

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده مهندسی کامپیوتر

پایان نامه کارشناسی ارشد مهندسی هوش مصنوعی

# OCR متون چاپی فارسی به کمک یادگیری عمیق

نگارنده: مرضیه رحمتی

اساتید راهنما

دکتر منصور فاتح

دکتر محسن رضوانی

اساتید مشاور

دکتر علیرضا تجری

دکتر وحید ابوالقاسمی

شهریور ۹۸

ماحصل آموخته‌هایم را تقدیم می‌کنم به آنان که مهر آسمانی‌شان آرام‌بخش آلام زمینی‌ام است

به استوارترین تکیه‌گاهم، دستان پر مهر پدرم

به سبزترین نگاه زندگیم، چشمان سبز مادرم

که هرچه آموختم در کتب عشق شما آموختم و هرچه بگوختم قطره‌ای از دریای بی‌کران مهربانیان را پاس توانم بگویم .

امروز،ستی‌ام به امید شماست و فردا کلید باغ به‌ستم رضای شما

را آوردی کران سنگ ترا از این ارزان‌داشتم تا به خاک پایتان نثار کنم، باشد که حاصل تلاشم نسیم کوزه‌خوار مستقیمتان را

بزداید .

بوسه بر دستان پر مهرتان

اکنون که بیاری پروردگار و یاری و راهبانی اساتید بزرگ موفق به پایان این رساله شده‌ام و وظیفه خود دانسته‌ام که نهایت

پاسکزاری را از تمامی عزیزانی که در این راه به من کمک کرده‌اند را به عمل آورم:

در آغاز از اساتید بزرگ و دانشمند جناب آقایان دکتر فاتح و دکتر رضوانی که راهبانی این پایان‌نامه را به عهده داشته‌اند کمال

تشکر را دارم.

از جناب آقایان دکتر تجریمی و دکتر ابوالقاسمی که اساتید مشاور این پایان‌نامه بوده‌اند نیز قدردانی می‌نمایم.

از داوران گرامی که زحمت داوری و تصحیح این پایان‌نامه را به عهده داشتند کمال سپاس را دارم.

خالصانه از تمامی اساتید و معلمان و مدرسانی که در مقطع مختلف تحصیلی به من علم آموخته و مرا از سرچشمه دانایی سیراب

کرده‌اند تشکر می‌کنم.

مرضیه رحمتی

شهریور ۹۸

## تعمیرنامه

اینجانب مرضیه رحمتی دانشجوی دوره کارشناسی ارشد (دکتری) رشته مهندسی هوش مصنوعی دانشکده کامپیوتر دانشگاه صنعتی شاهرود نویسنده پایان‌نامه OCR متون چاپی فارسی به کمک یادگیری عمیق تحت راهنمایی دکتر منصور فاتح و دکتر محسن رضوانی متعهد می‌شوم.

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان‌نامه ، در مواردی که از موجود زنده ( یا بافتهای آنها ) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ ۱۳۹۸/۰۶/۱۷

امضای دانشجو

### مالکیت نتایج و حق نشر

کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است ) متعلق به دانشگاه صنعتی شاهرود می‌باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود . استفاده از اطلاعات و نتایج موجود در پایان‌نامه بدون ذکر مرجع مجاز نمی‌باشد.

## حکیده

سیستم OCR یک سیستم تشخیص متن نوری است، که به دلیل توسعه تکنولوژی، به طور گسترده مورد استفاده قرار گرفته است. تمرکز اصلی سیستم‌های OCR موجود بر زبان‌های لاتین بوده است. در مطالعات اخیر، این سیستم‌ها، برای زبان‌هایی با رسم‌الخط شکسته استفاده شده‌اند که با چالش‌هایی روبرو هستند. در این پایان‌نامه، یک سیستم OCR مبتنی بر یادگیری عمیق با ساختار LSTM برای زبان فارسی ارائه شده است و برخی از پارامترهای مهم سیستم پیشنهادی مورد بررسی قرار گرفته است. سیستم OCR پیشنهادی چالش‌های عدم تشخیص زیر کلمه "لا" و نیم‌فاصله را حل می‌کند. همچنین، یک الگوریتم پیش‌پردازش برای حذف کشیدگی غیر متعارف حروف با استفاده از تکنیک‌های پردازش تصویر ارائه شده است. علاوه بر این، در این پایان‌نامه، یک مجموعه داده جدید ارائه شده است که حاوی پنج میلیون تصویر با هشت فونت رایج در زبان فارسی و ۱۰ اندازه قلم است. دقت سیستم پیشنهادی در مقایسه با سیستم OCR موجود در تزرت، دو درصد افزایش یافته است. نتایج آزمایشات نشان می‌دهد که سیستم پیشنهادی دارای دقت ۹۹,۶۸۸ درصد در سطح کلمه است.

**کلمات کلیدی:** سیستم OCR فارسی، یادگیری عمیق، پردازش تصویر، کشیدگی، چاپی.

## لیست مقالات مستخرج از پایان نامه

- 1- A Printed Persian OCR System using Deep Learning (2019), IET Image Processing, status: under review.

## فهرست مطالب

۵	فهرست جداول
۵	فهرست شکل‌ها
۱	فصل ۱: مقدمه
۲	۱-۱ مقدمه
۲	۲-۱ بیان مسئله و تعاریف
۳	۳-۱ چالش‌های موجود در سیستم‌های OCR
۴	۴-۱ چالش‌های موجود در سیستم‌های OCR فارسی
۵	۵-۱ چالش‌های حل نشده در سیستم‌های OCR فارسی
۸	۶-۱ هدف پایان‌نامه
۹	۷-۱ ساختار پایان‌نامه
۱۱	فصل ۲: کارهای انجام شده
۱۲	۱-۲ مقدمه
۱۲	۲-۲ سیستم OCR
۱۴	۳-۲ یادگیری عمیق
۱۷	۴-۲ کارهای انجام شده
۱۷	۱-۴-۲ کارهای مرتبط با زبان‌هایی دارای رسم‌الخط شکسته
۲۱	۲-۴-۲ کاربردهای سیستم‌های OCR
۲۳	۵-۲ جمع‌بندی



### فصل ۳: روش پیشنهادی

۲۷

۱-۳ مقدمه ..... ۲۸

۲-۳ پیش پردازش ..... ۲۸

۱-۲-۳ تخمین حد آستانه بر اساس قطر عمودی خط پس زمینه ..... ۳۰

۲-۲-۳ تخمین حد آستانه بر اساس ارتفاع خط ..... ۳۴

۳-۲-۳ حذف کشیدگی نامتعارف ..... ۳۶

۳-۳ آموزش سیستم OCR ..... ۳۹

۱-۳-۳ مدل یادگیری عمیق پیشنهادی ..... ۴۱

۴-۳ جمع بندی ..... ۴۲

### فصل ۴: نتایج و آزمایشات

۴۵

۱-۴ مقدمه ..... ۴۶

۲-۴ مجموعه داده ..... ۴۶

۳-۴ نتایج تشخیص و حذف کشیدگی نامتعارف ..... ۴۷

۴-۴ نتایج سیستم OCR پیشنهادی ..... ۴۹

۵-۴ نتایج سیستم OCR پیشنهادی پس از اعمال الگوریتم حذف کشیدگی ..... ۵۴

### فصل ۵: نتیجه گیری

۵۹

۱-۵ مقدمه ..... ۶۰

۲-۵ جمع بندی و نتیجه گیری ..... ۶۰

۳-۵ پیشنهاد برای کارهای آتی ..... ۶۱

۶۲

مراجع

۶۶

لغت نامه

مرتب بر اساس حروف الفبای فارسی ..... ۶۷

مرتب بر اساس حروف الفبای انگلیسی ..... ۶۹

## فهرست جداول

جدول ۱-۲. مقایسه روش‌های ذکر شده. ....	۲۴
جدول ۱-۳. شرح لایه‌های شبکه پیشنهادی. ....	۴۳
جدول ۱-۴. مقایسه دقت بر اساس تعداد لایه‌های LSTM به کار گرفته در مدل آموزشی، بر اساس حرف و کلمه، در تکرار ۱۵۰۰۰ام. ....	۴۹
جدول ۲-۴. مقایسه دقت بر اساس توابع فعال‌ساز مختلف به کار گرفته در لایه کانولوشن در مدل آموزشی، بر اساس حرف و کلمه، در تکرار ۱۵۰۰۰ام. ....	۵۰
جدول ۳-۴. دقت سیستم OCR پیشنهادی به ازای همه فونت‌ها، بر اساس حرف و کلمه، در تکرار ۱۵۰۰۰۰ام. ....	۵۱
جدول ۴-۴. دقت سیستم OCR پیشنهادی به ازای هر فونت، بر اساس حرف و کلمه، در تکرار ۱۵۰۰۰۰ام. ....	۵۲
جدول ۵-۴. دقت سیستم OCR پیشنهادی به ازای هر افزایش اندازه، بر اساس حرف و کلمه، در تکرار ۱۵۰۰۰۰ام. ....	۵۳
جدول ۶-۴. دقت سیستم OCR پیشنهادی به ازای سه اندازه آموزش داده نشده و یک اندازه آموزش داده شده، بر اساس حرف و کلمه، در تکرار ۱۵۰۰۰۰۰ام. ....	۵۴
جدول ۷-۴. دقت سیستم OCR پیشنهادی به ازای همه فونت‌ها و بعد از اعمال الگوریتم ۳-۳، بر اساس حرف و کلمه، در تکرار ۱۵۰۰۰۰۰ام. ....	۵۶
جدول ۸-۴. ماتریس کانفیوژن سیستم OCR پیشنهادی. ....	۵۷

## فهرست شکل‌ها

- شکل ۱-۱. چالش‌های حل نشده سیستم OCR فارسی. ۷
- شکل ۱-۲. روند کلی سیستم‌های OCR. ۱۳
- شکل ۲-۲. معماری سلول LSTM. ۱۶
- شکل ۱-۳. نمونه‌ای از این تشخیص کشیدگی نامتعارف توسط سیستم‌های OCR. ۲۸
- شکل ۲-۳. فلوجارت الگوریتم ۱-۳ به همراه یک مثال. ۳۳
- شکل ۳-۳. فلوجارت الگوریتم ۲-۳ به همراه یک مثال. ۳۵
- شکل ۴-۳. فلوجارت الگوریتم ۳-۳ به همراه یک مثال. ۳۸
- شکل ۵-۳. معماری شبکه عصبی عمیق پیشنهادی برای آموزش سیستم OCR فارسی. ۴۲
- شکل ۱-۴. قسمتی از یک مجموعه داده پیشنهادی. ۴۷
- شکل ۲-۴. نمونه‌ای از اعمال الگوریتم پیشنهادی حذف کشیدگی. ۴۸
- شکل ۳-۴. نمودار خطای سیستم OCR فارسی چاپی. ۵۱
- شکل ۴-۴. یک نمونه از اعمال و عدم اعمال الگوریتم ۳-۳ در OCR. ۵۵

فصل ۱ : مقدمه

## ۱-۱ مقدمه

امروزه تشخیص متون، یک موضوع مهم و چالش برانگیز در بخش شناسایی الگو و هوش مصنوعی است [۱]، [۲]، [۳]. بسیاری از کتاب‌ها و اسناد، فقط در فرمت‌های تصویری مانند BMP، TIFF، JPG، PNG و غیره در دسترس هستند. یکی از مشکلات اسناد در فرمت تصویری این است که امکان جستجو و ویرایش کلمات در این اسناد وجود ندارد. از این رو، استفاده از سیستم‌های تشخیص متن اجتناب‌ناپذیر است و نیاز به تبدیل تصاویر متنی به متون قابل ویرایش وجود دارد [۴]. برای رفع این نیاز، دو روش کلی وجود دارد. در روش اول، کاربر می‌تواند متن درون تصاویر را تایپ کند. این روش، زمان‌بر و هزینه‌بر است. در روش دوم، کاربر می‌تواند با استفاده از سیستم‌های صفحه‌خوان<sup>۱</sup> مختلف، به صورت خودکار، تصاویر متنی را به متن قابل ویرایش تبدیل کند.

ورودی یک سیستم صفحه‌خوان، تصاویر یک سند است. این تصاویر، با استفاده از اسکنر، دوربین یا نرم‌افزارهای خاص تولید می‌شوند. سپس، سیستم‌های صفحه‌خوان از طریق یک فرآیند شناخته‌شده به نام "تشخیص نوری حروف (OCR)"<sup>۲</sup>، هر تصویر از سند را تجزیه و تحلیل می‌کند. سپس، این سیستم تلاش می‌کند که حروف چاپ شده روی صفحه را تشخیص دهد. خروجی این سیستم‌ها، یک فایل متنی، حاوی اعداد و حروف تشخیص داده شده از تصاویر سند است [۵].

## ۱-۲ بیان مسئله و تعاریف

بازشناسی نوری حروف (OCR)، فرآیند تجزیه و تحلیل خودکار تصاویر حاصل از اسناد به زبان‌های مختلف است [۶]. به عبارت دیگر، OCR، فرآیند طبقه‌بندی خودکار الگوهای نوری درون یک تصویر دیجیتال حاوی حروف و اعداد است [۵]. سیستم‌های OCR، قادر به تبدیل خودکار تصاویر متون چاپی به متون قابل ویرایش و قابل جستجو با حجم کم هستند [۶].

---

<sup>۱</sup> Page Reading

<sup>۲</sup> Optical Character Recognition

سیستم‌های OCR از نظر نوع ورودی، به سیستم‌های تشخیص متون چاپی و متون دست‌نویس تقسیم می‌شوند [۱۵]. این نکته نیز حائز اهمیت است که این متون چاپی و دست‌نویس، دارای دو نوع مختلف گسسته و پیوسته هستند. متون پیوسته، متونی مانند خاطرات، نامه و غیره هستند که ما به طور روزانه آن‌ها را می‌نویسیم. اما متون گسسته، متونی هستند که در آن‌ها حروف به صورت جدا از هم نوشته می‌شوند. فرم‌های اطلاعاتی که نام و نام‌خانوادگی خود را در آن‌ها به صورت هر حرف داخل یک کادر نوشته می‌شود، نمونه‌ای از متون گسسته هستند. بدیهی است که توسعه سیستم‌های OCR برای متون گسسته، بسیار راحت‌تر از متون پیوسته است. دسته‌بندی دیگری برای این سیستم‌ها نیز وجود دارد. در این نوع دسته‌بندی، سیستم‌های OCR به دو گروه برخط<sup>۱</sup> و برون‌خط<sup>۲</sup> تقسیم شده‌اند. ورودی سیستم‌های OCR برخط، مختصات حرکت قلم است. در این دسته از سیستم‌ها، فرآیند تشخیص، در طول نوشتن رخ می‌دهد. از طرف دیگر، ورودی سیستم‌های برون‌خط، تصاویر چاپی یا دست‌نویس هستند [۱].

از جمله مهم‌ترین برنامه‌های کاربردی که سیستم‌های OCR در آن نقش مهمی دارند، می‌توان به شناسایی پلاک خودرو [۷، ۸، ۹، ۱۰]، خواندن شماره کنتور گاز [۱۱]، کمک به نابینایان [۱۲]، پردازش چک‌های بانکی [۱۳، ۱۴، ۱۵] و مرتب‌سازی خودکار<sup>۳</sup> [۱۶] اشاره کرد.

### ۱-۳ چالش‌های موجود در سیستم‌های OCR

چالش‌هایی در بازیابی متون درون تصاویر وجود دارد. وجود تصاویری حاوی متونی با چندین اندازه، چندین فونت و چندین رزولوشن، از چالش‌های این سیستم‌ها است. برای رفع این چالش‌ها، راه‌حل‌هایی ارائه شده است. به عنوان مثال، برای آموزش برخی از این سیستم‌ها، از مجموعه داده‌هایی حاوی متونی با چندین

---

<sup>۱</sup> Online

<sup>۲</sup> Offline

<sup>۳</sup> Automatic Sorting

اندازه، چندین فونت و چندین رزولوشن بهره گرفته شده است و آموزش این سیستم‌ها به صورت همزمان [۱۷، ۱۸، ۱۹، ۲۰، ۲۱] یا غیر همزمان [۱۰، ۱۴، ۲۲] است.

یکی دیگر از راه‌حل‌های رفع چالش‌های ذکر شده، تشخیص فونت و اندازه، قبل از شروع فرآیند تشخیص متن است. تشخیص فونت یکی از مراحل پیش‌پردازش در سیستم‌های OCR است، که تاثیر زیادی بر روی عملکرد آن‌ها دارد. زمانی که فونت و اندازه، قبل از فرآیند تشخیص شناسایی شود، شکل ظاهری کلمات و اندازه متن برای سیستم مشخص است و سیستم فقط کافی است که به تشخیص متن بپردازد [۲۳].

از دیگر چالش‌های بازیابی متن درون تصاویر، وجود پس‌زمینه‌های پیچیده، نویزهای مختلف، کیفیت پایین و چرخش تصاویر است. حذف و رفع این چالش‌ها، در مرحله پیش‌پردازش و توسط الگوریتم‌های مختلف انجام می‌شود [۲۰، ۲۱، ۲۴، ۲۵]. از دیگر چالش‌ها در این زمینه، وجود تصاویری حاوی متون به چند زبان مختلف است. برای رفع این چالش، می‌توان هنگام آموزش سیستم از تصاویر حاوی متونی با چند زبان بهره گرفت [۲۶].

## ۴-۱ چالش‌های موجود در سیستم‌های OCR فارسی

افزایش روزافزون فایل‌های غیر قابل ویرایش مانند PDF، کاربرد سیستم‌های OCR را به صورت قابل توجهی افزایش داده است. هم‌چنین، فرآیند تشخیص این سیستم‌ها به صورت خودکار و بدون دخالت انسان انجام می‌شود. از این رو، این ویژگی باعث کاربردی‌تر شدن این سیستم‌ها شده است [۲۶]. پیچیدگی این فرآیند در سیستم‌های OCR، برای زبان‌های گوناگون، متفاوت است. توسعه سیستم‌های OCR برای زبان‌های لاتین مانند زبان انگلیسی، به دلیل گسستگی حروف در رسم‌الخط این نوع از زبان‌ها، آسان‌تر است [۴]. از این رو، در سال‌های اخیر، پژوهش‌های بسیاری در زمینه توسعه سیستم‌های OCR برای زبان‌های لاتین



صورت گرفته است. اما در مورد زبان‌هایی با رسم‌الخط شکسته<sup>۱</sup>، مانند زبان فارسی، عربی و اردو<sup>۲</sup>، باید از تکنیک‌های خاصی برای بازشناسی متون استفاده کرد.

