولیه، مراحل آموزش، و ارزیابی عملکرد مدل‌ها با استفاده از داده‌های آزمایشی توضیح داده خواهد شد. همچنین، چالش‌های احتمالی در طول آموزش و راهکارهای به‌کاررفته برای بهینه‌سازی عملکرد مدل‌ها مورد بحث قرار می‌گیرد.

#### ۴-۲- مدل تشخیص برچسب

همانطور که گفته شد، برای تشخیص برچسب ارزش مواد غذایی، از مدلی به نام یولو ورژن ۸ استفاده شد. در این قسمت به نحوه آموزش و همچنین آماده‌سازی داده برای این مدل می‌پردازیم. برای مدل اولیه، از YOLOv8s استفاده شد. چندین نوع مدل وجود دارد که برای آموزش باید از یکی از این مدل‌ها استفاده کرد. این مدل‌ها به شرح زیر می‌باشند.

جدول ۲- انواع مدل‌های یولو





Model	size (pixels)	mAP <sup>val</sup> 50-95	Speed CPU ONNX (ms)	Speed A100 TensorRT (ms)	params (M)	FLOPs (B)
<a href="#">YOLOv8n</a>	640	37.3	80.4	0.99	3.2	8.7
<a href="#">YOLOv8s</a>	640	44.9	128.4	1.20	11.2	28.6
<a href="#">YOLOv8m</a>	640	50.2	234.7	1.83	25.9	78.9
<a href="#">YOLOv8l</a>	640	52.9	375.2	2.39	43.7	165.2
<a href="#">YOLOv8x</a>	640	53.9	479.1	3.53	68.2	257.8

این جدولی است که در گیت‌هاب<sup>۹۶</sup> یولو قرار دارد [۱۵]. به دلیل راحتی کار، و وجود فقط یک کلاس در هر عکس (برچسب مواد غذایی) کار مدل راحت است، پس نیاز به مدل‌های سنگین نیست و در این پروژه مدل s انتخاب شده که ترکیبی از دقت مناسب و سرعت قابل قبول را دارد.

حال برای اینکه سرعت و دقت بیشتری در فرایند آموزش به مدل بدهیم، یک مدل از قبل آموزش دیده بر روی مجموعه داده کوکو<sup>۹۷</sup> استفاده می‌کنیم. مجموعه داده کوکو یک مجموعه داده با ۸۰ کلاس است. هیچ کدام از این کلاس‌ها، مربوط به برچسب نیست، اما وزن‌های مدل، در این حالت نزدیک تر به وزن‌های بعد از آموزش هستند، از این رو آموزشی سریع و با دقت خواهیم داشت.

آرگومان‌هایی<sup>۹۸</sup> که در آموزش مدل استفاده شد به این صورت هستند.

- ایپاک<sup>۹۹</sup>: تعداد کل دوره‌های آموزشی را نشان می‌دهد که هر دوره نشان دهنده یک عبور کامل از کل مجموعه داده است. تنظیم این مقدار می‌تواند بر مدت زمان تمرین و عملکرد مدل تأثیر بگذارد. در اینجا تعداد کل دوره‌ها، ۲۰ تنظیم شده است.
- اندازه دسته<sup>۱۰۰</sup>: با سه حالت تنظیم به عنوان یک عدد صحیح (به عنوان مثال، دسته=۱۶)، حالت خودکار برای ۶۰٪ استفاده از حافظه کارت گرافیک (دسته=۱-)، یا حالت خودکار با کسر استفاده مشخص (دسته=۰/۷). در این پروژه، اندازه دسته ۳۲ در نظر گرفته شده است.
- نرخ یادگیری<sup>۱۰۱</sup>: تنظیم این مقدار برای فرآیند بهینه‌سازی بسیار مهم است و بر سرعت به‌روزرسانی وزن‌های مدل تأثیر می‌گذارد. در اینجا نرخ یادگیری ۰/۰۱ تنظیم شده است.

<sup>96</sup> GitHub

<sup>97</sup> COCO dataset

<sup>98</sup> Arguments

<sup>99</sup> Epoc

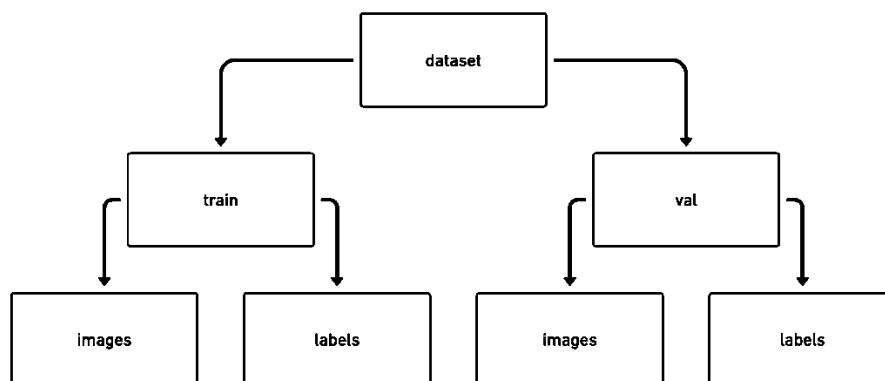
<sup>100</sup> Batch size

<sup>101</sup> Learning rate



آرگومان‌هایی که یولو می‌گیرد، بیش از ۳۰ تا است، این سه آرگومان دستی تنظیم شدند و مابقی به صورت پیش‌فرض باقی ماند.

هر مدل هوش مصنوعی، قالب دریافت مجموعه داده خود را دارد. یولو نیز از این قضیه مستثنی نیست. قالبی که یولو برای مجموعه داده دریافت می‌کند، به صورت درختی، به شکل زیر است.



تصویر ۱۲- قالب قابل قبول یولو

و برچسب‌ها باید به صورت یک فایل تکست<sup>۱۰۲</sup> برای عکس متناظر در پوشه عکس باشند. در ادامه مثالی برای درک بهتر خواهیم زد.



تصویر ۱۳- عملیات تشخیص برچسب

همان‌طور که دیده می‌شود، شماره کلاس در گوشه سمت چپ و بالای تصویر نوشته شده است. همچنین برای مشخص کردن قاب، از ۴ عدد استفاده می‌شود. طول و عرض نقطه بالا و چپ قاب و طول و عرض نقطه پایین سمت راست قاب. حال محتوای فایل برچسب را در ادامه مشاهده خواهید کرد.

