

## تشخیص کاراکترهای نوری و کاربرد آن در سازمان ارتش

بهشاد پنجه زاده

دانشجوی دکتری بیومکانیک ورزشی، دانشکده تربیت بدنی و علوم ورزشی، دانشگاه آزاد اسلامی، واحد تهران مرکزی، تهران، ایران.

حیدر آزادی نکو

کارشناسی ارشد دانشگاه آزاد اسلامی، واحد علوم تحقیقات کردستان

### چکیده

هدف این مطالعه بررسی تشخیص کاراکترهای نوری (OCR) و کاربردهای آن در دنیای واقعی، به ویژه در سازمان‌های نظامی است. فناوری (OCR) استخراج اطلاعات متنی را از اسناد اسکن شده، تصاویر یا سایر منابع دیجیتالی امکان پذیر می‌کند. از بینایی کامپیوتر و یادگیری ماشین برای تبدیل متن چاپ شده یا دست نویس به داده‌های قابل ویرایش و جستجو استفاده می‌کند. این فرآیند شامل اسکن تصویر ورودی، تجزیه و تحلیل آن، شناسایی کاراکترها یا کلمات و تبدیل آن‌ها به متن قابل خواندن توسط ماشین است. (OCR) کاربردهای مختلفی مانند دیجیتالی کردن اسناد، استخراج متن، ورود داده‌ها، ترجمه، تجزیه و تحلیل، تشخیص پلاک خودرو، کشف تقلب، مدیریت اطلاعات، دسترسی برای افراد کم بینا و پردازش اسناد دارد. به طور کلی، فناوری (OCR) داده کاوی و پردازش اسناد را متحول کرده است و کارایی و دقت را در طیف گسترده‌ای از برنامه‌ها ارائه می‌دهد. در نتیجه، فناوری (OCR) در عملیات نظامی برای پردازش کارآمد و دقیق اطلاعات متنی، افزایش قابلیت‌های پردازش داده‌ها، سرعت بازیابی اطلاعات، آگاهی از موقعیت و سادگی عملیات بسیار مهم است. با تبدیل اطلاعات چاپی یا دست نویس به فرمت دیجیتال، (OCR) از اطلاعات نظامی، نظارت و عملیات تصمیم گیری پشتیبانی می‌کند.

**واژگان کلیدی:** تشخیص کاراکتر نوری (OCR)، ارتش.

## مقدمه

تشخیص کاراکتر نوری (OCR) یک فناوری است که تبدیل متن چاپ شده یا دست نویس را به متن کدگذاری شده توسط ماشین تسهیل می‌کند. فناوری (OCR) در صنایع مختلفی مانند دیجیتالی کردن اسناد، بایگانی، ورود خودکار داده‌ها، انتشار الکترونیکی، سنتز متن به گفتار و غیره کاربرد دارد. فرآیند (OCR) مستلزم اسکن یا گرفتن تصاویر اسناد متنی، و سپس تجزیه و تحلیل و تفسیر کاراکترهای درون آن تصاویر برای استخراج اطلاعات متنی است. فرآیند (OCR) معمولی شامل مراحل زیر است:

ضبط تصویر: اسناد متنی با استفاده از اسکنرها، دوربین‌ها یا سایر دستگاه‌های ضبط تصویر گرفته می‌شوند و در نتیجه نمایش دیجیتالی سند ایجاد می‌شود. پیش پردازش: تصویر گرفته شده تحت تکنیک‌های مختلف پیش پردازش قرار می‌گیرد تا کیفیت را افزایش داده و آن را برای تشخیص کاراکتر آماده کند. این ممکن است شامل کارهایی مانند حذف نویز، چرخش تصویر، تصحیح اعوجاج و دوتایی (تبدیل تصویر به سیاه و سفید) باشد. محلی سازی متن: الگوریتم‌های (OCR) تصویر از پیش پردازش شده را برای مشخص کردن قسمت‌های حاوی متن تجزیه و تحلیل می‌کنند و به جداسازی متن از سایر عناصر تصویر مانند گرافیک، جداول و غیره کمک می‌کنند. تقسیم بندی کاراکترها: پس از تعریف مناطق متن، مرحله بعدی تقسیم کاراکترهای منفرد در آن مناطق با جدا کردن هر کاراکتر از متن اطراف است. استخراج ویژگی: ویژگی‌های مختلف کاراکترهای تقسیم‌بندی شده برای ایجاد یک نمایش عددی که ویژگی‌های متمایز آن‌ها را نشان می‌دهد، استخراج می‌شوند که می‌تواند شامل عرض، شکل، ارتفاع و پارامترهای مرتبط باشد. طبقه‌بندی کاراکترها: ویژگی‌های استخراج شده توسط الگوریتم‌های طبقه‌بندی مانند شبکه‌های عصبی یا مدل‌های آماری برای شناسایی و طبقه‌بندی هر شخصیت با مقایسه ویژگی‌ها با الگوهای از پیش آموزش دیده یا مدل‌های کاراکتر استفاده می‌شوند. پس از پردازش: پس از طبقه بندی کاراکترها، مراحل پس از پردازش اضافی ممکن است برای افزایش دقت و قابلیت اطمینان متن شناخته شده، شامل فعالیت‌هایی مانند بررسی املا، تصحیح خطا، و تجزیه و تحلیل مبتنی بر زمینه اجرا شود. خروجی متن: سپس متن شناسایی شده به عنوان خروجی تولید می‌شود که می‌تواند بیشتر پردازش، ذخیره یا برای برنامه‌هایی مانند جستجوی متن، داده کاوی، ترجمه زبان و بازیابی اطلاعات مورد استفاده قرار گیرد (Hussain et al, 2009; Plamondon and Srihari, 2000; Tappert et al, 1990; Clausner et al, 2017). بخش‌های بعدی به تفصیل به هر مرحله از تشخیص کاراکترهای نوری می‌پردازد و اهمیت شبکه‌های عصبی عمیق در هوش مصنوعی و پیشرفت‌های اساسی آن‌ها در (OCR) را توضیح می‌دهد. علاوه بر این، نمونه‌هایی نشان دهنده کاربرد فناوری (OCR) در صنایع مختلف و سناریوهای دنیای واقعی ارائه شده است. در نهایت، در بخش نتیجه‌گیری، کاربرد و ضرورت استفاده از فناوری OCR در ارتش جمهوری اسلامی ایران توضیح داده می‌شود.

## تشخیص کاراکترهای نوری

### ثبت تصویر

اکتساب تصویر اولین مرحله در فرآیند تشخیص کاراکتر نوری است که در آن یک تصویر حاوی متن گرفته می‌شود و برای تجزیه و تحلیل بیشتر و استخراج متن آماده می‌شود. هدف به دست آوردن یک تصویر با کیفیت بالا است که بتواند به طور موثر توسط الگوریتم‌های (OCR) پردازش شود. در اینجا توضیح مختصری از فرآیند ثبت تصویر در (OCR) آورده شده است:

۱. تصویربرداری: فرآیند با گرفتن تصویری که حاوی متنی است که باید تشخیص داده شود، شروع می‌شود. این کار را می‌توان با استفاده از دستگاه‌های مختلف مانند اسکنر، دوربین یا سخت افزار تخصصی (OCR) انجام داد.

۲. پیش پردازش تصویر: هنگامی که تصویر گرفته می‌شود، تحت پیش پردازش قرار می‌گیرد تا کیفیت آن افزایش یابد و برای تجزیه و تحلیل (OCR) مناسب شود. این معمولاً شامل چندین مرحله است، از جمله:

- تمیز کردن تصویر: حذف هرگونه نویز، مصنوعات، یا عناصر ناخواسته از تصویر، مانند خط و خش، یا به هم ریختگی پس زمینه.
  - افزایش کنتراست: تنظیم کنتراست تصویر برای بهبود تمایز بین متن و پس زمینه و اطمینان از دقت (OCR) بهتر.
  - چرخش تصویر: اصلاح هر گونه شیب یا انحراف در تصویر ناشی از دوربین یا فرآیند اسکن برای اطمینان از تراز مناسب متن.
  - دوتایی سازی<sup>۱</sup>: تبدیل تصویر به فرمت باینری، که در آن پیکسل های متن به صورت سیاه و پس زمینه به صورت سفید نمایش داده می شوند. این مراحل پردازش بعدی را ساده می کند.
۳. وضوح و مقیاس: وضوح تصویر برای اطمینان از عملکرد (OCR) بهینه تنظیم می شود. تصاویر با وضوح بسیار پایین ممکن است منجر به از دست دادن جزئیات متن شوند، در حالی که تصاویر با وضوح بسیار بالا ممکن است زمان پردازش و منابع مورد نیاز را افزایش دهند. مقیاس گذاری نیز برای عادی سازی اندازه تصویر انجام می شود و آن را با الگوریتم های (OCR) مورد استفاده سازگار می کند.
۴. کاهش نویز: تکنیک های مختلفی برای کاهش نویز یا مصنوعات که می توانند با دقت (OCR) تداخل داشته باشند، اعمال می شوند. این می تواند شامل فیلترها، الگوریتم های هموارسازی یا روش های آستانه تطبیقی برای تشخیص متن از نویز پس زمینه باشد.
۵. فشرده سازی تصویر: در برخی موارد، از تکنیک های فشرده سازی تصویر برای کاهش اندازه فایل تصویر بدون تاثیر قابل توجهی بر دقت (OCR) استفاده می شود. الگوریتم های فشرده سازی رایج شامل فرمت های (JPEG)، (PNG)، یا (TIFF) هستند.
۶. فرمت و رمزگذاری تصویر: تصویر معمولاً در فرمت استاندارد مانند (JPEG)، (PNG)، (TIFF) یا (PDF) ذخیره می شود. انتخاب فرمت به نیازهای خاص سیستم یا برنامه (OCR) بستگی دارد.
- توجه به این نکته مهم است که سیستم ها و برنامه های OCR مختلف ممکن است تغییرات خاصی یا مراحل اضافی در فرآیند دریافت تصویر داشته باشند. مراحل ذکر شده در بالا یک نمای کلی ارائه می دهد (McElhaney, 2008).

### پیش پردازش تصویر

پیش پردازش یک مرحله مهم در تشخیص کاراکتر نوری (OCR) است که شامل آماده سازی تصویر برای استخراج متن دقیق است. هدف آن افزایش کیفیت تصویر، حذف نویز، مصنوعات و افزایش خوانایی متن است. در زیر مراحل پیش پردازش (OCR) توضیح داده شده است:

تمیز کردن تصویر: بر حذف عناصر ناخواسته مانند نویز، خط و خش و به هم ریختگی پس زمینه با استفاده از فیلترها یا عملیات مورفولوژیکی تمرکز می کند.

<sup>1</sup> . Binarization.

دوتایی سازی تصویر: تصویر را به فرم باینری تبدیل می‌کند، متن سیاه و پس‌زمینه سفید است و به جداسازی متن از پس‌زمینه کمک می‌کند.

تصحیح اعوجاج: زوایای انحرافی را برای تراز کردن متن به صورت افقی یا عمودی برای تشخیص دقیق کاراکترها شناسایی و تصحیح می‌کند.

کاهش نویز: عناصر ناخواسته را برای افزایش دقت OCR با استفاده از روش‌هایی مانند فیلترهای میانی یا صاف کردن گاوسی سرکوب می‌کند.

تقویت کنتراست: با تنظیم کنتراست تصویر یا هیستوگرام برای تشخیص بهتر کاراکتر، تمایز متن-پس زمینه را بهبود می‌بخشد (Lin et al, 2015).

مقیاس‌بندی و عادی‌سازی: تصاویر را به وضوح استاندارد سازگار با الگوریتم‌های (OCR) تغییر اندازه می‌دهد و تغییرات اندازه، جهت یا نسبت ابعاد را در نظر می‌گیرد.

روش‌های پیش پردازش می‌تواند بر اساس سیستم (OCR)، ویژگی‌های تصویر و نیازهای کاربردی متفاوت باشد (Bieniecki et al, 2007; Gonzalez and Woods, 2017; Alginahi, 2010).

## محلی سازی متن

محلی سازی متن یک مرحله مهم در تشخیص کاراکتر نوری است که شامل شناسایی مناطقی از یک تصویر است که حاوی متن است. هدف شناسایی دقیق مناطق متن و جداسازی آن‌ها از پس زمینه یا سایر عناصر غیر متنی است. در اینجا توضیح مختصری از فرآیند محلی سازی متن در (OCR) آورده شده است:

۱. تشخیص لبه: تکنیک‌های تشخیص لبه اغلب به عنوان اولین گام در محلی سازی متن استفاده می‌شود. این روش‌ها تغییرات ناگهانی در شدت یا رنگ را در تصویر شناسایی می‌کنند که می‌تواند با مرزهای متن مطابقت داشته باشد. الگوریتم‌های رایج تشخیص لبه شامل آشکارساز لبه (Canny)، عملگر (Sobel) یا عملگر (LoG)<sup>۱</sup> است (CANNY, 1987).

۲. تجزیه و تحلیل اجزای متصل: تجزیه و تحلیل اجزای متصل پیکسل‌های مجاور را که مناطق متصل را تشکیل می‌دهند، گروه‌بندی می‌کند. در زمینه محلی سازی متن، این مرحله به شناسایی مناطق متن بالقوه بر اساس اتصال پیکسل‌های لبه کمک می‌کند. مناطقی با اتصال بالا و نسبت ابعاد مناسب بیشتر دارای متن هستند.

۳. استخراج نامزدهای متن: پس از شناسایی اجزای متصل، معیارهای اضافی برای فیلتر کردن مناطق غیر متنی و استخراج نامزدهای متن اعمال می‌شود. این معیارها ممکن است شامل ویژگی‌هایی مانند نسبت ابعاد، ارتفاع، عرض یا سایر ویژگی‌های آماری باشد که معمولاً با متن مرتبط است.

<sup>1</sup> . Laplacian of Gaussian.

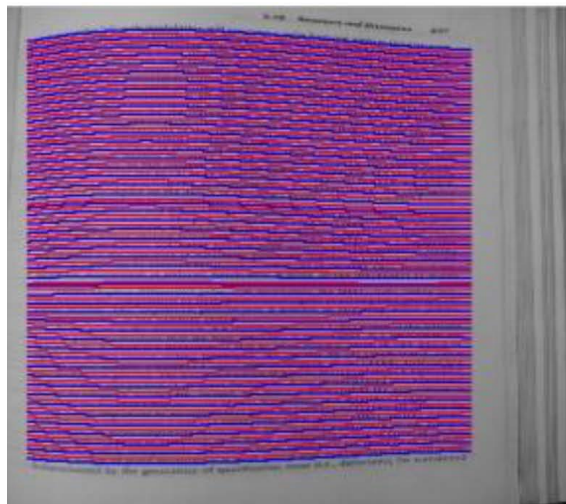
۴. تصدیق متن: نامزدهای متن استخراج شده برای تأیید این که آیا آن‌ها مناطق متن واقعی هستند یا خیر، تأیید می‌شوند. این مرحله می‌تواند شامل تکنیک‌های مختلفی مانند تجزیه و تحلیل بافت، سازگاری عرض ضربه، تقسیم‌بندی کاراکتر یا الگوریتم‌های یادگیری ماشینی برای تمایز بین مناطق متنی و غیر متنی باشد.

۵. تصحیح متن: در برخی موارد، مناطق متن شناسایی شده ممکن است کج یا مخدوش شوند. تکنیک‌های تصحیح متن برای هم‌ترازی مناسب نواحی متن، جبران اثرات چرخش یا پرسپکتیو به کار می‌رود. این تضمین می‌کند که متن در قالب قابل خواندن برای پردازش OCR بعدی باشد.

۶. حذف غیر متنی: در نهایت، عناصر غیر متنی باقیمانده یا موارد مثبت کاذب از مناطق متن شناسایی شده حذف می‌شوند. این را می‌توان با استفاده از روش‌های اکتشافی، روش‌های مبتنی بر قانون، یا رویکردهای یادگیری ماشینی برای تمایز بین عناصر متنی و غیر متنی انجام داد.

توجه به این نکته مهم است که تکنیک‌های محلی سازی متن می‌تواند بسته به سیستم OCR خاص، ویژگی‌های تصویر و الزامات برنامه متفاوت باشد (Gonzalez and Woods, 2017؛ Plamondon and Srihari, 2000؛ Fawzi et al, 2015).

در شکل ۱ نمونه‌ای از محلی سازی یک تصویر از متن یک کتاب نشان داده شده است. همانطوری که در تصویر مشخص است پس از محلی سازی مکان متن، با رنگ بنفش مشخص شده است.



شکل ۱: شناسایی مناطقی از تصویر است که حاوی متن است (Fawzi et al. 2015).