زبان‌های دارای رسم‌الخط شکسته، دارای پیچیدگی ذاتی و تعداد ویژگی بیشتری نسبت به زبان‌های لاتین هستند. بنابراین، سیستم‌های OCR کمتری برای این نوع زبان‌ها توسعه داده شده است [۴]. از جمله ویژگی‌های این زبان‌ها، می‌توان به پیوستگی حروف در یک کلمه، نوشتن راست به چپ حروف، نوشتن چپ به راست اعداد، شباهت شکلی حروف و اعداد، شکل‌های متفاوت یک حرف با توجه به موقعیت آن در کلمه و وجود فونت‌های زیاد و متنوع اشاره کرد [۱۷]. برای غلبه بر این ویژگی‌ها، از روش‌های متعددی استفاده شده است. برخی از این روش‌ها، الگوریتم‌های K-means [۲۷]، ماشین بردار پشتیبان<sup>۳</sup> [۲۷، ۲۸، ۲۹، ۳۰]، کاهش ابعاد ویژگی‌های استخراج شده [۲۱]، شبکه‌های عصبی [۱۸، ۱۹] و الگوریتم ژنتیک [۳۲] هستند. اما همچنان چالش‌هایی در سیستم‌های OCR مربوط به این زبان‌ها وجود دارد.

## ۵-۱ چالش‌های حل نشده در سیستم‌های OCR فارسی

همان‌طور که در بخش قبل توضیح داده شد، سیستم‌های OCR فارسی هنوز به طور کامل توسعه داده نشده‌اند و با چالش‌هایی مواجه هستند. در این بخش، به توضیح برخی از این چالش‌ها پرداخته شده است. علاوه بر پیچیدگی ذاتی موجود در حروف الفبای زبان فارسی، تعدادی قواعد نگارشی برای آراستگی ظاهری و مرتب‌سازی متون نیز وجود دارد. یکی از این قواعد، وجود حروف کشیده در متون است. این کشیدگی، یکی از قابلیت‌های مهم نرم‌افزارهای واژه‌پرداز است که با نام "کشیدگی حاصل از تراز کردن متون"<sup>۴</sup> شناخته می‌شود. نحوه تراز کردن متون بدین صورت است که برخی حروف در کلمات موجود در یک سطر کشیده می‌شوند، تا فاصله سطر از راست و چپ هم‌اندازه شود. تراز کردن، باعث افزایش خوانایی متون فارسی

---

<sup>۱</sup> Cursive

<sup>۲</sup> Urdu

<sup>۳</sup> Support Vector Machines

<sup>۴</sup> Justify

می‌شود و شامل سه حالت حداقل، متوسط و حداکثر است. تفاوت این سه حالت، در اندازه کشیدگی حروف است. در این پایان‌نامه، ما برای جلوگیری از تکرار نام طولانی این نوع کشیدگی، آن را "کشیدگی نامتعارف" نامگذاری می‌کنیم. شکل ۱-۱ الف و شکل ۱-۱ ب، نمونه‌ای از تاثیر اعمال و عدم اعمال این نوع کشیدگی بر روی حروف را نشان می‌دهد.

یکی دیگر از این قواعد نگارشی زبان فارسی، استفاده از نیم‌فاصله، جهت خوانایی بهتر و پرهیز از تغییر ساختار معنایی جمله است. به‌طور طبیعی، حروفی که قابلیت اتصال به حرف بعدی را ندارند، اگر کنار حرف دیگری قرار گیرند، به اندازه نیم‌فاصله از آن جدا می‌شوند [۳۳]. نیم‌فاصله در نگارش کلمات و افعال مرکب، فعل‌های پیشوندی، کلمات ساخته شده با پیشوند یا پسوند یا هر دو و ترکیب‌های عطفی استفاده می‌شود. تشخیص نیم‌فاصله به وسیله سیستم‌های OCR، به بهبود عملکرد آن‌ها کمک می‌کند.

همچنین، طبق بررسی‌های انجام گرفته، برخی از سیستم‌های OCR موجود، قادر به تشخیص درست زیر کلمه "لا" نیستند و آن را به صورت برعکس تشخیص می‌دهند. شکل‌های ۱-۱ ج و ۱-۱ د، نمونه‌هایی از اعمال و عدم اعمال ویژگی نیم‌فاصله بر روی کلمات فارسی و شکل‌های ۱-۱ ه و ۱-۱ و، تشخیص درست و نادرست زیر کلمه "لا" را نشان می‌دهند.

چهار سیستم OCR به طور گسترده در زبان عربی و فارسی استفاده می‌شود [۴، ۳۴]. یکی از آن‌ها تزرکت<sup>۱</sup> است، که از یادگیری عمیق<sup>۲</sup>، فرمت UTF-8، چند زبانه بودن<sup>۳</sup> و مدیریت زبان‌های راست به چپ<sup>۴</sup> پشتیبانی می‌کند و یک ابزار منبع باز<sup>۵</sup> است [۳۴]. این جعبه‌ابزار در سال‌های بین ۱۹۸۴ تا ۱۹۹۴

---

<sup>۱</sup> Tesseract

<sup>۲</sup> Deep Learning

<sup>۳</sup> Multi-Language

<sup>۴</sup> RTL Language

<sup>۵</sup> Open Source



شکل ۱-۱. چالش‌های حل نشده سیستم OCR فارسی. الف) بدون کشیدگی نامتعارف. ب) با کشیدگی

نامتعارف. ج) بدون نیم‌فاصله. د) با نیم‌فاصله. ه) تشخیص درست "لا" و تشخیص نادرست "لا".

تولید شده است [۲۲]. تزرتک، در نسخه‌های اولیه، فقط برای زبان انگلیسی استفاده می‌شد، اما این جعبه‌ابزار در طول زمان به‌روزرسانی شده است، به‌طوری‌که امروزه، بیش از ۱۰۰ زبان را پشتیبانی می‌کند. این نرم‌افزار در نسخه‌های مختلفی ارائه شده است. در حال حاضر، نسخه‌ی بتای ۴,۰,۰,۱ آن منتشر شده است و برای استفاده، در دسترس است. این نسخه، از یادگیری عمیق استفاده می‌کند و مبتنی بر مدل LSTM<sup>۱</sup> است. سیستم OCR فارسی موجود در این جعبه ابزار، دارای چالش‌های حل نشده ذکر شده، یعنی عدم تشخیص کامل و صحیح کشیدگی نامتعارف، نیم‌فاصله و زیر کلمه "لا" است.

در این پایان‌نامه، برای آموزش سیستم OCR از جعبه ابزار تزرتک بهره گرفته شده است و سعی شده است که چالش‌های حل نشده ذکر شده رفع شوند. برای رفع چالش‌ها، از یک مدل یادگیری عمیق مبتنی بر LSTM و تکنیک‌های پردازش تصویر استفاده شده است. همچنین، در فرآیند آموزش از یک مجموعه داده پیشنهادی نیز بهره گرفته شده است. علاوه بر حل چالش‌های موجود، دقت سیستم OCR فارسی پیشنهادی نیز افزایش یافته است.

<sup>۱</sup> Long short-term memory

## ۱-۶ هدف پایان نامه

همان طور که در بخش قبل گفته شد، چندین چالش حل نشده در OCR زبان فارسی وجود دارد، که هدف این پایان نامه رفع چالش های ذکر شده، به همراه افزایش دقت تشخیص در سیستم OCR فارسی است. بر این اساس، این پایان نامه، بر دو مرحله پیش پردازش و آموزش سیستم OCR فارسی تمرکز دارد. در مرحله اول، انجام پیش پردازشی روی تصاویر ورودی این سیستم ها انجام گرفته است. یک الگوریتم پیش پردازش، برای حذف کشیدگی نامتعارف از تصاویری حاوی متون فارسی ارائه شده است. الگوریتم حذف کشیدگی نامتعارف پیشنهادی در این پایان نامه، کشیدگی نامتعارف در حروف تصاویر را حذف می کند. این الگوریتم، از تکنیک های پردازش تصویر استفاده می کند، که برای تعیین حد آستانه جهت تشخیص کشیدگی نامتعارف، از دو پارامتر اصلی یک تصویر متنی، به نام های قطر عمودی خط پس زمینه<sup>۱</sup> و ارتفاع خط<sup>۲</sup> بهره گرفته شده است.

همچنین، در این پایان نامه، سعی شده است که با ساخت یک مجموعه داده جدید حاوی تعدادی کافی کلمه با نیم فاصله و زیر کلمه "لا" و استفاده از آن در روند آموزش سیستم OCR، چالش های عدم تشخیص نیم فاصله و "لا" در متون رفع شوند.

در مرحله دوم، برای آموزش سیستم OCR فارسی، از یادگیری عمیق استفاده شده است. در یادگیری عمیق، به صورت خودکار، مهندسی ویژگی انجام می شود. بدین معنا که شبکه عمیق به استخراج ویژگی و الگوهای مناسب از داده های آموزشی می پردازد. این ویژگی اساسی یادگیری عمیق، از دخالت انسان در استخراج ویژگی و وابستگی نتایج به این مرحله جلوگیری می کند [۳۵]. همچنین، برخی از پارامترهای مهم شبکه ی عصبی عمیق مورد بررسی قرار گرفته است و از بهترین پارامترها، برای آموزش استفاده شده است.

---

<sup>۱</sup> Vertical Diameter Baseline

<sup>۲</sup> Height Line

دلیل دیگر برای استفاده از شبکه‌های عصبی و به ویژه LSTM، که یکی از شبکه‌های عصبی مکرر است، این است که این تحقیق با داده‌های زبانی سر و کار دارد و هنگام استفاده از این نوع داده‌ها، اغلب با دنباله‌ای از حروف، کلمات یا جملات مواجه هستیم و در بیشتر موارد، ترتیب این دنباله اهمیت دارد. شبکه‌های عصبی مکرر این امکان را فراهم می‌کنند تا یک دنباله با طول نامشخص به یک بردار با اندازه ثابت نگاشت شود. در عین حال، بسیاری از خواص نحوی و ساختاری دنباله ورودی حفظ خواهند شد.

به طور خلاصه، آزمایش‌های زیر در این پایان‌نامه انجام شده و مورد بررسی قرار گرفته است:

- یک روش برای آموزش سیستم OCR فارسی ارائه شده است. این سیستم از یادگیری عمیق و جعبه‌ابزار تزرکت استفاده می‌کند. این سیستم، دارای دقت بالاتری نسبت به سیستم OCR موجود در تزرکت است.

- سه چالش موجود در سیستم OCR فارسی تزرکت، یعنی عدم تشخیص کامل و صحیح کشیدگی نامتعارف، نیم‌فاصله و زیر کلمه "لا"، با تکنیک پردازش تصویر، تغییر لایه‌های شبکه یادگیری عمیق و یک مجموعه‌داده مناسب برطرف شده است.

در نهایت، دقت سیستم OCR فارسی پیشنهادی در مقایسه با سیستم OCR فارسی موجود در تزرکت، دو درصد افزایش داشته است.

## ۷-۱ ساختار پایان‌نامه

این پایان‌نامه در پنج فصل تدوین شده است. در فصل دوم، مروری کلی بر سیستم‌های OCR موجود برای زبان‌های فارسی، عربی و اردو انجام شده است. در فصل سوم، روش پیشنهادی برای حذف کشیدگی و در نهایت، مدل جدیدی برای آموزش سیستم OCR ارائه شده است. در ادامه و در فصل چهارم، نتایج آزمایش‌های انجام شده در این پایان‌نامه ارائه شده است. فصل پنجم، به جمع‌بندی و نتیجه‌گیری اختصاص داده شده است.



فصل ۲: کارهای انجام شده

## ۱-۲ مقدمه

استفاده روز افزون از تکنولوژی، سبب پیشرفت‌های چشمگیری در استفاده از سیستم‌های هوشمند شده است. سیستم‌های OCR نیز با توجه به افزایش اسناد تصویری، از این قاعده مستثنی نیستند. بنابراین، این سیستم‌ها باید بهبود یابند تا بتوانند نیاز کاربران را مرتفع سازند، که در این زمینه تحقیقاتی انجام شده است. در این بخش از پایان‌نامه، ابتدا به شرح کلی یک سیستم OCR پرداخته شده است. سپس، برخی پژوهش‌های مرتبط با سیستم‌های OCR گردآوری شده است.

## ۲-۲ سیستم OCR

OCR به معنای نویسه‌خوان نوری است و وظیفه آن تشخیص خودکار متن‌ها در تصاویر و یا اسناد و تبدیل آن به متن قابل ویرایش و جستجو در رایانه می‌باشد. یک تصویر یا یک سند ممکن است از نظر انسان ارزش اطلاعاتی بسیاری داشته باشد اما از دید کامپیوتر، آن سند فقط متشکل از چند پیکسل ساده است که یک تصویر را تشکیل داده است. برای این که بتوان از اطلاعات نوشتاری موجود در تصاویر یا اسناد در رایانه استفاده کرد، باید از سیستم‌های OCR استفاده کرد.

سیستم OCR، در ابتدا برای شناسایی ارقام و حروف گسسته در اسناد چاپی استفاده می‌شد. نرم‌افزارهای نویسه‌خوان نوری، متن آن اسناد چاپی را می‌خواندند و آن را به قالب قابل ویرایش در کامپیوتر تبدیل می‌کردند. با این که تصاویر اسکن شده در رایانه بسیار سنگین هستند و امکان جستجو در آن‌ها وجود ندارد، اما فایل‌هایی که خروجی سیستم‌های OCR هستند، بسیار کم حجم‌تر از تصاویر بوده و می‌توان به راحتی یک متن را در آن جستجو کرد.

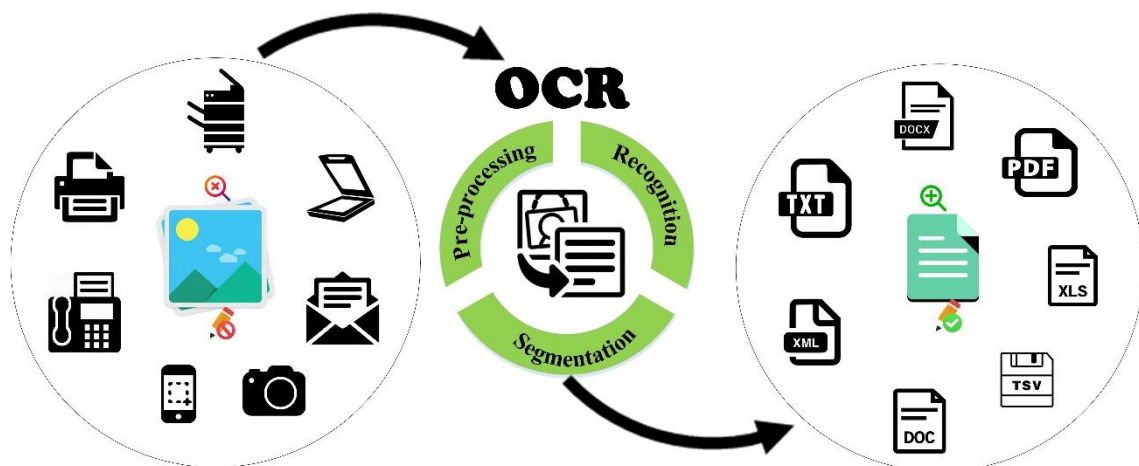
سیستم‌های OCR، همانند بسیاری از نرم‌افزارهای هوشمند دیگر، دارای پیچیدگی‌های زیادی هستند که میزان این پیچیدگی در زبان‌های گوناگون متفاوت است. به عنوان مثال، زبان‌هایی که حروف آن به صورت جدا از هم و مجزا نوشته می‌شوند، آسان‌تر از زبان‌هایی هستند که در آن‌ها، برای نوشتن یک کلمه، حروف به یکدیگر می‌چسبند. به عنوان مثال، تشخیص کلمات انگلیسی بسیار آسان‌تر از تشخیص کلمات فارسی



و یا عربی است.

در حالت کلی، سیستم‌های OCR شامل سه مرحله اساسی پیش‌پردازش، تقسیم‌بندی و تشخیص است [۴]. شکل کلی OCR در شکل ۱-۲ نشان داده شده است.

مرحله پیش‌پردازش، شامل کلیه اعمالی است که روی سیگنال تصویری خام صورت می‌گیرد تا موجب تسهیل روند اجرای فازهای بعدی گردد. این سیستم‌ها، اغلب پیش‌پردازش‌هایی روی تصاویر انجام می‌دهند تا کیفیت تصاویر بهبود یابد. برخی از این تکنیک‌ها، سیاه و سفید کردن تصاویر، چرخش تصاویر، حذف نویز، هموار سازی، نرم کردن لبه‌ها<sup>۱</sup>، دودویی کردن<sup>۲</sup> تصاویر، تشخیص زبان، تشخیص فونت، حذف کشیدگی نامتعارف حروف و نظیر این‌ها هستند [۳۶]. استفاده از این الگوریتم‌های پیش‌پردازش، دقت تشخیص را افزایش می‌دهد.



شکل ۱-۲. روند کلی سیستم‌های OCR.

اغلب الگوریتم‌های تشخیص متن، به دلیل تسهیل در انجام کار، فقط بر روی تصاویر دودویی عمل می‌کنند. بنابراین، اولین و مهم‌ترین کار در مرحله پیش‌پردازش، دودویی کردن تصاویر است [۳۷]. علاوه بر این، دودویی کردن، به میزان قابل توجهی فرآیند تشخیص حروف را تحت تأثیر قرار می‌دهد.

<sup>۱</sup> Smoothing

<sup>۲</sup> Binarization

مرحله دوم در سیستم OCR، تقسیم‌بندی تصاویر ورودی است. در این مرحله، خروجی مرحله قبل، به تعدادی پاراگراف، خط، جمله، کلمه و حتی حرف تقسیم می‌شود تا سیستم OCR بتواند مجموعه‌ای از ویژگی‌های معنادار را برای مرحله بعدی استخراج کند [۴]. این نکته حائز اهمیت است که گام تقسیم‌بندی به راحتی می‌تواند تحت تأثیر نوع فونت و اندازه قلم قرار گیرد. در نتیجه، برای ایجاد یک سیستم OCR قوی، باید سیستم OCR را با مجموعه داده‌ای شامل انواع فونت‌ها و اندازه‌های مختلف ارزیابی کرد [۳۸].

مرحله آخر، شامل روش‌هایی برای تطبیق هر یک از الگوهای بدست آمده از مرحله استخراج ویژگی با یکی از کلاس‌های فضای الگوهای مورد بحث است. این عمل از طریق کمینه ساختن فاصله بردار ویژگی‌های هر الگو نسبت به یکی از بردارهای مرجع موجود در پایگاه‌داده‌های سیستم انجام می‌گیرد. به عبارت دیگر، این مرحله، وظیفه تشخیص بخش‌های شناسایی شده در مرحله دوم، و در نهایت تولید خروجی نهایی را دارد. برای تشخیص بخش‌های شناسایی شده، می‌توان از روش‌های مختلفی مانند KNN<sup>۱</sup>، SVM<sup>۲</sup>، ANN<sup>۳</sup> و LSTM استفاده کرد [۳۶، ۳۷، ۳۹]. بر اساس تحقیقات انجام شده، استفاده از LSTM می‌تواند دقت را به طور قابل توجهی افزایش دهد. این امر به دلیل وجود سه ویژگی مکانیزم به‌روزرسانی خودکار<sup>۳</sup>، سلول حافظه<sup>۴</sup> و پس‌انتشار خطای پویا<sup>۵</sup> در این لایه است [۳۹].