0 0.483 0.656 0.962 0.510

تصویر ۱۴- یک ردیف از برچسب هر عکس

در هر ردیف، ۵ عدد وجود دارد. عدد اول شماره کلاس، که در اینجا چون فقط یک کلاس موجود است (کلاس برچسب ارزش محصول) پس کلاس صفر، نشان دهنده آن است. چهار عددی همه که برای نشان دادن قاب می باشد به ترتیبی که ذکر شد، در اینجا نوشته می شود.

فایل data.yaml است که در داخل پوشه Dataset قرار می گیرد و آدرس پوشه های train، test و validation داخل آن قرار می گیرد.

## ۴-۳- مدل تشخیص کلمات

همانطور که در فصل های گذشته اشاره شد، برای تشخیص کلمات از مدل گرفت استفاده شده است. تشخیص کلمات داخل برچسب از این رو مهم است که ما نیاز نداریم تمامی کلمات و جملات داخل برچسب را بخوانیم و فقط آن بخشی که برای ما مهم است، خوانده می شود.

برای مثال، شکل زیر عملکرد مدل گرفت قبل از آموزش بر روی مجموعه داده جدید را نشان می دهد.



تصویر ۱۵- مدل گرفت، قبل از آموزش

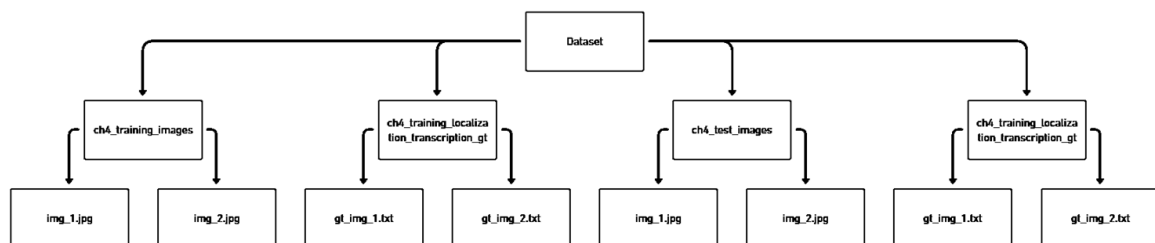
همان طور که دیده می شود، جملات کم اهمیت مانند «متوسط»، «زیاد» و «IFDA» که به آن ها نیازی نداریم خوانده شده اند اما کلمات و اعداد کلیدی مانند «قند» و «نمک» و یکسری از اعداد که رکن اساسی این پروژه محسوب می شوند، خوانده نشده است. از این رو نیاز داریم که این مدل را به آن صورت که مدنظر خودمان هست، آموزش دهیم.

فایل data.yaml این مدل، حاوی آرگومان های ورودی هنگام آموزش می باشد که چند مورد از این آرگومان هایی که در تنظیم مدل مهم تر هستند، توضیح داده می شود.



- مدل پیش‌زمینه<sup>۱۰۳</sup>: این پارامتر مشخص می‌کند که کدام مدل پیش‌زمینه برای استخراج ویژگی‌ها استفاده می‌شود. در اینجا، از معماری وی‌جی‌جی<sup>۱۰۴</sup> استفاده شده است که یک انتخاب رایج و کارآمد برای مدل‌های تشخیص متن مانند گرفت است.
- مسیر فایل مدل پیش‌آموزش‌دیده<sup>۱۰۵</sup>: استفاده از مدل پیش‌آموزش‌دیده بر روی داده‌های قبلی، می‌تواند باعث تسریع فرایند آموزش و افزایش دقت مدل نهایی شود. در اینجا مدل پیش‌فرضی که در گیت‌هاب گرفت بود<sup>۱۰۶</sup>، به عنوان مدل پیش‌آموزش‌دیده انتخاب گردید.
- نرخ یادگیری: همان‌طور که گفته شد، یکی از حیاتی‌ترین پارامترها در فرآیند آموزش مدل‌های یادگیری عمیق است. این پارامتر سرعت به‌روزرسانی وزن‌های شبکه در طول آموزش را تعیین می‌کند و تأثیر مستقیم بر روی همگرایی و عملکرد نهایی مدل دارد و در اینجا  $0.001$  قرار گرفت که پیش‌فرض خود مدل بود.
- اندازه دسته: این پارامتر نیز توضیح داده شد و نشان‌دهنده تعداد نمونه‌هایی است که در هر تکرار از آموزش پردازش می‌شوند. این پارامتر در اینجا  $5$  در نظر گرفته شد.
- تعداد دور: تعداد دورهای آموزشی که برای این مدل در نظر گرفته شده،  $10$  هزار دور است.

قالب مجموعه داده ای که در این مدل استفاده شد، با مدل قبلی کمی متفاوت است. ابتدا تصویری برای درک بهتر مجموعه داده در زیر نشان داده شده است.



تصویر ۱۶- قالب قابل قبول گرفت

در این مجموعه داده چهار پوشه وجود دارد. دو پوشه برای آموزش و دو پوشه برای ارزیابی مدل. یکی از دو پوشه، حاوی تصاویر می‌باشد و دیگری شامل فایل‌های برچسب که به فرمت تکست هستند. در ادامه یک نمونه از تصاویر و برچسب‌ها نشان داده شده است.

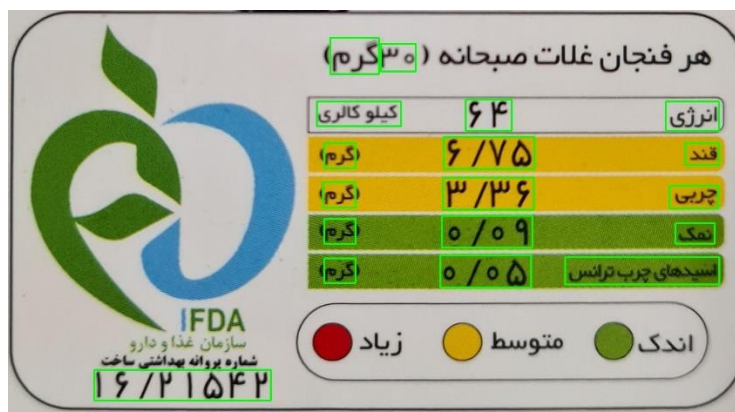
<sup>103</sup> Backbone

<sup>104</sup> VGG

<sup>105</sup> Pretrained (checkpoint)