## تقسیم‌بندی کاراکتر

تقسیم‌بندی کاراکتر یک مرحله مهم در تشخیص کاراکتر نوری است که شامل جدا کردن کاراکترهای فردی از یک تصویر متنی است. تقسیم بندی دقیق کاراکترها برای تشخیص بعدی کاراکترها و استخراج دقیق متن ضروری است. در اینجا توضیح مختصری از فرآیند تقسیم کاراکتر در (OCR) آمده است:

۱. تجزیه و تحلیل مؤلفه‌های متصل: تجزیه و تحلیل مؤلفه‌های متصل، همانطور که در فرآیند محلی سازی متن ذکر شد، اغلب گام اولیه در تقسیم بندی کاراکترها است. اجزای متصل بر اساس اتصال پیکسل‌های پیش زمینه شناسایی می‌شوند. در این مرحله، مناطق متن بیشتر به اجزای کوچک‌تر متصل تقسیم می‌شوند که به طور بالقوه با کاراکترهای جداگانه مطابقت دارند.

۲. تجزیه و تحلیل کانتور: تجزیه و تحلیل کانتور بر روی اجزای متصل انجام می‌شود تا اطلاعات مرزی آن‌ها استخراج شود. خطوط می‌توانند نشانه‌های ارزشمندی در مورد شکل و مرزهای کاراکتر ارائه دهند. تکنیک‌هایی مانند دنبال کردن کانتور، هموارسازی کانتور یا تقریب کانتور معمولاً در تقسیم بندی کاراکترها استفاده می‌شود.

۳. استخراج جعبه مرزی: جعبه‌های مرزی حول اجزای متصل جداگانه محاسبه می‌شوند تا کاراکترها را محصور کنند. این جعبه‌ها اطلاعات مکانی در مورد مکان و اندازه هر کاراکتر ارائه می‌دهند. استخراج جعبه مرزی می‌تواند شامل جعبه‌های محدود مستطیلی ساده یا اشکال مرزبندی پیچیده‌تر بر اساس خطوط کاراکتر باشد.

۴. اکتشافی و قوانین: روش‌های اکتشافی و مبتنی بر قوانین اغلب برای اصلاح نتایج تقسیم بندی کاراکترها استفاده می‌شود. این تکنیک‌ها از دانش دامنه و قوانین خاص برای اصلاح یا تصحیح تقسیم بندی اولیه استفاده می‌کنند. به عنوان مثال، قوانین را می‌توان بر اساس عرض خط کاراکتر، نسبت ارتفاع به عرض، یا روابط فضایی بین کاراکترهای مجاور تعریف کرد.

۵. روش‌های یادگیری ماشین: تکنیک‌های یادگیری ماشین، به‌ویژه الگوریتم‌های یادگیری تحت نظارت، می‌توانند برای آموزش مدل‌هایی برای تقسیم‌بندی کاراکترها استفاده شوند. این مدل‌ها از داده‌های کاراکتر برچسب‌گذاری شده یاد می‌گیرند و می‌توانند مرزهای کاراکتر را بر اساس ویژگی‌های مختلف یا نمایش‌های تصویری به‌طور دقیق پیش‌بینی کنند. الگوریتم‌های رایج یادگیری ماشینی که برای تقسیم‌بندی کاراکتر استفاده می‌شوند شامل ماشین‌های بردار پشتیبان (SVM)، جنگل‌های تصادفی یا شبکه‌های عصبی کانولوشنال (CNN) هستند.

۶. پس پردازش: تکنیک‌های پس پردازش را می‌توان برای اصلاح بیشتر نتایج تقسیم بندی کاراکترها به کار برد. این ممکن است شامل ادغام یا تقسیم بخش‌های کاراکتر مجاور بر اساس معیارهای خاص، صاف کردن خطوط کاراکترها، یا اعمال عملیات مورفولوژیکی برای اصلاح اشکال کاراکتر باشد.

توجه به این نکته مهم است که فرآیند تقسیم کاراکتر بسته به سیستم (OCR) خاص، زبان، نوع فونت و ویژگی‌های تصویر می‌تواند متفاوت باشد (Hussain et al, 2015؛ Gonzalez and Woods, 2017؛ da Silva and Mendonca).

## استخراج ویژگی

استخراج ویژگی یک مرحله مهم در تشخیص کاراکتر نوری است که شامل استخراج ویژگی‌های مرتبط از کاراکترهای تقسیم‌بندی شده یا مناطق متنی است. این ویژگی‌ها به عنوان نمایش‌های متمایزی عمل می‌کنند که ویژگی‌های متمایز هر کاراکتر را نشان می‌دهند و تشخیص دقیق را تسهیل می‌کنند. در اینجا توضیح مختصری از فرآیند استخراج ویژگی در OCR آورده شده است:

۱. ویژگی‌های از پیش تعریف شده: ویژگی‌های از پیش تعریف شده، نمایش‌های دست ساز هستند که بر اساس دانش قبلی از ساختار و ویژگی‌های کاراکتر طراحی شده‌اند. این ویژگی‌ها اغلب از شکل، کانتور یا سایر ویژگی‌های هندسی کاراکتر به دست می‌آیند. ویژگی‌های از پیش تعریف شده رایج عبارتند از:

- هیستوگرام گرادیان‌های جهت‌دار (HOG): این ویژگی توزیع جهت‌گیری‌های گرادیان را در یک تصویر کاراکتر توصیف می‌کند و اطلاعات مربوط به شکل را می‌گیرد.
- تبدیل ویژگی ثابت در مقیاس (SIFT): ویژگی‌های محلی متمایز را از شخصیت‌ها استخراج می‌کند، که در برابر تغییرات مقیاس، چرخش یا شرایط نوری قوی هستند.
- لحظه‌های هو: لحظه‌های هو مجموعه‌ای از هفت لحظه ثابت هستند که ویژگی‌های شکل یک کاراکتر را مشخص می‌کنند.

۲. ویژگی‌های آماری: ویژگی‌های آماری از ویژگی‌های آماری مناطق کاراکتر یا شدت پیکسل‌ها به دست می‌آیند. این ویژگی‌ها می‌توانند اطلاعات بافتی و تغییرات درون کاراکترها را ثبت کنند. ویژگی‌های آماری رایج عبارتند از:

- میانگین و واریانس: میانگین و واریانس شدت پیکسل‌ها در یک منطقه کاراکتر بینش‌هایی را در مورد روشنایی و کنتراست کلی آن ارائه می‌دهد.
- ماتریس همزمانی: ماتریس همزمانی رابطه فضایی بین شدت پیکسل در یک کاراکتر را توصیف می‌کند و اطلاعات مربوط به بافت را می‌گیرد.
- رمزگذاری طول اجرا (RLE): ویژگی‌ها را بر اساس طول اجرای پیکسل‌های متوالی در یک کاراکتر استخراج می‌کند که می‌تواند نشان دهنده ضربات کاراکتر باشد.

۳. ویژگی‌های یادگیری عمیق: با ظهور یادگیری عمیق، ویژگی‌ها را می‌توان به طور خودکار از داده‌ها با استفاده از شبکه‌های عصبی عمیق یاد گرفت. شبکه‌های عصبی کانولوشنال (CNN) به ویژه در استخراج ویژگی‌های سلسله مراتبی از تصاویر کاراکتر موثر هستند. ویژگی‌ها را می‌توان از لایه‌های میانی شبکه به دست آورد یا به صورت سرتاسری در یک چارچوب شناسایی یکپارچه آموخت.

۴. رویکردهای ترکیبی: رویکردهای ترکیبی ویژگی‌های از پیش تعریف شده و ویژگی‌های آموخته شده را ترکیب می‌کنند تا از نمایش‌های دست ساز و داده محور بهره مند شوند. این رویکردها از نقاط قوت ویژگی‌های از پیش تعریف شده در گرفتن ویژگی‌های کاراکتری خاص استفاده می‌کنند و در عین حال از قدرت یادگیری عمیق برای یادگیری ویژگی‌های پیچیده‌تر و متمایزتر استفاده می‌کنند.



۵. عادی سازی ویژگی ها: پس از استخراج ویژگی ها، تکنیک های عادی سازی ممکن است برای افزایش قدرت تمایز و بهبود استحکام ویژگی ها اعمال شود. نرمال سازی می تواند شامل تکنیک هایی مانند عادی سازی میانگین، استاندارد سازی، یا مقیاس بندی به یک محدوده خاص باشد.

توجه به این نکته مهم است که انتخاب ویژگی ها به سیستم OCR خاص، مجموعه کاراکترها، زبان، انواع فونت و الزامات برنامه بستگی دارد (Zhanget al. 2005؛ Ali and S. Mallaiah, 2021؛ Srivastava et al, 2022؛ Gonzalez and Woods, 2017).