## ۳-۲ یادگیری عمیق

یادگیری عمیق، زیرمجموعه‌ای از الگوریتم‌های یادگیری ماشین است. این نوع یادگیری، به معنی استفاده از شبکه‌های عصبی داده‌محور با تعداد لایه‌های مخفی زیاد است. در یادگیری عمیق، به صورت خودکار، مهندسی ویژگی انجام می‌شود. بدین معنا که شبکه، به صورت خودکار و بدون نیاز به انسان، به دنبال

---

<sup>۱</sup> k-Nearest Neighbor

<sup>۲</sup> Artificial Neural Network

<sup>۳</sup> Automatic Update

<sup>۴</sup> Memory Cell

<sup>۵</sup> Dynamic Error Backpropagation

بهترین ویژگی‌ها، برای تمیز دادن کلاس‌های مختلف داده می‌گردد. در نهایت، این ویژگی‌های بدست آمده می‌توانند داده ورودی را به بهترین وجه و بالاترین دقت طبقه‌بندی کنند. این ویژگی اساسی یادگیری عمیق، از وابستگی نتایج به مرحله استخراج ویژگی توسط انسان جلوگیری می‌کند [۳۵].

یک گروه مشهور از شبکه‌های عصبی، شبکه‌های عصبی مکرر<sup>۱</sup> هستند. به‌طور کلی، شبکه عصبی مکرر RNN، تابعی است که یک ورودی با طول غیرثابت مانند یک جمله را به صورت دنباله‌ای از  $n$  بردار  $d_{in}$  بعدی دریافت می‌کند و یک بردار خروجی  $y$  با ابعاد  $d_{out}$  را باز می‌گرداند، که در رابطه (۲-۱) نشان داده شده است.

$$y_n = \text{RNN}(x_{1:n}) \quad (1-2)$$

$$x_i \in R^{d_{in}} \quad y_i \in R^{d_{out}}$$

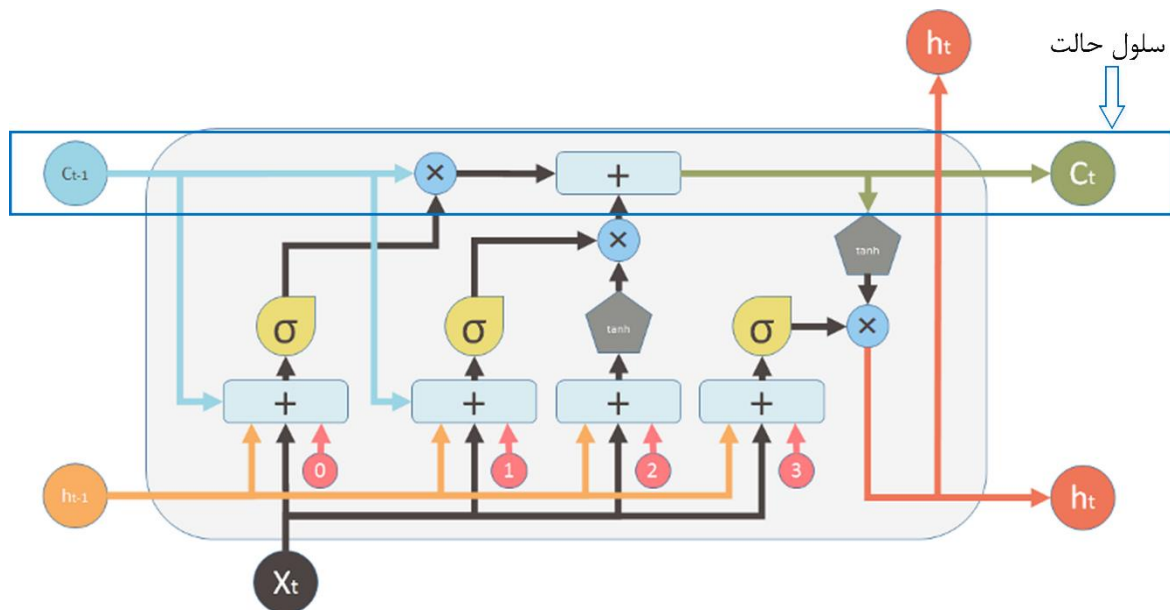
RNN استاندارد، توسط پس‌انتشار خطا آموزش می‌بیند. این شبکه، در یادگیری دنباله‌های بزرگ، دچار خطا می‌شود. این روند آموزش، موجب ایجاد مشکل ناپدید شدن یا انفجار گرادیان<sup>۲</sup> می‌شود. برای حل این مشکل، سلول RNN با چند دروازه جایگزین می‌شود. LSTM نوع خاصی از RNN است، که توانایی یادگیری وابستگی‌های بلند مدت را دارد. این کار به صورت پیش‌فرض و عادی برای LSTM وجود دارد و ساختار آن به صورتی است که اطلاعات خیلی دور را به خوبی یاد می‌گیرد [۴۰].

همه شبکه‌های عصبی، به شکل دنباله‌ای (زنجیره‌ای) تکرار شونده از ماژول‌های شبکه عصبی هستند. در شبکه‌های عصبی مکرر استاندارد، این ماژول‌های تکرار شونده ساختار ساده‌ای دارند. ماژول‌های تکرار شونده می‌توانند ساختار متفاوتی داشته باشند. این ساختارها، طبق قواعد ویژه‌ای، با هم در تعامل و ارتباط هستند. نمونه‌ای از یک شبکه LSTM در شکل ۲-۲ آورده شده است.

---

<sup>۱</sup> Recurrent Neural Network

<sup>۲</sup> Vanishing Gradient



شکل ۲-۲. معماری سلول LSTM [۴۰].

عنصر اصلی LSTM، سلول حالت است که به صورت خط افقی در بالای شکل ۲-۲ نشان داده شده است. این سلول مشخص می‌کند که چه مقدار از اطلاعات سلول حالت LSTM مرحله‌ی قبل، که با  $h_{t-1}$  نام‌گذاری شده‌است، باید در سلول حالت کنونی  $C_{t-1}$  نگه‌داشته شود. این تصمیم‌گیری، توسط یک لایه سیگموئید<sup>۱</sup> انجام می‌شود. این عملیات در معادله (۲-۲) نشان داده شد.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (۲-۲)$$

سپس، طبق معادله‌های (۳-۲)، (۴-۲) و (۵-۲)، LSTM باید تصمیم بگیرد که چه اطلاعات جدیدی به سلول حالت افزوده شود. این تصمیم، دو بخشی است. در بخش اول، باید تصمیم گرفته شود که چه اطلاعاتی باید به‌روز شوند. این تصمیم توسط یک لایه سیگموئید گرفته می‌شود. در بخش دوم، باید تصمیم گرفته شود که چه اطلاعاتی باید به سلول حالت افزوده شود. این تصمیم توسط یک لایه تانژانت هیپربولیک گرفته می‌شود.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (۳-۲)$$

$$C_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (۴-۲)$$

<sup>۱</sup> Sigmoid

$$\tilde{C}_t = f_t * C_{t-1} + i_t * C_t \quad (5-2)$$

در نهایت، باید مقدار خروجی مشخص شود. این خروجی، با در نظر گرفتن مقدار موجود سلول حالت محاسبه خواهد شد، ولی باید از فیلتر مشخصی عبور کند. بدین صورت که، ابتدا یک لایه سیگموئید تصمیم می‌گیرد که چه مقدار از سلول حالت باید به خروجی برده شود. سپس، برای نگاشت خروجی به محدوده بین ۱- تا ۱+ باید به یک لایه تانژانت هیپربولیک ارسال می‌شود. سپس، خروجی در یک لایه سیگموئید ضرب می‌شود تا فقط بخش مورد نظر به خروجی نهایی برده شود. این عملیات در رابطه‌های (۶-۲) و (۷-۲) نشان داده شده‌اند [۴۰].

$$O_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_o) \quad (6-2)$$

$$h_t = O_t * \tanh(C_t) \quad (7-2)$$

## ۴-۲ کارهای انجام شده

در این بخش، به بررسی برخی از کارهای انجام شده در زمینه سیستم‌های OCR می‌پردازیم. از آنجایی که زبان فارسی، یک زبان با رسم‌الخط شکسته است، این بخش، بیشتر به مرور تحقیقات مرتبط با این نوع زبان‌ها می‌پردازد. همچنین، در این فصل، برخی از کاربردهای سیستم OCR نیز مورد بررسی قرار گرفته‌اند.

### ۱-۴-۲ کارهای مرتبط با زبان‌هایی دارای رسم‌الخط شکسته

رفع چالش‌های موجود در بازیابی متن، در مرحله اول سیستم‌های OCR، یعنی پیش‌پردازش انجام می‌شود. در بازیابی متن از تصویر، چالش‌هایی مانند اندازه‌های متفاوت، فونت‌های مختلف، جهت‌گیری، پس‌زمینه پیچیده و غیره وجود دارد. در سیستم‌های OCR، این نکته حائز اهمیت است که دقت می‌تواند تا حد زیادی وابسته به الگوریتم پیش‌پردازش تصویر و الگوریتم تقسیم‌بندی باشد [۴۱].

مرحله دوم سیستم‌های OCR، یعنی تقسیم‌بندی خطوط متن، یک مرحله مهم و حساس در سیستم‌های OCR است. این امر، دلایل زیادی مانند پیچیدگی پس‌زمینه و چرخش متن دارد. در مرجع [۴۱]،

الگوریتمی برای تقسیم‌بندی ساده و قوی متون دست‌نوشته و چاپی به زبان اردو ارائه شده است. مجموعه- داده استفاده شده در این مقاله، شامل ۸۰ صفحه از متون چاپی شامل ۶۸۷ خط و ۴۸ صفحه از متون دست‌نوشته شامل ۴۹۵ خط است. دقت روش پیشنهادی این مقاله، بر روی مجموعه‌داده متون چاپی برابر ۹۹,۱۹ درصد و برای مجموعه‌داده دست‌نوشته برابر ۹۸,۱ درصد است.

تشخیص حروف در زبان اردو، به دلیل ماهیت ذاتی شکسته بودن این زبان، یک مشکل پیچیده است. وجود فونت‌ها و اندازه‌های مختلف حروف، بر این مشکل می‌افزایند. در مرجع [۴۲]، از شبکه عصبی کانولوشن<sup>۱</sup> و شبکه LSTM، برای شناسایی حروف زبان اردو، در اندازه‌های مختلف استفاده شده است. برای ارزیابی مقیاس‌پذیری روش ارائه شده، ساختار LSTM-RNN، بر روی مجموعه ویژگی‌های از قبل استخراج شده اعمال شده است. عملکرد کلی شبکه‌ها بین ۹۰ تا ۹۸,۸ درصد بوده است.

اکثر سیستم‌های OCR موجود برای زبان اردو، از تشخیص متن در یک اندازه واحد پشتیبانی می‌کنند. در مرجع [۴۳]، یک سیستم OCR مستقل از اندازه، برای تشخیص متون در تصاویر اسناد به زبان اردو ارائه شده است. در این مقاله، از اندازه‌های ۱۴ تا ۲۸ برای آموزش استفاده شده است و سه اندازه‌ی ۱۴، ۱۶ و ۲۲ به عنوان سه اندازه اصلی در نظر گرفته شده‌اند. روش کار بدین گونه است که هنگام استخراج متن از یک تصویر، با استفاده از تکنیک نزدیک‌ترین همسایه، اندازه متن، به یکی از سه اندازه اصلی تغییر می‌کند تا بتوان متن را به خوبی استخراج کرد.

یکی از بزرگ‌ترین چالش‌ها در طبقه‌بندی و تشخیص متن، وجود ویژگی‌های زیاد و اضافی برای هر داده است. انتخاب یک زیر مجموعه موثر از ویژگی‌های مهم هر داده، کارایی سیستم را بالا می‌برد. در مرجع [۲۱]، برای کاهش ابعاد ویژگی داده‌ها، یک روش ترکیبی با استفاده از شبکه عصبی مصنوعی و الگوریتم ژنتیک کوانتومی<sup>۲</sup> پیشنهاد شده است که می‌تواند برای تشخیص حروف دست‌نویس فارسی استفاده شود. هدف اصلی این مقاله، شناسایی حروف دست‌نویس فارسی با دقت بالاتر و سرعت بیشتر بوده است.

---

<sup>۱</sup> Convolutional Neural Network

<sup>۲</sup> Genetic Quantum Algorithm

زبان کُردی دارای گویش‌های مختلفی است و هنگام نوشتن آن، چندین رسم‌الخط استفاده می‌شود. در مرجع [۱۷]، یک سیستم OCR برای گویش کردی ارائه شده است. در این گویش، از لغات فارسی و عربی به طور گسترده استفاده شده است. این لغات به هم پیوسته هستند و با ساختار راست به چپ در گویش کردی استفاده شده‌اند. در این تحقیق، یک روش مبتنی بر بخش‌بندی زیر کلمات پیشنهاد شده است. در واقع روش پیشنهادی، ویژگی به هم پیوستگی لغات را در نظر گرفته است. دقت روش پیشنهادی بر روی اسناد به زبان کردی، ۹۰٫۸۲ درصد است.

تشخیص فونت یکی از مراحل پیش‌پردازش در سیستم‌های OCR است که می‌تواند عملکرد آن‌ها را تا حد زیادی تحت تاثیر قرار دهد. در مرجع [۲۳]، دو روش برای تشخیص فونت در زبان فارسی ارائه شده است. در روش اول، برای استخراج ویژگی از تصاویر از فیلتر Gabor، برای کاهش ابعاد ویژگی‌ها از الگوریتم PCA<sup>۱</sup> و برای طبقه‌بندی از یک شبکه عصبی پرسپترون<sup>۲</sup> استفاده شده است. در روش دوم، از الگوریتم Random Forest برای تشخیص فونت استفاده شده است. برای ارزیابی روش پیشنهادی، از یک مجموعه داده با ۱۰ فونت مشهور فارسی استفاده شده است. معیار F-measure برای روش اول و دوم به ترتیب برابر ۹۸٫۷۰ درصد و ۹۶٫۹۵ درصد است.

شناسایی اسکرپیت در تجزیه و تحلیل اسناد و فیلم‌ها نقش مهمی دارد. با توجه به کیفیت پایین تصاویر، پس‌زمینه پیچیده و وجود شکل‌های مختلف حروف در برخی از اسکرپیت‌ها، مانند یونانی، فارسی، و غیره، تشخیص متن یکی از چالش‌های موجود است. در مرجع [۱۸]، بر روی مسئله شناسایی اسکرپیت‌ها در تصاویر متنی با پس‌زمینه پیچیده و اسکرپیت‌های ویدئویی تمرکز شده است. روش پیشنهادی در این مقاله، استخراج و ترکیب ویژگی‌های محلی و سراسری با استفاده از چارچوب CNN-LSTM است که اسکرپیت‌های شناسایی شده را به صورت پویا ارزیابی می‌کند. مجموعه داده‌های استفاده شده در این مقاله، SIW-13، CVSI2015، ICDAR-17، و MLe2e هستند.

---

<sup>۱</sup> Principal Component Analysis

<sup>۲</sup> Perceptron Neural Network

تشخیص حروف دست‌نویس، یک موضوع اساسی در بخش تشخیص الگو و هوش مصنوعی بوده و همچنان به عنوان یک چالش مهم، در حال بررسی است. با توجه به کاربرد عظیم این موضوع، تشخیص حروف دست‌نویس برای زبان‌های مختلفی انجام شده است. زبان عربی، یک زبان متنوع است که تحقیقات گسترده‌ای روی آن انجام شده است. در مرجع [۳]، یک مدل شبکه عصبی کانولوشن برای شناسایی اعداد دست‌نویس در این زبان ارائه شده است. در این مقاله، از مجموعه داده‌ی CMATERDB 3.3.1 استفاده شده است. روش پیشنهادی در برابر بیش‌برازش<sup>۱</sup> بسیار توانمند است. هم‌چنین، با تغییر تابع فعال‌ساز، مشکل ناپدید شدن گرادیان از بین رفته است. دقت روش پیشنهادی در این مقاله، ۹۹،۴ درصد است.

ساخت سیستم‌های قدرتمند تشخیص متن، برای زبان‌هایی با خط شکسته مانند اردو، همواره چالش برانگیز بوده است. از جمله مشکلات این زبان‌ها، پیچیدگی ذاتی و عدم وجود داده‌هایی حاوی اطلاعات کافی است. در مرجع [۴۴]، یک تکنیک تشخیص حروف، برای زبان اردو و فونت نستعلیق ارائه شده است. این روش، براساس تشخیص حروف و نوع اتصالات آن‌ها است و اثربخشی معماری ترکیبی CNN-RNN در تشخیص متن اردو از اسناد چاپی را به خوبی به نمایش می‌گذارد. مجموعه داده استفاده شده در این مقاله، مجموعه داده چاپی به نام UPTI است. هم‌چنین، آن‌ها یک مجموعه داده جدید ارائه دادند. این مجموعه داده، تصاویر اسکن شده نشریات چاپ شده اردو در فونت‌های مختلف است. دقت تشخیص حروف در این دو مجموعه داده، به ترتیب ۹۵،۵۸ درصد و ۹۸،۳۷ درصد است. به علاوه، در مرجع [۴۵]، برای یادگیری ویژگی‌ها و طبقه‌بندی فونت نستعلیق در زبان اردو، از ترکیب شبکه‌های عصبی کانولوشن و شبکه‌های عصبی مکرر استفاده شده است. مجموعه داده استفاده شده در این مقاله نیز مجموعه داده چاپی UPTI بوده است. دقت تشخیص در روش پیشنهادی بر روی این مجموعه داده، ۹۸،۱۲ درصد است.

---

<sup>۱</sup> Overfitting



## ۲-۴-۲ کاربردهای سیستم‌های OCR

سیستم بانکی در سراسر جهان از وابستگی شدید به نیروی انسانی و اسناد کتبی رنج می‌برد. بدین ترتیب، فرآیندهای بانکداری متداول، بسیار خسته‌کننده و وقت‌گیر هستند. روش‌های موجود برای تراکنش‌های مرتبط با چک، باعث می‌شود که پردازش به تعویق بیفتد، زیرا جزئیات تراکنش‌ها باید به صورت دستی وارد کامپیوتر شوند. بنابراین، نیاز به وجود سیستم‌های OCR، در زمینه‌های مختلف مانند ورود و تشخیص داده‌ها حس می‌شود. هدف این سیستم‌ها، ترکیب تکنیک‌های یادگیری ماشین، برای بهبود فرآیندهای بانکی است که می‌تواند از طریق پردازش چک به صورت خودکار انجام شود. در مرجع [۳]، روشی برای تشخیص حروف دست‌نویس مبتنی بر شبکه عصبی کانولوشن دوبعدی ارائه شده است. در این روش، تشخیص الگو با استفاده از یادگیری ماشین انجام شده است. مجموعه داده استفاده شده در این مقاله، مجموعه داده استاندارد EMNIST است.

یکی از کاربردهای مهم سیستم OCR، تشخیص پلاک‌های خودرو است. برای شناسایی شماره پلاک از روش‌های مختلفی بهره گرفته شده است. در مرجع [۷]، ساختار جدیدی برای یک طبقه‌بند شبکه عصبی کانولوشن، برای تشخیص کاراکترهای شماره پلاک خودروهای ایرانی، از تصویر پلاک پیشنهاد شده است. آن‌ها، ساختار پیشنهادی خود را با حدود ۲۲,۴۰۰ تصویر ارزیابی کردند و به دقت ۹۹,۷۳ درصد دست یافته‌اند. در مرجع [۸]، با استفاده از شبکه عصبی هاپفیلد<sup>۱</sup>، روشی برای تشخیص پلاک خودرو در یک تصویر ویدئویی پیشنهاد شده است. در مرجع [۲۵]، با استفاده از دو ابزار تزرتک و Transym به استخراج پلاک خودرو پرداخته شده است. نتایج بدست آمده در این تحقیق نشان می‌دهد که ابزار تزرتک در استخراج شماره پلاک خودرو نسبت به Transym قوی‌تر بوده است، در حالی که Transym در تشخیص متن از انواع دیگر تصاویر بهتر از تزرتک عمل کرده است.