<sup>106</sup> CRAFT\_clr\_amp\_14000.pth





تصویر ۱۷- تشخیص کلمات و ارقام به وسیله گرفت

852, 73, 981, 73, 981, 169, 852, 169, گرم  
836, 352, 919, 352, 919, 418, 836, 418, گرم  
833, 456, 919, 456, 919, 523, 833, 523, گرم  
836, 553, 919, 553, 919, 618, 836, 618, گرم  
836, 657, 919, 657, 919, 722, 836, 722, گرم  
817, 242, 1040, 242, 1040, 307, 817, 307, کیلو کالری  
1211, 237, 1333, 237, 1333, 319, 1211, 319, ۶۴  
1153, 335, 1393, 335, 1393, 420, 1153, 420, ۶/۷۵  
1151, 443, 1391, 443, 1391, 528, 1151, 528, ۳/۳۶  
1148, 544, 1387, 544, 1387, 629, 1148, 629, ۰/۰۹  
1144, 649, 1384, 649, 1384, 734, 1144, 734, ۰/۰۵  
228, 952, 695, 952, 695, 1033, 228, 1033, ۱۶/۳۱۵۴۳  
1740, 242, 1881, 242, 1881, 319, 1740, 319, انرژی  
1792, 352, 1881, 352, 1881, 410, 1792, 410, فند  
1750, 465, 1879, 465, 1879, 525, 1750, 525, جریب  
1766, 564, 1870, 564, 1870, 618, 1766, 618, نمک  
1474, 654, 1882, 654, 1882, 726, 1474, 726, اسیدهای جرب ترانس،  
986, 89, 1079, 89, 1079, 162, 986, 162, ۳۰

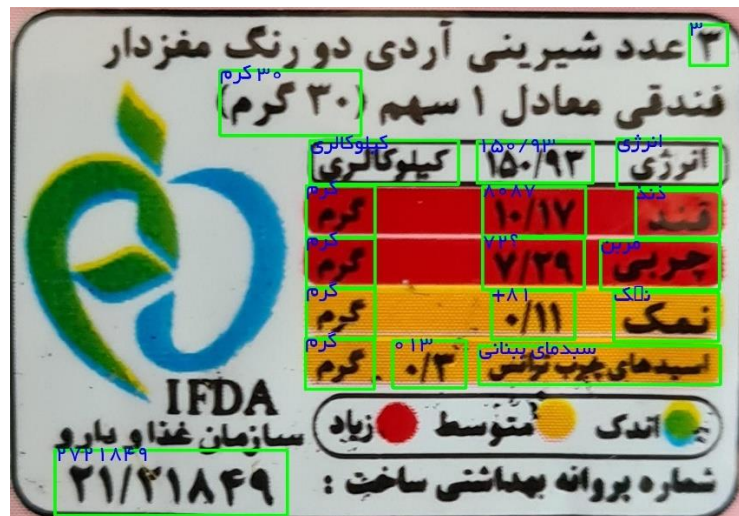
تصویر ۱۸- برچسب‌های تصویر قبل

در تصویر اول ۱۸ عدد کادر سبزرنگ وجود دارد که هر کدام نشان دهنده یک ردیف از فایل تکستی است که در تصویر بعدی نشان داده شده است. می‌باشد. هر ردیف از برجسب، ۹ ویژگی دارد. ۸ ویژگی اول به ترتیب دو به دو، طول و عرض گوشه‌های سمت چپ بالا، سمت راست بالا، سمت راست پایین و سمت چپ پایین می‌باشند. ویژگی آخر نشان دهنده نوشته‌ی داخل هر کادر است. این نوشته به تشخیص کلمه کمک می‌کند.

#### ۴-۴- مدخل خواندن کلمات

گام آخر در استخراج اطلاعات برچسب ارزش مواد غذایی، خواندن اطلاعاتیست که مدل استخراج کلمات به دست آورده. این اطلاعات با استفاده از یک مدل اوسی آر خوانده می‌شود که قبلاً به آن اشاره شد. این مدل از پیش آموزش دیده، عملکرد ضعیفی بر روی داده‌های مجموعه داده این پروژه دارد. در اینجا عملکرد این مدل، قبل از آموزش را مشاهده می‌کنیم.





تصویر ۱۹- خروجی مدل اوسی آر قبل از آموزش

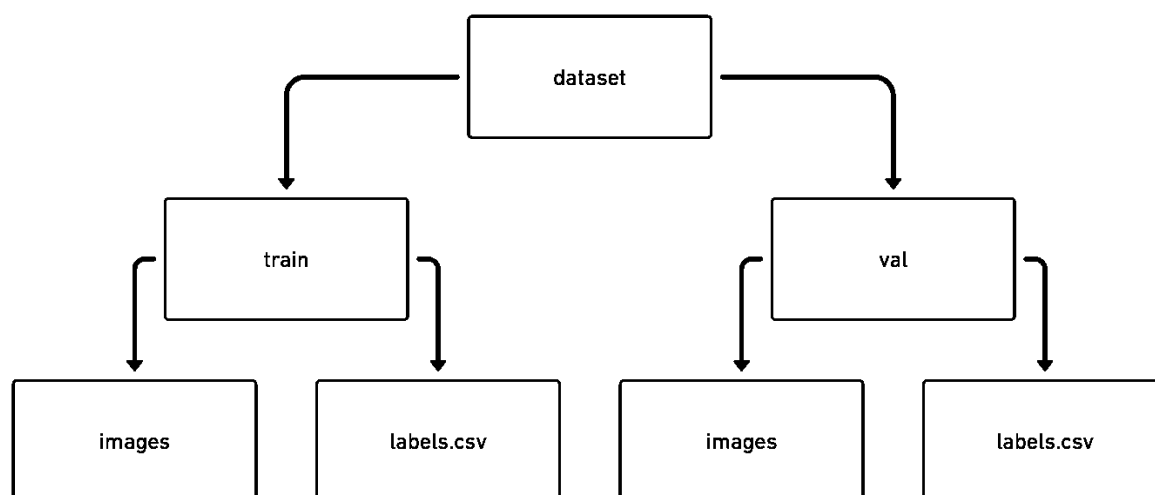
همانطور که در این تصویر مشاهده می کنید، عملکرد مدل از پیش آماده ضعیف است، دلایل مختلفی می تواند داشته باشد. مهم ترین دلایل به شرح زیر است.