### طبقه بندی کاراکتر

طبقه بندی کاراکتر یک گام اساسی در تشخیص کاراکترهای نوری (OCR) است که شامل تخصیص یک برجسب کاراکتر خاص به هر کاراکتر تقسیم بندی شده و استخراج شده توسط ویژگی است. هدف شناسایی و طبقه بندی دقیق کاراکترها بر اساس ویژگی های استخراج شده آنهاست. در اینجا توضیح مختصری از فرآیند طبقه بندی کاراکترها در (OCR) آورده شده است:

۱. آماده سازی داده های آموزشی: فرآیند طبقه بندی کاراکترها معمولاً با تهیه داده های آموزشی آغاز می شود. این شامل جمع آوری مجموعه داده های برجسب گذاری شده از کاراکترها است، که در آن هر کاراکتر با برجسب حقیقت زمینی مربوطه مرتبط است. مجموعه داده باید طیف گسترده ای از کاراکترها، فونت ها، اندازه ها و تغییرات را پوشش دهد تا از عملکرد طبقه بندی قوی اطمینان حاصل شود.

۲. انتخاب ویژگی: در مرحله طبقه بندی کاراکتر، مجموعه ای از ویژگی های مربوطه از کاراکترهای تقسیم شده انتخاب یا استخراج می شود. این ویژگی ها می توانند ویژگی های دست ساز، ویژگی های آماری یا ویژگی های یادگیری عمیق، همانطور که در بخش قبلی در مورد استخراج ویژگی بحث شد، از پیش تعریف شده باشند.

۳. آموزش طبقه بندی کننده: پس از انتخاب ویژگی ها، یک مدل طبقه بندی با استفاده از یادگیری ماشین یا تکنیک های یادگیری عمیق آموزش داده می شود. الگوریتم های طبقه بندی مختلفی را می توان مورد استفاده قرار داد، از جمله:

- ماشین های بردار پشتیبانی (SVM): SVM ها طبقه بندی کننده های محبوبی هستند که هدفشان یافتن بهترین هایپرپلن جداکننده کلاس های مختلف در فضای ویژگی است.
- جنگل های تصادفی: جنگل های تصادفی مجموعه ای از درخت های تصمیم را برای انجام طبقه بندی بر اساس ترکیبی از ویژگی ها می سازند.
- شبکه های عصبی کانولوشن (CNN): CNN ها مدل های یادگیری عمیق هستند که ویژگی های سلسله مراتبی را از تصاویر کاراکترها یاد می گیرند و طبقه بندی آنها به انتها را انجام می دهند.

۴. آموزش و اعتبارسنجی مدل: مدل طبقه بندی با استفاده از مجموعه داده آموزشی برجسب دار آموزش داده می شود. این شامل بهینه سازی پارامترهای مدل برای به حداقل رساندن خطای طبقه بندی است. سپس مدل آموزش دیده با استفاده از یک مجموعه داده اعتبار سنجی جداگانه برای ارزیابی عملکرد آن و اطمینان از تعمیم اعتبارسنجی می شود.



۵. تشخیص کاراکتر: در طی تشخیص کاراکتر، طبقه‌بندی‌کننده آموزش‌دیده به ویژگی‌های استخراج‌شده کاراکترهای دیده نشده اعمال می‌شود. طبقه‌بندی‌کننده بر اساس الگوهای آموخته شده و اطلاعات متمایز به هر کاراکتر یک برچسب اختصاص می‌دهد. سپس کاراکترهای شناسایی شده به نمایش متن مربوطه خود نگاشت می‌شوند.

۶. پس پردازش: در برخی موارد، تکنیک‌های پس پردازش برای اصلاح نتایج تشخیص کاراکتر استفاده می‌شود. این ممکن است شامل الگوریتم‌های تصحیح خطا، مدل‌های زبان یا تحلیل زمینه‌ای برای بهبود دقت تشخیص و رسیدگی به خطاهای احتمالی باشد. توجه به این نکته مهم است که انتخاب الگوریتم‌ها و تکنیک‌های طبقه‌بندی ممکن است به عواملی مانند پیچیدگی مجموعه کاراکترها، زبان، اندازه مجموعه داده‌ها و الزامات محاسباتی بستگی داشته باشد (Phangtriastu et al, 2017؛ Srivastava et al, 2022).

## پس پردازش

پس پردازش در تشخیص کاراکتر نوری به تکنیک‌هایی اشاره دارد که برای اصلاح نتایج تشخیص پس از طبقه‌بندی کاراکترها اعمال می‌شود. هدف از پس پردازش بهبود دقت و قابلیت اطمینان متن شناخته شده است. در اینجا توضیح مختصری از فرآیند پس از پردازش در (OCR) آورده شده است:

۱. تصحیح خطا: تکنیک‌های تصحیح خطا برای تصحیح کاراکترها یا کلمات طبقه‌بندی شده اشتباه اعمال می‌شود. این تکنیک‌ها می‌تواند شامل موارد زیر باشد:

- تصحیح مبتنی بر فرهنگ لغت: یک فرهنگ لغت یا فهرست کلمات برای بررسی وجود کلمه شناسایی شده در واژگان زبان استفاده می‌شود. اگر کلمه شناسایی شده یافت نشد، می‌توان آن را بر اساس نزدیک‌ترین کلمه منطبق در فرهنگ لغت جایگزین یا تصحیح کرد.
- مدل‌های زبان آماری: مدل‌های زبان آماری از بافت و ویژگی‌های آماری زبان برای تخمین احتمال توالی کلمات معین استفاده می‌کنند. این مدل‌ها می‌توانند به تصحیح خطاها بر اساس احتمال وجود توالی کلمات در متن اطراف کمک کنند.

۲. یکپارچه سازی مدل زبان: مدل‌های زبانی را می‌توان برای بهبود انسجام کلی و صحت دستوری متن شناخته شده به کار برد. مدل‌های زبان الگوهای آماری و ساختار یک زبان خاص را به تصویر می‌کشند و امکان پیش‌بینی دقیق‌تر توالی کلمات را فراهم می‌کنند.

۳. رمزگشایی پسین: تکنیک‌های رمزگشایی پسین از اطلاعات آماری از مدل‌های متن و زبان شناخته شده برای تصمیم‌گیری آگاهانه‌تر استفاده می‌کنند. این می‌تواند شامل استفاده از مدل‌های احتمالی، مانند مدل‌های مارکوف پنهان (HMMs) یا میدان‌های تصادفی شرطی (CRFs)، برای تخمین محتمل‌ترین توالی کلمات یا کاراکترها بر اساس خروجی تشخیص باشد.

۴. چک کردن املا: الگوریتم‌های املا را می‌توان برای تشخیص و تصحیح غلط‌های املائی در متن شناخته شده اعمال کرد. این الگوریتم‌ها کلمات شناسایی شده را با یک فرهنگ لغت مقایسه می‌کنند و اصلاحاتی را برای کلمات غلط املائی پیشنهاد می‌کنند.

۵. امتیازدهی اعتماد: امتیازدهی اعتماد، سطح اطمینان یا احتمال صحت را برای هر کاراکتر یا کلمه شناسایی شده اندازه گیری می کند. این می تواند برای شناسایی و پرچم گذاری نتایج تشخیص نامشخص یا کم اعتماد برای بررسی یا تأیید بیشتر استفاده شود. تکنیک های پس پردازش را می توان برای سیستم های (OCR)، زبان ها و الزامات برنامه های کاربردی خاص تنظیم کرد. انتخاب و ترکیب تکنیک ها به سطح مورد نظر از دقت و ویژگی های کار تشخیص بستگی دارد (Kissos and Dershowitz, 1994; Sun and Jelinek, 1994; Kissos and Dershowitz, 2016; Magdy and K. Darwish, 2006; Smith, 2007; 2017). در شکل ۲ نمونه ای از ماتریس سردرگمی<sup>۱</sup> برای یک کاراکتر نمایش داده شده است.

```
<wrongSegment string="ر">
  <correctSegment popularity="14">ز</correctSegment>
  <correctSegment popularity="11">ي</correctSegment>
  <correctSegment popularity="10">س</correctSegment>
  <correctSegment popularity="2">و</correctSegment>
  <correctSegment popularity="2">م</correctSegment>
  <correctSegment popularity="1">ه</correctSegment>
  <correctSegment popularity="1">ل</correctSegment>
  <correctSegment popularity="1">ا</correctSegment>
</wrongSegment>
```

شکل ۲: گزیده ای از ماتریس سردرگمی برای کاراکتر “ر” (Kissos and Dershowitz, 2017).