---

<sup>۱</sup> Hopfield Neural Network

از کاربردهای نوین سیستم‌های OCR، می‌توان به استفاده از آن در تبدیل تصویر به متن و سیستم‌های تبدیل متن به گفتار، برای نابینایان اشاره کرد. در مرجع [۱۲]، با استفاده از سیستم‌های OCR، نابینایان می‌توانند از کتاب‌هایی که به صورت معمولی و نه خط بریل نوشته شده‌اند، بهره بگیرند. همچنین، در این مقاله، یک نمونه دستگاه معرفی شده که مراحل تصویر برداری از متن کتاب تا استخراج متن از تصویر و خواندن آن فرد نابینا از طریق بلندگو را به سهولت انجام می‌دهد.

تحلیل تصاویر با فونت‌های قدیمی و تشخیص متون حک شده در کتیبه‌های هخامنشی، از کاربردهای جالب توجه سیستم‌های OCR است. در مرجع [۴۶]، روشی برای استخراج این متون ارائه شده است. در این مقاله، از تکنیک‌های پردازش تصویر برای حذف نویز استفاده شده است. خروجی نهایی سیستم، شامل استخراج فونت باستانی و تبدیل تصویر کتیبه به متن قابل ویرایش برای زبان فارسی و انگلیسی است. در این مقاله از جعبه ابزار ترزکت، نسخه ۳،۰۴ استفاده شده و دقت به دست آمده ۹۲ درصد است. همچنین، در مرجع [۴۷]، برای تشخیص متون کتاب‌های چاپی قدیمی، از ترکیب خوشه‌بندی زیر کلمات و یک شبکه عصبی LSTM، به عنوان یک طبقه‌بند برای کاهش میزان خطا استفاده شده است. در روش پیشنهادی، خوشه‌بندی زیر کلمات با یک شبکه عصبی ترکیب شده‌اند و به عنوان یک طبقه‌بند جدید در نظر گرفته شده‌اند. این روش، نرخ خطا را تا ۰،۰۵ درصد کاهش داده است.

تشخیص سریع و خودکار آدرس گیرنده‌ی پاکت‌نامه، برای مرتب‌سازی پاکت‌نامه‌ها، موضوع مهمی برای ادارات پست است. در مرجع [۱۶]، یک روش سریع برای تشخیص آدرس گیرنده از روی تصویر پاکت نامه‌ی پستی ارائه می‌شود. در این روش، پس از یافتن بلوک آدرس گیرنده و جدا کردن خطوط آن از یکدیگر، بلوک حاوی کدپستی گیرنده پیدا می‌شود و تصویر ارقام از روی تصویر پاکت‌نامه جدا می‌شود. سپس ویژگی‌های استخراج‌شده از تصویر ارقام، جهت تشخیص کد پستی به شبکه عصبی داده می‌شود. در نهایت، از روی کد پستی بدست آمده و پایگاه‌داده موجود در اداره پست، به سرعت آدرس موجود روی پاکت‌نامه تشخیص داده می‌شود.

امروزه، تعداد زیادی تصویر در دنیای دیجیتال وجود دارد. این تصاویر ممکن است با دستگاه‌های مختلف، مانند اسکنر، دوربین و غیره تولید شده باشد. در این میان توجه کمتری به تصاویر حاصل از نگارفت<sup>۱</sup> موبایل شده است. در مرجع [۲۴]، یک سیستم تشخیص متن، از تصاویر دریافت شده از طریق یک تلفن همراه هوشمند ارائه شده است. در این مقاله، ابتدا یک پیش‌پردازش برای استخراج و بهبود ناحیه‌ی حاوی متن، بر اساس تجزیه و تحلیل مولفه‌های متصل ارائه شده است. سپس از یک شبکه عصبی مکرر دوطرفه<sup>۲</sup>، برای تشخیص متن استفاده شده است. مجموعه داده استفاده شده در این مقاله، مجموعه داده‌ای به نام ICDAR2015 است. نتایج تجربی نشان داده است که شبکه پیشنهادی دارای سرعت محاسباتی بیشتری نسبت به LSTM است که اغلب در سیستم OCR استفاده می‌شود.

## ۵-۲ جمع‌بندی

در این فصل به بررسی کارهای انجام شده قبلی در سیستم‌های OCR پرداخته شد. همان‌طور که اشاره شد، بازیابی متن از تصویر دارای مشکلاتی فراوانی مانند وجود اندازه‌ها و فونت‌های مختلف، پس‌زمینه پیچیده و غیره است. رفع این مشکلات در مراحل پیش‌پردازش انجام می‌شود. بنابراین، این مرحله در افزایش کارایی سیستم‌های OCR نقش مهمی دارد. از این رو، الگوریتم‌های مختلفی در این زمینه ارائه شده است. همچنین، سیستم‌های OCR، دارای کاربردهای مختلفی مانند پردازش چک‌های بانکی، استخراج پلاک خودرو، تبدیل تصویر به گفتار برای نابینان و غیره است. در این بخش، به کاربرد این نوع سیستم‌ها اشاره شده و برخی از مقالات مرتبط با آن کاربردها بررسی شده‌اند. جدول ۱-۲، روش‌های شرح داده شده در مقالات بررسی شده را به طور خلاصه مقایسه می‌کند.

در فصل بعد، سعی شده است تا برخی چالش‌های موجود در سیستم‌های OCR فارسی رفع شود. سپس، با یافتن پارامترهای بهینه و استفاده از آن‌ها در شبکه عمیق جهت آموزش سیستم‌های OCR، به دقت

---

<sup>۱</sup> Screenshot

<sup>۲</sup> Bidirectional Recurrent Neural Network

جدول ۱-۲. مقایسه روش‌های ذکر شده.

توضیحات			مرجع
کارهای مرتبط با زبان‌هایی دارای رسم‌الخط شکسته			
دقت (درصد)	روش	زبان	
مجموعه داده ۱: ۹۹/۱۹ مجموعه داده ۲: ۹۸/۱	RCNN	اردو	[۴۱]
۹۰-۹۸/۸	LSTM-RNN	اردو	[۴۲]
۹۰/۸۲	LSTM	کردی	[۱۷]
۹۸/۷۰	Gabor, PCA, PNN	فارسی	[۲۳]
۹۹/۴	CNN	عربی	[۳]
مجموعه داده ۱: ۹۵/۵۸ مجموعه داده ۲: ۹۸/۳۷	CNN-RNN	اردو	[۴۴]
۹۸/۱۲	RCNN	اردو	[۴۵]
کاربردهای سیستم‌های OCR			مرجع
دقت (درصد)	روش	نوع کاربرد	
۱۰۰	Nave Bayes	پردازش چک‌های بانکی	[۱۵]
۷۳	CNN	تشخیص پلاک خودرو	[۷]

۹۲	تکنیک‌های پردازش تصویر	تشخیص متن از کتیبه‌های هخامنشی	[۴۶]
----	---------------------------	--------------------------------------	------



## فصل ۳: روش مشاهدی

## ۳-۱ مقدمه

در فصل‌های پیش به تعریف مسئله پرداخته شد و در ادامه، کارهای انجام شده مورد بررسی قرار گرفت. در این فصل، روش پیشنهادی با جزئیات مورد بررسی قرار می‌گیرد. از این رو، ابتدا در بخش ۳-۲، الگوریتمی برای حذف کشیدگی نامتعارف موجود در تصاویری حاوی متون چاپی فارسی، با استفاده از یک حد آستانه ارائه می‌شود. سپس، در بخش ۳-۳، یک روش یادگیری عمیق برای آموزش سیستم OCR چاپی فارسی، با پارامترهایی دارای مقدار بهینه پیشنهاد می‌شود.

## ۳-۲ پیش‌پردازش

حذف کشیدگی نامتعارف حروف، یکی از پیش‌پردازش‌های لازم برای آماده‌سازی تصاویر ورودی برای سیستم‌های OCR فارسی است. در صورتی که این کشیدگی به درستی تشخیص داده و حذف نشود، فرآیند تشخیص را با چالش‌هایی روبرو می‌کند. عدم حذف این کشیدگی، موجب تشخیص یک "سین" اضافه یا یک گسستگی در میان حروف یک کلمه می‌شود. برای نمایش این خطاها، متونی مختلف، حاوی کلماتی با کشیدگی نامتعارف، برای بررسی تشخیص این نوع کشیدگی، به سیستم OCR داده شده است. دو نمونه از خروجی این سیستم، هنگام مواجهه با کلمه‌ی "است" دارای کشیدگی نامتعارف، در شکل ۳-۱ آورده شده است. در شکل ۳-۱-الف، حرف "سین" و در شکل ۳-۱-ب، یک گسستگی، بجای کشیدگی نامتعارف تشخیص داده شده است.

است است  
الف ب

شکل ۳-۱. نمونه‌ای از تشخیص کشیدگی نامتعارف توسط سیستم‌های OCR. الف) تشخیص یک "سین" اضافه بجای کشیدگی نامتعارف. ب) تشخیص یک گسستگی در کلمه بجای کشیدگی نامتعارف.



در این بخش، برای حل این چالش، یک روش مبتنی بر پردازش تصویر ارائه شده است، که از ترکیب دو ویژگی اصلی یک تصویر متنی، به نام‌های قطر عمودی خط پس‌زمینه و ارتفاع خط بهره گرفته است. هنگام نوشتن یک متن فارسی، تعداد زیادی از پیکسل‌های هر حرف، روی یک خط قرار می‌گیرند، که به آن خط پس‌زمینه گفته می‌شود. قطر این خط، یکی از ویژگی‌های مهم در تصاویر متنی است. این نکته حائز اهمیت است که قطر عمودی خط پس‌زمینه به اندازه و نوع فونت بستگی دارد و با افزایش اندازه قلم، اندازه آن نیز افزایش می‌یابد.

پیش از انجام فرآیند حذف کشیدگی نامتعارف حروف، باید مکان آن در کلمه مشخص شود. میزان کشیدگی نامتعارف حروف را می‌توان بر اساس دو پارامتر قطر عمودی خط پس‌زمینه و ارتفاع خط تخمین زد. در بسیاری از فونت‌ها مانند B Nazanin، می‌توان تخمین مناسبی از کشیدگی متعارف را به کمک پارامتر اول، یعنی قطر عمودی خط پس‌زمینه محاسبه کرد. اما در برخی از فونت‌ها مانند فونت Titr، تخمین این کشیدگی به کمک پارامتر اول میسر نیست و از این رو باید از پارامتر دوم، یعنی ارتفاع خط بهره گرفت. در هر صورت، حداقل تخمین حاصل از این دو پارامتر، به عنوان اندازه‌ی کشیدگی متعارف در نظر گرفته می‌شود.

لازم به ذکر است که تمام کشیدگی‌های موجود در یک متن فارسی، کشیدگی‌های نامتعارف نیستند. بسیاری از این کشیدگی‌ها جزو بخش اصلی کلمات بشمار می‌آیند. از این رو، باید کشیدگی‌های نامتعارف، از کشیدگی‌های متعارف تفکیک شوند. این تفکیک بر اساس اندازه‌ی کشیدگی صورت می‌پذیرد. کشیدگی‌هایی با اندازه‌های بالا، می‌توانند جزو کشیدگی‌های نامتعارف باشند.

برای پرهیز از اشتباه گرفتن کشیدگی‌های متعارف و نامتعارف، نیاز به تعریف یک حد آستانه وجود دارد. این حد آستانه، نشان دهنده حداکثر اندازه کشیدگی متعارف حروف است، بدین معنی که هر کشیدگی که از این حد آستانه بیشتر باشد، یک کشیدگی نامتعارف است. بنابراین، برای تشخیص کشیدگی‌های نامتعارف، تعیین میزان حد آستانه بسیار مهم و حیاتی است و عدم انتخاب مقدار مناسب برای آن، حذف کشیدگی‌های

نامتعارف را با چالش همراه می‌کند.

استفاده از ارتباط بین میزان کشیدگی حروف در تصاویری حاوی متونی با کشیدگی متعارف و دو پارامتر ذکر شده، تخمین کشیدگی متعارف را میسر می‌سازد. روش‌های تخمین حد آستانه با استفاده از دو پارامتر قطر عمودی خط پس‌زمینه و ارتفاع خط در بخش‌های ۱-۲-۳ و ۲-۲-۳ با جزئیات شرح داده می‌شوند.

### ۱-۲-۳ تخمین حد آستانه بر اساس قطر عمودی خط پس‌زمینه

همانطور که در بخش قبل توضیح داده شد، تعیین حد آستانه برای تشخیص نوع کشیدگی، یکی از مهم‌ترین مراحل انجام حذف کشیدگی نامتعارف است و در برخی از فونت‌ها، این حد آستانه، بر اساس قطر عمودی خط پس‌زمینه محاسبه می‌شود. در این بخش، به شرح الگوریتم ۱-۳ پرداخته شده است. این الگوریتم، حد آستانه را بر مبنای قطر عمودی خط پس‌زمینه محاسبه می‌کند.

الگوریتم ۱-۳. تخمین حد آستانه بر اساس قطر عمودی خط پس‌زمینه

۱. تبدیل تصویر خاکستری به سیاه و سفید (تصویر دودویی).
۲. اتخاذ مقدار صفر برای پس‌زمینه و مقدار یک برای پیکسل‌های متنی.
۳. محاسبه مجموع شدت روشنایی‌های عمودی در هر ستون از ماتریس تصویر و رسم نمودار پراکندگی آن.
۴. محاسبه هیستوگرام مقادیر حاصل از مرحله قبل.
۵. اتخاذ پر تکرارترین مقدار در هیستوگرام به عنوان قطر عمودی خط پس‌زمینه.
۶. تعیین مکان و مقدار کشیدگی بر اساس قطر عمودی خط پس‌زمینه.

شرح الگوریتم ۱-۳: تبدیل یک تصویر به تصویر دودویی، عملکرد الگوریتم‌های پردازش تصویر را تسهیل و بهبود می‌بخشد. از این رو، اغلب الگوریتم‌های پردازش تصویر، قبل از انجام هرگونه پردازشی، از این روش استفاده می‌کنند. بنابراین، یکی از مهم‌ترین کارها در مرحله پیش‌پردازش، دودویی کردن تصاویر است. در نتیجه در مرحله اول، برای ساده‌سازی پردازش‌های بعدی، تصویر خاکستری به تصویر سیاه و سفید

(دودویی) تبدیل شده و در مرحله دوم از این الگوریتم، تصویر معکوس می‌شود. بدین معنا که پیکسل‌های متنی مقدار صفر و پیکسل‌های پس‌زمینه مقدار یک را اتخاذ می‌کنند.

بدیهی است که هر تصویر می‌تواند به عنوان یک ماتریس در نظر گرفته شود. هر تصویر خاکستری یا سیاه و سفید را می‌توان به صورت یک ماتریس دو بعدی نمایش داد. آدرس هر درایه در این ماتریس، مختصات یک پیکسل از تصویر و مقادیر موجود در هر درایه، نشان دهنده شدت روشنایی تصویر در آن مختصات است.

در مرحله سوم، در هر ستون از ماتریس حاصل از تصویر، حاصل جمع شدت روشنایی‌های عمودی تصویر محاسبه می‌شود. در مرحله چهارم، نمودار هیستوگرام بر اساس این مقادیر رسم می‌شود. نمودار هیستوگرام، یکی از پرکاربردترین نمودارهای ستونی است و داده را به یک محدوده‌ای خاص محدود می‌کند. به عبارت دیگر، این نمودار، یک نگاه کلی به فراوانی داده‌های گروه‌بندی شده دارد. محور افقی نمودار هیستوگرام، گروه‌های موجود در مجموعه داده را نشان می‌دهد. محور عمودی در این نمودار، نشان دهنده مقدار فراوانی مرتبط با هر گروه است. نقطه اوج در نمودار هیستوگرام، نشان می‌دهد که کدام بازه از داده‌ها، دارای فراوانی بیشتری هستند. در مرحله پنجم نیز، به نقطه اوج نمودار هیستوگرام حاصل از حاصل جمع شدت روشنایی‌های عمودی تصویر نیاز است. در واقع، بنا به تعریف خط پس‌زمینه که در بخش ۳-۲ ارائه شد، تجمع پیکسل‌های متنی در این ناحیه بیشتر است. بنابراین، این مقدار به عنوان قطر عمودی خط پس‌زمینه لحاظ می‌شود.

بعد از محاسبه قطر عمودی خط پس‌زمینه، باید مکان کشیدگی‌های موجود در این تصویر مشخص شوند. از این رو، در مرحله ششم، مجدداً نیاز به استفاده از نمودار بدست آمده در مرحله سوم است. با توجه به این نمودار، بخش‌هایی که یک خط ممتد با مقداری ثابت و به اندازه قطر عمودی خط پس‌زمینه است، مکان کشیدگی‌های تصویر را نشان می‌دهد.

پس از تعیین مکان کشیدگی، در مرحله بعدی باید اندازه‌ی کشیدگی‌های مشخص شده در مرحله قبل محاسبه شود. همان‌طور که گفته شد، در برخی از فونت‌ها، میزان کشیدگی متعارف، متناسب با قطر عمودی خط پس‌زمینه است. در این مرحله، تصاویری با متون، اندازه‌ها و فونت‌های مختلف بررسی شده‌اند و میزان کشیدگی‌های موجود در آن‌ها، با ضرایب تناسب مختلف نسبت به قطر عمودی خط پس‌زمینه بررسی شده است. در نهایت، مقدار ۶ به عنوان ضریب تناسب بین میزان کشیدگی متعارف حروف و قطر عمودی خط پس‌زمینه در نظر گرفته شده است.