- رنگ پس زمینه: اعداد و ارقام با رنگ پس زمینه به خوبی تشخیص داده نمی شوند زیرا تشخیص کد رنگی این رنگ ها دشوارتر است و مدل از قبل بر روی این پس زمینه ها آموزش ندیده.
- فونت مختلف: با اینکه در مصوبه ی درج این برچسب، فونت و اندازه نوشتار تایین شده، اما برخی از شرکت ها با فونت هایی غیر از فونت ذکر شده، اطلاعات را درج می کنند.
- داشتن نویز یا مات بودن: بعضی از برچسب ها دارای نویز هستند یا برخی اوقات عکاس، با لغزش دست عکسی را گرفته که موجب مات شدن تصویر شده است. در تصویر بالا نویز را روی دو کلمه ی «چرب ترانس». همچنین در این تصویر، برخی از حروف کمی مات هستند.

حال تصویر زیر که خروجی مدل، پس از آموزش است را مشاهده می کنیم.







تصویر ۲۱- قالب قابل قبول مدل اوسی آر

در این مجموعه داده دو پوشه آموزش و ارزیابی و در هر پوشه، تصاویر و یک فایل سی‌اس‌وی<sup>۱۰۷</sup> وجود دارند. تصاویر مربوط به کلماتی است که در برچسب‌های ارزش مواد غذایی وجود دارند و از تصاویر استخراج شده‌اند. فایل سی‌اس‌وی شامل دو ستون است. ستون آدرس و کلمه. آدرس هر کدام از این عکس‌ها در ستون آدرس و نوشته داخل هر عکس، داخل ستون کلمه نوشته می‌شود.

#### ۴-۵- جمع بندی

در این فصل، فرآیند استفاده از مدل‌های استفاده شده در این پروژه برای تشخیص و خواندن برچسب‌های ارزش مواد غذایی توضیح داده شده است. مدل تشخیص برچسب، برای تشخیص برچسب‌ها به کار گرفته شد و با استفاده از یک مدل پیش‌آموزش‌دیده، دقت و سرعت آن بهینه شد. سپس، مدل یافتن کلمات، برای تشخیص کلمات کلیدی داخل برچسب‌ها مورد استفاده قرار گرفت و عملکرد آن قبل و بعد از آموزش بررسی شد. در نهایت، مدل خوانش متن، برای خواندن متون استخراج‌شده به کار رفت و چالش‌های مربوط به تشخیص حروف و اعداد بررسی و نتایج بهبود یافته‌ی مدل پس از آموزش ارائه شد.



## فصل پنجم

### ارزیابی مدل‌ها و استفاده از اطلاعات

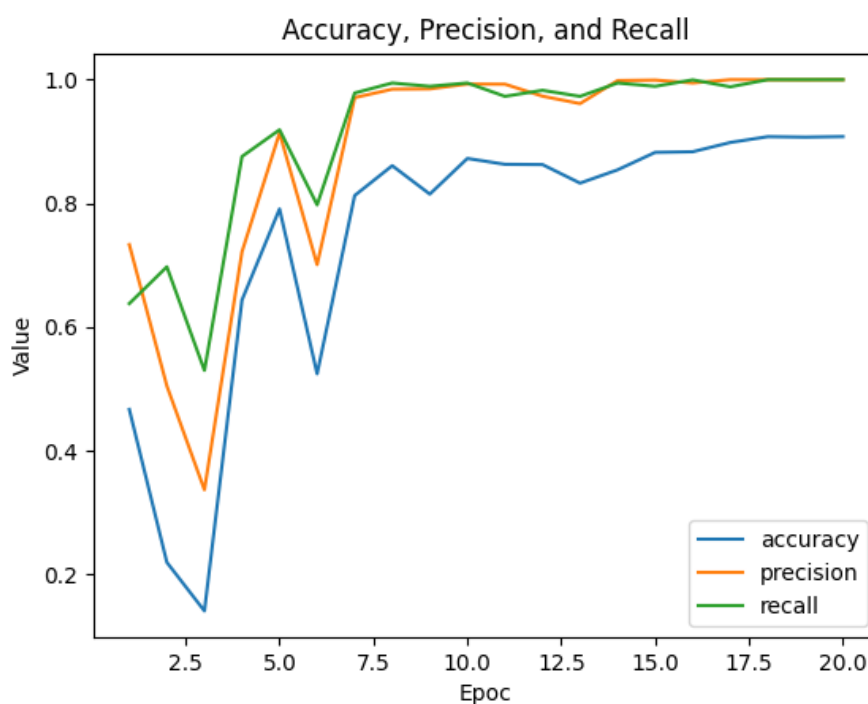
#### ۵-۱- مقدمه

این فصل به بررسی نتایج به‌دست‌آمده و معیارهای ارزیابی این پروژه اختصاص دارد. در این بخش، به‌منظور ارزیابی دقیق عملکرد مدل، از چندین معیار کلیدی استفاده شده است که شامل دقت، صحت، و بازخوانی می‌باشند. این معیارها به‌طور خاص برای سنجش توانایی مدل در پردازش و تحلیل داده‌ها مورد استفاده قرار گرفته‌اند. لازم به ذکر است که تمامی دقت‌های گزارش‌شده در این فصل، مرتبط با داده‌های آزمون هستند که ۲۰ درصد از کل داده‌های اصلی پروژه را تشکیل می‌دهند. ارزیابی دقیق مدل با استفاده از این داده‌ها، امکان تحلیل جامع‌تری از عملکرد مدل در شرایط واقعی را فراهم می‌سازد و به تعیین نقاط قوت و ضعف آن کمک می‌کند. همچنین، نتایج این ارزیابی‌ها می‌تواند راه‌گشای بهینه‌سازی‌های آتی و بهبود مدل‌های مشابه در پروژه‌های آینده باشد.

#### ۵-۲- ارزیابی مدل تشخیص برچسب

نمودار زیر، سه معیار دقت، صحت و بازخوانی را برای مدل یولو ورژن ۸ که ضیفه جدا کردن برچسب ارزش محصولات غذایی از عکس داشت، را نشان می‌دهد.





تصویر ۲۲- نمودار سه معیار دقت، صحت و بازخوانی برای مدل یولو

این نمودار نشان می‌دهد که مدل پس از چند دوره آموزشی، عملکرد بسیار خوبی را از نظر دقت، صحت و بازخوانی به دست آورده است. اگرچه در دوره‌های اولیه نوساناتی دیده می‌شود، اما با ادامه آموزش، مدل به سرعت یاد می‌گیرد و این نوسانات کاهش می‌یابند. در نهایت، مدل به یک عملکرد پایدار و بالا در تمامی این معیارها می‌رسد، که نشان‌دهنده‌ی آموزش موفقیت‌آمیز مدل است.