## خروجی متن

خروجی متن در تشخیص کاراکتر نوری (OCR) به مرحله نهایی فرآیند (OCR) اشاره دارد، جایی که کاراکترهای شناسایی شده برای پردازش یا ارائه بیشتر به یک قالب متنی تبدیل می شوند. فرآیند خروجی متن شامل تبدیل دنباله کاراکترهای شناسایی شده به یک سند متنی قابل خواندن و قابل استفاده است. در شکل ۳ نمونه ای از خروجی یک سیستم تشخیص کاراکتر نوری بر روی تصویری از یک یادداشت دست نویس نمایش داده شده است. در اینجا توضیح مختصری از فرآیند خروجی متن در (OCR) آورده شده است:

۱. یکپارچه سازی دنباله کاراکترها: کاراکترهای فردی که در مرحله طبقه بندی کاراکترها شناسایی می شوند، ترکیب می شوند تا دنباله ای را تشکیل دهند که نشان دهنده متن شناخته شده است. این دنباله می تواند شامل فاصله ها، علائم نگارشی و سایر کاراکترهای خاص باشد.

۲. تشخیص مرز کلمات: فرآیند خروجی متن شامل تعیین مرزهای بین کلمات در دنباله کاراکترهای شناسایی شده است. تشخیص مرز کلمات می تواند بر اساس نشانه های مختلفی مانند فاصله ها، علائم نقطه گذاری یا مدل های زبان آماری باشد.

<sup>1</sup> . confusion matrix.

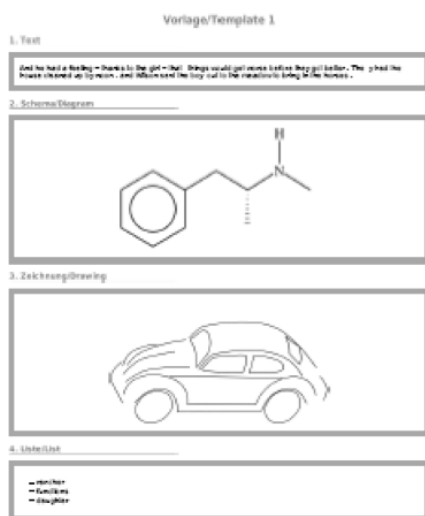
۳. زبان و قالب بندی متن: متن شناسایی شده ممکن است تحت پردازش اضافی قرار گیرد تا از نمایش و قالب بندی مناسب زبان اطمینان حاصل شود. این می تواند شامل وظایفی مانند:

- شناسایی زبان: تکنیک های شناسایی زبان را می توان برای تعیین زبان متن شناسایی شده به کار برد، که می تواند به پردازش یا تجزیه و تحلیل بعدی خاص زبان کمک کند.
- قالب بندی متن: وظایف قالب بندی شامل تنظیم طرح، فونت، سبک و سایر جنبه های بصری متن شناسایی شده برای مطابقت با سند اصلی یا فرمت خروجی دلخواه است.

۴. پس پردازش: همانطور که قبلا ذکر شد، تکنیک های پس پردازش، مانند تصحیح خطا و بررسی املا، می تواند برای اصلاح متن شناخته شده و بهبود دقت و خوانایی آن استفاده شود.

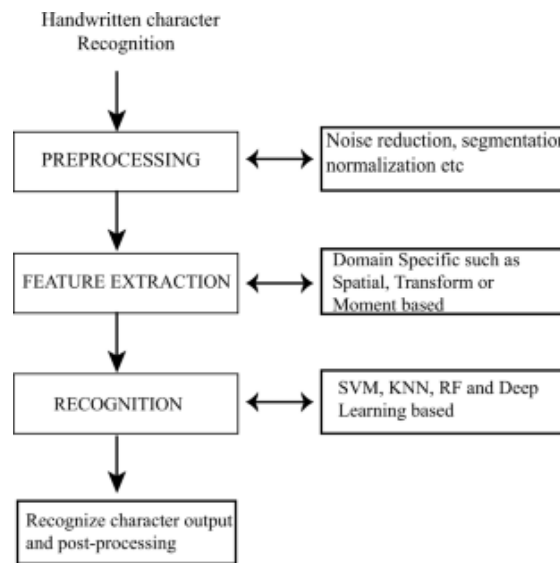
۵. فرمت های خروجی متن: مرحله نهایی شامل تبدیل متن شناسایی شده به یک فرمت خروجی خاص بر اساس الزامات برنامه است. فرمت های خروجی رایج عبارتند از: متن ساده (ASCII یا Unicode)، فرمت های متن غنی (RTF)، (HTML)، (XML)، یا فرمت های سند خاص (مانند PDF یا DOC).

توجه به این نکته ضروری است که فرآیند خروجی متن می تواند تحت تأثیر عواملی مانند کیفیت سیستم (OCR)، پیچیدگی زبان، طرح سند و وجود نویز یا مصنوعات در سند ورودی باشد (Inunganbi, 2023؛ Magdy and K. Darwish, 2006؛ Srivastava et al. 2022).



شکل ۳: خروجی یک سیستم (OCR) پس از پردازش تصویر یادداشت دست نویسی (Hussain et al, 2015).

بطور کلی همانطوری که در شکل ۴ نمایش داده شده است، می توان مراحل تشخیص کاراکتر نوری را به صورت خلاصه به مراحل پیش پردازش، استخراج ویژگی، تشخیص کاراکتر و پس پردازش تقسیم کرد.



شکل ۴: فرآیند تشخیص کارکتر نوری در یک تصویر (Inunganbi, 2023).

### شبکه‌های عصبی عمیق و کاربرد آن در تشخیص کاراکترهای نوری

شبکه‌های عصبی عمیق (DNN) پیشرفت‌های چشم‌گیری در حوزه هوش مصنوعی و یادگیری ماشین داشته‌اند. موفقیت آن‌ها را می‌توان به ظرفیت آن‌ها در به دست آوردن مستقل بازنمایی‌های سلسله مراتبی از داده‌های خام، تسهیل استخراج ویژگی‌ها و الگوهای پیچیده‌ای که ممکن است طراحان انسانی را فراری دهند، نسبت داد. این قابلیت DNN‌ها را قادر می‌سازد تا به طور موثر داده‌های با ابعاد بالا و بدون ساختار مانند تصاویر، صدا و متن را پردازش کنند. به این ترتیب، یک مرور مختصر از (DNN) در زیر ارائه شده است (I. G. and Y. B. and A, 2016).

### شبکه‌های عصبی عمیق

شبکه‌های عصبی عمیق دسته‌ای از مدل‌های یادگیری ماشینی هستند که از ساختار و عملکرد مغز انسان الهام گرفته شده‌اند. آن‌ها از لایه‌های متعددی از نورون‌های مصنوعی به هم پیوسته به نام گره‌ها یا واحدها تشکیل شده‌اند که به صورت سلسله مراتبی سازماندهی شده‌اند. هر نورون سیگنال‌های ورودی را دریافت می‌کند، یک تبدیل ریاضی برای آن‌ها اعمال می‌کند و خروجی تولید می‌کند که به لایه بعدی نورون‌ها ارسال می‌شود. از طریق فرآیندی به نام آموزش، DNN‌ها یاد می‌گیرند که الگوها را تشخیص دهند و از داده‌های ورودی پیش‌بینی کنند.

بلوک اصلی یک شبکه عصبی عمیق، نورون مصنوعی است که به عنوان پرسپترون نیز شناخته می‌شود. چندین ورودی می‌گیرد، برای هر ورودی وزن اعمال می‌کند و با استفاده از یک تابع فعال‌سازی برای تولید خروجی آن‌ها را ترکیب می‌کند. تابع فعال‌سازی غیرخطی‌ها را به شبکه معرفی می‌کند و آن را قادر می‌سازد تا روابط پیچیده بین ورودی‌ها و خروجی‌ها را بیاموزد.

DNN‌ها می‌توانند از لایه‌های مختلفی مانند لایه‌های ورودی، پنهان و خروجی تشکیل شوند. لایه ورودی داده‌های خام را دریافت می‌کند و در لایه‌های پنهان حرکت می‌کند. لایه‌های مخفی داده‌ها را پردازش کرده و ویژگی‌های سطح بالاتر را استخراج می‌کنند.