این نکته حائز اهمیت است که برای اجتناب از حذف کشیدگی‌های متعارف، مقدار این ضریب کمی بزرگ‌تر از میزان حاصل در بررسی‌های میدانی در نظر گرفته شده است. ارتباط بین کشیدگی‌های متعارف و قطر عمودی خط پس‌زمینه در رابطه (۱-۳) نشان داده شده است. در صورتی که اندازه‌ی یک کشیدگی کوچک‌تر از مقدار حد آستانه محاسبه شده توسط رابطه (۱-۳) باشد، این کشیدگی، یک کشیدگی متعارف و جزئی از بخش اصلی کلمه است و نباید حذف شود. در صورت حذف کشیدگی‌های متعارف، کلمه از بین می‌رود و تصویر خراب می‌شود. در غیر این صورت، اگر میزان کشیدگی برابر یا بیشتر از مقدار حد آستانه باشد، این کشیدگی به عنوان یک کشیدگی نامتعارف تشخیص داده می‌شود که باید حذف شود.

$$\text{حد آستانه} = 6 * \text{قطر عمودی پس‌زمینه} \quad (1-3)$$

شکل ۲-۳، فلوجارت الگوریتم ۱-۳ به همراه یک مثال را نشان می‌دهد. طبق الگوریتم ۱-۳، در ابتدا باید تصویر "است" به یک تصویر دودویی تبدیل شود. با مقایسه تصویر اول و دوم در شکل ۲-۳، واضح است که مقدار عددی پیکسل‌های متنی و پیکسل‌های پس‌زمینه عوض و تصویر متنی معکوس شده است. سپس، در تصویر سوم، حاصل جمع شدت روشنایی این تصویر در هر ستون محاسبه و رسم شده است. در تصویر بعدی، نمودار هیستوگرام حاصل از تصویر کلمه "است" رسم شده است. مرکز دسته‌ای که مقدار فراوانی آن ماکزیمم تعداد فراوانی در این نمودار است، عدد پنج است. این مقدار، نشان دهنده تجمع پیکسل‌های متنی در پنج سطر متوالی است، که در تصویر سوم به وضوح دیده می‌شود. سپس، قطر عمودی خط پس‌زمینه

بر روی تصویر دودویی کلمه "است" با یک فلش قرمز، نشان داده شده است. در نهایت، همان طور که در شکل می بینید، اندازه کشیدگی حرف "سین" در این کلمه، کمتر از ۶ برابر قطر عمودی خط پس زمینه است.



شکل ۳-۲. فلوجارت الگوریتم ۳-۱ به همراه یک مثال. تشخیص حد آستانه براساس قطر عمودی خط پس زمینه.

### ۲-۲-۳ تخمین حد آستانه بر اساس ارتفاع خط

به دلیل ضخامت زیاد خط پس‌زمینه در برخی فونت‌ها مانند فونت Titr، نمی‌توان از ویژگی قطر عمودی خط پس‌زمینه برای محاسبه حد آستانه استفاده کرد. به همین دلیل، در این قسمت، از یکی دیگر از ویژگی‌های مهم تصاویر متنی، به نام ارتفاع خط برای تعیین حد آستانه استفاده شده است. در این بخش ما الگوریتم ۲-۳ را شرح می‌دهیم. این الگوریتم، حد آستانه را بر مبنای ارتفاع خط در یک تصویر متنی محاسبه می‌کند.

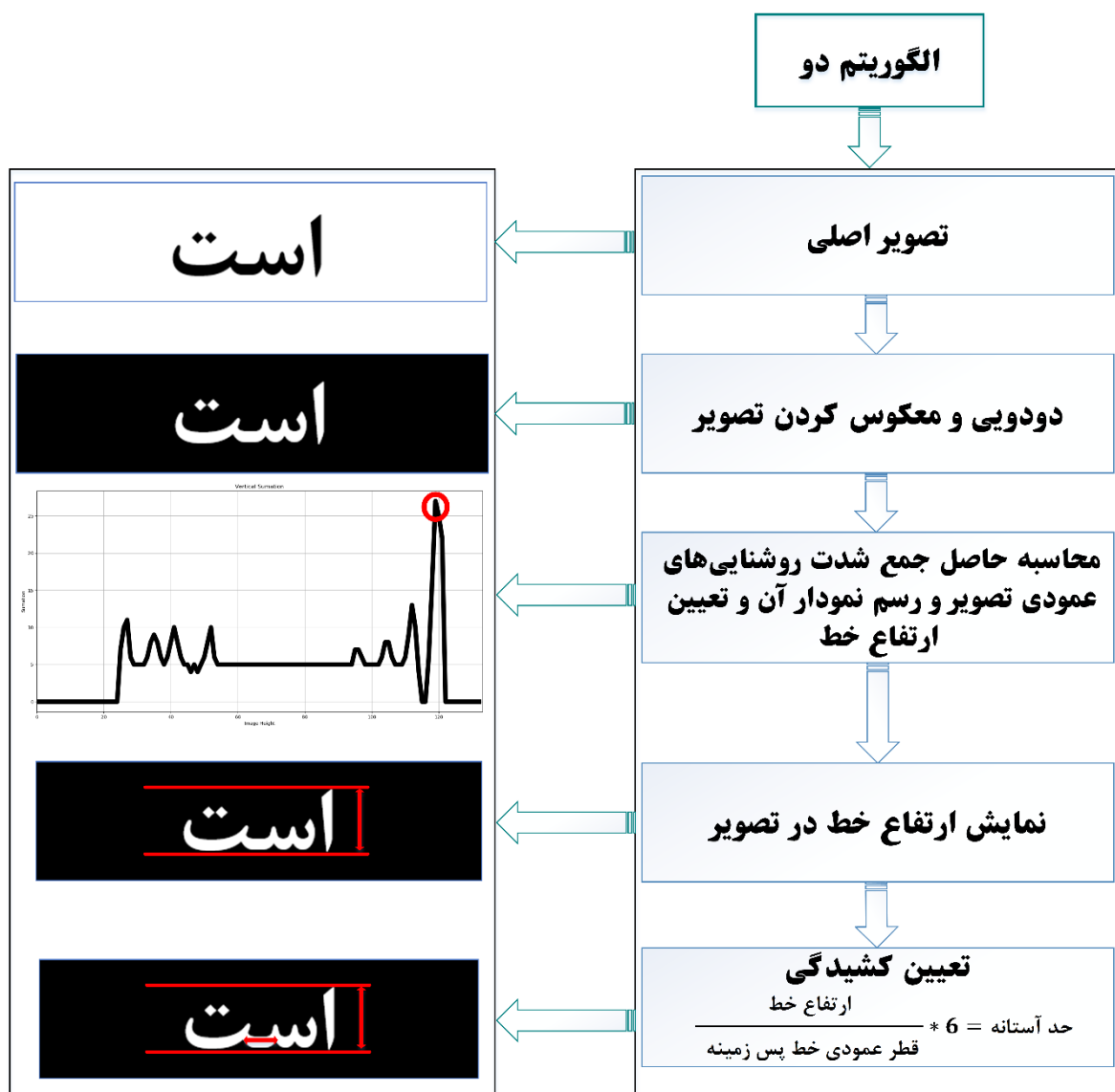
#### الگوریتم ۲-۳. تخمین حد آستانه بر اساس ارتفاع خط

۱. تبدیل تصویر خاکستری به سیاه و سفید (تصویر دودویی).
۲. اتخاذ مقدار صفر برای پس‌زمینه و مقدار یک برای پیکسل‌های متنی.
۳. محاسبه مجموع شدت روشنایی‌های عمودی در هر ستون از ماتریس تصویر و رسم نمودار پراکندگی آن.
۴. اتخاذ ماکزیمم نمودار مرحله ۳ به عنوان ارتفاع خط.
۵. تعیین مکان و مقدار کشیدگی بر اساس ارتفاع خط.

شرح الگوریتم ۲-۳: سه مرحله ابتدایی الگوریتم ۲-۳، همانند الگوریتم ۱-۳ است. مهم‌ترین مرحله در این الگوریتم، مرحله ۴، یعنی محاسبه ارتفاع خط است. در این الگوریتم، برای محاسبه ارتفاع خط، فاصله بین بالاترین و پایین‌ترین شدت روشنایی تصویر در هر ستون عمودی ماتریس تصویر محاسبه می‌شود و سپس، حداکثر این مقادیر به عنوان ارتفاع خط در نظر گرفته می‌شود. مرحله پنجم از الگوریتم ۲-۳، برای محاسبه مکان و اندازه کشیدگی، همانند مرحله ۶ در الگوریتم ۱-۳ عمل می‌کند. سپس، با بررسی تصاویری با متون، اندازه‌ها و فونت‌های مختلف، رابطه (۲-۳) برای محاسبه حد آستانه کشیدگی‌های متعارف در تصاویری با فونت‌هایی مانند Titr، استفاده شده است.

$$\text{حد آستانه} = 6 * \frac{\text{ارتفاع خط}}{\text{قطر عمودی خط پس‌زمینه}} \quad (2-3)$$

لازم به ذکر است که برای محاسبه قطر عمودی خط پس‌زمینه، از همان روش ذکر شده در الگوریتم ۱-۳ استفاده شده است. در صورتی که یک اندازه‌ی یک کشیدگی کوچک‌تر از مقدار حد آستانه محاسبه شده توسط رابطه (۲-۳) باشد، این کشیدگی، یک کشیدگی متعارف و جزئی از بخش اصلی کلمه است و نباید حذف شود، چرا که در صورت حذف آن، کلمه از بین می‌رود و تصویر خراب می‌شود. در غیر این صورت، اگر میزان کشیدگی برابر یا بیشتر از مقدار حد آستانه باشد، این کشیدگی به عنوان یک کشیدگی نامتعارف تشخیص داده می‌شود که باید حذف شود.



شکل ۳-۳. فلوچارت الگوریتم ۲-۳ به همراه یک مثال. تشخیص حد آستانه براساس ارتفاع خط.

شکل ۳-۳، فلوجارت الگوریتم ۲-۳ به همراه یک مثال را نشان می‌دهد. همانند الگوریتم ۱-۳، در ابتدا باید تصویر "است" به یک تصویر دودویی تبدیل شود. سپس، تصویر معکوس می‌شود که این عملیات در تصویر ۱ و ۲ نشان داده شده است. در تصویر سوم، حاصل جمع شدت روشنایی این تصویر در هر ستون محاسبه و رسم شده است. ماکزیمم نقطه این نمودار به عنوان ارتفاع خط در نظر گرفته می‌شود که در تصویر سوم در شکل ۳-۳، با یک دایره قرمز رنگ مشخص شده است. سپس، در تصویر چهارم، ارتفاع کلمه "است" با یک فلش قرمز رنگ مشخص شده است. در نهایت، در تصویر پنجم، طبق رابطه (۲-۳)، حد آستانه کشیدگی متعارف که بر اساس ارتفاع خط محاسبه شده است، روی حرف "سین" دارای کشیدگی نشان داده شده است. همان طور که می‌بینید، اندازه کشیدگی این حرف، حداکثر برابر اندازه حد آستانه است.

### ۳-۲-۳ حذف کشیدگی نامتعارف

پس از تعیین مقدار حد آستانه بر اساس الگوریتم‌های ۱-۳ و ۲-۳، مینیمم این دو مقدار، به عنوان مقدار نهایی حد آستانه در نظر گرفته می‌شود. سپس، باید کشیدگی‌های بزرگ‌تر از حد آستانه مشخص شوند و برای حذف کشیدگی نامتعارف، به الگوریتم حذف ارسال شوند.

#### الگوریتم ۳-۳. حذف کشیدگی نامتعارف

۱. استخراج دو پارامتر قطر عمودی خط پس‌زمینه و ارتفاع خط.
۲. اتخاذ حداقل مقدار دو پارامتر به عنوان حد آستانه.
۳. تعیین مقدار کشیدگی‌های بیشتر از حد آستانه (کشیدگی‌های نامتعارف).
۴. محاسبه یک‌دوم حد آستانه به عنوان  $x$ .
۵.  $x$  پیکسل از ابتدا و  $x$  پیکسل از انتهای کشیدگی حفظ و حذف ناحیه بین آن دو به کمک عملیات شیفت به راست تصویر.
۶. ذخیره‌ی تصویر جدید به عنوان یک تصویر بدون کشیدگی نامتعارف.

شرح الگوریتم ۳-۳: در الگوریتم ۳-۳، برای حذف کشیدگی‌های نامتعارف، در مرحله اول باید دو پارامتر قطر عمودی خط پس‌زمینه و ارتفاع خط، طبق الگوریتم‌های ۱-۳ و ۲-۳ استخراج شوند. سپس، در مرحله دوم، حداقل مقدار این دو پارامتر به عنوان حد آستانه در نظر گرفته می‌شود.



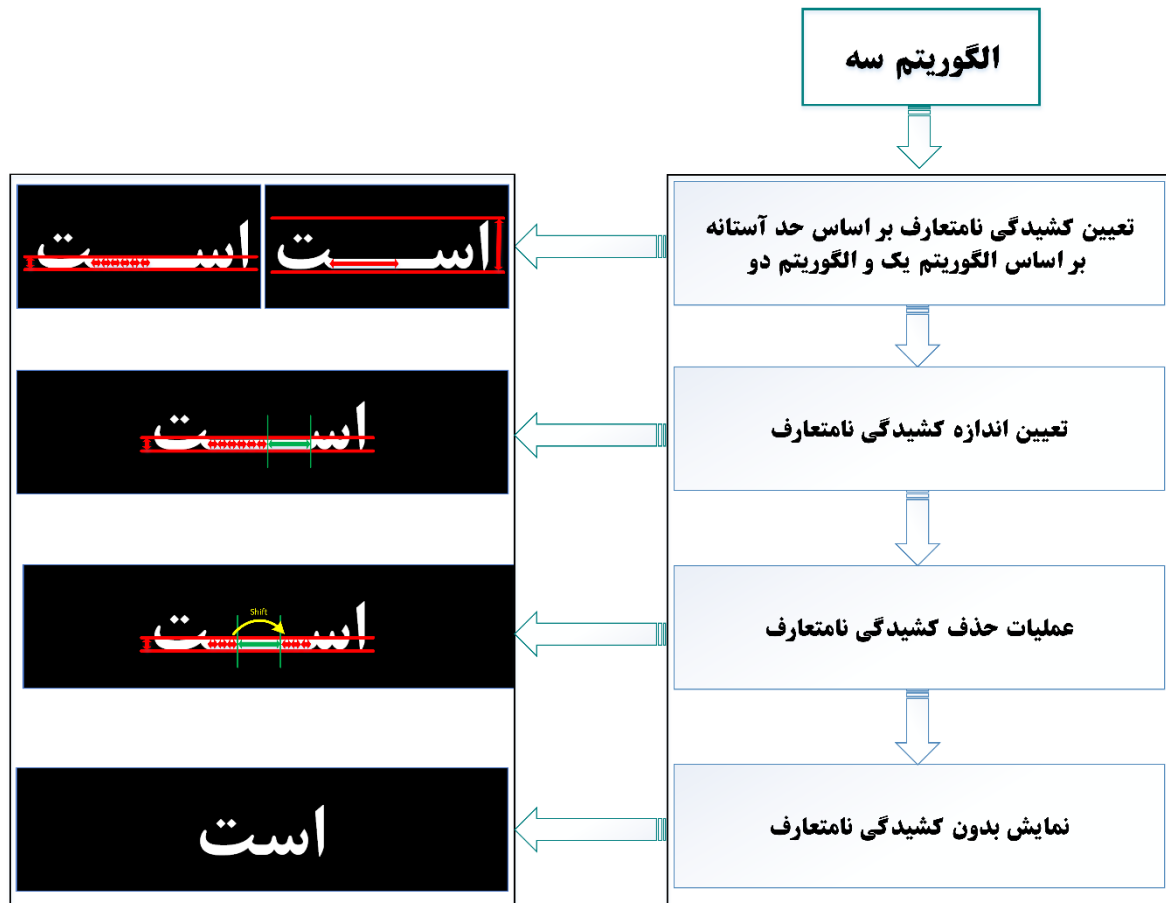
در مرحله سوم، برای تعیین کشیدگی‌های نامتعارف، مقدار کشیدگی‌های موجود در تصویر با حد آستانه مقایسه می‌شوند. در صورتی که اندازه کشیدگی بیشتر از حد آستانه باشد، آن کشیدگی به عنوان یک کشیدگی نامتعارف در نظر گرفته می‌شود و برای حذف، وارد مرحله بعد می‌شود.

حذف کشیدگی نامتعارف از میانه ناحیه کشیده شده در تصویر اتفاق می‌افتد. چون با حذف کشیدگی از ابتدا یا انتهای حرف کشیده شده، ممکن است بخشی از اجزای اصلی حرف مانند دندان یا نقطه (نقاط) آن حذف شود. بنابراین، بهتر است که حذف کشیدگی نامتعارف از وسط حرف انجام شود. برای این منظور  $x$  را برابر مقدار حد آستانه و  $y$  را برابر مقدار کشیدگی نامتعارف در نظر بگیرید. از آنجایی که اندازه هر کشیدگی پس از انجام فرآیند حذف، باید برابر حد آستانه ( $x$ ) باشد، بنابراین باید به اندازه  $y - x$  از کشیدگی نامتعارف حذف شود. برای روشن شدن این مطلب، ما کشیدگی نامتعارف را به سه ناحیه تقسیم می‌کنیم. ناحیه ابتدایی و انتهایی، دارای تعداد  $x$  پیکسل و تعداد پیکسل‌های ناحیه میانی برابر  $y - x$  است. با توجه به توضیحات داده شده، روشن است که ناحیه‌ای که باید حذف شود، ناحیه میانی است. در نهایت، این ناحیه با استفاده از عملیات شیفت به راست تصویر حذف می‌شود و ناحیه ابتدایی و انتهایی به هم متصل می‌شوند.

در پایان، تصویر جدید به عنوان یک تصویر با کشیدگی‌های متعارف ذخیره شده و جهت آموزش سیستم OCR مورد استفاده قرار می‌گیرد.

فلوچارت الگوریتم ۳-۳ به همراه یک مثال، در شکل ۳-۴ نشان داده شده است. همان‌طور که در شکل ۳-۴ نشان داده شده است، در ابتدا با توجه به نوع فونت، باید حد آستانه بر اساس یکی از دو رابطه (۳-۱) و (۳-۲) محاسبه شده و کوچکترین حد آستانه بدست آمده از این دو معادله به عنوان حد آستانه نهایی در نظر گرفته شود. در اینجا، حد آستانه بر اساس رابطه (۳-۱) و ویژگی قطر عمودی خط پس‌زمینه محاسبه شده است. سپس، اندازه کشیدگی موجود در حرف "سین" در کلمه "است"، با حد آستانه بدست آمده،

یعنی شش برابر قطر عمودی خط پس‌زمینه مقایسه می‌شود و همان‌طور که ملاحظه می‌شود، کشیدگی از این مقدار بیشتر است.



شکل ۳-۴. فلوجارت الگوریتم ۳-۳ به همراه یک مثال. حذف کشیدگی نامتعارف.

مقدار کشیدگی نامتعارف که در این حرف با یک فلش سبز رنگ نشان داده شده است باید حذف شود. طبق توضیحات الگوریتم ۳-۳، حذف این کشیدگی باید از وسط حرف اتفاق بیفتد. در تصویر سوم، منطقه‌ای که توسط الگوریتم ۳-۳ برای حذف مشخص شده است، با رنگ سبز نشان داده شده است که بعد از حذف آن، با عملیات شیفت، دو ناحیه ابتدایی انتهایی کشیدگی به هم می‌چسبند. در نهایت، تصویر آخر، کلمه "است" با یک کشیدگی متعارف است.