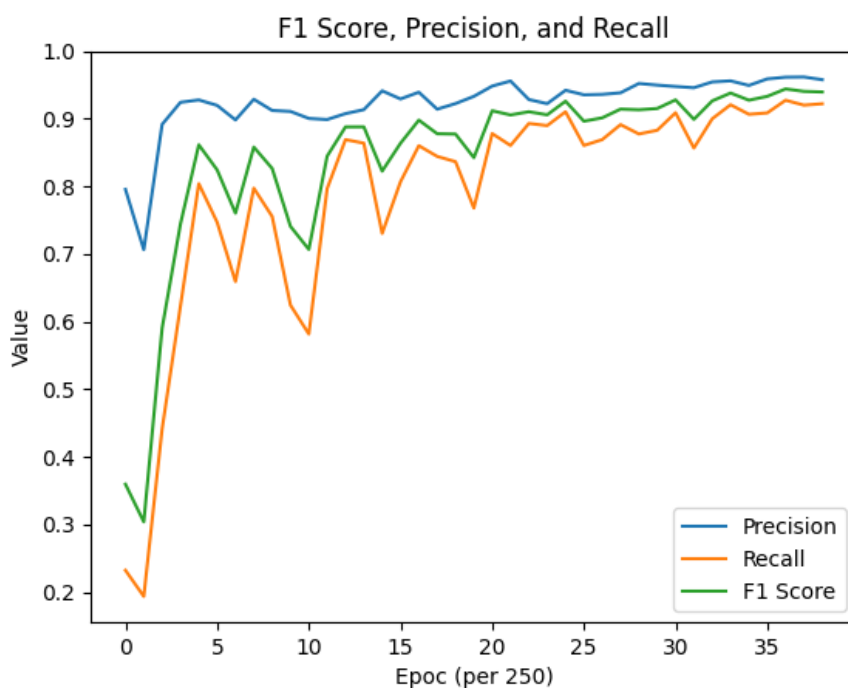
اعداد دقیق اولین و آخرین دوره آموزشی این مدل، در جدول زیر نمایان است.

جدول ۳- مقایسه آمار اولین و آخرین دوره آموزشی در یولو

	اولین دوره	آخرین دوره
دقت	۴۶/۶٪	۹۰/۷٪
صحت	۷۳/۳٪	۹۷/۸٪
بازخوانی	۶۳/۷٪	۹۸/۱٪

### ۵-۳- ارزیابی مدل تشخیص کلمات

نمودار زیر، سه معیار دقت، صحت و بازخوانی را برای مدل گرفت که ضیفه خواندن کلمات مدنظر بر روی برچسب ارزش محصولات غذایی را داشت، را نشان می‌دهد.



تصویر ۲۳- نمودار سه معیار دقت، صحت و میانگین هارمونیک برای مدل گرفت

نمودار نشان‌دهنده تغییرات صحت، بازخوانی و میانگین هارمونیک در طول ۱۰ هزار دوره آموزش مدل است. در ابتدای آموزش، هر سه معیار به سرعت بهبود یافته و به مقادیر بالاتری دست می‌یابند. صحت (خط آبی) معمولاً بالاتر از بازخوانی (خط نارنجی) است و در اکثر دوره‌ها به بیش از ۰/۹ می‌رسد. در مقابل، بازخوانی در ابتدای آموزش نوسانات بیشتری دارد اما به تدریج به ثبات رسیده و نزدیک به ۰/۹ می‌شود. F1 score (خط سبز) که میانگین هارمونیک این دو معیار است، به طور پیوسته بهبود می‌یابد و بین صحت و بازخوانی قرار دارد. به طور کلی، این نمودار نشان می‌دهد که مدل با گذشت زمان به دقت و قابلیت بازیابی مناسبی دست یافته است.

اعداد دقیق اولین و آخرین دوره آموزشی این مدل، در جدول زیر نمایان است.

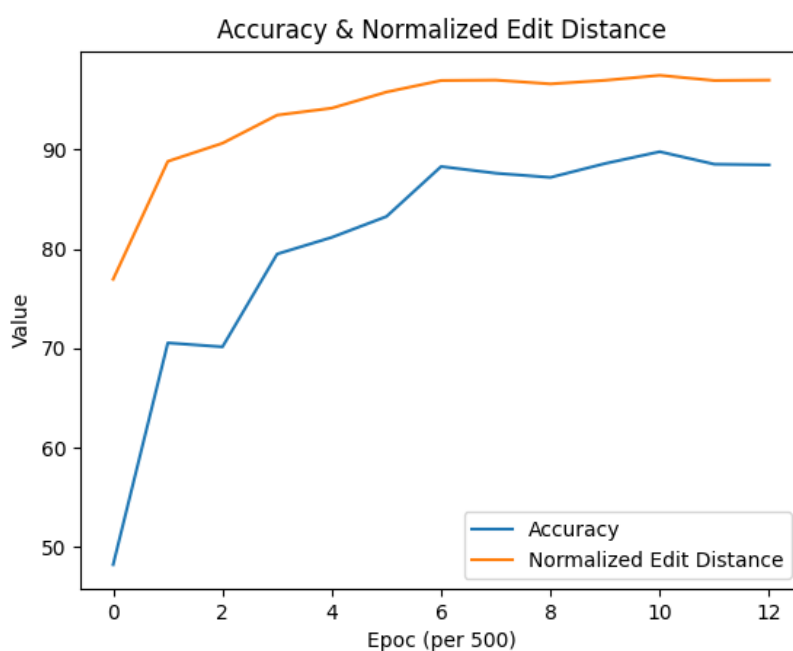


جدول ۴- مقایسه آمار اولین و آخرین دوره آموزشی در کرفت

	اولین دوره	آخرین دوره
میانگین هارمونیک	۳۵/۹٪	۹۲/۲٪
صحت	۷۹/۵٪	۹۵/۷٪
بازخوانی	۲۳/۲٪	۹۳/۹٪

## ۴-۵- ارزیابی مدل خوانش حروف و ارقام

نمودار زیر، رابطه بین دو معیار عملکرد مدل یعنی دقت و میانگین خطای نرمالیزه شده را در طول تعداد دوره‌های مختلف نمایش می‌دهد.