لایه خروجی خروجی نهایی را تولید می کند، مانند برچسب های طبقه بندی یا پیش بینی های عددی (I. G. and Y. B. and A, 2016)؛ Le Cun et al, 2015). در اینجا، دلایل موفقیت اخیر شبکه های عصبی عمیق در هوش مصنوعی را بررسی می کنیم. این موفقیت را می توان ناشی از عوامل زیر دانست:

۱. یادگیری بازنمایی: DNN ها می توانند نمایش های سلسله مراتبی داده ها را با تبدیل مکرر ورودی در هر لایه بیاموزند. هر لایه از نورون ها به طور فزاینده ای ویژگی های انتزاعی و پیچیده را جذب می کند و به شبکه اجازه می دهد تا روابط پیچیده را در داده ها مدل کند. این توانایی برای یادگیری خودکار بازنمایی ها، DNN ها را قادر می سازد تا مجموعه داده های متنوع و چالش برانگیز را به طور موثر مدیریت کنند.

۲. مقیاس پذیری: شبکه های عصبی عمیق می توانند برای رسیدگی به حجم زیادی از داده ها و وظایف پیچیده، مقیاس شوند. با در دسترس بودن منابع محاسباتی قدرتمند و قابلیت های پردازش موازی، DNN ها را می توان بر روی مجموعه های داده عظیم آموزش داد و در نتیجه مدل هایی را به خوبی به نمونه های دیده نشده تعمیم داد. علاوه بر این، عمق شبکه ها به مدل های گویاتر اجازه می دهد و آن ها را قادر می سازد تا مرزهای پیچیده تصمیم گیری را بیاموزند.

۳. در دسترس بودن مجموعه داده های بزرگ برچسب گذاری شده: موفقیت DNN ها با در دسترس بودن مجموعه داده های بزرگ حاشیه نویسی تسهیل شده است. یادگیری نظارت شده، جایی که DNN ها از نمونه های برچسب گذاری شده یاد می گیرند، نیروی محرکه ای در آموزش مدل های دقیق بوده است. با داده های برچسب دار بیشتر، DNN ها می توانند نمایش های ظریف تری را بیاموزند و بهتر به نمونه های جدید تعمیم دهند.

۴. پیشرفت در الگوریتم های بهینه سازی: آموزش شبکه های عصبی عمیق به دلیل تعداد زیاد پارامترها و ماهیت غیر محدب مسئله بهینه سازی یک کار چالش برانگیز است. با این حال، پیشرفت ها در الگوریتم های بهینه سازی، مانند نزول گرادیان تصادفی (SGD) و انواع آن، همراه با تکنیک هایی مانند نرمال سازی دسته ای و نرخ های یادگیری تطبیقی، آموزش مدل های عمیق را به طور موثر امکان پذیر کرده اند.

۵. افزایش قدرت محاسباتی: شبکه های عصبی عمیق به منابع محاسباتی قابل توجهی برای آموزش و استنتاج نیاز دارند. در دسترس بودن پردازنده های گرافیکی قدرتمند و سخت افزار تخصصی، همراه با چارچوب های محاسباتی موازی مانند (Tensor Flow) و (PyTorch)، آموزش و استقرار شبکه های عصبی عمیق را تسریع کرده است (LeCun et al, 2015).

### شبکه های عصبی عمیق در (OCR)

شبکه های عصبی عمیق سیستم های تشخیص کاراکتر نوری (OCR) را به طور قابل توجهی پیشرفته کرده اند که امکان تشخیص دقیق و کارآمد متن از روی تصاویر را فراهم می کند. (OCR) نقش مهمی در برنامه های مختلف از جمله دیجیتالی کردن اسناد، استخراج متن از تصاویر و ورود خودکار داده ها دارد. شبکه های عصبی عمیق در بهبود عملکرد (OCR) با استفاده از توانایی خود در یادگیری نمایش های سلسله مراتبی از داده های خام بسیار مؤثر بوده اند. شبکه های عصبی عمیق در تشخیص کاراکتر نوری از ترکیبی از شبکه های عصبی کانولوشن (CNN) و شبکه های عصبی بازگشتی (RNN) بهره می برد. این روش شامل استخراج ویژگی ها از تصاویر با استفاده از (CNN) و سپس استفاده از RNN برای رمزگشایی این ویژگی ها و تولید خروجی های متنی است. ترکیبی از (CNN) و

(RNN) امکان عملکرد (OCR) قوی و دقیق را فراهم می کند، چرا که شبکه های عصبی عمیق توسعه سیستم های (OCR) سرتاسر را تسهیل کرده اند، جایی که کل خط لوله (OCR)، از تشخیص مکان متن تا تشخیص متن، در یک مدل واحد ادغام شده است. سیستم های (OCR) انتها به انتها از ویژگی های سلسله مراتبی و آگاه از زمینه ای که توسط شبکه های عصبی عمیق آموخته شده اند برای پردازش مستقیم تصاویر و خروجی متن شناسایی شده استفاده می کنند. این خط لوله OCR را ساده می کند، انتشار خطا را کاهش می دهد و دقت کلی را بهبود می بخشد. در اینجا توضیح مفصلی از مراحل کار ارائه شده است:

۱. استخراج ویژگی: شبکه های عصبی عمیق، به ویژه شبکه های عصبی کانولوشن (CNN) در استخراج ویژگی از تصاویر برتری دارند. در OCR، شبکه های عصبی عمیق بر روی مجموعه داده های بزرگی از تصاویر متن برچسب گذاری شده آموزش داده می شوند تا ویژگی های متمایزکننده ای را بیاموزند که کاراکترها و الگوهای متن متفاوت را متمایز می کنند. این شبکه ها به طور خودکار یاد می گیرند که نشانه های بصری سطح پایین، مانند لبه ها، گوشه ها، و بافت ها، و همچنین ویژگی های معنایی سطح بالا، مانند شکل ها و ساختارهای کاراکترها را ثبت کنند. معماری (CNN) معمولاً از چندین لایه کانولوشن و به دنبال آن لایه های ادغام تشکیل شده است. لایه های کانولوشن فیلترهایی را روی تصویر ورودی اعمال می کنند و الگوهای محلی و روابط فضایی را ثبت می کنند. لایه های ادغام شده نقشه های ویژگی را پایین می آورند و ابعاد فضایی آن ها را کاهش می دهند و در عین حال برجسته ترین اطلاعات را حفظ می کنند. ساختار سلسله مراتبی CNN ها با افزایش عمق شبکه، امکان استخراج ویژگی های انتزاعی فزاینده را فراهم می کند.

۲. محلی سازی متن: پس از استخراج ویژگی ها، تکنیک های پیشنهاد منطقه برای شناسایی مناطق متنی بالقوه در تصویر استفاده می شود. این تکنیک ها، مانند جستجوی انتخابی یا رویکردهای پنجره کشویی، مناطق نامزدی را پیشنهاد می کنند که احتمالاً حاوی متن هستند. شبکه های عصبی عمیق را می توان برای محلی سازی متن استفاده کرد که شامل شناسایی و بومی سازی مناطق در یک تصویر است که حاوی متن است. با استفاده از CNN ها، سیستم های OCR می توانند مناطق متن را با تجزیه و تحلیل ویژگی های آموخته شده و شناسایی الگوهای نشان دهنده حضور متن شناسایی کنند. معماری های مختلف (CNN)، مانند (Faster R-CNN)، (SSD) و (YOLO)، برای محلی سازی دقیق و کارآمد متن استفاده شده اند.

۳. تشخیص متن: شبکه های عصبی عمیق به طور گسترده برای تشخیص متن استفاده می شوند، جایی که هدف تبدیل نواحی متن شناسایی شده در تصاویر به نمایش های متنی واقعی است. شبکه های عصبی بازگشتی (RNN) معمولاً همراه با (CNN) برای رمزگشایی ویژگی های استخراج شده از مناطق متنی و تولید خروجی های متنی مربوطه استفاده می شوند. RNN ها مانند شبکه های حافظه کوتاه مدت (LSTM)، مدل سازی وابستگی های متوالی و اطلاعات متنی را امکان پذیر می کنند و تشخیص دقیق متن را در سیستم های (OCR) تسهیل می کنند. برای آموزش (RNN) برای تشخیص متن، تابع  $loss$  طبقه بندی زمانی اتصال گرا (CTC) اغلب استفاده می شود. CTC به RNN اجازه می دهد تا مستقیماً از جفت های ورودی-خروجی بیاموزد، بدون اینکه نیاز به هم ترازای صریح بین دنباله های ورودی و برچسب های هدف داشته باشد. این آن را برای کارهای (OCR) که مکان و طول متن ممکن است متفاوت باشد، مناسب می کند.