## ۳-۳ آموزش سیستم OCR

تصاویر بسیاری حاوی متون چاپی در دنیای دیجیتال وجود دارد. همه این صفحات می‌توانند ورودی سیستم‌های OCR باشند. عملکرد این سیستم‌ها وقتی که با طیف وسیعی از انواع فونت‌ها، اندازه‌ها و کیفیت‌های چاپ روبرو می‌شوند، بسیار دقیق‌تر می‌شود. از این رو، سیستم ارائه شده باید نسبت به این موارد مقاوم باشد. برای مقاوم‌سازی سیستم در برابر فونت‌های مختلف، از آموزش هر فونت به صورت مجزا یا آموزش چندین فونت با هم استفاده می‌شود. در سیستم‌های مبتنی بر یادگیری عمیق، غالباً از سیستم‌هایی با آموزش چندین فونت با هم استفاده می‌شود. با این عمل، نیازی به تشخیص نوع فونت نیست و خطای تشخیص نوع فونت از سیستم حذف می‌شود. از این رو، استفاده از سیستم‌هایی با آموزش چندین فونت با هم، از یک مزیت عمده‌ی مقاوم‌سازی در برابر نوع فونت برخوردار هستند. سیستم ارائه شده در این مقاله، به صورت چندین فونت باهم، آموزش داده شده است. هم‌چنین، برای مقاوم‌سازی سیستم در برابر اندازه‌های مختلف، از مجموعه‌داده‌ای با اندازه‌های متفاوت و رایج در مقالات و کتاب‌ها استفاده شده است. آموزش OCR توسط تزرکت، شامل سه مرحله است که در ادامه شرح داده شده‌اند. مرحله اول، تولید مجموعه‌داده آموزشی است. فایل متنی تهیه شده، به تصاویری با اندازه‌ها و فونت‌های متفاوت تبدیل می‌شود. روش‌ها و نرم‌افزارهای زیادی مانند وب‌سایت‌های برخط، عکس گرفتن از صفحه نمایش و غیره برای تبدیل متن به تصویر وجود دارد. در این تحقیق، از جعبه‌ابزار تزرکت برای تبدیل متن به تصویر استفاده شده است. ورودی تزرکت برای تولید این تصاویر، یک فایل متنی، مسیر فونت‌ها، تعیین زبان ورودی، مسیر فایل لازم جهت آموزش OCR و اندازه موردنظر است. خروجی این مرحله از تزرکت، به ازای هر فونت، شامل تصاویری با فرمت Tiff، فایلی با فرمت box، فایلی با فرمت lstmf، فایلی متنی حاوی مسیر فایل‌های lstmf ایجاد شده، فایلی با فرمت fas.traineddata و یک فایل با فرمت unicharset حاوی حروف الفبا، اعداد و علائم زبان فارسی است. در فایل box، مختصات مکان هر حرف موجود در تصاویر مشخص شده است. از آنجایی که شبکه LSTM، از یک مجموعه داده ورودی به همراه برچسب آن‌ها، تغذیه می‌کند، در اینجا

تصاویر Tiff تولید شده به عنوان ورودی و محتوای فایل box به عنوان برچسب ورودی، محسوب می‌شوند. فایل lstmf، ترکیبی از دو خروجی تصاویر و فایل‌های box است که در روند آموزش، بجای تصاویر و فایل box، از این فایل استفاده می‌شود. فایل fas.traineddata ترکیبی از چندین فایل ایجاد شده مرتبط با فایل متنی ورودی است. در نسخه‌های قبلی تزکت، تمامی فایل‌های آموزشی مورد نیاز برای ایجاد فایل fas.traineddata و همچنین فایل box، باید به صورت جداگانه ساخته می‌شدند. درحالی که در تزکت نسخه ۱،۰،۰،۴، این فایل‌ها با یک دستور و به صورت خودکار ساخته می‌شوند. همچنین در این نسخه، در صورت وجود کاراکترهای غیرمجاز، هنگام تولید تصاویر در متن داده‌های آموزشی، خطا اعلام می‌شود. در نهایت، مرحله اول، مجموعه داده تولید شده را به حدود چهار میلیون تصویر تبدیل می‌کند. هر تصویر، به طور متوسط شامل ۴۰ خط و هر خط حاوی حداکثر ۲۰ کلمه است.

مرحله دوم، آموزش OCR است. ورودی این مرحله، فایل fas.traineddata تولید شده در مرحله قبل، ساختار شبکه، مسیر ذخیره خروجی، فایل متنی حاوی مسیر فایل‌های lstmf تولید شده در مرحله قبل، نرخ یادگیری و تعداد تکرار است. در روند آموزش، با افزایش دقت در هر بار تکرار، یک فایل آموزشی ایجاد می‌شود. تعداد این فایل‌ها وابسته به تعداد افزایش یا کاهش در تغییر دقت است و این فایل‌ها حاوی اطلاعاتی از فرآیند آموزش تا زمان ساخت خود هستند. روند آموزش طولانی است و گاهی بیش از یک روز طول می‌کشد. با ایجاد این فایل‌های آموزشی، در صورت قطع روند آموزش، نیاز به آموزش مجدد نیست. مرحله سوم، مرحله تولید خروجی نهایی است. در این مرحله، یک فایل جدید fas.traineddata براساس دو مرحله‌ی پیش ایجاد و فرآیند آموزش، نهایی می‌شود. با تکمیل روند آموزش، امکان ارزیابی سیستم OCR فارسی وجود دارد. ارزیابی، با جایگزینی این فایل با فایل fas.traineddata موجود در تزکت انجام می‌شود.

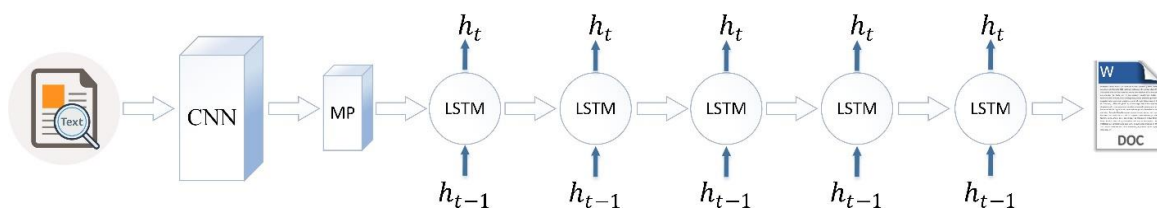
### ۳-۳-۱ مدل یادگیری عمیق پیشنهادی

در این بخش، مدل مبتنی بر یادگیری عمیق برای آموزش سیستم OCR فارسی ارائه شده است. این مدل، بر اساس دو روش شبکه‌های عصبی کانولوشن و شبکه‌های عصبی مکرر عمل می‌کند. نسخه توسعه یافته مدل یادگیری عمیق در تزرکت، نسخه ۴,۰,۰,۱ است. از این مدل برای آموزش سیستم OCR فارسی استفاده می‌شود. در این ساختار، ابتدا ویژگی‌های تصاویر با استفاده از لایه کانولوشن استخراج می‌شوند و سپس، این ویژگی‌ها، شبکه‌های عصبی مکرر را تغذیه می‌کنند. در نهایت، یک لایه softmax که همان لایه خروجی است، خروجی شبکه‌های عصبی مکرر را به کلاس‌های موجود طبقه‌بندی می‌کند.

شبکه موجود در نسخه تزرکت، دارای یک لایه کانولوشن با ۱۶ نرون خروجی با کرنل  $3 \times 3$ ، یک لایه Max Pooling با کرنل  $3 \times 3$  و ۴ لایه LSTM به ترتیب با ۴۸، ۹۶، ۹۶ و ۲۵۶ سلول مخفی است. در لایه کانولوشن این مدل، از تابع فعال‌ساز تانژانت هیپربولیک استفاده شده است. طبقه‌بند استفاده شده در لایه آخر شبکه آن‌ها، softmax است. تعداد کلاس‌های خروجی این مدل، به تعداد کاراکترهای موجود در مجموعه حروف و علائم زبان فارسی، یعنی ۱۱۰ است. همچنین، نرخ یادگیری شبکه عصبی پیشنهادی  $20 \times e - 4$  است.

ساختار مدل پیشنهادی در این پایان‌نامه، تا حدودی شبیه به ساختار شبکه استفاده شده در تزرکت است. تفاوت‌های شبکه پیشنهادی در این پایان‌نامه با شبکه موجود در تزرکت، تابع فعال‌ساز در لایه کانولوشن، تعداد لایه‌های LSTM استفاده شده و تعداد سلول‌های مخفی در لایه‌های LSTM است.

شبکه پیشنهادی در این پایان‌نامه از تصاویر متون فارسی تغذیه می‌کند و متن تشخیص داده شده را به عنوان خروجی برمی‌گرداند. این ساختار شامل هفت لایه است. در میان این لایه‌ها، یک لایه کانولوشن دو بعدی، یک لایه Max Pooling و پنج لایه LSTM است. این معماری در شکل ۳-۵ نشان داده شده است.



شکل ۳-۵. معماری شبکه عصبی عمیق پیشنهادی برای آموزش سیستم OCR فارسی.

ورودی لایه کانولوشن دویبعدی، تصاویر تهیه شده توسط تزرکت در مرحله اول است. این لایه، ۱۶ نقشه ویژگی استخراج و به لایه Max Pooling ارسال می‌کند. سپس، با استفاده از این لایه، تعداد ویژگی‌های استخراج شده کاهش می‌یابد. بدین صورت که، این لایه با یک کرنل  $3 \times 3$  با گام یک روی هر تصویر حرکت می‌کند و از بین آن ماتریس  $3 \times 3$ ، یک ویژگی را انتخاب می‌کند. در این روش، ما برای ساخت شبکه عصبی مکرر از پنج لایه LSTM استفاده کرده‌ایم. لایه‌های LSTM، از خروجی لایه Max Pooling تغذیه می‌کنند. سپس، این لایه‌ها، به دلیل داشتن سلول حالت، برای استخراج ویژگی از متون بکار می‌روند و با مهندسی ویژگی خودکار خود، برای هر داده متنی یک سری ویژگی استخراج می‌کنند. در نهایت، این اطلاعات به لایه تماماً متصل softmax جهت طبقه‌بندی ارسال می‌شوند و این لایه، آن‌ها را به ۱۱۰ کلاس موجود طبقه‌بندی می‌کند. جدول ۳-۱، لایه‌های شبکه پیشنهادی را به طور خلاصه شرح می‌دهد.

### ۳-۱ جمع‌بندی

در این فصل، روش پیشنهادی و نوآوری این پایان‌نامه مورد بررسی قرار گرفت. روش پیشنهادی شامل دو بخش پیش‌پردازش و آموزش سیستم است. برای بخش پیش‌پردازش، الگوریتمی ارائه شد که همه کشیدگی‌های نامتعارف موجود در تصاویر متنی را بر اساس دو پارامتر اصلی یک تصویر متنی، پیدا می‌کرد. این پارامترها، قطر عمودی خط پس‌زمینه و ارتفاع خط است. سپس، یک الگوریتم جهت حذف کشیدگی‌های نامتعارف یافت شده، ارائه شد. سپس، یک مدل مبتنی بر یادگیری عمیق با پارامترهای بهینه، جهت آموزش سیستم OCR فارسی ارائه شد.

جدول ۱-۳. شرح لایه‌های شبکه پیشنهادی.

توضیحات	لایه
۱۶ نرون خروجی - کرنل $3 \times 3$ ، تابع فعال‌ساز $\tanh$	کانولوشن
کرنل $3 \times 3$	MaxPooling
۳۲ سلول مخفی	LSTM
۶۴ سلول مخفی	LSTM
۱۲۸ سلول مخفی	LSTM
۲۵۶ سلول مخفی	LSTM
۵۱۲ سلول مخفی	LSTM
لایه $\text{softmax}$ با ۱۱۰ نرون خروجی	خروجی





## فصل ۴: نتایج و آزمایش‌ها

## ۴-۱ مقدمه

در این فصل، عملکرد روش پیشنهادی ارزیابی شده است. در ابتدا نحوه کار الگوریتم حذف کشیدگی نامتعارف مورد بررسی قرار گرفته است. سپس، آزمایش‌های مختلفی بر روی مدل پیشنهادی، مانند تاثیر تغییر پارامترهای مختلف، بررسی اندازه‌های مختلف متون فارسی و بررسی انواع فونت‌های مختلف در حین آموزش انجام شده است. همچنین، در این بخش، مقیاس‌پذیری OCR پیشنهادی بر روی اندازه‌های مختلف و به ویژه اندازه‌های آموزش داده نشده نیز بررسی شده است. در نهایت، تاثیر الگوریتم پیش‌پردازش بر روی سیستم OCR فارسی، همراه با یک مثال مورد بررسی قرار گرفته است.

## ۴-۲ مجموعه داده

برای آموزش و ارزیابی نتایج بدست آمده، یک مجموعه داده جامع و استاندارد نیاز است. از آنجایی که مجموعه داده‌ای برای زبان فارسی وجود ندارد، یک مجموعه داده جهت ارزیابی‌های مختلف برای زبان فارسی در این پایان‌نامه تهیه شده است. برای ساخت این مجموعه داده، از متون فارسی درون وبسایت‌های مختلف استفاده شده است. همچنین، در ساخت آن، از پیکره همشهری نیز بهره گرفته شده است [۴۸]. پیکره همشهری دارای فرمت Unicode نیست و بیشتر کلمات عربی هستند. همچنین، در این مجموعه داده نیم‌فاصله‌ها رعایت نشده‌اند. بنابراین، ابتدا باید متون جمع‌آوری شده را پیش‌پردازش و سپس از آن برای ورودی مرحله اول تزرتکت، یعنی ساخت تصاویر استفاده کرد.

مجموعه داده پیشنهادی شامل همه انواع کلمه یعنی اسم، فعل، صفت، عدد و غیره است. جملات آن نیز از قواعد ساختاری و معنایی زبان فارسی پیروی می‌کنند. این مجموعه داده شامل ۵,۵۵۰,۰۶۳ کلمه و ۱۰۲,۵۳۶,۴۹۳ حرف است، که از این تعداد دارای ۳۵۰,۰۰۰ کلمه غیرتکراری است. از این مجموعه داده در هشت فونت مختلف و ۱۰ اندازه مختلف، جهت آموزش سیستم OCR فارسی تهیه شده است. سرانجام، به طور تصادفی، ۷۵ درصد از این مجموعه داده برای آموزش، ۱۵ درصد برای آزمون و ۱۰ درصد آن برای ارزیابی در نظر گرفته شده است. قسمتی از یک صفحه از این مجموعه داده، در شکل ۴-۱ نشان داده شده

است.

بند کوهک واقع در ۳۶ کیلومتری جنوب خاوری زابل وارد خاک ایران (مرز دو کشور) می شود و از این نقطه به دو شعبه سیستان و پریشان تقسیم می شود که شعبه سیستان وارد خاک ایران می گردد. هیرمند حوزه آبریز وسیعی را شامل شده که اندازه این حوزه ۱۵۰۰۰۰ کیلومتر مربع برآورد می گردد. آبدهی آن در ابتدای ورود به دشت در خاک افغانستان ۱۱۶۰۰ میلیون متر مکعب تخمین زده شده است که نیمی از این آبدهی قبل از رسیدن به مرز مشترک دو کشور در خاک افغانستان با استفاده

شکل ۱-۴. قسمتی از یک مجموعه داده پیشنهادی.

در این پایان نامه، برای پیاده سازی الگوریتم حذف کشیدگی نامتعارف پیشنهادی، از زبان برنامه نویسی پایتون نسخه ۳،۷ بهره گرفته شده است و برای آموزش سیستم OCR پیشنهادی از تزرکت نسخه ۴،۰،۰،۱ استفاده شده است. در این سیستم از بهینه ساز Adam و نرخ یادگیری  $10 \times e - 4$  استفاده شده است. برای تسهیل در فرآیند تولید مجموعه داده و آموزش سیستم OCR، ما از سرور HP ProLiant DL380 Gen9، که دارای یک پردازنده ۵۶ هسته ای و ۲۵۶ گیگابایت حافظه است، بهره گرفته ایم.

### ۳-۴ نتایج تشخیص و حذف کشیدگی نامتعارف

همانطور که در فصل اول توضیح داده شد، یک راه برای افزایش خوانایی و زیبایی متن، تراز کردن متن است. استفاده از این ویژگی، سبب ایجاد کشیدگی برخی در برخی از حروف می شود. این نوع کشیدگی، در هنگام تشخیص متن، یک نویز به حساب می آید. در این پایان نامه، برای حذف و رفع این نویز، الگوریتم هایی ارائه شد. در این بخش نتایج حاصل از تشخیص و حذف کشیدگی نامتعارف با این الگوریتم ها مورد بررسی قرار گرفته است.

در الگوریتم پیش پردازش حذف کشیدگی، مهم ترین نکته یافتن یک حد آستانه مناسب، جهت پیدا کردن کشیدگی های نامتعارف بود، که توسط رابطه های (۱-۳) و (۲-۳) بدست آمد. سپس، هر کشیدگی موجود که از این حد آستانه بیشتر بود، به عنوان یک کشیدگی نامتعارف در نظر گرفته و حذف می شد. شکل ۴-۲، نمونه ای از اعمال این الگوریتم است. شکل ۴-۲-الف و شکل ۴-۲-ب، تصاویری با کشیدگی های نامتعارف و به ترتیب با فونت های B Nazanin و Titr هستند.

هوش مصنوعی را باید عرصه پهناور تلاقی و ملاقات بسیاری از دانش ها، علوم و فنون قدیم و جدید دانست.

الف

**هوش مصنوعی را باید عرصه پهناور تلاقی و ملاقات بسیاری از دانش ها، علوم و فنون قدیم و جدید دانست.**

ب

هوش مصنوعی را باید عرصه پهناور تلاقی و ملاقات بسیاری از دانش ها، علوم و فنون قدیم و جدید دانست.

ج

**هوش مصنوعی را باید عرصه پهناور تلاقی و ملاقات بسیاری از دانش ها، علوم و فنون قدیم و جدید دانست.**

د

شکل ۴-۲. نمونه ای از اعمال الگوریتم پیشنهادی حذف کشیدگی. الف) یک متن با کشیدگی نامتعارف و فونت B Nazanin. ب) یک متن با کشیدگی نامتعارف و فونت Titr. ج) خروجی الگوریتم بعد از اعمال تصویر ۴-۲-الف. د) خروجی الگوریتم بعد از اعمال تصویر ۴-۲-ب.

رنگ قرمز برخی از حروف در شکل ۴-۲، نشان دهنده کشیدگی های نامتعارف است. همان طور که در بخش قبل توضیح داده شده، برای محاسبه حد آستانه جهت رفع کشیدگی های نامتعارف در شکل ۴-۲-الف، از ویژگی قطر عمودی خط پس زمینه و معادله (۱-۳) و برای محاسبه حد آستانه جهت رفع کشیدگی های

نامتعارف در شکل ۴-۲-ب، از ویژگی ارتفاع خط و معادله (۲-۳) استفاده شده است. خروجی این الگوریتم، بعد از حذف کشیدگی‌های نامتعارف در شکل‌های ۴-۲-ج و ۴-۲-د آورده شده است. با مقایسه شکل‌های ۴-۲-الف و ۴-۲-ب با شکل‌های ۴-۲-ج و ۴-۲-د می‌توان نتیجه گرفت که کشیدگی‌های قرمز رنگ، بعد از اعمال الگوریتم حذف کشیدگی نامتعارف پیشنهادی، کوچک‌تر شده‌اند و به کشیدگی‌هایی با اندازه متعارف تبدیل شده‌اند.