تصویر ۲۴- نمودار معیارهای دقت و میانگین خطای نرمالیزه شده در OCR

- **میانگین خطای نرمالیزه شده:** این مقدار با افزایش تعداد ایپاک‌ها ابتدا به سرعت افزایش می‌یابد و سپس به یک حد ثابتی نزدیک به ۹۶ درصد می‌رسد. این نشان می‌دهد که مدل به تدریج در حال بهبود است و خطاهای نرمالیزه شده کمتر می‌شود، اما بعد از چند دوره به یک ثبات نسبی می‌رسد.
- **دقت:** دقت نیز به تدریج افزایش می‌یابد و از حدود ۴۸ درصد در ابتدا به بیش از ۸۸ درصد می‌رسد. پس از یک سری دوره‌ها، دقت نیز به یک مقدار نسبتاً ثابت می‌رسد.

اما دلیل اینکه دقت از میانگین خطای نرمالیزه شده کمتر است، چیست؟

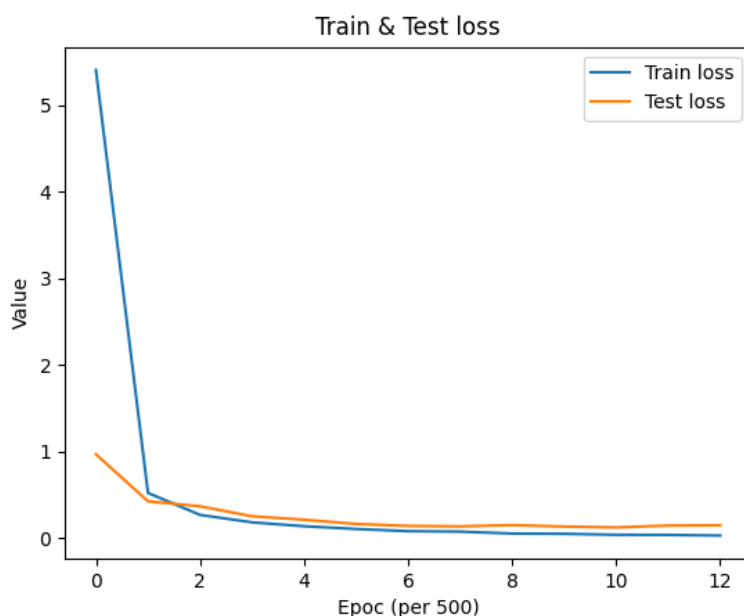
- **دقت** یک معیار بسیار سخت گیرانه است. اگر حتی یک کاراکتر در پیش بینی مدل اشتباه باشد، کل پیش بینی به عنوان اشتباه در نظر گرفته می شود و در نتیجه دقت کاهش می یابد.
- **میانگین خطای نرمالیزه شده** انعطاف پذیرتر است. این معیار با محاسبه فاصله ویرایشی، میزان نزدیکی پیش بینی به جواب درست را اندازه گیری می کند. بنابراین، اگر پیش بینی به جواب درست نزدیک باشد، حتی اگر کاملاً صحیح نباشد، همچنان مقدار نسبتاً بالایی برای این معیار به دست می آید.

اعداد دقیق اولین و آخرین دوره آموزشی این مدل، در جدول زیر نمایان است.

جدول ۵- مقایسه آمار دقت و میانگین خطای نرمالیزه شده در اولین و آخرین دوره

	اولین دوره	آخرین دوره
دقت	۴۸/۲٪	۸۸/۴٪
میانگین خطای نرمالیزه شده	۷۶/۹٪	۹۶/۹٪

تابع زیان این مدل برا داده های آموزش و آزمایش نیز در ادامه به نمایش درآمده اند.



تصویر ۲۵- تابع زیان داده های آموزش و آزمایش برای مدل OCR



هر دو خط مربوط به توابع زیان، روند کاهشی را نشان می‌دهند که این نشان دهنده مناسب بود آموزش و کمتر شدن خطای به وجود آمده در مدل است. مشاهده می‌شود که پس از یک تعداد دور، مقدار تابع زیان برای داده‌ها تغییری نکرده است و این درست جاییست که دیگر نیاز به آموزش نداریم چون با آموزش بیشتر، مدل بیش‌برازش می‌شود و نمی‌تواند داده‌های جدید را درست پیش‌بینی کند.

جدول ۶- مقایسه آمار مقادیر توابع زیان داده‌های آموزش و آزمایش در اولین و آخرین دوره

	اولین دوره	آخرین دوره
آموزش	۵/۴۰	۰/۰۴
آزمایش	۰/۹۷	۰/۱۴

## ۵-۵- تجزیه و تحلیل انرژی محصول

در این پروژه، بعد از اینکه مقادیر و کلمات نوشته شده بر روی برچسب ارزش غذایی خوانده شدند، حال با توجه به موقعیت کلمه «انرژی» در تصویر پیدا شد، عدد نوشته شده در سطر این کلمه، به عنوان انرژی هر سهم از محصول پیدا می‌شود. حال با استفاده از این عدد و مقدار MET مخصوص هر فعالیت بدنی و با فرض وزن ۷۵ کیلوگرم، در یک جدول اچ‌تی‌ام‌ال، اطلاعات مربوط به زمان مورد نیاز هر فعالیت بدنی برای سوزاندن آن سهم از محصول، ذکر شده است.

برای نمونه به جدول زیر دقت کنید.



Calories Burned by Exercise

EXERCISE TYPE	AMOUNT OF TIME NEEDED TO BURN 195.0 KCAL PER MINUTE
Slow walk	45
Fast walk	33
Slow bike	19
Moderate bike	15
Fast bike	12
Basketball	19
Bodyweight exercises	19
Slow swim	18
Moderate swim	15
Jump rope	14
Moderate run	12

تصویر ۲۶ - خروجی برنامه بعد از تجزیه و تحلیل انرژی هر سهم محصول

### ۲-۳- اطلاعات فنی محصول

با استفاده از سلنیوم، لینک سایت استعلام سیب سلامت همراه با پروانه بهداشتی محصول به عنوان کوئری<sup>۱۰۸</sup> به این سایت فرستاده شد و پس از دریافت اطلاعات فنی محصول، تصویری از این صفحه به عنوان خروجی در کنار سایر خروجی‌ها (تصویر محصول به همراه کلمات خوانده شده و جدول سوزاندن کالری دریافتی برحسب فعالیت بدنی) با همان نام ذخیره می‌شود.