۴. رمزگشایی و تولید خروجی: هنگامی که (RNN) نواحی متن را پردازش کرد، از یک الگوریتم رمزگشایی برای تبدیل دنباله خروجی کاراکترها به متن شناسایی شده نهایی استفاده می شود. تکنیک های رمزگشایی شامل جستجوی پرتو یا رمزگشایی حریصانه است

که محتمل ترین دنباله کاراکترها را بر اساس احتمالات خروجی (RNN) انتخاب می کند. همچنین می توان از مدل های زبان برای بهبود دقت متن رمزگشایی شده استفاده کرد (Jaderberg et al, 2015؛ Graves et al, 2006؛ I. G. and Y. B. and A, 2016).

## کاربردها

فناوری تشخیص کاراکتر نوری را می توان برای تبدیل متن چاپی یا دست نویس به فرمت دیجیتال استفاده کرد که امکان ثبت کارآمد و دقیق اطلاعات را در سناریوهای مختلف دنیای واقعی فراهم می کند. در اینجا چند نمونه آورده شده است:

۱. دیجیتالی کردن اسناد: از (OCR) می توان برای اسکن و تبدیل اسناد فیزیکی مانند فاکتورها، رسیدها، قراردادهای و فرم ها به فایل های دیجیتال قابل ویرایش و جستجو استفاده کرد. این امکان بایگانی، بازیابی و تجزیه و تحلیل آسان اطلاعات موجود در این اسناد را فراهم می کند.

۲. اسکن کارت ویزیت: (OCR) را می توان در برنامه های تلفن همراه یا اسکنرها برای گرفتن اطلاعات تماس از کارت های ویزیت استفاده کرد. این فناوری نام، شماره تلفن، آدرس ایمیل و سایر جزئیات مربوطه را استخراج می کند، که سپس می تواند به طور خودکار در یک دفترچه آدرس دیجیتال یا سیستم (CRM) ذخیره شود.

۳. استخراج داده ها از فاکتورها: (OCR) می تواند اطلاعات کلیدی را از فاکتورها استخراج کند، مانند شماره فاکتور، جزئیات فروشنده، توضیحات اقلام، مقادیر و قیمت ها. این داده ها را می توان به طور خودکار به سیستم های حسابداری یا برنامه ریزی منابع سازمانی (ERP) وارد کرد و نیاز به ورود دستی داده ها را از بین برد و خطاها را کاهش داد.

۴. اسکن پاسپورت و شناسه: (OCR) را می توان در اسکنرهای گذرنامه یا برنامه های تلفن همراه برای گرفتن داده ها از گذرنامه ها، گواهینامه های رانندگی و سایر اسناد شناسایی استفاده کرد. این فناوری می تواند نام، تاریخ تولد، شماره پاسپورت و سایر اطلاعات مرتبط را استخراج کند و فرآیندهای تأیید هویت سریع تر و دقیق تر را تسهیل کند.

۵. پردازش رسید: (OCR) می تواند اطلاعاتی را از رسیدهای خرده فروشی استخراج کند، از جمله نام فروشگاه، تاریخ معامله، اقلام خریداری شده، قیمت ها و جزئیات پرداخت. این داده ها را می توان برای ردیابی هزینه، بازپرداخت یا ادغام با نرم افزار حسابداری استفاده کرد.

۶. دیجیتالی کردن کتاب: از (OCR) می توان برای تبدیل کتاب های چاپی یا نسخه های خطی به متون دیجیتال قابل جستجو استفاده کرد. با اسکن و پردازش صفحات، نرم افزار (OCR) می تواند کاراکترها را تشخیص دهد و نسخه های الکترونیکی قابل ویرایش ایجاد کند و دسترسی و جستجوی اطلاعات خاص در متون را آسان تر کند.

۷. تبدیل یادداشت دست نویس: فناوری (OCR) تا حدی پیشرفت کرده است که می تواند متن دست نویس را تشخیص داده و به فرمت دیجیتال تبدیل کند. این می تواند برای دیجیتالی کردن یادداشت ها، نامه ها یا فرم های دست نویس مفید باشد و امکان ذخیره سازی، ویرایش و بازیابی آسان اطلاعات را فراهم کند.



۸. تشخیص خودکار پلاک خودرو: (OCR) را می‌توان در سیستم‌هایی استفاده کرد که تصاویر پلاک‌ها را ضبط و پردازش می‌کنند. این فناوری نویسه‌های الفبایی روی صفحات را تشخیص می‌دهد و امکان شناسایی و ردیابی خودکار وسایل نقلیه را در برنامه‌هایی مانند جمع‌آوری عوارض، مدیریت پارکینگ و اجرای قانون فراهم می‌کند.

۹. تشخیص تقلب: (OCR) را می‌توان در سیستم‌های تشخیص تقلب برای تجزیه و تحلیل و تأیید صحت اسناد، مانند گذرنامه، گواهینامه رانندگی، یا کارت شناسایی به کار برد. با مقایسه متن استخراج شده با متن اصلی، به شناسایی اسناد جعلی یا دستکاری شده کمک می‌کند.

۱۰. قابلیت دسترسی برای افراد دارای اختلال بینایی: OCR نقش مهمی در دسترسی به مواد چاپی برای افراد دارای اختلالات بینایی ایفا می‌کند. با تبدیل متن چاپ شده به گفتار یا خط بریل، OCR به افراد کم بینا امکان دسترسی و درک محتوای نوشته شده را می‌دهد. این‌ها تنها چند نمونه از نحوه استفاده از (OCR) برای ضبط اطلاعات از دنیای فیزیکی و تبدیل آن به فرمت دیجیتال است که امکان مدیریت، تجزیه و تحلیل و ادغام آسان‌تر با سایر سیستم‌ها را فراهم می‌کند.

### بحث و نتیجه‌گیری

فناوری تشخیص کاراکتر نوری (OCR) با تسهیل پردازش کارآمد و دقیق اطلاعات متنی، نقش مهمی در عملیات نظامی ایفا می‌کند. قابلیت‌های پردازش داده‌ها را افزایش می‌دهد و منجر به بازیابی سریع‌تر اطلاعات، افزایش آگاهی موقعیتی و ساده‌سازی فرآیندهای عملیاتی می‌شود. (OCR) با تبدیل داده‌های چاپی یا دست‌نویس به فرمت‌های دیجیتال، پشتیبانی از اطلاعات نظامی، نظارت و تلاش‌های تصمیم‌گیری به این مهم دست می‌یابد. در اینجا برخی از کاربردهای کلیدی (OCR) در ارتش آورده شده است: تجزیه و تحلیل اسناد: (OCR) برای استخراج اطلاعات از اسناد نظامی مختلف مانند نقشه‌ها، دستورالعمل‌ها، گزارش‌های اطلاعاتی و مواد ضبط شده دشمن استفاده می‌شود. این تبدیل متن به فرم دیجیتال، بازیابی، تجزیه و تحلیل و ترجمه سریع داده‌های مهم را امکان پذیر می‌کند.

تشخیص خودکار پلاک خودرو (ALPR): سیستم‌های (ALPR) نظامی از (OCR) برای شناسایی و نظارت بر وسایل نقلیه استفاده می‌کنند. (OCR) با تجزیه و تحلیل تصاویر یا ویدیوهای دوربین‌های نظارتی یا هواپیماهای بدون سرنشین، کاراکترهای پلاک خودرو را به داده‌های قابل خواندن توسط ماشین تشخیص داده و تبدیل می‌کند. این به ردیابی حرکات وسیله نقلیه، ردیابی وسایل نقلیه مشکوک/ تحت تعقیب، و تقویت امنیت محیطی کمک می‌کند.

اطلاعات و شناسایی: (OCR) به پردازش و تجزیه و تحلیل حجم زیادی از داده‌های متنی جمع‌آوری شده از ارتباطات، اسناد و رسانه‌های اجتماعی رهگیری شده برای مأموریت‌های اطلاعاتی و شناسایی کمک می‌کند. با تبدیل متن به فرمت‌های دیجیتال قابل

جستجو، (OCR) به تحلیل گران کمک می کند تا اطلاعات مربوطه را به سرعت استخراج کنند، الگوها را شناسایی کنند و بینش هایی را برای تصمیم گیری به دست آورند.

ترجمه اسناد: (OCR) ترجمه خودکار اسناد را در عملیات نظامی به چند زبانه تسهیل می کند. با تشخیص و تبدیل متن از اسناد زبان خارجی به فرمت های دیجیتال، نرم افزار (OCR) ترجمه، تجزیه و تحلیل و کمک به جمع آوری اطلاعات و آگاهی از موقعیت را تسریع می کند.

مدیریت پرسنل و هویت: (OCR) با پرسنل نظامی و سیستم های مدیریت هویت برای پردازش مدارک شناسایی مانند گذرنامه، شناسه های نظامی و ویزا یکپارچه شده است. داده های مربوطه را استخراج و تأیید می کند، فرآیندهای احراز هویت و کنترل دسترسی را ساده می کند.