## ۴-۴ نتایج سیستم OCR پیشنهادی

یکی از پارامترهای تاثیرگذار ورودی سیستم ترزکت، جهت آموزش مدل، ساختار شبکه است. ساختار شبکه پیشنهادی در این پایان‌نامه، شامل یک لایه کانولوشن، یک لایه Max Pooling و پنج لایه LSTM است. لایه کانولوشن برای استخراج ویژگی‌های تصویر و لایه LSTM برای استخراج ویژگی‌های متنی درون تصاویر استفاده شده است. تعداد لایه‌های LSTM، می‌تواند یکی از مهم‌ترین پارامترهای موثر در آموزش مدل پیشنهادی باشد. به همین دلیل، تعداد مختلف این لایه برای آموزش شبکه پیشنهادی مورد بررسی قرار گرفته است و نتایج حاصل از آن در جدول ۴-۱ آمده است.

همان‌طور که در جدول ۴-۱ نشان داده شده است، با افزایش تعداد لایه‌های LSTM دقت شبکه افزایش یافته است، اما افزایش دقت، بعد از اضافه کردن لایه ششم بسیار ناچیز است. از طرفی، با افزایش تعداد این لایه‌ها، زمان لازم جهت آموزش به طور قابل توجهی افزایش می‌یابد. بنابراین، علت انتخاب تعداد پنج لایه LSTM برای مدل پیشنهادی، این است که زمان اجرای کمتر به دقت بالاتر ترجیح داده شده است.

جدول ۴-۱. مقایسه دقت بر اساس تعداد لایه‌های LSTM بکار گرفته در مدل آموزشی، بر اساس حرف و کلمه،

در تکرار ۱۵۰۰۰ام.

تعداد لایه‌های LSTM	دقت آموزش (%)	دقت آزمون (%)
---------------------	---------------	---------------

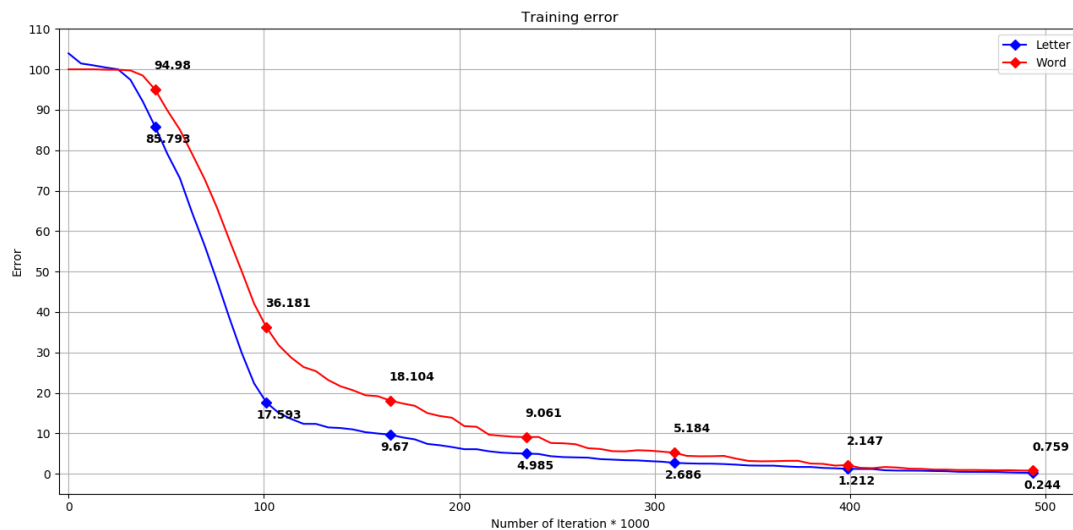
کلمه	حرف	کلمه	حرف	
۷۵,۳۰۹	۷۷,۳۳۱	۸۸,۸۷۷	۹۰,۸۱۹	۳
۸۰,۳۶۵	۸۳,۲۳۶	۸۶,۸۳۵	۹۰,۴۳۰	۴
۹۲,۶۱۲	۹۳,۱۲۶	۹۳,۶۲۵	۹۵,۲۵۳	۵
۹۲,۹۷۵	۹۳,۵۵۰	۹۴,۸۰۴	۹۷,۱۹۹	۶

بعد از یافتن تعداد بهینه لایه‌های LSTM، از پارامترهای تاثیرگذار دیگر بر روی دقت شبکه، می‌تواند تابع فعال‌ساز لایه کانولوشن است. جدول ۲-۴، تاثیر توابع فعال‌ساز مختلف بر روی دقت شبکه را نشان می‌دهد. همان‌طور که جدول ۲-۴ نشان می‌دهد، تابع فعال‌ساز Relu بهترین عملکرد و بالاترین دقت را به همراه دارد. بنابراین، تابع فعال‌ساز Relu برای لایه کانولوشن مدل پیشنهادی انتخاب شده است. جدول ۲-۴. مقایسه دقت بر اساس توابع فعال‌ساز مختلف به کار گرفته در لایه کانولوشن در مدل آموزشی، بر اساس حرف و کلمه، در تکرار ۱۵۰۰۰م.

دقت آموزش (%)		توابع فعال‌ساز
کلمه	حرف	
۹۳,۶۶۶	۹۳,۳۲۸	<b>Relu</b>
۹۵,۸۲۴	۹۳,۵۵۰	<b>Tanh</b>
۹۶,۷۳۵	۹۴,۷۱۲	<b>Sigmoid</b>
۹۸,۳۶۲	۹۸,۶۰۱	<b>Softmax</b>

نمودار خطای سیستم پیشنهادی در شکل ۳-۴ نشان داده شده است. طبق شکل ۳-۴، با افزایش تعداد تکرار مراحل آموزش، دقت سیستم نیز افزایش یافته است. با افزایش تعداد تکرارها، شیب نمودار به سمت

صفر میل کرده است. در حقیقت در تکرار پانصد هزارم، شیب خط خیلی به صفر نزدیک است. بنابراین، آموزش پایان یافته است.



شکل ۴-۳. نمودار خطای سیستم OCR فارسی چاپی.

زمانی که یک سیستم به صورت همزمان، چندین فونت را با هم آموزش می‌بیند، شبکه با شکل‌ها و ظاهرهای مختلفی از حروف و کلمات آشنا می‌شود. بنابراین، هنگام برخورد با کلمات با ظواهر مختلف، راحت‌تر و با دقت بیشتر آن‌ها را تشخیص می‌دهد. در نتیجه، دقت تشخیص سیستم هنگام آموزش همزمان چندین فونت با آموزش هر فونت به صورت مجزا، بیشتر خواهد بود. جداول ۴-۳ و ۴-۴، به ترتیب نتایج آموزش و آزمون سیستم پیشنهادی برای همه فونت‌ها و به ازای هر فونت را نشان می‌دهند.

جدول ۴-۳. دقت سیستم OCR پیشنهادی به ازای همه فونت‌ها، بر اساس حرف و کلمه، در تکرار ۱۵۰۰۰۰۰م.

دقت آزمون (%)		دقت آموزش (%)		همه فونت‌ها
کلمه	حرف	کلمه	حرف	
۹۸,۴۸۵	۹۹,۱۸۸	۹۹,۲۴۱	۹۹,۷۵۶	

همان‌طور که در جدول‌های ۳-۴ و ۴-۴ نشان داده شده است، دقت سیستم پیشنهادی بر روی هشت فونت مختلف به طور همزمان از دقت آموزش ۹۰ درصد فونت‌ها به صورت جداگانه بیشتر است.

جدول ۴-۴. دقت سیستم OCR پیشنهادی به ازای هر فونت، بر اساس حرف و کلمه، در تکرار ۱۵۰۰۰۰۰م.

دقت آزمون (%)		دقت آموزش (%)		نام فونت
کلمه	حرف	کلمه	حرف	
۹۹,۱۵۱	۹۹,۶۸۱	۹۹,۳۸۸	۹۹,۷۱۱	<b>Arial</b>
۹۸,۴۶۴	۹۹,۱۵۷	۹۸,۸۷۷	۹۹,۳۴۱	<b>Badr</b>
۹۸,۴۵۱	۹۹,۲۰۱	۹۸,۷۵۶	۹۹,۳۵۲	<b>Lutus</b>
۹۹,۰۲۷	۹۹,۳۷۹۸	۹۹,۱۵۶	۹۹,۵۱۵	<b>B Nazanin</b>
۹۸,۷۶۱	۹۹,۲۹۹	۹۸,۸۵۷	۹۹,۴۷۱	<b>Mitra</b>
۹۹,۰۲۹	۹۹,۷۳۳	۹۹,۲۷۹	۹۹,۷۷۶	<b>Teraffic</b>
۹۹,۰۶۶	۹۹,۵۶۴	۹۸,۸۲۱	۹۹,۷۲۴	<b>Titr</b>
۹۹,۸۳۶	۹۹,۱۸۴	۹۹,۰۳۷	۹۹,۳۴۶	<b>Homa</b>

مشابه با حالت آموزش چندین فونت، در صورتی که سیستم به طور همزمان چندین اندازه را آموزش ببیند، با اندازه‌های مختلف حروف و کلمات آشنا می‌شود و در برخورد با اندازه‌های مختلف متن، می‌تواند آن را به آسانی و با خطای کمتر تشخیص دهد. ما برای آموزش سیستم پیشنهادی، ابتدا از مجموعه داده‌ای شامل تنها یک اندازه استفاده کرده‌ایم. سپس، متونی با اندازه‌های بالاتر را به ترتیب به مجموعه داده افزوده و سیستم را دوباره آموزش دادیم. همان‌طور که از جدول ۴-۵ مشخص است، با افزایش تعداد اندازه‌ها، دقت سیستم در فاز آموزش بیشتر شده است.



جدول ۴-۵. دقت سیستم OCR پیشنهادی به ازای هر افزایش اندازه، بر اساس حرف و کلمه، در تکرار

۱۵۰۰۰۰۰ام.

دقت آموزش (%)		اندازه قلم
کلمه	حرف	
۹۹,۶۶۵	۹۹,۶۶۸	۱۰
۹۹,۸۲۸	۹۹,۷۸۱	۱۱-۱۰
۹۸,۰۳۲	۹۹,۸۳۹	۱۲-۱۱-۱۰
۹۸,۳۴۶	۹۹,۶۷۸	۱۳-۱۲-۱۱-۱۰
۹۸,۴۰۶	۹۹,۳۲۹	۱۴-۱۳-۱۲-۱۱-۱۰
۹۹,۵۹۷	۹۹,۴۶۹	۱۵-۱۴-۱۳-۱۲-۱۱-۱۰
۹۸,۷۹۹	۹۹,۵۰۹	۱۶-۱۵-۱۴-۱۳-۱۲-۱۱-۱۰
۹۸,۹۹۰	۹۹,۴۰۱	۱۸-۱۶-۱۵-۱۴-۱۳-۱۲-۱۱-۱۰
۹۹,۱۳۷	۹۹,۳۵۴	۲۰-۱۸-۱۶-۱۵-۱۴-۱۳-۱۲-۱۱-۱۰
۹۹,۲۰۵	۹۹,۱۲۶	۲۴-۲۰-۱۸-۱۶-۱۵-۱۴-۱۳-۱۲-۱۱-۱۰

سیستم OCR پیشنهادی در این پایان‌نامه، در برابر ویژگی اندازه متون، مقیاس‌پذیر است. بدین معنی که، این سیستم علاوه بر ارزیابی در برابر اندازه‌های آموزش داده شده، در ارزیابی با اندازه‌های آموزش داده نشده، نیز به خوبی عمل می‌کند. برای اثبات این قضیه، ما سه اندازه آموزش داده نشده ۹، ۱۷ و ۳۰ و یک اندازه آموزش داده شده ۱۳، را بررسی کردیم و نتایج در جدول ۴-۶ آمده است. همان‌طور که در این جدول می‌بیند، دقت سیستم OCR فارسی پیشنهادی در برابر این اندازه‌ها نیز معقول است.

جدول ۴-۶. دقت سیستم OCR پیشنهادی به ازای سه اندازه آموزش داده نشده و یک اندازه آموزش داده شده، بر اساس حرف و کلمه، در تکرار ۱۵۰۰۰۰۰ام.

دقت آزمون (%)		اندازه‌های مورد بررسی
حرف	کلمه	
۹۹,۷۳۸	۹۸,۱۹۰	۱۳ (آموزش داده شده)
۹۹,۷۳۱	۹۸,۱۶۰	۹
۹۷,۴۴۳	۹۸,۳۴۰	۱۷
۹۹,۶۳۵	۹۷,۹۸۰	۳۰

## ۵-۴ نتایج سیستم OCR پیشنهادی پس از اعمال الگوریتم حذف

### کشیدگی

یکی از پیش‌پردازش‌های انجام شده روی تصاویر موجود در مجموعه داده پیشنهادی، رفع کشیدگی‌های نامتعارف در آن‌ها بود. همان‌طور که در فصل اول توضیح داده شده، در صورتی که تصویری با این نوع کشیدگی به سیستم داده شود، سیستم به جای آن، یک "سین" اضافه و یا یک گسستگی در کلمه ایجاد می‌کند که این سبب تشخیص نادرست آن کلمه می‌شود. بنابراین، تشخیص و حذف کشیدگی نامتعارف از تصاویر با استفاده از الگوریتم‌های ۱-۳، ۲-۳ و ۳-۳، تاثیر بسزایی در عملکرد سیستم پیشنهادی داشته است.

نمونه‌ای از خروجی سیستم OCR پیشنهادی، در برابر تصاویر با کشیدگی نامتعارف، با اعمال و بدون اعمال الگوریتم پیش‌پردازش پیشنهادی جهت رفع این نوع کشیدگی، در شکل ۴-۴ آورده شده است. تصویر ۴-۴ الف، یک تصویر با کشیدگی نامتعارف است. در صورتی که این تصویر به سیستم OCR پیشنهادی داده شود، خروجی حاصل به صورت تصویر ۴-۴ ب خواهد بود. همان‌طور که می‌بینید در این تصویر، کلماتی

که دارای کشیدگی نامتعارف بوده‌اند، به درستی تشخیص داده نشده‌اند و این سبب کاهش دقت در سیستم OCR می‌شود. حال اگر تصویر ۴-۴-الف، ابتدا به الگوریتم پیش‌پردازش پیشنهادی برای حذف کشیدگی نامتعارف داده شود و سپس به سیستم OCR اعمال شود، الگوریتم تمام کشیدگی‌های نامتعارف موجود را حذف می‌کند و یک تصویر عادی با کشیدگی‌هایی متعارف، مانند شکل ۴-۴-ج به سیستم می‌دهد و سیستم متن درون آن را با کم‌ترین خطا بازیابی می‌کند.



شکل ۴-۴. یک نمونه از اعمال و عدم اعمال الگوریتم ۳-۳ در OCR. الف) یک متن با کشیدگی‌های نامتعارف.

ب) تشخیص OCR قبل از اعمال الگوریتم ۳-۳. ج) تشخیص OCR بعد از اعمال الگوریتم ۳-۳.

همچنین، دقت حاصل از سیستم OCR پیشنهادی، بعد از اعمال الگوریتم ۳-۳، در جدول ۴-۷ ارائه شده است. با مقایسه جدول ۴-۳ و جدول ۴-۷، می‌توان گفت که این الگوریتم، باعث افزایش دقت سیستم OCR پیشنهادی شده است.

جدول ۴-۷. دقت سیستم OCR پیشنهادی به ازای همه فونت‌ها و بعد از اعمال الگوریتم ۳-۳، بر اساس حرف و کلمه، در تکرار ۱۵۰۰۰۰۰ام.

دقت آزمون (%)		همه فونت‌ها
کلمه	حرف	
۹۹,۰۸۵	۹۹,۶۸۸	

یکی از معیارهای مطرح در ارزیابی، ماتریس درهم‌ریختگی<sup>۱</sup> است. جدول ۴-۶ نشان‌دهنده ماتریس درهم‌ریختگی سیستم OCR پیشنهادی، برای یک متن ده صفحه‌ای و با سه فونت مختلف است. در حالت کلی، در ماتریس درهم‌ریختگی، سطر و ستون باید شامل همه حروف زبان فارسی، اعداد و نشانه‌ها باشد که تعداد آن‌ها، ۱۱۰ عدد است. به منظور کاهش تعداد سطرها و ستون‌های این ماتریس، سطرها و ستون‌هایی که شامل کاراکترهایی با درصد تشخیص خطای صفر، از جدول حذف شده‌اند. با انجام این کار، هم حجم ماتریس کاهش یافته است و هم از حضور مقادیر تکراری (سطرها و ستون‌هایی با مقادیر صفر) جلوگیری شده است. بنابراین، در جدول ۴-۶، در این ماتریس، سطرها نشان‌دهنده کاراکترهای اصلی و ستون‌ها، نشان‌دهنده کاراکترهای تشخیص داده شده توسط سیستم است. اعداد درون هر درایه از این ماتریس، درصد خطای تشخیص کاراکتر سطر به جای کاراکتر ستون، توسط سیستم پیشنهادی است. درصد خطای تشخیص کاراکتر با استفاده از معادله (۴-۱) محاسبه می‌شود.

$$(۴-۱) \quad \text{تعداد تشخیص اشتباه سیستم} \\ \text{خطای تشخیص کاراکتر} = \frac{\text{تعداد تکرار کاراکتر در متن}}{\text{تعداد تکرار کاراکتر در متن}} \times 100$$

<sup>۱</sup> Confusion Matrix



ى	ف	م	ه	و	ا	[	ء	
.	.	.	.	.	.	.	.	٩
.	.	.	.	.	.	.	٤٩,٣٩٤	ع
.	٠,٤٤٧	.	٠,٠٨٩	.	.	.	.	غ
.	.	.	.	٠,٨٨٥	.	.	.	,
.	.	.	.	.	.	٨,٣٣	.	١
.	.	.	.	.	.	.	.	«
.	.	.	.	٨٤,٣٦٧	.	٨,٣٣	.	ث
.	.	.	.	.	.	.	.	٣
.	.	.	.	.	.	١٥,٣٨٥	.	(
.	.	.	.	.	.	.	.	٧
٠,٠٤٧	.	.	.	.	.	.	.	ت
.	.	.	.	.	١,٩٩١	.	.	0
.	.	.	.	.	.	.	.	٤
.	.	.	.	.	.	.	.	.
٠,٠٢٤	.	٠,٠٩٤	.	.	.	.	.	حذف

## فصل ۵: نتیجه گیری

## ۱-۵ مقدمه

در این فصل، الگوریتم پیش‌پردازش ارائه شده و مدل پیشنهادی برای آموزش سیستم OCR را جمع‌بندی و نقاط قوت و ضعف آن را بررسی می‌کنیم. همچنین پیشنهادهایی برای کارهای آتی ارائه می‌دهیم.

## ۲-۵ جمع‌بندی و نتیجه‌گیری

در این پایان‌نامه، یک الگوریتم پیش‌پردازش، به همراه یک نسخه توسعه یافته از سیستم OCR فارسی موجود در ترزکت نسخه ۴,۰,۰,۱ ارائه شد. روش پیشنهادی بر دو بخش پیش‌پردازش و آموزش سیستم OCR تمرکز دارد.