تصویر زیر، نمای کلی وب سایت پس از استعلام محصول می‌باشد.



پروژه آیت نام

IFDA سازمان غذا و دارو

ثبت شرکت به مسئول

شماره پروانه ساخت: ۵۰۲۰۴۹۵

نوع فرآورده: نوشیدنی مالت با طعم طعم حاوی شربت ذرت با فروکتوز بالا

نام تجاری یا برند: تالینس

مشخصات بسته بندی:

نام کارخانه: تک در چادران

نظای کارخانه: شرکت صنعتی چادران - صنعت ۲

تلفن کارخانه: ۰۲۱۳۳۳۳۳۳۳

تاریخ صدور: ۱۳۹۹/۰۲/۰۴

تاریخ انقضا: ۱۴۰۹/۰۲/۰۴

وضعیت پروانه: معتبر

© ۱۳۹۴ - کلیه حقوق مادی و معنوی به سازمان غذا و داروی کشور تعلق دارد

تصویر ۲۷- اطلاعات فنی و نمای وبسایت پس از استعلام محصول

## ۵-۶- جمع بندی

در فصل پنجم، ارزیابی و تحلیل عملکرد مدل‌های مختلف مورد استفاده در پروژه به‌طور جامع مورد بررسی قرار گرفت. معیارهای کلیدی مانند دقت، صحت، و بازخوانی به عنوان شاخص‌های اصلی برای سنجش کارایی مدل‌ها در نظر گرفته شدند. این فصل نشان می‌دهد که مدل‌ها پس از چندین دوره آموزشی به عملکرد مطلوبی دست یافته‌اند و نوسانات اولیه به تدریج کاهش یافتند. نتایج به‌دست‌آمده از این ارزیابی‌ها، نقاط قوت و ضعف مدل‌ها را روشن کرده و زمینه‌ساز بهبود و بهینه‌سازی‌های آینده شده است. همچنین، در این فصل به تحلیل دقیق انرژی محصولات و اطلاعات فنی مربوط به آن‌ها نیز پرداخته شد.



## فصل ششم

### جمع بندی و پیشنهادهایی برای بهبود پژوهش

#### ۶-۱- جمع بندی کلی

این پروژه با هدف توسعه یک سیستم جامع برای تحلیل برچسب ارزش غذایی محصولات و ارزیابی آن‌ها از جنبه‌های مختلف طراحی شده است. مراحل اصلی پروژه شامل پردازش تصویر، تشخیص برچسب‌ها، خوانش کلمات و ارقام، و در نهایت تحلیل اطلاعات مربوط به انرژی محصولات غذایی بوده است.

ابتدا، از مدل‌های پیشرفته‌ی یادگیری عمیق مانند یولو برای تشخیص برچسب‌های غذایی و مدل کرفت برای خواندن کلمات و ارقام استفاده شده است. این مدل‌ها پس از چندین دوره آموزشی، عملکردی بالا در دقت، صحت و بازخوانی از خود نشان داده‌اند. ارزیابی‌های انجام‌شده بر روی این مدل‌ها نقاط قوت و ضعف آن‌ها را مشخص کرده و منجر به بهینه‌سازی‌های مؤثر در روند پروژه شده است.

در ادامه، اطلاعات مربوط به انرژی محصولات از طریق تحلیل متنی استخراج و با استفاده از فرمول‌های محاسباتی، زمان مورد نیاز برای سوزاندن کالری دریافتی از هر محصول با فعالیت‌های بدنی مختلف محاسبه و در قالب یک جدول ارائه شده است. همچنین، با استفاده از ابزارهایی نظیر سلنیوم، اطلاعات فنی محصولات از وبسایت‌های مرتبط جمع‌آوری و به عنوان بخشی از گزارش نهایی ارائه شده است.

به‌طور کلی، این پروژه نشان‌دهنده یکپارچگی در استفاده از تکنیک‌های یادگیری ماشین، پردازش تصویر، و استخراج اطلاعات وب بوده و قابلیت‌های خود را در تحلیل و ارائه اطلاعات کاربردی در حوزه سلامت و تغذیه به‌خوبی به نمایش گذاشته است.



## ۶-۲- بهبودهای قابل اعمال

با توجه به گستردگی زیاد پروژه، نقاط و ظرفیت بالایی برای بهبود وجود دارد و اینکه که پروژه به طور کمال و تماماً بهینه پیاده سازی نشده و جای کار زیادی باقیست. در ادامه، لیستی از بهبودهای ممکن مشاهده می‌شود.

- **مجموعه داده:** با اینکه تا کنون مجموعه داده عظیمی جمع آوری شده که حدوداً ۱۰۰۰ برچسب ارزش غذایی محصول دارد اما در زمینه یادگیری ماشین و یادگیری عمیق، چیزی به نام داده کافی وجود ندارد و هرچه تعداد داده‌ها بیشتر باشد، دقت مدل بالاتر خواهد رفت. همچنین در این پروژه تمرکز بر روی برچسب ارزش غذایی استاندارد بود و برچسب‌های غیر استاندارد به درستی خوانده نمی‌شوند. به همین دلیل، اضافه کردن این امکان به این پروژه، باعث کامل‌تر شدن مجموعه داده می‌شود.
- **بهبود مدل یولو:** در این پروژه از یولو ورژن ۸ استفاده شد. یکی از این دلایل، بهتر بودن پشتیبانی و مستندات<sup>۱۰۹</sup> خوب این مدل بود. مدل‌های بهتر چه در این زمان و چه در آینده معرفی خواهند شد که با استفاده از آنها، می‌توان دقت بخش اول این پروژه را از مقدار فعلی بهبود بخشید.
- **پیاده سازی تکنیک‌های داده افزایی<sup>۱۱۰</sup>:** به کارگیری تکنیک‌های داده‌افزایی می‌تواند باعث افزایش تنوع داده‌های ورودی و بهبود عملکرد مدل‌ها شود. مثلاً تغییراتی در رنگ، اندازه، چرخش و غیره بر روی تصاویر برچسب‌ها اعمال شود تا مدل‌ها برای داده‌های واقعی‌تر آموزش ببینند.
- **خواندن برچسب‌های کج و زاویه دار:** در این پروژه، تصویری که کاربر باید وارد کند، تصویری بدون زاویه و چرخش است. می‌توان با استفاده از الگوریتم‌های مختلف، این مشکل را نیز برطرف و زاویه و چرخش تصاویر را مدیریت کرد.
- **بهینه‌سازی پردازش متن و استخراج اطلاعات:** می‌توان از الگوریتم‌های پیشرفته‌تر پردازش زبان طبیعی<sup>۱۱۱</sup> برای تحلیل دقیق‌تر متن‌های موجود بر روی برچسب‌ها استفاده کرد. این امر می‌تواند به بهبود دقت در تشخیص و تفسیر اعداد و کلمات کمک کند و خروجی معنی‌دارتری به کاربر نشان دهد.
- **استفاده از روش‌های یادگیری تجمعی<sup>۱۱۲</sup>:** ترکیب چند مدل مختلف و استفاده از روش‌های یادگیری تجمعی می‌تواند به بهبود دقت نهایی پروژه منجر شود. مثلاً ترکیب خروجی چندین مدل مختلف برای تشخیص برچسب‌ها و کلمات می‌تواند نتیجه نهایی دقیق‌تری را ارائه دهد.
- **افزایش سرعت و کارایی پردازش:** بهینه‌سازی کد و استفاده از سخت‌افزارهای قوی‌تر مانند GPUها یا TPUها می‌تواند به افزایش سرعت پردازش و کاهش زمان اجرای کل سیستم کمک کند. این به‌ویژه در مواقعی که نیاز به پردازش تعداد زیادی تصویر یا متن داریم، بسیار مفید است.
- **بهبود رابط کاربری و تجربه کاربر:** بیشتر وقت این پروژه، صرف برچسب زدن مجموعه داده و آموزش مدل‌ها شد در نتیجه شاهد رابط کاربری حرفه‌ای نیستیم. طراحی یک رابط کاربری بهتر و ساده‌تر برای