تجزیه و تحلیل پزشکی قانونی: (OCR) در تجزیه و تحلیل پزشکی قانونی برای تفسیر متن از شواهد صحنه جرم مانند یادداشت های دست نویس، نامه ها یا اسناد استفاده می شود. با تبدیل متن به فرمت های دیجیتال، نرم افزار (OCR) به رمزگشایی و تجزیه و تحلیل محتوا کمک می کند و به طور بالقوه بینش های ارزشمندی را برای تحقیق ارائه می دهد.

مدیریت اطلاعات: فناوری (OCR) سوابق مبتنی بر کاغذ را در سیستم های مدیریت اطلاعات نظامی شامل پرونده های پرسنل، اسناد لجستیک و گزارش های عملیاتی دیجیتالی و پردازش می کند. این تبدیل ذخیره، بازیابی و تجزیه و تحلیل کارآمد اطلاعات را امکان پذیر می کند و کارایی عملیاتی و تصمیم گیری را افزایش می دهد. اذعان به این نکته ضروری است که کاربرد (OCR) در ارتش می تواند



بر اساس الزامات کشور، شاخه و مأموریت متفاوت باشد. این فناوری ممکن است در سیستم‌های نرم افزاری نظامی موجود ادغام شود یا بسته به سناریوی خاص به عنوان برنامه های کاربردی مستقل مورد استفاده قرار گیرد.

#### منابع

- F. Hussain *et al.*, "Optical Character Recognition: An Illustrated Guide to the," *J. Vis. Impair. Blind.*, vol. 84, no. 10, pp. 507–509, 2009.
- R. Plamondon and S. N. Srihari, "On-line and Off-line Handwriting Recognition," *IEEE Trans. Pattern Anal. Mach.*

- Intell.*, vol. 22, no. 1, pp. 63–84, 2000, [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=824821>
- C. C. Tappert, C. Y. Suen, and T. Wakahara, “The State of the Art in On-Line Handwriting Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 8, pp. 787–808, 1990, doi: 10.1109/34.57669.
- C. Clausner, A. Antonacopoulos, and S. Pletschacher, “ICDAR2017 Competition on Recognition of Documents with Complex Layouts - RDCL2017,” *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 1, pp. 1404–1410, 2017, doi: 10.1109/ICDAR.2017.229.
- R. M. McElhaney, “Algorithms for graphics and image processing,” *Proc. IEEE*, vol. 71, no. 9, pp. 1116–1117, 2008, doi: 10.1109/proc.1983.12734.
- S. C. F. Lin *et al.*, “Image enhancement using the averaging histogram equalization (AVHEQ) approach for contrast improvement and brightness preservation,” *Comput. Electr. Eng.*, vol. 46, pp. 356–370, 2015, doi: 10.1016/j.compeleceng.2015.06.001.
- W. Bieniecki, S. Grabowski, and W. Rozenberg, “Image preprocessing for improving OCR accuracy,” *Proceeding 3rd Int. Conf. Young Sci. "Perspective Technol. Methods MEMS Des. MEMSTECH 2007*, pp. 75–80, 2007, doi: 10.1109/MEMSTECH.2007.4283429.
- Y. Alginahi, “Preprocessing Techniques in Character Recognition,” *Character Recognit.*, 2010, doi: 10.5772/9776.
- R. Gonzalez and R. Woods, *Digital Image Processing 4th Edition*. Pearson; 4th edition (March 20, 2017), 2017.
- J. CANNY, “A Computational Approach to Edge Detection,” *Readings Comput. Vis.*, pp. 184–203, 1987, doi: 10.1016/b978-0-08-051581-6.50024-6.
- M. Fawzi *et al.*, “Rectification of camera captured document images for camera-based OCR technology,” *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 2015-Novem, pp. 1226–1230, 2015, doi: 10.1109/ICDAR.2015.7333959.
- R. Hussain, A. Raza, I. Siddiqi, K. Khurshid, and C. Djeddi, “A comprehensive survey of handwritten document benchmarks: structure, usage and evaluation,” *Eurasip J. Image Video Process.*, vol. 2015, no. 1, pp. 1–24, 2015, doi: 10.1186/s13640-015-0102-5.
- E. A. B. da Silva and G. V. Mendonca, “Digital Image Processing,” *Electr. Eng. Handb.*, pp. 891–910, 2004, doi: 10.1016/B978-012170960-0/50064-5.
- S. Srivastava, A. Verma, and S. Sharma, “Optical Character Recognition Techniques: A Review,” *2022 IEEE Int. Students' Conf. Electr. Electron. Comput. Sci. SCEECS 2022*, 2022, doi: 10.1109/SCEECS54111.2022.9740911.
- A. A. A. Ali and S. Mallaiah, “Intelligent handwritten recognition using hybrid CNN architectures based-SVM classifier with dropout,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3294–3300, 2022, doi: 10.1016/j.jksuci.2021.01.012.
- P. Zhang, T. D. Bui, and C. Y. Suen, “Hybrid feature extraction and feature selection for improving recognition accuracy of handwritten numerals,” *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 2005, pp. 136–140, 2005, doi: 10.1109/ICDAR.2005.129.
- M. R. Phangtriastu, J. Harefa, and D. F. Tanoto, “Comparison between Neural Network and Support Vector Machine in Optical Character Recognition,” *Procedia Comput. Sci.*, vol. 116, pp. 351–357, 2017, doi: 10.1016/j.procs.2017.10.061.
- D. X. Sun and F. Jelinek, *Statistical Methods for Speech Recognition*, vol. 94, no. 446. 1999. doi: 10.2307/2670189.
- I. Kissos and N. Dershowitz, “Image and text correction using language models,” *1st IEEE Int. Work. Arab. Scr. Anal.*

- Recognition, ASAR 2017*, pp. 158–162, 2017, doi: 10.1109/ASAR.2017.8067779.
- R. Smith, “An overview of the tesseract OCR engine,” *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 2, pp. 629–633, 2007, doi: 10.1109/ICDAR.2007.4376991.
- W. Magdy and K. Darwish, “Arabic OCR error correction using character segment correction, language modeling, and shallow morphology,” *COLING/ACL 2006 - EMNLP 2006 2006 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, pp. 408–414, 2006, doi: 10.3115/1610075.1610132.
- I. Kissos and N. Dershowitz, “OCR Error Correction Using Character Correction and Feature-Based Word Classification,” *Proc. - 12th IAPR Int. Work. Doc. Anal. Syst. DAS 2016*, pp. 198–203, 2016, doi: 10.1109/DAS.2016.44.
- S. Inunganbi, “A systematic review on handwritten document analysis and recognition,” *Multimed. Tools Appl.*, Jun. 2023, doi: 10.1007/s11042-023-15326-9.
- I. G. and Y. B. and A. Courville, *Deep learning*, vol. 29, no. 7553. 2016. [Online]. Available: <http://deeplearning.net/>
- Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, “Connectionist temporal classification,” in *Proceedings of the 23rd international conference on Machine learning - ICML '06*, 2006, pp. 369–376. doi: 10.1145/1143844.1143891.
- M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, “Deep structured output learning for unconstrained text recognition,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, 2015.



## Recognition of optical characters and its application in military organizations

**Behshad Panjezadeh**

PhD student in Sports Biomechanics, Faculty of Physical  
Education and Sports Sciences, Islamic Azad University,  
Central Tehran Branch, Tehran, Iran.

**Heydar Azadi Neko**

Master's degree in Research Science from Islamic Azad  
University, Kurdistan.

### Abstract

The purpose of this article is to examine optical character recognition (OCR) and its applications in the real world, particularly in military organizations. OCR technology enables the extraction of textual information from scanned documents, images, or other digital sources. It uses computer vision and machine learning to convert printed or handwritten text into editable and searchable data. The process involves scanning the input image, analyzing it, identifying characters or words, and converting them into machine-readable text. OCR has various applications such as document digitization, text extraction, data entry, translation, analysis, license plate recognition, fraud detection, information management, accessibility for the visually impaired, and document processing. Overall, OCR technology has transformed data mining and document processing, offering efficiency and accuracy in a wide range of applications. In conclusion, OCR technology is crucial in military operations for efficient and accurate processing of textual information, enhancing data processing capabilities, information retrieval speed, situational awareness, and operational simplicity. By converting printed or handwritten information into digital format, OCR supports military intelligence, surveillance, and decision-making operations.

**Keywords:** optical character recognition (OCR), army.