در بخش اول، یک الگوریتم برای حذف کشیدگی نامتعارف حاصل از تراز کردن متون، از تصویر متون فارسی پیشنهاد شده است که این پیش‌پردازش سبب افزایش دقت در عملکرد این سیستم‌ها می‌شود. این الگوریتم از تکنیک‌های پردازش تصویر بهره گرفته است. همچنین، برای تعیین حد آستانه جهت تشخیص کشیدگی نامتعارف در این الگوریتم‌ها، از ترکیب دو ویژگی اصلی تصاویر متنی، به نام‌های قطر عمودی خط پس‌زمینه و ارتفاع خط استفاده شده است.

در بخش دوم، دقت سیستم OCR به وسیله توسعه شبکه عمیق آن و ارائه یک مجموعه داده پیشنهادی افزایش یافت. به منظور آموزش سیستم OCR و بهبود عملکرد آن، تمرکز بر روی تغییر لایه فعال‌ساز لایه کانولوشن و تغییر تعداد لایه‌های LSTM در مدل پیشنهادی بود.

در این پایان‌نامه، برای آموزش و ارزیابی عملکرد سیستم پیشنهادی، به دلیل عدم وجود یک مجموعه داده جامع و استاندارد برای زبان فارسی، ما یک مجموعه داده جدید، شامل انواع کلمه مانند اسم، فعل، حرف، عدد و غیره با هشت نوع فونت در ۱۰ اندازه مختلف تهیه کردیم.

در نهایت، با اعمال الگوریتم پیشنهادی، استفاده از شبکه عمیق جدید و با بهره گرفتن از مجموعه داده جدید، ما توانستیم دقت سیستم OCR را به طور قابل توجهی افزایش دهیم.



## ۳-۵ پیشنهاد برای کارهای آتی

در این پایان‌نامه، فقط یک الگوریتم برای پیش‌پردازش ارائه شده است. پیش‌پردازش‌های دیگری مانند حذف نویز، حذف پس‌زمینه پیچیده، حذف اعراب وجود دارد که کمک شایانی به سیستم‌های تشخیص متن می‌کند. در کارهای آتی می‌توان با استفاده از این الگوریتم‌ها، دقت سیستم‌های OCR را افزایش داد. از دیگر کارهای آتی، می‌توان به آموزش یک سیستم OCR چندزبانه اشاره کرد. برای آموزش چنین سیستمی باید از یک مجموعه داده استاندارد و جامع، حاوی متونی چندزبانه بهره گرفت.

## مراجعه

- [1] Bhatia E.N. (2014) "Optical Character Recognition techniques: a review" **International Journal of advanced research in computer science and software engineering**, 4, 5.
- [2] Soomro W.J., Ismaili I.A., and Shoro G.M. (2018) "Optical Character Recognition System for Sindhi Text: a survey" **University of Sindh Journal of Information and Communication Technology**, 2, 2, pp81-87.
- [3] Ashiquzzaman A., Tushar A.K., Rahman A., and Mohsin F. (2019) "An Efficient Recognition Method for Handwritten Arabic Numerals using CNN with Data Augmentation and Dropout" **Data Management, Analytics and Innovation**, pp299-309.
- [4] Alkhateeb F. and Doush I.A. (2017) "Arabic Optical Character Recognition software: a review" **Pattern Recognition and Image Analysis**, 27, 4, pp763-776.
- [5] Chaudhuri A., Mandaviya K., Badelia P., and Ghosh S.K. (2017) "Optical Character Recognition systems" **Optical Character Recognition Systems for Different Languages with Soft Computing**, pp9-41.
- [۶] طلائی م.م و امامی م. (۱۳۹۶) "بهبود بازشناسی متن فارسی: ارائه یک روش تکمیلی بر پایه حذف نقاط حروف" سومین کنفرانس بین‌المللی بازشناسی الگو و تحلیل تصویر ایران، شهرکرد، ایران.
- [۷] مظاهری م. و برکتین ب. (۱۳۹۵) "بازشناسی کاراکترهای شماره پلاک خودروی ایرانی مبتنی بر شبکه عصبی پیچشی" کنفرانس بین‌المللی مهندسی و علوم کامپیوتر، نجف آباد، ایران.
- [۸] صمدی بهرامی ه. و جبرئیلی ع. (۱۳۹۴) "تشخیص و دنبال کردن پلاک خودرو در یک تصویر ویدئویی با استفاده از ویژگی رنگی HSV و OCR آن با شبکه عصبی هاپفیلد؛ به صورت بلادرنگ" سومین کنفرانس بین‌المللی پژوهش‌های کاربردی در مهندسی کامپیوتر و فن-آوری اطلاعات، تهران، ایران.
- [9] Roy A. and Ghoshal D.P. (2011) "Number Plate Recognition for use in different countries using an improved segmentation" **In 2011 2<sup>nd</sup> National Conference on Emerging Trends and Applications in Computer Science**, pp1-5.
- [10] Wen Y., Lu Y., Yan J., Zhou Z., von Deneen K.M., and Shi P. (2011) "An algorithm for license plate recognition applied to intelligent transportation system" **IEEE Transactions on Intelligent Transportation Systems**, 12, 3, pp830-845.
- [۱۱] صابری ر. و طوسی‌زاده س. (۱۳۹۵) "شناسایی موقعیت و استخراج اعداد کنتور گاز از تصاویر واقعی با استفاده از الگوریتم MSER" دومین همایش چشم‌انداز تکنولوژی کامپیوتر و شبکه در ۲۰۳۰، میبد، ایران.

[۱۲] صیادی ا. (۱۳۹۴) "کاربرد هوش مصنوعی و سیستم‌های OCR در خواندن متون برای نابینایان" کنفرانس بین‌المللی پژوهش‌های کاربردی در فناوری اطلاعات، کامپیوتر و مخابرات، تربت حیدریه، ایران.

[۱۳] عابدفرد س.، رفیعی ع. و اسعدی ا. (۱۳۹۴) "طبقه‌بندی و پس‌پردازش چک‌های بانکی ایرانی با استفاده از تجزیه کننده‌ی تصحیح خط" سومین کنفرانس ملی و اولین کنفرانس بین‌المللی پژوهش‌هایی کاربردی در مهندسی برق، مکانیک و مکاترونیک، تهران، ایران.

[۱۴] عابدفرد س. و رفیعی ع. (۱۳۹۴) "طبقه‌بندی چک‌های بانکی ایرانی با استفاده از روش K نزدیک‌ترین همسایگی" چهارمین کنفرانس ملی ایده‌های نو در مهندسی برق، اصفهان، ایران.

[15] Srivastava S., Priyadarshini J., Gopal S., Gupta S., and Dayal H.S. (2019) "Optical Character Recognition on Bank Cheques Using 2D Convolution Neural Network" **Applications of Artificial Intelligence Techniques in Engineering**, pp589-596.

[۱۶] برومندنیاع، یوسفی‌رامندی ه. و خزایی ک. (۱۳۹۲) "شناسایی الگوی کدپستی از روی تصویر پاکت‌نامه و تشخیص ارقام کدپستی از روس‌نامه‌های دست‌نویس فارسی" اولین همایش ملی کاربرد سیستم‌های هوشمند (محاسبات نرم) در علوم و صنایع، قوچان، ایران.

[17] Yaseen R. and Hassani H. (2018) "Kurdish Optical Character Recognition" **UKH Journal of Science and Engineering**, 2, 1, pp18-27.

[18] Bhunia A.K., Konwer A., Bhowmick A., Roy P.P., and Pal U. (2019) "Script identification in natural scene image and video frames using an attention based Convolutional-LSTM network" **Pattern Recognition**, 85, pp172-184.

[19] Hassanpour H., Samadiani N., and Akbarzadeh F. (2017) "A Modified Self-organizing Map Neural Network to Recognize Multi-font Printed Persian Numerals" **International Journal of Engineering**, 30, 11, pp1700-1706.

[20] Mirza N.M. (2018) "Printed Arabic Characters Recognition Based on Minimum Distance Classifier Technique" **Iraqi Journal of Science**, 59, 2A, pp762-770.

[21] Aranian M.J., Sarvaghad-Moghaddam M., and Houshmand M. (2017) "Feature dimensionality reduction for recognition of Persian handwritten letters using a combination of quantum genetic algorithm and neural network" **Majlesi Journal of Electrical Engineering**, 11, 2.

[22] Smith R. "An overview of the Tesseract OCR engine" **In Ninth International Conference on Document Analysis and Recognition (ICDAR)**, 2, pp629-633.

[23] Eghbali K., Veisi H., Mirzaie M., and Behbahani Y.M. (2017) "Font recognition for persian optical character recognition system" **In 10<sup>th</sup> Iranian Conference on Machine Vision and Image Processing (MVIP)**, pp252-257.

- [24] Bahi H.E. and Zatni A. (2017) "Segmentation and recognition of text images acquired by a mobile phones" **International Journal of Tomography Simulation**, **30, 4**, pp95-107.
- [25] Patel C., Patel A., and Patel D. (2012) "Optical character recognition by open source OCR tool tesseract: A case study" **International Journal of Computer Applications**, **55, 10**, pp50-56.
- [26] Mathew M., Singh A.K., and Jawahar C.V. (2012) "Multilingual OCR for indic scripts" **12th IAPR Workshop on Document Analysis Systems (DAS)**, pp186-191.

[۲۷] صالح پور م. (۱۳۹۱) "بازشناسی ارقام دست‌نویس فارسی مقاوم در برابر چرخش و تغییر مقیاس توسط طبقه‌بندی کننده svm فازی مبتنی بر خوشه‌بندی K-means" **نخستین کنفرانس بین‌المللی پردازش خط و زبان فارسی**، سمنان، ایران.

[۲۸] اکبری ی. جلیلی، فروزنده م.ج. و صدری ع.ج. (۱۳۹۱) "بررسی تاثیر ارتقا تصویر و اصلاح شیب در بهبود نرخ بازشناسی ارقام جدا شده از اسناد دست‌نویس فارسی" **نخستین کنفرانس بین‌المللی پردازش خط و زبان فارسی**، سمنان، ایران.

[۲۹] صدیقی ن.، سلیمانی ایوری م. و خسروی ع.ج. (۱۳۹۳) "کاهش ویژگی توسط توده ذرات دودویی برای بازشناسی ارقام دست‌نویس فارسی" **سیستم‌های هوشمند در مهندسی برق**، شماره ۱.

[۳۰] کریم‌زاده بجستانی س. و محمدی عنبران ع. (۱۳۹۵) "بازشناسی حروف مجزای فارسی با استفاده از تبدیل موجک و ماشین بردار پشتیبان فازی" **کنفرانس ملی دانش و فناوری علوم مهندسی ایران**، تهران، ایران.

[۳۱] سلطان‌زاده ح. و رحمتی م. (۱۳۸۲) "شناسایی ارقام دست‌نویس فارسی با استفاده از گرادیان و دسته‌بندی کننده ماشین بردار پشتیبان" **نهمین کنفرانس سالانه انجمن کامپیوتر ایران**، تهران، ایران.

[۳۲] یاسایی س. و حاتم ا. (۱۳۹۵) "انتخاب ویژگی‌های مؤثر در بازشناسی ارقام دست‌نویس فارسی با الگوریتم تکاملی ژنتیک" **اولین کنفرانس بین‌المللی دستاوردهای نوین پژوهشی در مهندسی برق و کامپیوتر**، تهران، ایران.

[۳۳] ستایش ح. (۱۳۸۹) "نیم‌فاصله و نقش آن در نگارش کلمات مرکب" **سبک‌شناسی نظم و نثر فارسی (بهار ادب)**، تهران، ایران.

- [34] Tafti A.P., Baghaie A., Assefi M., Arabnia H.R., Yu Z., and Peissig P. (2016) "OCR as a service: an experimental evaluation of Google Docs OCR, Tesseract, ABBYY FineReader, and Transym" **International Symposium on Visual Computing**, pp735-746.

- [35] Hochreiter S. and Jürgen S. (1997) "Long short-term memory" **Neural computation**, **9**, **8**, pp1735-1780.
- [36] Plötz, T. and Fink G.A. (2009) "Markov models for offline handwriting recognition: a survey" **International Journal on Document Analysis and Recognition (IJ DAR)**, **12**, **4**, pp269.
- [37] Gupta M.R., Jacobson N.P., and Garcia E.K. (2007) "OCR binarization and image preprocessing for searching historical documents" **Pattern Recognition**, **40**, **2**, pp389-397.
- [38] Zahour A., Likforman-Sulem L., Boussellaa W., and Taconet B. (2007) "Text line segmentation of historical arabic documents" **In Ninth International Conference on Document Analysis and Recognition (ICDAR)**, **1**, pp138-142.
- [39] Senior A.W. and Robinson A.J. (1996) "Forward-backward retraining of recurrent neural network" **In Advances in Neural Information Processing Systems**, pp743-749.
- [40] Patterson J. and Gibson A. (2017) "Deep learning: A practitioner's approach" **O'Reilly Media, Inc.**
- [41] Malik S. A., Maqsood M., Aadil F., and Khan M.F. (2019) "An Efficient Segmentation Technique for Urdu Optical Character Recognizer (OCR)" **In Future of Information and Communication Conference**, pp131-141.
- [42] Naseer A. and Zafar K. (2018) "Meta features-based scale invariant OCR decision making using LSTM-RNN" **Computational and Mathematical Organization Theory**, pp1-19.
- [43] Akram Q.U.A., Hussain S. (2017) "Ligature-based font size independent OCR for Noori Nastalique writing style" **In 2017 1<sup>st</sup> International Workshop on Arabic Script Analysis and Recognition (ASAR)**, pp129-133.
- [44] Akram Q.U.A. and Hussain S. (2019) "Improving Urdu Recognition Using Character-Based Artistic Features of Nastalique Calligraphy" **IEEE Access**, **7**, pp8495-8507.
- [45] Naz S., Umar A.I., Ahmad R., Siddiqi I., Ahmed S.B., Razzak M.I., and Shafait F. (2017) "Urdu Nastaliq recognition using convolutional recursive deep learning" **Neuro computing**, **243**, pp80-87.
- [46] Mousavi S.M.H. and Lyashenko V. (2017) "Extracting Old Persian cuneiform font out of noisy images (handwritten or inscription)" **In 10<sup>th</sup> Iranian Conference on Machine Vision and Image Processing (MVIP)**, pp241-246.
- [47] Soheili M.R., Yousefi M.R., Kabir E., and Stricker D. (2017) "Merging clustering and classification results for whole book recognition" **In 2017 10<sup>th</sup> Iranian Conference on Machine Vision and Image Processing (MVIP)**, pp134-138.
- [48] AleAhmad A., Amiri H., Darrudi E., Rahgozar M., and Oroumchian F. (2009) "Hamshahri: A standard Persian text collection" **Knowledge-Based Systems**, **22**, **5**, pp382-387.

نفت نامہ

## مرتب بر اساس حروف الفبای فارسی

عنوان فارسی	عنوان انگلیسی
ارتفاع خط	Height Line
اردو	Urdu
الگوریتم ژنتیک کوانتومی	Genetic Quantum Algorithm
انفجار گرادیان	Vanishing Gradient
برخط	Online
برون خط	Offline
به روزرسانی خودکار	Automatic Update
بیش‌برازش	Overfitting
بینایی ماشین	Machine vision
پس انتشار خطای پویا	Dynamic Error Backpropagation
تایید امضا	Verify signatures
تزرکت	Tesseract
تشخیص نوری حروف	Optical Character Recognition (OCR)
چند زبانه	Multi-Language
دودویی کردن	Binarization
سلول حافظه	Memory Cell
سیگموئید	Sigmoid
شبکه عصبی	Neural Network (NN)
شبکه عصبی پرسپترون	Perceptron Neural Network (PNN)
شبکه عصبی کانولوشن	Convolutional Neural Network (CNN)
شبکه عصبی مکرر دوطرفه	Bidirectional Recurrent Neural Network (BRNN)
شبکه های عصبی مکرر	Recurrent Neural Network (RNN)
شکسته	Cursive
صفحه‌خوان	Page Reading
قطر عمودی خط پس زمینه	Vertical Diameter Baseline
کشیدگی حاصل از تراز کردن متون	Justify
ماتریس درهم‌ریختگی	Confusion Matrix

Support Vector Machines (SVM)	ماشین بردار پشتیبان
RTL Language	مدیریت زبان های راست به چپ
Sorting and Routing Ruto	مسیریابی
Open Source	منبع باز
Smoothing	نرم کردن
Deep Learning	یادگیری عمیق



## مرتب بر اساس حروف الفبای انگلیسی

عنوان فارسی	عنوان انگلیسی
به روزرسانی خودکار	Automatic Update
شبکه عصبی مکرر دوطرفه	Bidirectional Recurrent Neural Network (BRNN)
دودویی کردن	Binarization
ماتریس درهم‌ریختگی	Confusion Matrix
شبکه عصبی کانولوشن	Convolutional Neural Network (CNN)
شکسته	Cursive
یادگیری عمیق	Deep Learning
پس انتشار خطای پویا	Dynamic Error Backpropagation
الگوریتم ژنتیک کوانتومی	Genetic Quantum Algorithm
ارتفاع خط	Height Line
کشیدگی حاصل از تراز کردن متون	Justify
بینایی ماشین	Machine vision
سلول حافظه	Memory Cell
چند زبانه	Multi-Language
شبکه عصبی	Neural Network (NN)
برون خط	Offline
برخط	Online
منبع باز	Open Source
تشخیص نوری حروف	Optical Character Recognition (OCR)
بیش‌برازش	Overfitting
صفحه‌خوان	Page Reading
شبکه عصبی پرسپترون	Perceptron Neural Network (PNN)
شبکه عصبی مکرر	Recurrent Neural Network (RNN)
مدیریت زبان های راست به چپ	RTL Language
سیگموئید	Sigmoid
نرم کردن	Smoothing
مسیریابی	Sorting and Routing Ruto
ماشین بردار پشتیبان	Support Vector Machines (SVM)

Tesseract	تزرکت
Urdu	اردو
Vanishing Gradient	انفجار گرادیان
Verify signatures	تایید امضا
Vertical Diameter Baseline	قطر عمودی خط پس زمینه

# Abstract

OCR is an optical text recognition system that has been widely used in different applications. The main focus of existing OCR systems has been on Latin languages. In recent studies, OCR systems, used for language with cursive style have been introduced, followed by some challenges. In this thesis, we proposed an LSTM-based OCR system for the Persian language. We investigated some involved parameters in the proposed system. The proposed OCR system addressed the challenge false recognition of sub-word “LA” and “Zero-width non-breaking space”. In addition, we presented a preprocessing algorithm to remove “justification” using an image-processing-based technique. Moreover, in this thesis, a new dataset was generated, comprising of five million images with 8 popular Persian fonts and in 10 various font sizes. The experimental results with our generated dataset shows that the accuracy of the proposed OCR is increased by 2%, compared to the Persian OCR currently available in Tesseract. Moreover, the proposed system has an accuracy of 99.688 at the letter level.

**Keywords:** Persian OCR, Deep Learning, Image Processing, Justify, Printed.



Shahrood University of  
Technology

Faculty of Computer Engineering  
M.Sc. Thesis in Artificial Intelligence Engineering

# **A Printed Persian OCR System using Deep Learning**

By: Marziye Rahmati

Supervisors:

Dr. Mansoor Fateh

Dr. Mohsen Rezvani

Advisors:

Dr. Alireza Tajary

Dr. Vahid Abolghasemi

August 2019