<sup>109</sup> Documentation

<sup>110</sup> Augmentation

<sup>111</sup> Natural language processing

<sup>112</sup> Ensemble learning



کاربران نهایی می‌تواند استفاده از سیستم را راحت‌تر و جذاب‌تر کند. به‌علاوه، ارائه نتایج به‌صورت بصری و تعاملی می‌تواند تجربه کاربر را بهبود بخشد.

- **راه‌اندازی برنامه برای موبایل:** با توجه به اینکه تصاویر از طریق دوربین به برنامه داده می‌شوند و این دوربین عموماً دوربین تلفن همراه است، پس اگر بتوان نرم‌افزاری برای تلفن همراه نوشت که کاربران بتوانند به صورت مستقیم و در نرم‌افزار تصویر برداری کرده و به صورت آنی، نتایج را مشاهده کنند، تجربه کاربری بسیار بهتری به مصرف‌کننده انتقال داده می‌شود.
- **تحلیل کامل و جامع تر:** در این پروژه فقط مشاهده شد که با مصرف هرسهم از محصول، چه مدت زمان از هر فعالیت بدنی نیاز است تا آن مقدار انرژی دفع شود. اما این انتهای کار نیست. در این پروژه می‌توان با مقدار اندازه‌گیری و تحلیل ۴ ماده غذایی یعنی قند، نمک، چربی و اسیدهای چرب ترانس، تحلیل جامع و کامل‌تری بر اطلاعات داشت. همچنین با مقایسه انرژی هرسهم و مقدار هرسهم از محصول برحسب گرم، که هربار خوانده می‌شود، می‌توان از سالم یا ناسالم بودن آن محصول، اطمینان حاصل کرد.
- **استفاده از رنگ‌ها:** رنگ‌ها در این پروژه، نقش مزاحمی در خواندن اطلاعات داشتند و این پروسه را سخت‌تر کردند. اما می‌توان به آن‌ها نیز به عنوان داده نگاه کرد. با استخراج رنگ‌ها از برچسب، می‌توان اطلاعات مفیدی از برچسب ارزش مواد غذایی استخراج کرد و تحلیلی بهتر داشت.



## منابع

- [۱] General Department of Supervision and Evaluation of Food Products and F. a. D. O. cosmetics, "Instruction and guide for inserting nutritional color indicator." [Online]. Available: [https://fdo.skums.ac.ir/Dorsapax/Data/Sub\\_5/File/neshangar%20rangi.pdf](https://fdo.skums.ac.ir/Dorsapax/Data/Sub_5/File/neshangar%20rangi.pdf)
- [۲] "Roboflow." <https://roboflow.com/> (accessed).
- [۳] "Label studio." <https://labelstud.io/> (accessed).
- [۴] "Saitak number convertor." <https://www.saitak.com/number> (accessed).
- [۵] "Kaggle." <https://www.kaggle.com/> (accessed).
- [۶] "EasyOCR repository." <https://github.com/JaidedAI/EasyOCR> (accessed).
- [۷] "Craft repository." <https://github.com/clovaai/CRAFT-pytorch> (accessed).
- [۸] "Deep text recognition benchmark repository." <https://github.com/clovaai/deep-text-recognition-benchmark> (accessed).
- [۹] "How Do You Calculate Calories Burned During Exercise?" [https://www.medicinenet.com/how\\_to\\_calculate\\_calories\\_burned\\_during\\_exercise/article.htm](https://www.medicinenet.com/how_to_calculate_calories_burned_during_exercise/article.htm) (accessed).
- [۱۰] "Metabolic Equivalent." <https://www.whyiexercise.com/metabolic-equivalent.html> (accessed).
- [۱۱] "Scan Nutrition Labels to Track Foods Faster." <https://macrofactorapp.com/label-scanner-announcement/> (accessed).
- [۱۲] J. REIBRING, "Photo OCR for Nutrition Labels," 2017. [Online]. Available: <https://odr.chalmers.se/bitstream/20.500.12380/256649/1/256649.pdf>
- [۱۳] R .S. Nate Matsunaga, "Image processing for the extraction of nutritional information from food labels," Santa Clara University, 2015. [Online]. Available: [https://scholarcommons.scu.edu/cgi/viewcontent.cgi?article=1041&context=cseng\\_senior](https://scholarcommons.scu.edu/cgi/viewcontent.cgi?article=1041&context=cseng_senior)
- [۱۴] T. Zaman, "Vision Based Extraction of Nutrition Information from Skewed Nutrition Labels," 2016. [Online]. Available: <https://digitalcommons.usu.edu/etd/4893>
- [۱۵] "Yolov8." <https://github.com/ultralytics/ultralytics> (accessed).

